Social Construction of Urban Space: Using LLMs to Identify Neighborhood Boundaries From Craigslist Ads

Adam Visokay[1] **Ruth Bagley**[4] **Ian Kennedy**[2] **Chris Hess**[3]
Kyle Crowder[1] **Rob Voigt**[4] **Denis Peskoff**[1,5]

Sociology
[1]University of Washington
[2]University of Illinois Chicago
[3]Kennesaw State University

Linguistics
[4]Northwestern University
Information
[5]University of California, Berkeley

Abstract

Rental listings offer a window into how urban space is socially constructed through language. We analyze Chicago Craigslist rental advertisements from 2018 to 2024 to examine how listing agents characterize neighborhoods, identifying mismatches between institutional boundaries and neighborhood claims. Through manual and large language model annotation, we classify unstructured listings from Craigslist according to their neighborhood. Further geospatial analysis reveals three distinct patterns: properties with conflicting neighborhood designations due to competing spatial definitions, border properties with valid claims to adjacent neighborhoods, and "reputation laundering" where listings claim association with distant, desirable neighborhoods. Through topic modeling, we identify patterns that correlate with spatial positioning: listings further from neighborhood centers emphasize different amenities than centrally-located units. Natural language processing techniques reveal how definitions of urban spaces are contested in ways that traditional methods overlook.

1 Contested Neighborhood Boundaries

Neighborhood location matters for a wide range of individual and collective outcomes (Sampson et al., 2002; Sharkey and Faber, 2014; Minh et al., 2017; Chyn and Katz, 2021). Beyond objective demographic characteristics, the subjective features of a neighborhood—its reputation, status, or stigma—shape resident satisfaction, place attachment, and overall well-being (Tran et al., 2020; Kullberg et al., 2010; Permentier et al., 2011; Otero et al., 2024). Neighborhood reputation also structures the economic value of property, patterns of investment, and the residential mobility that drives neighborhood stratification (Krysan and Crowder, 2017; Evans and Lee, 2020; Kirk, 2024; Korver-Glenn and Mayorga, 2024).

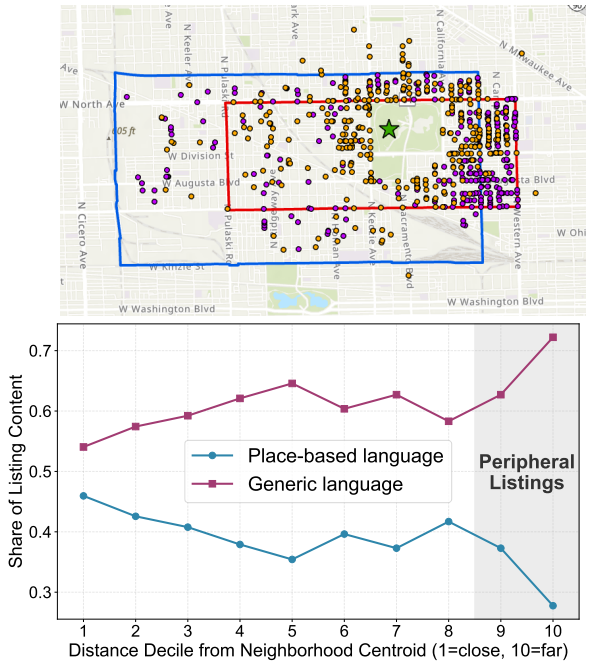

Figure 1: Extracting neighborhoods from unstructured rental listings with LLMs (RQ1, Section 3) provides insight into the social construction of space (*RQ2*, top)[1] and allows us to study how language changes relative to distance from neighborhood centers (*RQ3*, bottom).[2]

We define **neighborhood identity** not just as a boundary in space, but as the spatial area that corresponds with the patterns of social activity and perceptions of people living in the area. This conceptualization builds on a lineage of research utilizing user-generated content (UGC) to define urban space (Plangprasopchok and Lerman, 2009; Hollenstein and Purves, 2010; Hiippala et al., 2019;

[1] *Conflicting conceptions* of the Humboldt Park neighborhood according to the official City of Chicago limits (blue), Zillow's boundary (red) and rental advertisements (points). Orange points depict unit listings claiming to be Humboldt Park while purple points claim elsewhere. The green star denotes the LLM-defined social center of Humboldt park.

[2] Across Chicago, the share of place-based language decreases and generic "boilerplate" language increases in listings for units further from the social center of the neighborhood.

Proceedings of the 7th Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), pages 307–321

Brunila et al., 2023). Because this identity is socially constructed through interaction and language, it is inherently fluid and contested. Identifying a singular "true" neighborhood boundary is not our aim, but rather, mapping the contours of contestation: the systematic slippage between institutional maps and the spatial claims made by social actors.

Neighborhoods are socially constructed through interaction and language (Zelner, 2015; Hohle, 2023; Stuart et al., 2024), yet traditional urban research often relies on rigid administrative boundaries as a proxy for neighborhood identity. In practice, listing agents frequently navigate a tradeoff between geographic fidelity and reputational leverage, substituting symbolic identity for physical proximity when properties sit at the periphery of desirable areas. We characterize this behavior not as random noise, but as a systematic distortion of urban space. Still, an important question remains: do spatial claims that depart from institutional maps always represent strategic "reputation laundering" (Stuart et al., 2024), or might they reflect legitimate disagreements arising from the inherent ambiguity of socially constructed boundaries?

A related challenge for urban sociology has been the difficulty of observing and classifying such distortions at scale. Conventional data is blind to the strategic manipulation of spatial labels, and traditional NLP methods like string-matching struggle to distinguish between casual mentions and strong locational claims. Large Language Models (LLMs) provide a transformative opportunity to recover this latent social variable: claimed neighborhood identity. By evaluating LLMs on Chicago Craigslist rental advertisements (2018–2024), this work provides answers to the following Research Questions:

- **(RQ1) Measurement Viability:** Can zero-shot LLMs accurately identify specific neighborhood claims (vs. mere mentions) in unstructured rental listings compared to more traditional string-matching?
- **(RQ2) Social Location:** Where are neighborhoods actually located according to listing claims, and can we define a meaningful "social center" for each neighborhood?
- **(RQ3) Linguistic Substitution:** How does marketing language vary with spatial location? Specifically, does place-based language change as listings move farther from their claimed neighborhood center?

Section 2 describes our spatially-anchored corpus of Chicago listings. Section 3 evaluates LLM performance against string-matching baselines **(RQ1)**. Section 4 develops a framework for identifying neighborhood "social centers" **(RQ2)**. Section 5 and Section 5.2 analyze the semantic structure and statistical associations between unit language and spatial positioning **(RQ3)**. Finally, Section 6 contextualizes these findings within the broader field of computational social science.

2 Craigslist Housing Advertisements

We use data collected from Chicago Craigslist rental advertisements from 2018 to 2024 to identify listings with mismatches between the neighborhood containing the unit and the neighborhood claimed by the listing agent.[3]

2.1 Why Craigslist?

These rental listings offer a particularly valuable lens for examining how neighborhoods are socially constructed and contested because Craigslist's platform design creates an unusually unconstrained environment for spatial classification. Unlike many digital platforms that restrict users to predetermined administrative or commonly recognized neighborhood boundaries, Craigslist allows advertisers to freely designate any neighborhood label in their listings with unstructured text. This feature transforms rental advertisements into sites of boundary-making where the socio-spatial imaginary of the city becomes visible.

The neighborhood fields in these listings represent more than mere locational information—they reveal how real estate actors actively participate in constructing, reinforcing, or challenging existing spatial hierarchies. When landlords and property managers assign neighborhood labels to their listings, they engage in acts of spatial categorization that reflect both market strategies and internalized cognitive maps of urban space. These choices may align with officially recognized boundaries, reproducing understandings of place, or deliberately transgress established spatial categories to claim association with perceived higher-status areas.

[3]We have been actively collecting data in Chicago since 2018, providing a rich window into the discursive construction of urban space. It includes all available advertisements each day from December 2018 until June 2024 using the web-scraper `Helena` (Chasins et al., 2018; Hess and Chasins, 2022). Occasional changes to the architecture of the Craiglist website result in limited periods of data loss, the longest of which was from August 2019 to early October 2019.

Preceding computational analysis of Craigslist rental listings explores price distributions (Boeing and Waddell, 2017), neighborhood descriptions (Kennedy et al., 2020; Besbris et al., 2021), housing policy interventions (Boeing et al., 2021), exclusionary language (Stewart et al., 2023), affordability (Hess et al., 2023), and home security (Somashekhar et al., 2024). Holistically, research shows that rental listings on Craigslist align with and appear to reproduce social inequality. Recent work has begun to focus on the importance of neighborhood names and the places those names describe (Schachter et al., 2024), with specific focus on contested naming: when advertisements use neighborhood names that seem to diverge from the name most local residents would use for that space.

2.2 Why Chicago?

We focus on Chicago because it stands as a quintessential "city of neighborhoods," where locally recognized community areas hold exceptional cultural, economic, and social significance (Hwang and Sampson, 2014). Chicago is an ideal site for examining the social construction of urban space due to the high salience of its neighborhood boundaries. While the city maintains 77 officially recognized community areas, these rigid, non-overlapping boundaries often fail to represent the fluidity of neighborhood identity in reality. By using Chicago's stable institutional definitions as a point of comparison, we can more effectively identify and quantify how real estate actors use language to challenge or reinforce existing spatial hierarchies. This neighborhood orientation is so deeply embedded in Chicago's social fabric that it shapes how residents understand their place in the city, influences social networks, and structures daily mobility patterns (Kaysen, 2024).

By analyzing patterns in how these spatial designations align with or diverge from official boundaries, we can observe in real time the processes through which neighborhood reputations are reinforced or contested. The negotiated quality of these spatial boundaries becomes particularly visible when examining cases where advertisers claim association with neighborhoods other than those in which units are formally located, according to administrative boundaries. Such instances of spatial repositioning offer a window into the dynamics that shape how urban space is valued, categorized, and ultimately experienced by various stakeholders from listing agents to residents or potential tenants to municipal administrators.

3 Detecting Neighborhoods with LLMs

Online rental advertisements are generally unstructured and vary widely between listings. Distinguishing between which neighborhoods are mentioned in a listing from which neighborhood(s) the listing claims to be in is a nuanced task of great importance to social scientists interested in the social construction of urban space. This answer to "which neighborhood does this advertisement claim the unit to be in?" is not always obvious, even to a human annotator. For example,

> ...this fully restored **East Lakeview** property sits on a beautiful tree-lined street located in the heart of the popular **Wrigleyville** neighborhood ...

It is clear based on the language that this advertisement is not merely mentioning these neighborhoods, but staking a strong claim to being located in both. Wrigleyville is a Chicago neighborhood within the larger neighborhood of East Lakeview, so this claim is entirely coherent. However, reconciling such competing claims at scale is a principal challenge inherent to this particular task.

Furthermore, some listings contain mentions of several irrelevant neighborhoods, even dozens like this example:

> ...Disclaimer: Pricing, availability, and specials are subject to change at any time and without notice. HotSpot Rentals services the following neighborhoods: South Loop, Printers Row, Near North, River North, Gold Coast, West Loop, Fulton River District, West Loop Gate, The Loop, Streeterville, Lakeshore East, New East Side, Old Town, Medical District, University Village, Near North, River West, Lincoln Park South, Lakeview, Uptown, Ukrainian Village, Wicker Park, Edgewater, Ravenswood, Bronzeville, Logan Square.

Making the distinction between neighborhood mentions and strong claims that a unit is in a particular neighborhood is a nuance that large language models are particularly well-suited for compared to existing methods. To make this comparison, first we label the full corpus using a bespoke string-matching approach which serves as the baseline

"best practice" which we compare to the Language Model-based labeling. Both sets of labels are evaluated against a subset of 200 manually labeled neighborhood listings. These manual labels were produced by authors of this article.

3.1 Manual Labeling

The process for creating our validation set of "gold standard" labels considers three sources of information for each advertisement in the following order. First, if the title field includes a neighborhood name, that becomes the manual label for the neighborhood claim. Then if there is no claim in the title, we consider the body of the listing. This is the largest source of text in each advertisement, and also the most ambiguous with respect to identifying strong claims. When faced with multiple claims – as in the quote above – we take the first claim as the manual label. Finally, if neither the title nor body fields contain a neighborhood claim, we extract a neighborhood claim from the neighborhood field, if it exists. An advertisement only receives the 'unknown' label if there is not a strong claim in any of the three fields. We follow the same prioritization scheme in the string-matching and LM labeling procedures.

3.2 Data Pre-Processing

Before labeling we performed standard data preprocessing and de-duplication on the raw text. We removed common boilerplate text that appeared in virtually all Craigslist listings (such as "QR Code Link to This Post"), corrected Unicode translation errors to ensure consistent character rendering, and precisely mapped listings onto Zillow and City of Chicago neighborhood boundaries to confirm geographical validity. We also performed thorough de-duplication of listing entries by title and body text, retaining only the most recent version when multiple entries existed, as Craigslist saves edited listings as separate posts. This preparation distilled a clean dataset of 30,531 unique listings from an initial corpus of n=128,764 initial observations.

3.3 String-Match Labeling Neighborhoods

To determine which Chicago neighborhood a rental advertisement belongs to based on the text in the post we begin with a list of 192 distinct neighborhoods from Zillow, a real-estate marketplace company which provides commercial neighborhood lists for major US cities. Then, we manually construct a comprehensive list of regular expressions that can match the official name and its alternatives (e.g. *wriggleyville, wriglyville, wrigglyville, wrigley ville* for *wrigleyville*). These regex patterns are designed to be flexible with spaces and case-insensitive. Following the same protocol as the manual annotations, we use a function which tries to match neighborhoods in the listing title, body and dedicated neighborhood field. When multiple neighborhoods match, we select the label which appears earliest in the text.

3.4 Language Model Labeling Neighborhoods

We prompt GPT-4.1 mini as a high quality relative to cost option.[4] Table 1 contains our prompt.

```
BASE_PROMPT
Extract the Chicago neighborhood from the rental text.
Rules:
- NEVER use the address or zip code to determine the neighborhood
- Choose only explicitly stated neighborhood from possible responses
- If neighborhood is unclear mark it as [unknown]
- Format precision: "lakeview" but not "x, y, z" for "lakeview residence near x, y, z"
text: {body}
Possible responses: {zillow_list}
Return only:
label: [neighborhood]
```

Table 1: Prompt format for extracting Chicago neighborhood claims from rental listings.

We create three separate labeling workflows, one for each title, body and neighborhood field. We input the data independently to avoid leakage. A Zillow neighborhood list is provided to constrain possible responses to items in the list. In the case of unknowns, we prioritize the label from the title, then the body, then the neighborhood field. Only if all three are unknown is the final neighborhood claim labeled 'unknown'.

3.5 Label Post-Processing

While we engineer the prompt to provide structured output in the form `label:response`, the outputs still require post-processing. We implement a multi-stage post-processing pipeline to standardize the three LLM labels for each advertisement and assess model performance against manual validation. Through manual review we flag instances of minor spelling discrepancies (e.g. 'lakeview' instead of

[4] We compare a sample of regular vs mini vs nano, and other non-OpenAI options. Reasoning models are unnecessary for this extraction task.

Metric	String Match	GPT-4.1
Accuracy	0.79	**0.85**
Macro Average F1	0.62	**0.70**
Weighted Average F1	0.77	**0.85**

Table 2: GPT-4.1 outperforms the string match-based keyword search on the neighborhood claim labeling task. While the performance gain is marginal, the LLM labeling process was cheap, fast and can be scaled up.

'lake view', 'wrigglyville' instead of 'wrigleyville'). For text normalization, we construct a replacement dictionary to correct such instances. We implement the same priority-based claim-selection algorithm as is used for manual annotation and string matching. The post-processing system first examines the listing title label; if no neighborhood is identified, it analyzes the listing body label, and if still unsuccessful, it checks the neighborhood field label; and only then returns 'unknown' classification. This aligns with how potential renters process listing information, prioritizing the most prominent textual elements. When faced with compound neighborhood designations, we prioritize the first neighborhood when multiple were present, just as we do for manual labeling and string-matching. For outputs that contain separators (e.g. 'uptown, ravenswood' or 'avondale/logan square'), we extract the first neighborhood claim before a separator.

3.6 Evaluation of Labeling Task

We compare the string-match and LLM labels to the manual labels in the 200-item validation set. Performance metrics are shown in Table 2. While accuracy scores are comparable (GPT-4 mini's 85% is marginally better than the string matching's 79%), the disagreements largely reflect the inherent ambiguity of neighborhood classification – a challenge even human annotators face when listings claim multiple neighborhoods. A notable advantage of the LLM approach is its efficiency and scalability. Unlike string matching on keywords, which requires extensive manual configuration of locale specific patterns, the pre-trained LLM can work with a zero-shot prompt, making it adaptable.

4 Geospatial Analysis of Neighborhoods

Although Craiglists rental listings do not comprehensively show every available property in the Chicago area, there is still substantial coverage across the city, as seen in Figure 2. As a result, we take these rental listings to be a reasonably representative sampling, which we use to identify neighborhoods as they might be conceived of by the people, or at least the realtors, of Chicago.

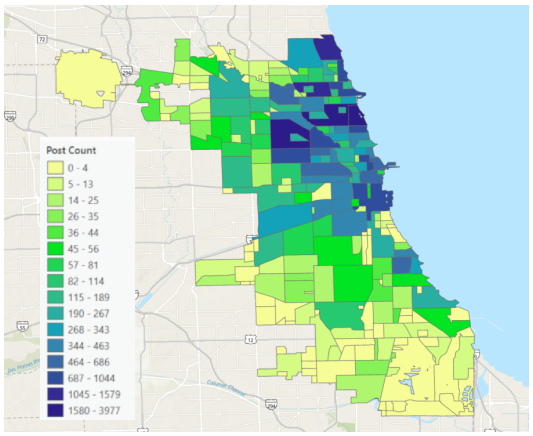

Figure 2: The city of Chicago and its constituent neighborhoods, as defined by Zillow, colored to represent the number of posts for properties located in each area

Neighborhood boundaries are nebulous and hard to define; even the City of Chicago and Zillow, both with access to a great deal of data/information, have developed quite different maps of neighborhood boundaries. Figure 1 shows that while the official Zillow and city boundaries of Humboldt Park include most of the posts claiming to be in that neighborhood, none of the borders are the same. In addition, borders between neighborhoods are likely less rigid than an official boundary might suggest; despite being located within the formal boundaries of Humboldt Park, there are a number of listings, mainly around the edges, claiming to be part of other nearby neighborhoods, primarily Ukrainian Village, Wicker Park, or West Town.

Using the rental listings, we conceptualize the "social center" of the neighborhood as the centroid of all listings that claim to be located within that neighborhood. We use `geopandas` to identify the centroid of all posts claiming to be in the same neighborhood using the latitude and longitude coordinates of all property locations. The map of Humboldt Park suggests this is an effective method, because the social center (represented by the green star) is in fact located in the eponymous park which is considered to be the heart of the neighborhood.

However, not all posts claiming to be in a given neighborhood may have the same strength to their claim; advertisements for apartments at the center of a popular neighborhood like Logan Square likely have different characteristics than posts for units on the fringes of the neighborhood. The posts on the

fringes might even be trying to seem more desirable by claiming to be in a more popular neighborhood, whereas it might be more of an objective description of location for a property actually located on Logan Square. We conceptualize this by identifying how far from the social center a post is, using three metrics for distance: 1) Raw Distance: Euclidean distance between the latitude and longitude coordinates of the post and that of the neighborhood centroid; 2) Relative Distance: raw distance for all posts in the neighborhood projected onto a [0,1] interval using min-max scaling; 3) Z-scored Distance: z-scored distance for posts claiming to be in the same neighborhood

We use these measures to distinguish a specific subset of the posts: those on the periphery of a neighborhood. We define peripheral posts as those that are in the furthest 20% of posts from the centroid for a given neighborhood (for any neighborhood labels with at least 5 posts).

4.1 Comparison of Neighborhood Boundaries

In order to get a more concrete understanding of different conceptions of neighborhoods, we identify two more neighborhoods to explore more in depth: Logan Square and Pilsen.

Logan Square is a well-known neighborhood in Chicago, and also a neighborhood label where a large number (n=2495) of listings claim to be. We can see in Figure 4 that the City of Chicago boundaries for Logan Square contain primarily postings claiming to be in the neighborhood (orange points) without many posts claiming to be elsewhere (purple). The Zillow boundaries do encompass Logan Square claims that the Chicago boundaries do not, but the areas excluded by Chicago also have a higher concentration of claims of being in other neighborhoods. In addition, there are a number of listings claiming to be in Logan Square that are outside both formal boundaries–these posts would certainly be part of the 'peripheral' posts we defined earlier. This kind of posting merits further exploration to better understand what is happening in listings that claim to be part of a neighborhood when that might be more likely to be contested.

Although the blue City of Chicago boundaries appear to be a better approximation for Logan Square in Figure 4, this is not the case for every neighborhood. Pilsen, shown in Figure 3, is another well-known neighborhood within Chicago, but it is not included as a distinct neighborhood by the City of Chicago. However, even though Zillow

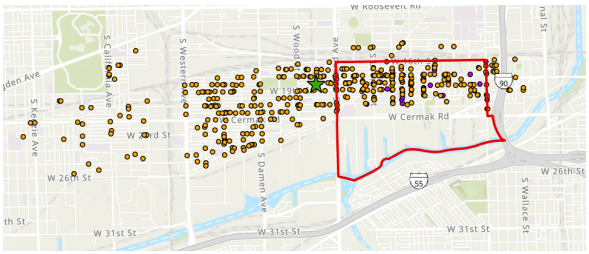

Figure 3: Contested boundaries for Pilsen neighborhood according to Zillow's definition (red). Orange points depict listings claiming to be in Pilsen, purple points are listings claiming to be elsewhere). The green star depicts the Pilsen neighborhood centroid. This is an example of what we call *border stretching*, demonstrating how static boundary systems may not capture the neighborhood identities as experienced by the people living in them.

does include Pilsen as an area, it also does not seem to define it in the same way as Craigslist advertisements. Many of the posts claiming to be in Pilsen appear in a clustered way outside of the Zillow boundaries, and even the centroid of the Craigslist-defined neighborhood is outside the Zillow bounds. This could represent a developing neighborhood identity, which may not have been as strong at the time of the map creation, and supports the need for a more dynamic model of neighborhoods.

5 Content Analysis of Rental Listing Text

We use the `tomotopy` Python package to identify latent topics in the content posted in Craigslist rental advertisements. `tomotopy` uses Gibbs-sampling and is based on the LDA approach described in (Blei et al., 2003) and (Newman et al., 2009). We prepared the text corpus by combining listing titles and body text from the Craigslist dataset. Standard preprocessing was applied using NLTK – lowercasing, removing alphanumeric characters, tokenization, custom stopword filtering, and lemmatization. We remove common real estate jargon that would otherwise dominate the topic distributions without providing meaningful differentiation between the content in different listing types. For the LDA implementation we selected hyperparameters that allow for moderate document sparsity ($\alpha = 0.1$) and greater topic-word concentration ($\eta = 0.01$). We trained for 100 iterations and tried $k = 5, 7$ and 10 topics which returned coherence scores of `0.7344`, `0.7425` and `0.7509`, respectively. After manual review we decided to focus on the $k = 7$ results for our remaining analysis as these had clearer

Topics	Common Words
1. Furnished Short-Term *3.6% of corpus*	lease, home, month, furnished, mo, amenity, community, view, apartment, offer, blueground, access, neighborhood, enjoy, stay, loop
2. Rental Terms *25.9% of corpus*	fee, lease, included, tenant, credit, street, pay, deposit, application, large, move, heat, per, gas, security, pet, utility
3. Property Search *1.1% of corpus*	apartment, rental, property, place, hill, cheap, grove, find, apt, height, center, agency, search, local, credit
4. Spanish Language *2.1% of corpus*	apartment, rental, property, place, hill, cheap, alquilar, propiedades, buscar, alquileres, height, search, agency, google
5. Amenities *35.0% of corpus*	floor, central, new, appliance, dishwasher, heat, large, space, air, stainless, feature, modern, steel, living, cat, closet
6. Listing Conditions *13.3% of corpus*	subject, unit, change, price, center, property, onsite, hour, availability, special, dog, pricing, studio, fitness, housing, amenity
7. Contact Information *18.9% of corpus*	contact, info, show, click, feature, view, call, id, renovated, , included, closet

Table 3: Topic interpretations based LDA topic modeling on Chicago Craigslist rental listings.

separation than the $k = 5$ topics and are more interpretable than the $k = 10$ results.

5.1 Topic Interpretations

We identify seven distinct topics in the Craigslist advertisements, shown in Table 3. These topics reveal distinct patterns in how rental listings are framed, which have important implications for understanding the conception of neighborhood reputation and representation in the online rental market.

Topic 1 focuses on furnished short-term rentals, highlighting amenities and comfort with words like “furnished," “month," and “stay," suggesting a market segment catering to temporary residents seeking turnkey living situations. Topic 2 centers on rental requirements and financial considerations, with terms like “fee," “credit," “deposit," and “application," reflecting the administrative and financial aspects of renting. Topics 3 and 4 both relate to property search, with Topic 4 specifically including Spanish-language terms like “alquilar" and “propiedades," indicating efforts to reach Spanish-speaking audiences in Chicago’s rental market. Topic 5, the most prevalent across the corpus at approximately 35% of document content, focuses on apartment features and amenities such as “appliance," “stainless," “modern," and “dishwasher," underscoring the prevalence of interior quality in marketing rental properties. Topic 6 addresses listing conditions and availability, featuring terms related to pricing, special offers, and property policies. Finally, Topic 7 concentrates on contact information and viewing arrangements, with words like “contact," “show," “click," and “schedule," facilitating the connection between potential renters and property managers. The distribution of these topics across advertisements reveals how Chicago’s rental market is presented online, with physical features and financial terms dominating the discourse.

5.2 Regression Analysis

We estimate the relationship between listing characteristics and proximity to the social center of an associated neighborhood using OLS regression with relative distance to neighborhood center as our primary dependent variable. Our model includes unit characteristics (bedrooms, bathrooms, rent, and square footage) and the topic proportions identified in our LDA analysis. For a full table of regression outputs see Table 5 in the Appendix.

This analysis reveals several patterns in spatial representations. Advertisements with more bedrooms/bathrooms are associated with being further from neighborhood centers, while higher-priced listings are closer to their respective social centers. Most notably, listings with higher proportions of Topic 3 (Property Search) content exhibit increased relative distance from the center of the neighborhood (+0.20, $p<0.001$). Conversely, listings emphasizing apartment amenities (Topic 5) are associated with being closer to the centroid (-0.03, $p<0.01$). These findings indicate that misrepresentation is not random but correlates with specific marketing approaches in rental listings.

6 Results

Our analysis of over 30,000 unique Chicago rental listings reveals that neighborhood boundaries are actively renegotiated through strategic linguistic claims. By establishing a “social center” for neighborhoods based on the density of textual claims rather than rigid municipal boundaries, we demonstrate how agents navigate the tradeoff between geographic reality and reputational leverage. The following sections detail our findings in relation to our primary research questions.

6.1 Labeling Viability (RQ1)

We find that Large Language Models are highly effective for identifying neighborhood claims within unstructured text, outperforming traditional string-matching techniques. While the accuracy gain is incremental (85% for GPT-4.1 mini vs. 79% for string-matching), the LLM approach can scale beyond Chicago to other urban contexts without requiring reconfiguration or the same level of local real estate knowledge. Our labeling system prioritizes precision by selecting the single strongest neighborhood claim, addressing the inherent ambiguity found in listings that mention multiple areas.

6.2 Defining Social Centers (RQ2)

Our geospatial analysis reveals that the "social center" of a neighborhood—defined as the centroid of all property listings claiming that identity—often aligns with local landmarks, such as the eponymous park in Humboldt Park. However, these Craigslist-defined centers frequently diverge from institutional boundaries. We identify significant variation in representation: neighborhoods like Lake View are heavily overrepresented in claims relative to their Zillow-defined footprints, while areas such as Englewood and North Austin are significantly underrepresented, suggesting a lack of strong neighborhood identity or the presence of territorial stigma in the rental market.

6.3 Linguistic Substitution (RQ3)

Our analysis characterizes spatial misrepresentation not as random market noise, but as a predictable distortion of urban space. We identify three distinct patterns of spatial claim discrepancies: (1) *Conflicting Conceptions*, where stakeholders disagree on boundaries (e.g. Humboldt Park in Figure 1); (2) *Border Stretching*, where listings claim adjacent, plausible neighborhoods (e.g. Pilsen in Figure 3); and (3) *Reputation Laundering*, where properties or peripheral areas claim association with distant, desirable neighborhoods.

To identify these patterns systematically, we operationalize "peripheral claims" as those located in the furthest 20% from a neighborhood's social centroid. By comparing these LLM-labeled claims against institutional Zillow boundaries, we find evidence of systematic over and underrepresentation. Specifically, Lake View is significantly overrepresented–claimed more frequently than geographic distributions would predict–while neighborhoods such as Englewood, North Austin, and Hanson Park are significantly underrepresented. This pattern is consistent with reputation laundering: agents appear to distance properties from stigmatized neighborhood names while strategically claiming higher-status alternatives.

This substitution is statistically observable through a compensatory language pattern. Regression results show that as properties move further from the neighborhood social center, listing agents substitute symbolic identity for physical proximity. For instance, generic property search language (Topic 3) is positively associated with relative distance from centroid ($+0.20, p < 0.001$), while specific interior amenity language (Topic 5) is negatively associated ($-0.03, p < 0.01$). Further, compared to central listings, the language used in peripheral listings shifts from location specific, non-portable amenities (Topics 1,5) toward more abstract and generic property-search language (Topics 2,3,4,6,7), a pattern illustrated in Figure 1.

7 Implications Beyond Sociology

The use of LLMs has exploded in recent years, and they can be seen by the general public as a simple, reliable solution to many routine problems. However, it remains an open question how powerful they may be in interdisciplinary research (Ziems et al., 2024). In order to better understand the impact of NLP on a broader scale and help address a specific question in the field of Urban Sociology, we demonstrate the effectiveness of LLMs at a notoriously difficult task: identifying where rental listings claim to be located on a large scale, in order to inform our understanding of processes of social construction of urban space.

However, although the impact of language models on this task may be transformative in the ability to quickly expand the scope of analysis, the quantifiable improvement on simple algorithms designed by experts is perhaps more incremental. In addition, the LLM labeling was certainly not perfectly accurate, suggesting that there is still room for improvement in large language models that might not be captured by tests and benchmarks developed solely within the field of NLP, and that interdisciplinary collaboration could lead to improvements both in NLP methodology and in making research questions and analyses in a broad range of fields more tractable and scalable.

8 Limitations

Our analysis of Craigslist rental listings provides valuable insights into neighborhood claims and social construction, but several important limitations should be acknowledged. The process of collecting data from Craigslist presents inherent challenges regarding post uniqueness and identification. Despite our deduplication efforts, the platform's structure makes it difficult to definitively determine which posts represent truly unique listings versus slightly modified versions of the same property.

While our dataset offers substantial coverage across Chicago neighborhoods, it cannot claim to be fully representative of the entire rental market. Craigslist represents just one segment of available rental properties, potentially skewing toward certain price points or property types. Additionally, our data may over represent certain management companies and realtors who post frequently on the platform, as opposed to "mom and pop" owners. These high-volume posters are more likely to use standardized language across multiple listings, which may introduce uniformity in how neighborhoods are described that doesn't reflect broader market patterns.

A fundamental challenge in this research is the absence of an authoritative catalog of Chicago's neighborhoods. As we argue, such a catalog is conceptually impossible. For practical purposes, we relied on Zillow's neighborhood boundaries as our primary reference, but these designations do conflict with local understandings. For example, questions arise about whether "West Loop" constitutes its own neighborhood distinct from "West Loop Gate," or whether "East Rogers Park" should be considered separate from "Rogers Park." These ambiguities reflect the socially constructed nature of neighborhoods themselves. Also, many units along Lake Michigan have a view of the water, and therefore advertise "views of the lake" or "lake views" which can be impossible to distinguish from listings claiming to be a "lake view" without considering the corresponding latitude and longitude coordinates.

Furthermore, assigning a single neighborhood label to each listing proved challenging when advertisements contained multiple, sometimes competing neighborhood claims. Our hierarchical labeling approach (prioritizing title, then body, then neighborhood field) necessarily simplifies what can be complex spatial positioning strategies employed by listing agents. A rental advertisement might strategically claim association with multiple neighborhoods simultaneously, a nuance our single-label framework cannot fully capture. For instance, a listing located on the border between Logan Square and Humboldt Park might leverage both neighborhood identities depending on the perceived audience and market conditions.

These limitations highlight the inherent complexity of studying socially constructed spatial boundaries through digital traces and suggest opportunities for future research employing more nuanced approaches to neighborhood identification and classification.

Ethics Statement

Deciding how to approach this analysis is a non-trivial decision and following the extensive work in sociology and economics is important. We spoke experts in both fields in scoping and executing this project.

We collect data from Craigslist which, in some cases, contains specific information about individual posters. Craigslist is a public forum whose housing section should not contain much information irrelevant to the housing ads themselves. We do not release these data publicly per the non-commercial use terms of service and ensure no PII appear in any of our writing, results or figures.

Another limitation when working with pre-trained foundation models like GPT-4 is a lack of reproducibility, as we do not have access to the training data or the weights. In the interest of reproducibility, we keep annotation costs under $100.

We utilized multiple Generative AI tools (OpenAI's GPT-4/5 and Anthropic's Claude 4.5 Opus/Claude Code) in the production of this manuscript, in the following ways: 1) producing computer code for data cleaning and analysis and 2) iteratively improving the concision and clarity of the writing. We have carefully reviewed all aspects of the manuscript for accuracy and coherence. All scientific insights, analysis and interpretation of data and scientific conclusions are made solely by the authors.

9 Acknowledgements

We thank Drew Messamore, Diag Davenport, and Michael Reher for providing valuable suggestions on an earlier version of this manuscript. Adam gratefully acknowledges the resources provided

by the International Max Planck Research School for Population, Health and Data Science (IMPRS-PHDS). We are also grateful to the University of Washington's Department of Sociology writing workshop for their comments and feedback and to the University of California, Berkeley Bellwether Postdoctoral program.

References

Kenan Alkiek, Anna Wegmann, Jian Zhu, and David Jurgens. 2025. Neurobiber: Fast and interpretable stylistic feature extraction. *arXiv preprint arXiv:2502.18590*.

Abdulkareem Alsudais. 2020. Incorrect data in the widely used inside airbnb dataset. *Decision Support Systems*, 141:113453.

Taylor Berg-Kirkpatrick, David Burkett, and Dan Klein. 2012. An empirical investigation of statistical significance in nlp. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 995–1005.

Max Besbris, Ariela Schachter, and John Kuk. 2021. The unequal availability of rental housing information across neighborhoods. *Demography*, 58(4):1197–1221.

Douglas Biber. 1991. *Variation across speech and writing*. Cambridge University Press.

Abeba Birhane, Atoosa Kasirzadeh, David Leslie, and Sandra Wachter. 2023. Science in the age of large language models. *Nature Reviews Physics*, pages 1–4.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Geoff Boeing and Paul Waddell. 2017. New insights into rental housing markets across the united states: Web scraping and analyzing craigslist rental listings. *Journal of Planning Education and Research*, 37(4):457–476.

Geoff Boeing, Max Besbris, Ariela Schachter, and John Kuk. 2021. Housing search in the age of big data: smarter cities or the same old blind spots? *Housing Policy Debate*, 31(1):112–126.

Dallas Card, Peter Henderson, Urvashi Khandelwal, Robin Jia, Kyle Mahowald, and Dan Jurafsky. 2020. With little power comes great responsibility. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9263–9274, Online. Association for Computational Linguistics.

Sarah E. Chasins, Maria Mueller, and Rastislav Bodik. 2018. Rousillon: Scraping distributed hierarchical web data. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pages 963–975.

Eric Chyn and Lawrence F. Katz. 2021. Neighborhoods matter: Assessing the evidence for place effects. *Journal of Economic Perspectives*, 35(4):197–222.

Megan Evans and Barrett A. Lee. 2020. Neighborhood reputations as symbolic and stratifying mechanisms in the urban hierarchy. *Sociology Compass*, 14(10):1–15.

George Galster and Erin Godfrey. 2005. By words and deeds: Racial steering by real estate agents in the us in 2000. *Journal of the American Planning Association*, 71(3):251–268.

Edward L. Glaeser, Michael Luca, and Erica Moszkowski. 2020. Gentrification and neighborhood change: Evidence from yelp. *National Bureau of Economic Research*.

Igal Hendel, Aviv Nevo, and François Ortalo-Magné. 2009. The relative performance of real estate marketing platforms: Mls versus fsbomadison. com. *American Economic Review*, 99(5):1878–1898.

Chris Hess and Sarah E. Chasins. 2022. Informing housing policy through web automation: Lessons for designing programming tools for domain experts. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–9.

Chris Hess, Rebecca J. Walter, Ian Kennedy, Arthur Acolin, Alex Ramiller, and Kyle Crowder. 2023. Segmented information, segregated outcomes: Housing affordability and neighborhood representation on a voucher-focused online housing platform and three mainstream alternatives. *Housing Policy Debate*, 33(6):1511–1535.

Randolph Hohle. 2023. Rusty gardens: stigma and the making of a new place reputation in buffalo, new york. *American Journal of Cultural Sociology*, 11(2):193–219.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

Biqing Huang and Ronald Rutherford. 2007. Who you going to call? performance of realtors and non-realtors in a mls setting. *The Journal of Real Estate Finance and Economics*, 35:77–93.

Jackelyn Hwang and Robert J. Sampson. 2014. Divergent pathways of gentrification: Racial inequality and the social order of renewal in chicago neighborhoods. *American Sociological Review*, 79(4):726–751.

Ronda Kaysen. 2024. Apartment rent renewal rates are rising. *The New York Times*.

Ian Kennedy, Chris Hess, Amandalynne Paullada, and Sarah Chasins. 2020. Racialized discourse in seattle rental ad texts. *Social Forces*, 99(4):1432–1456.

Richard Kirk. 2024. Legitimising displacement: Academic discourse, territorial stigmatisation and gentrification. *Urban Studies*, 61(13):2492–2512.

Elizabeth Korver-Glenn and Sarah Mayorga. 2024. *A good reputation: how residents fight for an American barrio*. University of Chicago Press.

Maria Krysan and Kyle Crowder. 2017. *Cycle of Segregation: Social Processes and Residential Stratification*. Russell Sage Foundation, New York.

Agneta Kullberg, Toomas Timpka, Tommy Svensson, Nadine Karlsson, and Kent Lindqvist. 2010. Does the perceived neighborhood reputation contribute to neighborhood differences in social trust and residential wellbeing? *Journal of Community Psychology*, 38(5):591–606.

Anita Minh, Nazeem Muhajarine, Magdalena Janus, Marni Brownell, and Martin Guhn. 2017. A review of neighborhood effects and early child development: How, where, and for whom, do neighborhoods matter? *Health & Place*, 46:155–174.

Burt L. Monroe, Michael P. Colaresi, and Kevin M. Quinn. 2017. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403.

David Newman, Arthur Asuncion, Padhraic Smyth, and Max Welling. 2009. Distributed algorithms for topic models. *Journal of Machine Learning Research*, 10:1801–1828.

NYC Office of Special Enforcement. 2023. Short-term rental registration and verification by booking services.

Gabriel Otero, Quentin Ramond, María Luisa Méndez, Rafael Carranza, Felipe Link, and Javier Ruiz-Tagle. 2024. The damages of stigma, the benefits of prestige: Examining the consequences of perceived residential reputations on neighbourhood attachment. *Urban Studies*, 61(3):462–494.

Jeffrey Nathaniel Parker. 2019. *That Kind of Neighborhood: Creating, Contesting, and Commodifying Place Reputation*. Ph.D. thesis, The University of Chicago.

M. Permentier, G. Bolt, and M. Van Ham. 2011. Determinants of neighbourhood satisfaction and perception of neighbourhood reputation. *Urban Studies*, 48(5):977–996.

Kirsten Robertson and Antony Doig. 2010a. An empirical investigation of variations in real-estate marketing language over a market cycle. *Housing, Theory and Society*, 27(2):178–189.

Kirsten Robertson and Antony Doig. 2010b. An empirical investigation of variations in real-estate marketing language over a market cycle. *Housing, Theory and Society*, 27(2):178–189.

Robert J. Sampson, Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology*, 28(1):443–478.

Ariela Schachter, John Kuk, Max Besbris, and Lydia Ho. 2024. What's in a name? place misrepresentation and neighbourhood stigma in the online rental market. *Urban Studies*, 61(16):3050–3068.

Youngme Seo, JongHo Im, and Brian Mikelbank. 2020. Does the written word matter? the role of uncovering and utilizing information from written comments in housing ads. *Journal of Housing Research*, 29(2):133–155.

Patrick Sharkey and Jacob W. Faber. 2014. Where, when, why, and for whom do residential contexts matter? moving away from the dichotomous understanding of neighborhood effects. *Annual Review of Sociology*, 40(1):559–579.

Mahesh Somashekhar, Chris Hess, Ian Kennedy, and Kyle Crowder. 2024. How do real estate actors advertise in mixed-income neighborhoods? the importance of home security. *Socius*, 10:23780231241260253.

Remy Stewart, Chris Hess, Ian Kennedy, and Kyle Crowder. 2023. Move-in fees as a residential sorting mechanism within online rental markets. *Cityscape (Washington, DC)*, 25(1):239.

Forrest Stuart, Charles R. Collins, Bocar Wade, Rebecca D. Gleit, and Caylin Louis Moore. 2024. Where do neighbourhood reputations come from? analysing chicago community areas using a systematic neighbourhood reputation score, 1985–2020. *Urban Studies*, pages 00420980241297088.

Emma Tran, Kim Blankenship, Shannon Whittaker, Alana Rosenberg, Penelope Schlesinger, Trace Kershaw, and Danya Keene. 2020. My neighborhood has a good reputation: Associations between spatial stigma and health. *Health & Place*, 64:102392.

United States Census Bureau, the. 2023. American community survey 5-year data (2009-2021).

Mohammadzaman Zamani and H. Andrew Schwartz. 2017. Using Twitter language to predict the real estate market. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 28–33, Valencia, Spain. Association for Computational Linguistics.

Sarah Zelner. 2015. The perpetuation of neighborhood reputation: An interactionist approach. *Symbolic Interaction*, 38(4):575–593.

Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. Can large language models transform computational social science? *Computational Linguistics*, 50(1):237–291.

Livia Hollenstein and Ross Purves. 2010. Exploring place through user-generated content: Using Flickr to describe city cores. *Journal of Spatial Information Science*, 1:1–18.

Anon Plangprasopchok and Kristina Lerman. 2009. Constructing folksonomies from user-specified relations on Flickr. In *Proceedings of the 18th International Conference on World Wide Web*, pages 781–790.

Mikael Brunila, Jack LaViolette, Sky CH-Wang, Priyanka Verma, Clara Féré, and Grant McKenzie. 2023. Toward a Critical Toponymy Framework for Named Entity Recognition: A Case Study of Airbnb in New York City. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 4676–4695, Singapore. Association for Computational Linguistics.

Tuomo Hiippala, Anna Hausmann, Henrikki Tenkanen, and Tuuli Toivonen. 2019. Exploring the Linguistic Landscape of Geotagged Social Media Content in Urban Environments. *Digital Scholarship in the Humanities*, 34(2):290–309.

A Full Performance Metrics for Neighborhood Claim Validation

A.1 Topic Modeling Results

Working with more than 35,000 unique and often verbose advertisements, we fit a topic model on $k = 25$ topics to identify main themes in the rental listings. We do not provide interpretations of the topic, but show the common words in Table 6.

	String Match			GPT-4 Mini			
Neighborhood	**Prec.**	**Recall**	**F1**	**Prec.**	**Recall**	**F1**	**Support**
the loop	0.12	0.50	0.20	1.00	1.00	1.00	2
rogers park	1.00	1.00	1.00	0.75	1.00	0.86	3
lake view	0.63	0.86	0.73	0.94	0.73	0.82	22
ranch triangle	–	0.00	0.00	–	0.00	0.00	1
lincoln park	0.90	1.00	0.95	0.94	0.89	0.91	18
fulton river district	–	0.00	0.00	1.00	1.00	1.00	1
south loop	1.00	1.00	1.00	1.00	1.00	1.00	3
park west	–	0.00	0.00	–	0.00	0.00	1
west town	1.00	1.00	1.00	1.00	1.00	1.00	1
logan square	0.88	0.88	0.88	0.89	1.00	0.94	8
lake view east	–	0.00	0.00	0.90	0.90	0.90	10
university village - little italy	–	0.00	0.00	–	0.00	0.00	2
uptown	1.00	0.83	0.91	1.00	0.83	0.91	6
wicker park	1.00	1.00	1.00	1.00	1.00	1.00	4
portage park	1.00	1.00	1.00	1.00	1.00	1.00	2
old town	1.00	0.83	0.91	1.00	0.83	0.91	6
avondale	–	0.00	0.00	–	0.00	0.00	1
west loop gate	–	0.00	0.00	1.00	1.00	1.00	4
streeterville	1.00	1.00	1.00	1.00	1.00	1.00	2
ravenswood	1.00	0.80	0.89	1.00	0.80	0.89	5
wrigleyville	1.00	0.67	0.80	1.00	1.00	1.00	3
river north	0.83	0.71	0.77	0.86	0.86	0.86	7
buena park	1.00	1.00	1.00	0.75	1.00	0.86	3
bucktown	1.00	1.00	1.00	1.00	1.00	1.00	7
kenwood	–	0.00	0.00	1.00	1.00	1.00	1
old irving park	1.00	1.00	1.00	1.00	1.00	1.00	1
edgewater	1.00	1.00	1.00	1.00	1.00	1.00	5
pilsen	1.00	1.00	1.00	1.00	1.00	1.00	7
garfield ridge	1.00	1.00	1.00	1.00	1.00	1.00	1
humboldt park	1.00	1.00	1.00	1.00	1.00	1.00	1
andersonville	1.00	1.00	1.00	1.00	0.67	0.80	3
west rogers park	1.00	1.00	1.00	1.00	1.00	1.00	1
ukrainian village	1.00	1.00	1.00	1.00	1.00	1.00	1
gold coast	0.75	1.00	0.86	0.75	1.00	0.86	3
hyde park	1.00	1.00	1.00	1.00	1.00	1.00	1
roscoe village	1.00	1.00	1.00	0.33	1.00	0.50	2
east garfield park	1.00	1.00	1.00	1.00	1.00	1.00	1
albany park	1.00	1.00	1.00	1.00	1.00	1.00	1
lincoln square	0.33	1.00	0.50	0.00	0.00	0.00	1
south shore	1.00	1.00	1.00	1.00	1.00	1.00	1
hermosa	1.00	1.00	1.00	0.50	1.00	0.67	1
ravenswood manor	–	0.00	0.00	–	0.00	0.00	1
unknown	0.78	0.88	0.82	0.50	0.50	0.50	8
Average Metrics							**Total: 163**
Accuracy		0.79			0.85		
Macro Avg	0.76	0.74	0.62	0.78	0.81	0.70	
Weighted Avg	0.88	0.79	0.77	0.91	0.85	0.85	

Table 4: Detailed performance metrics by neighborhood for string matching and GPT-4 Mini classification methods. The Support column indicates the number of test samples for each neighborhood. Dashes indicate cases where precision could not be calculated due to zero predictions.

Variable	Coef.	Std. Err.	t	P>|t|
Intercept	0.10	0.01	10.40	0.00
Bedrooms	0.02	0.00	8.83	0.00
Bathrooms	0.02	0.00	6.25	0.00
Rent	0.00	0.00	-14.56	0.00
Square Footage	0.00	0.00	2.17	0.03
Topic 2 Rental	0.03	0.01	3.40	0.00
Topic 3 Property	0.20	0.01	14.99	0.00
Topic 4 Spanish	0.10	0.01	9.42	0.00
Topic 5 Amenities	-0.03	0.01	-2.98	0.00
Topic 6 Condition	0.02	0.01	1.86	0.06
Topic 7 Contact	0.05	0.01	5.56	0.00

Table 5: Regression results demonstrating how relative distance from the center of the neighborhood associates with rental unit characteristics.

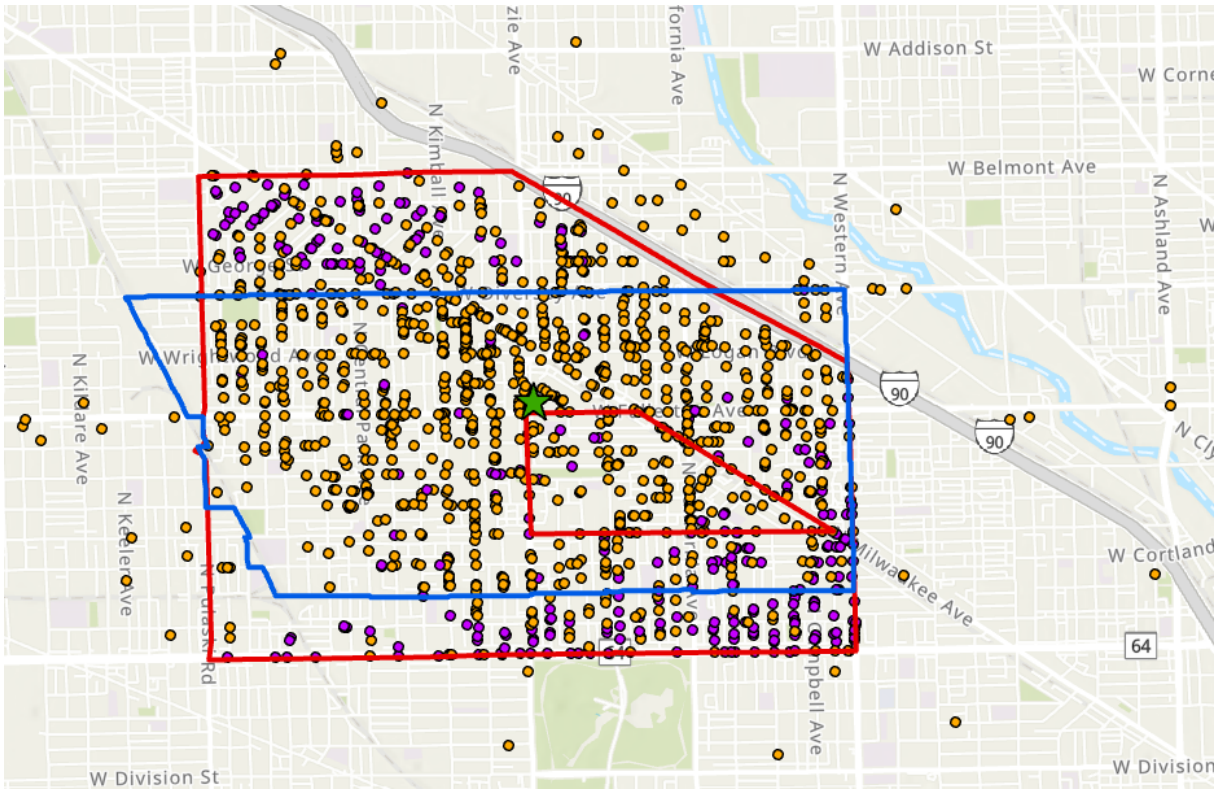

Figure 4: Contested boundaries for Logan Square neighborhood according to the official City of Chicago limits (blue), Zillow's definition (red), and claims in Craigslist rental listings (orange points for listings claiming to be in the neighborhood, purple if claiming to be elsewhere); the center of our Craigslist-defined neighborhood is represented by a green star. Logan Square represents a popular neighborhood with many postings and a number of nearby listings claiming to be within the neighborhood.

Table 2: Topics for $k = 25$

Topic	Common Words
Topic 1	hyde, property, si, terrace, la, estos, con, mac, elli, street
Topic 2	large, space, floor, living, dining, closet, bedroom, heat, storage, central
Topic 3	contact, call, schedule, photo, tour, please, unit, info, show, actual
Topic 4	space, walk, home, lease, great, living, street, one, loft, large
Topic 5	real, estate, star, text, tour, community, view, lounge, apartment, finish
Topic 6	lease, mo, blueground, month, furnished, amenity, stay, offer, view, access
Topic 7	cat, feature, floor, call, dishwasher, one, fee, lakeview, n, dog
Topic 8	modern, heat, property, central, bus, appliance, call, gas, cat, air
Topic 9	apartment, rental, lake, property, place, hill, alquilar, cheap, new, grove
Topic 10	unit, special, availability, subject, dog, weight, onsite, please, price, various
Topic 11	view, center, lake, amenity, free, onsite, community, michigan, call, window
Topic 12	fee, credit, deposit, tenant, new, security, included, move, pay, pet
Topic 13	hyde, apartment, regent, horas, onsite, market, la, est, e, museum
Topic 14	apartment, rental, lake, property, place, hill, cheap, find, new, grove
Topic 15	studio, property, bjb, internet, complimentary, e, change, picture, subject, fitness
Topic 16	fee, mile, white, home, tile, neighborhood, make, company, new, tenant
Topic 17	housing, water, included, heat, opportunity, landstar, equal, group, act, feature
Topic 18	contact, show, info, price, availability, email, renovated, included, click, web
Topic 19	lake, bus, block, cta, walk, red, minute, away, stop, distance
Topic 20	info, contact, show, click, il, id, feature, friendly, text, n
Topic 21	amenity, view, center, pool, fitness, luxury, outdoor, lounge, garage, window
Topic 22	logan, banker, coldwell, square, blue, real, estate, please, opportunity, equal
Topic 23	hyde, village, height, center, drexel, grand, nuestros, river, maintenance, m
Topic 24	fee, per, cat, lease, dog, application, deposit, one, heat, gas
Topic 25	new, appliance, stainless, steel, floor, central, large, feature, granite, dishwasher

Table 6: Topic clusters from LDA topic modeling on Chicago Craigslist rental listings, $k = 25$.