Text Mining and Spatial Analysis of Yelp Data to Support Socially Vibrant Cities

Alexander Bendeck, Clio Andris
{abendeck3,clio}@gatech.edu
College of Computing, Georgia Institute of Technology
Atlanta, Georgia, USA

ABSTRACT

Points of interest (POIs) are an important data layer in computational urban models. Crowdsourced reviews of services and experiences afforded by POIs can help us understand their value. Reviews can also provide information on how social ties, such as family, romantic, friendship, and professional relationships, are supported by local amenities. However, existing data-driven POI research is mostly oriented toward studies of accessibility, amenity locations, and recommendation systems. In this paper, we use computational text mining methods to analyze user reviews from the *Yelp Open Dataset* in eight cities to discover how people use POIs for social interaction. We geolocate the results and spatially analyze the locations of POIs with reviews mentioning relationship keywords.

Our analysis shows that different parts of cities host different types of relationships, and in some cases there is little overlap. We also find that certain POIs support different types of relationships more than others. We also share an interactive online tool that lets users select a relationship type of interest (e.g., "family") and search for POIs whose reviews mention these relationships. Urban planners can use these findings to reflect upon what kinds of places help support ties, which ties may need more places for their outings, and how a city can evaluate whether its social infrastructure supply is meeting the demands of residents and visitors.

CCS CONCEPTS

• **Information systems** → *Data mining*; • **Human-centered computing** → Visualization systems and tools.

KEYWORDS

Yelp, Text Mining, Urban Computing, Points of Interest, Spatial Analysis, Social Relationships

ACM Reference Format:
Alexander Bendeck, Clio Andris. 2022. Text Mining and Spatial Analysis of Yelp Data to Support Socially Vibrant Cities. In *Proceedings of The 11th International Workshop on Urban Computing (UrbComp '22).* ACM, New York, NY, USA, 10 pages.

UrbComp '22, August 15, 2022, Washington, DC

ACM ISBN 978-X-XXXX-XXXX-X/YY/MM…$15.00

1 INTRODUCTION

Points of interest (POIs) such as parks, pubs, cafes, religious institutions, community centers, and bookstores support events and serendipitous interactions in the built environment [22, 23, 26, 34]. POIs allow people to meet, have experiences, form memories, and perform activities whether they are alone, in pairs, or in groups. Certain cities may have better opportunities for convening at POIs, while others may lack such infrastructure. In addition, some cities may have POIs that are conducive to certain relationships (such as professional ties, or parent-young child relationships) while others may lack places for certain types of ties to gather.

In this work, we show how POIs provide experiences for social life and personal relationships through the following research questions:

(1) Which relationship words appear in reviews of points of interest (POIs)? Which types of relationships use POIs in which types of categories (e.g., restaurants or shops)?
(2) How do cities differ in their ability to collectively support certain types of relationships (e.g., romantic or friendship)? Are POIs that mention certain relationships concentrated in certain parts of the city?

We use POI data and reviews about POIs from the online crowdsourcing review site *Yelp* via the Yelp Open Dataset. We use basic natural language processing (NLP) techniques to find the POIs and POI types (i.e., Yelp-based POI categories: food, nightlife, etc.) where relationships are mentioned (such as "friend", "mother", etc.) under four relationship-type headings (family, romantic, friendship, and professional). The study area is eight metropolitan statistical areas in North America: Atlanta, GA; Austin, TX; Boulder, CO; Boston, MA; Columbus, OH; Orlando, FL; Portland, OR; and Vancouver, BC. We detect spatial clusters of relationship words using Getis-Ord Gi* statistics and report on how spatial contexts (e.g., suburban, urban, special use zoning) tend to attract different types of social ties.

Our results show that over 150 relationship-related words occur per 1,000 Yelp reviews, confirming that these reviews are a fruitful source of information about social relationships. As in prior work which found that spatial homophily can be found in location-based social network (LBSN) data [13, 15, 49], we also find that different parts of the city cater to different ties. Romantic relationships are often located in downtown areas, while reviews referencing children are dispersed in suburban areas; professional relationships tend to be concentrated downtown and, in some cases, near universities. Restaurants play a major role in social spaces, and gender-oriented relationship keywords are frequently associated with certain kinds of POIs.

The main contributions of this work include:

Figure 1: This Yelp review mentions the keyword "wife" as part of a user's experience at an Atlanta restaurant.

- A Python-based geo-text mining pipeline applied to the Yelp Open Dataset to analyze the occurrence of social relationship words for 160,000+ POIs in eight cities.
- A synthesis of relationship occurrences at the city level, POI category level, and individual POI level using metrics such as relationship word rate, Gini impurity, and Jaccard similarity score. This explicitly links social relationships and geographic space through crowdsourced evidence.
- Detection of statistically-significant geo-clusters at the POI level to respond to existing theories of how urban form (density, land use, etc.) produces different experiences and offerings in the city.
- An interactive visual tool for exploring POIs based on relationship word frequency, with example use case scenarios.
- Arguments towards a social relationship oriented urban computing research agenda that can help support specific relationships' (not just individuals') access to amenities to improve quality of life.

The remainder of this paper is organized as follows. We next provide a brief literature review (Section 2) and subsequently describe the dataset and analysis methods (Section 3). We then report on our results and findings (Sections 4 and 5) and finally describe how these findings can be useful in practice and lead to future work (Section 6).

2 LITERATURE REVIEW

2.1 Relationships and Urban Space

A place that can accommodate groups or pairs of individuals offers opportunities to build new social relationships or revitalize existing relationships and social networks [17]. Different communities demand different kinds of third places (i.e., places besides home and work) based on proximity, affordability, and culture [51]. Mazumdar et al. (2018) systematically reviewed 2,042 publications and found that POIs, along with walkability and land use, were the most crucial built environmental factors to social capital and relationship-building. Theoretically, our work uses a large, dynamic data set to respond to theories that the built environment is configured to exclude certain, often deprived, individuals by making destinations geographically inaccessible, socially exclusive, or too costly [37]. It extends classic studies on how elements such as shade and park benches allow for social interaction [10].

2.2 LBSN and POI Data Analysis

The urban computing community has developed multiple ways to harness and reconcile POIs across different data sources [20, 30, 35] and contributors [3]. Urban computing research has also emphasized how POIs can be used in recommendation systems [16, 46] to improve user experience, especially for mobile applications. Our analysis leverages a location-based social network (LBSN) to learn more about place and space using personal experiences via user-generated text content [39]. Prior LBSN data analyses have used text from geolocated Tweets to show that certain parts of a city are quiet, deprived, joyful, etc. [31]. Geolocated social media data has also been used to tell us more about place and about social ties separately, under the headings of urban computing and computational social science [1, 6, 12], respectively. Similarly, large POI datasets from Twitter, Foursquare, Yelp, Facebook, Weibo, Open Street Map, Flickr, etc. have been used to examine how culture and weather affects when POIs are open [40] and how their patronage varies throughout the day and year [21], as well as to recommend new POIs for people to visit [46]. Researchers have also leveraged the semantics of POI labels (and especially POIs with multiple labels) and co-occurrence of words in reviews to show that similar POIs tend to cluster geographically [15, 42], and to show the dynamics of neighborhood change over time [27]. This follows past findings that POI types can be used to segment geographies into different distinct areas [47]. Although past LBSN research has analyzed social ties in the aggregate, we disambiguate between types of ties (e.g., mother, uncle, friend) to help distinguish how different relationships access amenities.

2.3 Yelp Data

Yelp has amassed 244 million cumulative reviews (as of Dec. 31, 2021), comprised of 18% home and local services, 18% restaurants, 16% shopping, 11% beauty and fitness, etc. [44]. According to Yelp, Inc., about one third of Yelp users are 18-34, 35-54 and 55+, respectively; two thirds of users have a college degree. Reviewers give a 5-star rating (out of 5) in about half of all reviews. From our analysis of relationship keyword frequency, "husband" and "boyfriend" were mentioned more often than "wife" and "girlfriend"; we speculate that women leave Yelp reviews more frequently than men in our data sample, but this is inconclusive as we do not assume that any single reviewer is in a heterosexual partnership.

Prior research has shown that Yelp reviews, for example, tend to be well-trusted and that users "tend to act on the information" they learn from these reviews, such as deciding (not) to patronize a business [28]. Moreover, denizen reviewers tend to post with good intentions in mind – that is, with the idea that they can help others with their reviews [28]. These reviews are read by the the public and can influence a user's perception about a business [11, 28]. Compared with Google Maps restaurant reviews, Li and Hecht (2021) found that Yelp tends to have lower restaurant review ratings, due in part to higher satisfaction with chain restaurants that pervade Google Maps.

In addition, Yelp data on restaurants has been analyzed to show that reviews are not necessarily reliable indicators for food safety concerns [2]. The analysis of Yelp data has also shown that some reviews tend to be created by bots [38] or individuals who did not patronize a POI, and that aspects of Yelp reviews can be used to predict a POI's rating [7].

3 METHODS

3.1 Data Description

We conducted our analysis on the Yelp Open Dataset, which contains over 8.6 million reviews for around 160,000 businesses and is publicly available on the web [45]. Reviews are free to post and can be read by anyone who visits the Yelp website. The entire dataset includes posts from October 2004 to January 2021, but we use roughly 7.5 million reviews posted between January 1, 2010 and December 31, 2019 (Table 1) to examine a more compact time period and avoid irregularities due to the onset of the COVID-19 pandemic in the U.S. in early 2020.

The median number of reviews per business is 15 and the mean is 47.9; 34% of POIs have fewer than 10 reviews while a few outliers have many reviews. The POI with the most reviews is Voodoo Doughnut-Old Town in Portland, OR with 8,479, followed by two other restaurants, Screen Door in Portland and Mike's Pastry in Boston, MA which have over 6,600 reviews each. Twenty four percent of businesses had no relationship words in their reviews and over two-thirds of businesses had fewer than five relationship word occurrences total, perhaps due to few total reviews. In Section 4, we account for this skew by normalizing the relationship word count for each business to get the expected word count per 1,000 reviews, which we call the *relationship word rate*. When normalizing, we only consider POIs with at least 30 reviews and at least 30 relationship word occurrences, in order to avoid spurious or insignificant results.

City	# POIs	# Reviews	Median # Reviews per POI
Boston, MA	36,019	1,731,248	15
Portland, OR	28,301	1,367,612	16
Austin, TX	24,487	1,289,078	17
Orlando, FL	21,912	990,606	14
Atlanta, GA	18,092	987,780	17
Vancouver, BC	17,305	564,826	14
Columbus, OH	11,260	375,874	13
Boulder, CO	3,199	123,229	14

Table 1: Cities in the Yelp Open Dataset.

Each POI was tagged with one or more of 22 top-level categories (e.g. "Food"), as listed in the Yelp Fusion API Category List [43]. We considered only businesses within eight categories deemed likely to provide a place for social activity (see Table 2). For instance, we removed the category "Home Services" (e.g., plumbing, pool cleaners, etc.). We note the distinction between "Restaurants" and "Food", as "Restaurants" are sit-down establishments while "Food" refers to other places where food and drinks are sold, such as coffee shops and food trucks.

Some POIs (24%) belong to multiple categories, leading to "double-counting" of businesses and unreliable category breakdowns. Upon further inspection, we discovered that multi-categorization was largely due to many restaurants also being tagged as belonging to another category, such as "Food" or "Nightlife". We modified the categorizations so that any POI tagged as a "Restaurant" could not belong to any other category simultaneously, reducing the number of POIs with 2+ categories significantly (6%). Due to the difficulty of assigning unique categories to POIs (for instance, should a cosmetics store fall under "Beauty & Spas" or "Shopping"?), we leave the remaining POIs with multiple categories. We do not expect this to bias our POI category breakdowns in any systematic way because of the inconsistent nature of the remaining category overlaps.

Category	# POIs
Restaurants	50,755
Shopping	25,583
Beauty & Spas	16,505
Food	13,139
Active Life	8,905
Hotels & Travel	5,335
Arts & Entertainment	4,330
Nightlife	3,443

Table 2: POI Category Count.

3.1.1 *Relationship keywords.*

We sourced keywords from the American Time Use Survey [25] to represent social relationship types. This list was augmented by eight graduate researchers (mostly in their 20s) who were asked to add any colloquial keywords that they thought were missing from this list. They added keywords such as "bae", "boo", and "roommate". Similar to prior work [8], we grouped all resulting words into four "bins": family, romantic, friendships, and professional (Table 3), according to interpersonal relationship types from sociology and business management sectors [18, 50]. We added multiple forms of each relationship word onto our list of relationship words (e.g., "friend" and "friends") to account for word variations to avoid the need for word stemming during text mining.

Type	Words
Family	child(ren), kid(s), daughter(s), son(s), parent(s), mother, mom, father, dad, brother(s), sister(s), siblings, aunt(s), uncle(s), niece(s), nephew(s), cousin(s), grandchild(ren), grandmother, grandma, grandfather, grandpa, grandparents
Romantic	partner, relationship, date, boo, bae, sweetheart, fiance, fiancee, girlfriend, gf, boyfriend, bf, spouse, husband, wife
Friendship	bff, friend(s), buddy, buddies, pal(s), housemate(s), roommate(s), flatmate(s)
Professional	neighbor(s), classmate(s), teacher(s), coworker(s), colleague(s), client(s), boss

Table 3: Relationship Keywords.

3.2 Analytical Methods

To detect relationship words in Yelp reviews, we converted words to lower case, and replaced punctuation marks with spaces, as reviews contain possessive forms of the words (e.g., "we came for my friend's birthday"). We then used a two-word tokenization approach to detect relationship words preceded by the word "my" (or "our" for words related to children, since parents often refer to their children this way). We do this to avoid instances where authors are not referring to their own relationships (e.g., "I am a client at this salon"). This also helps with edge case of POIs with a relationship word in their name, such as the Atlanta theater *Dad's Garage*. We removed any duplicate tokens within individual reviews, to capture the prevalence of relationship words across, not within, reviews. Lastly, we counted the occurrences of our relationship keywords per POI.

To find which types of POIs cater to the most diverse set of relationships and which types of relationships use the most diverse set of POIs, we use Gini impurity. We use Jaccard similarity to find when specific pairs of relationships (like brother–sister) frequent the same POIs.

We geolocated POIs using longitude/latitude coordinates provided in the Yelp Dataset. We removed five POIs that fell outside reasonable metropolitan area boundaries. To find whether POIs or POI types cluster, we performed an adapted version of quadrat analysis where we divided POIs into 50 clusters per city using nearest neighbor analysis. We chose 50 in order to capture changes across a large metropolitan area without being too granular for a smaller area, such as Boulder (see [33]). For each city, we computed the error value $\left(\frac{(\text{Expected}-\text{Actual})^2}{\text{Actual}}\right)$ for each cluster and for each of the four types of relationships (actual values), and we compared each value to the average percentages for the four types for the entire city (expected values).

Text mining was performed in Python using the popular `pandas` and `nltk` libraries, and we performed spatial statistics and mapping in ArcMap 7.2.

4 RESULTS

4.1 Common Relationship Words By City

The most common relationship keyword across the eight cities is "friend", with 33.8 occurrences per 1,000 reviews, followed by "husband" (26.0), "wife" (17.7), "boyfriend" (12.5), and "daughter" (9.6) (Table 7). Cities' occurrences of relationship words in their reviews range from 184.5 per 1,000 reviews (Orlando) to 152.6 (Vancouver) (Table 4). Orlando may have many relationship words because of its tourism industry that attracts families to Disney World and Universal Studios; accordingly, Orlando's (and to a lesser extent Columbus's) relationship words are largely family-related.

Metro	Words	Family	Rom.	Friends	Profes.
Orlando	184.5	128.5	23.8	29.3	2.9
Columbus	175.0	113.5	22.9	35.4	3.2
Boston	172.4	89.3	31.9	47.4	3.8
Atlanta	164.8	86.5	26.3	46.9	5.1
Austin	163.8	97.7	23.5	39.1	3.5
Portland	158.5	96.7	24.5	34.8	2.5
Boulder	157.3	99.2	23.3	31.8	3.0
Vancouver	152.6	68.8	29.4	51.1	3.4

Table 4: Relationship Word Occurrences Per 1,000 Reviews by City.

To quantify the diversity of types of relationships in each city's reviews, we use the *Gini impurity* measure (as in [9, 48]) ranging from 0 to 1, where higher values indicate greater impurity. We find that Orlando (0.47) and Columbus (0.52) have the lowest impurities (perhaps due to many family words). Vancouver (0.65), Boston (0.62), and Atlanta (0.62) have high friendship and family word rates, resulting in relatively high impurity. Thus, cities like Vancouver, Atlanta, and Boston are more likely to have "something for everyone" while Orlando may emphasize families.

4.2 How POIs Serve Relationships

Next, for each POI category, we analyze how often each type of relationship word occurs per 1,000 reviews (Table 5) and find that *nightlife* (0.63 Gini impurity) supports the widest variety of relationships, followed by *restaurants* (0.61); each has relatively high occurrences of friendship and romantic-related relationships. The *active life* (0.43) and *hotels & travel* (0.46) POIs have the least variety, each with a disproportionate number of family-related relationships.

To find which types of relationships have the widest variety of options and amenities in the built environment, we compute the Gini impurity for columns in Table 5. Friendship POIs have an impurity of 0.86 and each other type has an impurity of 0.87 (not shown in table), indicating that no relationship type has a strong preference toward a POI type. This finding may ease concerns that a variety of POIs could favor one type of relationship over another.

Category	Family	Rom.	Friends	Profes.	Gini
Active	127.5	17.1	27.4	2.3	0.43
Arts	109.2	22.2	42.0	1.9	0.54
Beauty	78.4	14.3	37.0	2.5	0.56
Food	71.2	18.2	31.3	2.6	0.58
Hotels	107.5	18.6	22.9	3.5	0.46
Nightlife	52.3	25.4	67.0	2.1	0.63
Restaurants	97.1	32.1	46.5	3.8	0.61
Shopping	104.5	18.4	24.7	2.3	0.47

Table 5: Relationship Words of Each Type per 1,000 Reviews of Each POI Category.

Next, we provide further examples that support these summary statistics.

4.2.1 Restaurants. Restaurants tend to have the most relationship word occurrences and receive the most reviews. In Atlanta, restaurants such as Mary Mac's Tea Room and Poor Calvin's have 753 and 736 relationship word occurrences in their reviews, respectively. In Austin, Moonshine Patio Bar & Grill (835 occurrences) and Franklin Barbecue (711) have the most occurrences. In Boston, Mike's Pastry has 803 occurrences and Neptune Oyster has 790. In Portland, restaurants Screen Door (1,475 relationship words in its reviews) and Pok Pok (1,007) are most popular.

4.2.2 POIs for children. POIs with the highest relationship word occurrence rate (words per 1,000 reviews) are geared towards children, and are often indoor playgrounds for parties and events. In Atlanta, the POIs HippoHopp, Catch Air, and Leapin' Lizards Play & Party Center each have a relationship word rate of 575 or more, and often mention the words "child", "son", and "daughter". Boston's most popularly-reviewed playspace is VinKari Safari, with a relationship word rate of 741. Austin's popular playspaces are Epic Fun

(769) and iPlay Austin (597), and Portland's are Dizzy Castle (766) and G6 Airpark (692).

4.2.3 Friends and couples. Bars and nightclubs were popular for friendships. In Atlanta, bars such as The Dive Bar and Happy Karaoke had the most friendship-related words (250 rate). In Boston, nightclubs called Liquor Store (313) and Tunnel (298) were the most popular POIs for friendships. In Portland, nightclubs Dirty Nightlife (210) and The Nest Lounge (203) frequently mentioned friends. In Austin, the top POI for friendship was a tattoo shop called Electric 13 Tattoo (216), followed by nightclubs Trophy Club and Pure (both 200). Other than jewelry shops, flower shops and restaurants, POIs with the highest rates of romantic words included Healing Hands Massage and Wellness (246) (Atlanta), Zen Blend Massage (189) (Boston), Portland Tub and Tan (161) (Portland). Popular restaurants included Amuse! (169) (Atlanta), The Melting Pot (234) (Austin) and Agave Mexican Grille (218) (Boston) and Shari's Cafe (139) (Portland).

4.2.4 Professional ties. Whereas friends engage in a wide variety of activities, food establishments are important spaces for professional ties. Over three-quarters of POIs that mention "coworker" are restaurants, compared to just over 50% for reviews that mention "friend". "Friends" use shopping POIs almost three times as much as do "coworkers" (12.2% vs. 3.8%). Beauty & Spas, Arts & Entertainment, and Active Life POIs also comprise a larger proportion of friendship reviews. This result may not be suprising, since professional ties may have less investment in recreation and out-of-work activities. Still, there seems to be less evidence that coworkers are mentioned in a variety of POI types; the Gini impurity is 0.68 for "friend" and 0.38 for "coworker".

4.2.5 Gender, couples and family. There are more diverse POIs for relationships that include a female than a male[1]. Restaurants are more prominent with male-leaning relationships including "son", "father", and "brother" (Figure 2), suggesting that there are fewer options for this group. However, in couple relationships such as boyfriend and husband, this pattern as not as pronounced. Of POIs that include "daughter", 12% are Beauty & Spas POIs, compared to 7% including "son" (see Table 7). Using Gini impurity, we find the presence of "daughter" across POI categories has an impurity of 0.70, compared to 0.65 for "son"; "mother" (0.66), "father" (0.54); and "sister" (0.65) vs. "brother" (0.51). Yet, this trend may change in the future: the increased representation of "son" at the salon rather than "brother" or "father" may signal increasing gender inclusivity, or male-leaning interest in POIs like salons.

The Gini impurity is 0.54 for "girlfriend" vs. 0.57 for "boyfriend", and 0.61 for "wife" vs. 0.66 for "husband". These figures are perhaps more comparable because POIs that are good for boyfriends (or husbands) are also good for girlfriends (or wives), whereas this is less true for pairs like sister/brother. Using the *Jaccard similarity index* (where 1 indicates complete overlap), there is the most overlap for husband/wife (0.42) followed by boyfriend/girlfriend (0.34), daughter/son (0.33), and then mother/father (0.24), and finally sister/brother (0.21).

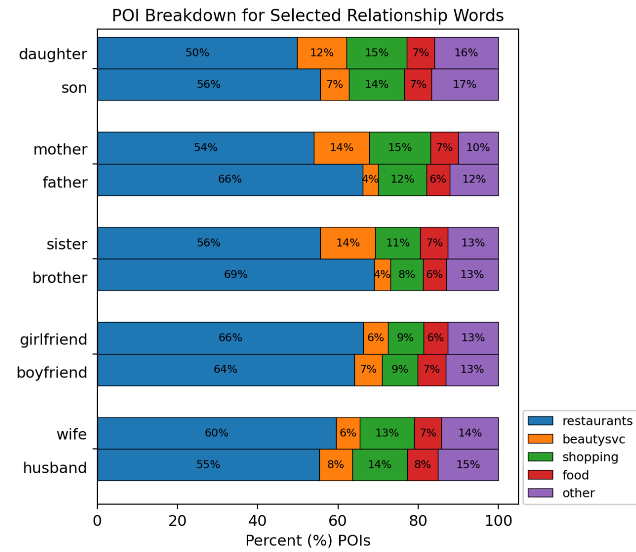

Figure 2: POI distribution for pairs of relationship words: "daughter" (n =29,919 POIs) vs. "son" (n =25,686), "mother" (n =27,850) vs. "father" (n =13,792), "sister" (n =19,376) vs. "brother" (n =10,833), "girlfriend" (n =21,264 POIs) vs. "boyfriend" (n =33,373), and "wife" (n =40,156) vs. "husband" (n =48,845). *Other* includes active life categories (including sports), nightlife, hotels, and arts establishments.

4.3 POI & Relationship Concentrations in Cities

We now examine individual cities' POI distributions and use hot spot detection to measure whether POIs that serve the four high-level relationship categories are concentrated in specific (and unique) parts of the city. Using the Getis-Ord Gi* statistic with a 2-kilometer search threshold (see Bivan and Wong 2018), we find that each city has hot spots whose locations differ depending on the type of relationship. POIs with many romantic relationship keywords per review were often densely clustered in downtown areas, while establishments with high rates of family keywords were more spread out geographically (Figure 3). Prior research discovered a similar pattern in several cities including Las Vegas, NV, Pittsburgh, PA, and Phoenix, AZ using Yelp data reviews for restaurants [32]. This configuration is likely due to more families living in the suburbs, and couples wanting to spend time together in dense, unique, historic downtowns [41].

In one example of relationship type clustering, Atlanta's core is divided into areas that are notable for romantic ties and for professional ties, with some overlapping area (Figure 4). Romantic hotspots are found in gentrifying areas of Atlanta that have changed as a result of the Atlanta Beltline (an extensive walking path) development. Professional ties are found in areas near universities and the hotel district. The wealthy neighborhood of Buckhead, north of the city, has POIs with both types of ties.

[1]Gender should not be approached as a binary construction as this excludes individuals who do not identify as male or female. Here, gendered terms are analyzed because they occurred often in reviews. References to "male" or "female" include anyone who identifies as male or female.

We now describe whether POIs of one of the four types of relationships tend to cluster more than others, and whether this clustering happens more in certain cities (Table 6). Using ANOVA to test significance, we found no significant difference in the distribution of keyword frequency or percentage of POI keywords in a category across cities. However, there were significant differences in clustering between the four relationship word types for both frequency (F statistic: 17.45, p-value: 1.38e-06) and percentage (F statistic: 7.506, p-value: 0.0007). POIs related to romantic and professional relationships tended to cluster the most. Family-related POIs tended to be the least clustered in Atlanta and Orlando, perhaps due to suburbanization and tourist attractions, respectively. Friendship-related POIs had notable clustering in Boston and Austin, presumably because POIs near universities in these cities attract groups of friends.

City	**Keyword Frequency (Counts)**			
	Family	Romantic	Friendship	Professional
Atlanta	516.76	53	127.77	27.87
Austin	217.38	34.12	88.17	62.05
Boston	199.91	56.38	83.82	18.37
Boulder	126.96	87.52	186.24	46.29
Columbus	215.05	80.04	136.31	42.71
Orlando	408.89	60.61	149.95	23.13
Portland	326.5	72.45	218.71	39.5
Vancouver	425.28	166.28	110.54	145.82
City	**Percentage of Keywords in a Category**			
	Family	Romantic	Friendship	Professional
Atlanta	1.39	0.42	1.24	0.33
Austin	0.65	0.22	0.59	0.59
Boston	0.66	0.5	0.67	0.24
Boulder	0.58	0.34	0.75	0.45
Columbus	0.59	0.39	0.89	0.39
Orlando	1.11	0.46	1.19	0.37
Portland	1.06	0.45	1.75	0.37
Vancouver	0.85	0.72	1.12	1.2

Table 6: Error values for expected vs. actual keyword frequencies and percentages of keywords in one relationship type.

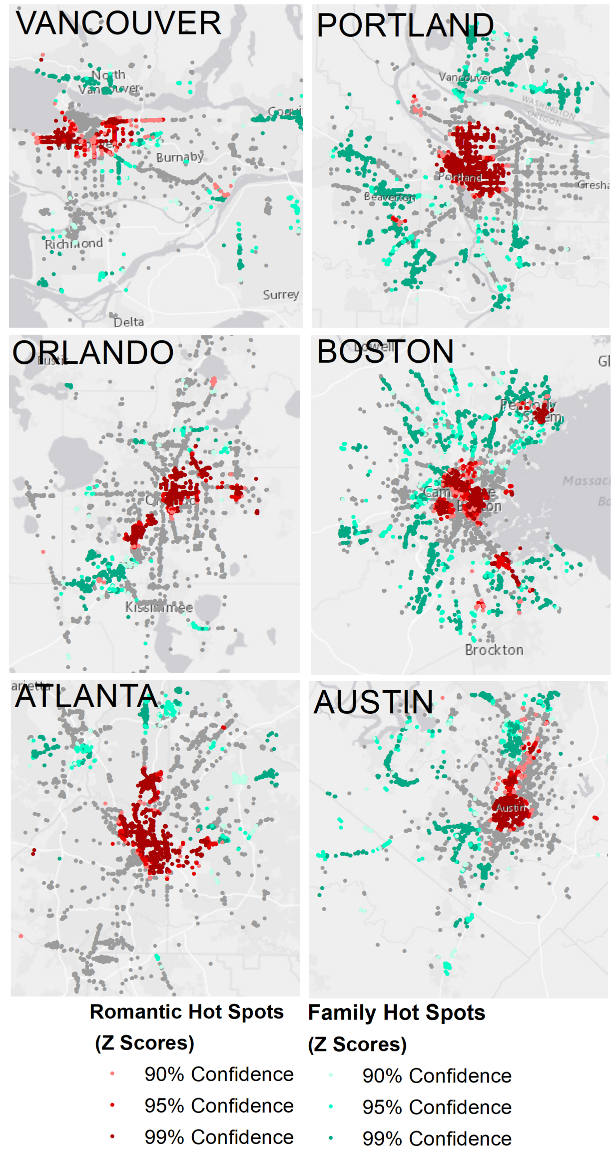

Figure 3: Geographic distribution for family and romantic relationships hot spots from the normalized Z scores of the Getis-Ord Gi* statistic. Although there could be overlap on the map due to drawing order, there is virtually no overlap between the two types of hot spots.

5 INTERACTIVE TOOL

5.1 Overview

In this section, we introduce an interactive online tool[2] that allows users to explore a city's POIs by both POI type and the type of relationships they support. It is built with paradigms of interactive visualization in mind that build on the framework of providing a broad overview, letting users zoom and filter, and allowing users to seek finer details on individual entities [36]. For each of the eight cities, the tool maps POIs whose reviews contain relationship words of the desired type(s) in their reviews. The tool is written in the R programming language and uses the `shiny` and `leaflet` packages to render the user interface as a front-facing web visualization. Our intended users include locals and tourists looking to find a venue to patronize under certain social contexts, as well as professionals aiming to understand how city layouts support different social relationships or where to put new businesses.

5.2 Features

The tool's interface is shown in Figure 5. Users first select one of the eight cities represented in the Yelp dataset. They can then select the relationship types they are interested in (e.g., "professional" or "romantic") using checkboxes, and optionally filter by POI type (e.g., "Restaurants" or "Arts & Entertainment") using a dropdown

[2]Viewable at: https://doi.org/10.6084/m9.figshare.20352963

menu. The user can view the *raw* number of relationship words, relationship *word rate* (number of words per 1000 reviews), or for specific relationship types, the percentage of relationship keywords in their reviews under that type. Users can limit the number of POIs to show on the map, wherein POIs with the highest relationship word rate (or raw count, or percent of words under the selected type) are drawn first.

The tool draws points on the map that represent the top businesses for the chosen relationship word settings. POIs are symbolized by one of four different colors according to relationship type, and POIs with higher values of the selected value of interest (relationship word rate, raw count, or percent under the selected type) are rendered with higher opacity. Users can hover over POI points to see tooltips that include the name of the business and details about the relationship word occurrences in its reviews.

The tool also has a sortable table of the POIs that are currently in view, which includes additional attributes such as total number of reviews. Users can select a POI in the table and "jump" to it on the map. In this way, the tool supports *exploration* of different neighborhoods and serves as a *reference* through search capabilities (i.e., the ability to look up specific POIs). By providing zoomed out snapshots of cities' POIs, it also allows users to visually distinguish the prevalent relationships supported by various parts of a city and thus valuate different parts of the city with relationships in mind.

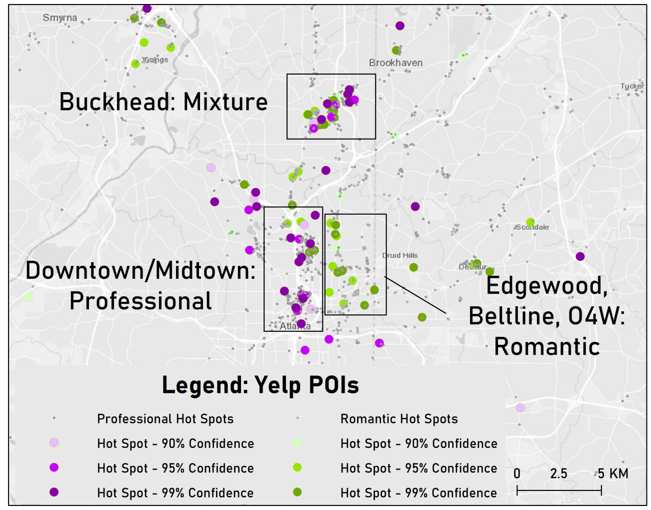

Figure 4: Geographic distribution for professional and romantic relationships in Atlanta shows distinctive areas for each type of tie in the downtown core, with a notable overlap in a northern neighborhood.

5.3 Example Use Case Scenarios for Decision Support

5.3.1 Scenario 1: Urban Planning. Xin is an urban planner in Portland. She is planning on creating a new zoning plan for the city that supports a new family community center in the suburbs. She knows that some suburbs are oriented more towards families with children, and she sees using the tool that there is a cluster of family-oriented POIs in Beaverton. The displayed map is helpful because Census data did not provide this information, travel diaries mentioned destinations but did not capture enough households, and GPS trace data did not include information on children. While she will also consider site suitability and equity factors, Xin feels more confident about placing the new community center in an area with POIs that already serves families, in order to allow for trip-chaining and convenience.

Over the past 15 years, Xin has also worked on a "complete streets" project in downtown Portland to promote walkability, bikeability, and safe corridors for pedestrians. She wonders if tourists, locals, children, friends, and couples all enjoy these areas or if the downtown is only serving a small segment of the population (e.g., tourists in hotels), as other cities' central business districts (CBDs) have struggled to support this variety. She uses the tool and finds that indeed, a variety of relationships are being supported by a mixture of POIs. She is especially surprised to see so many family-oriented POIs in the downtown area. She visually compares these results to Columbus, OH's map and sees that their downtown has very few relationship-oriented POIs.

5.3.2 Scenario 2: Business Lunch. Jim lives in Atlanta and is meeting a business associate for lunch. He wants to find an appropriate restaurant for this meeting, ideally close to the downtown Hotel District since the associate is staying near there. Jim opens the tool, selects Atlanta, and checks only the box for "Professional" relationship words. He then filters to see businesses in the "Food & Restaurants" category. He zooms in to the map to focus on the downtown area that includes the Hotel District and hovers over a few POIs in this area. The displayed restaurants all have a similar percentage of "Professional" relationship words, so Jim changes the view options to instead show the top businesses with the highest "Professional" relationship word rate. Now, one particular restaurant in the Hotel District stands out as having a particularly high rate. Hovering over the restaurant, Jim finds that it is called Sear and has a rate of almost 40 "Professional" words per 1000 reviews. The top other restaurants nearby have a rate only around 25. Jim does a quick online search for Sear, is satisfied with its accommodations, and decides to make a lunch reservation there for himself and his associate.

5.3.3 Scenario 3: Finding a Neighborhood. Pedro is looking to move to Boston for the summer. They would like to be in an area that has *nightlife* and may be good for a resident who lives alone. They visit the tool, select Boston, and choose to filter the top businesses in the "Nightlife" category. Pedro decides they are interested in locales that support friendships in order to help them connect with new people in Boston. They select the checkbox for "Friendship" relationship words and find a cluster of POIs near Northeastern University and Fenway, indicating that these areas are good for nightlife. Pedro then changes the dropdown from "Nightlife" to choose "Arts & Entertainment" and notices that there are many nearby options for these types of POIs as well. Using the "Table of Businesses" tab, they find that the Museum of Science, Fenway Park, and Isabella Stewart Gardner Museum are among the most popular options, and they decide to search for the Museum of Science's hours and address through a search engine. They make plans to go there once they arrive. In all, Pedro now feels more confident about moving to the Northeastern-Fenway area, since this part of the city seems to be a place where their needs can be supported through the nearby POIs.

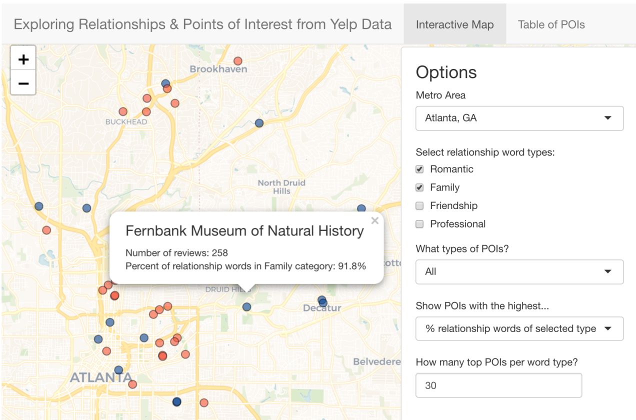

Figure 5: This is the interface for our R Shiny tool to search for POIs supporting specific relationship types. "Romantic" POIs are shown in red, while "Family" POIs are in blue.

6 DISCUSSION AND STUDY LIMITATIONS

6.1 Implications for Urban Planning and Users

Showing how cities are spatially segmented by the different relationships they serve can help urban planners foster a sense of community and provide spaces for people to live, work, and play–which are primary planning goals [5, 24, 29]. Since our approach attaches a new attribute to POIs, it can help planners valuate places differently–not through high or low traffic, crime, or wealth, but rather how they perform in service to social life. Planners can see which kinds of relationships have sufficient places to spend time together (i.e., if urban spaces offer "something for everyone") or if new venues are needed. In terms of decision support, having more explicit links between public spaces and personal ties can be used to inform decisions on what to build, preserve, or retrofit to support specific joint and group activities. For the typical user who may explore a city with our visualization tool, a relationship-oriented approach makes the city more personal by giving users a chance to see the city through the lens of their own social networks. It puts the user's social life and interpersonal relationship health in a focal position, helping them make more informed decisions about how to create memorable experiences with others. This opportunity is especially important given the isolation and loneliness felt through the COVID-19 pandemic, as well as other prevailing trends in digital social life.

6.2 Limitations & Future Work

This study had several limitations. First, we rely on user generated content, which may not accurately reflect the actual social ties that are associated with or that benefit from a certain POI. Next, the keywords we chose may not have sufficiently captured the variety of relationships that use POIs. We also may neglect to capture certain regional dialects or relationship-oriented words that the authors have not associated with relationships, or are unknown to the authors. We asked students for their input on relationship words to include beyond those from the American Time Use Survey, but did not capture a wide range of geographic and demographic groups. We use English-oriented keywords and thus may miss relationship words that are not in English, such as *abuela* for "grandmother" in Spanish. We also do not account for other relationship word edge cases, such as women using the word "girlfriend" to refer to a platonic friendship. The term "partner" is also ambiguous as to whether it refers to a professional or romantic relationship, although the latter is likely more common and increasingly so. This leaves an opportunity for future work to take a more nuanced and inclusive approach to dealing with relationship words.

Regarding technical limitations, our spatial analysis does not consider the density of POIs that do not have relationship keywords, and thus, is a self-selected sample. The Yelp dataset also only contains information for a limited number of cities, and its users tend to be a select set of individuals. Our analysis includes places such as kids' hair cut services and jewelry stores that, while important, may not adequately provide *third place* experiences. Further work should also disaggregate the restaurant/food categories into subtypes, such as upscale dining, sports bars, hot pot restaurants, etc. that may be used by pairs and groups for different purposes. From such an analysis (perhaps in future work), we may be able to find stronger connections between relationship types and locales. In the future, location-based social networks that suggest POIs could also make suggestions based on who the user wants to spend time with.

Finally, this type of analysis and supporting tool may reinforce or "brand" particular spaces as being suitable or unsuitable for certain types of activities. Users should be cautioned that these results only present one perspective of cities and social network relationships, and a lack of POIs or information about a place does not imply that a place is not suitable for social life or activities (and vice versa).

To support study replication and future work in this space, we have released the text mining pipeline and processed dataset used for this study's analysis on GitHub[3] for public use.

7 CONCLUSION

In this study, we use basic natural language processing techniques to conduct text mining on the Yelp Open Dataset and investigate various phenomena about the occurrence of social relationship words in the reviews of POIs in eight cities. We show that online reviews can show how a city or a type of place – not just an individual business – serves residents and visitors. Our findings suggest that place reviews are a valuable source of relationship data and common terms are related to friendship, couples, and children. POI categories (i.e., restaurant vs. shopping) support different types of relationships; active life POIs (e.g., a golf course) have high rates of family words, while nightlife POIs have many friendship words in their reviews. We also find that female-related terms tend to have a broader distribution across POI types. Lastly, certain neighborhoods or districts of individual cities often support specific relationship types, which supports longstanding theories that cities are segmented and that different urban forms serve different purposes.

[3]https://github.com/AlexanderBendeck/yelp-relationships-geography

REFERENCES

[1] Lada A Adamic and Eytan Adar. 2003. Friends and neighbors on the web. *Social Networks* 25, 3 (2003), 211–230.

[2] Kristen M Altenburger and Daniel E Ho. 2019. Is Yelp Actually Cleaning Up the Restaurant Industry? A Re-Analysis on the Relative Usefulness of Consumer Reviews. In *The World Wide Web Conference*. 2543–2550.

[3] Jennings Anderson, Dipto Sarkar, and Leysia Palen. 2019. Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information* 8, 5 (2019), 232.

[4] Roger S Bivand and David WS Wong. 2018. Comparing implementations of global and local indicators of spatial association. *Test* 27, 3 (2018), 716–748.

[5] Adam Boessen, John R Hipp, Carter T Butts, Nicholas N Nagle, and Emily J Smith. 2018. The built environment, spatial scale, and social networks: Do land uses matter for personal network structure? *Environment and Planning B: Urban Analytics and City Science* 45, 3 (2018), 400–416.

[6] Moira Burke and Robert E Kraut. 2016. The relationship between Facebook use and well-being depends on communication type and tie strength. *Journal of Computer-Mediated Communication* 21, 4 (2016), 265–281.

[7] Yifan Chen and Fanzeng Xia. 2020. Restaurants' Rating Prediction Using Yelp Dataset. In *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*. IEEE, 113–117.

[8] Minje Choi, Ceren Budak, Daniel M Romero, and David Jurgens. 2021. More than Meets the Tie: Examining the Role of Interpersonal Relationships in Social Networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 105–116.

[9] Saptarshi Das, Shamik Sural, Jaideep Vaidya, and Vijayalakshmi Atluri. 2018. Using Gini Impurity to Mine Attribute-Based Access Control Policies with Environment Attributes. In *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies* (Indianapolis, Indiana, USA) *(SACMAT '18)*. ACM, New York, NY, USA, 213–215.

[10] Jan Gehl and Birgitte Svarre. 2013. *How To Study Public Life.* Island Press, Washington, DC.

[11] Hartwig H Hochmair, Levente Juhász, and Sreten Cvetojevic. 2018. Data quality of points of interest in selected mapping and social media platforms. In *LBS 2018: 14th International Conference on Location Based Services*. Springer, 293–313.

[12] Jason J Jones, Jaime E Settle, Robert M Bond, Christopher J Fariss, Cameron Marlow, and James H Fowler. 2013. Inferring tie strength from online directed behavior. *PloS one* 8, 1 (2013), e52168.

[13] Neal Lathia, Daniele Quercia, and Jon Crowcroft. 2012. The hidden image of the city: sensing community well-being from urban mobility. In *International conference on pervasive computing*. Springer, 91–98.

[14] Hanlin Li and Brent Hecht. 2021. 3 Stars on Yelp, 4 Stars on Google Maps: A Cross-Platform Examination of Restaurant Ratings. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

[15] Xi Liu, Clio Andris, and Sohrab Rahimi. 2019. Place niche and its regional variability: Measuring spatial context patterns for points of interest with representation learning. *Computers, Environment and Urban Systems* 75 (2019), 146–160.

[16] Yaqiong Liu and Hock Soon Seah. 2015. Points of interest recommendation from GPS trajectories. *International Journal of Geographical Information Science* 29, 6 (2015), 953–979.

[17] Judith Mair and Michelle Duffy. 2020. The Role of Festival Networks in Regional Community Building. In *Located Research*, Angela Campbell, Michelle Duffy, and Beth Edmondson (Eds.). Palgrave McMillan, Singapore, 89–116.

[18] Alexandra Marin and Keith N. Hampton. 2007. Simplifying the personal network name generator. *Field Methods* 19, 2 (2007), 163–193.

[19] Soumya Mazumdar, Vincent Learnihan, Thomas Cochrane, and Rachel Davey. 2018. The built environment and social capital: A systematic review. *Environment and Behavior* 50, 2 (2018), 119–158.

[20] Grant McKenzie, Krzysztof Janowicz, and Benjamin Adams. 2014. A weighted multi-attribute method for matching user-generated points of interest. *Cartography and Geographic Information Science* 41, 2 (2014), 125–137.

[21] Grant McKenzie, Krzysztof Janowicz, Song Gao, Jiue-An Yang, and Yingjie Hu. 2015. POI Pulse: A Multi-granular, Semantic Signature–Based Information Observatory for the Interactive Visualization of Big Geosocial Data. *Cartographica* 50, 6 (2015), 71–85.

[22] Vikas Mehta. 2007. Lively streets: Determining environmental characteristics to support social behavior. *Journal of Planning Education and Research* 27, 2 (2007), 165–187.

[23] Vikas Mehta and Jennifer K Bosson. 2010. Third places and the social life of streets. *Environment and Behavior* 42, 6 (2010), 779–805.

[24] Kostas Mouratidis. 2018. Built environment and social well-being: How does urban form affect social life and personal relationships? *Cities* 74 (2018), 7–20.

[25] U.S. Bureau of Labor Statistics. 2020. *American Time Use Survey (ATUS).* https://www.bls.gov/tus/database.htm

[26] Ray Oldenburg. 1989. *The Great Good Place: Cafes, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community.* Paragon House, New York.

[27] Alexander W Olson, Fernando Calderon-Figueroa, Olimpia Bidian, Daniel Silver, and Scott Sanner. 2021. Reading the city through its neighbourhoods: Deep text embeddings of Yelp reviews as a basis for determining similarity and change. *Cities* 110 (2021), 103045.

[28] Anish Parikh, Carl Behnke, Mihaela Vorvoreanu, Barbara Almanza, and Doug Nelson. 2014. Motives for reading and articulating user-generated restaurant reviews on Yelp.com. *Journal of Hospitality and Tourism Technology* 5, 2 (2014), 160–176.

[29] Deirdre Pfeiffer and Scott Cloutier. 2016. Planning for happy neighborhoods. *Journal of the American Planning Association* 82, 3 (2016), 267–279.

[30] Mateusz Piech, Aleksander Smywinski-Pohl, Robert Marcjan, and Leszek Siwik. 2020. Towards automatic points of interest matching. *ISPRS International Journal of Geo-Information* 9, 5 (2020), 291.

[31] Daniele Quercia. 2013. Urban*: Crowdsourcing for the good of London. In *Proceedings of the 22nd International Conference on World Wide Web*. 591–592.

[32] Sohrab Rahimi, Clio Andris, and Xi Liu. 2017. Using Yelp to find romance in the city: A case of restaurants in four cities. In *Proc. UrbanGIS'17: 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*. ACM, 1–8.

[33] Rafael G Ramos, Bráulio FA Silva, Keith C Clarke, and Marcos Prates. 2021. Too fine to be good? Issues of granularity, uniformity and error in spatial crime analysis. *Journal of Quantitative Criminology* 37, 2 (2021), 419–443.

[34] Mark S Rosenbaum. 2006. Exploring the social supportive role of third places in consumers' lives. *Journal of Service Research* 9, 1 (2006), 59–72.

[35] Tatjana Scheffler, Rafael Schirru, and Paul Lehmann. 2012. Matching Points of Interest from Different Social Networking Sites. In *KI 2012: Advances in Artificial Intelligence*, Birte Glimm and Antonio Krüger (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 245–248.

[36] Ben Shneiderman. 2003. The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization*. Elsevier, 364–371.

[37] David Sibley. 2002. *Geographies of Exclusion: Society and Difference in the West.* Routledge, Abingdon, UK.

[38] Andre Sihombing and Alvis Cheuk Ming Fong. 2019. Fake review detection on Yelp dataset using classification techniques in machine learning. In *2019 International Conference on Contemporary Computing and Informatics (IC3I)*. IEEE, 64–68.

[39] Thiago H Silva, Aline Carneiro Viana, Fabrício Benevenuto, Leandro Villas, Juliana Salles, Antonio Loureiro, and Daniele Quercia. 2019. Urban computing leveraging location-based social network data: a survey. *ACM Computing Surveys (CSUR)* 52, 1 (2019), 1–39.

[40] Kevin Sparks, Gautam Thakur, Amol Pasarkar, and Marie Urban. 2020. A global analysis of cities' geosocial temporal signatures for points of interest hours of operation. *International Journal of Geographical Information Science* 34, 4 (2020), 759–776.

[41] Emily Talen and Hyesun Jeong. 2019. Does the classic American main street still exist? An exploratory look. *Journal of Urban Design* 24, 1 (2019), 78–98.

[42] Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Song Gao. 2017. From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–10.

[43] Yelp, Inc. 2022. *All Category List - Yelp Fusion.* https://www.yelp.com/developers/documentation/v3/all_category_list

[44] Yelp, Inc. 2022. *Fast Facts.* https://www.yelp-press.com/company/fast-facts/default.aspx

[45] Yelp, Inc. 2022. *Yelp Open Dataset: An All-Purpose Dataset for Learning.* https://www.yelp.com/dataset

[46] Josh Jia-Ching Ying, Eric Hsueh-Chan Lu, Wen-Ning Kuo, and Vincent S. Tseng. 2012. Urban Point-of-Interest Recommendation by Mining User Check-in Behaviors. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing* (Beijing, China) *(UrbComp '12)*. Association for Computing Machinery, New York, NY, USA, 63–70. https://doi.org/10.1145/2346496.2346507

[47] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 186–194.

[48] Ye Yuan, Liji Wu, and Xiangmin Zhang. 2021. Gini-Impurity Index Analysis. *IEEE Transactions on Information Forensics and Security* 16 (2021), 3154–3169.

[49] Ke Zhang and Konstantinos Pelechrinis. 2014. Understanding spatial homophily: the case of peer influence and social selection. In *Proceedings of the 23rd International Conference on World Wide Web*. 271–282.

[50] Ran Zhang and Phil Carspecken. 2013. Content inference fields in intersubjective space: Transpersonal position, illocution, and logic in the analysis of human interactions. *Qualitative Research: A Reader in Philosophy, Core Concepts, and Practice* (2013), 201–242.

[51] Sharon Zukin. 1998. Urban lifestyles: diversity and standardisation in spaces of consumption. *Urban Studies* 35, 5/6 (1998), 825–839.

8 APPENDIX: COUNTS OF RELATIONSHIP WORDS PER 1000 REVIEWS

Word	All Cities (n=7.5M)	Atlanta (n=988K)	Austin (n=1.3M)	Boston (n=1.7M)	Boulder (n=123K)	Columbus (n=376K)	Orlando (n=991K)	Portland (n=1.4M)	Vancouver (n=565K)
friend	33.80	38.17	29.80	36.80	27.87	31.68	25.94	29.25	50.02
husband	26.00	23.44	27.71	23.07	26.34	33.49	33.66	26.56	17.09
wife	17.66	14.34	17.08	16.19	18.34	25.50	24.31	17.52	12.63
boyfriend	12.51	12.00	10.70	15.05	11.18	11.24	12.57	10.70	14.61
daughter	9.61	7.73	11.36	8.29	10.21	9.89	13.70	10.48	4.66
mother	8.47	8.06	7.60	9.01	6.99	9.36	9.59	7.97	8.38
son	8.22	6.54	10.51	6.73	8.39	9.43	11.31	8.77	3.94
child	6.67	4.88	8.80	5.25	6.40	6.78	9.28	7.41	3.73
girlfriend	5.30	4.73	4.65	6.38	4.97	5.28	5.02	4.51	6.52
sister	4.44	4.68	3.76	4.87	3.51	4.99	4.59	4.10	4.58
father	3.21	2.56	2.95	3.49	2.93	3.91	3.68	3.02	3.21
partner	2.41	1.89	1.47	1.85	2.32	1.84	1.48	3.58	5.52
coworker	2.30	3.51	1.82	2.59	1.76	2.43	1.97	1.59	2.76
parent	2.08	1.57	1.81	2.23	1.90	2.65	2.37	1.67	2.92
brother	2.02	1.93	1.65	2.11	1.84	2.28	2.29	1.86	2.35
fiance(é)	1.89	1.75	1.77	2.83	1.55	1.93	1.50	1.53	1.28
housemate	1.19	0.98	1.83	1.93	1.16	0.85	0.59	0.81	0.33
neighbor	1.08	1.42	2.39	0.78	0.64	0.58	0.58	1.07	0.16
cousin	1.00	1.31	0.71	1.12	0.89	0.73	1.04	0.81	1.34
date	0.90	1.51	0.70	0.95	0.75	0.98	0.65	0.65	1.18
grandmother	0.77	0.83	0.74	0.76	0.59	0.87	0.75	0.77	0.81
niece	0.56	0.53	0.47	0.54	0.36	0.77	0.82	0.56	0.34
client	0.46	0.47	0.82	0.39	0.51	0.21	0.30	0.50	0.28
boss	0.45	0.54	0.46	0.54	0.43	0.44	0.40	0.36	0.30
aunt	0.44	0.41	0.34	0.52	0.36	0.44	0.47	0.35	0.58
nephew	0.35	0.30	0.33	0.30	0.18	0.51	0.57	0.32	0.25
spouse	0.30	0.33	0.26	0.21	0.31	0.34	0.33	0.32	0.38
grandfather	0.19	0.16	0.19	0.18	0.23	0.21	0.20	0.23	0.18
uncle	0.19	0.17	0.16	0.21	0.19	0.19	0.16	0.20	0.26
grandparents	0.14	0.09	0.10	0.15	0.17	0.22	0.13	0.14	0.23
bff	0.14	0.21	0.15	0.12	0.03	0.13	0.12	0.14	0.17
relationship	0.10	0.11	0.16	0.12	0.09	0.05	0.05	0.10	0.06
boo	0.09	0.19	0.09	0.05	0.07	0.06	0.08	0.09	0.05
siblings	0.06	0.07	0.04	0.06	0.01	0.04	0.07	0.07	0.08
classmate	0.06	0.04	0.07	0.06	0.05	0.10	0.03	0.03	0.09
grandchild	0.04	0.02	0.04	0.04	0.06	0.05	0.07	0.06	0.02
teacher	0.04	0.02	0.05	0.07	0.05	0.00	0.02	0.03	0.05
professor	0.02	0.03	0.02	0.04	0.00	0.04	0.02	0.01	0.00

Table 7: Word Counts per 1,000 Words Found in Reviews, By City; n is the Number of Reviews per City. Yelpers in Vancouver used "friend" about fifty times per 1000 reviews, almost twice the rate that "friend" was used in reviews for Orlando POIs. In Vancouver and Portland, the word "partner" (5.52 and 3.58 rate, respectively) is used more often than in reviews for Orlando (1.48) and Austin (1.47), which are in conservative states with recent legislation opposing the use of gender-inclusive language such as "partner". The word "neighbor" is most common in Austin, signaling perhaps more social events with those living nearby.