

Understanding Gendered Spaces Using Social Media Data

Aljoharah Alfayez^{1,2(✉)}, Zeyad Awwad^{1,2(✉)}, Cortni Kerr^{3(✉)},
Najat Alrashed^{1,2(✉)}, Sarah Williams^{3(✉)}, and Areej Al-Wabil^{1,2(✉)}

¹ Center for Complex Engineering Systems (CCES),
King Abdulaziz City for Science and Technology (KACST),
Riyadh, Saudi Arabia
{a.alfayez, z.alawwad, n.alrashed}@cces-kacst-mit.org,
areej@mit.edu

² MIT, Cambridge, USA

³ Civic Data Design Lab, MIT, Cambridge, MA, USA
{chkerr, sew}@mit.edu

Abstract. In Saudi Arabia, gender shapes cities in a way that is not commonly found in other cities due to Saudi Arabia's imposed gender segregation. This segregation policy drives both genders to different areas of the city in different ways, influencing the emergence of gendered spaces. In this paper, we utilize social media data to better understand gendered spaces throughout the city of Riyadh. For our analysis, we developed an algorithm to perform gender annotation based on users' first names. The method, optimized for English and Arabic language names, was applied to a sample of over 120,000 geotagged tweets between November 2016 and January 2017. The customer demographics of Foursquare venues were estimated based on the gender ratio of reviewers. Areas with a high degree of gender concentration in these datasets were used to identify gendered spaces. The correlation between gendered space identified from tweets and Foursquare venues was used to examine the link between amenities and gender-specific mobility habits in Riyadh. Throughout our analysis, we aim to identify ways in which government policies and the organization of businesses and services with similar customer demographics impact the mobility patterns of women and men and lead to the emergence of gendered spaces in Riyadh.

Keywords: Gendered spaces · Gender annotation · Gender segregation · Social media · Foursquare · Twitter · Urban planning

1 Introduction

In this work, our objective is to explore how gender information in social media can be used to understand the effects of gender segregation and mobility restriction on women in Riyadh, Saudi Arabia. Gender based variations of mobility is an established phenomenon across the world [1]. However, the depth of understanding this phenomenon in Saudi Arabia is inadequate [2]. By identifying and comparing gendered spaces from geotagged tweets and Foursquare venue reviews, we examine how female dominant

spaces differ from male dominant spaces and record general observations of where they tend to cluster most. To infer the gender of geotagged tweets in Saudi Arabia, we have developed an algorithm that relies heavily on Arabic specific language features to boost performance. We have also used Foursquare reviews and check-ins to understand the customer demographics of Foursquare venues.

Additionally, we describe a method we used to identify gendered spaces in the city. As a final step, we overlap female gendered spaces emerging from geotagged tweets and from our Foursquare demographic results to find patterns that emerge from those two data sets. We replicate this process with male spaces. Finally, we compare female patterns to male patterns. This work can be used as a first step to identify the innate differences between mobility patterns of men and women along with the implications of female mobility restrictions and segregation on the socio-spatial patterns of Riyadh.

2 Background

2.1 Gender Sensitivity in Urban Planning

The role of gender in urban planning theory and practice increased significantly toward the end of the 20th century, leading to the development of the concepts such as gendered spaces and gendered mobility [3–6]. This evolution in theory and practice occurred in response to rising awareness that women experience cities in different ways than men. The term “gendered spaces” is used here to refer to the spatial segregation of men and women at both the architectural and geographic scales [4, 7]. For example, a house may have specific women’s and men’s spaces and different parts of the city can be used differentially by men and women, whether it is a predominately male athletic stadium or a market frequented mostly by women. Spain (1993) argues that “Women’s position within society, whether measured as power, prestige, economic position, or social rank, is related to spatial segregation insofar as existing physical arrangements facilitate or inhibit the exchange of knowledge between those with greater and those with lesser status” [4].

Gendered spaces have developed across many different contexts and cultures [4]. Despite a burgeoning literature on gendered spaces in U.S. and European contexts, limited work has been published that explores gender and space in Saudi Arabia at the city scale. Saudi Arabia is unique when it comes to gendered spaces. There is a national policy of gender segregation in public spaces dictating where and with whom women may occupy urban space. For example, restaurants and cafes have separate seating sections for men only and for women and families. Likewise, most schools and universities are gender segregated and athletic stadiums are primarily for men only. Moreover, women are prohibited from driving in Saudi Arabia, limiting their mobility options and potentially influencing the emergence of gendered spaces.

2.2 Gender and Mobility

Evidence shows that women and men travel within cities differently. Identifying gendered spaces in Riyadh is the first step in understanding gendered mobility in

the city. According to Law (1999), “Daily mobility incorporates a range of issues central to human geography, including the use of (unequally distributed) resources, the experience of social interactions in transport-related settings and participation in a system of cultural beliefs and practices.” [10] Gendered differences in travel patterns hold across the Global North and Global South contexts [8]. Much of the quantitative work on gendered mobility converges around the conclusion that women have shorter commute times than men [5, 8, 9]. However, context dictates whether this a positive or negative finding, when taking into consideration differential costs of travel. Moreover, this quantitative work is confined largely to national-scale, home-to-work commuting population surveys [11]. This type of data does not reflect non-work mobility and is also unavailable in Saudi Arabia. As an alternative, social media data is an opportunity to understand the spatial distribution of gendered spaces and mobility at the city scale.

2.3 Gender in Social Media

Social media are defined as tools for communication, where users can create and share their content with others through social networks. These web-based tools generate vast amounts of readily available data about populations around the world, triggering researchers to explore them. In Saudi Arabia, social media has become a major source of data. In 2016, it was reported that Saudi Arabia had the highest penetration rate of Twitter users in the world [12]. A 2012 study reported that the number of registered twitter users in Saudi Arabia grew over 93% in six months, reaching 2.9 M. Additionally, Riyadh, Saudi Arabia’s capital city, was described as the tenth most active city by number of posted tweets worldwide [13]. However, as informative as social media data is, it lacks some basic demographic information that can be highly useful in different domains. For this reason, many researchers have focused their efforts on finding ways of inferring some of this missing information.

Identifying user demographics from social media data is valuable and applicable in many domains, from business marketing to sociolinguistic analysis to behavior analysis. Several methods have been employed to infer data from social media. Most methods rely on using classification and regression techniques on certain characteristics found in a user’s profile, such as names and language. From these characteristics, researchers have developed techniques to predict different demographic information, such as ethnicity, gender and political preference [16, 17, 19]. In particular, gender annotation using names has been a well explored theme, specifically in the context of social media [17–21]. The use of gender-name associations to label social media data is generally regarded as the most reliable gender inference method, specifically with tweets [17]. Gender inference accuracy is reported between 80 and 85% for English language names [16].

In the literature, gender inference from social data was used for understanding the Twitter population. A study of commuting patterns in Toronto, Canada found compelling evidence that Twitter data could be used to detect the demographic compositions of communities [14]. In a 2011 study, researchers utilized the U.S. Social Security Administration’s database of birth records to classify the gender of 71% of the 3.2 million Twitter users in their study [19]. A 2016 study found that pairing web traffic

demographic data with Twitter data in a non-fully-supervised regression model produced highly accurate gender classification, with moderately accurate ethnicity and political classification [18].

However, despite the growth of Twitter around world and across several language contexts, there is limited material on non-English gender inference. However, retraining the methods used for English-language names and incorporating language-specific identifiers has been shown to produce successful results [16]. This has yet to be applied in a significant manner to Arabic, which is the 6th most commonly used language on the platform [13]. In the next section, we shall describe our Arabic-optimized method of inferring the gender of geotagged tweets and Foursquare venue reviews in Riyadh.

3 Gender Annotation

3.1 Datasets

The analysis of this paper was applied on two datasets: 123,977 geotagged tweets collected from November 2016 to January 2017 using the Twitter API and 73,747 Foursquare venues collected from the Foursquare website in November 2016. With both data sets, only locations within a bounding box of 24.25° to 25.25° latitude and 46.25° to 47.25° longitude were considered in this study. After applying the name-based gender annotation methodology described in Sect. 3.2, a total of 18,302 tweets were classified as female and 32,839 were classified as male. The methodology described for Foursquare venues identified 1,057 male-oriented venues and 483 female-oriented venues

Both datasets contain implicit biases that should be kept into consideration. Since Foursquare data is crowdsourced, it relies on businesses' staff and customers to add venues to the database. Businesses that attract different demographics may not be equally represented in their data. Furthermore, the Foursquare data includes private reviews which do not publicly display the user's gender. With the Twitter data set, the user's name is self-reported so there is a possibility that it does not accurately reflect their gender [22, 23]. Due to a combination of sampling bias and potential limitations of the annotation method, both datasets have a significantly greater proportion of male users than female users.

3.2 Gender Annotation

As previously mentioned, the collected Foursquare data of Riyadh's amenities include a gender attribute in the reviews. Thus, the data set did not require the use of the name gender annotation algorithm described below. However, we applied a set of filters to determine if a venue was oriented toward a single gender. First, we removed venues that did not have at least 10 reviews with a gender attribute. This step reduced the number of venues to 2,915. Although this filtration step removed the majority of venues from our data set, venues with fewer reviews might not provide representative samples of their customer's demographics. For the remaining venues, we looked at the gender distribution of reviewers; If over 70% of reviewers were male or female,

then the venue would be annotated with its respective dominant gender. After applying this step, 1,057 venues were labeled as male and 483 venues were labeled as female. The total number of gender annotated venues we have generated was 1,540 from a total of over 73,000 venues.

We have relied on a three-step method to gender annotate mainly Arabic names from geotagged tweets. Here is a description of our method:

- (1) Preprocessing: Initial cleaning to prepare names for annotation.
 - (a) Replace specific characters with a simpler variant (see Table 1).

Table 1. Replaced characters

| | | | | | | | |
|-----------------------|---|---|---|---|---|---|---------|
| Replaced characters | أ | إ | أ | و | ى | ة | _ -/. |
| Replacement character | ا | | | و | ي | ه | (space) |

- (b) Remove any characters not in the basic English and Arabic alphabets.
- (2) Annotation: Determines the gender of a name if a set of tests return exactly one unique gender (excluding null results).
 - (a) Split the name by spaces and check the database for a match, beginning with the first substring and adding more substrings if a match is not found.
 - (b) Check if the first name begins with certain prefixes (Table 2).

Table 2. Sample of prefixes

| | | |
|---------|-----------------|-----|
| Arabic | أبو | عبد |
| English | Abu (Father of) | Abd |

- (c) Check the name for certain keywords that indicate gender (Table 3).

Table 3. Sample of keywords

| | | | | | |
|---------|-----|------|----------------|--------------------|--------------|
| Arabic | سيد | سيدة | ام | بنت | بن |
| English | Mr | Mrs | Om (Mother of) | Bint (Daughter of) | Bin (Son of) |

- (3) Refinement: After running the annotation, names that were not annotated were exported and the most frequent 10–15% were manually labelled and added to the database.

Although this method can be applied to annotate any name in the database regardless of its language, it also leverages identifiers that are unique to Arabic since language-specific features can significantly improve the rate of annotation [16]. The database of names was initially populated from a list of over 1000 gender-annotated Arabic names with both Arabic and English spelling variants retrieved from the website Behind the Name [24]. Using the initial database, the method could identify the gender of approximately 34% of geotagged tweets, almost 75% of which were male. Although the database was further expanded by including names from various sources such as

university registries, we found that the refinement step was the fastest way to increase our method's performance.

After applying the refinement step, the method could annotate 41% of tweets primarily by identifying a larger number of female tweets. Although the gender balance was improved, female users still represented only 35% of all geotagged tweets. Previous studies have estimated that Twitter is more female biased [20, 25] but due to cultural norms, we suspect that women in Saudi Arabia are more likely to have private accounts. As a result, users of social media platforms are unlikely to be representative of the population.

To reduce the rate of false positives in our results, gender-ambiguous names were removed from the database whenever they were identified. These included names that apply to both men and women (e.g. 'Sam'), or names that become indistinguishable when written in English (e.g. 'Alaa', which can refer to both the male علاء or the female علا). The quality of the results relies on the accuracy of self-declared names entered by Twitter users. Although this source of error is difficult to control for due to the nature of social media data, we checked the profiles of a random sample of annotated tweets to manually identify the user's gender (for example by looking at profile pictures and Arabic pronouns they use) and found the names generally appeared to be a reliable indication of the user's gender and rarely produced false positives.

4 Gendered Spaces

Prior literature has proposed that “all space is already gendered: we live, experience and in some cases have reinforced for us, our gendered identities in particular built environments which assume certain things already about codes of gender” [26]. In the context of Riyadh, such gender norms and characteristics are often enforced. The following section will explore how gender annotated data from Twitter and Foursquare can be used to identify Riyadh's gendered spaces and investigate regions of overlap between the results of the two datasets.

4.1 Spatial Distribution of Tweets and Foursquare Data

Method. To accurately reflect the raw data with minimal processing, we initially visualized the gender balance of Twitter and Foursquare data using grid-based heatmaps. The visualizations were displayed using Leaflet, an open-source JavaScript library commonly used to integrate geospatial data with interactive maps. The data points were aggregated using grid cells with side lengths of 0.01° by 0.01° (approximately 1×1 km) to determine the spatial distributions of genders based on Twitter (Fig. 1) and Foursquare data (Fig. 2). Using a higher resolution grid is generally preferred to reveal differences in gender balance but since the data is sparse, we found that using significantly smaller cells led to frequent gaps in the heatmap and a high sensitivity to outliers.

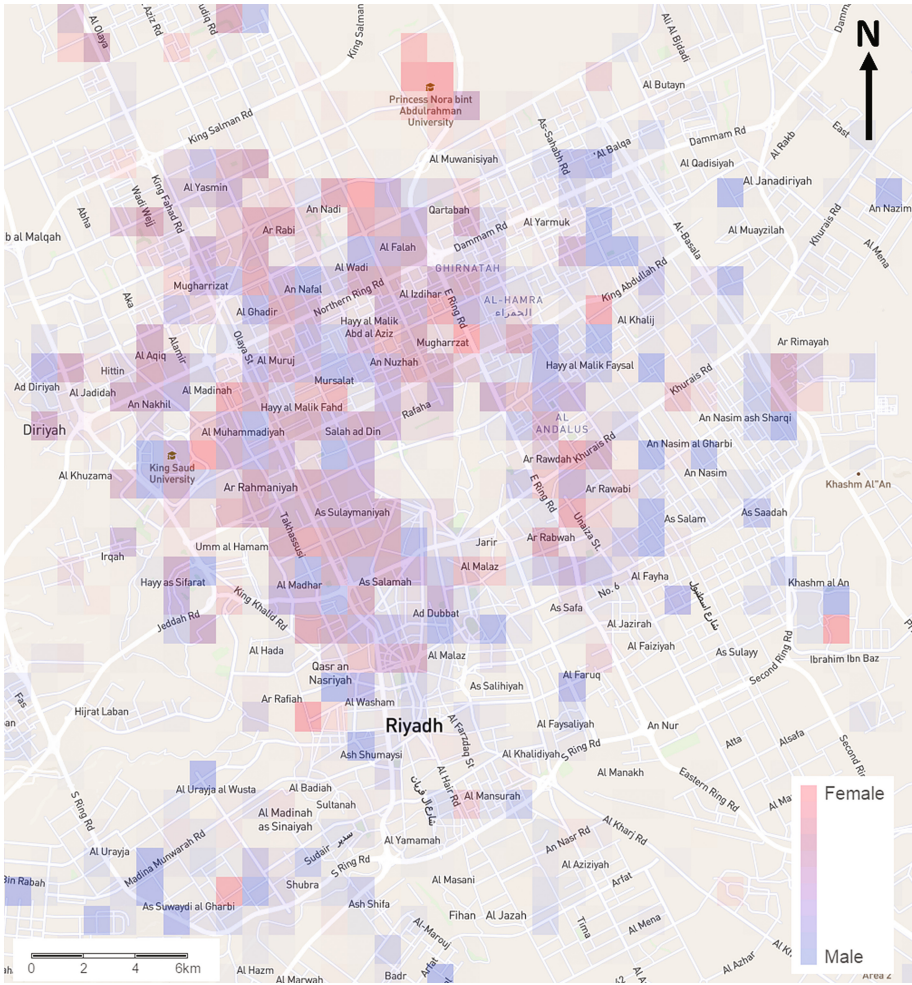


Fig. 1. Spatial distribution of gender annotated geotagged tweets (Color figure online)

The gender balance of cells in Figs. 1 and 2 is reflected in the color scale which ranging from red (female) to blue (male) with intermediate ratios shown in purple tones. The opacity of a cell varies based on the number of geotagged tweets or check-ins within that cell. In Fig. 1, the color intensity increases proportionally to the number of geotagged tweets up to a maximum of 250 tweets per cell. Since the number of reviews is not necessarily proportional to the number of visits, our analysis of Foursquare data used reviews only to determine the gender ratio. The ratio was used to distribute the check-ins between male and female users, resulting in a total of nearly 713,000 and 1,915,000 check-ins for women and men respectively. In Fig. 2, the color intensity increases proportionally to the number of check-ins up to a maximum of 20,000 check-ins per cell.

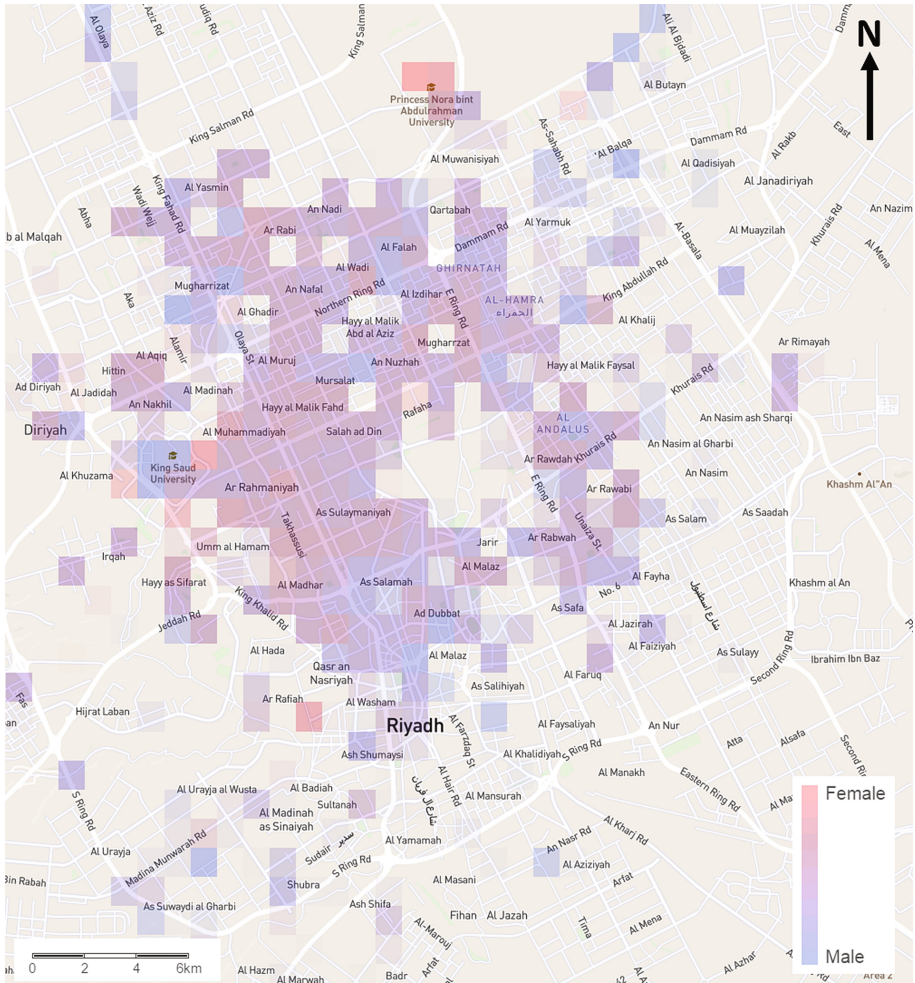


Fig. 2. Spatial distribution of gender-oriented venues weighted by check-ins (Color figure online)

Results. A few trends are apparent from the raw data visualization in Figs. 1 and 2. Since Twitter and Foursquare activity appears to correlate with population density, activity drops off in the more sparsely populated northern zone and southeastern zones. Additionally, a large gap in the middle of the city corresponds to a military air base with very little commercial and residential activity nearby. The level of social media activity in certain areas may also be affected by demographic factors such as age, nationality and socioeconomic status or a lack of amenities that tend to appear on Foursquare.

In addition, major segregated landmarks reflect their imposed gender segregation. For example, the grid cells in the northeast quadrant that cover the female-only Princess Noura University appear prominently in red. Likewise, the King Fahad stadium and the male campus of King Saud University were identified as male spaces. Furthermore,

areas of high activity tend to emerge around busy commercial areas in the center of the city. Commonly visited shopping districts located around Al-Olaya and King Fahad roads are predominantly female spaces. We also observed that female activity on Twitter is generally more concentrated in the center of the city whereas male activity is generally more spread out. The difference was measured using the spatial dispersion (Eq. 1) which was approximately 20% higher for male tweets (0.119°) than female tweets (0.099°).

$$\sqrt{\frac{\sum_{i=1}^N (x_i - x_{centroid})^2 + (y_i - y_{centroid})^2}{N}} \tag{1}$$

Since the data is sparse, this approach limits in the resolution that can be used in practice. Without applying a smoothing method to the data, outliers can easily cause individual cells to conflict with the gender balance of the surrounding area. The sharp transitions between values and sensitivity to outliers make it difficult to identify large regions that demonstrate a consistent gender imbalance. These limitations were addressed by using the smoothing method described in Sect. 4.2.

4.2 Identifying Gendered Spaces from Social Media Data

Method. The Twitter and Foursquare data both capture the locations visited by a small sample of the population. The ability to perform high-resolution spatial aggregation is limited by the sparseness of these datasets since the values at different locations are highly sensitive to outliers. This results in sharp transitions between one grid cell to another and tends to limit the resolutions that can be used in practice. These issues can be minimized by applying a Gaussian filter to each data point, a method that consists of defining values for a region surrounding each tweet or venue by defining a Gaussian distribution that peaks at the original location (Fig. 3a).

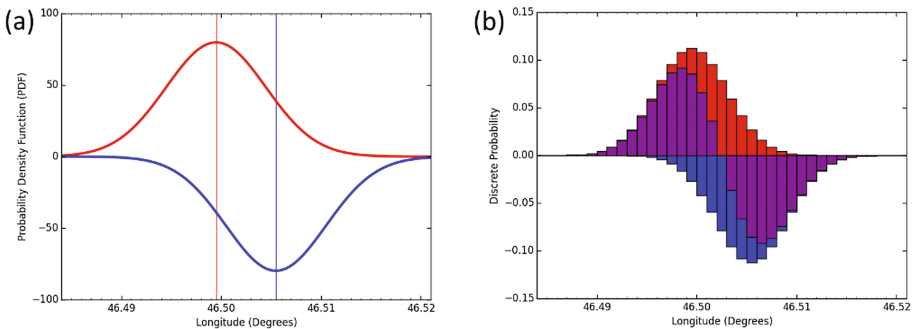


Fig. 3. (a) The results of a Gaussian filter applied to a female (red) and male (blue) location along the x axis with the exact location represented by thin vertical lines. (b) The discrete probability distribution of the same female (red) and male (blue) locations along with the distribution of the sum of all distributions (purple) (Color figure online)

The results were rasterized by discretizing the two spatial dimensions to cells with fixed widths that determine the resolution of the raster. The area under the curve of the probability density function was used to define a discrete probability value for each cell (Fig. 3b) along each dimension. The Gaussian filter is applied to each dimension independently and the value of each 2D grid cell is the product of the corresponding x and y values. This process is applied to each gender-annotated tweet and Foursquare venue to generate a raster using a sum of all the discrete probability distributions. The sum of the distributions were initially generated separately for males and females to produce a pair of aggregate distributions for both Twitter and Foursquare users. This methodology was used to define smooth gender distributions using the gender balance over a large area.

For the Twitter data, we generated an additional raster to represent the difference in gender concentration throughout the city. This was performed by using opposite signs

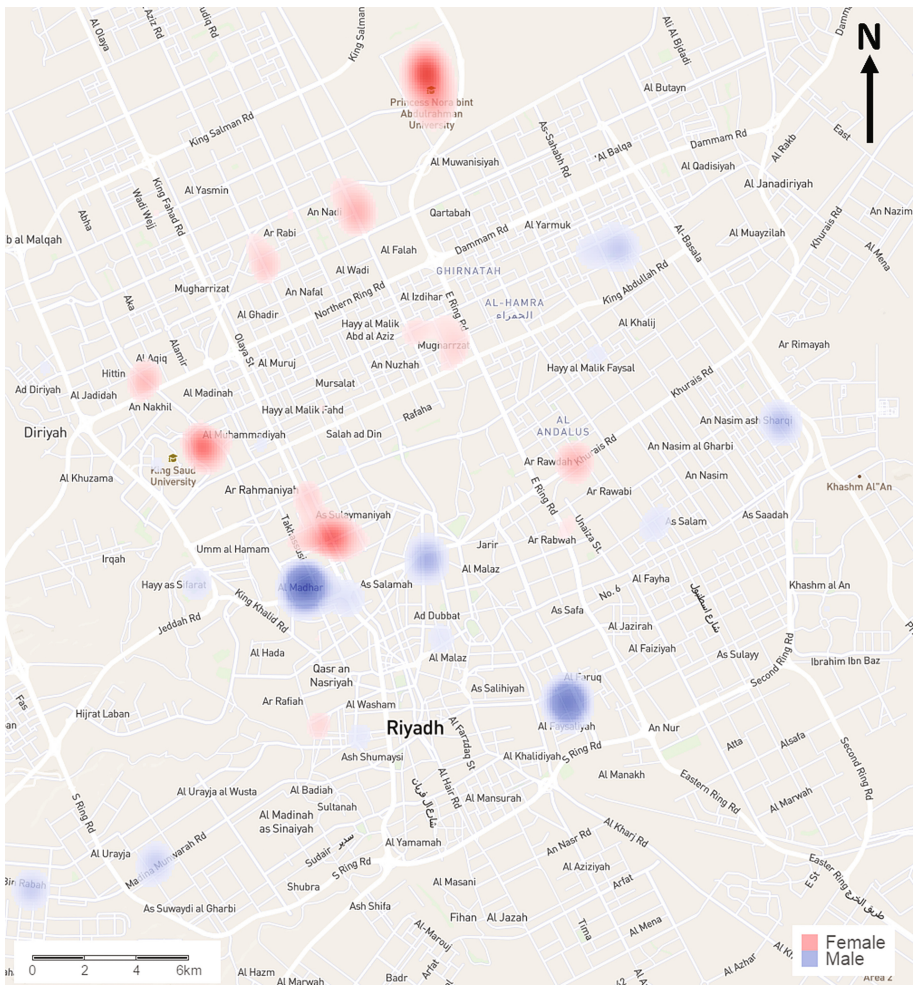


Fig. 4. The distribution of gendered spaces identified from Twitter data

for the genders and taking the sum of the distributions (Fig. 3b). This approach was used to distinguish areas with a high gender imbalance from areas that have a large number of tweets from both genders. This approach was not used with the Foursquare data since female-oriented venues accounted for less than 3.2% of all gender-oriented venues and the combined distribution was always dominated by male-oriented venues. As a result, our analysis was applied to each gender separately to identify areas of high concentration for male- and female-oriented venues.

Results. In Fig. 4, gender-annotated tweets were used to identify areas with a high gender imbalance. A cell width of 0.001 (approximately 0.1 km) was used for both axes to produce a rasterized distribution with a 2D spatial resolution of approximately 0.01 km². A threshold of ± 0.5 was used to isolate areas with a high gender imbalance.

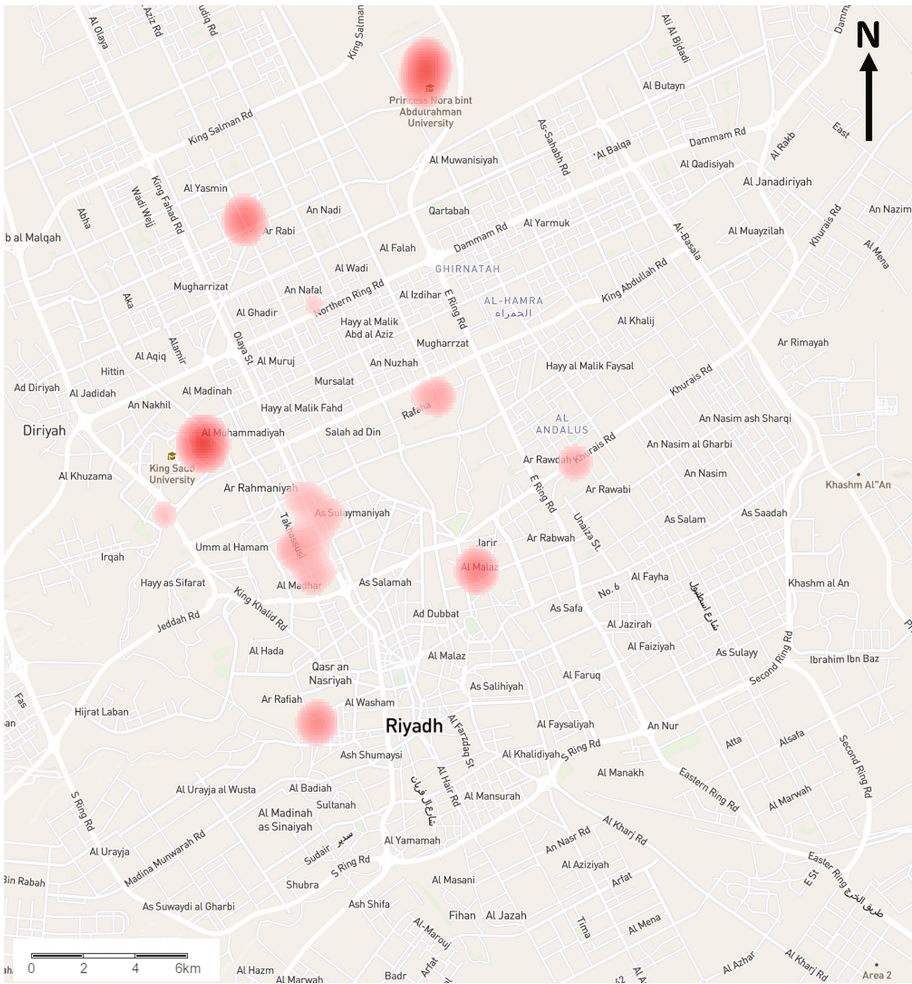


Fig. 5. The distribution of gendered spaces identified from female-oriented Foursquare venues

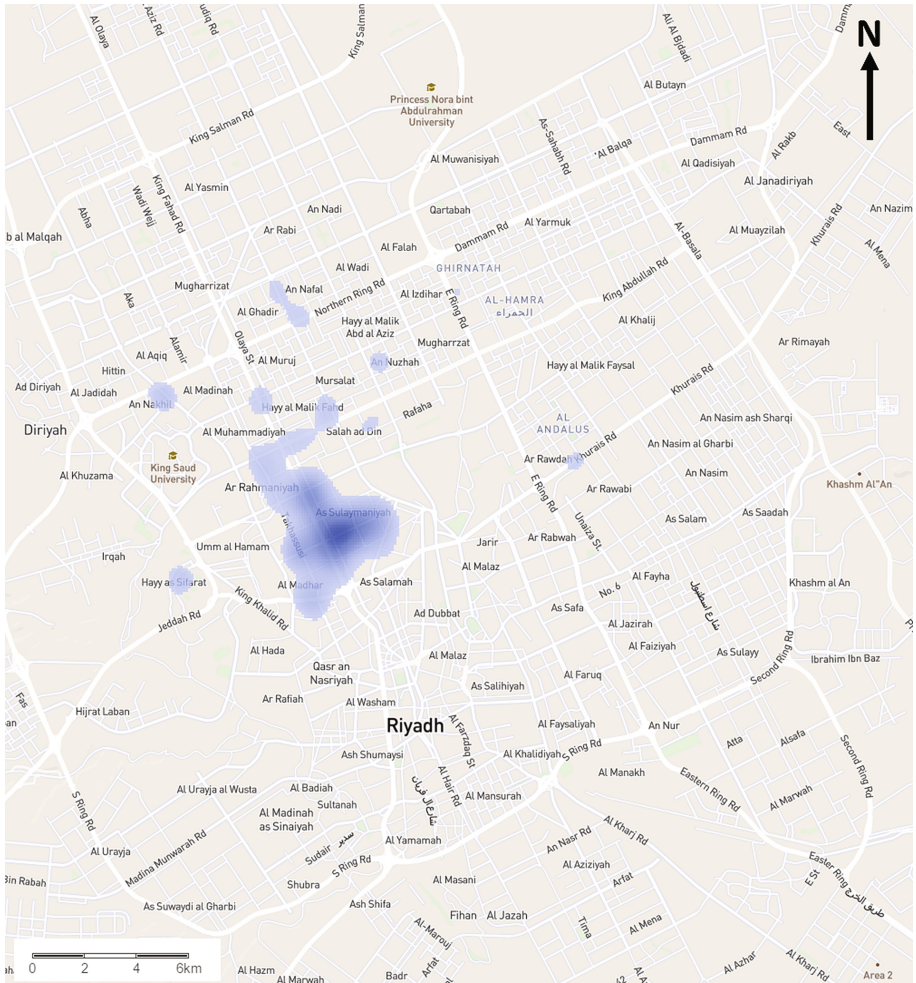


Fig. 6. The distribution of gendered spaces identified from male-oriented Foursquare venues

The smoothing process reveals more distinctive gendered spaces in the city, with some of the patterns from the raw data emerging still. First, as shown in Fig. 4, there is a distinct female cluster around the women's only Princess Noura University. Likewise, a female cluster appears on the female campus of King Saud University and a male cluster on the male campus of King Saud University. Female spaces also emerge in the previously mentioned Olaya commercial district and King Fahad road's shopping areas. A distinct male space emerges in the industrial Alfaruq and Alfaisaliyah areas. There are several factories located here that primarily employ men. Overall, it appears that male spaces are spread across a larger span of the city, whereas female spaces are located more closely to one another, Fig. 4.

Figures 5 and 6 reveal the highly-concentrated areas of gendered venues on Foursquare. Female-oriented venues (Fig. 5) appear more dispersed than the concentrations

of male-orientated venues. However, the most female concentrated clusters appear at the gender segregated campuses of Princess Noura and King Saud universities. There also seems to be a high density of male clusters (Fig. 6) around the commercial Olaya road area, this cluster appears in the twitter data but not as pronounced.

4.3 Overlapping Gendered Spaces from Tweets and Foursquare

To investigate the relationship between gendered spaces identified using Twitter and Foursquare data, this section will compare areas of overlap between the results of the two datasets. This will be examined quantitatively by computing the percent overlap (Table 4, Fig. 7) and qualitatively by identifying the different regions of overlap (Figs. 8 and 9) and some of the popular venues within them.

The results in Table 4 suggest that areas visited by female Twitter users in Riyadh generally tend to exhibit a significantly higher correlation with Foursquare venues in comparison to the correlation between the areas visited by male Twitter users and Foursquare venues. Although this pattern holds for areas with a high concentration of

Table 4. A comparison of the top 2,500 cells with the highest gender concentration from each data set showing the number of overlapping cells and their percent of the total area

| Twitter | Foursquare | Area of overlap (number of cells) | Percent overlap of total area |
|---------|------------|-----------------------------------|-------------------------------|
| Female | Female | 1,230 | 32.63% |
| Female | Male | 549 | 12.33% |
| Male | Male | 331 | 7.09% |
| Male | Female | 170 | 3.52% |

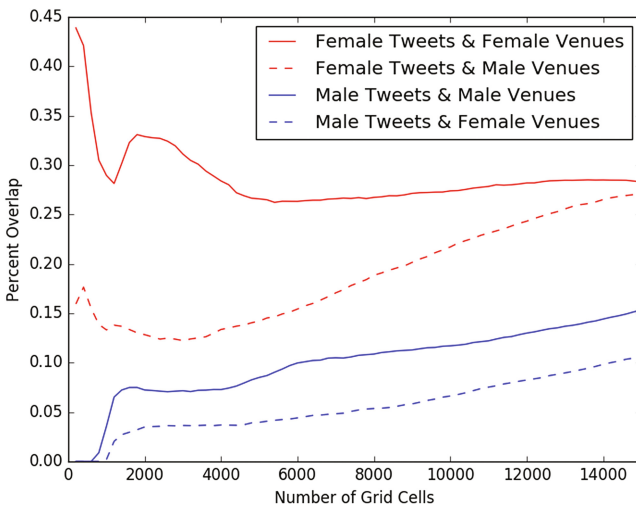


Fig. 7. The percent overlap of gendered spaces for Twitter and Foursquare data when different numbers of grid cells are used to define regions with the highest concentration

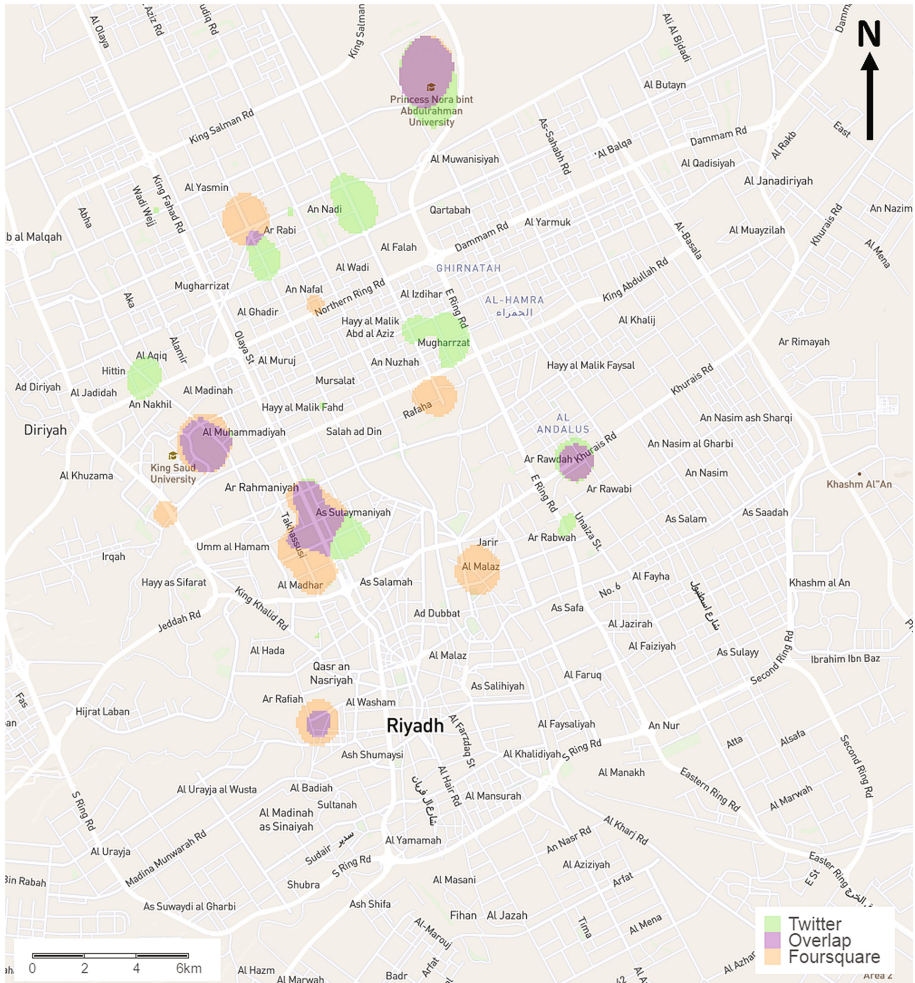


Fig. 8. Overlap of Foursquare and Twitter female spaces

venues regardless of the venues’ targeted gender, areas with a high concentration of female tweets tend to overlap more with female venues and areas with a high concentration of male tweets tend to overlap more with male venues.

The findings in Table 4 and Fig. 7 suggest that gendered spaces for women in Riyadh tend to form in highly clustered venue spaces. Female tweets are most strongly correlated with areas containing a high concentration of female-oriented venues but they also exhibit a significant correlation with male-related venues. By contrast, areas with a high concentration of male tweets exhibit a much lower rate of overlap with venues regardless of gender. This suggests that areas of male activity in Riyadh are less likely to be driven by the density of venues. Since women in Riyadh are not allowed to drive and do not have any currently available public transportation options, it is



Fig. 9. Overlap of Foursquare and Twitter male spaces

possible that constraints on female mobility could lead women to prefer co-located venues and encourage venues that target women to form clusters.

The percent overlap between the areas of highest concentration changes depending on the coverage of the city, measured by the number of grid cells considered. However, the order of the results remains consistent across a broad range of coverage values. The overlap between female tweets and female venues is greatest when only the top 2,000 grid cells are included and stabilizes at approximately 30% as additional cells are added. The rate of overlap for female tweets and male venues gradually increases with the number of cells and converges with the overlap of female venues when the top 15,000 cells are considered. The percent overlap of male tweets and venues of either gender also increases but remains significantly lower regardless of the degree of coverage.

For consistency, the following results in Figs. 8 and 9 are based on the top 2,500 grid cells for each data set.

Application of this method resulted in six areas of overlap for females (Fig. 8). Overlap emerges at three female university campuses: Princess Noura University and King Saud University's primary and preparatory campuses. Both data sets also identified a high concentration of women at a commercial area along Khurais Road that includes popular stores, cafes and an all-women's gym. One of the largest areas of overlap was found along King Fahad Road and Olaya Street in the city's primary commercial hub which contains several shopping malls, popular restaurants and cafes. A small area of overlap was also found on King Abdulaziz Road in an area containing schools, restaurants and a private tertiary care center (the Kingdom Hospital).

A correlation between male space identified by Twitter and male-oriented Four-square venues was only observed at two locations (Fig. 9). The first was the Al-Madhar area near the city's commercial hub, a region with substantial number of popular restaurants and coffee shops. A small area of overlap was observed in the Diplomatic Quarter (Hayy As-Sifarat). This neighborhood is home primarily to foreign embassies along with popular parks where sporting activities and social events are held.

5 Discussion

When applied to data from social media platforms, the methodology described above can help identify the impacts of Saudi Arabia's unique gender segregation policy on Riyadh's built environment and urban dynamics. The initial heatmap visualizations of the raw Twitter and Foursquare datasets displayed regions of gender imbalance throughout the city but didn't identify broad gendered spaces. A Gaussian filter was used to identify agglomerations with smooth transitions between values to identify gendered spaces from both Twitter activity and Foursquare venues. This analysis accurately picked up some of Riyadh's gendered landmarks such as male and female university campuses, male-only sporting venues, male- and female-oriented commercial areas and male-oriented industrial areas.

Furthermore, a comparison of the two datasets revealed the extent to which Twitter gendered spaces are correlated with single-gender oriented Foursquare venues. Areas with a high concentration of female tweets were more correlated with female-oriented venues than concentrations of male tweets were correlated with male-orientated venues. For female spaces, the overlap between the two datasets was 33%, whereas for males the overlap was less than 10%. This finding is potentially an indication that limits on women's mobility in Riyadh encourages the clustering of venues that cater specifically to them. Further testing of this hypothesis is necessary and might be explored using additional data sets that could identify the relationship between gendered spaces and mobility patterns.

The contributions of this study for future research are twofold: improvements to gender annotation for Arabic names and a new methodology to identify gendered spaces from social media data. This methodology could be combined with the advances in Arabic sentiment analysis [25] to analyze tweet content as a barometer for public opinion on key social or policy issues. There is also potential for utilizing gender

annotated social media data in a comparative study with other Saudi Arabian cities and neighboring Gulf countries which share a similar culture but do not impose the same gender segregation policies as Saudi Arabia. Comparing the gendered spaces results with other results could give more insight on the implications of mobility limitation on women and gender segregation.

Due to the unique cultural context in Saudi Arabia, there is a high degree of spatial segregation between the destinations of females and males in Riyadh. Understanding the distribution of Riyadh's gendered spaces is meaningful for city planning, and specifically for the transit planning associated with the ongoing development of Riyadh's first public transit system. The system, which will include rail and bus, is expected to be the largest mass-transit system built from scratch when it opens in late 2018 [25].

Currently, Riyadh faces serious traffic congestion challenges [29]. The new metro system could alleviate congestion and lessen the dependence of the city's 6 million residents on fuel consuming vehicles. The metro also has the potential to alter women's dependence on male drivers and taxi services. Given Saudi Arabia's policy that prohibits women from driving, women are expected to be a significant user population of the metro [30–33]. By identifying gendered spaces, this paper takes the first step toward a nuanced analysis of gendered mobility in Riyadh. Future work could assess the impact of the metro system on women's transit to these places or design bus routes specifically for women based on the distribution of gendered spaces.

6 Conclusion

The varying roles and division of labor between genders appear at various scales in urban planning and design - individual buildings, neighborhoods, cities, and regions - and in the different domains of city design, such as housing, public facilities, transportation, streets and open space, employment and commercial/services areas. In this paper, we presented an overview of an innovative use of social media data to identify gendered spaces in the city of Riyadh. Utilizing 51,000 of geotagged tweets that were then gender annotated, we identified gendered spaces in Riyadh and opportunities for future research on gendered spaces and mobility in Riyadh and other cities, especially where social media users are Arabic speakers. The outcomes of this research are intended to allow researchers in urban design and city planning to understand how Saudi Arabia's social policies shape urban dynamics.

References

1. Uteng, T.: Gender and mobility in the developing world. World development report, 7778105-1299699968583 (2012)
2. Aldalbahi, M., Walker, G.: Riyadh transportation history and developing vision. *Procedia-Soc. Behav. Sci.* **216**, 163–171 (2016)
3. Sandercock, L., Forsyth, A.: A gender agenda: new directions for planning theory. *J. Am. Plann. Assoc.* **58**(1), 49–59 (1992). doi:[10.1080/01944369208975534](https://doi.org/10.1080/01944369208975534)

4. Spain, D.: Gendered spaces and women's status. *Sociol. Theor.* **11**(2), 137 (1993). doi:[10.2307/202139](https://doi.org/10.2307/202139)
5. Hanson, S.: Gender and mobility: new approaches for informing sustainability. *Gend. Place Cult.* **17**(1), 5–23 (2010)
6. McDowell, L.: Towards an understanding of the gender division of urban space. *Environ. Plann. D Soc. Space* **1**(1), 59–72 (1983)
7. Spain, D.: *Gendered Spaces*. University of North Carolina Press, Chapel Hill (1992)
8. Loukaitou-Sideris, A.: A gendered view of mobility and transport: next steps and future directions. *Town Plann. Rev.* **87**(5), 547–565 (2016)
9. Crane, R.: Is there a quiet revolution in women's travel? Revisiting the gender gap in commuting. *J. Am. Plann. Assoc.* **73**(3), 298–316 (2007)
10. Law, R.: Beyond “women and transport”: towards new geographies of gender and daily mobility. *Prog. Hum. Geogr.* **23**(4), 567–588 (1999)
11. Cresswell, T., Uteng, T.P.: *Gendered mobilities: towards an holistic understanding*. *Gendered Mobilities*, pp. 1–12. Aldershot, England (2008)
12. Globalmediainsight.com: Saudi Arabia social media statistics 2016 - official GMI blog (2017). <http://www.globalmediainsight.com/blog/saudi-arabia-social-media-statistics/>. Accessed 27 Feb 2017
13. SemioCast: SemioCast—Geolocation of Twitter users, July 2012. http://semioCast.com/en/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US. Accessed 27 Feb 2017
14. Liu, W., Al Zamal, F., Ruths, D.: Using social media to infer gender composition of commuter populations. In: *Proceedings of the When the City Meets the Citizen Workshop, The International Conference on Weblogs and Social Media, May 2012*
15. Bergsma, S., Dredze, M., Van Durme, B., Wilson, T., Yarowsky, D.: Broadly improving user classification via communication-based name and location clustering on Twitter. In: *HLT-NAACL*, pp. 1010–1019 (2013)
16. Ciot, M., Sonderegger, M., Ruths, D.: Gender inference of Twitter users in non-english contexts. In: *EMNLP*, pp. 1136–1145, October 2013
17. Al Zamal, F., Liu, W., Ruths, D.: Homophily and latent attribute inference: inferring latent attributes of Twitter users from neighbors. In: *ICWSM*, p. 270 (2012)
18. Culotta, A., Kumar, N.R., Cutler, J.: Predicting Twitter user demographics using distant supervision from website traffic data. *J. Artif. Intell. Res. (JAIR)* **55**, 389–408 (2016)
19. Mislove, A., Lehmann, S., Ahn, Y.Y., Onnela, J.P., Rosenquist, J.N.: Understanding the demographics of Twitter users. In: *ICWSM*, vol. 11, p. 5 (2011)
20. Burger, J.D., Henderson, J., Kim, G., Zarrella, G.: Discriminating gender on Twitter. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*, pp. 1301–1309, July 2011
21. Liu, W., Ruths, D.: What's in a name? Using first names as features for gender inference in Twitter. In: *AAAI Spring Symposium: Analyzing Microtext*, vol. 13, p. 1, March 2013
22. Zafarani, R., Tang, L., Liu, H.: User identification across social media. *ACM Trans. Knowl. Discov. Data (TKDD)* **10**(2), 16 (2015)
23. Novak, J., Raghavan, P., Tomkins, A.: Anti-aliasing on the web. In: *Proceedings of the 13th International Conference on World Wide Web*, pp. 30–39. ACM, May 2004
24. Campbell, M.: Behind the name: meaning of names, baby name meanings (2017). *Behindthename.com*. <http://www.behindthename.com/>. Accessed 27 Feb 2017
25. Heil, B., Piskorski, M.: New Twitter research: men follow men and nobody tweets. *Harvard Bus. Rev.* **1**, 2009 (2009)

26. Long, J.: 'There's no place like home?': domestic life, gendered space and online access. In: Proceedings of the International Symposium on Women and ICT: Creating Global Transformation, p. 12. ACM, June 2005
27. alOwisheq, A., alHumoud, S., alTwaresh, N., alBuhairi, T.: Arabic sentiment analysis resources: a survey. In: Meiselwitz, G. (ed.) SCSM 2016. LNCS, vol. 9742, pp. 267–278. Springer, Cham (2016). doi:[10.1007/978-3-319-39910-2_25](https://doi.org/10.1007/978-3-319-39910-2_25)
28. Varinsky, D.: Saudi Arabia's Riyadh metro will be largest urban-transit system ever built from scratch. <http://www.businessinsider.com/saudi-arabias-riyadh-metro-will-be-the-largest-urban-transit-system-ever-built-from-scratch-2016-5>. Accessed 20 Feb 2017, 20 May 2016
29. Chodrow, P.S., Al-Awwad, Z., Jiang, S., González, M.C.: Demand and congestion in multiplex transportation networks. PLoS ONE 11(9), e0161738 (2016). doi:[10.1371/journal.pone.0161738](https://doi.org/10.1371/journal.pone.0161738)
30. Beasley, K.: For Saudi women, new subway will mean more than a cool ride. <http://www.npr.org/sections/parallels/2013/07/30/207077269/for-saudi-women-new-subway-will-mean-more-than-a-cool-ride>. Accessed 20 Feb 2017, (n.d.)
31. Kirk, M.: What a new metro in Saudi Arabia will mean for women. <http://www.citylab.com/commute/2017/02/what-riyadhs-new-metro-will-mean-for-women/515859/>. Accessed 20 Feb 2017, 7 Feb 2017
32. Winter, C.: For women, saudi arabia's new metro may mean greater mobility. Bloomberg.com. <https://www.bloomberg.com/news/articles/2013-08-16/for-women-saudi-arabias-new-metro-may-mean-greater-mobility>. Accessed 16 Aug 2013
33. Asmahan Alghamdi: المرأة تنتظره لأسبابها الخاصة «المترو». <http://www.alriyadh.com/854435>. Accessed 24 July 2013