11-1-2018

Hidden in Plain Sight: A Machine Learning Approach for Detecting Prostitution Activity in Phoenix, Arizona

Edward Helderop

Jessica Huff

Fred Morstatter

Anthony Grubesic

Danielle Wallace

Hidden in Plain Sight: A Machine Learning Approach for Detecting Prostitution Activity in Phoenix, Arizona

Edward Helderop[1], Jessica Huff[2], Fred Morstatter[3], Anthony Grubesic[1] & Danielle Wallace[2]

1 Center for Spatial Reasoning & Policy Analytics, College of Public Service & Community Solutions, Arizona State University, Phoenix, AZ 85004, USA

2 School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ 85004, USA

3 Information Sciences Institute, Viterbi School of Engineering, University of Southern California, Marina del Rey, Los Angeles, CA 90292, USA

Abstract

Prostitution has been a topic of study for decades, yet many questions remain about where prostitution occurs. Difficulty in identifying prostitution activity is often attributed to the hidden and seemingly victimless nature of the crime. Despite numerous challenges associated with policing street prostitution, these encounters become more difficult to identify when they take place indoors, especially in locations away from public view, such as hotels. The purpose of this paper is to develop a strategy for identifying hotel facilities and surrounding areas that may be experiencing elevated levels of prostitution activity using high-volume, user-generated data, namely hotel reviews written by guests and posted to Travelocity.com. A unique synthesis of methods including data mining, natural language processing, machine learning, and basic spatial analysis are combined to identify facilities that may require additional law enforcement resources and/or social/health service outreach. Prostitution hotspots are identified within the city of Phoenix, Arizona and policy implications are discussed.

Keywords

Natural language processing, Machine learning, Prostitution, Spatial analysis, Phoenix

Introduction

Although prostitution is frequently studied, and has been for years, there remain a paucity of studies that investigate the locations at which prostitution occurs (McCutcheon et al. 2015). At face value, prostitution may seem like a victimless crime, where prostitutes and their customers consent to engage in illegal behavior. However, recent work suggests that prostitution and sex trafficking are strongly linked in the

United States (Kempadoo et al. 2015; Farley 2007; Heiges 2009).[1] To be clear, not all prostitutes are victims of human trafficking, nor are all victims of human trafficking forced into prostitution, but there is an underlying co-dependency at work (Butcher 2003; Kempadoo et al. 2015). Furthermore, elements of victimization for prostitutes, who are at considerable risk of violence from customers, are real (Weitzer 1999). From the perspective of prevention and law enforcement, unless the act, solicitation, or advertisement of prostitution is visible and occurring in public, it is rarely reported or effectively policed (Weitzer, 1999). As a result, much of the existing research on prostitution has largely focused on street activity.

The problem with this narrow focus is that prostitution occurs both publicly and privately. When done in public, prostitutes ply their trade visibly (often on the street) to acquire customers. When done in private, prostitution can occur virtually anywhere. However, much of this private activity takes place in hotels. As detailed by Weitzer (2000a), hotels are frequently complicit in the prostitution trade because of the financial benefits associated with the patronage of prostitutes and their clients. Local hotels make ideal locations to engage in prostitution for several reasons. First, they provide quick access to private, comfortable spaces where sexual transactions can take place. second, most hotels have basic security resources that help to minimize (or at least mitigate) violent behavior. Finally, hotels can be used discreetly, because vehicular and/or foot traffic into and out of the facility are the norm.

The use of non-public facilities for prostitution combined with most customers' preferences for discrete transactions limits the information available to law enforcement agencies, social workers, and health professionals about prostitution activity and its participants. As a result, in locations where prostitution is a crime targeted for enforcement, this lack of information means that agencies often misdirect resources and policing efforts to combat prostitution (Weitzer 2010). In communities where prostitution is not targeted for enforcement, efforts to provide social services or healthcare support for prostitutes can also be misdirected and/or inefficient (Jeal and Salisbury 2007; Harcourt et al. 2010).

Recent work, which leverages the power of spatial analytics blended with crowd-sourced, publicly-available datasets has improved basic capabilities for identifying crime hotspots and holds great promise for deepening our understanding of prostitution activity. For general crime applications, these data can include mobile phone check-in data taken from cell towers (establishing anonymous footfall data), as well as enormous amounts of textual data harvested from Twitter. These datasets can be used to build models that identify existing crime hotspots or predict their likely future locations. Bogomolov et al. (2014), for example, used cell tower footfall data and historical crime data to predict future crime locations in London. Other studies incorporate textual data.

1 For a more thorough review on this topic, readers should consult Gozdziak and Collett (2005) and Schauer and Wheaton (2006).

Likely hit-and-run incident locations were identified using georeferenced Twitter data (Wang et al. 2012). Similarly, Gerber (2014) and Malleson and Andresen (2015) built models that predict crime hotspots for dozens of major crimes in US cities that automatically harvest and parse Twitter data.

The purpose of this paper is to develop a more efficient strategy for identifying facilities and/or areas that may be experiencing elevated levels of prostitution activity, away from the public eye. To accomplish this, we detail a methodological framework that incorporates data mining, natural language processing (NLP), machine learning, and spatial analytics to provide insight on the spatial footprint of hotel-based prostitution activity in Phoenix, Arizona. Specifically, we use a large database of Travelocity hotel reviews, written by guests, to build a classifier that is predictive of prostitution activities in/around hotels throughout the city.

This work is important for several reasons. First, in addition to identifying locations where prostitution is occurring behind closed doors, it simultaneously identifies locales where crucial public health and/or social service outreach activities can be targeted to support a vulnerable population. Second, in communities where prostitution is policed, the developed detection system can help to better allocate limited law enforcement resources to mitigate prostitution activity. Finally, the developed detection system can also help inform community policies related to nuisance facilities and how to minimize their impacts to surrounding neighborhoods (Farber 1998; Mansfield et al. 2001; Prior and Croft 2012).

Background

Prostitution and Community Standards

By definition, prostitution refers to the practice or occupation of engaging in sexual activity with someone for payment (Oxford 2018). In the United States prostitution is usually illegal, although community standards associated with prostitution vary considerably. For example, in the state of Nevada, prostitution is permitted in rural counties, geographically removed from the state's population centers (Fig. 1). Although this is the only location in the United States where prostitution is allowed, there are considerable variations in how prostitution is penalized and how it is viewed by residents at the community level.[2] For example, in relatively liberal locales, such as San Francisco, CA and Berkeley, CA, recent ballot measures have suggested that a fairly large majority of residents wanted law enforcement agencies to deprioritize the policing of prostitution. Specifically, the 2008 ballot measure in San Francisco, which stipulated

2 Consider, for example, the variations in penalties for prostitutes between Arizona and Texas. In Arizona, there are no financial penalties, but the first, second, third offenses and fourth offenses confer 15, 30, 60 and 180 days of jail time, respectively. In Texas, the first, second and third offenses confer up to 180 days and/or $2000, 1 year and/$4000, 2 years and/or $10,000, respectively (Procon 2017). There are equally diverse penalties for customers, pimps and brothel owners across all 50 states.

that police would discontinue enforcing all laws against prostitution, received support from 42% of the voters (Weitzer 2009). In 2004, 36% of voters in Berkeley voted for police to give prostitution enforcement the lowest policing priority (Weitzer 2009). Obviously, there are significant differences between decriminalization and legalization, but these local efforts are certainly indicative of divergent attitudes toward prostitution throughout the United States. This is also evident in the highly variable penalties for engaging in the sex trade at the state level (Procon 2017).

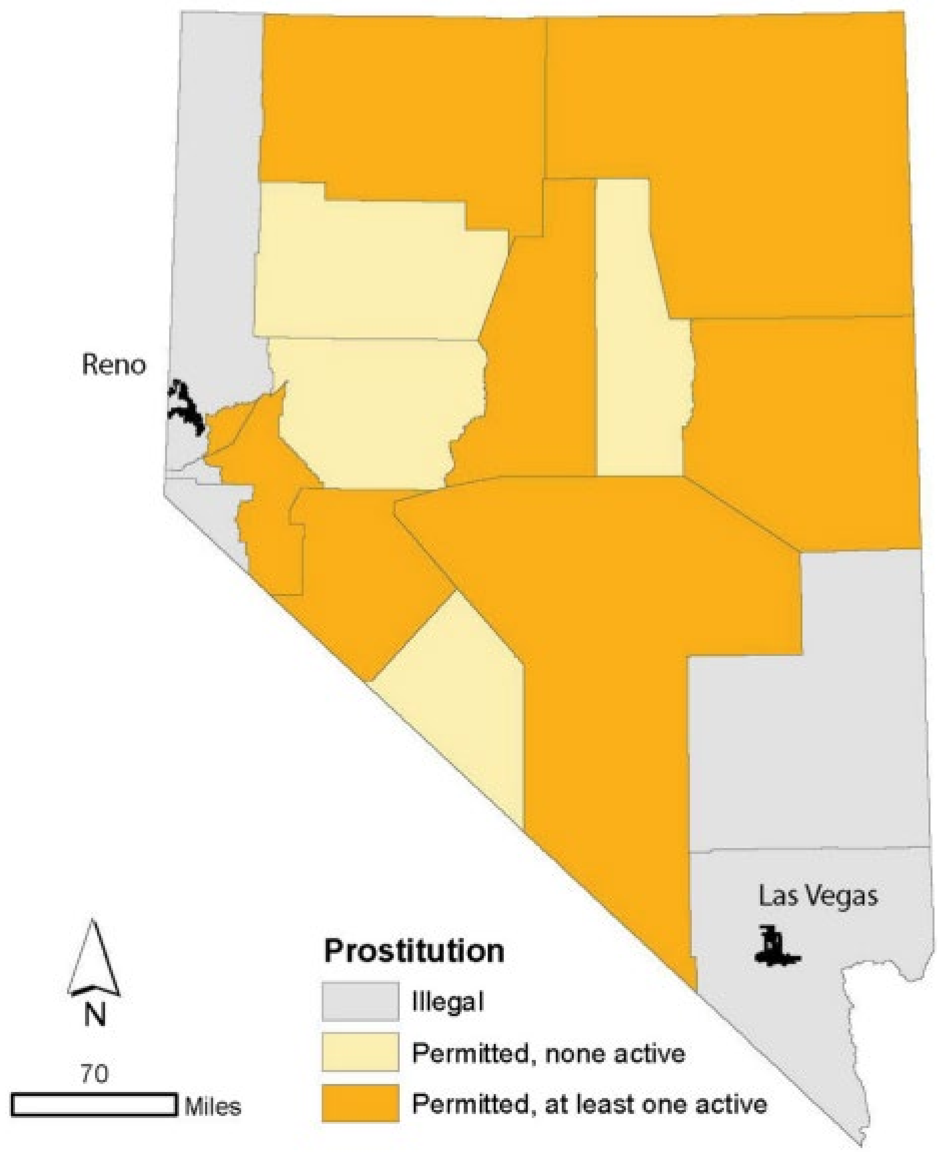

Figure 1. Prostitution and brothel activity in the state of Nevada. These are the only location where prostitution is legal in the United States.

There is no single, clear-cut reason as to why there is such divergence between states and the criminal penalties they levy against prostitutes, clients of prostitutes, pimps, and brothel owners. However, Weitzer (2010) argues that moral crusades against prostitution have been exceedingly effective in many places. For example, in the

United States, Weitzer (2010) suggests the efforts against trafficking and sex work is dominated by a coalition of the religious right and abolitionist feminists.[3] Given the cultural and religious diversity of the U.S., it is no surprise that the presence and influence of these groups varies geographically. For more details and a snapshot of these groups across states, readers should consult the recent collaboration between the Washington Post and the Public Religion Research Institute (Chokshi 2015).

Policing Prostitution

Considering that prostitution is illegal in all but one U.S. state, one obvious question concerns the size and significance of the sex trade in the U.S. In a recent report by Dank et al. (2014), the economics of this underground industry were made clearer. For example, in 2007, the sex trade in Atlanta was worth $290 million; similarly in Miami, where it was worth $235 million (Dank et al. 2014). In short, the fact that prostitution is illegal has not prevented it from becoming a $14.6 billion dollar industry in the United States (Havocscope 2015). Given the pervasive nature of the sex trade, why is policing prostitution so difficult? In part, it is challenging because of its pervasive nature, the diversity of the prostitution market, and the unintended consequences that accompany harsh police crackdowns (Sanders 2004; Weitzer 2000a). Street prostitution in particular is highly visible and is often subject to police attention when it occurs in areas that non-prostitution participants deem unacceptable (Krüsi et al. 2016; Larsen 1992).

Another outstanding challenge in policing prostitution is the difficulty in detecting and intervening in prostitution that occurs indoors – even though a majority of prostitutes may work this way (Cunningham and Kendall 2011; Benson and Matthews 2000; Matthews 2005; Weitzer 2000b). Indoor prostitution is more difficult for the police to identify and mitigate for several reasons. First, there is a low likelihood that either the prostitute or the client will seek assistance from law enforcement, even when the transaction does not go smoothly. Second, bystanders are also less likely to call law enforcement – residents or businesses in an area may only be aware of indoor prostitution rarely, and may not always contact police even when aware of it occurring (Cunningham and Kendall 2011; Benson and Matthews 2000; Prior and Croft 2012; Weitzer 2000b).

When interventions from law enforcement do occur, there can be a range of unintended consequences (Hubbard 1997; Lowman 2000; Sanders 2004; Weitzer 2000a). Crackdowns may displace prostitutes to new and potentially unsafe areas and encourage them to engage in unsafe practices to avoid police detection (Hubbard 1997;

3 In the U.S., the religious right refers to a loose coalition of evangelical Protestants, Latter Day Saints, Roman Catholics, and other groups that support socially conservative policies on prostitution, contraception, abortion and homosexuality. Weitzer (2010, 64) defines abolitionist feminists as those "who argue that the sex industry should be eliminated because of its objectification and oppressive treatment of women, considered to be inherent in sex for sale."

Larsen 1992; Sanders 2004; Weitzer 2000a). This can greatly exacerbate the existing vulnerabilities of prostitutes, who already suffer from high levels of victimization and violence (Barnard 1993; Church et al. 2001; Lowman 2000; Sanders 2004). Notably, many prostitutes working indoors are less likely to report being victimized when compared to those working on the street (Church et al. 2001).

From a spatial perspective, there are some systematic patterns associated with prostitution worth noting. Prostitution is generally concentrated or contained within geographic areas characterized by a revolving door of police crackdowns and prostitution arrests (Weitzer 2000a). That said, it is also true that the geographic accounts of prostitution activity strongly reflect where the police enforce laws, which laws are enforced, and the types of individuals the police choose to enforce them against (Hubbard 1997). Because of this variability, it is likely that no two cities have an identical footprint of prostitution activity.

It is also important to acknowledge that prostitution policies overwhelmingly involve criminalization, where prostitutes are viewed as offenders instead of victims (Barnard 1993; Day and Ward 2007; Krüsi et al. 2016; Larsen 1992; Sanders 2004; Weitzer 2000a). Criminalization alienates prostitutes from the protective potential of the police and other social services (Lowman 2000; Shannon et al. 2008). Thus, despite the illegality of prostitution throughout most of the United States and an aggressive criminalization strategy by many law enforcement agencies, prostitution continues to occur. To put this in perspective, somewhat dated estimates of law enforcement efforts to combat prostitution suggest that the price tag approaches $200 million annually for the U.S. (Weitzer 2000a; Matthews 2000). Again, much of this effort is dedicated to policing street prostitution, largely disconnecting law enforcement efforts from indoor activity.

Prostitution and Public Health

There are numerous public health implications associated with the underground sex trade, for both prostitutes and their clients, encompassing a range of physical, social and psychological issues. Willis and Levy (2002) note that child prostitution is a global health crisis, with nearly 10 million children actively prostituted and an estimated 1 million children forced into prostitution each year. From a physical perspective, the specter of sexually transmitted disease (STD) is a significant concern for prostitutes and their clients. From HIV/AIDS to syphilis and gonorrhea, working prostitutes and their clients face real risks when engaging in unprotected sex (Gil et al. 1996; Van den Hoek et al. 1989; Farley and Kelly 2000; Milrod and Monto 2016). It is important to note that both prostitutes and their clients are typically aware of the risks associated with unprotected sex, but research suggests that there is a significant link between the use of alcohol and the willingness to have unprotected sex for more money (Gossop et al. 1995). Physical violence is also a concern for working prostitutes. Farley and Barkan (1998) reported that 82% of the 130 prostitutes interviewed in San Francisco had been physically assaulted and 68% had been raped.

The physical violence faced by working prostitutes also contributes to mental health issues, including post-traumatic stress disorder (PTSD). PTSD severity was significantly associated with both assault and the number of times raped while engaged in prostitution (Farley and Barkan 1998). Depression is also a significant mental health concern for prostitutes (Chudakov et al. 2002). In fact, homeless youth involved in prostitution are at a greater risk for both depression and suicide (Yates et al. 1991).

Finally, the geographic locations in which prostitution enforcement is conducted can also impact prostitutes' access to important health services (Hubbard 1997; Shannon et al. 2008). For example, research suggests that prostitutes in Vancouver, BC actively avoided many locations with needle exchange programs because of a heavy police presence in those areas. This highlights the need to develop a deeper understanding of where these vulnerable populations reside (and work), and for the creation of strategies to develop public health interventions that reach their intended audience.

Prostitution and Community Impacts

Although the bulk of current law enforcement efforts directed at prostitution in the United States seek to reduce its visibility, the manner in which these laws are enforced can be uneven. Specifically, empirical evidence suggests that prostitution is more acceptable in some places than others. For example, street prostitution is often relegated to areas with low socioeconomic status, resulting in charges of class bias (Hubbard 1997; Larsen 1992). Moreover, because the bulk of street prostitution occurs in disadvantaged areas with intense social problems, many view prostitutes as a cause, rather than a symptom, of disorderly communities (Hubbard 1997).

Selective policing and the associated concentration of activity into subsets of urban space has fueled a range of responses at the community level. A core issue for street prostitution is whether or not it represents a genuine environmental nuisance for communities and their residents. Hubbard (1998) details the impacts of community protests against female street prostitution in Birmingham and Bradford (UK) and the factors driving these protests. In short, Hubbard's (1998) work suggests that "not-in-my-backyard" (NIMBY) sentiments, reflecting anxiety about street prostitution activity and its perceived connections to drugs, crime, degrading quality of life and local amenities, largely fueled community picketing. Interestingly, these community protests did not yield a reduction in street prostitution activity, they merely displaced it. Displacement and deflection will continue unabated without the ability to connect prostitutes with a viable exit strategy, something designed to help them explore alternative career paths (Hubbard 1998).

The indoor sex trade is also important to consider – the bulk of which occurs in hotels, private residences, strip clubs and massage parlors (Weitzer 2010). Interestingly, this indoor activity has been linked to the noxious facilities literature. While the bulk of work on noxious facilities is tied to environmental and economic concerns (e.g., pollution, property values) (Opaluch et al. 1993; Anstine 2003), mental health facilities, halfway houses, and sex premises are also considered noxious facilities (Grubesic and Murray 2008; Hubbard et al. 2013; Smith and Hanham 1981). In a study of sex premises in Sydney, Australia, Hubbard et al. (2013) found that these facilities did not pose a genuine nuisance to local communities and that distance from these premises were poor predictors of residents' experiences of nuisance. Their conclusion was that although many residents had moral objections to the activities taking place within the sex premises, they were not directly impacted (Hubbard et al. 2013). This is not to say, however, that the presence of busy sex premises in a neighborhood lacks impact. As detailed by McCleary and Weinstein (2009), there are secondary, ambient impacts in the neighborhood. Specifically, in a case study of Sioux City, Iowa, results suggested that when an off-site adult business opened, the ambient crime risk doubled when compared to a control area. Further, the risk of victimization drastically increased during the night-time hours.

In this context, the obvious problem for both law enforcement agencies and social/health service providers is that unlike strip clubs or related commercial sex premises, locations hosting indoor prostitution do not advertise this activity with flashing lights and neon signs, at least not in the United States.[4] As a result, there is little geographic evidence to inform agencies where to pursue intervention or interdiction strategies. As mentioned previously, these strategies will differ between communities, especially for those that either emphasize/do not emphasize the policing of prostitution.

It is also important to note that the nature of sex work is changing rapidly. As the internet continues to grow in popularity for soliciting and consuming sex (Cunningham and Kendall 2011), indoor prostitution is likely to grow. As detailed previously, prostitutes are often victimized by violent clients (Raphael and Shapiro 2004). However, research suggests that victimization is less likely among prostitutes working indoors than those working on the street (Church et al. 2001; Lowman 2000; Weitzer 2000b). In fact, Benson and Matthews (2000) note that most British vice squads responsible for enforcing prostitution offenses only address indoor prostitution in response to a complaint. Again, this means that much of this activity remains hidden from the public, law enforcement, and social service providers. In their study of massage parlors operating in Mesa, Arizona, Huff et al. (2018) found online user review forums to be a promising strategy to distinguish between legally operating massage businesses and those providing illicit sexual services.

Given the significant obstacles faced by law enforcement agencies, public health officials, and community groups for identifying indoor prostitution activity, this paper proposes a strategy and analytical framework for identifying hotels experiencing high

4 For more details on the impacts of red-light districts, prostitution, neighborhood effects and municipal regulation in Belgium, see Boels and Verhage (2016) and Prior and Croft (2012).

rates of prostitution. Although this is only one facet of the much larger underground sex industry in most urban areas, it is an important one given the range of potential issues it can present to prostitutes and customers (e.g., violence and victimization), neighborhood residents (crime, perceived decay of amenities), as well as social and health service providers looking to improve and enhance their community outreach efforts. In the next section, we provide details on the data and methods used for identifying hotel facilities where indoor prostitution might be taking place. This is followed by a case study Phoenix, Arizona. We conclude with a discussion of the results and policy implications.

Study Area, Data and Methods

Study Area

The city of Phoenix, Arizona serves as the study area for this analysis. According to recent Census (2017) estimates, Phoenix is the largest city in Arizona and is the fifth largest city in the United States, with a population of 1,615,017. The Phoenix metropolitan area, which includes the cities of Mesa, Scottsdale, and Chandler, is the twelfth largest metropolitan area in the U.S., with a population of 4.49 million (Census 2017). The urban morphology of Phoenix is representative of many newer cities in the Southwest (e.g., Las Vegas, Albuquerque), with medium-density central business districts, major arterials that are four lanes wide, and endless ribbons of freeways. Interstate 10 (east/west), a nationally important trucking corridor, runs through Phoenix, where it connects to Interstate 17 (north/south). Phoenix is also a destination city for major sporting events, including the Super Bowl (2015), the NCAA basketball tournament (2017), the Phoenix Open (golf) and baseball spring training for numerous major league teams. Given this high level of activity, drawing tourists from all over the U.S., the city lends itself to the study of prostitution.

Data

The analytical framework developed in this paper relies upon four major data sources: 1) prostitution related arrest data from the Phoenix Police Department, 2) hotel locations in Phoenix, 3) hotel characteristics (e.g., price), and 4) online hotel reviews written by guests and posted on Travelocity. This constellation of data is critical for generating a more holistic snapshot of prostitution activity and its local context. Core to our efforts are the arrest data for prostitution in Phoenix. Consider, for example, that 12 % of prostitution arrests in Phoenix between 2012 and 2014 occurred in hotels (Wallace n.d.). Although this represents just over 10% of the prostitution arrests in the city, it is twice the national average (Roe-Sepowitz et al. 2011). More importantly, these data likely represent a conservative baseline of indoor prostitution activity in Phoenix, for all of the reasons detailed above.

The Travelocity data, which include hotel locations (i.e., street address), hotel pricing, star rating, and the customer reviews provide the second, third and fourth pillars of the analysis. Again, this hotel information provides a unique snapshot of the site and situation associated with each hotel, along with basic data concerning its operational context. One hundred and twenty-five hotels within the city of Phoenix were included for the analysis, corresponding to the complete population of hotels during the years that prostitution arrest data were collected. To pair the hotel and arrest data, each hotel was assigned a 250-m buffer, and any prostitution arrest occurring within a buffer was attributed to the associated hotel. On the rare occasion when an arrest occurred in more than one buffer (6%), the assignment was made to the nearest hotel.

The prostitution arrest data were obtained from the Phoenix Police Department and included all prostitution-related arrests from January 1st, 2012 to December 31st, 2014 (3 years). The prostitution data are very limited, including only the year of arrest, location of arrest, and reason for arrest. No individual identifying information was contained in these data. Although limited, the arrest data are interesting. In particular, the reason for arrest provided contextual information on the crime event. These reasons included: 1) escort license violation, 2) escort violation, 3) prostitution, 4) prostitution of a child, 5) prostitution solicitation, and 6) attempted prostitution solicitation. There were a total of 3387 prostitution-related arrests during the 3 year period, 1506 of which occurred in the catchment area of a hotel. Hotels in this study ranged from having 0 prostitution arrests to 262, averaging one arrest every 4 days.

As detailed previously, information for each hotel was mined from multiple internet sites. Price data were gathered by randomly selecting three nights (the same three nights for each hotel) that were at least 7 days but not more than 28 days in the future. For each hotel, on each night, the price of the cheapest available room was extracted from Hotels.com and averaged to arrive at a typical price for a given hotel. Hotel review data were gathered from Travelocity.com for all the reviews that were written from January 1, 2006 to December 31, 2015 for each hotel. The scraping was performed using a custom script written in Autohotkey v. 1.1.24.01 and the web-scraping program Mozenda v. 4.0.159.

Methods

Review Processing and Modeling

The reviews were pre-processed using R v. 3.2.2, such that each review was separated into the date it was written, the review title, the score (from one to five stars, assigned by the reviewer), and the written body of the review. It is important to note that these fields are optional for review writers, except for the score and the date each review was submitted. For the purposes of our analysis, only the actual body of each review was used. Additionally, every review for a given hotel was combined into one block of text – ensuring that the model was agnostic to review counts.

The next pre-processing step was tokenization, which involves splitting each review block into individual atomic words called tokens. For example, the sentence “This hotel was nice.” would become [“This”, “hotel”, “was”, “nice”, “.”]. Then, for each token, punctuation is stripped both within and around words (the word “e-mail” becomes “email”, for example, and the “.” is eliminated). Next, case variation is removed (“This” becomes “this”). Finally, all stopwords are removed. These are words that are common enough in English that their appearances in a review are superfluous. Stopwords include “the”, “a”, “have”, etc. In total, 543 such stopwords were removed.[5] In addition to the stopwords, tokens that were typos, URLs, or non-English words were also removed, resulting in a total removal of 2% of the tokens in the dataset.

One common way to represent text is through the use of word vectors (Mikolov et al. 2013). Word vectors are useful for this type of analysis. In short, the representation of each word is learned from the word’s positionality compared to other words. The goal of word embedding is to learn an n-dimensional vector representation for each word such that words that appear closer together in the text are embedded closer together in vector space. While the resulting word vectors are not interpretable, they carry important semantic information regarding the words in the corpus (Levy et al. 2015). By employing word vectors to generate feature vectors, we are hypothesizing that there are regions of this embedding where the words are more indicative of a hotel with high levels of prostitution than others.

Our efforts to exploit this information is based on pre-trained embedding made available by Facebook.[6] We leveraged pre-trained vectors because they are trained on a larger set of text data than our hotel reviews, and thus the positioning of the words will be informed by a larger body of information. The vectors in this embedding are trained on the entirety of English Wikipedia, and contain 300 dimensions (Bojanowski et al. 2016). Specifically, for each of the 125 pre-processed review blocks (corresponding to each individual hotel), a 300-dimension vector, d, was created, with all of its entries initialized to 0. Then, for each token in the review block, its word vector, v, was queried within the pre-trained embedding. Next, v was added to d, and d was divided by the number of tokens in the review block. This results in a feature vector that we then used to classify the review blocks.

Finally, in an attempt to predict hotels that have especially high or low prostitution activity, two steps are taken. First, a medoid clustering routine (Grubesic 2006; Kaufman and Rousseeuw 2009) was used to group the hotels into one of three clusters based on their prostitution-related activity. Cluster 1 hotels (n = 16) contained hotels that had no activity. Cluster 2 (n = 87) contained hotels that had low prostitution activity, and Cluster 3 (n = 22) contained the hotels that had high prostitution activity. After this process, we next wanted to build a predictive model that could classify a hotel as a Cluster 1 or Cluster 3 hotel (see Fig. 2 for a conceptual schematic of how this model is

5 This list was obtained from the “MySQL Stopwords” list at http://www.ranks.nl/stopwords/

6 https://github.com/facebookresearch/fastText/blob/master/pretrained-vectors.md

built using vectorized hotel reviews).[7] We chose a random forest classifier for this task (Breiman 2001). We chose this classifier for two reasons. First, it is an ensemble classifier, meaning it leverages multiple classifiers and aggregates them to make a final prediction. Ensemble classifiers are known for giving superior performance to non-ensemble methods on problems with less data (Domingos 2012). Second, due to the way that the random forest algorithm aggregates columns across the different trees it leverages, it is resilient to noise and high dimensionality (Breiman 2001). This is particularly useful as many of the 300 dimensions in our pre-trained vectors may not be useful for the classification task. A random forest classifier was built for the Cluster 1 (no prostitution activity) and Cluster 3 (high prostitution activity) hotels using cross validation. A predictive baseline was built using just hotel price and location data to predict hotel cluster membership. The classifier we test was built using price, location, and the review vectors to predict the same. In short, improvements over the baseline model will indicate that the addition of the review feature vectors either helped, or did not help predict which hotels are likely to have exceptionally low and high prostitution activity. These classifiers were built using scikit-learn, a machine learning library for Python (Pedregosa et al. 2011).

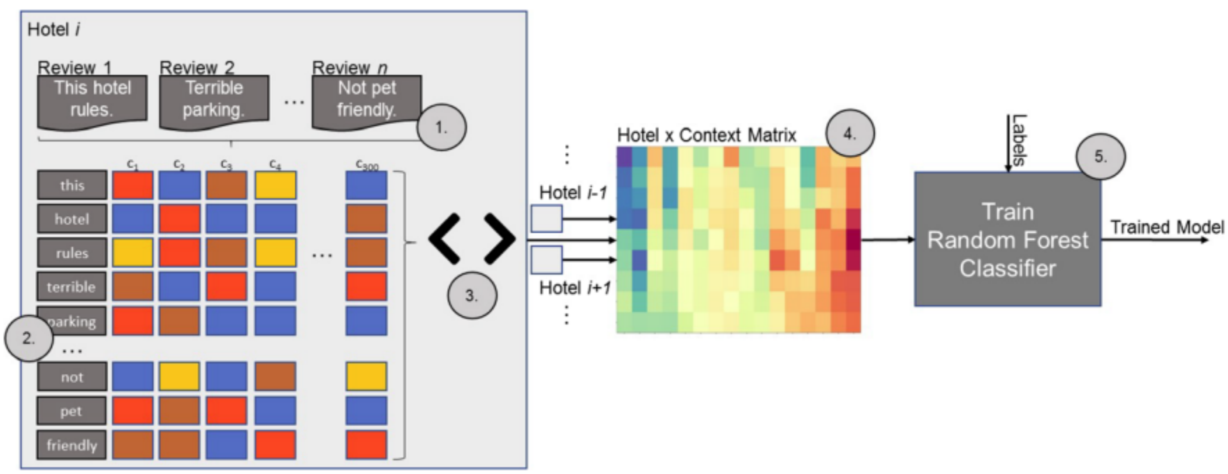

Fig. 2 This figure shows the five steps taken to train the classifier, each step denoted by a circle. In step 1, all reviews for a particular hotel are combined into a single document. In step 2, each word's corresponding vector representation is looked up. In step 3, the vectors for all the words in a given document are aggregated. Repeating this step for each hotel results in the hotels' representation. In step 4, all of the hotel representations are combined with their labels. Finally, in step 5, those are fed into the algorithm that trains the classifier

Results

Descriptive statistics for the three clusters of hotels are found in Table 1. The number of prostitution-related arrests for an individual hotel ranged from 0 to 262. All of

7 Predictive models are different from exploratory models in many ways. For a more thorough discussion, readers should consult Shmueli (2010) or Mack and Grubesic (2009). In short, issues of multicollinearity, heteroskedasticity and endogeneity are important to correct in explanatory models, but this is not the case for good forecasting models.

the hotels that have 0 prostitution-related arrests were assigned to Cluster 1. Cluster 2 consisted of hotels with 1–49 prostitution arrests, while Cluster 3 had hotels with 50–262 total arrests over the 3 years of data. While all hotels with 0 prostitution-related arrests were assigned to Cluster 1, the cutoff between Clusters 2 and 3 was determined by sensitivity analysis. Specifically, three factors were given weight in determining a cutoff value of 50 prostitution-related arrests for Cluster 3 membership – the analysis attempted to maximize feature distance between Clusters 2 and 3, minimize the within-cluster distances, and keep the sizes of Clusters 2 and 0 similar. The average nightly price of a hotel was negatively correlated with the number of prostitution-related arrests associated with that hotel. For example, the Cluster 3 hotels (high prostitution activity) were the least expensive, on average, at $60.43 per night. The Cluster 1 hotels (low prostitution activity), cost an average of $81.27 per night and the Cluster 1 hotels (no prostitution activity) cost an average of $98.65 per night.

Table 1 Descriptive statistics of the three clusters that hotels were assigned to based on those hotels' numbers of prostitution arrests that occurred nearby

	n	Average arrests	Arrest range	Average nightly room price	Room price range	Price standard deviation
Cluster 1 (No prostitution)	16	0	0–0	$98.65	$50–$149	31.95
Cluster 2 (Low prostitution)	87	12	1–49	$81.27	$39–$150	36.67
Cluster 3 (High prostitution)	22	106	51–262	$60.43	$44–$86	20.42

Cluster membership was determined by medoid clustering

When it comes to predicting cluster membership, the baseline random forest classifier using hotel price and location (but ignoring the feature vectors generated from the hotel reviews) predicted cluster membership with an 8.9% error rate, where error rate is defined as 100% - accuracy. However, when using the review block feature vectors in the model, the error rate decreased by 1.0 percentage points, indicating approximately an 11% improvement.

The best performing model was a classifier that used both hotel location and the *feature vectors*, but ignored price (Table 2). In this model, cluster membership was predicted with an error rate of only 6.3%, a 29% improvement over the baseline model. This is a significant improvement and indicates that the amalgamation of geographic information, hotel characteristics and associated hotel reviews provide a powerful combination for predicting prostitution activity in Phoenix – significantly better than the simple baseline model. Of note, however, is that the inclusion of hotel pricing information did not help. While price was negatively correlated with prostitution arrests proximal to a hotel, the variance of price for all three clusters was quite large, with significant overlap between clusters. This additional noise may have served to reduce the model's accuracy.

Geographic Trends

Figure 3 displays the density of prostitution arrests from 2012 to 2014 along with the spatial distribution of the hotels used in our study. There are three distinct clusters of high density prostitution arrests, each of which occurs along a major interstate highway corridor and in neighborhoods that could be classified as a mix of industrial and commercial, with proximal areas of low-grade residential.[8] The westernmost cluster, which is located along the east/west corridor of I-10, is in the heart of the Coronado Commerceplex. The Commerceplex is an area comprised of warehouses, light industry, cheap hotels and fast food restaurants (Fig. 4). It also presents an easy rest stop location for long-haul truck drivers seeking food, fuel and sleep. If one closely inspects the satellite imagery in Fig. 4, multiple semi-trucks can be seen in the hotel parking lots. Amazingly, when exploring the local geographic context for this location, we stumbled upon a Google Maps pin referring to this location as “hookers paradise” (Fig. 4).[9] More importantly, many of the Travelocity reviews that were mined for the hotels in this area (Baymont Inn Suites, Comfort Inn and Travelers Inn; all Cluster 3 hotels), had references to “hookers paradise” in the reviews. Clearly, this is an area with a reputation – significant enough for it to be included in Google Maps and noted publicly in the hotel reviews for the area.

Table 2 Random forest classifier results built from several combinations of explanatory variables

Baseline models	Error rate	Improvement over baseline
Price	25.30%	−184%
Location	9.50%	−7%
Price + Location	*8.90%*	*0%*
Feature vector models		
Vectors	7.30%	18%
Price + Vectors	8.90%	0%
Location + Vectors	**6.30%**	**29.20%**
Price + Location + Vectors	7.90%	11%

Baseline models include those that use price and/or location but ignore review block feature vectors, while the feature vector models include the review block data. The best performing baseline model included both price and location and is indicated by italics in the table below. Any model that included the review block data performed the same as or better than the best baseline model, while the very best model used feature vector data and location, but ignored price (bolded in the table)

8 Low-grade residential is a catch-all term for neighborhoods dominated by poor-quality housing, which includes trailer parks, poorly constructed multi-family apartment buildings and single-family houses.

9 This was not placed in the Commerceplex by the authors.

The second hot spot is located along the Interstate 17 corridor and Indian School Road in Phoenix (Fig. 5). This is an area with an abundance of low-grade residential housing that includes numerous trailer parks and deteriorated apartment complexes. This is also the location of two Cluster 3 hotels, including a Motel 6 and a Travel Inn. The Motel 6 complex is particularly large, and includes three detached wings, two pools and ample parking for long-haul truck drivers. Additionally, of all our study hotels, this Motel 6 had the largest number of associated prostitution arrests – 262 over the course of the three-year arrest dataset. While this hotel is undoubtedly contributing to the existence of this cluster of prostitution arrests, the area's residential profile is also a contributing factor.

The third hot spot, located along Interstate 17 and Bell Road in north Phoenix (Fig. 6), has a similar composition to the Indian School Road hot spot. Although the adjacent residential areas are less deteriorated than the other hot spots, the Bell Road location is also home to several Cluster 3 hotels, including another Motel 6, a Marriott Fairfield Inn & Suites, Red Roof Inn and the Bell Hotel and Suites. The usual selection of fast food restaurants are also found here, as well as ample parking for long-haul truckers, especially at the Red Roof Inn, which is located on the west side of I-17.

An important aspect of the results worth reiterating is the relationship between hotels, prostitution arrests, hot spots, and hotel clusters. Prostitution arrests were attributed to hotels if they were within 250 m of a facility. Hotel clusters (Clusters 1–3) were statistically derived using the number of prostitution arrests attributed to each hotel. The kernel density maps are generated to help identify the geographic footprint of prostitution arrest densities and highlight outliers (i.e., hot spots). With this basic analytical framework in mind, and given the spatial distribution of prostitution arrests in Phoenix, one would expect that at least some arrests would be attributed to nearby hotels. However, as detailed in Fig. 3, the highest density hot spots of prostitution arrests have a strong correlation to Cluster 3 hotels. This is not a coincidence. It is reasonable to conclude that some amount of prostitution is occurring within all of the Cluster 3 hotels located in these areas. Furthermore, while there may be several different brands of hotels in these hot spots, they all have similar profiles: 1) inexpensive, 2) few amenities, 3) ample parking for long haul truck drivers, 4) nearby fast food restaurants, 5) rooms that ingress/egress directly to the parking lot instead of to a central hallway.

Discussion and Conclusion

Many municipalities have a vested interest in determining the spatial distribution of prostitution activity within their borders. Whether those municipalities are primarily interested in policing prostitution, reducing human trafficking, or applying social outreach and/or health intervention programs, understanding where prostitution is occurring is an important component of effective and efficient resource use. Significant amounts of prostitution occur indoors, sometimes in private residences, but more commonly in hotels. Indoor prostitution is notoriously difficult to combat because it is

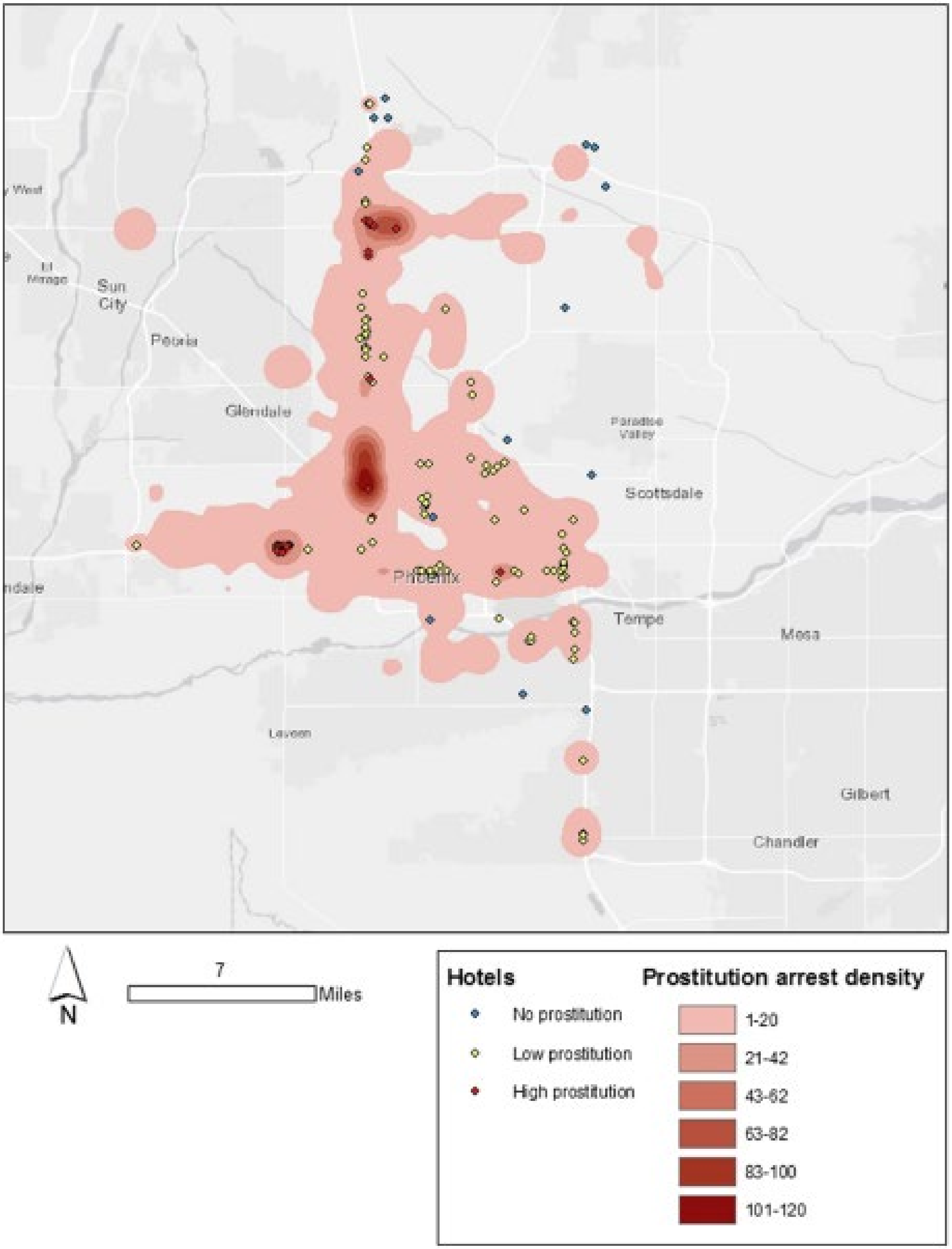

Fig. 3 A kernel density of prostitution arrests in Phoenix from 2012 to 2014 as well as the 125 hotels used for this study. Prostitution arrest density is highest along major interstates but away from the central business districts of Phoenix and its neighboring cities. Three hotspots emerge, each indicated by dark red in the kernel density

underreported – making it difficult for law enforcement to recognize (Raphael and Shapiro 2004). Furthermore, aggressive crackdowns on street prostitution often drive it elsewhere; sometimes to adjacent neighborhoods, but also off the streets into nearby hotels (Cunningham and Shah 2014). To that end, the identification of prostitution in more aggregate dimensions, such as neighborhoods, is insufficient for both law enforcement and social outreach services.

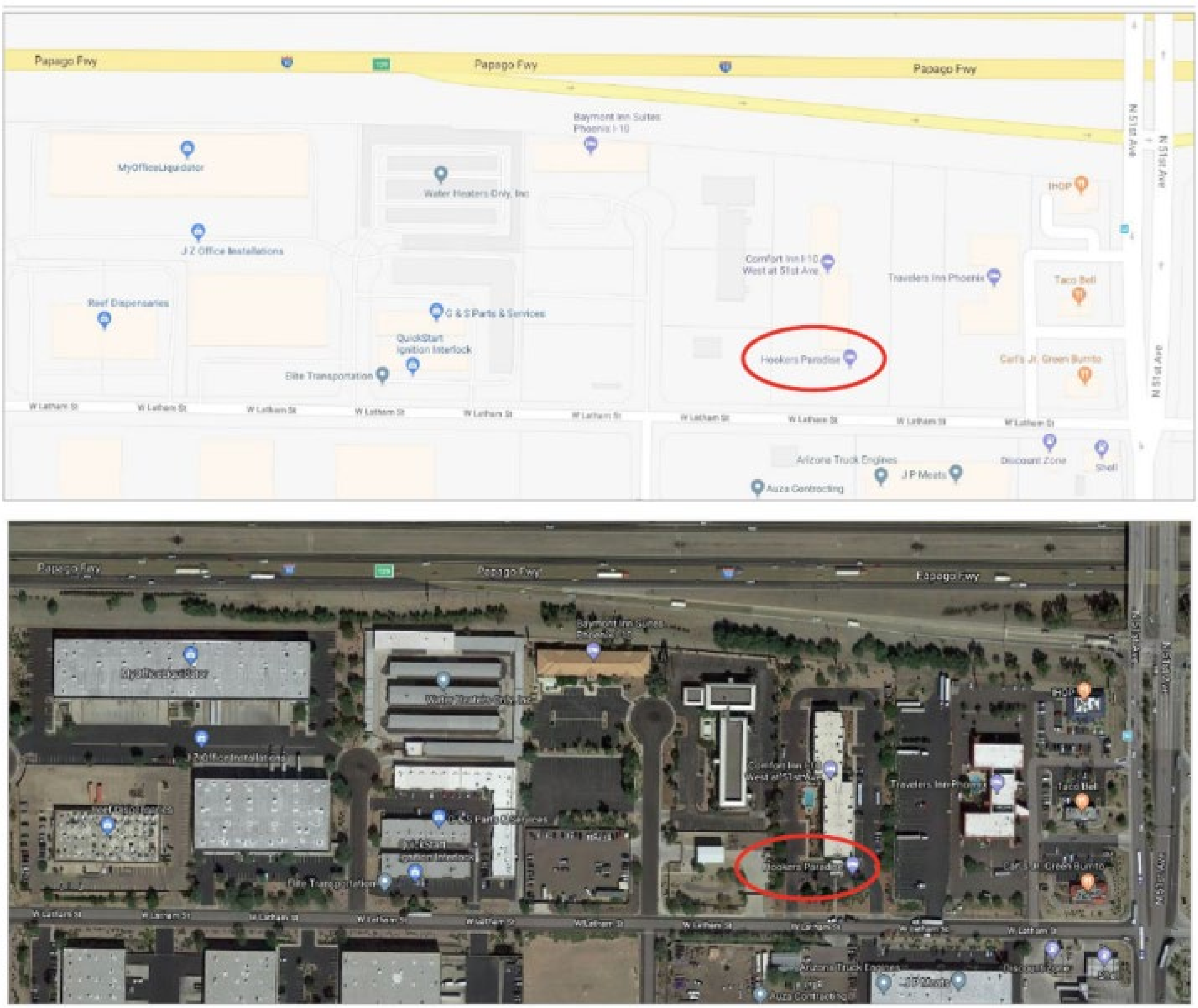

Fig. 4 A screenshot showing the geographic context of the westernmost prostitution arrest cluster along I-10 in Phoenix, AZ. Note the presence of fast food restaurants, semi-truck parking, and chain motels. There is also a user-generated map pin designating an area as 'Hooker's Paradise.' Source: Google Maps, Google Earth

As a result, the ability to identify indoor prostitution and the hotels within which it is occurring is critical. As detailed in this paper, while this can be done fairly well by simply looking at the hotel's location and price, it is not enough. Admittedly, a cheap hotel located in neighborhoods associated with frequent street prostitution makes for an attractive location for indoors prostitution as well. However, the addition of supplementary information, including hotel reviews, dramatically increases the predictive accuracy of models seeking to identify hotels with high levels of prostitution activity. With a predictive error rate reduction from 8.9% (using price and location) to 6.3% (using location and the review vectors), the model incorporating the hotel reviews is almost 30% more effective than the baseline model. Additionally, this approach demonstrates a method by which other high-volume user-generated data (such as Twitter posts that mention a given hotel) can be incorporated, offering an additional pathway to model refinement.

Law enforcement agencies, social outreach services, and health service providers are frequently working in a resource-constrained environment in combating prostitution and helping prostitutes. Too few people, not enough time, and a general mandate to do more, with less. While a 30% predictive improvement may not be jaw-dropping, any improvement to their capabilities for identifying problem areas and allocating their resources more efficiently are critical. For example, the identification of prostitution hot spots and associated hotels could not only inform the optimal location(s) for mobile women's health clinics, but it could also inform healthcare providers in distributing free condoms or other prophylactics more efficiently. Law enforcement agencies could use the improved models for narrowing their search for human sex trafficking victims, or could more effectively reduce the rate of violence against sex workers. Even something as simple as regular patrols in and around high-prostitution hotels could reduce the frequency of crimes related to prostitution.

As detailed previously, there are community effects to consider as well. Areas with high levels of street prostitution are strongly correlated with street drug use (Young et al. 2000), adolescent alcohol use (Winstanley et al. 2008), and a loss of business income (Steenbeek et al. 2012). Again, while these negative externalities are more commonly associated with street prostitution, effectively combating them requires tackling indoor prostitution as well. For instance, Huff et al. (2018) found that massage parlors providing illicit sexual services were associated with increased neighborhood crime and physical disorder. Thus, understanding where prostitution is occurring – both publicly and privately – can inform more effective policy and district development efforts. Revitalizing previously blighted areas of a city leads to more tax revenue, more tourism, more business investment, and a more satisfied populace (Schill et al. 2002).

Limitations and Future Work

This study serves as a proof of concept for the method of incorporating hotel reviews written by guests into a model that predicts where indoor prostitution may be occurring. It is still limited in scope, but this was by design. Although this work leverages 3 years of prostitution-related calls, only 125 hotels are included, all of which fall in the Phoenix metro area. Future work will be needed to verify the extent to which these results generalize beyond the Phoenix area.

In particular, the 'black box' nature of producing feature vectors from hotel review blocks does call into question one potential issue. In using pre-trained vectors produced from the body of English-language Wikipedia articles, we ignore differences in how the

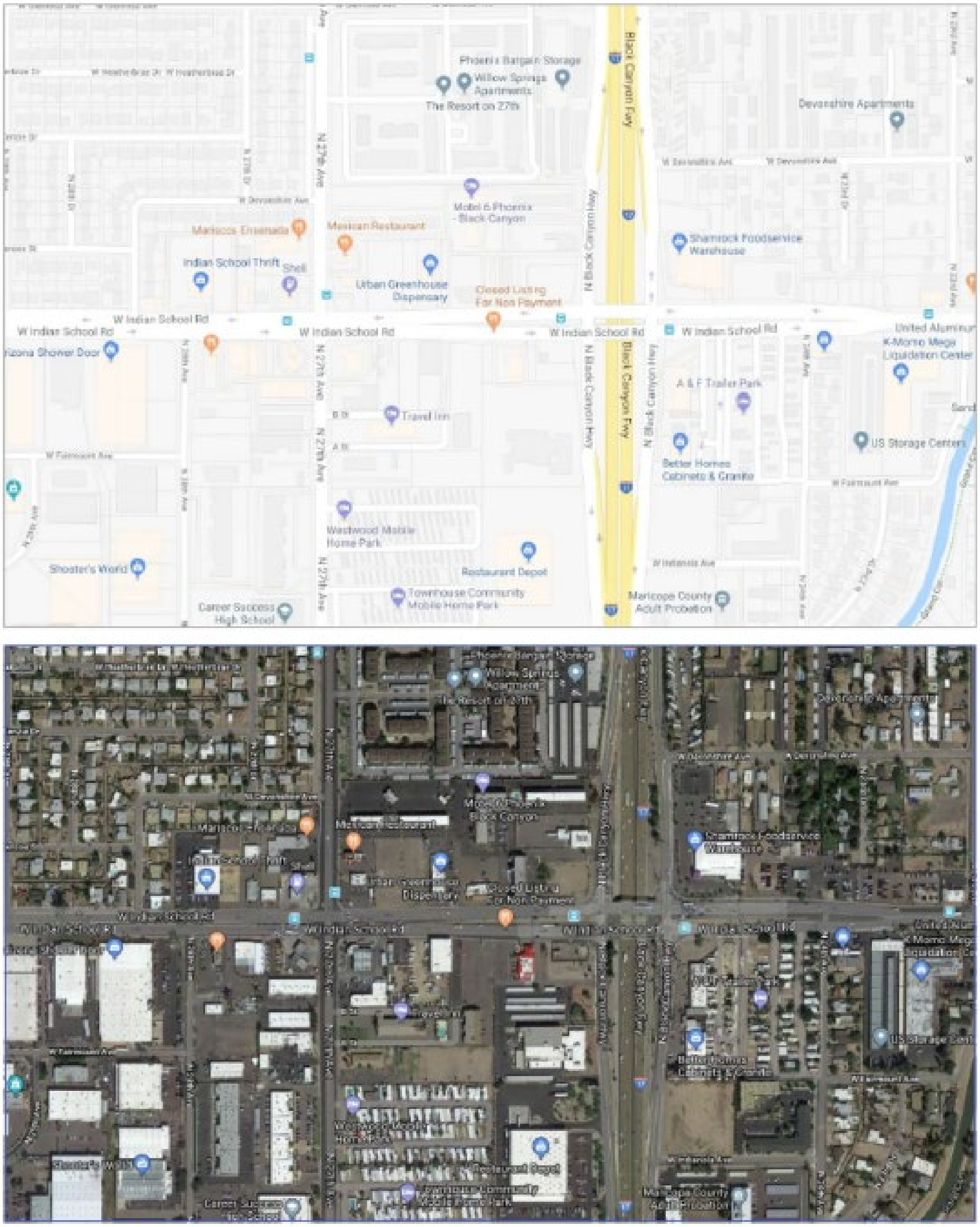

Fig. 5 A screenshot of the central prostitution arrest cluster, near I-17 and Indian School Road in Phoenix, AZ. This area contains trailer parks, several large and deteriorating apartment complexes, and cheap motels. Source: Google Maps, Google Earth

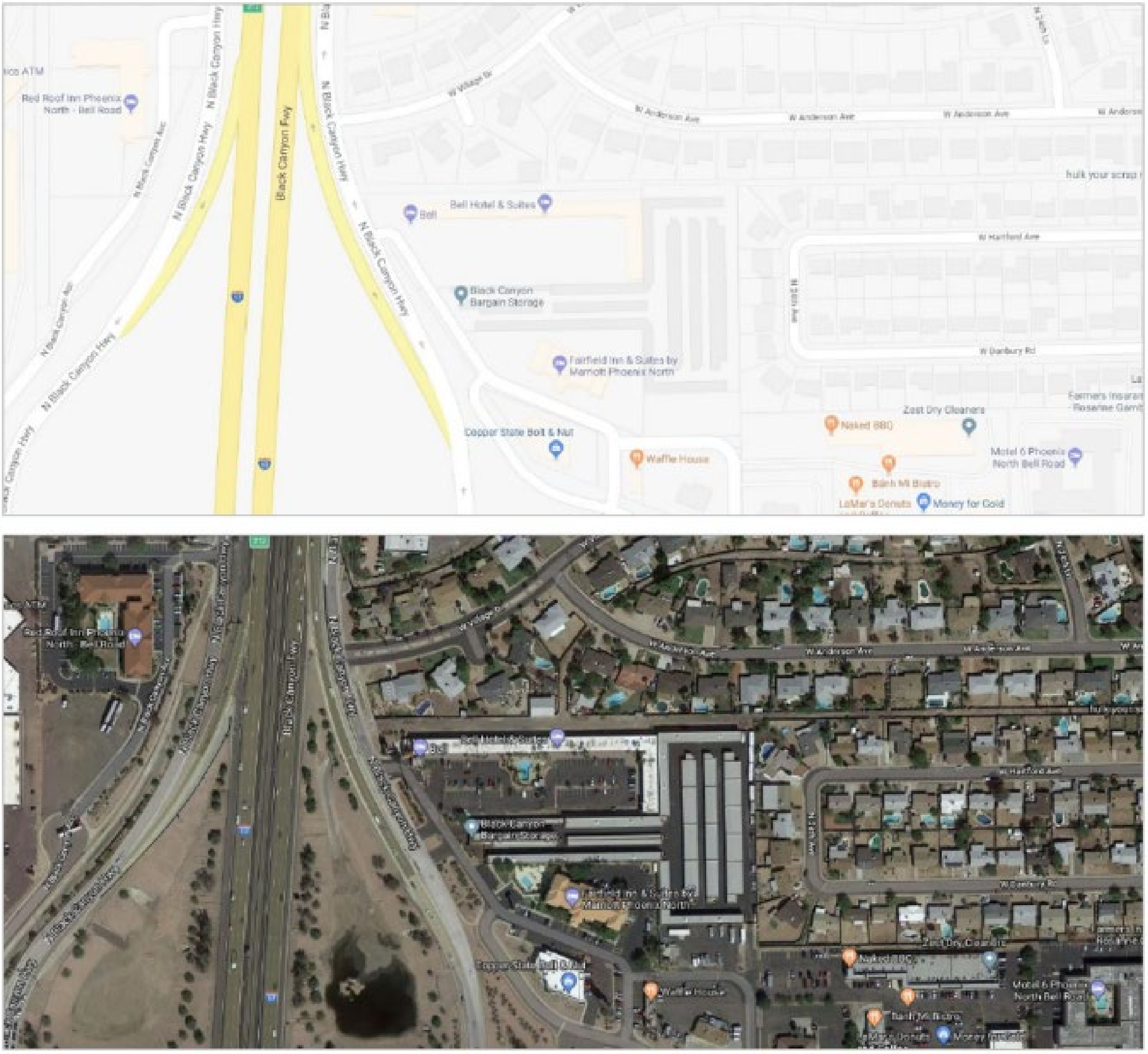

Fig. 6 The northernmost prostitution arrest cluster in Phoenix, AZ. Like the others, it is home to semi-truck parking, large, cheap motels, and fast food restaurants. It is worth noting that the residential areas adjacent to this area are not nearly as deteriorated as those in the other two clusters. Source: Google Maps, Google Earth

typical hotel review is written versus the typical Wikipedia article. A hotel review is likely to be written more casually, perhaps with colloquialisms that are not used in Wikipedia, or with certain sentence structures that could change the meaning of a review in a more formal context. However, this risk is relatively minimal as there are few words that would be used commonly in both scenarios but with different enough meanings to skew the results. In addition, by using pre-trained vectors, we avoid the pitfalls of training our own using our review data (primarily a lack of accuracy, as it takes far more data than we used to generate reliable training vectors). Finally, this study only examined reviews written in English. While reviews written in other languages were rare, they could add fidelity to the model, but would require using different vectors and a different training process.

Additionally, by using the hotel reviews themselves, some bias is introduced to the predictive model. Many guests stay in these hotels and do not write any review – or they write a review that is posted to a website other than Travelocity.com. This model does not capture any of these guests' sentiments, and there may be persistent bias in the individual attributes that lead to a guest writing a review. Furthermore, reviews may be written more or less often for particular types of hotels, or after having particular experiences (e.g., reviews may be more likely to be written after a negative experience). There may also be systematic biases in the types of people who write hotel reviews in the first place, as has been discovered using other types of social media (Liu et al. 2014). However, individual biases should be largely consistent across hotels, minimizing representational error. In addition, there is little reason to suspect that guests' attitudes toward prostitution are consistently different between the different types of hotels.

It is also important to acknowledge that the prostitution data was limited to arrests. Given that this excludes any police contact regarding prostitution not resulting in arrest, this data likely underestimates illegal prostitution activity. There are other ways that this data does not represent the true level of prostitution activity at a hotel. For instance, a simple threat to call the police from a hotel manager may be enough to stop an incident of prostitution or remove the activity from the hotel. Again, this shows the arrest data underestimates prostitution activity. Finally, because the data does not include information on the specifics of the arrest, we are unable to determine if the arrest happened inside the hotel or somewhere else on the grounds. This information could have been used to understand what guests may be seeing while staying at the hotel should the arrest occur on grounds of the hotel rather than in the privacy of a hotel room.

As for future work, the generalizability of these results should be confirmed. Performing a similar study in multiple cities with hotel review data would allow for broader conclusions to be drawn about the efficacy of using hotel review data in identifying hotels that are commonly host to indoor prostitution. At that point, this methodological framework could be used by police agencies to more efficiently leverage their limited resources in their fight against human trafficking and violence directed at prostitutes. In addition, public outreach or public policy groups could better target hotspots in their efforts to help trafficked prostitutes escape their traffickers and rehabilitate.

More broadly, this study lays the groundwork for other uses of high-volume user-generated data to solve spatial problems that suffer from resource constraints or imperfect information. Highly granular economic information, for example, could be gathered via business reviews for a given neighborhood – these could be further used to tease out temporal patterns as reviews change over time. This could allow local government to better anticipate gentrification well before property prices change. Alternatively, an adapted approach could access real-time public sentiment to local conditions through Twitter posts, leading to everything from more effective emergency response during a disaster to public opinion on proposed legislation.

Conclusion

Our goal was to determine whether the use of publicly written hotel review data could help identify hotels that fall on either end of extreme prostitution-related activity. The results shown here indicate that the addition of hotel review feature vectors do improve a model's ability to predict whether a hotel is likely to be home to particularly high or particularly low levels of prostitution-related activity. This study adds to the growing body of literature that suggests that the properly applied use of crowdsourced data (hotel reviews, Twitter data, mobile network activity, etc.) adds to the ability to predict crime hotspots. Not only does this beget more effective application of limited police resources but also allows for more effective outreach programs directed at helping and rehabilitating street prostitutes.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Anstine, J. (2003). Property values in a low populated area when dual noxious facilities are present. Growth and Change, 34 (3), 345–358.

Barnard, M. A. (1993). Violence and vulnerability: conditions of work for streetworking prostitutes. Sociology of Health & Illness, 15(5), 683–705. https://doi.org/10.1111/j.1467-9566.1993.tb00368.x.

Benson, C., & Matthews, R. (2000). Police and prostitution: Vice squads in Britain. In R. Weitzer (Ed.), Sex for sale: Prostitution, pornography, and the sex industry (pp. 245–264). New York: Routledge.

Boels, D., & Verhage, A. (2016). Prostitution in the neighbourhood: impact on residents and implications for municipal regulation. International Journal of Law, Crime and Justice, 46, 43–56.

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014). Once upon a crime: Towards crime prediction from demographics and mobile data. In Proceedings of the 16th international conference on multimodal interaction (pp. 427–434). New York: ACM. https://doi.org/10.1145/2663204.2663254.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching Word Vectors with Subword Information. ArXiv:1607.04606 [Cs]. Retrieved from http://arxiv.org/abs/1607.04606. Retrieved October 17, 2018.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Butcher, K. (2003). Confusion between prostitution and sex trafficking. The Lancet 361, (9373), 1983.

Census Bureau. (2017). State population totals and components of change: 2010–2017. Retrieved from: https://www.census.gov/data/datasets/2017/demo/popest/state-total.html. Retrieved October 17, 2018.

Chokshi, N. (2015). The religious states of America, in 22 maps. Washington Post. URL: https://tinyurl.com/ybalkfn3

Chudakov, B., Ilan, K., Belmaker, R. H., & Cwikel, J. (2002). The motivation and mental health of sex workers. Journal of Sex & Marital Therapy, 28(4), 305–315.

Church, S., Henderson, M., Barnard, M., & Hart, G. (2001). Violence by clients towards female prostitutes in different work settings: questionnaire survey. BMJ (Clinical Research Ed.), 322(7285), 524–525. https://doi.org/10.1136/bmj.322.7285.524.

Cunningham, S., & Kendall, T. D. (2011). Prostitution 2.0: The changing face of sex work. Journal of Urban Economics, 69(3), 273–287. https://doi.org/10.1016/j.jue.2010.12.001.

Cunningham, S., & Shah, M. (2014). Decriminalizing indoor prostitution: Implications for sexual violence and public health (working paper no. 20281). National Bureau of Economic Research. https://doi.org/10.3386/w20281

Dank, M.L., Khan, B., Downey, P.M., Kotonias, C., Mayer, D., Owens, C., ... & Yu, L. (2014). Estimating the size and structure of the underground commercial sex economy in eight major US cities (p. 348). Washington, DC: Urban Institute.

Day, S. E., & Ward, H. (2007). British policy makes sex workers vulnerable. BMJ (Clinical Research Ed.), 334, 185–187. https://doi.org/10.1136/bmj.39104.638785.59.

Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10), 78–87.

Farber, S. (1998). Undesirable facilities and property values: a summary of empirical studies. Ecological Economics, 24(1), 1–14.

Farley, M. (2007). Prostitution and trafficking in Nevada: Making the connections. San Francisco: Prostitution Research & Education.

Farley, M., & Barkan, H. (1998). Prostitution, violence, and posttraumatic stress disorder. Women & Health, 27(3), 37–49.

Farley, M., & Kelly, V. (2000). Prostitution: a critical review of the medical and social sciences literature. Women & Criminal Justice, 11(4), 29–64.

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. Decision Support Systems, 61, 115–125. https://doi.org/10.1016/j.dss.2014.02.003.

Gil, V. E., Wang, M. S., Anderson, A. F., Lin, G. M., & Wu, Z. O. (1996). Prostitutes, prostitution and STD/HIV transmission in mainland China. Social Science & Medicine, 42(1), 141–152.

Gossop, M., Powis, B., Griffiths, P., & Strang, J. (1995). Female prostitutes in South London: use of heroin, cocaine and alcohol, and their relationship to health risk behaviours. AIDS Care, 7(3), 253–260.

Gozdziak, E. M., & Collett, E. A. (2005). Research on human trafficking in North America: a review of literature. International Migration, 43(1–2), 99–128.

Grubesic, T. H. (2006). On the application of fuzzy clustering for crime hot spot detection. Journal of Quantitative Criminology, 22(1), 77–105.

Grubesic, T. H., & Murray, A. T. (2008). Sex offender residency and spatial equity. Applied Spatial Analysis and Policy, 1(3), 175–192.

Harcourt, C., O'connor, J., Egger, S., Fairley, C. K., Wand, H., Chen, M. Y., ... & Donovan, B. (2010). The decriminalisation of prostitution is associated with better coverage of health promotion programs for sex workers. Australian and New Zealand Journal of Public Health, 34(5), 482–486.

Havocscope. (2015). Prostitution: Prices and statistics of the global sex trade. Havocscope Books. Digital. 1st ed. Retrieved from: https://www.havocscope.com/prostitution-book/. Retrieved October 17, 2018.

Heiges, M. (2009). From the inside out: reforming state and local prostitution enforcement to combat sex trafficking in the United States and abroad. Minnesota Law Review, 94, 428.

Hubbard, P. (1997). Red-light districts and toleration zones: geographies of female street prostitution in England and Wales. The Royal Geographical Society, 29(2), 129–140. https://doi.org/10.1111/j.1475-4762.1997.tb00015.x.

Hubbard, P. (1998). Community action and the displacement of street prostitution: evidence from British cities. Geoforum, 29(3), 269–286.

Hubbard, P., Boydell, S., Crofts, P., Prior, J., & Searle, G. (2013). Noxious neighbours? Interrogating the impacts of sex premises in residential areas. Environment and Planning A, 45(1), 126–141.

Huff, J., Wallace, D., Riggs, C., Katz, C. & Choate, D. (2018). Examining the association between massage parlors and neighborhood crime. Crime & Delinquency. Online first: https://doi.org/10.1177/0011128718787154.

Jeal, N., & Salisbury, C. (2007). Health needs and service use of parlour-based prostitutes compared with street-based prostitutes: a cross-sectional survey. BJOG: An International Journal of Obstetrics & Gynaecology, 114(7), 875–881.

Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: An introduction to cluster analysis (Vol. 344). Hoboken: Wiley.

Kempadoo, K., Sanghera, J., & Pattanaik, B. (2015). Trafficking and prostitution reconsidered: New perspectives on migration, sex work, and human rights. Abingdon: Routledge.

Krüsi, A., Kerr, T., Taylor, C., Rhodes, T., & Shannon, K. (2016). "They won"t change it back in their heads that we're trash': the intersection of sex work-related stigma and evolving policing strategies. Sociology of Health & Illness, 38(7), 1–14. https://doi.org/10.1111/1467-9566.12436.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. Transactions of the Association for Computational Linguistics, 3(0), 211–225.

Larsen, E. N. (1992). The politics of prostitution control: interest group politics in four Canadian cities. International Journal of Urban and Regional Research, 16, 169–189.

Liu, Y., Kliman-Silver, C., & Mislove, A. (2014). The tweets they are a-Changin: evolution of Twitter users and behavior. ICWSM, 30, 5–314.

Lowman, J. (2000). Violence and the outlaw status of (street) prostitution in Canada. Violence Against Women, 6, 987–1011.

Mack, E. A., & Grubesic, T. H. (2009). Forecasting broadband provision. Information Economics and Policy, 21(4), 297–311.

Malleson, N., & Andresen, M. A. (2015). Spatio-temporal crime hotspots and the ambient population. Crime Science, 4(1), 10. https://doi.org/10.1186/s40163-015-0023-8.

Mansfield, C., Van Houtven, G., & Huber, J. (2001). The efficiency of political mechanisms for siting nuisance facilities: are opponents more likely to participate than supporters? The Journal of Real Estate Finance and Economics, 22(2), 141–161.

Matthews, R. (2005). Policing prostitution: ten years on. British Journal of Criminology, 45(6), 877–895. https://doi.org/10.1093/bjc/azi046.

McCleary, R., & Weinstein, A. C. (2009). Do "off-site" adult businesses have secondary effects? Legal doctrine, social theory, and empirical evidence. Law & Policy, 31(2), 217–235.

McCutcheon, J. C., Mustaine, E. E., & Tewksbury, R. (2016). Working the stroll: prostitutes, Pimps, and Johns. Deviant Behavior, 37(12), 1449–1458. https://doi.org/10.1080/01639625.2016.1185882.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. ArXiv:1301.3781 [Cs]. Retrieved from http://arxiv.org/abs/1301.3781. Retrieved October 17, 2018.

Milrod, C., & Monto, M. (2016). Condom use, sexual risk, and self-reported STI in a sample of older male clients of heterosexual prostitution in the United States. American Journal of Men's Health, 10(4), 296–305.

Opaluch, J. J., Swallow, S. K., Weaver, T., Wessells, C. W., & Wichelns, D. (1993). Evaluating impacts from noxious facilities: including public preferences in current siting mechanisms. Journal of Environmental Economics and Management, 24(1), 41–59.

Oxford [Oxford English Dictionary]. (2018). Prostitution. URL: https://en.oxforddictionaries.com/definition/prostitution

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011).

Scikitlearn: machine learning in python. Journal of Machine Learning Research, 12, 2825–2830.

Prior, J., & Croft, P. (2012). Effects of sex premises on neighbourhoods: residents, local planning and the geographies of a controversial land use. New Zealand Geographer, 68, 130–140. https://doi.org/10.1111/j.1745-7939.2012.01228.x.

Procon. (2017). US federal and state prostitution laws and related punishments. URL: https://prostitution.procon.org/view.resource.php?resourceID=000119

Raphael, J., & Shapiro, D. L. (2004). Violence in indoor and outdoor prostitution venues. Violence Against Women, 10(2), 126–139. https://doi.org/10.1177/1077801203260529.

Roe-Sepowitz, D. E., Hickle, K. E., Loubert, M. P., & Egan, T. (2011). Adult prostitution recidivism: risk factors and impact of a diversion program. Journal of Offender Rehabilitation, 50(5), 272–285. https://doi.org/10.1080/10509674.2011.574205.

Sanders, T. (2004). The risks of street prostitution: punters, police and protesters. Urban Studies, 41(9), 1703–1717. https://doi.org/10.1080/0042098042000243110.

Schauer, E. J., & Wheaton, E. M. (2006). Sex trafficking into the United States: a literature review. Criminal Justice Review, 31(2), 146–169.

Schill, M. H., Ellen, I. G., Schwartz, A. E., & Voicu, I. (2002). Revitalizing inner-city neighborhoods: New York city's ten-year plan. Housing Policy Debate, 13(3), 529–566. https://doi.org/10.1080/10511482.2002.9521454.

Shannon, K., Rusch, M., Shoveller, J., Alexson, D., Gibson, K., & Tyndall, M. W. (2008). Mapping violence and policing as an environmental-structural barrier to health service and syringe availability among substance-using women in street-level sex work. International Journal of Drug Policy, 19(2), 140–147. https://doi.org/10.1016/j.drugpo.2007.11.024.

Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289–310.

Smith, C. J., & Hanham, R. Q. (1981). Any place but here! Mental health facilities as noxious neighbors. The Professional Geographer, 33(3), 326–334.

Steenbeek, W., Völker, B., Flap, H., & Oort, F. v. (2012). Local businesses as attractors or preventers of neighborhood disorder. Journal of Research in Crime and Delinquency, 49(2), 213–248. https://doi.org/10.1177/0022427810397948.

Van den Hoek, J. A., van Haastrecht, H. J., Scheeringa-Troost, B., Goudsmit, J., & Coutinho, R. A. (1989). HIV infection and STD in drug addicted prostitutes in Amsterdam: potential for heterosexual HIV transmission. Genitourinary Medicine, 65(3), 146–150.

Wallace, D. (n.d.). Prostitution at hotels in Phoenix, AZ: Final report prepared for the Sandra Day O'Connor Institute. Demanding Justice Arizona. Retrieved from: https://sharedhope.org/wpcontent/uploads/2015/05/Arizona-FieldAssessment-Report-ONLINE.pdf. Retrieved October 17, 2018.

Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic crime prediction using events extracted from twitter posts. In Social computing, behavioral - cultural modeling and prediction (pp. 231–238). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-29047-3_28.

Weitzer, R. (1999). Prostitution control in America: rethinking public policy. Crime, Law and Social Change, 32(1), 83–102.

Weitzer, R. (2000a). The politics of prostitution in America. In R. Weitzer (Ed.), Sex for sale: Prostitution, pornography, and the sex industry (pp. 159–180). New York: Routledge.

Weitzer, R. (2000b). Why we need more research on sex work. In R. Weitzer (Ed.), Sex for sale: Prostitution, pornography, and the sex.

Weitzer, R. (Ed.). (2009). Sex for sale: Prostitution, pornography, and the sex industry. Routledge.

Weitzer, R. (2010). The movement to criminalize sex work in the United States. Journal of Law and Society, 37(1), 61–84.

Willis, B. M., & Levy, B. S. (2002). Child prostitution: global health burden, research needs, and interventions. The Lancet, 359(9315), 1417–1422.

Winstanley, E. L., Steinwachs, D. M., Ensminger, M. E., Latkin, C. A., Stitzer, M. L., & Olsen, Y. (2008). The association of self-reported neighborhood disorganization and social capital with adolescent alcohol and drug use, dependence, and access to treatment. Drug and Alcohol Dependence, 92(1), 173–182. https://doi.org/10.1016/j.drugalcdep.2007.07.012.

Yates, G. L., MacKenzie, R. G., Pennbridge, J., & Swofford, A. (1991). A risk profile comparison of homeless youth involved in prostitution and homeless youth not involved. Journal of Adolescent Health, 12(7), 545–548.

Young, A., Boyd, C., & Hubbell, A. (2000). Prostitution, drug use, and coping with psychological distress. Journal of Drug Issues, 30(4), 789–800. https://doi.org/10.1177/002204260003000407