Using Social Media to Assess Neighborhood Social Disorganization
A Case Study in the United Kingdom

Diogo F. Pacheco, Marcos Oliveira, Ronaldo Menezes
BioComplex Laboratory, School of Computing, Florida Institute of Technology, Melbourne, FL, USA
{dpacheco, moliveira}@biocomplexlab.org, rmenezes@cs.fit.edu

Abstract

Urban communities can benefit from behavior regulation of their members in the interest of collective values. The absence of such control is related to the concept of social disorganization and is hypothesized to be associated with crime and anti-social behavior in neighborhoods. Social disorganization is, however, hard to quantify due to the lack of data and the inherent complexity that emerges from social interactions. Notably, geolocated social media provides a real-time assessment of places via the examination of the digital footprints left by users. In this paper, we introduce a measure for social disorganization by analyzing geotagged posts on Twitter. We propose to characterize the social disorganization of a place by evaluating the entropy of individuals' opinions about certain subjects. As a case study, we used tweets related to football in the UK, given its ubiquity in that country, which makes its supporters as proxies for the social characteristics of those places. We found that our proposed measure can reasonably explain the variation of the occurrence of crime across regions in UK and that our measure better explains the variation of crime among places with higher social disorganization.

Introduction

Crime takes place unevenly across places in cities. This characteristic of crime is argued by ecological criminologists to be the outcome of features in the social fabric of the places themselves (Henry and Einstadter 2006). Social ecologists highlight the active role of place in criminal occurrence and consider crime as the product of social disorganization which is produced by social changes, such as immigration, rural-urban migration, high social mobility, among others (Sampson, Raudenbush, and Earls 1997). These changes undermine social arrangements, such as traditional control institutions, traditional stable structures, and established coping behavior (Kubrin and Weitzer 2003). A deficient social structure in a community drives people to compete rather than to cooperate. The residents of cooperative communities have the capability to organize themselves, and to share expectations for the social control of public space, a concept called collective efficacy (Sampson, Raudenbush, and Earls 1997). The breakdown of such community control leads to a disorganized community insulated from conventional norms and prone to criminal activities (Henry and Einstadter 2006).

The place-oriented view of crime advocates for an analysis of the social structure of communities and its relationship with offenses, but the lack of data has hindered a broader assessment of these theories. Notably, the increasing amount of localized data available from social media have the potential to help in these analyses. An example of this is Twitter, an online micro-blogging social network which has grown significantly since its founding in 2006; as of the third quarter of 2016, Twitter claimed to have 317 million monthly active users. Twitter has become a general-content platform and has been used by researchers to understand and to model human behavior. In particular, the platform has allowed a better understanding of different aspects of crime (Wang, Gerber, and Brown 2012; Gerber 2014; Williams, Burnap, and Sloan 2016; Chen, Cho, and young Jang 2015; Wang and Gerber 2015; Patton 2015; Niu, Zhang, and Ebert 2015).

In this paper we propose a measure of social disorganization in places by examining geotagged tweets. We characterize a place using the entropy of different types of tweets in the region. In order to have a proxy of the social features in places, we gathered tweets about one of the most popular sports in the world: football (aka soccer). Due to its popularity, it is hard to imagine the behavior of the supporters as independent from other aspects of society, particularly in places such as Brazil, Italy, Germany, and the United Kingdom; football, however, has an estimated fan base of 3.5 billion individuals, and it is played by over 250 million people. We used football-related tweets from the UK to examine the relationship between our measure of social disorganization and criminal occurrence. For our measurement, we considered mentions to clubs' official Twitter account in the tweets. First, we evaluated the correlation between our proposed measure and the crime rate in places in the UK. Then, we isolated the population effect from users by performing a partial correlation. Finally, we built regression models to assess the contribution and the significance of our proposal to estimate social disorganization in the form of criminal activity.

Related Works

Twitter has been used independently to understand the dynamics of football and crime. The behavior of users in the platform has been examined to improve the prediction of offenses (Wang, Gerber, and Brown 2012; Gerber 2014; Wang and Gerber 2015; Chen, Cho, and young Jang 2015; Williams, Burnap, and Sloan 2016) and to estimate the population in at-risk regions (Malleson and Andresen 2015a). From the perspective of crime, gangs have been analyzed by tracking information from the tweets of gang members (Patton 2015). Moreover, Twitter has helped law enforcement officials to improve the awareness about the regions in a city with respect to crime-related tweets (Niu, Zhang, and Ebert 2015). Last, the prediction of social unrest has been linked to Twitter and other social media activity (Compton et al. 2013; Xu et al. 2014).

The almost real-time usage of Twitter during football matches allows analysis of tweets in order to automatically extract events occurring in a match (Van Oorschot, Van Erp, and Dijkshoorn 2012; Esmin et al. 2014). Such proxy can be also retrieved when the clubs' fandom in Twitter are analyzed, which allows the extraction of the real-world structure of supporters (Weller and Bruns 2013; Coche 2014). Moreover, tweets can be used to derive a proxy of the way people support their national squad during tournaments, such as the FIFA World Cup (Pacheco et al. 2015).

More recently, tweets were used to characterize football supporters based on the amount of attention they give to a set of national clubs (Pacheco et al. 2016). The authors aggregated individual's characterization to yield a collective characterization of the clubs they support. Their approach was able to capture the real rank of clubs' supporters, and to identify levels of rivalries among clubs. In this work, we use a similar methodology, but this time we group individuals' characterization by regions. In other words, we characterize geographical regions according to tweets mentioning football clubs.

Methods

Characterizing Regions

As proposed in (Pacheco et al. 2016), one can characterize users based on the amount of attention given by them to a set of football clubs. This work proposes to characterize a region as the aggregation of its users. A user belongs to all regions where he/she tweeted. Therefore, supporters (Twitter users) can be characterized in two ways:

Global – one characterization per user based on all his/her tweets. Focus on a single user characterization, and how it influences the regions where he/she interacts.

Local – a different characterization per region he/she has tweeted; focus on regions.

In this work, we use the local characterization to understand the role of regions over users' behavior and their social disorganization.

In this work, a region is a Lower Layer Super Output Area (LSOA), which is a hierarchy proposed to improve the reporting of small area statistics about geographical areas of England and Wales. Thus, each type of event (crime occurrences or tweets) is aggregated by LSOA.

Formally, we measure the frequencies of mentions to clubs by users in each region. The normalized contingency matrix $\hat{U} = [\hat{u}_{ij}]_{m \times n}$ represents m users and their relationship with n clubs, where each row represents a user in a region. The unit of attention is a tweet, so multiple mentions in a tweet are considered "divided attention." Thus, the attention from user i to club j is given by:

$$u_{ij} = \sum_{t \in \tau_i} \frac{\mathcal{W}_j(t)}{C_t}, \tag{1}$$

where $\mathcal{W}_j(\cdot)$ is a function that returns the number of mentions club j receives in a tweet, and $C_t = \sum_{j=1}^{n} \mathcal{W}_j(t)$ is the number of mentions in that tweet. Finally, to ensure all users are treated equally regardless the number of tweets they send, we normalize them so the rows of the matrix $\hat{U}$ sum to 1.

$$\hat{u}_{ij} = \frac{u_{ij}}{\sum_{j=1}^{n} u_{ij}} \tag{2}$$

Regions, on the other hand, are characterized by the aggregation of their supporters. Therefore, the region-characterization matrix $K = [k_{ij}]_{r \times n}$ represents r regions in relation to n clubs, and it is given by

$$K = (L^{\mathsf{T}} L)^{-1} L^{\mathsf{T}} \hat{U}, \tag{3}$$

where $L = [l_{ij}]_{m \times n}$ is a membership-indicator matrix identifying to which region each user belongs. In other words, l_{ij} is 1 if user i tweeted in region j, and is 0 otherwise. Finally, the rows of K represent regions as probability distributions.

Data sources

To characterize regions based on football supporters, we used a Twitter dataset capturing part of the 2014/2015 English/Welsh Premier League. Tweets were gathered using the *Twitter Streaming API* by tracking verified Twitter accounts (@) and hashtags (#) from clubs. We do not need to apply natural language processing to identify mentions to clubs since accounts and hashtags are indexed by Twitter as specific entities. Therefore, a mention is a simple lookup comparing these entities and the official terms for each club. Table 1 shows all terms used and which clubs were tracked.

Table 2 presents some statistics about our dataset. It is a subset of a larger dataset where we selected only geo-tagged tweets within the LSOA regions. Around 90,000 tweets were collected over a 3-month period, representing around 25,000 different users.

The dataset containing the crime-related events that occurred in the UK was retrieved from the Open Data project (*data.gov.uk*) filtered by the period of the Twitter dataset, i.e. from February to May of 2015. Table 3 describes the different types of crime in this dataset and the numbers of crimes that occurred in the period considered.

Table 1: Official Twitter accounts and hashtags of the football clubs considered in our analysis.

Club	Account	Hashtag
Arsenal	@arsenal	#AFC
Aston Villa	@AVFCOfficial	#AVFC
Burnley	@BurnleyOfficial	#BURNLEYFC
Chelsea	@ChelseaFC	#CFC
Crystal Palace	@CPFC	#CPFC
Everton	@Everton	#EFC
Hull	@HullCity	#HCAFC
Leicester	@LCFC	#LCFC
Liverpool	@LFC	#LFC
Man City	@MCFC	#MCFC
Man Utd	@ManUtd	#MUFC
Newcastle	@NUFC	#NUFC
Qpr	@QPRFC	#QPR
Southampton	@SouthamptonFC	#SAINTSFC
Spurs	@SpursOfficial	#THFC
Stoke	@stokecity	#SCFC
Sunderland	@SunderlandAFC	#SAFC
Swansea	@SwansOfficial	#SCAFC
West Brom	@WBAFCofficial	#WBA
West Ham	@whufc_official	#WHUFC

Table 2: Statistics from the football Twitter dataset.

Statistic	Value
Start Data Collection	07/Feb/15
Finish Data Collection	07/May/15
Number of Days	89
Tweets with Mentions	89,416
Users with Mentions	24,974
Tweets per User	3.58
Tweets per Day	1,005

The Entropy of Football Supporters

Our hypothesis is based on the social disorganization theory which states that crime is related to the disorganization of the society in a region (Henry and Einstadter 2006). Thus, we need to define social disorder in a football context. The rate of similarity among supporters may be a proxy for disorder. For instance, places with similar individuals are less prone to conflict than those with people with conflicting ideas. Hence, we argue that regions with single club supporters are more socially organized than places containing supporters from several clubs.

Thus, we use entropy to measure the level of social-football disorganization in a region. The supporter probability distribution of a region $\mathbf{k}_i$ (rows of K, see Eq. 3) is the aggregation of all supporters' behavior ($\hat{\mathbf{u}}_i$) in that region. Therefore, we can calculate the normalized entropy $\hat{S}_i$ of a region i as follows:

$$S_i = -\sum_{j=1}^{n} k_{ij} \cdot \log k_{ij} \quad \text{with} \quad \hat{S}_i = \frac{S_i}{S^+}, \tag{4}$$

where S^+ is the maximum entropy that a region can assume. S^+ happens when supporters are evenly distributed among all clubs (maximum social-football disorder), i.e. when $k_{ij} = {}^1/_n$ for all j clubs.

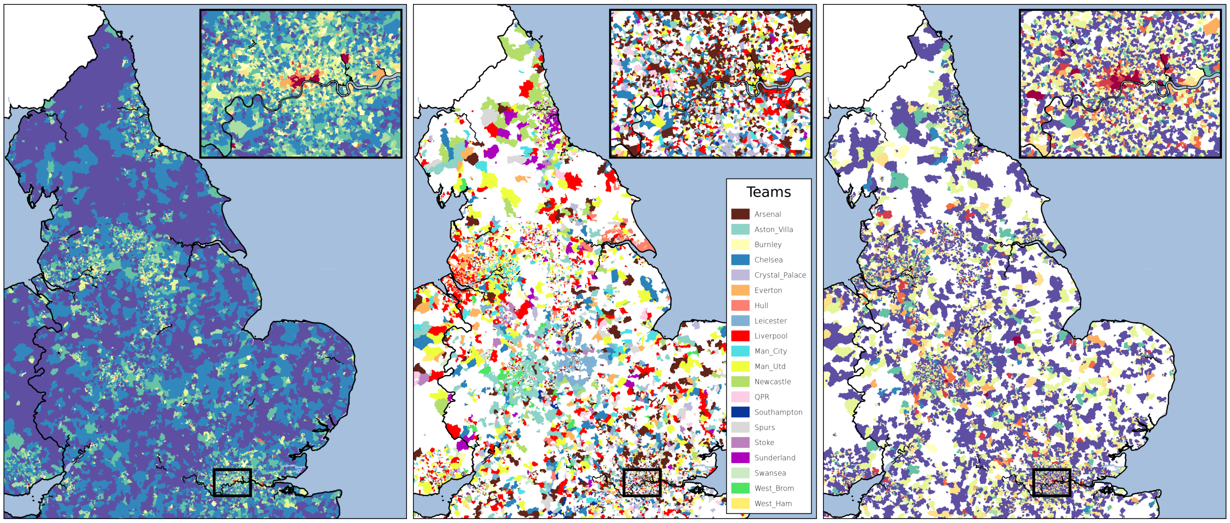
Results and Discussion

In this paper, we proposed to measure the entropy of football supporters' diversity in a region (Equation 4), and to analyze to which extent it can be used to measure social disorganization. Since social disorganization has already been shown to correlate with crime activity (Henry and Einstadter 2006), we use football as a proxy for social disorganization and correlate it with crime.

Figure 1(a) depicts the spatial distribution of crimes in the UK during the period of our dataset by aggregating criminal events within each LSOA. The resident population size in each LSOA is normally distributed with an average (standard deviation) equal to 1584 (279). Still, the heatmap in this figure shows the existence of spots of high criminal rate, an aspect of crime supported by the literature of crime concentration (Henry and Einstadter 2006). Figure 1(b) shows the spatial distribution of the clubs with the most supporters in each region when tweets are aggregated in the same LSOAs. Although many clubs are the most popular in many different regions, some patterns can be seen by examining the clubs and their spatial neighbors, for instance: Newcastle (green) and Sunderland (purple) are prevalent in their regions in the Northern UK; Liverpool (red) and Everton (orange) are predominant in the Liverpool area; and there are some clubs quite popular everywhere on the map, like Arsenal (magenta) and Manchester United (yellow). Such simple analyses suggests the region characterization based on tweets is able to capture football supporters' preferences. Moreover, Figure 1(c) shows the spatial distribution of the entropy of football supporters based on tweets in each LSOA with at least two users. Although the patterns in the entire map are not clear, the inner map (London area) presents visual similarities with the inset of the criminal map in Figure 1(a).

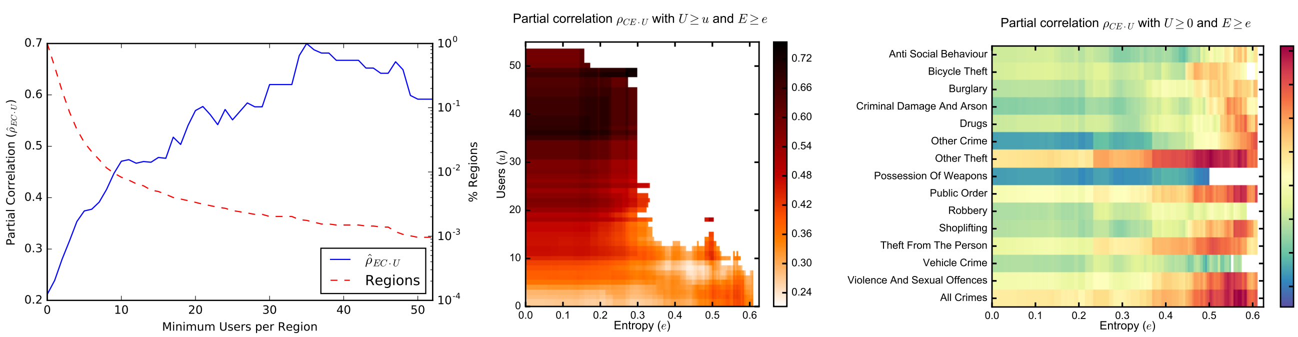
Due to the evidence that population size and crime rate are related (Bettencourt et al. 2007), an estimator for the actual population size (*i.e.* transient population) in each place would probably already reveal a relationship with crime occurrence in a given region. The population information in each LSOA from censuses can not capture these dynamics, since census takes into account only the *residents* of the regions. In fact, Twitter has been used to find such transient population, an additional piece of information that improves crime prediction (Malleson and Andresen 2015a).

In order to address the contribution of supporters' diversity on crime estimation, we need first to analyze how the number of tweets and users in a region correlate with the number of crimes. This would be the base for a null model that provides alternative information regarding tweets; this null model possibly is able to capture the influence of a transient population of regions on crime. To this extent, we calculated the correlations $\hat{\rho}_{TC}$, $\hat{\rho}_{UC}$, and $\hat{\rho}_{EC}$ between the number of tweets, the number of users, and the supporters' entropy, respectively, with crime activity in each region, as shown in Table 3. The correlation values for entropy $\hat{\rho}_{EC}$ are higher than for users $\hat{\rho}_{UC}$, that are also higher than for tweets



(a) (b) (c)

Figure 1: Plots of UK with inner plot from London area of: (a) crime, (b) regions labeled according to highlighted football clubs, and (c) entropy of football supporters.



(a) (b) (c)

Figure 2: Our proxy to social disorganization better explains crime rate in disorganized and populated places. (a) The partial correlation analysis $\hat{\rho}_{EC\cdot U}$ between *all crimes* and *supporters entropy*, controlling for *number of users*, increases as the minimum users increases, but the cover percentage of regions decreases. The entropy of regions in UK is (b) positively correlated with crime rate and this correlation tends to increase as we filter locations by the amount of users and entropy. In fact, regardless of crime type, (c) places with higher entropy present stronger correlation between crime and entropy. The white parts in (b) and (c) represent correlation with high p-values or the lack of points with a certain entropy value.

$\hat{\rho}_{TC}$, regardless of crime type taken into account. Thus, apparently the number of users in a region explains more crime activity than the number of their tweets, and the entropy of supporter diversity explains crime even more than the number of users. However, this direct comparison may lead to wrong conclusions, especially when the random variables could be correlated between each other. Thus, we also calculated the partial correlation $\hat{\rho}_{EC\cdot U}$ between the entropy of supporter diversity and crime activity, controlling for the effect of the number of users. The results show that the entropy correlation is not driven by a population effect.

The use of entropy as a proxy of disorder of a place and the correlation values found provide support to the social disorganization theory. However, due to the many aspects of crime and criminology, two points need to be raised: *(i)* social disorganization is not the only factor that explains crime, i.e. other factors can also drive the increase of crime in places with high or low levels of disorder; and *(ii)* the places without disorder are also subject to criminal occurrences. Due to these intrinsic difficulties in the theories from criminology, we need to analyze the user and entropy ranges in which crime rate is better explained.

The correlations shown in Table 3 considered regions with at least 5 users. Figure 2(a) addresses the impact of the minimum number of users in a region and the partial correlation $\hat{\rho}_{EC\cdot U}$ for all crimes. Intuitively, as we increase the minimum users' constraint, the number of regions considered in the calculation decreases significantly. On the other hand,

Table 3: Statistics, correlations and regression summary per type of crime. The number of crimes within the period Feb/15 – May/15. Correlations and regressions considering regions with at least 5 users. The correlations $\hat{\rho}_{TC}$, $\hat{\rho}_{UC}$, and $\hat{\rho}_{EC}$ between the tweets count, the number of users, and the entropy of supporters in regions of UK, respectively, with criminal occurrence are increasingly stronger. The partial correlation $\hat{\rho}_{EC \cdot U}$ between entropy and crime whilst controlling the effect of the number of users in a region is always greater than or equal to $\hat{\rho}_{EC}$. The adjusted R^2 for the model $crime \sim user + entropy$ is significantly greater than for the model $crime \sim user$. Regression coefficients are the intercept α, user β_u and entropy β_e.

Type of Crime	# Crimes	$\hat{\rho}_{TC}$	$\hat{\rho}_{UC}$	$\hat{\rho}_{EC}$	$\hat{\rho}_{EC \cdot U}$	Adj.$R^2_{C \sim U}$	Adj.$R^2_{C \sim U+E}$	α	β_U	β_E
Anti-Social Behavior	593,238	0.06	0.12	0.24	0.25^c	0.013^c	0.073^c	-4.19	0.46^c	133.49^c
Bicycle Theft	26,177	0.01	0.06	0.25	0.25^c	0.002	0.063^c	-3.20^c	0.03^b	19.30^c
Burglary	129,974	0.01	0.07	0.26	0.26^c	0.003^a	0.070^c	2.53^c	0.03^b	13.42^c
Criminal Damage and Arson	170,953	0.05	0.10	0.24	0.24^c	0.008^c	0.065^c	3.52^c	0.05^c	17.82^c
Drugs	46,907	**0.08**	**0.15**	0.26	0.27^c	**0.023^c**	0.090^c	-2.26^b	0.08^c	18.68^c
Other Crime	18,582	0.04	0.09	0.19	0.19^c	0.007^c	0.043^c	-2.07^c	0.03^c	9.01^c
Other Theft	158,332	0.06	0.12	**0.38**	**0.39^c**	0.014^c	**0.163^c**	-29.75^c	0.29^c	127.02^c
Possession of Weapons	7,214	0.03	0.06	0.17	0.17^c	0.002^a	0.031^c	-0.81^b	0.01^a	4.28^c
Public Order	57,969	0.06	**0.15**	0.31	0.32^c	0.021^c	0.121^c	-4.22^c	0.09^c	28.12^c
Robbery	16,245	0.04	0.09	0.25	0.26^c	0.008^c	0.073^c	-1.24^c	0.02^c	8.31^c
Shoplifting	111,714	0.05	0.12	0.26	0.26^c	0.013^c	0.080^c	-17.83^c	0.30^c	93.04^c
Theft From the Person	25,960	**0.08**	0.12	0.33	0.33^c	0.014^c	0.124^c	-19.83^c	0.18^c	68.84^c
Vehicle Crime	115,826	0.02	0.05	0.21	0.21^c	0.001	0.045^c	3.06^c	0.02^a	12.43^c
Violence and Sexual Offenses	304,348	0.05	0.12	0.32	0.33^c	0.014^c	0.122^c	-10.34^c	0.24^c	88.49^c
All Crimes	1,783,439	0.06	0.14	0.34	0.35^c	0.019^c	0.140^c	-86.62^c	1.85^c	642.26^c

Significance codes: $^a\rho < 0.10$, $^b\rho < 0.05$, and $^c\rho < 0.01$. For adjusted R^2, significance based on F test.

$\hat{\rho}_{EC \cdot U}$ increases as we keep the more populated regions, suggesting our proxy for social disorganization explains crime rates better in more populated areas, or at least, in regions more represented on Twitter. Figure 2(c) depicts the relationship between entropy versus the amount of crime in each LSOA. This plot shows that entropy explains crime better when places with higher entropy are taken into account, a finding that holds true regardless of crime type. Yet, Figure 2(b) shows that the compound effect of increasing both users and entropy also reflects in increasing the partial correlation between all crimes and entropy of supporters. Therefore, our proxy works better explaining crime in places more populated and more disorganized.

In order to assess the explanation power of the entropy of supporters over crime, we constructed two linear models:

$C = \alpha + \beta_U U + \epsilon$ in which the variation in the number offenses in a region is explained by the number of Twitter users in that region.

$C = \alpha + \beta_U U + \beta_E E + \epsilon$ in which the variation in the number of offenses in a region is explained by the number of users and our measure of social disorganization.

Table 3 shows the adjusted R^2 for both models and the regression coefficients of the model . As expected, the contribution of number of users (β_U) is significant for most types of crime. Moreover, the contribution of the entropy of supporters (β_E) is significant regardless the type of crime, and its addition to regression causes a significant increase in the variance explained by the model (adjusted R^2). Nonetheless, we do not expect that football can explain all types and occurrences of crime, or that football is the sole component that leads to crime, since it is known that many other factors may lead to crime, such as population education level, social unrest, and economic opportunities, to name just a few (Malleson and Andresen 2015b).

Conclusions

We proposed a measure for social disorganization in a region by analyzing the entropy of online social media data from users in that region. We carried out experiments using football supporters conversations and the diversity of clubs they mention in their geolocated tweets; more specifically, we used football-related tweets from the UK. We observed a significant correlation between the number of users and crime, and between the entropy of supporters and crime. Then, we measured the partial correlation between them to confirm that the entropy correlation was not inflated by the population effect. Finally, we used regression models to confirm the contribution of both, the number of users and their entropy, to model crime activity. The coefficients for entropy are statistically significant, regardless the type of crime. For instance, when considering regions with at least 5 users and for all crimes, the model incorporating the entropy explains 7 times more the variance of crime (adj. R^2) than the model without it. We also found that our proposed measure of social disorganization explains better the variation of crime among regions with higher disorganization and larger population. This work is a first attempt to create a framework to assess the levels of social disorganization in locations by using social media. It is worth noticing that this is the contribution of this paper and that the example with the UK football is a case study to demonstrate that certain subjects (in this case football) can be used to quantify social disorganization.

Although we found a positive correlation between our measure of social disorganization and crime, we aggregated longitudinal data in such way that the entropy of places and the amount of crimes were analyzed without taking into account their variations over time. This was the case mainly due to the temporal granularity of the data sets, i.e. monthly, provided by the police forces in the UK. Still, one of the benefits of our proposal is the capability to assess places in a real-time fashion; thus, as future work, and by the possession of richer criminal data sets, we want to find the minimum time window to extract significant characterizations of places, as well as to capture the movement of social disorganization in places over time and its relationship with crime mobility. We also want to test the impact of characterizing regions based on a global approach; that is, based on people's behavior everywhere, and not on the behavior of people in a particular region (local approach as defined in Methods section). Moreover, we intend to examine other factors that can also be used to quantify social disorganization and how they can aggregate value to the entropy of football in a predictive model. For instance, if we look at what people eat and calculate the entropy of their choices, would we get similar correlation numbers, or is there something special about football? How about the diversity of the languages spoken in a region? Can these seemly unrelated social factors be combined in a calculation of social disorganization? It is generally difficult to get several datasets related to different factors for the same geographical regions. However, if such datasets are made available, the approach proposed here can be easily applied.

Acknowledgments

MO and DP would like to thank CAPES–Brazil for the financial support under grants 1032/13-5 and 0544/13-2.

References

Bettencourt, L. M.; Lobo, J.; Helbing, D.; Kühnert, C.; and West, G. B. 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences* 104(17):7301–7306.

Chen, X.; Cho, Y.; and young Jang, S. 2015. Crime prediction using twitter sentiment and weather. In *Systems and Information Engineering Design Symposium (SIEDS), 2015*, 63–68. IEEE.

Coche, R. 2014. What women's soccer fans want: a twitter study. *Soccer & Society* 15(4):449–471.

Compton, R.; Lee, C.-K.; Lu, T.-C.; de Silva, L.; and Macy, M. 2013. Detecting future social unrest in unprocessed twitter data:"emerging phenomena and big data". In *2013 IEEE International Conference On Intelligence and Security Informatics (ISI)*, 56–60. IEEE.

Esmin, A. A.; Júnior, R. S.; Santos, W. S.; Botaro, C. O.; and Nobre, T. P. 2014. Real-time summarization of scheduled soccer games from twitter stream. In *Natural Language Processing and Information Systems*. Springer. 220–223.

Gerber, M. S. 2014. Predicting crime using twitter and kernel density estimation. *Decision Support Systems* 61:115–125.

Henry, S., and Einstadter, W. J. 2006. *Criminological theory: An analysis of its underlying assumptions*. Rowman & Littlefield Publishers.

Kubrin, C. E., and Weitzer, R. 2003. New Directions in Social Disorganization Theory. *Journal of Research in Crime and Delinquency* 40(4):374–402.

Malleson, N., and Andresen, M. A. 2015a. The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science* 42(2):112–121.

Malleson, N., and Andresen, M. A. 2015b. Spatio-temporal crime hotspots and the ambient population. *Crime Science* 4(1):10.

Niu, R.; Zhang, J.; and Ebert, D. S. 2015. Classification and visualization of crime-related tweets. In *The Summer Undergraduate Research Fellowship (SURF) Symposium*.

Pacheco, D. F.; Lima-neto, F. B.; Moyano, L. G.; and Menezes, R. 2015. Football Conversations : What Twitter Reveals about the 2014 World Cup. In *Brazilian Workshop on Social Network Analysis and Mining (CSBC 2015 - BraSNAM)*.

Pacheco, D. F.; Pinheiro, D.; Lima-Neto, F. B.; Ribeiro, E.; and Menezes, R. 2016. Characterization of Football Supporters from Twitter Conversations. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI' 16)*, 169–176.

Patton, D. U. 2015. Gang violence, crime, and substance use on twitter: A snapshot of gang communications in detroit. In *Society for Social Work and Research 19th Annual Conference: The Social and Behavioral Importance of Increased Longevity*. Sswr.

Sampson, R. J.; Raudenbush, S. W.; and Earls, F. 1997. Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. *Science* 277(5328):918–924.

Van Oorschot, G.; Van Erp, M.; and Dijkshoorn, C. 2012. Automatic extraction of soccer game events from twitter. In *Proc. of the Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web*, volume 17.

Wang, M., and Gerber, M. S. 2015. Using twitter for next-place prediction, with an application to crime prediction. In *Computational Intelligence, 2015 IEEE Symposium Series on*, 941–948. IEEE.

Wang, X.; Gerber, M. S.; and Brown, D. E. 2012. Automatic crime prediction using events extracted from twitter posts. In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer. 231–238.

Weller, K., and Bruns, A. 2013. Friends or followers? the relations between german soccer clubs and their fans on twitter. *Selected Papers of Internet Research* 1–6.

Williams, M. L.; Burnap, P.; and Sloan, L. 2016. Crime sensing with big data: The affordances and limitations of using open source communications to estimate crime patterns. *British Journal of Criminology*.

Xu, J.; Lu, T.-C.; Compton, R.; and Allen, D. 2014. Civil unrest prediction: A tumblr-based exploration. In *Social computing, behavioral-cultural modeling and prediction*. Springer. 403–411.