Relationships Between Crime and Twitter Activity Around Stadiums

Alina Ristea, Chad Langford
Department of Geoinformatics, Doctoral College
University of Salzburg
Salzburg, Austria
mihaela-alina.ristea@sbg.ac.at; chad.langford@sbg.ac.at

Michael Leitner
Department of Geography and Anthropology
Louisiana State University
Baton Rouge, USA
mleitne@lsu.edu

Abstract—Research shows that public events can be violence attractors or generators. Recent studies focus on the effect of crowd events on spatial crime analysis, using environmental criminology theories as a background. Also, social media analyses show possible influences of social media in crowd movements. This study displays and correlates spatial patterns of crimes and tweets around two stadiums in Manchester, United Kingdom. Moreover, text analysis of tweets is used for extracting violent tweets. For this study the data used are: aggregated monthly crime data analyzed for 2km around the two stadiums and geo-located tweets for the same study area. Spatial interpolation and spatial statistics are used to determine relationships between datasets. Results show differentiated influence of Twitter data on crime for the two stadiums, benefits of using tweets subsets, and they also support the importance of using disaggregated crime types.

Keywords- crime; stadiums; Twitter; spatial correlation; violent tweets

I. INTRODUCTION

Environmental criminology provides an important theoretical foundation in exploring relationships between events and crimes, including Routine Activity Theory [1] for place-based explanations of crime or Crime Pattern Theory [2], related to the rise of crime in specific areas. Recently, researchers found spatial crime correlations in sport events using various approaches, different regions of interest, and types of crimes [3-6]. In football related crimes, researchers emphasize connections with criminal damage, theft-and-handling or violent offenses in Elland Road Stadium, Leeds [7], and also violent crime and theft and handling for Wembley Stadium, Northern London [8]. Additionally, it has been shown that social media platforms escalate the collective action of violent incidents after events [9]. Moreover, a growing amount of literature suggests and explains spatial relationships between tweets and crime density [10-16], and also includes social media information in crime prediction models [17-19].

There is, however, little research showing how social media can help improve the understanding of spatial crime patterns for events. This study focuses on the spatial correlation between crimes and Twitter activity in the immediate proximity of two stadiums in Manchester, United Kingdom. The analysis shows spatial relationships between crimes and geo-located tweets and it uses text mining for identifying violent tweets and football topics.

II. DATA

A. Data

For this study the data used include aggregated monthly crime data from police.co.uk for 2012, excluding the months with no football games (June, July), analyzed for the area around the two stadiums, Manchester United (Old Trafford) and Manchester City (Etihad) located in the area of Greater Manchester. Because of the low temporal resolution of crime data, this study cannot use crimes occurring just in days when football matches are played in both stadiums. In addition, Twitter data are processed and correlated with crime data to find relationships between crime and social media spatial clusters, as well as relationships between violent tweets and disaggregated crime types. The temporal resolution of the Twitter data is fine, which was helpful on extracting tweets subset only from football game days. The geo-located tweets used were obtained through the Twitter Streaming Application in 2012 [20]. Although Twitter data are often used in research, the tweets that are received via the streaming application represent 1-10% of all tweets [21-22].

B. Study area

This study considers two stadiums of the best teams in 2012 Premier League Championship [23], located in Manchester, United Kingdom (Figure 1). Manchester United fans are highly involved in disorder and violence, as well as Tottenham Hotspur, Leeds United, and Liverpool fans [24].

III. METHODOLOGY

The analyzed areas are represented by a 500m Euclidean distance grid around two stadiums up to a distance of 2km. However, the Twitter data have a valuable temporal scale, so besides using all geo-located tweets from 2012, we extracted just the home game days for both stadiums and we used this new variable in addition to the all geo-located messages to test the correlation with crime. Also, the study highlights text analytics by extracting tweets which include, as a minimum, one violent word, using a dictionary [25].

This research was funded by the Austrian Science Fund (FWF) through the Doctoral College GIScience at the University of Salzburg (DK W 1237-N23).

Authorized licensed use limited to: University of Melbourne. Downloaded on May 04,2022 at 09:17:50 UTC from IEEE Xplore. Restrictions apply.

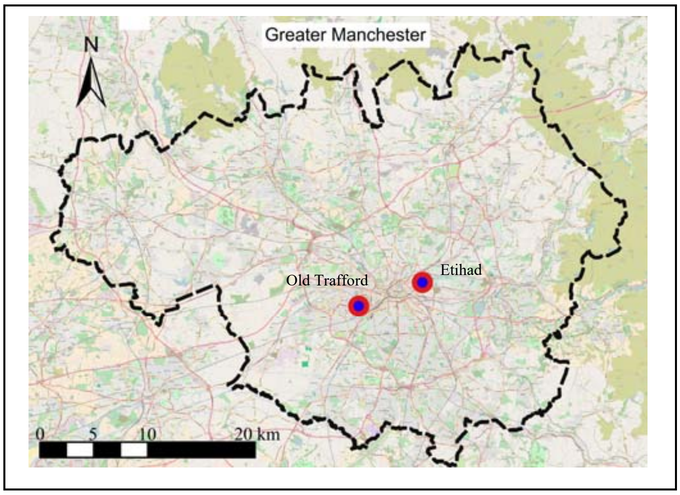

Figure 1. Study area: Greater Manchester, United Kingdom

Crime and Twitter data are the major variables and they were pre-processed regarding crime types, temporal filtering for home game days, semantic analysis for violent tweets extraction using bag of words filtering. The aggregated crimes volume is higher for the Manchester City area, with 10,957 total crimes, than for the Manchester United area, with only 6,892 crimes.

Three Twitter subsets are used in this analysis, namely all geo-located tweets from 2012, excluding June and July (123,911 for United, 158,900 for City), tweets from football home game days in 2012 (15,739 for United, 18,841 for City), and violent tweets - subset of messages including violent words (6,445 for United, 8,404 for City). The Pearson correlation, bivariate spatial autocorrelation [26], and spatial interpolation methods, such as hot-spot [27] and heat map analyses were applied to the major datasets.

IV. Results and Discussions

Preliminary results show relationships between tweets and crime occurrences, mostly for Manchester City stadium surroundings. One explanation might be related with land use and mobility, such as locations of pubs, leisure facilities, or transport stations [28]. Firstly, we tested the crime distribution per month and crime type (Figure 2) and we calculated the correlation of monthly crime from each type between the two study areas. Considering the amount of crimes per month and then compare those monthly values between the stadium areas, the Pearson correlation value was r=0.46.

As expected, Anti social behavior (r=0.84) and Other theft (r=0.48) crime types have stronger variance and correlation between the two stadiums, which confirms football hooliganism in the United Kingdom. The Anti social behavior crime type (2,055 crimes for United, 3,880 for City) covers a wide range of unacceptable activity that causes harm, such as vandalism, street drinking, or dumping of rubbish, common before or after football games. Stadium areas become crowded during the event time frame, which increase crime opportunity when guardians are distracted [1], e.g. for crime types such as Other theft (982 crimes for United, 1,491 for City), namely for snatching property that the victim is carrying. This result supports the importance of disaggregated crime types and the fact that each crime type exhibits different patterns [29].

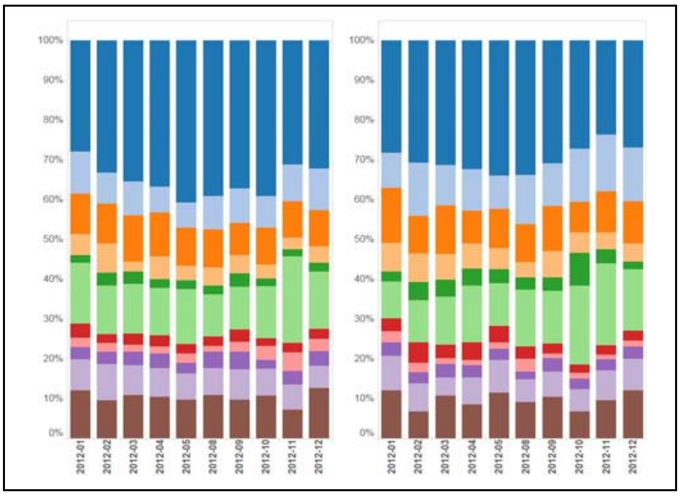

Figure 2. Monthly crime distribution around Manchester City (left) and Manchester United (right) stadiums: Antisocial behavior (at the top), followed by Burglary, Criminal damage and arson, Drugs, Other crime, Other theft, Public disorder and weapons, Roberry, Shoplifting, Vehicle crime, and Violent crime (at the bottom)

We also tested the Pearson correlation for the Twitter data between the two stadiums datasets and received the following results: geo located tweets (r=0.78), game tweets (r=0.56), and violent tweets (r=0.85). Thus far, according to the Pearson correlation aggregated crime occurrences between stadiums are not similar on a monthly basis, although two crime types show similarities as well as Twitter activity, mainly all geo-located and violent tweets.

Secondly, the analysis shows spatial relationships between various crime types and social media behavior. After running spatial autocorrelation on each of the variables the results were positive. None of the spatial autocorrelations are strong, but it is expected considering the complex crime patterns and that crimes are monthly aggregated. Moran's I values in the Manchester United area (Table I) are higher using the violent tweets subset as lag for aggregated crimes and Anti social behavior compared with the other tweets subsets. At the same time, values are always under 0.2, and considering game tweets they are not statistically significant. It is noticed that Anti social behavior has the highest value in relationship with violent tweets (0.16).

Crime types	**Twitter data**		
	Geo located tweets	***Game tweets***	***Violent tweets***
Aggregated crime types	0.1374**	0.0106	0.1545**
Anti social behavior	0.1383**	0.0070	0.1652**
Other theft	0.1402**	0.0427*	0.1169***

*** p=0.001; ** p =0.01; *p=0.1

TABLE I. ManChester United - Bivariate Spatial Autocorrelation Moran's I index

Authorized licensed use limited to: University of Melbourne. Downloaded on May 04,2022 at 09:17:50 UTC from IEEE Xplore. Restrictions apply.

For the Manchester City study area all tweets subsets are significant when correlated with crime (Table II). It is worth mentioning that Other theft has the highest Moran's I value when considering violent tweets (0.33). Between all geo-located tweets and violent tweets the index is not much different, which is worthwhile to be discussed in more detail with football home game day crime dataset analysis.

Crime types	Twitter data		
	Geo located tweets	*Game tweets*	*Violent tweets*
Aggregated crime types	0.2958***	0.2136***	0.3096***
Anti social behavior	0.2428***	0.1751***	0.2544***
Other theft	0.3189***	0.2392***	0.3303***

*** p=0.001; ** p =0.01; *p=0.1

TABLE II. MANCHESTER CITY - BIVARIATE SPATIAL AUTOCORRELATION MORAN'S I INDEX

Thus, we showed that the Manchester United area has a less uniform spatial distribution of crime and tweets (Table I) than the Manchester City area (Table II). Also, violent tweets have a higher spatial correlation with crimes than other all geo-located tweets for 2012 and tweets from football home game days. However since the values only represent monthly crime data, it is possible that the resultant spatial distribution refer to substantially different patterns.

Figure 3 shows the heat map overlay between geo-located tweets and all aggregated crime types. The heat map for Manchester United indicates that a strong intersecting region of geo-located tweets and various highly irregular crime types is located very close or inside the stadium. On the other hand, the heat map for Manchester City shows a weak intersecting region of geo-located tweets and various highly irregular crime types directly where the stadium is located. This may be due to a spatial shifting that occurs within the crime data due to spatial aggregation techniques introduced to increase privacy.

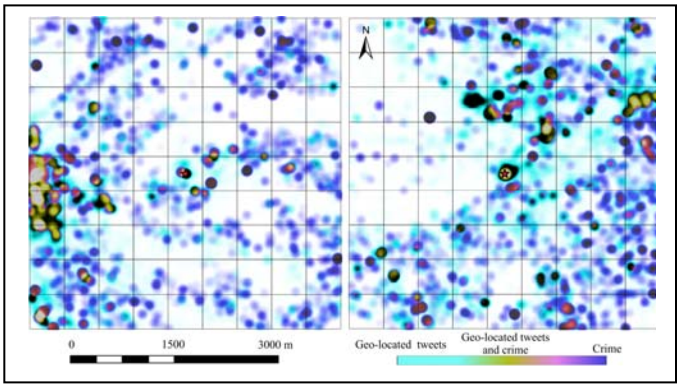

Figure 3. Heat map between geo-located tweets and aggregated crime types for Manchester City (left) and Manchester United (right)

As discussed, violent tweets possess the most significant spatial connection with crime occurrences. Figure 4 represents heat maps for the two study areas and emphasizes the spatial connection between violent tweets and all aggregated crime types. Visually, the situation is similar when using all geo-located tweets for both study areas. However, Etihad Stadium is not shown as a common hot spot for the two variables, with violent tweets focusing more in the surroundings. It is noticed that in the 2km area around Old Trafford Stadium there exists a large space with low crime in the west, which is unexpected. One of the reasons is supported by the "neighborhood watchers", where the Urmston and Stretford Neighbourhood Teams communities are responsible for the area including the Trafford Centre shopping complex and Trafford Park Industrial Estate, one of the biggest in Europe. The policing priorities for the area are specifically targeting "antisocial behavior", "domestic burglary", and "vehicle crime". On the other hand, the land use is mostly suburban and industrial. Violent messages are more numerous but less focused in the western part of the study area, compared with all geo-located tweets.

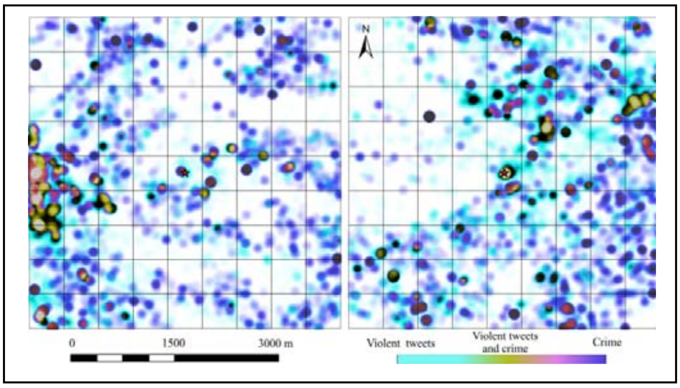

Figure 4. Heat map between violent tweets and all aggregated crime types for Manchester City (left) and Manchester United (right)

Cartographic representations (Figure 5) show similar patterns between Anti social behavior and violent tweets compared to the other two map sets (Figures 3, 4), which can become confusing without a statistical interpretation. Anti social behavior has the highest Moran's I index for Manchester United compared with the other crimes for this stadium (0.16).

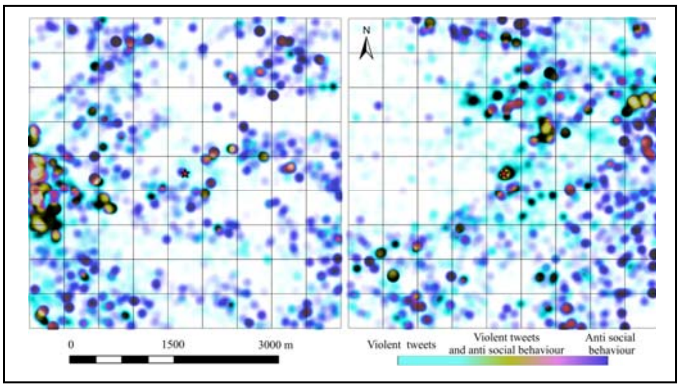

Figure 5. Heat map between violent tweets and Anti social behavior crime types for Manchester City (left) and Manchester United (right)

In contrast, Manchester City has a lower spatial correlation (0.25) compared with Other theft or aggregated crime types. Crime hot spots are more clustered and violent messages are more spread out which might lead to a lower correlation. For Etihad stadium, Other theft shows the highest spatial

Authorized licensed use limited to: University of Melbourne. Downloaded on May 04,2022 at 09:17:50 UTC from IEEE Xplore. Restrictions apply.

correlation with Twitter activity. This crime type can also refer to pickpocketing, which is frequent around the stadium. Offenders work in groups and use distraction methods, this being another application of the Routine Activity Theory [1], including many suitable targets before and after a football game.

Crime types and their spatial relations in crowd based events can vary at different places, as briefly mentioned in the related research section, and as shown at a small resolution in this analysis between two areas of about 2km around stadiums. The same variability can be found also for social media messages. It is important to highlight the need of investigating the spatial structure, e.g. land use, and attributes, e.g. economic factors, for the study cases. However, spatial heterogeneity is not detailed in this study and can be the focus in some future research.

V. Conclusion and Future Work

Crime-tweets relationships analysis shows the possibility of using tweets as one explanatory variable for crime occurrences. Antisocial behavior crime type occurrences and Twitter data distributions follow a similar monthly pattern when correlated between the 2km stadium buffers in Manchester, UK.

Twitter activity has a positive spatial autocorrelation with the density of aggregated crimes around the two selected stadiums and this correlation is higher for Manchester City than for Manchester United. We believe that the results would show more detail using daily crime data. Also, this study introduces the semantic analysis for violent tweets extraction, using a violent words corpus, along with testing messages only from game days. From the three Twitter subsets, violent tweets display higher spatial correlation with the crime data.

This research shows the potential of using geo-located tweets to discuss spatial crime patterns around two stadiums in the same city. This study shows an overall relationship between crime and tweets around stadiums, considering the limitation due to the monthly aggregated crimes. The results strongly suggest studying this issue further in the future focusing on the influence of football games for specific time periods on crime, i.e. a fixed number of hours before and after the event. Moreover, a detailed analysis can focus on the possibility of introducing Twitter data in crime prediction models for football events timeframe.

References

[1] Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. American sociological review, 588-608.

[2] Brantingham, P. J., & Brantingham, P. L. (1981). Environmental criminology. Sage Publications Beverly Hills, CA.

[3] Marie, O. (2015). Police and thieves in the stadium: measuring the (multiple) effects of football matches on crime. *Journal of the Royal Statistical Society: Series A (Statistics in Society*

[4] Yu, Y., Mckinney, C. N., Caudill, S. B., & Mixon Jr, F. G. (2016). Athletic contests and individual robberies: an analysis based on hourly crime data. *Applied Economics, 48* (8), 723-730.

[5] Struse, S. P., & Montolio, D. (2014). The effect of football matches on crime patterns in Barcelona.

[6] Caruso, R., & Di Domizio, M. (2013). International hostility and aggressiveness on the soccer pitch: Evidence from European Championships and World Cups for the period 2000–2012. *International Area Studies Review, 16* (3), 262-273.

[7] Kurland, J., Tilley, N., & Johnson, S. D. (2014). The Football 'Hotspot'Matrix. *Football Hooliganism, Fan Behaviour and Crime: Contemporary Issues*, 21

[8] Kurland, J., Johnson, S. D., & Tilley, N. (2013). Offenses around stadiums: A natural experiment on crime attraction and generation. *Journal of research in crime and delinquency*, 0022427812471349.

[9] Howard, Philip N. and Duffy, Aiden and Freelon, Deen and Hussain, Muzammil M. and Mari, Will and Maziad, Marwa, Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring? (2011). Available at SSRN: https://ssrn.com/abstract=2595096 or http://dx.doi.org/10.2139/ssrn.2595096

[10] Malleson, N., & Andresen, M. A. (2016). Exploring the impact of ambient population measures on London crime hotspots. Journal of Criminal Justice, 46, 52-63.

[11] Malleson, N., & Andresen, M. A. (2015). 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.

[12] Williams, M. L., Burnap, P., & 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*, azw031

[13] Bendler, J., Brandt, T., Wagner, S., & Neumann, D. (2014). Investigating crime-to-twitter relationships in urban environments-facilitating a virtual neighborhood watch.

[14] Bendler, J., Ratku, A., & Neumann, D. (2014). Crime Mapping through Geo-Spatial Social Media Activity.

[15] Corso, A. J. (2015). Toward Predictive Crime Analysis via Social Media, Big Data, and GIS Spatial Correlation. *iConference 2015 Proceedings*

[16] Kounadi, O., Ristea, A., Leitner, M., & Langford, C. (2017). Population at risk: using areal interpolation and Twitter messages to create population models for burglaries and robberies. *Cartography and Geographic Information Science*, 1-15

[17] Featherstone, C. (2013b). The relevance of social media as it applies in South Africa to crime prediction. In *IST-Africa Conference and Exhibition (IST-Africa), 2013* (pp. 1-7). IEEE.

[18] 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). Springer

[19] Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems, 61*, 115-125.

[20] Twitter Streaming API. (2014). *Available: https://dev.twitter.com/docs/streaming-apis (accessed Oct 2014).*

[21] Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. *arXiv preprint arXiv:1306.5204*.

[22] Gao, J. (2016). FINAL REPORT Mining Transportation Information from Social Media for Planned and Unplanned Events.

[23] http://uk.soccerway.com/national/england/premier-league/2011-2012/regular-season/r14829/

[24] James, Mark, and Geoff Pearson. "Legal Responses to Football Crowd Disorder and Violence in England and Wales." Legal Responses to Football Hooliganism in Europe. TMC Asser Press, 2016. 35-52.

[25] Violence vocabulary word list. (2016). *Available online: https://myvocabulary.com/word-list/violence-vocabulary/ (accessed on 1 june 2016).*

[26] Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis, 27* (2), 93-115

[27] Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal, 21* (1), 4-28

Authorized licensed use limited to: University of Melbourne. Downloaded on May 04,2022 at 09:17:50 UTC from IEEE Xplore. Restrictions apply.

[28] Andresen, M. A., & Linning, S. J. (2012). The (in) appropriateness of aggregating across crime types. *Applied Geography, 35* (1), 275-282.

[29] Scholes-Balog, K. E., Hemphill, S. A., Kremer, P. J., & Toumbourou, J. W. (2016). Relationships between sport participation, problem alcohol use, and violence: A longitudinal study of young adults in Australia. *Journal of interpersonal violence, 31* (8), 1501-1530

Authorized licensed use limited to: University of Melbourne. Downloaded on May 04,2022 at 09:17:50 UTC from IEEE Xplore. Restrictions apply.