

Football fan locality- An analysis of football fans tweet locations

Harris N¹, James P¹

¹ Newcastle University, School of Civil Engineering and Geoscience, Newcastle-upon-Tyne, NE1 7RU

November 11, 2014

Summary

This paper looks at the validity of using social media as a dataset for spatial analysis and demonstrates the use of geo-located tweets to investigate the locality of football club fan-bases

KEYWORDS: Social media, Locality analysis, Football

1. Introduction

Twitter is a free social networking and micro-blogging service that enables its millions of users to send and receive messages of up to 140 characters (tweets). Twitter to date has over 284 million users and processes over 500 million tweets a day. Many recent events have been documented live via Twitter users on the ground. For example the hashtag #OccupyCentral became a live feed on information of the recent protest in Hong Kong with 700+ tweets a minute being sent at the peak of the protests (Boehler, 2014).

Approximately 3% of tweets also include location information in the metadata of the tweet. This amounts to 15 million tweets a day that contain location (Dredze *et al.*, 2013). Twitter has been used to map the effects of an earthquake (Yin *et al.*, 2012), the spread of disease (Lan *et al.*, 2012) and many other applications.

Tweets using the official club hashtags from football clubs provide a potentially rich data source to test the assumptions often made about the location of supporters. This is a hotly debated topic on forums and in pubs and assumptions about fan bases such as Manchester United supporters being predominantly based in the South East and their rivals Manchester City being locally based are widely quoted in the press (MEN, 2007). It is also a common assumption that the locality of football fans is inversely proportional to success, with teams that win trophies attracting fans from across the UK. (Dudley, 2014).

1.1 Football Tweets

Data is collected by listening on the official hashtag of each club in the football league e.g. Manchester United's official hashtag is #mufc. The location, the team in question and the time, are captured and stored in a PostGIS relational database. To date we have recorded several million football tweets spanning the last 3 seasons.

1.2 Locality

In order to understand the locality of football fans a definition of "local" is needed. In the food industry the term "local" is often used and definitions exist for the term local. CAMRA (2014) defines a local ale as one produced within 30 miles of the pub. *Farmer's Markets in Clark County* (2014) only count produce within 3 hours travel time as local. Waitrose (2014) considers local recipes as recipes using ingredients from the same county.

2. Analysis

In order to analyse the tweets a subset of the data was created using only tweets made during the October, 2014 international break (6/10/14 00:00 -16/10/14 00:00). This would discount tweets from fans travelling to and from away football matches and reduce tweets about contentious incidents in a game like a dubious red card. A total of 3000 tweets were collected for the Premier League and 1500 for the Championship.

2.1 Tweet distance as a Euclidean measure

Initially the distance between tweet and club was analysed, with the home ground of each club acting as their reference point. Figure 1 shows the results for Premiership clubs. The results of this analysis are inconclusive although they suggest that there is some truth in the assumptions with the bottom of the table filled with the “glory clubs” of Manchester United, Arsenal and Liverpool. More unexpectedly, it also includes Crystal Palace and Swansea City. Swansea’s low score is probably due to their generic official hashtag of “#swans”. At the top of the table are some of the teams that have been recently promoted. Manchester City sits third perhaps suggesting that there is some truth in them being locally supported.

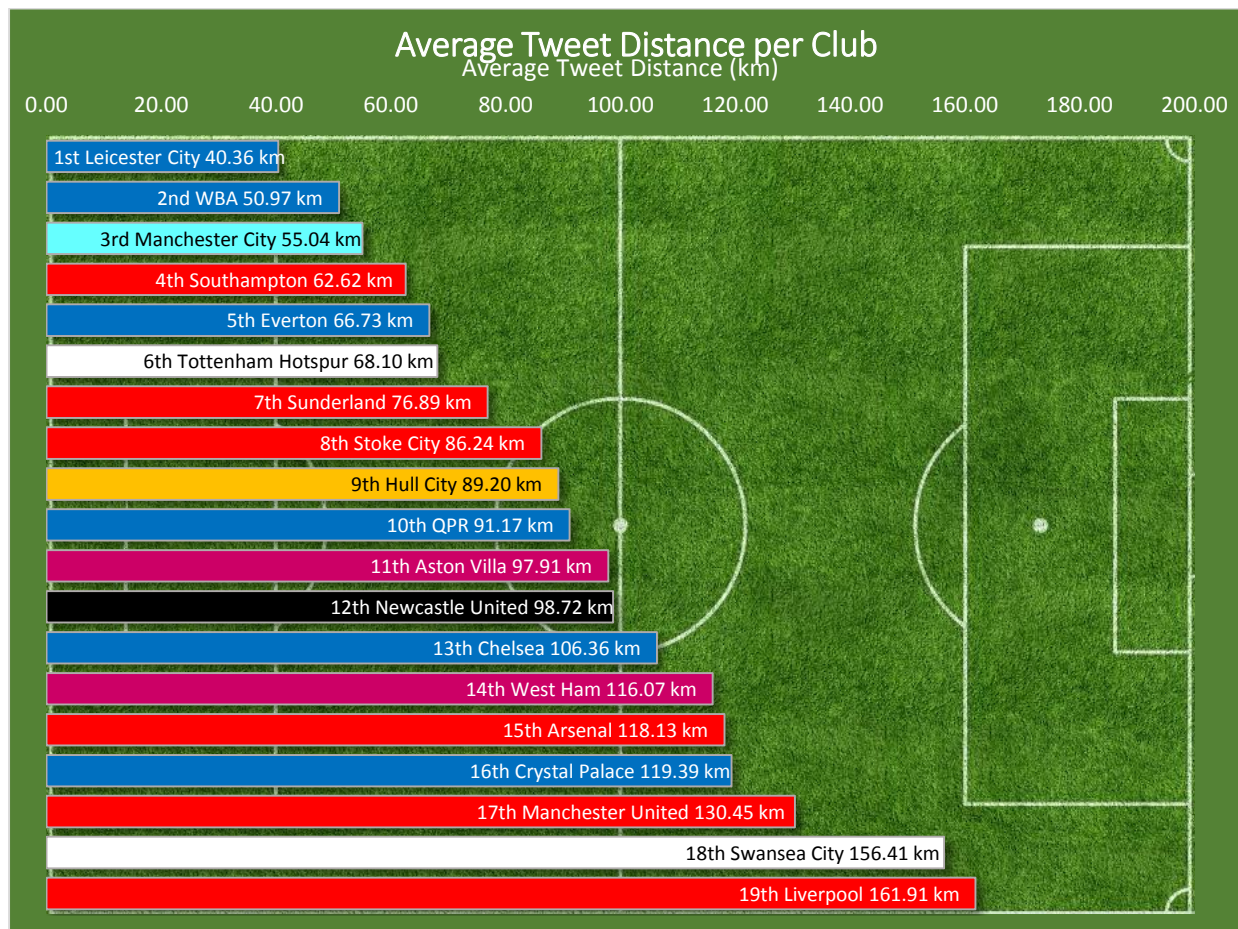


Figure 1: Average Tweet distance per club

Another assumption often quoted by football pundits (*Football club and fan relationships better in lower leagues*, 2014) is that Premiership teams attract supporters outside the immediate geographic area whereas teams lower down the division are more likely to be locally supported. To test this hypothesis the Championship tweets from the same weekend were analysed in the same fashion and compared to the Premiership clubs. Figure 2 goes some way to supporting this assumption with only one Premier League side making it into the top 10 and only 7 in the top half of the table. The worst performing club is surprisingly Blackpool FC which may be due to the fact that at the time of data collection Blackpool were heavily in the news as a result of public disagreements between the owner, manager and fans (SkySports, 2014).

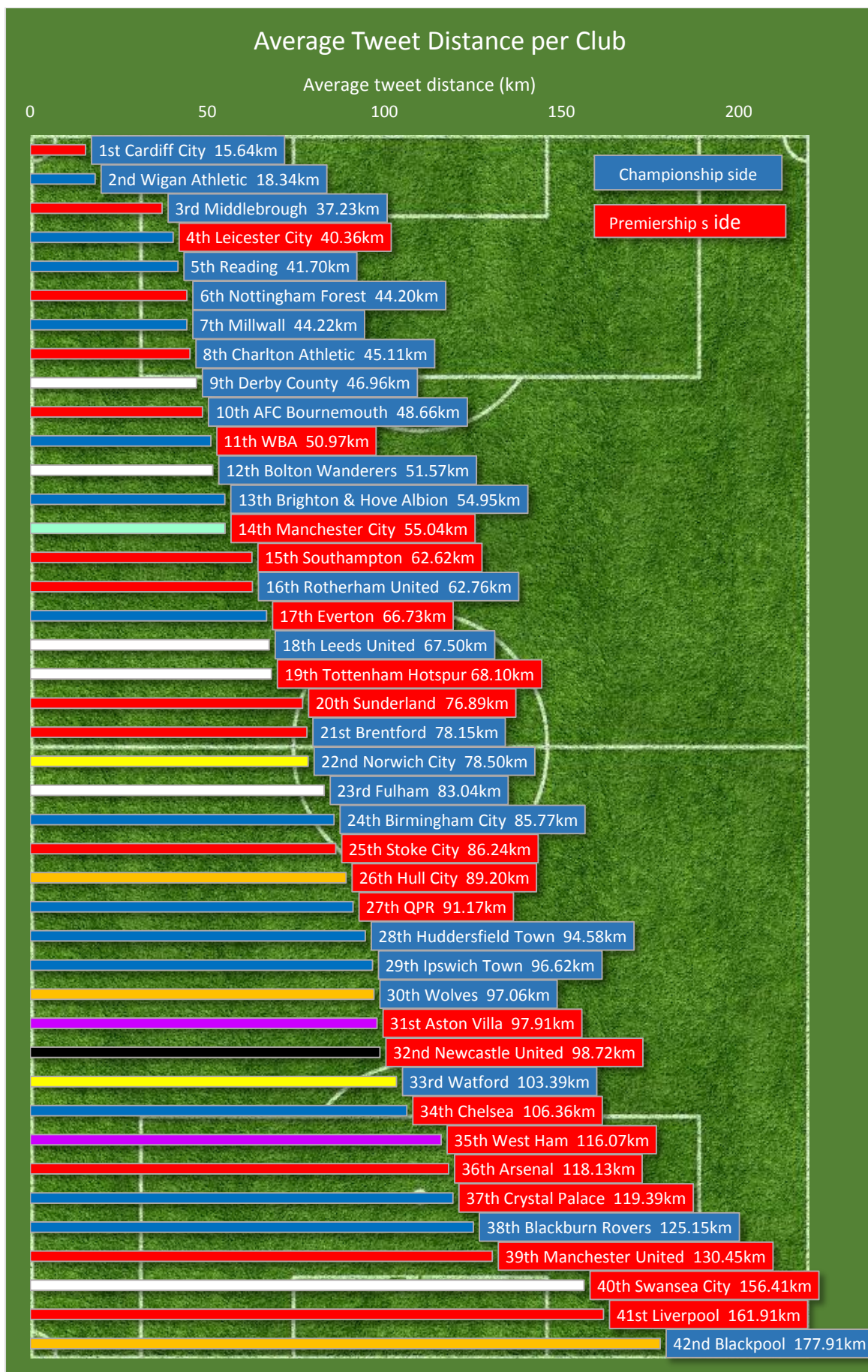


Figure 2 Average tweet distance per club

To further test this hypothesis the average distance from tweet to club per division was computed and the results shown in Figure 3. Here the Championship clearly out scores the Premier League rivals with a difference in distance of 40km in average between the two divisions.

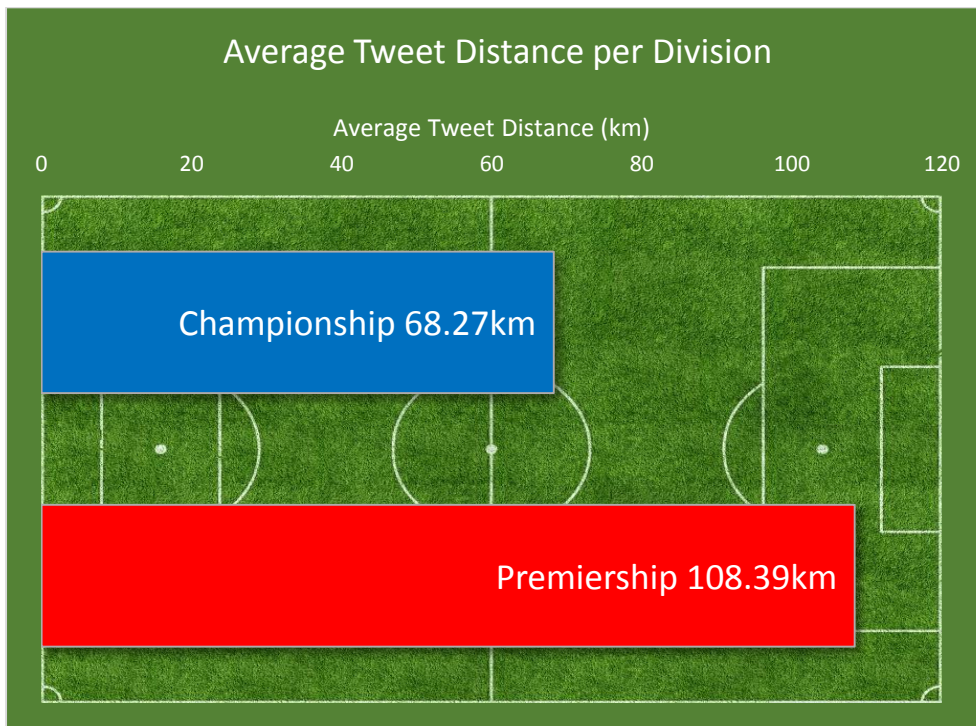


Figure 3 Average tweet distance per division

2.2 Proportion of local tweets

This basic analysis applies a 30 mile buffer around the club as a measure of locality, however if a club is more isolated, like Norwich City, it is possible to be more than 30 miles from the club but it still be the nearest football league club. Therefore a tweet was considered local if it was within 30 miles of the ground **or was the nearest** football league club. Figure 4 shows the results of this test. Again the results are inconclusive with Liverpool and Arsenal and Manchester United in the lower half but accompanied by some other teams like West Ham and Crystal Palace. At the top of the table we have Manchester City and a smattering of some of the smaller clubs.

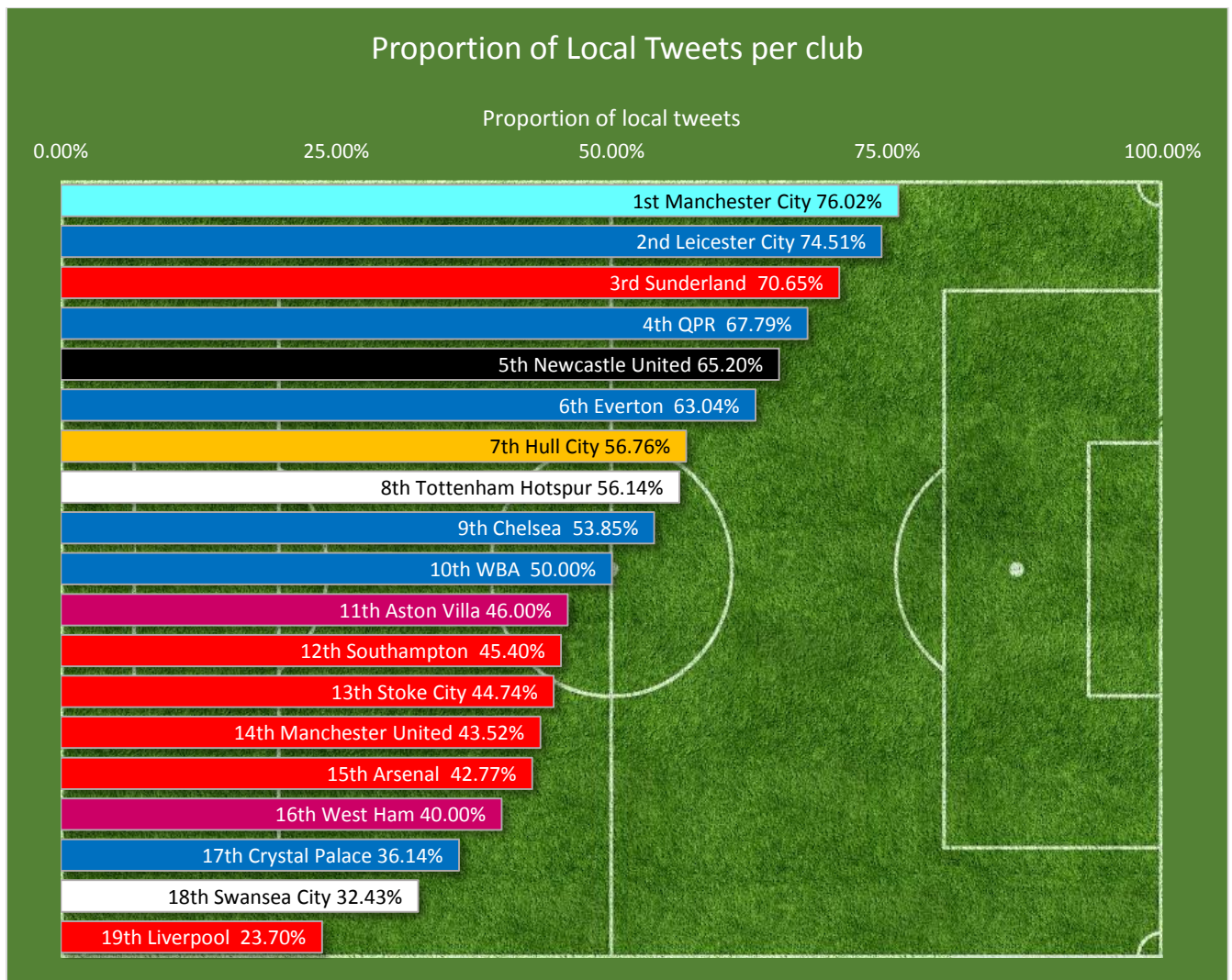


Figure 4 Proportion of local tweets

Again this test was repeated including Championship sides with the results shown in Figure 5. This supports the hypothesis that teams in lower divisions attract a local fan base with only two Premier League clubs now in the top 10 and only 6 in the top half of the table.

Proportion of local tweets per club



Figure 5 Proportion of local tweets

Finally the proportion of local tweets per division was looked at and the results show in Figure 6. Again the Championship scored significantly higher than the Premier League.

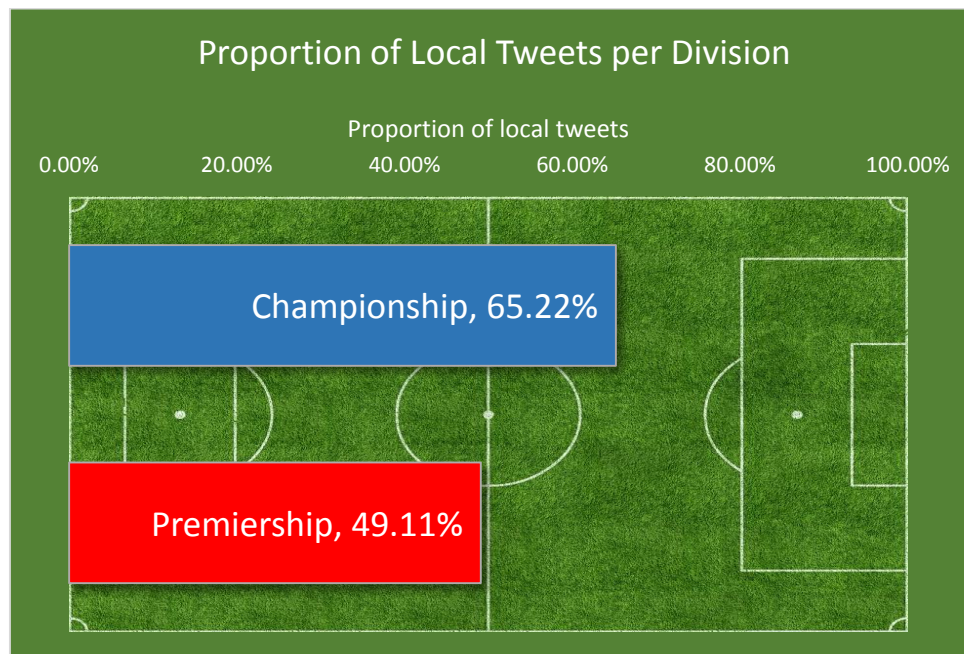


Figure 6 Proportion of local tweets per division

3. Conclusions

This analysis of the tweets for this one weekend broadly supports the hypothesis that Championship teams have a more local fan base than their Premier League rivals.

When the results from both the distance and proportion tests are combined (Table 1) the results suggest that there is some truth to the assumption that the more successful teams have a wider fan base. The top end of the table is dominated by lesser teams with Manchester City and Everton being the exceptions, both of which are a city's second team which may result in their high placing. At the foot of the table, whilst it is Manchester United which has the reputation for the most widespread fan base, it is actually Liverpool who performed worst in both tests. The two notable anomalies in the bottom five are Swansea City and Crystal Palace. It is unclear from the tweet data itself why Crystal Palace does so poorly.

Table 1: Summary of results per club

Pos	Club	Rank on Average Distance	Rank on Proportion of local tweets	Lowest rank score	Average Rank
1	Leicester City	1	2	1	1.5
2	Manchester City	3	1	1	2
3	Sunderland	7	3	3	5
4	Everton	5	6	5	5.5
5	WBA	2	10	2	6
6	QPR	10	4	4	7
7	Tottenham Hotspur	6	8	6	7
8	Southampton	4	12	4	8
9	Hull City	9	7	7	8
10	Newcastle United	12	5	5	8.5
11	Stoke City	8	13	8	10.5
12	Chelsea	13	9	9	11
13	Aston Villa	11	11	11	11
14	West Ham	14	16	14	15

15	Arsenal	15	15	15	15
16	Manchester United	17	14	14	15.5
17	Crystal Palace	16	17	16	16.5
18	Swansea City	18	18	18	18
19	Liverpool	19	19	19	19

Given that the Twitter data is a simple point data set any number of further analyses could be applied to it depending on the nature of the selected concept of “local”. Looking at the proportion of supporters within a region e.g. the South East or considering network travel time rather than distance may also yield some interesting results.

Lovelace *et al.* (2014) discuss the limitations of using social media as a data source in spatial analysis and whilst it is not without its uncertainties it is important not to dismiss it outright as a concept. The sheer volume of tweets provides a rich data source for the research community. Further information on football tweets can be found at: <http://ceg-sense.ncl.ac.uk/footballtweet>.

4. Biographies

Neil Harris is a researcher in Geomatics at Newcastle University. Whose research centres around the collection, management, analysis and visualisation of geospatial sensor and social media data.

Phil James is a Senior Lecturer in GIS. His research centres on the use of Geospatial data to solve and support Engineering problems. Phil is the academic lead on Newcastle's Urban Observatory that integrates real time urban sensors across multiple scales with geospatial data and engineering models using the Cloud.

5. References

- Boehler, P. (2014) *How Hong Kong's #OccupyCentral became a global topic on Twitter*. Available at: <http://www.scmp.com/news/hong-kong/article/1604409/infographic-how-hong-kongs-occupycentral-became-global-topic-twitter> (Accessed: 3/11/2014).
- CAMRA (2014) *CAMRA LocAle*. Available at: <http://www.camra.org.uk/locale> (Accessed: 6/11/2014).
- Dredze, M., Paul, M.J., Bergsma, S. and Tran, H. (2013) 'Carmen: A Twitter Geolocation System with Applications to Public Health'.
- Dudley, B. (2014) 'Sing when you're losing', *Supporters Not Customers*. Available at: <http://supportersnotcustomers.com/2014/04/22/sing-when-youre-losing/>.
- Farmer's Markets in Clark County* (2014). Available at: <http://ext100.wsu.edu/clark/agriculture/small-farms/farmersmarkets/> (Accessed: 6/11/2014).
- Football club and fan relationships better in lower leagues* (2014). Available at: <http://www.staffs.ac.uk/news/football-club-and-fan-relationships-better-in-lower-leagues-staffs-university-researcher-suggests-tcm4233521.jsp> (Accessed: 6/11/2014).
- Lan, R., Lieberman, M.D. and Samet, H. (2012) 'The picture of health: map-based, collaborative spatio-temporal disease tracking'. pp. 27-35.
- Lovelace, R., Malleson, N., Harland, K. and Birkin, M. (2014) 'Can Social media data be useful in spatial modelling?', *GISRUk Glasgow*.
- MEN (2007) *Ben salutes Blue Manchester*. Available at: <http://www.manchestereveningnews.co.uk/sport/football/football-news/ben-salutes-blue-manchester-1117827> (Accessed: 6/11/2014).
- SkySports (2014) *Sky Bet Championship: Blackpool chairman Karl Oyston insists protests won't work* (Accessed: 6/11/2014).
- Waitrose (2014) *The Waitrose Small Producers' Charter*. Available at: http://www.waitrose.com/content/waitrose/en/home/inspiration/about_waitrose/the_waitrose_way/small_producers_charter.html.html (Accessed: 6/11/2014).
- Yin, J., Lampert, A., Cameron, M., Robinson, B. and Power, R. (2012) 'Using Social Media to Enhance Emergency Situation Awareness', *IEEE Intelligent Systems*.