Interpreting the 2011 London Riots from Twitter Metadata

Marc Cheong
Clayton School of IT,
Monash University, Australia
E-mail: marc.cheong@monash.edu

Sid Ray
Clayton School of IT,
Monash University, Australia
E-mail: sid.ray@monash.edu

David Green
Clayton School of IT,
Monash University, Australia
E-mail: david.green@monash.edu

***Abstract*—Social media have rapidly become one of the principal venues for personal and public communication. This makes them rich sources of information about real-world events. As a case study, we used Twitter metadata to investigate social dimensions of the 2011 London riots. The results showed that Twitter-based commentary and participation in the London Riots are closely linked to the real-world manifestation of the riots (e.g. in terms of geographic presence). Twitter metadata on users and their messages during the riots can be used to generate useful inferences which allows us to gain a better insight into intents, information-sharing behavior, and demographics of both the rioters and observers of the riots. Pattern recognition approaches can be used to further reveal latent properties from the acquired inferences.**

***Keywords*-Twitter; London Riots; spatial correlation; demographics; follower-friend ratio; Student's t-statistic; clustering.**

I. Introduction

The 2011 London Riots are a good example of how social media and communications technology played a part in influencing, reporting, and catalyzing real-world events. In brief, the riots took place from 6–11 August 2011 (inclusive), erupting from "a peaceful protest over the police killing of a Tottenham man, Mark Duggan" [1]. On the 7th of August, the rioting, looting, and arson started in parts of London, which subsequently spread to Birmingham, Manchester and Liverpool [1]. The riots continued spreading until 10th August, after which it started to ebb due in part to bad weather and police response [1], [2]. Consequently, the rioting resulted in over 1,000 arrests [2], a rough estimate of £100 million in damages [3], and a few collateral deaths [1]. Theories on the cause of the riots range from "police prejudices, a lack of social mobility, unemployment..." to "welfare dependency... teenage pregnancies... [and] consumerism" [1].

What is notable about the London Riots is the prominent use of social media and mobile technology. BlackBerry Messenger, Facebook, and Twitter were among the oft-mentioned technologies used in the riots for "inciting public disorder" [2]. An example of this is reported by Adams [4], who observed the usage of Twitter to "encourage violence".

The use of such modern communication technologies, Twitter in particular, is a significant shift from traditional forms of communication during riot events. Physical presence and "shouting through a megaphone" [4] were necessary to organize and participate in a riot in prior years. Before the 2011 London Riots, several other riots were catalyzed by the usage of modern forms of communication, namely the 2005 Paris riots which were partly sparked by usage of blogs, and the 2005 Cronulla riots fueled by inflammatory text messages. However, what makes the 2011 London Riots unique compared to other riots is the prevalent usage of Twitter by the rioters and also people talking about the riot.

Twitter, a microblogging service, contains a wealth of metadata [5] about its users and the messages (or *tweets*) traversing its vast network. Such metadata, from both the domains of *users* and *messages*, are a treasure trove of hidden information which we can use to learn more about the riots (and people's reactions to them) as they unfolded.

II. Study Objectives

The broad aim of this paper is to study the commentary and participation in the London Riots as observed via Twitter, and to link observations from both the *user* and *message* domains on Twitter to real-world happenings. Specifically, our study has the following subgoals, which address gaps in extant research:

- **Goal 1:** Constructing a corpus of Twitter message metadata pertaining to the London Riots, together with associated user metadata from their authors (*Section IV*).
- **Goal 2:** Inferring the demographic properties of gender and geographical location of Twitter users tweeting about the London Riots. (*Section V*).
- **Goal 3:** Ascertaining the presence of spatial correlations between real-life riot activity and localized Twitter chatter in England (*Section VI*).
- **Goal 4:** Studying user messaging intent and Twitter online presence during and after the riots (*Section VII*).
- **Goal 5:** Discovering hidden patterns by clustering; given the feature space of user demography, online presence, and messaging intent (*Section VIII*).

III. Literature Review

In prior literature [5], [6], we have illustrated how latent metadata on Twitter can be used to provide useful inferences about the demography, online presence, and messaging habits of Twitter's user base. Of interest is a study in a Twitter-based framework to chronicle civilian response to terrorism events [7], which generates a wealth of information (such as mobility status of a person, and sentiments during terrorism events) from such hidden metadata. On a related topic, Twitter's use during "mass convergence [and] emergency events" – such as disasters and mass political conventions – have been studied by Hughes & Palen [8], who observed that Twitter

978-1-4673-5119-5/12/$31.00

exhibits traits of "information dissemination... broadcasting and brokerage" [8]. In other words, messages during such events exhibit properties of information sharing (URLs) and interpersonal communication (`@user` messages). Similar research [9] was conducted on the 2009 Canadian Red River Valley floods to investigate social information sharing and self-organization by users affected by the tragedy. Meanwhile, other studies such as [10] use Twitter to track the geographic spread of earthquakes over time.

Closely related to our paper is a study conducted by Tonkin & Tourte [11] on Twitter's role in the London Riots. The authors analyze tweets composed during the riots (without going into detail in terms of the users domain). The aims of [11] were, in a nutshell: to see if Twitter was "used as an organizational tool during the riots" ; to discover motivations behind retweets; and postulate potential uses of "real-time data from Twitter" during such events [11]. The study found insufficient evidence to claim that Twitter was "...a central organizational tool to promote [rioting]" [11]. Also, "irrelevant tweets [were found to] die out", and "Twitter users retweeted to show support for their beliefs in others commentaries" about the London riots [11]. Twitter was proven useful as a medium in "spreading word about subsequent events" [11], with the prevalent use of `#hashtags` to collate messages of a similar subtheme pertaining to the riots (e.g. `#riotcleanup` to discuss post-riot cleanup efforts).

IV. Data Collection and Sample

To accomplish **Goal 1**, the Twitter Streaming API was used to collect tweets on the London Riots and their user/message metadata. Using a Perl script, we invoked the `filter` interface to the Streaming API[1] to capture tweets with the hashtag `#londonriots` (similar to [11]), inclusive of user and message metadata. Our simple script contained a feedback loop which automatically attempted to reconnect to Twitter in the event of errors or loss of connectivity.

As our data capture began chronologically in the middle of the riots, we were able to capture Twitter chatter at the tail end of the riots, including its immediate aftermath. The properties of our captured data are listed in Table I.

TABLE I
Summary of Captured London Riots Data.

Property	Statistics
Observation period	Start: Tue Aug 09 2011, 05:03:32 (UTC) End: Mon Aug 15 2011, 02:46:40 (UTC)
Number of records	503,865 messages 254,690 unique users (~307MB uncompressed data)
Chronological message distribution	Aug 09 2011: 262093 Aug 10 2011: 113188 Aug 11 2011: 64589 Aug 12 2011: 38997 Aug 13 2011: 15550 Aug 14 2011: 9061 Aug 15 2011: 387

[1] `http://apiwiki.twitter.com`

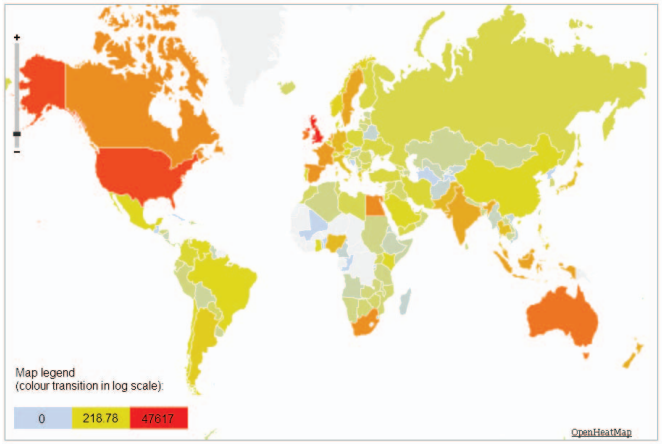

Fig. 1. Geographic heat-map of the authors contributing to tweets in the London Riots dataset. The color intensities represent the number of messages per country.. Color values are interpolated logarithmically from the seed colors in the map legend (bottom-left). Generated with *OpenHeatMap* [12].

V. Metadata-based Inferences on Demographic Attributes

A. Gender

The presence of real names on Twitter user profiles allow us to potentially infer a user's gender [5], [6], which is an interesting demographic property to study in relation to the London Riots [1]. We apply a frequency-based gender inference algorithm [6], which uses 130 years' worth of US Social Security Administration first name data to classify gender.

Out of 254,690 unique user records from our dataset, there were 1,654 records with blank first names which were omitted from our analysis. Table II details the results of the *GenderFromName* algorithm when applied to the remaining 253,036 names.

TABLE II
Distribution of genders found from first names in the London Riots dataset.

Gender	Count	Percentage
Male ♂	112,052	44.28%
Female ♀	80,417	31.78%
Unassigned	60,567	23.94%
Total	**253,036**	**100%**

The proportion of males found in the London Riots dataset exceed those of females by approximately 10 percent. This agrees with observations in news media: more males participate in the riots compared to females [1].

B. Geographic Location

The motivation behind studying geographic location is taken from existing studies on crisis/convergence events [9], [10] where geographic information found in Twitter user metadata formed the basis of studying how events spread spatially in the real-world.

For the London Riots data, we apply a two-phase geo-location approach [6] on the user and message meta-data to determine country information for the authors of `#londonriots` messages. We found a total of 82,049 tweets, containing some form of location information embedded within their metadata (either latitude/longitude pairs, or free-form location strings). With this information, we constructed a *heatmap* – i.e. coloring country polygons based on frequency of tweets within a country – to visualize the geographic distribution of tweets pertaining to the London Riots.

This heatmap, generated with *OpenHeatMap* [12], is illustrated in Figure 1. From the heatmap, we observe that the majority of messages in our dataset of the London Riots originate from the United Kingdom – 47,617 tweets (a majority of 58.03 %) – as we would naturally expect. A secondary proportion of messages originate from developing countries or countries with close relations to the United Kingdom, such as the United States (9.17%), Australia (3.10%), and the republic of Ireland (2.99%).

From the three observations conducted in this section (*Section V A–B*), we have accomplished **Goal 2** by transforming basic Twitter metadata into real-world properties of gender and geographic distribution, for characterization of Twitter observers and participants of the London Riots.

VI. Correlation between Riot Activity with Tweet Locations

To visualize the Twitter activity close to the heart of the riots, we draw upon the work by Rogers *et al.* from *The Guardian* [13], who pinpointed every verified riot-affected location on a Google Maps mash-up. Given the availability of accurate latitude/longitude pairs found in our dataset, we used a similar Google Maps mash-up (Figure 3) to accurately pinpoint clusters of Twitter activity related to the riots using a Google Maps mash-up.

Metadata in the London Riots Twitter dataset are parsed to find accurate geographic coordinates within the United Kingdom. The coordinate bounding box has the longitude range of $[-8.1647°, 1.7245°]$ and latitude range of $[49.9553°, 60.6311°]$. Using 6,720 locations found in metadata, we constructed the Google Maps mash-up (Figure 3); with each yellow dot representing a single tweet.

By comparing Figures 2 and 3, we observe the following:

- Generally, locations of Twitter chatter found in Figure 3 are close to the actual outbreaks of the riots as documented in Figure 2 [13].
- Riot events – and correspondingly tweets – are concentrated around the most-affected areas: London, Birmingham, Bristol, Cardiff, Liverpool, Manchester, and Leeds.
- Although confirmed reports of riots are absent in major cities such as Newcastle-upon-Tyne, Southampton, and Dublin, chatter about the riots are mildly concentrated amongst these areas.

To statistically test for any correlation between the presence of `#londonriots` tweets and documented riot outbreaks

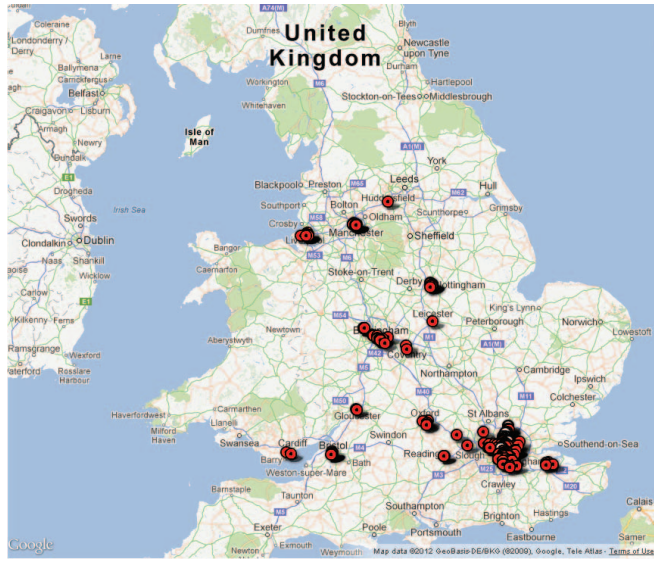

Fig. 2. Comprehensive map from The Guardian chronicling "what has happened where as rioting spreads across England" [13]. Each red dot indicates a reported case of rioting activity.

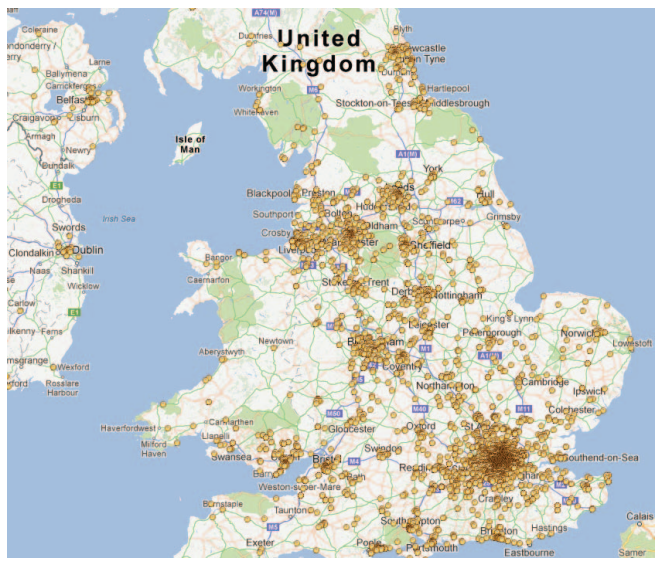

Fig. 3. Visualization of locations found in Twitter metadata from our London Riots dataset, originating from the United Kingdom. Each yellow dot represents a single tweet.

[13], we first partition the map into squares of unit $1° \times 1°$. We considered only map squares containing parts of England and Wales; we removed squares that are fully located in bodies of water, as well as those that fall completely in Ireland or Scotland.

For each of the valid map squares, we count both the number of tweets and the number of reported riot outbreaks which fall within. We then calculate the Pearson product-moment correlation coefficient, r [14] between the two samples: tweet count (per map square) versus real-life riot events (per map square). We used Students t-test [15] to check for statistical

significance, i.e. to refute the null hypothesis that *there is zero correlation between the number of tweets per map square and the corresponding number of reported riot outbreaks*.

From the partitioning into 37 map squares of unit $1° \times 1°$ ($\sim$12100km^2), we obtain $r = 0.970$; which is statistically significant at the 1% level (with a Student's t-value of $23.77 > 2.44$)[2].

We repeated this experiment, this time partitioning the map into squares of unit $0.5° \times 0.5°$ ($\sim$3025km^2), resulting in 114 such squares. This time, we obtain $r = 0.937$, which is again statistically significant at the 1% level (with a Student's t-value of $28.40 > 2.36$)[3]. From the two versions of this experiment, there is strong enough evidence to refute the null hypothesis.

To conclude this section, we have accomplished **Goal 3** by identifying a possible correlation between the number of tweets in a given area and the frequency of nearby riot events (at the 1% significance level). These results confirm that Twitter metadata can be an accurate source of location information, useful in accurately pinpointing locations of real-world events [7], [9], [10].

VII. Online Presence and Messaging Behavior of London Riots Tweet Authors

A. Users' Device Class, Mobility, and Spam

The identification of `source` strings (i.e. strings identifying the software used in composing a tweet) have been performed in existing Twitter research, e.g. [5], [6]. *Device classes* are groupings of similar `source` strings based on the platform a particular client software runs on.

Figure 4 illustrates the distribution of the different device classes found in the metadata of the London Riots dataset. The noteworthy properties of this distribution were:

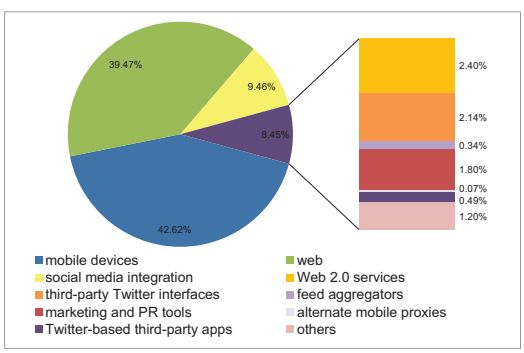

Fig. 4. Distribution of device classes inferred from Twitter client source strings, in the London Riots dataset. Observe that almost half of the users compose London Riot-related tweets from a mobile device.

- The usage of *mobile* clients constitute the majority, indicating a tendency by the participants contributing to tweets on the riots to participate while mobile or 'on the move'. This, to a certain extent, matches the observations from news reports [1], [2] which pinpoint mobile technology as a catalyst for riot participation.
- *Social media* clients and the *web* interface together contribute to half of the total participation during the riots; manifested in the form of Twitter comments or discussions by users not actively involved in the riots (e.g. Londoners at home, or overseas commenters).
- The proportion of other non-prevalent device classes – such as *bots* and *suspicious/rogue applications* – are virtually absent from our sample (categorized as *Other* in Figure 3). However, this does not exclude the possibility of spammers capitalizing on the `#londonriots` hashtag to publish spam tweets.
- In the sample, there is one hitherto unreported software client – *Donate Your Account* – which is the 28th most commonly seen `source` string in the dataset (amounting to 730 tweets). Upon investigation, this is a website which lets you 'lend' your Twitter account to a campaign, allowing it to broadcast tweets on your behalf. In the case of the London Riots, the accounts are 'borrowed' by user `@CitizenRadio`, a political podcast, to broadcast `#londonriots` tweets.

[2]A minimum $t = 2.43$ is needed to show 1% statistical significance, with $37 - 2 = 35$ degrees of freedom.

[3]A minimum $t = 2.36$ is needed to show 1% statistical significance, with $114 - 2 = 112$ degrees of freedom.

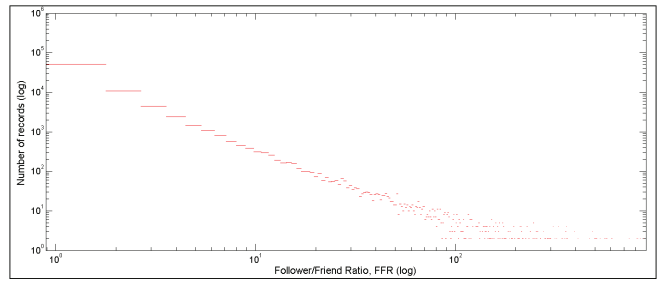

Fig. 5. Distribution of follower/friend ratio (FFR) of unique users in the London Riots dataset *sans* outliers (with a disproportionately high FFR).

B. Users' Friend/Follower Ratio and User Influence

With the availability of user metadata in our London Riots dataset, we are also able to obtain statistics of tweets' authors. Extant research on Twitter in the London Riots focus mostly on the messages rather than users [11].

We will now analyze the distribution of the follower/friend ratio (FFR) of users in the London Riots dataset. In short, the FFR for each unique user is simply the ratio of Twitter *followers* to *friends*. A plot of all user FFRs is generated (Figure 5), *sans* outliers that constitute only $\sim$0.04% of the dataset (for clarity).

The FFR distribution of users found in the London Riots dataset follows a *power law*, exhibiting characteristics of a scale-free network [16]. From our interpretation, such users range from users with a balanced ratio of followers-to-friends; to high-profile Twitter users commenting or breaking news of the riots (with a disproportionate ratio of followers-to-friends).

Studying the long tail of the FFR distribution, the outliers

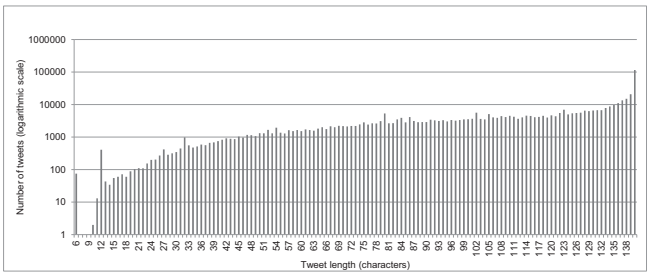

Fig. 6. Distribution of tweet length in the London Riots dataset.

with disproportionately large FFRs are high-profile Twitter users, in one of several categories: *celebrities* expressing their views (e.g. Cheryl Cole); *news sources and media organizations* (e.g. Al-Jazeera/BBC); *academics/writers* (e.g. Neil Gaiman); *politicians/pundits* (e.g. `@ConservativeHome`); *satirical* Twitter accounts (e.g. `@Prince___Harry`); *automated bots* (e.g. `@toptweets`); and *Twitter users created in response to the riots* who rapidly achieved a high FFR (e.g. `@Riotcleanup`, a private initiative to promote cleaning up after the riots).

C. Message Statistics

Interesting summary statistics can be obtained from the 503,865 tweets (and message metadata). One such statistic, useful in the context of the London Riots, is the *message length*, as a summary of the amount of information conveyed in a tweet. The frequency distribution of message lengths in our dataset is shown in Figure 6.

From this distribution, we can make several observations on the messaging behavior during the London Riots:

- There is a spike in the graph (Figure 6) for length 12, i.e. these messages contain nothing useful but a hashtag string `#londonriots` (amounting to 12 characters). We deduce that the only *raison d'être* of such messages is for a user to 'contribute' to the overall Twitter chatter on the riots without adding any useful information. Another possibility is simply for a rioter to indicate his presence (and availability) with the mere presence of a hashtag.
- There is a spike of tweets (an increase of an order of magnitude) at the 140-character boundary, caused by the truncation of long messages.
- Compared to the distribution of everyday tweets [17], the histogram representation of the London Riots dataset does not contain a bimodal distribution. Such a bimodal distribution is characteristic of everyday tweets containing a dichotomy of both short and long messages, but absent in `#londonriots` tweets. The reason is that more information – e.g. URLs and personal communication – are conveyed in a single tweet about the London Riots which are "mass convergence" events [8], [9].

For the sake of completeness, we performed a cursory examination of common entities found in London Riots tweets; complementing the results in [11]. The most commonly occurring Twitter accounts mentioned *within* tweets (in the form of `@user`) are from major news outlets (e.g. BBC and ITV), and campaigns to promote recovery (e.g. `@riotcleanup` as described in *Section VII B*). As for `#hashtags`, a variety of synonyms were used to categorize the London Riots tweets; our dataset corroborates the findings in [11]. Place name `#hashtags` (such as `#liverpool`) also occur frequently in tweets, similar to tweets found in other crisis events such as earthquakes [10].

This section, in summary, has accomplished research **Goal 4** by firstly characterizing Twitter online presence during the London Riots via device classes, and consequently pinpointing users who are likely to be rioters. Next, an analysis of FFRs have illustrated that the users in our London Riots dataset follow a power-law distribution, and users with abnormally high FFRs tend to be high-profile commenters. Finally, by investigating message length and frequently-occurring entities, we are able to study the messaging intent of users during the riots.

VIII. Clustering to Detect Patterns in London Riots Tweets

To answer the last research goal (**Goal 5**), we performed pattern recognition (using unsupervised clustering) to reveal hidden commonalities among the collection of users and messages in our dataset. In the same vein as [5], [7], [18], we use *Viscovery SOMine* to perform unsupervised self-organizing map (SOM) [19] clustering, and merging the final result clusters with Ward [20] clustering. In brief, SOM works by projecting input from higher dimensions onto maps of two-dimensions, where similar features are spatially located close by. Ward's bottom-up hierarchical clustering method [20] is used to cluster the output maps into regions.

The input data to *Viscovery SOMine* consisted of all 503,865 records within the London Riots dataset. The set of input features for each record consists of 12 features, including (but not limited to) the ones described in *Sections V* and *VII*. Briefly, the feature space includes: *gender*, *country*, *device class*, *profile customization* [5], *FFR*, *user activity ratio* [6], *total user posts* [6], *message length*, and presence of special notations (`@users`, `#hashtags`, `RTs`, and URLs). Clustering the 503,865 records in the London Riots dataset using *Viscovery SOMine*'s SOM-Ward implementation resulted in a total of three clusters (Figure 7).

Cluster I (light-blue) constitutes the majority of the records (66.09%). From the features found in the majority of records in this cluster, we deduce that this cluster consists of 'random chatter' regarding the London Riots that originate from a wide variety of origins. The metadata provided on Twitter for these records however cannot be reveal much about the users behind these tweets, possibly due to anonymization (e.g. invalid locations and non-human names).

Cluster II (pink) is the second largest, containing 33.75% of total records, contains almost all tweets from the United Kingdom (`GB`) as well as tweets with no apparent location

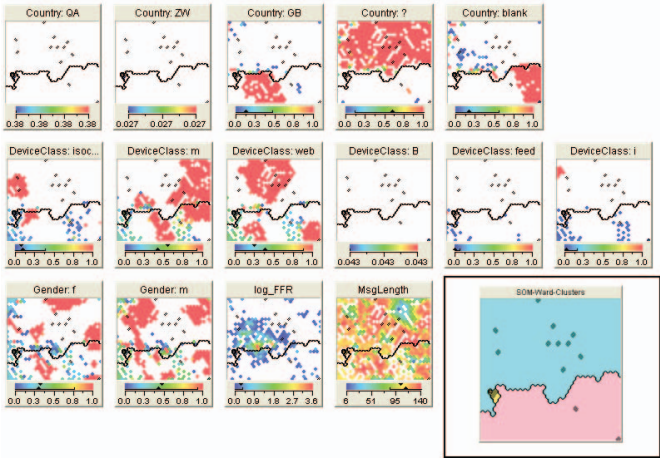

Fig. 7. Maps resulting from SOM-Ward clustering of metadata. The overall clustering is in the inset (*bottom-right*). Maps of interesting features include: *First row from left*: Country found in user metadata [in order: *Qatar* (Middle-East), *Zimbabwe* (developing nation) *United Kingdom* (riot-affected)], unknown country, no location metadata found.
Second row from left: Device class inferred [in order: *social media integration*, *mobile*, *web*, *bots*, *feed aggregators*, *Web 2.0 services*].
Third row from left: Gender [in order: *female*, *male*]; FFR; message length.

information. Narrowing down on the surface of the map where the British tweets are concentrated, there is a visibly significant proportion of male users, agreeing with existing news coverage [1]. In the same map region as the British tweets, the distribution of message lengths (not shown) tend toward the 140-character limit, suggesting a high information content per tweet. Device classes found in the same map region tend to be either web or mobile devices. There is also a prevalence of mobile devices in this cluster. We surmise that this cluster contains the users who are actively involved in the riot, using Twitter primarily via mobile phones (earlier documented as catalysts for participation in the riots [1], [2]).

Cluster III (yellow) is the smallest, comprising merely 0.16% of the overall input data size. However, this cluster, which appears to be anomalous with respect to the rest of the clusters found, exhibits a quaint property in terms of the origin of tweets. Tweets contained within this cluster entirely originate from developing nations and the Middle-East. This might suggest that users from these countries share a common concern between them about these riots, which is exhibited in their common feature space in terms of metadata.

In essence, we have accomplished **Goal 5** in this section by identifying broad clusters of users who contributed to `#londonriots` tweets in our dataset.

IX. Conclusion and Further Work

This study clearly shows that analysis of social metadata can yield useful insights about major social events. Firstly, there are more males than females in London Riots' tweet authorship. This is in contrast to the pattern of female-majority in general [6]. A huge statistical correlation was found between the tweet origins and the real-world riot locations. The proportion of tweets from mobile devices was the highest, suggesting a possibility of their catalyzing riots. On the other hand, from analyzing FFRs and message summaries from high-profile Twitter users, we detect another segment of tweets that are focused on commentary and recovery initiatives. By clustering the user and message properties using Kohonen's SOM, we obtained three clusters. Each one of them exhibits unique spatial and behavioral characteristics. Further experimental investigations are required to draw decisive conclusions on behavioral patterns, with emphasis on their clustering.

References

[1] J. May, "Burning issues," *The Age*, vol. August 13, pp. 15–17, 2011.
[2] J. Meikle and S. Jones, "UK riots: More than 1,000 arrests strain legal system to the limit," *The Guardian*, vol. August 10, 2011.
[3] A. Lock, "Insurers say London riot losses 'well over £100m'," *City A.M.*, vol. August 9, 2011.
[4] W. L. Adams, "Were Twitter or BlackBerrys Used to Fan Flames of London's Riots?" *Time Magazine*, vol. Aug 28, 2011.
[5] M. Cheong and V. Lee, "Twitmographics: Learning the emergent properties of the Twitter community," in *From Sociology to Computing in Social Networks: Theory, Foundations and Applications*, ser. Lecture Notes in Social Networks. Springer-Verlag, 2010, vol. 1, pp. 323–342.
[6] M. Cheong, S. Ray, and D. Green, "Large-scale Socio-demographic Pattern Discovery on Microblog Metadata," in *Proc. SoCPaR 2012*, 2012.
[7] M. Cheong and V. Lee, "A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter," *Information Systems Frontiers*, vol. 13, no. 1, pp. 45–59, 2011.
[8] A. Hughes and L. Palen, "Twitter adoption and use in mass convergence and emergency events," in *Proc. ISCRAM 2009*, 2009.
[9] K. Starbird, L. Palen, A. Hughes, and S. Vieweg, "Chatter on The Red: What hazards threat reveals about the social life of microblogged information," in *Proc. CSCW 2010*, 2010.
[10] B. D. Longueville, R. S. Smith, and G. Luraschi, "'OMG, from here, I can see the flames!': a use case of mining Location Based Social Networks to acquire spatiotemporal data on forest fires," in *Proc. LBSN '09*, 2009, pp. 73–80.
[11] E. Tonkin and G. Tourte, "Twitter, information sharing and the London riots?" *Bulletin of the American Society for Information Science and Technology*, vol. 38, no. 2, pp. 49–57, 2012.
[12] P. Warden, "OpenHeatMap," Available from http://www.openheatmap.com/, 2012.
[13] S. Rogers, A. Sedghi, and L. Evans, "UK riots: every verified incident - interactive map," [Online]. Available: http://www.guardian.co.uk/news/datablog/interactive/2011/aug/09/uk-riots-incident-map, August 11 2011.
[14] K. Pearson, "Contributions to the mathematical theory of evolution. iii. regression, heredity, and panmixia." *Proceedings of the Royal Society of London*, vol. 59, no. 353-358, pp. 69–71, 1895.
[15] Student, "The probable error of a mean," *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908.
[16] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
[17] M. Yoshida, T. Inui, and M. Yamamoto, "Analysis of Tweets including URLs on Twitter (translated)," in *Proc. DEIM Forum 2010*, 2010.
[18] Z. Yao, T. Eklund, and B. Back, "Using SOM-Ward clustering and predictive analytics for conducting customer segmentation," in *Proc. 2010 IEEE International Conference on Data Mining Workshops*, 2010.
[19] T. Kohonen, *Self-Organization and Associative Memory*. Berlin: Springer, 1984.
[20] J. H. Ward, Jr., "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.