Temporal Analysis of Activity Patterns of Editors in Collaborative Mapping Project of OpenStreetMap

Taha Yasseri
Oxford Internet Institute
University of Oxford
Oxford, UK
taha.yasseri@oii.ox.ac.uk

Giovanni Quattrone
Dept. of Computer Science
University College London
London WC1E 6BT, UK
g.quattrone@cs.ucl.ac.uk

Afra Mashhadi
Bell Labs, Alcatel-Lucent
Dublin, Ireland
afra.mashhadi@alcatel-lucent.com

ABSTRACT

In the recent years Wikis have become an attractive platform for social studies of the human behaviour. Containing millions records of edits across the globe, collaborative systems such as Wikipedia have allowed researchers to gain a better understanding of editors participation and their activity patterns. However, contributions made to Geo-wikis —wiki-based collaborative mapping projects—differ from systems such as Wikipedia in a fundamental way due to spatial dimension of the content that limits the contributors to a set of those who posses local knowledge about a specific area and therefore cross-platform studies and comparisons are required to build a comprehensive image of online open collaboration phenomena. In this work, we study the temporal behavioural pattern of OpenStreetMap editors, a successful example of geo-wiki, for two European capital cities. We categorise different type of temporal patterns and report on the historical trend within a period of 7 years of the project age. We also draw a comparison with the previously observed editing activity patterns of Wikipedia.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications – Spatial Databases and GIS; H.5.3 [**Group and Organization Interfaces**]: Collaborative computing, computer-supported cooperative work

General Terms

Human Factors, Measurement

Keywords

OpenStreetMap, Geo-wiki, Mass Collaboration, Wikipedia, Eigenbehaviour, Principal Component Analysis, Circadian Pattern

1. INTRODUCTION

In the recent years, the research on wikis and open, online collaboration systems has attracted a great deal of atten-

tion from social scientist, aiming to analyse behaviour of a large number of individuals and the interaction between them in a collaborative environment in details. The outcome of such research not only could provide a deeper understanding on human collaboration and its intrinsic features in general, but also could have great applications in innovating and improving new and existing collaboration platforms. A common finding has shown that in many of collaborative systems such as Wikipedia, the distribution of editors and contributions is highly skewed, with the majority of contributions coming from only about 2.5% of contributors [15]. Wikipedia as a predominant example of mass collaboration has been studied vastly and from multiple aspects, e.g., size and growth rate [2], conflict and editorial wars [21, 27], user reputation and article quality [23, 24], linguistic features and readability [25].

However, despite the great popularity that geo-wiki systems received in the last years, they have seen far less research attention, with the exception of recent studies on accuracy and coverage of OpenStreetMap[1] (OSM) where the accuracy is shown to be high [7, 14] and coverage to be non-uniformly distributed [28, 13]. However, users behaviour in this domain is by far less investigated subject and thus requires more research effort to be well understood. In this regard, Hristova et al [10] has taken the first steps to quantitatively understand users participation in geo-wikis. The authors show that in OSM the distribution of editors and contributions are as skewed as Wikipedia, with 95% of contributions made in the area of London, UK, attributed to less than 10% of users. Also Panciera et al [16] found that in Cyclopath[2] 5% of its users are responsible for the majority of its content. These results are even more interesting when one considers that geo-wikis systems are fundamentally different from classical web-based wikis as their spatial dimension limits the contributors to a set of those who posses local knowledge about a specific area (i.e., have physically visited the place). Therefore, not all the properties investigated on classical web-based wikis could be generalized to geo-wikis too. Furthermore, the recent popularity and widespread adoption of smart-phones has brought forward classical wiki systems as a candidate worth considering in the urban domain too, with citizens becoming cartographers, with geo-wikis such as OSM. Indeed, these services offer mobile applications that users can deploy directly on their smart-phones to generate geo-tagged content *on the go*. Given these characteristics,

---

[1] http://www.openstreetmap.org/
[2] http:// cyclopath.org/

the question we then ask ourselves is: *what are the temporal activity patterns of geo-wiki editors and how they differ from those of previously highlighted in Wikipedia?* Answers to this question not only illustrate interesting patterns in human dynamics, but also show how the rise in use of mobile devices has changed the work-flow of collaborative mapping projects. The effects and potentials of transformation towards "mobile crowdsourcing" could be well examined by studying characteristics of such platforms in details, with great implications in other areas related to ubiquitous and social computing. In this direction, this paper takes the first steps in identifying the *temporal* behavioural pattern of geo-wiki editors. In particular, it focuses on OSM, a very successful example of geo-wikis which boasts over one million users collectively building a free, openly accessible, editable map of the world.

Temporal patterns of human behaviour has been studied extensively in various fields. In the domain of web and networking, the circadian patterns of the Internet traffic have brought interesting information about individual habits of the Internet usage in different societies, assisting with utilisation of infrastructure [20, 1, 19]. In domain of social networking and collaborative systems, temporal features of online activities and among them circadian patterns have attracted great interest [27, 6, 22, 3]. Circadian patterns could explain many different observed phenomena in the cyberspace. One example is the anti-social attitude of heavy users of online applications, specifically depression and other mental health symptoms among power users of peer-production platforms like Wikipedia, as these attitudes are reported to have especial temporal fingerprints [5]. Eagle et al. [4] have studied the temporal behavioural pattern of human by monitoring communication, location and interactions of 100 test subjects. They extracted common "eigenbehaviour" structures from their data by using principle component analysis. In this paper we adopt similar methodology and analyse the activity logs of OSM editors for two European capital cities of London and Rome for a period of 7 and 5 years respectively. We follow the principal component analysis to identify the repeating temporal structures of editors in our data, where typical structures are represented by eigenpatterns. We study these behavioural structures over different stages of OSM. Our results show that despite the differences between the nature of contributions of geo-wikis and wikis, the OSM editors generally exhibit similar temporal behavioural pattern of contributions during the day as Wikipedians. However, we show that over the past years the editors behaviour has shifted towards afternoon (3 p.m. to 6 p.m.) activities from the late night activities perhaps corresponding to the success of OSM as a smart-phone application. This latter has a slower dynamics in the case of Rome compared to London.

2. MATERIALS AND METHODS

Data. The OSM dataset is freely available to download and contains the history of all edits (since 2005) on all spatial objects performed by all users. Spatial objects can be one of three types: *nodes*, *ways*, and *relations*. Nodes broadly refer to POIs, ways are representative of roads, and relations are used for grouping other objects together. For the purpose of this study, we selected two cities to analyse: London, UK, as an example of a very well represented large metropolitan city in OSM, and Rome, Italy, as an example of a large city which is steadily increasing its spatial representation in OSM. Furthermore, we focused only on nodes; in fact, we expect that editing pattern of nodes differ from those found in classical wiki systems due to simplicity of adding/editing a node on the fly in a more ubiquitous fashion. Finally, to consider only genuine users' contributions, we have looked into contributions that most likely correspond to bulk imports. Many bulk imports were detected in the whole dataset, with tens of thousands of edits done in a single day by a single user, spread throughout Greater London (e.g., more than 20,000 post boxes spread across all Greater London appeared in OSM in only one day in 2009 from the same user). We chose to discard such data for the reason that we intend to model genuine 'bottom-up', user-generated contributions, of which massive imports are not representative. Finally, we assume that contributions are made locally that is they come from users from that country who share the same timezone. Although this assumption is debatable as the Internet globalisation has created a worldwide user base with little guarantee of 'localness', studies such as [8, 12, 9] have shown that the likelihood of contributions to Wikipedia exponentially decreases as the geographical distance between the article and user locations increases. The datasets we are left with are summarised below:

City	#Users	#Edits	Time Window
London	607	534361	7 years
Rome	563	146000	5 years

Method. To capture the editing patterns of editors in OSM, we compute two different measures:

● *Circadian cycles.* We analyzed the periodic characteristics of temporal patterns of editors for each hour of 24 hours day by considering the ratio of the average number of edits that has been done within the considered time window on the average number of edits that has been done within all the 24 hours of the day, as specified below:

$$circadian_{\Delta t} = \frac{avg(\#Edits_{\Delta t})}{avg(\#Edits)}$$

where, $avg(\#Edits_{\Delta t})$ is average amount of edits occurred during the specific time interval Δt of the day and $avg(\#Edits)$ is the average number of edits occurred during the whole day.

● *Eigenbehaviours.* We used the principle components analysis (PCA) of our data focusing on *time of the day*. The principle component analysis has been extensively used in the past to identify the behavioural patterns of users mobility [4, 17]. Similarly, we adopt the same methodology in which we define the principal components as a set of vectors that span a 'behaviour space' and characterise the behavioural variation between users. More precisely, we create a matrix M where each row corresponds to a unique user in our dataset and each column corresponds to the time unit under study. Each entry m_{ij} of our matrix M contains the average ratio of edits that the user u_i has made during the time units t_j. Exactly as before, as time unit we considered one hour time interval. For each column of this matrix, we compute the empirical mean and the deviation of each column from this mean, after this we computed the covariance matrix B of the previous matrix. Finally, we extract the

repeating behaviour from this data structure by calculating the eigenvectors and eigenvalues of B. These extracted eigenbehaviours are the heavily weighted vectors that represent a type of repeated behaviour, such as editing during evening, night or afternoon. In order to get a robust set of eigenvectors, we needed to filter out editors with few edits leading to abnormal circadian patterns with picks of larger than 20% of all edits within a one hour window.

3. RESULTS

We first start by analysing the periodic characteristics of temporal patterns of editors inferring the circadian cycles of OSM edits. We considered every single edit performed since 2005/2007 on London/Rome and used the timestamps assigned to each edit to calculated the overall activity of users for the time of day. Figure 1 illustrates the circadian pattern. It is interesting to observe that the editing activities are to their minimum at early morning around 4 a.m., followed by a rapid increase up to noon. The activity shows a slight increase until around 9 p.m., where it start to decrease during night. This result corresponds closely to the circadian patterns previously observed in Wikipedia [26] as well as qualitatively corresponding to the other kind of human activities such as mobile calling behaviour [11], and the Internet instant messaging [18]. However, Rome editors show more activity around noon and less in the afternoon compared to London editors.
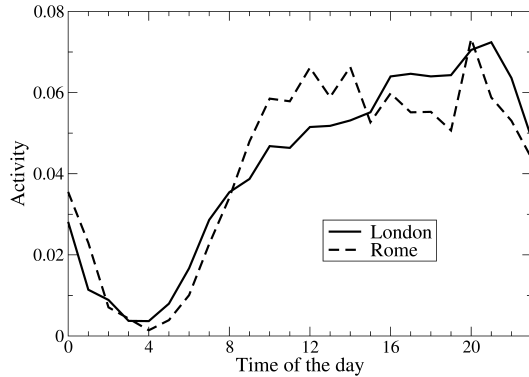

Figure 1: The circadian pattern of edits based on the aggregated data for the whole dataset period.

We further investigate the circadian activity patterns by performing PCA. The first 3 principal components V_1, V_2, V_3 extracted from the individual activity pattern of London editors are shown in Figure 2. The sum of the 3 corresponding eigenvalues exceeds 80% of the sum of the all spectrum, justifying the reduction of the space dimension from 24 to 3. It is possible to see that these 3 component roughly correspond to 3 type of editorial pattern: *(i)* V_1 corresponds to afternoon edits where the peak is from 2 p.m to 6 p.m, *(ii)* V_2 corresponds to morning and noon edits and finally where the peak of edits falls from 7 a.m to 2 p.m. *(iii)* V_3 corresponds to night edits where the peak of edits fall from 6 p.m. to midnight. Note that the eigenvectors indicate the direction of the most typical deviations from the average behaviour. We repeated the same analysis for the Rome data which resulted in a very similar pattern. However as the smaller size of the Rome dataset contributes to the level of noise dramatically, we proceed with our analysis based on the eigenvectors extracted from London data only.
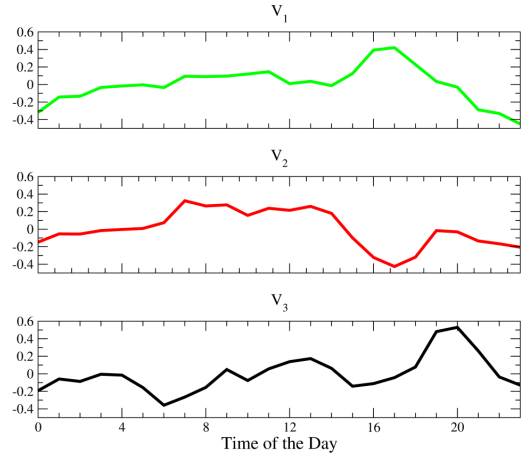

Figure 2: Three Top Eigenbehaviours constructed from individual activity patterns of London OSM editors.

In the next step we build up activity vectors for all edits within each year for both London and Rome and project them on the identified eigen-behaviours V_1, V_2, V_3. The projected values c_1, c_2, c_3 show to which extent the activity vector could be reproduced by the corresponding eigen-behaviour. c_i's calculated for London and Rome for different years are shown in Figure 3. In both cases a shift towards afternoon edit is evident; however faster in the case of London. Furthermore, we observe a subside effect in the edits associated to night time activities. We speculate two reasons for these trends. Firstly the widespread of Internet services on devices such as smart-phones in the recent years, may have had an impact on the temporal behaviour of users as they become less limited to stationary computers (e.g., at work or home). It is interesting to note that the lesser extend of afternoon projection for Rome can similarly be explained by a recent study[3] which has highlighted a much lower adoption rate of smart-phones in Italy (about 11%) in comparison to the UK (70%) market. Secondly, as OSM system ages, the saturation effect on maps may mean that less mass contributions from power-users is required, thus a decrease in night activities (i.e., a trend usually associated with power-users).

4. CONCLUSION

In this work we studied the temporal behavioural pattern of OSM editors as a representative example of geo-wiki systems. By performing Principal Component Analysis, we extracted the typical editorial behaviour temporal patterns. We showed that although OSM editors generally exhibit similar temporal behavioural pattern of contributions during the day as Wikipedians, but over the past years the editors behaviour has shifted towards afternoon activities from the late night activities perhaps corresponding to the success of OSM as a smart-phone application. However, this trend has a slower dynamics in the case of Rome compared to London. We believe these results not only shed light on the underlying mechanisms of collaborative mapping projects, but also could suggest elaborative strategies in order to increase the viability and quality of the project; facilitating more mobile contributions being one example. Note that higher quality of London map compared to Rome has been already reported [14], in accordance with our observation of
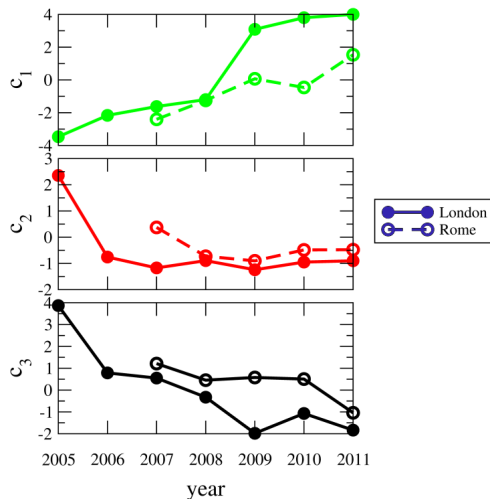
---

[3]http://tinyurl.com/ ylj9bvb

Figure 3: The trend of projections into the identified eigenbehaviours (shown in Figure 2) over years for London and Rome editors.

larger use of mobile devices by its users.

As future direction, we plan to go beyond this study by considering in our temporal analysis not only the daily pattern of user edits, but also the difference on daily pattern during working days and weekends for different user categories (e.g., power-users, early adopters). Furthermore, to gain a deeper understanding of geo-wikis editors behaviour, we plan to divide our data into roads versus POIs, hypothesising that editing POIs requires less skills and attention than roads, for this reason they may exhibit a different temporal pattern.

5. REFERENCES

[1] M. Ahmed, S. Traverso, P. Giaccone, E. Leonardi, and S. Niccolini. Analyzing the performance of lru caches under non-stationary traffic patterns. *CoRR*, 1301.4909, 2013.

[2] R. B. Almeida, B. Mozafari, and J. Cho. On the evolution of wikipedia. In *Proceedings of the International Conference on Weblogs and Social Media*, ICWSM'07, 2007.

[3] T. C. Daniel Preo tiuc Pietro. Mining user behaviours: A study of check-in patterns in location based social. In *International AAAI Conference of WebSci'13,*, 2013.

[4] N. Eagle and A. S. Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.

[5] S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.

[6] S. A. Golder, D. M. Wilkinson, and B. A. Huberman. Rhythms of social interaction: Messaging within a massive online network. In *Communities and Technologies 2007*, pages 41–66. Springer, 2007.

[7] M. Haklay, S. Basiouka, V. Antoniou, and A. Ather. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus Law to Volunteered Geographic Information. *Cartographic Journal, The*, 47(4):315–322, 2010.

[8] D. Hardy, J. Frew, and M. F. Goodchild. Volunteered geographic information production as a spatial process. *International Journal of Geographical Information Science*, 26(7):1191–1212, 2012.

[9] B. J. Hecht and D. Gergle. On the localness of user-generated content. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 229–232. ACM, 2010.

[10] D. Hristova, G. Quattrone, A. Mashhadi, and L. Capra. The life of the party: Impact of social mapping on openstreetmap. In *Proceedings of the AAAI International Conference on Weblogs and Social Media(ICWSM2013)*. AAAI, 2013.

[11] H.-H. Jo, M. Karsai, J. Kertész, and K. Kaski. Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics*, 14(1):013055, 2012.

[12] M. D. Lieberman and J. Lin. You are where you edit: Locating wikipedia contributors through edit histories. In *3rd International AAAI Conference on Weblogs and Social Media*, 2009.

[13] A. Mashhadi, G. Quattrone, and Capra. Putting ubiquitous crowd-sourcing into context. In *Proceedings of the 8th International Conference on Computer Supported Work and Social Computing (CSCW 2013))*. ACM, 2013.

[14] A. Mashhadi, G. Quattrone, L. Capra, and P. Mooney. On the accuracy of urban crowd-sourcing for maintaining large-scale geospatial databases. In *Proceedings of the 8th International Symposium on Wikis and Open Collaboration (WikiSym'12)*. ACM, 2012.

[15] K. Panciera, A. Halfaker, and L. Terveen. Wikipedians are born, not made. In *Proceedings of the ACM 2009 international conference on Supporting group work*, pages 51–60. ACM, 2009.

[16] K. Panciera, R. Priedhorsky, T. Erickson, and L. Terveen. Lurking? Cyclopaths? A quantitative lifestyle analysis of usr behaviour in a geowiki. In *Proceedings of CHI 2010: 28th ACM Conference on Human Factors in Computing Systems*, pages 1917–1926, 2010.

[17] J. Park, D.-S. Lee, and M. C. González. The eigenmode analysis of human motion. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(11):P11021, 2010.

[18] A. Pozdnoukhov and F. Walsh. Exploratory novelty identification in human activity data streams. In *Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming*, pages 59–62. ACM, 2010.

[19] D. H. Spennemann. The internet and daily life in australia: an exploration. *The Information Society*, 22(2):101–110, 2006.

[20] D. H. Spennemann, J. Atkinson, and D. Cornforth. Sessional, weekly and diurnal patterns of computer lab usage by students attending a regional university in australia. *Computers & Education*, 49(3):726–739, 2007.

[21] R. Sumi, T. Yasseri, A. Rung, A. Kornai, and J. Kertész. Edit wars in Wikipedia. In *2011 IEEE Third International Conference on Social Computing (SocialCom)*, pages 724–727, 2011.

[22] M. Thamm and A. Bleier. When politicians tweet. In *International AAAI Conference of WebSci'13,*, 2013.

[23] D. M. Wilkinson and B. A. Huberman. Assessing the value of cooperation in Wikipedia. *First Monday*, 12(4), 2007.

[24] G. Wu, M. Harrigan, and P. Cunningham. Characterizing Wikipedia pages using edit network motif profiles. In *Proceedings of the 3rd international workshop on Search and mining user-generated contents*, SMUC '11, pages 45–52, New York, NY, USA, 2011. ACM.

[25] T. Yasseri, A. Kornai, and J. Kertész. A practical approach to language complexity: a Wikipedia case study. *PloS ONE*, 7(11):e48386, 2012.

[26] T. Yasseri, R. Sumi, and J. Kertész. Circadian patterns of wikipedia editorial activity: A demographic analysis. *PloS one*, 7(1):e30091, 2012.

[27] T. Yasseri, R. Sumi, A. Rung, A. Kornai, and J. Kertész. Dynamics of conflicts in Wikipedia. *PloS ONE*, 7(6):e38869, 2012.

[28] D. Zielstra and A. Zipf. A comparative study of proprietary geodata and volunteered geographic information for germany. In *Proceedings of the 13th International Conference on Geographic Information Science*, 2010.