Mining Mobile Youth Cultures

Tobias Blanke, Giles Greenway, Jennifer Pybus and Mark Cote
Department of Digital Humanities, King's College London, United Kingdom
Email: tobias.blanke@kcl.ac.uk, giles.greenway@kcl.ac.uk, jennifer.pybus@kcl.ac.uk, mark.cote@kcl.ac.uk

***Abstract*—In this short paper we discuss our work on co-research devices with a young coder community, which help investigate big social data collected by mobile phones. The development was accompanied by focus groups and interviews on privacy attitudes.**

I. Introduction

Our social and cultural world is transformed through the unprecedented growth in the data we generate about ourselves. We exist in a data-centric society characterised by our information-rich environment and a 'quantified self' [1]. In order to better understand these changes, we have engaged in a project to investigate 'Big Social Data (BSD)', and the role it plays in digital culture.

BSD denotes what we generate in our everyday lives through new mediated, cultural and communicative practices, like the 4 B pieces of content shared daily on Facebook, the 200 M tweets, etc. Currently, however, working knowledge and power over BSD is almost exclusively in the hands of an elite few: social media giants like Google and Facebook and state surveillance. The 'Our Data Ourselves' project's[1] aim is to investigate the consumption of personal data in social media applications and open up the data for research in digital culture. The project researches and experiments with a collaborative BSD research environment for arts and humanities scholars and will leverage this new form of knowledge and power into community-based cultural and economic assets.

In particular, we concentrate on developing research about smart phone environments as used by youth cultures. New kinds of communities around social media are created here, the impact of which we are only beginning to understand. Data on youth behaviour is increasingly 'born-digital'. The second focus of the project is the engagement with these mobile youth communities. We work with youth coders in the Young Rewired State (YRS) network [2], an organisation that helps to network, principally through holding hackdays. Together with YRS, we have developed 'hackathons' to engage young hackers as co-researchers in the project and generate together an understanding of the mobile data produced in mobile social media environments. This will facilitate a better understanding of the kinds of social connections, information sharing and normative relations that are developing to and through BSD.

The third major challenge we address is how BSD can be transformed into data for research by those who generate it and their communities. The basic issue raised is whether BSD can be transformed into a research asset and become a creative resource for cultural and economic community development. The project thus identifies BSD as a distinct subset of big data, one particularly relevant to arts and humanities researchers.

[1] http://www.big-social-data.net

II. Background — Reality Mining

Sandy Pentland is widely recognised for his idea of using mobile phones for 'reality mining' [3]. In 2004, he and his colleagues analysed 350,000 hours of mobile phone data and have since captured in 'living laboratories' human culture on an unprecedented scale [3]. They investigated mobile phone data from the Ivory Coast to track commuter behaviour on public transport. The aim was to find patterns that would help reduce commuting time. This experiment is typical for a host of reality mining projects that discover patterns of real-life regularities in terms of mobility, lifestyle choices, opinions, etc. from born-digital traces of human life.

BSD was as a concept developed by Lev Manovich [4] and has since seen great interest as the subject of research in digital culture and society. One of the earlier examples in the UK is [5], discussing how Twitter has influenced the movement during the 2011 London riots. Twitter has been used both by rioters and the police to coordinate their actions. In Australia, Queensland University of Technology's 'Mapping Online Publics' [6] examines the use and impact of the social networking tool Twitter for public engagement. 'Our data ourselves' adds to the contemporary BSD research a specific perspective on born-digital youth cultures in smart phone environments. We aim to learn about the composition of the 'data footprint' of engaged youth, how their information circulates among different corporate organisations and how that data might develop from communities of consumption to a communities of research.

We approach a complex technological realm not just as an object of study for arts and humanities research but in order to transform it into research tools and a means for the production and circulation of knowledge. The technology developments, which we present in Section III, are thus complemented by focus groups and interviews (Section IV). In Section V, we introduce initial results of our on-going work.

III. Open Digital Culture Devices

This section presents devices we developed for our reality mining environment: an Android MobileMiner app and a social data commons.

A. MobileMiner App

For our experiments we chose to investigate BSD as generated by smartphones that use the Android operating system. Android has the advantage of allowing to install apps from unofficial sources without rooting or 'jail-breaking' devices. This will help reach a wider audience and has allowed our MobileMiner app to be developed in conjunction with YRS.

MobileMiner is intended to gather information about the device on which it is installed and other apps on it. The participants in our experiments have been selected from young coders, who are coming regularly to YRS events. They were issued with Android smartphones, on to which the app had been installed, and asked to assist with its development and testing. The app captures data that another third party might be already recording in order to investigate how digital culture is tracked in mobile environments.

MobileMiner uses the Android API, which provides functions that return the total number of bytes transmitted and received by a given app. MobileMiner polls these every half second, and logs periods where these values continually increase. This direct approach was suggested by one of our YRS coders, and always captures when an app has been active. More detailed information about network traffic can be obtained. Each app on an Android device corresponds to a user id in the operating system's underlying Linux kernel. The kernel provides virtual files, `/proc/<pid>/net/<protocol>`, where `<pid>` is the process id of a given app. and `<protocol>` is an internet protocol; usually `tcp` or `udp`. Every half-second, MobileMiner reads all files it is allowed to access for each running app. These follow a standard format, and thus allow the port, IP address and protocol of each network socket to be determined. This way, all activities of the Chrome web browser on the mobile, for instance, are detected, along with that of apps such as Facebook, Skype, Foursquare, Spotify, many games, etc. This approach is used by commercial providers, too, and has the advantage that it does not require permissions granted by the user when the app is installed.

However, MobileMiner also collects data that requires the users to grant explicit permission as a condition of installing it. The app, for instance, records the unique ids of each cell tower in the mobile network the device connects to, as well as the names and BSSIDs (unique identifiers) of wireless networks. The Android API supports these kinds of services very well. We decided against tracking more fine-grained location data provided by GPS at the moment, because it would be too invasive, and cause significant power consumption issues. Especially, the latter would have been an issue for our experiments, as the young coders often used relatively low-tech environments because of budget constraints.

We finally convert the cell tower data into very approximate location data using the gazetteer of cell tower locations provided by `opencellid.org`. The user is also asked to enable MobileMiner as an *accessibility service*. Normally, this would be used, e.g., to provide a text-to-speech service for the visually impaired. We employ it to record the times when internet-enabled apps push notifications to the user.

B. Social Data Commons

MobileMiner provides participants with the option to start and stop recording data whenever they desire, as well as an overview of which apps have active network connections. They can copy their data to an area of the device's file-system accessible when connected to a personal computer.

We intend to make the data collected by MobileMiner as open as possible (subject to privacy concerns), in what we call a social data commons. To this end, we host an instance of the open source data management application CKAN. CKAN has the advantages of a high degree of interoperability with other environments through common formats such as RDF, the built-in display of geo-spatial data, and the ability to federate with other instances.

Each new installation of MobileMiner requests a unique, but anonymous identifier from the CKAN instance via a custom Python plugin against which the uploaded data is stored. Every two minutes the app writes newly accumulated data to the mobile device's SQLite database. As the cell tower connected to a mobile device can change frequently, writing this data as soon as it is acquired, rather than in batches, would adversely effect power consumption. Once the device is connected to a wireless network, the app uploads newly acquired data to the CKAN instance via the plugin.

Participants in our experiments can also explore their data directly on the device through the app. The app provides, e.g., a list of the apps that open the most network sockets. Data on cell towers in the U.K. from `opencellid.org` is also included with the app, which allows a heat-map of frequently visited cells to be built up using OpenStreetMaps and the OpenLayers JavaScript library.

However, we have only started to explore social media and cultural analytics techniques in the project, which will be a major focus of its second phase — after we have collected enough data. At the moment, we have already developed an environment to investigate Twitter user behaviour based on the Neo4J graph database and a Hadoop cluster, which we jointly develop with the YRS coders.

IV. Investigation of Youth Digital Culture

We follow a holistic approach to community engagement and see our partners as co-researchers throughout the entire project cycle. They generate BSD in-community, and work with us to develop the tools and applications for both its capture and analysis, as just discussed. The end result will be an open environment for BSD research with tools, applications and an infrastructure available for widespread community use.

As seen, we have been working with our co-researchers from YRS to help us develop tools and applications to capture, visualise and understand key components of their BSD; specifically what they generate when they text, browse and produce content on their smartphones. But, we have also engaged them in a more traditional way through interviews and focus groups that delivered attitudes to privacy and data use. To date, we have conducted four, ninety-minute focus groups that accompanied one hackathon.

A. Focus Groups and Interviews

The focus groups and interviews were conducted prior to the hackathon with a total of 21 participants in an age range of 14 to 18. The participants were chosen from a list of young coders involved with YRS in the past. The focus groups and interviews targeted the complex relationship young people have with their online privacy and tried to raise a series of questions. How much control do users actually have? Is privacy in social media something that is experienced individually or collectively? Why is there so much apathy in relation to

the amount of data that young people are willing to share about themselves online? How can a more engaged strategy be formulated? Overall, the focus groups revealed issues of control with regard to one's personal data, changing attitudes towards privacy and finally strategies to address these issues and attitudes, as we will discuss in Section V.

B. Collaboration with YRS - Hackathons

During the hackathon members of YRS considered ways in which smartphones generate BSD; complementing the interviews and letting the young coders experience how BSD is used and exploited in mobile environments. The hackathons build on the requirements of BSD privacy, which the focus groups delivered, and allowed groups to (a) improve the MobileMiner app and produce creative ways to visualise the data it is capturing; to (b) think through the ways privacy agreements increase dataveillance on users; and to (c) finally consider the growing access to personal data that third parties are being granted; particularly via seemingly benign smartphone applications. The hackathon results are published through the YRS community.[2]

V. Initial Results

The project is on-going and will conclude in early 2015. This section presents initial results of our work on devices to develop and understand an open digital culture with big social data and the attitudes expressed in the interviews.

The MobileMiner app has proven to be a useful tool to track mobile phone activities, where the young coders' interactions with casual games provide a good example. Online games are one the foremost expression of digital culture and generally track social activities in the background. Many such games are available on the Google Play Store at no cost, and so are monetized by their developers through serving advertisements to the player [7]. In order to understand the interactions of young people with these kinds apps we can offer two typical examples, the first one exhibiting standard behaviour and the second heavier usage.

Three of the young coders have played the game *Don't tap the white tile*, which claims over 130 million downloads at the time of writing [8]. MobileMiner registered the game accessing the internet on each of player A, B and C's devices on 46, 53 and 42 occasions, over periods of 21, 2 and 3 three days respectively. Normalised plots of the number of internet access events at each hour of the day (Fig 1) show the players interacting with the game for limited periods during the early morning, and early and late evening. The game was played somewhat intensively, then abandoned. Commercial studies report that up to 95% of mobile apps see similar patterns of use [9].

This result is in complete contrast to player A's interaction with *The Line-Keep In* (Fig 2), which is a simple game involving navigating a dot through a vertically scrolling maze that has only around 12000 downloads [10]. An advert is served nearly every time the player collides with the maze. On closer inspection of the code, the game uses various tools

[2] http://hacks.youngrewiredstate.org/events/Kings%20College

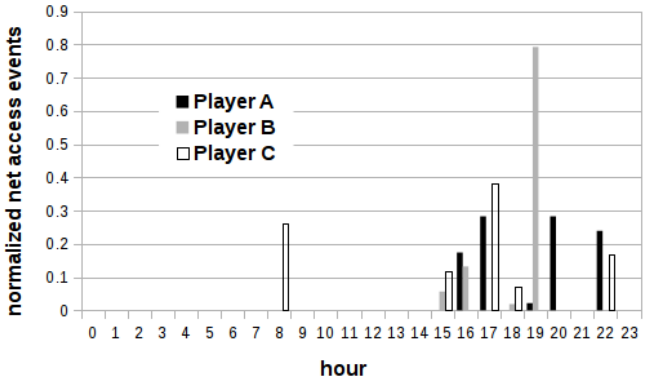

Fig. 1. Network access for the game *Don't tap the white tile*.

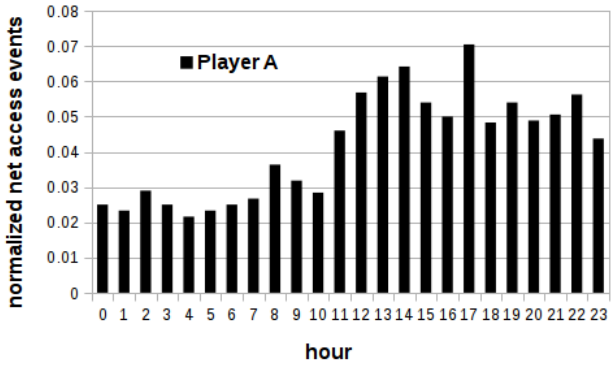

Fig. 2. Network access for the game *The Line*.

to gather deep statistics and push user messages. Permission to access the player's GPS location is requested, whereas the Mobile Country Code of a device's last connected cell tower would be sufficient to localize advertisements to the player's country. Despite the game's simplicity, the player's device interacted with the internet 1760 times over a period of 27 days. Each time of day registered significant activity.

Our app shows how easy it is to track behaviour using mobile devices. Both game examples demonstrate how advertisers could use the information from mobiles to push ads according to user location and behaviour, and to target communities. The second app's monitoring services are also used in a new type of apps that help parents monitor their children's behaviour [11].

The participants in our studies might not know the exact nature of how games and other apps track their digital behaviour but they understand some of the key issues involved well. This is the initial result of evaluating our interviews and focus groups.

The study participants were first of all generally concerned with the idea of having control over their own data. This was a common theme running through each of the focus groups. Many participants felt strongly that their intimate understanding of technology afforded both control over what

they chose to put out in the public domain and over their data that contain more intimate information such as location, address, likes, dislikes, friends and so on. One typical response exemplifies this belief:

> 'Being of kind of this generation and being tech savvy we have some control because we know how to have control, where as I know that my mum doesn't have any idea ...' (Participant A)

Conversely, there were other participants who appeared more critical and not nearly as confident, questioning how one could ever truly have any such control:

> 'People don't realise how large their digital footprints actually are. (...) it is incredibly easy to track down the personal details of someone ...' (Participant H)

Most of the participants agreed that privacy is relative to the amount of information that you are comfortable sharing. Some even expressed that 'if you have nothing to hide you have nothing to fear'. However, participants also worked through the idea that privacy is not something that should necessarily be understood as an individual construct but rather as something that is experienced collectively. In the words of one of female participant:

> Privacy is 'attached to other people ... so if someone you agree to connect with is open then you can be accessed through them cause it's kind of herd thing, you've all got to do it otherwise, one person is in trouble.' (Participant C)

Many of the young coders believed that their peers are too apathetic when it comes to sharing information about themselves. Some of the reasons for this apathy that were discussed (although are not limited to):

1) Resistance to and frustration with fear-mongering parents and teachers regarding online dangers.
2) Poor understanding about how technology functions and the ease by which information can be discovered about anyone.
3) The feeling that there is a great distance between the data that they produce and how it gets redeployed by marketers or sites like Facebook or Google.
4) Group pressure to use social media platforms to connect with friends.
5) Positive benefits associated with the circulation of big social data.

While most of our YRS participants believed that privacy is a concern, many agree that being on social media can make up the conditions for this historical moment:

> 'We have a few friends who aren't on social media but it's really hard core, (...) how do i talk to you, you aren't on Facebook!' (Participant A)

Discomfort with social media was mainly expressed with respect to the inability to delete what you have posted about yourself. One participant discussed how she had deleted an old Twitter account in 2011, but all the Tweets could still be easily accessed. Other concerns revolved around privacy agreements (for those who read them, which accounted for one third of our participants) and how little they inform or protect the user.

Some of our most active participants, who were very concerned about privacy, engaged in different activities to protect themselves. Their strategies ranged from deleting tags that reveal geolocation on photographs, to setting up proxies, to using alternative open source platforms, etc. However, many of these participants felt it was difficult to maintain some of these strategies, because they could not convince their peers to adopt such alternative and proactive approaches to privacy protection. Again, implying that a successful approach to privacy must address the collective as opposed to the individual.

VI. Conclusion and Future Work

The project has demonstrated how to connect with and research a community of young intensive users of social media in the mobile ecosystem. We have been developing a co-research platform with these users that opens the experience of BSD collected by mobile phones. The platform consists of an app to track mobile behaviour and a BSD commons to store the collected data. Section V has presented typical results such a platform can generate from tracking mobile game behaviour.

We have been using hackathons to co-develop tools with YRS that extract BSD from mobile phones. In the focus groups and interviews, the youth communities seemed well aware of the potential of BSD for research and commerce as well as the major issues associated with it. In our future work, we will work with the YRS communities to co-develop a set of analyses that will utilize BSD for humanities research in digital culture.

References

[1] D. Nafus and J. Sherman, "Big data, big questions— this one does not go up to 11: The quantified self movement as an alternative big data practice," *International Journal of Communication*, vol. 8, p. 11, 2014.

[2] "Young rewired state website." [Online]. Available: https://youngrewiredstate.org

[3] A. Pentland, *Social Physics: How Good Ideas Spread-The Lessons from a New Science*. Penguin, 2014.

[4] L. Manovich, "Trending: the promises and the challenges of big social data," 2011.

[5] R. Procter, F. Vis, and A. Voss, "Reading the riots on twitter: methodological innovation for the analysis of big data," *International journal of social research methodology*, vol. 16, no. 3, pp. 197–214, 2013.

[6] A. Bruns and J. Burgess, "Researching news discussion on twitter: New methodologies," *Journalism Studies*, vol. 13, no. 5-6, pp. 801–814, 2012.

[7] A. Tongaonkar, S. Dai, A. Nucci, and D. Song, "Understanding mobile app usage patterns using in-app advertisements," in *Proceedings of the 14th International Conference on Passive and Active Measurement*, ser. PAM'13. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 63–72. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-36516-4_7

[8] "Don't tap the white tile, google play." [Online]. Available: https://play.google.com/store/apps/details?id=com.umonistudio.tile

[9] "Marketing pilgrim blog." [Online]. Available: http://www.marketingpilgrim.com/2013/09/twenty-two-percent-of-mobile-apps-are-only-used-once.html

[10] "The line-keep in, google play." [Online]. Available: https://play.google.com/store/apps/details?id=com.onetouchgame.TheLine

[11] S. Brown, "6 apps and gadgets to keep track of your child, some controversial," 2014.