Application of multiple change point detection methods to large urban telecommunication networks.

Andrew Shields
Department of Computing,
Institute of Technology Tralee,
County Kerry, Ireland
andrew.shields@staff.ittralee.ie

Pat Doody
Department of Computing,
Institute of Technology Tralee,
County Kerry, Ireland
pat.doody@staff.ittralee.ie

Ted Scully
Department of Computing,
Cork Institute of Technology,
County Cork, Ireland
ted.scully@cit.ie

***Abstract*—An integral enabler of the smart city vision is the ability to effectively model collective population behaviour. The realisation of sustainable smart mobility is underpinned by the effective modelling of the spatial movements of the population. Furthermore, it is also crucial to identify significant deviations in collective behaviour over time. For example, a change in urban mobility patterns would subsequently impact traffic management systems.**

This paper focuses on the issue of modelling the collective behaviour of a population by utilizing mobile phone data and investigates the ability to identify significant deviations in behaviour over time. Mobile phone data facilitates the inference of real social networks from their call data records (CDR). We use this data to model collective behaviour and apply change-point detection algorithms, a category of anomaly detection, in order to identify statistically significant changes in collective behaviour over time. The result off the empirical analysis demonstrate that modern change point detection can accurately identify change points with an R^2 value of 0.9633.

Keywords— Change point detection, event prediction, telecommunication networks, spatio-temporal data mining.

I. Introduction

Mobile technologies and services provide a fundamental role in enabling smart cities. Data from mobile phones can support smart cities services such as sustainable smart mobility; disaster recovery and can provide a platform for gathering societal behavioural data. As mobile phone use has become more ubiquitous, it has impacted the way we communicate and interact. Currently there are in excess of 7 billion mobile cellular subscriptions worldwide, which have grown from less than 1 billion since the start of the new millennium in 2000. Mobile broadband use has achieved 47% penetration in 2015, a value that has increased by a factor of 12 since 2007 [1].

A. Call Data Records (CDR)

This ubiquity allows mobile technology to act as sensors of our environment. Call data records (CDR) are captured by the mobile phone operators and are used for billing purposes; this data reveals with whom and how often their customers communicate. Call data records also contain rich meta-data on each individual such as location, Internet usage and movement data. Augmenting this information with external data on customers such as age and gender makes mobile phone CDRs a rich source of data for research. The past few years have seen the rise of research based on the analysis of CDRs including the MIT Reality Mining Data Set [2] Nokia's Mobile Data Challenge (MDC) [3] and the Orange Challenge 4 Development (D4D) [4].

Because a mobile phone is a personal object it facilitates the inference of real social networks from the CDR data. Each mobile phone can be viewed as a representation of an individual and those whom form part of their social network. By being mobile, the phone user has the possibility to answer and make calls at any time, which facilitates the study of temporal patterns of communication. Another important factor is the positioning functionality inbuilt in phones such as cell towers, GPS and Wi-Fi locators; these allow the spatial movements of the user to be recorded [5].

B. Collective behaviour

Collective behaviour is a term describing the behaviour of a large number of people. A pivotal component of the smart city vision is the capability to model certain aspects of the collective behaviour of the population or a subgroup of a population [6],[7]. The continued increase in urbanisation and population growth will place significant pressure on existing urban infrastructures and resources. When analysing collective behaviour data, it is important to build a model that persists under normal situations, it is also important to be able to identify significant anomalies or deviations from normal collective behaviour. That is, we need to reliably detect events that are anomalous to the current model. For example, the early identification of a deviation from normal collective mobility patterns could provide valuable input to an intelligent traffic management system and water management [8].

C. Change point detection

Change point detection involves identifying points where the probability distribution of time series data undergoes statistically significant change. More specifically, they model the normal behaviour of the recorded data and then identify when the data deviates from this normal behaviour, thus a new normal behaviour will be established in the network and the model will then look for further deviations. Hence these networks are continually evolving and the model must continually adapt to theses evolving structures.

978-1-5386-1046-6/17/$31.00 ©2017 IEEE

D. Summary

In this paper we use mobile phone data for the application of change point algorithm as a means of identifying changes in mobile phone call and SMS data logs. We evaluate the application of the change point algorithm and analyse the performance of the models based on known significant events and compare these results to other research in the field to demonstrate the contribution of this work.

This reminder of this paper is laid out as follows; Section 2 describes related work and provides an overview of the change point detection algorithms utilised in this study. Section 3 describes the MDC mobile dataset, which are used in the evaluation. Section 4 describes the experimental methodology, including the pre-processing of the data. Section 5 presents a detailed empirical evaluation and finally conclusions are drawn and future work identified in Section 6.

II. Related Work

A. Background

The term 'smart city' has various definitions in the literature, however one criteria that makes a city 'smart' refers to the increasing extent to which pervasive and ubiquitous computing and digitally instrumented devices are embedded into the infrastructure of urban environments. For example, telecommunication networks, utility services, transport infrastructure, sensors, building management systems, etc. may be used to monitor, manage and regulate the infrastructure of the city, often in real-time.

Arguably the goal of any smart city is to create a system where data from the population and sensors embedded in the city infrastructure are routed into the central operating system of the city, allowing the infrastructure of the city to adapt and respond to changes [8], [9]. This vision of a smart city is still evolving but many cities utilise existing technologies moving towards this goal. For example, Rio de Janeiro has a large control centre, which applies data analytics to social media, sensors and surveillance cameras in an attempt to predict and control events taking place in the city. Other cities such as Santander and Singapore have invested in sensor networks to record a range of environmental and traffic conditions at locations across the cities. Intel and Dublin City Council announced that Dublin is also to get a sensor network for measuring city processes such as smart bins and the Dublin Dashboard which displays real time environment and transport data across the region [10], [11].

Smart phones have become ubiquitous and are used by most urban citizens; they may be used to interact with and within the city. This not only produces rich information on the citizens' interactions but also augments this with location and activity data. Connecting, integrating and analysing the information produced by these various forms of sensors, provides a more holistic interconnected understanding of the city that enhances efficiency and sustainability [12]. As these system move towards real-time sensing the data can be processed to better model and understand urban processes thereby forecasting future urban development [9]. Such data improves public services providing the infrastructure for business activity and growth and stimulates the service and knowledge economy.

While the ability to collate this data and produce accurate models is important it is equally important to be able to accuracy identify deviation from atypical patterns of behaviour or operation. Many different anomaly detection techniques and methods have been developed including statistical methods, streaming algorithms and machine learning algorithms which are beginning to show potential for commercial application [13],[14]. As with many approaches there is a balance to be found between the volume and complexity of the data available and the performance of the models. Statistical anomaly methods such as change point detection require large volumes of data to provide samples, from which accurate estimates of a network normal behaviour can be made; therefore, their application in mobile telecommunication networks is appropriate.

B. Change Point Detection

Change point analysis is a process of detecting distributional changes within time-ordered observations [15]. Such models try to estimate the point in a time series where the statistical properties of a sequence of observations undergo a significant change. Detecting such changes is important in many different application areas. Recent examples include climatology [16], bioinformatics applications [17], finance [18] .

Change point analysis operates on an ordered sequence of time series data points , $y1{:}n = (y1, \dots, yn)$. Where a number of change points exist m. A change point is detected within these points when the statistical properties of the points change in some way, usually a change in mean or variance. This can be used to detect multiple changes points within the sequence. In the case of multiple change points each change point will have a position typically divided along a time series $\tau 1{:}m = (\tau 1, \dots, \tau m)$. This results in splitting the data into a number of segments, identifying where a change occurred, generally defined as anomalous behaviour [19].

To explain how change point detection operates let us elucidate with a simplified example: In a telecommunication network a list of call interactions have been recorded of normal behaviour over a period of time. These interactions may vary during normal operations as in the transition from weekdays to weekends. Such changes are inherent in the normal behaviour of users and should therefore not be classed as significant events. Change points should be identified along the time series when an extreme change occurs from a response to a significant event, such as a natural disaster or change in regular routine such as holiday periods or major social events [20]. Detecting such extreme changes in social networks could provide a better understanding of patterns of city life and an early detection of impacts on city infrastructure. Several change point search algorithms have been implemented in various programming languages and platforms, most notably the binary segmentation algorithm [21], [22], the segment neighbourhood algorithm [23], [24] and more recently the Pruned Exact Linear Time (PELT) algorithm [19].

1) Binary Segmentation

The Binary Segmentation method facilitates the identification of multiple change points by iteratively applying the binary segmentation method on the dataset. It initially applies a single change point test statistic to the entire data; if a change point is identified the data is split in two at the change point location, resulting in two periods before and after the change. This process then continues on each of these identified periods, each time a new change point is discovered the method again divides the data into subsets split by each discovered change point location. This continues until no new subsets can be created, the final set of change points is the location of all the split points.

The Binary Segmentation search method is computationally efficient, resulting in a $O(n \log n)$ calculation. However, this efficiency comes at the cost of accuracy. Because of the binary splitting of the data at each change point, each new change point is dependent on the locations of the previous change points. The consequence is that the dataset will not be entirely examined, resulting in a subset of the locations being searched providing an approximation of the change point locations. This may not be an issue with a small number of change points, however, issues may become apparent when dealing with datasets containing many change points or where the distance between one change point and the next is small. Therefore binary segmentation is an approximate algorithm but is computationally fast[25].

2) Segmentation Neighbourhood

The segmentation neighbourhood method allows for adjusting the maximum number of change points to discover by altering the value of a parameter Q. It then calculates the cost of all possible optimal segmentations from 0 to Q change points. This searches over all Q possible change point locations. The computational cost for this method is $O(Qn^2)$. Therefore, this method performs poorly as the size of the data set increases particularly where the number of possible change points is large. Whilst this algorithm is exact, the computational complexity is considerably higher than that of binary segmentation.

3) Pruned Exact Linear Time (PELT)

PELT is a newer method which attempts to be computational efficient while maintaining accuracy. It achieves this by performing sequential searches of the data and it achieves efficiency by pruning paths that do not to lead to a solution.

PELT relies solely on a penalty value to determine the number of change point's m. As such the optimal segmentation is $F(n)$. The model commences by calculating $F(1)$ and then recursively calculates $F(2), \dots, F(n)$. At each iteration the optimal segmentation up to $\tau m + 1$ is stored. When the model reached the end of the series $F(n)$ the optimal segmentation for the entire data has been identified and the number and location of change points have been recorded. At each step the minimisation over τm covers all previous values. A more complete description of this method is described in [19].

III. Data source

A. 3.2 Mobile Data Challenge

The dataset used in this study was the Mobile Data Challenge (MDC), which was a research initiative where the goal was to create a large-scale mobile data resource that would facilitate community-based evaluation of mobile data analysis methodologies. One of the first tasks of the research was the data gathering exercise, which the research team termed the Lausanne Data Collection Campaign (LDCC) [26]. This was a data collection activity that recorded unique, longitudinal smartphone data from nearly 200 volunteers in the Lake Geneva region over a one-year period.

The LDCC study produced a large data set with a temporal dimension, and variety of data types. The study consists of 185 participants (38% female, 62% male), with the majority of participants aged between 22-33 year-old. This age demographic accounts for approximately two thirds of the population. The data generated during this study included location (GPS, WLAN), motion (accelerometer), proximity (Bluetooth), communication (phone call and SMS logs), multimedia (camera, media player), and application usage [3].

IV. Data Preperation

While previous research work has investigated the application of change-point detection techniques to Bluetooth interactions [27] little has been undertaken in applying these models to call and SMS data. While Bluetooth interactions give a good understanding of the physical interactions among individuals, traditionally this data can prove difficult to gather as it requires the use of custom application running on the phone and requires the user to continually operate the Bluetooth function [28]. Alternatively call and SMS interactions are readily available to cell phone network providers through log records and have been used in previous studies [20],[7],[13].

Specifically, in this work we aim to study a succinct group of individuals in order to evaluate the accuracy of change point detection techniques when compared to a pre-determined sequence of known significant events. To facilitate this study we adopt a methodology similar to that presented in [# provide reference for MIT paper], where the data analysed is related to an academic campus and significant events are identified from the academic calendar.

The initial task is to extract a distinct cohort of individuals from the MDC data set. We adopt an approach which focuses on identifying the patrons from the University of Geneva. The university cohort was chosen for a number of reasons (1) students and staff are a well-understood social group and comparison to the well-studied MIT data set will be possible. (2) The university is spread over six campuses in the metropolitan area of Geneva. We can therefore use GPS positional data to infer university populations. (3) The calendar of significant academic events of a university is publically available. For example, exams, semester start times public holidays etc. are all available from the university's academic calendar. Using GPS traces we can isolate the targeted group using spatial information.

For each of the campuses a radius of 250 meters from the centre point was analysed. This distance was selected, as it is broad enough to encompass each campus without identifying participants that are external to the university. The GPS table from the MDC data set was then queried to extract each GPS reading within this radius of each campus, the data set is filtered to only include individuals that have repeated GPS readings within the time frame under study. This is to reduce the likelihood of including transient or one time visitors to the campus; all university staff and students will typically have more than a single GPS reading associated with a campus. Subsequently, we identify the unique users associated with the filtered GPS coordinates. Next, all call log entry and SMS entry data for each unique user is retrieved and stored. It should be noted that the collective call and SMS data is not limited to on-campus interactions. Each call and SMS is time stamped. Finally, all the captured data is aggregated into weekly time slots as the university academic calendar deals in weekly events.

V. Empirical Evaluation

This section describes the application of the binary segmentation, segmentation neighbourhood and PELT change point detection algorithms to MDC dataset described in the previous section.

A. Methododology.

As described previously we have pre-processed the data to extract the university population only, distributed across the campuses of the University of Geneva. We selected a period of time that corresponds with the academic year 2009-2010. Table 1 outlines the list of significant events obtained from the academic calendar of the University of Geneva.

TABLE I. UNIVERSITY OF GENEVA SEMESTER DATES OF KNOWN SIGNIFICANT EVENTS.

Week Number	Description
0-2	Exams: August/September session
3	Start of Semester 1
11	Reading Week
16	End of Semester 1
17-21	Christmas Break
22-24	Exams
25	Start of Semester 2
30-31	Last Week before Mid Term Break
32-33	Easter
35	May bank holiday
39	End of Semester 2
40-43	Exams Semester 2

Upon initial visual inspection of the histogram in fig 3 we can see that there is at least one point where the data deviates significantly from the norm, week 19, this coincides with the middle of the Christmas break.

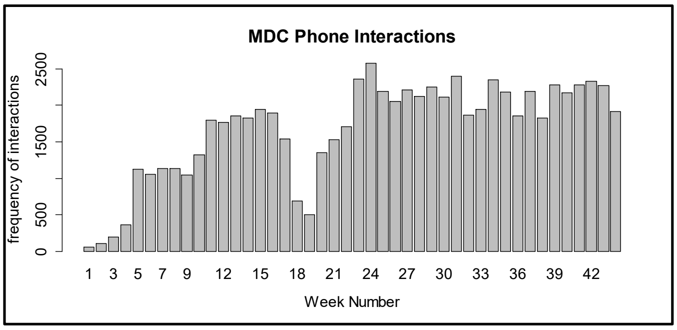

Fig. 1. Histogram displaying the number of interactions from the MDC call log data set aggregated by weeks for Geneva University Students.

1) Choice of penalty

The MDC dataset, as with real smart city sensor data does not have a pre identified event list. Therefore, we will not be limiting the number of change points by setting the Q parameter for binary segmentation and segmentation neighbourhood methods. However, the penalty value will be adjusted for all methods, providing a balance between the identification of many small change points, which may not be significant and identifying very few change points. The appropriate value of the penalty depends on the data and the question we need to answer.

There are many ways to choose the penalty value two of the most common are: (1) trial and error, iteratively try several different penalty values until one is found that is appropriate for the problem (2) "Elbow-plot", i.e. plot the number of change points identified against the penalty used. This creates a curve whereby small values of the penalty produces large (spurious) changes and as the penalty decreases these spurious changes drop off as only true changes are left before slowly converging to no changes for larger penalties. Method (1) above is a relatively ad-hoc approach and can prove time consuming. Method (2) provides a more objective method of selecting an appropriate penalty value. Therefore, we adopted elbow-plots in order to select an appropriate penalty. One should choose the penalty at the elbow of the plot. This is because at low penalties, spurious change points are generated that drop off as you increase the penalty. Generally, we extract the components on the steep slope; the components on the shallow slope contribute little to the solution. A problem with elbow plots is finding a sharp break. Interpretation may suffer from subjectivity and ambiguity, especially where there are either no clear breaks or two or more apparent breaks.

For our examples in Fig 2 binary segmentation the last significant drop occurs between change points 16 and 17, therefore we chose the corresponding penalty value of 40500. Binary segmentation does not search the entire solution space as it is a non-exact approximation method, therefore it detects no change point beyond 23. In segmentation neighbourhood the last drop is between the 18 and 19 change points; after this point we can see some spurious results with the number of change points rising to 43 and dropping to 23. Therefore, we select a corresponding penalty of 41200 and finally PELT has a final drop at change point 19 with a penalty selection of 41500.

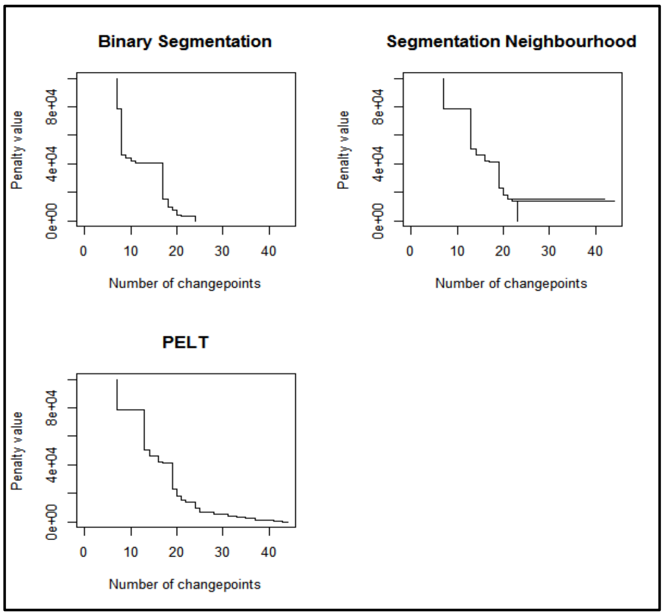

Fig. 2. Elbow plots for assessing penalty values.

2) Model Outputs

We now run the three change point methods on the data, the Q parameter for the binary segmentation and segmentation neighbourhood is set to the maximum possible value and our penalty values is set according to the elbow plot results, as outlined the previous section.

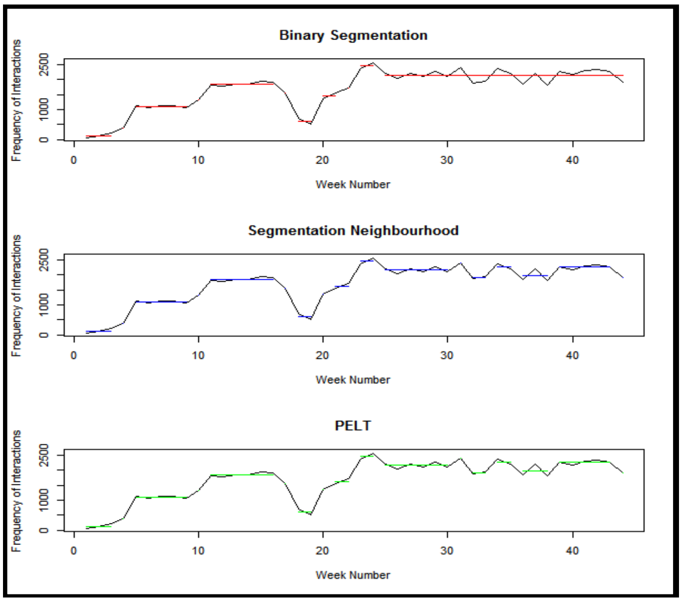

Fig. 3. Plot of the MDC dataset with horizontal lines for the underlying (fitted) mean change points with modified parameters.

Fig 3 shows the results for each of the methods with horizontal lines representing each of the detected change point locations, it is interesting to note the single horizontal line in semester two from the binary segmentation method, illustrating that no change points were detected. Binary segmentation returns change points at weeks (3, 4, 9, 10, 16, 17, 19, 21, 22, 24) segmentation neighbourhood returned change points at weeks (3, 4, 9, 10, 16, 17, 19, 20, 22, 24, 30, 31, 33, 35, 38, 43). And PELT returns change points identical to segmentation neighbourhood at (3, 4, 9, 10, 16, 17, 19, 20, 22, 24, 30, 31, 33, 35, 38, 43). This corresponds to the literature where segmentation neighbourhood and PELT both perform a full search of the dataset.

3) Results

The final stage of this analysis is to compare the discovered change point with our known significant events. Table 2 illustrates the expected significant events (as per University of Geneva academic Calendar) alongside the change point detected by the change point methods, grouped to the closest known event. From inspection of table 2 and fig. 3 it can be seen the change point methods have identified many of the significant events, with segmentation neighbourhood and PELT performing well over both semesters while binary segmentation only detecting events in semester 1 as previously discussed binary segmentation is an approximation method and does not search the entire space for change points.

TABLE II. RESULTS OF THE CHANGE POINT DETECTION METHODS

Expected Week	Desc	Bin Seg	Seg Neig	PELT
0-2	Exams: August/September session	-	-	-
3	Start of Semester 1	3, 4	3, 4	3, 4
11	Reading Week	9, 10	9, 10	9, 10
16	End of Semester 1	16	16	16
17-21	Christmas Break	17, 19, 21,	17, 19, 20	17, 19, 20
22-24	Exams	22, 24	22, 24	22, 24
25	Start of Semester 2	-	-	-
30-31	Last Week before Mid Term Break	-	30, 31	30, 31
32-33	Easter	-	33	33
35	May bank holiday	-	35	35
39	End of Semester 2	-	38	38

In order to evaluate the models accuracy against the expected known event we use the calculated root mean squared error (RMSE). Here we are interested in the fit of the model to the data; how close the data points from our known events are to the models predicted values.

Because RMSE is an absolute measure of fit, the square root of a variance, RMSE can be interpreted as the standard deviation of the unexplained variance, and has the useful property of being in the same units as the response variable, in this case weeks. The calculated RMSE between the expected events and the binary segmentation method is 16.0 the RMSE for the segmentation neighbourhood and PELT methods is 2.872281.

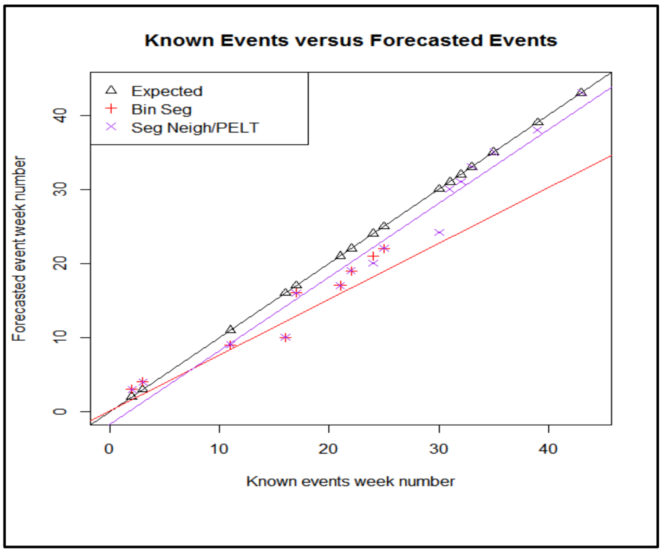

Fig. 4. Comparisons between the models and expected events.

While we have calculated the RMSE, which quantifies the accuracy of the models we would also like to quantify a models fit to the known events. This may be achieved by comparison to the expected significant events and the model counterparts. We performed linear regression on the data as shown in Fig 4. Segmentation neighbourhood and PELT have an adjusted R^2 value of 0.9633 with an intercept at -1.73042 while Binary segmentation has an adjusted R^2 value of 0.9409 and an intercept of 0.06366. By producing this scatter plot of the known events and the model predicated events it can be seen that the Segmentation neighbourhood and PELT models most closely match the known events, especially towards the end of the academic year. The Binary Segmentation model misses many change points, deviating in semester 2.

VI. CONCLUSIONS

The work outlined in this paper has focused on the application of change point detection algorithms to mobile data from Nokia's Mobile Data Challenge (MDC) in order to identify significant changes in collective behaviour over time. We have demonstrated that the change point detection models approach allows for the identification of significant events using data in the form of SMS and call interactions, the best RMSE for the methods is 2.872281, indicating that such a model could be used to identify events within a three-day period. In this experimental configuration, model parameters were adjusted. The selection of change points and the setting of the penalty values can significantly improve model performance.

An interesting extension of this work would be an investigation into multivariate methods to consider more variables such as internet usage, app usage and Bluetooth encounters which could lead to insights into the forming and separation of social groups associated with significant events. For example, what affect do orientation activities in a university have on forming social connections and how do these connections develop over the semester.

On a technical note there are many areas to investigate, firstly the application of alternative methods of change point detection. For example, machine-learning methods such as neural networks could be applied to learn and detect significant events. Another consideration is the application of these methods in a real world environment. Most research in this area has focused on static data sets, which have been previously gathered and accessed via a data warehouse or other centralised data store. Distributed paradigms should be investigated to allow for the efficient application of these methods in real smart city environments, minimising data transfer.

REFERENCES

[1] International Telecommunication Union, "ICT Facts and Figures 2015," Geneva, Switzerkand, 2015.

[2] N. Eagle and A. Pentland, "Social Serendipity : Proximity Sensing and Cueing," *Office*, no. May, 2004.

[3] I. Aad, M. Markus, J. K. Laurila, J. Blom, O. Dousse, D. Gatica-perez, O. Bornet, and J. Eberle, "The mobile data challenge: Big data for mobile computing research," in *Proceedings of the Workshop on the Nokia Mobile Data Challenge, in Conjunction with the 10th International Conference on Pervasive Computing*, 2012, pp. 1–8.

[4] Y. De Montjoye and Z. Smoreda, "D4D-Senegal: The Second Mobile Phone Data for Development Challenge," *arXiv Prepr. arXiv …*, no. 1, pp. 1–11, 2014.

[5] V. D. Blondel, A. Decuyper, and G. Krings, "A survey of results on mobile phone datasets analysis," 2015.

[6] M. A. Bayir, M. Demirbas, and N. Eagle, "Mobility profiler: A framework for discovering mobility profiles of cell phone users," *Pervasive Mob. Comput.*, vol. 6, no. 4, pp. 435–454, 2010.

[7] L. Gabrielli, B. Furletti, R. Trasarti, F. Giannotti, and D. Pedreschi, "City users' classification with mobile phone data," *Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015*, pp. 1007–1012, 2015.

[8] D. E. Difallah, P. Cudré-Mauroux, and S. A. Mckenna, "Scalable Anomaly Detection for Smart City Infrastructure Networks," *IEEE Internet Comput.*, 2013.

[9] M. Batty, K. Axhausen, G. Fosca, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali, "Smart Cities of the Future (Working Paper 188)," 2012.

[10] E. Moir, T. Moonen, and G. Clark, "WHAT ARE FUTURE CITIES? ORIGINS, MEANINGS AND USES," 2014.

[11] R. Kitchin, "The real-time city? Big data and smart urbanism," *GeoJournal*, vol. 79, no. 1, pp. 1–14, Feb. 2014.

[12] C. Ratti and A. Townsend, "The Social Nexus," *Sci. Am.*, vol. 305, no. 3, pp. 42–48, Aug. 2011.

[13] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási, "Uncovering individual and collective human dynamics from mobile phone records," *J. Phys. A Math. Theor*, vol. 41, pp. 224015–11, 2008.

[14] M. Thottan, G. Liu, and C. Ji, "Anomaly detection approaches for communication networks," *Networks*, vol. 2, pp. 1–19, 2010.

[15] D. S. Matteson and N. a. James, "A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data," *J. Am. Stat. Assoc.*, vol. 109, no. 505, pp. 334–345, 2014.

[16] J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, "A review and comparison of changepoint detection techniques for climate data," *J. Appl. Meteorol. Climatol.*, vol. 46, no. 6, pp. 900–915, 2007.

[17] V. M. R. Muggeo and G. Adelfio, "Efficient change point detection for genomic sequences of continuous measurements.," *Bioinformatics*, vol. 27, no. 2, pp. 161–166, 2011.

[18] A. Zeileis, A. Shah, and I. Patnaik, "Testing, monitoring, and dating structural changes in exchange rate regimes," *Comput. Stat. Data Anal.*, vol. 54, no. 6, pp. 1696–1706, 2010.

[19] E. I. A. Killick R, Fearnhead P, "Optimal detection of changepoints with a linear computational cost," *J. Am. Stat. Assoc.*, vol. 107, no. 500, pp. 1590–1598, 2012.

[20] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft, "Recommending social events from mobile phone location data," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 971–976, 2010.

[21] a J. Scott and M. Knott, "A Cluster Analysis Method for Grouping Means in the Analysis of Variance," *Biometrics*, vol. 30, no. 3, pp. 507–512, 1974.

[22] A. Sen and S. Srivastava, "On tests for detecting change in mean when variance is unknown," *Ann. Inst. Stat. Math.*, vol. 27, no. 1, pp. 479–486, 1975.

[23] I. E. Auger and C. E. Lawrence, "Algorithms for the optimal identification of segment neighborhoods," *Bull. Math. Biol.*, vol. 51, no. 1, pp. 39–54, 1989.

[24] J. Bai and P. Perron, "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, vol. 66, no. 1, pp. 47–78, 1998.

[25] R. Killick and I. Eckley, "changepoint: An R Package for changepoint analysis," *Lancaster Univ.*, vol. 58, no. 3, pp. 1–15, 2013.

[26] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila, "Towards rich mobile phone datasets: Lausanne data collection campaign," *Proc. ACM Int. Conf. Pervasive Serv.*, vol. Berlin, 2010.

[27] L. Peel and A. Clauset, "Detecting Change Points in the Large-Scale Structure of Evolving Networks," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 2914–2920.

[28] N. Eagle and A. S. Pentland, "Reality Mining : Sensing Complex Social Systems," *Methods*, no. June, 2005.