Visual analysis of social networks in space and time

Aidan Slingsby
giCentre
City University London
London, EC1V 0HB, UK
a.slingsby@city.ac.uk

Roger Beecham
giCentre
City University London
London, EC1V 0HB, UK
Roger.Beecham.1@city.ac.uk

Jo Wood
giCentre
City University London
London, EC1V 0HB, UK
j.d.wood@city.ac.uk

ABSTRACT
We designed and applied novel interactive visualisation to investigate how social networks – derived from smartphone logs – are embedded in time and space. Social networks were identified through direct calls between participants and calls to mutual contacts of participants. Direct contact between participants was sparse and deriving networks through mutual contacts helped enrich the social networks. Our resulting interactive visualisation tool offers four linked and coordinated views of spatial, temporal, individual and social network aspects of the data. Brushing and filtering techniques help us investigate how these aspects relate. We also simultaneously display some demographic and attitudinal variables to help add context to the behaviours we observe. Using these techniques, we were able to characterise spatial and temporal aspects of participants' social networks and suggest explanations for some of them. We reflect on the extent to which such analysis helps us understand social communication behaviour.

Categories and Subject Descriptors
E.0 [**Data**]: General

General Terms
Measurement, design, experimentation.

Keywords
Visual analytics, data visualisation, exploratory analysis information visualisation

1. INTRODUCTION
Many aspects of increasing numbers of people's lives are being organised through smartphone devices. Such devices and services record information about how, when and where we work, socialise, travel and relax, providing significant opportunities for social scientists and market researchers for identifying and understanding human behaviour [8, 4]. However, the heterogeneous and multivariate nature of such data, make it difficult to know which aspects to investigate. Visual analysis offers useful solutions [16, 1] and is widely advocated for studying human behaviour in time and space [5, 9]. We use highly interactive graphics to explore and filter data to help us identify and understand patterns of human behaviour.

Using call and GPS logs from the Lausanne Data Collection Campaign [12], we draw on techniques from visual analytics to explore three research questions:

- How can we characterise spatial and temporal aspects of participants' social networks using visual analytics?
- Can linking spatial, temporal and call connectivity patterns help us explain how participants construct their social networks?
- To what extent do smartphone device logs help us to understand social communication behaviour?

We designed novel, highly-interactive, coordinated graphical views that link social networks derived from call logs to spatial and temporal views. We use these to explore, characterise and try to explain the behaviour of participants in the study. We describe the design and how we identified our findings from it.

2. APPROACH
Our approach to visual analysis is that visualisation design and data analysis are not separate activities [18]. We iteratively design, prototype and test visualisation ideas, generating new research questions that inform the design process. At each iteration, the design becomes more refined, resulting in our final tool. Excel and SPSS were used for early exploratory ideas and analyses, but Processing [7, 6] – a lightweight graphically oriented set of Java libraries that facilitates the rapid prototyping of data visualisation designs – was subsequently used, along with other libraries developed from previous projects [17, 14]. An advantage of this approach is the design flexibility and the ability to leverage designs from previous projects. A disadvantage is that it can be more time-consuming than using off-the-shelf tools, but this is mitigated to some extent through our experience and use of existing libraries.

Normally, our work tries to draw out general trends and patterns from large datasets and we had considered using

This material was prepared for the Mobile Data Challenge 2012 (by Nokia) Workshop; June 18-19, 2012; Newcastle, UK.

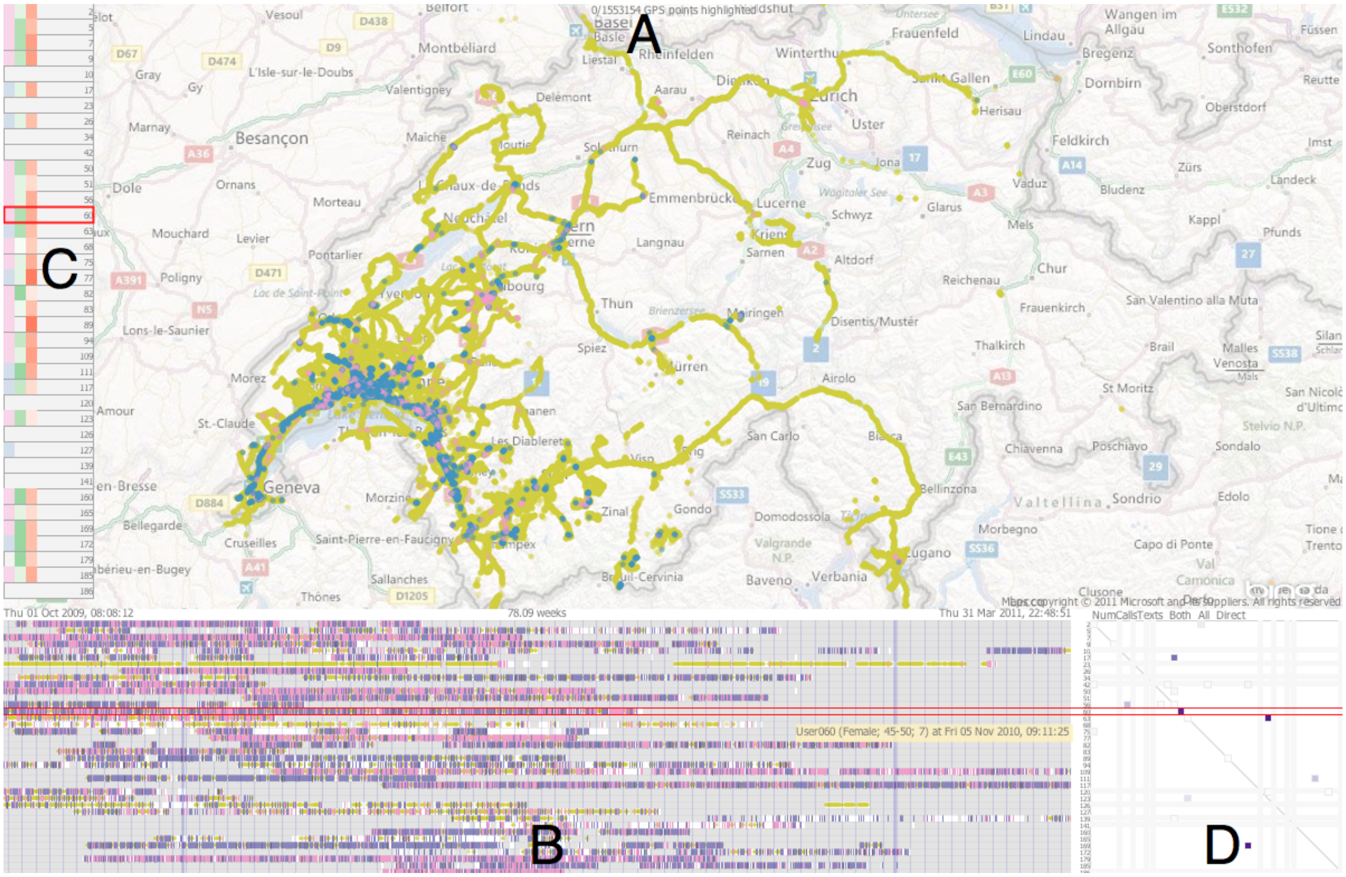

Figure 1: Screenshot of tool, with four coordinated views: [A] Zoomable map that shows all GPS points; blue and purple indicates where calls and texts were made. [B] Zoomable timeline (bottom) with one row per participant showing their GPS positions (yellow, calls (blue) and texts (purple). [C] List of participants where colours indicate gender (blue: male, pink: female), age (dark green: older) and a derived measure of social activity (dark red: more socially active). [D] Matrix (bottom right) of number of calls between participants. See our video at `https://vimeo.com/43245266`.

clustering techniques to identify different behavioural types. However, the relatively small and diverse sample of participants in this study – thirty-eight; nine of whom did not answer the questionnaire – makes such generalisations problematic. It became clear that *offering details of individuals* would supply context that might help produce tentative explanations of the behaviours observed. This focus on individuals is not scalable to many more participants, but is appropriate for this dataset.

We acknowledge that mobile call records represent only a part of a social network. Face-to-face contact, email and social networking websites are major aspects of social networks that we do not consider here.

3. DESIGN AND ANALYSIS

The design and data we chose to use were guided by our research questions. We used the call logs, GPS data and participants' phone numbers, personal survey responses and participation information. We identified calls between participants, between mutual contacts of participants (where at least one call was made) and georeferenced calls with GPS data based on their timestamps.

We were fortunate to already have a validated tool design for studying spatial and temporal aspects of GPS data. That design resulted from work with animal behaviouralists who were studying seagull behaviour from a sample of birds they had tracked with GPS loggers [14]. We took a user-centred approach to the design [14, 10]. The resulting tool – developed over an intensive two-week period of workshops, prototyping, feedback and evaluation – allowed the domain specialists to explore their data in a way that was not previously possible using a tool they helped design. The novelty of that work was the close involvement of the 'users' and evaluation based on their research questions. We used this validated design and code as a way of exploring space-time patterns of behaviour, modifying the design as necessary for addressing our research questions here.

The design is based on multiple coordinated views, to which we added participant attribute information, sorting and information about calling behaviour within the sample. ColorBrewer [3] helped us select appropriate colour schemes. The novelty of this work is the incorporation of social network information derived from the call logs.

As shown in Fig. 1, our design has four coordinated views: a map, timeline, characteristics of individual participants and a social network matrix. The video (`https://vimeo.com/43245266`) demonstrates the coordinated brushing [15] and selection across these views that enables locations within a particular temporal window, times within a spatial window and other relationship between the data represented by these four coordinated views.

When dealing with spatiotemporal data, multiple coordinated views have some advantages over other techniques often employed. Animating a map over time is a common and appealing technique because movement is depicted directly. However, our limited visual memory makes it difficult to identify trends [13], except perhaps over very short timescales. The space-time cube [11] is another popular technique, where the base of a cube represents 2D location and height represents time. However, it does not scale well beyond a few space-time paths and it suffers from a great deal of occlusion. Interactions that facilitate rotation and filtering can help, but they raise usability issues. Small multiples are often effective where data within adjacent time windows are ordered chronologically, row by row. This technique is limited to relatively few time windows and is only suitable where relevant movement is discernible between relatively few time windows (perhaps $\sim$30). Small multiples are unsuitable here, because we would need many more time windows to study movement at a temporal resolution of less than an hour over a timescale of more than a year.

The tool reads the required data from original data files, making it easy to replace with other data. The data volume supported depends on available RAM (the tool occupied 500Mb of RAM with supplied data loaded). The graphical design is not scalable to many more participants because one row/column is required for each participant (Fig. 1B and D), but if we had to work with many more participants, we would change our design appropriately.

3.1 Spatial view

The zoomable map displays all the GPS positions over the entirety of the study period. The GPS density surface (Fig. 2, centre) shows that most activity was focussed in Lausanne, around the Northern edge of Lake Geneva and Martigny. Colouring positions by participant (Fig. 2, right) enables the spatial extents of individuals' tracks to be identified, but occlusion where multiple GPS logs overlap is a problem. Brushing and filtering between views helps overcome this. Fig. 3 illustrates brushing to identify participant 63's use of space, then filtering and brushing on the timeline to reveal their use of space over a week.

3.2 Temporal view

The zoomable timeline shows participation, GPS positions, calls and texts. Fig. 4 shows participation was highest between the winter of 2009 and following summer, with a gap over the Christmas and New Year period. Zooming in, large gaps become apparent in the GPS records, probably due to participants being indoors and turning their GPS off (we were only able to georeference 6% of calls to within 6 hours of the call). In Fig. 1, participants with GPS locations but no calls are apparent (in yellow), those that make many voice calls but few text (blue) and heavy users of text messaging (purple).

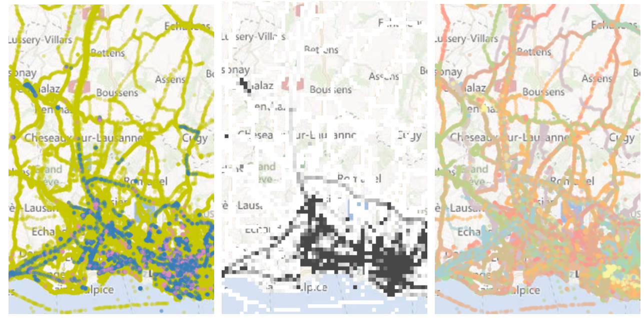

Figure 2: Zoomed-in alternative map views offered by the tool: GPS positions coloured by call activity (left), density (centre) and positions coloured by participant (right).

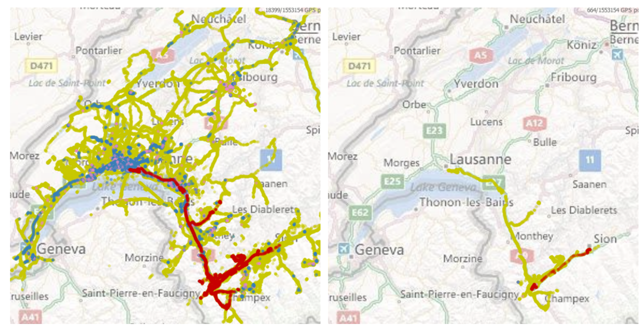

Figure 3: *Left:* Highlighting participant 63 through brushing. *Right:* Hiding other participants and brushing a week-long window on the timeline.

3.3 Participants view

Participants are listed down the left of the screen. Blue/pink, green and red indicate gender, age and recent social activity respectively (Fig. 1). Sorting by these characteristics in the participant view enables sorting in the timeline and matrix views. This view provides the basis for brushing users and the selection of one or more participants (Fig. 3). A tooltip provides detailed information about participants to help us interpret their behaviour.

3.4 Social network matrix

We use re-orderable matrices [2] to show the connectivity between participants, whose rows align with the timeline rows. We use three measures of social network activity: direct calls between participants (as a matrix), calls to/from mutual contacts of participants (as a matrix) and all calls made by each participant (as a barchart). These are summarised as number of calls ((Fig. 5, left), number of contacts (Fig. 8) and average call length and can be filtered by call type. Colours can be interactively rescaled to match the value-range of interest.

Fig. 5 (left) shows that direct calls/texts between participants are sparse and that most participants only have direct contact with one other participant. Calls between mutual

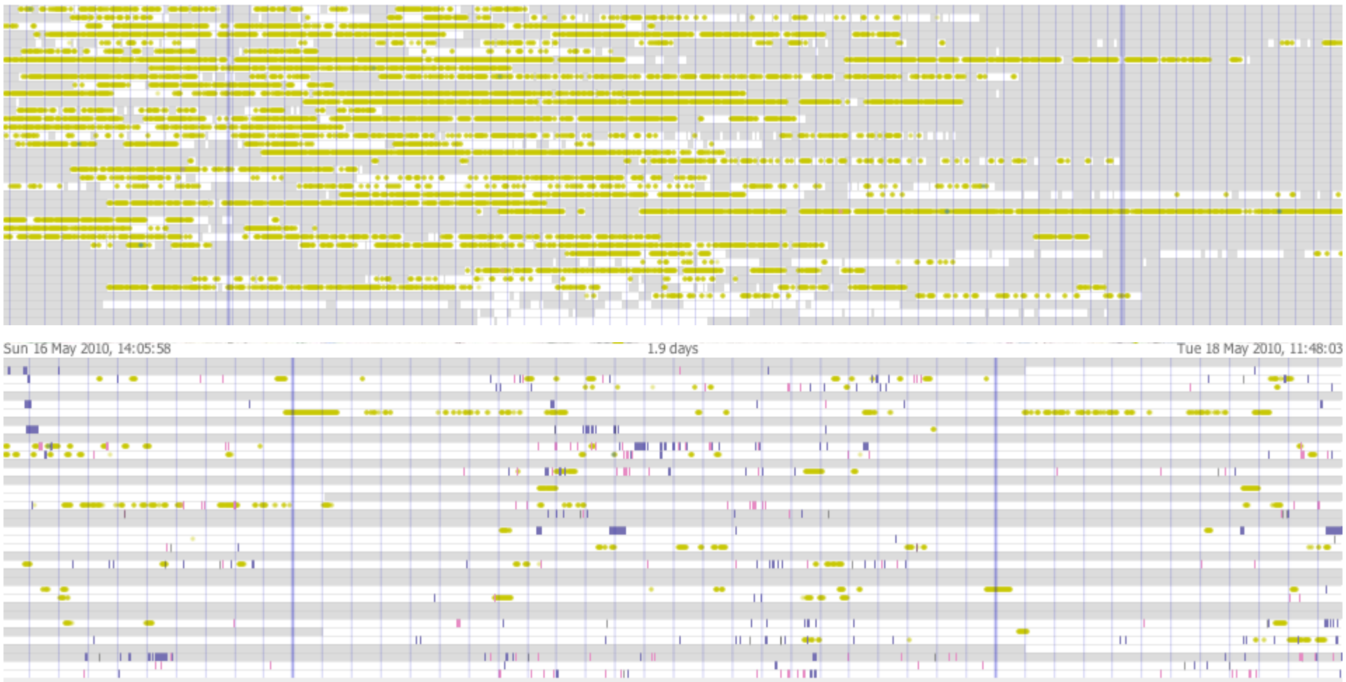

Figure 4: ***Top:*** **Timeline showing user participation (white) and GPS points (yellow).** ***Bottom:*** **Zooming-in (thin vertical lines represent hours) and superimposing calls/texts shows gaps in the GPS record and the associated difficulty in georeferencing calls.**

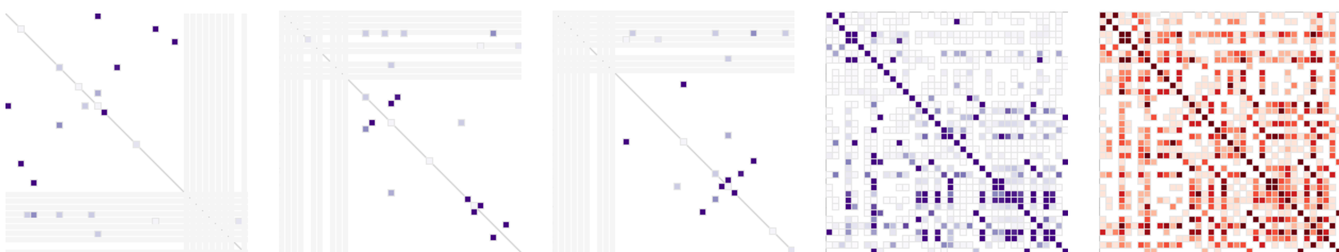

Figure 5: Social network matrices. ***From left to right:*** **Sorted by gender and coloured by number of calls/texts; sorted by age and coloured by number of calls/texts; sorted by social activity and coloured by number of calls/texts; sorted by social activity and coloured by calls to mutual contacts; sorted by social activity and coloured by number of mutual contacts.**

contacts are much more numerous (Fig. 5, right) and reveal a more nuanced set of social relations.

Sorting participants by their characteristics in the participants' view, also sorts timeline rows and the columns and rows of the matrix, helping us identify broad usage patterns. Sorting by gender in Fig. 5 (left) shows that of the direct calls to participants, most are between a female and male participant. The second matrix shows they are of a similar age and that most calls are between the youngest participants and the oldest participants. These contact patterns suggest that these pairs of participants are couples.

The third matrix in Fig. 5 is sorted by social activity and suggests that more calls are made between more socially-active participant. This pattern is also broadly reflected in the fourth matrix of calls to mutual friends. The fifth matrix broadly shows that more socially-active participants share more mutual contacts.

The matrix view also serves as a means to highlight calls made between pairs of participants or mutual friends. A participant who calls himself is apparent in Fig. 5 (left; on the diagonal line; participant 60). Selecting this cell highlights the calls on the map and timeline. Full call details are available as a tooltip. None were georeferenced, but they occurred throughout the participation period as either incoming text messages or outgoing calls. We speculate that this participant may have been using voice and text messaging as reminders.

Since call logs report both sent and received calls, and since failed calls/texts were also recorded, one would expect the matrix of calls between participants to be symmetrical. Fig. 6 shows that this is often not the case. Participant 123 only has eleven calls to/from participant 63 logged (indi-

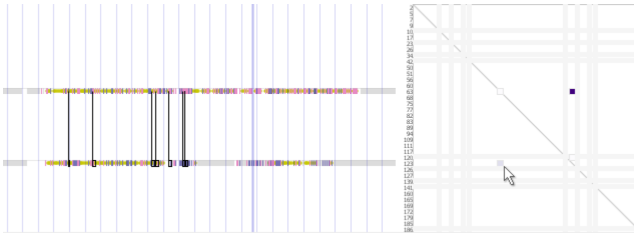

Figure 6: Asymmetry in the number of calls between participants 63 and 123.

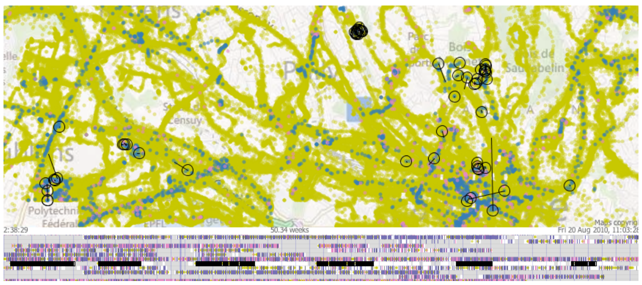

Figure 7: All participant 75's calls are highlighted.

cated with the cursor) compared to the 262 logged by participant 63. A contributing reason for this is that participant 123 had a break from the experiment (indicated by the grey in the timeline), but this does not explain all of the unaccounted calls.

3.5 Use of the linked views

Most of the visual analysis relies on the filtering, brushing and inspection of all four views. Selecting all participant 75's calls shows that they were made around the Lausanne area but in distinctive three/four week blocks (Fig. 7). White areas outside these blocks indicate the participant was participating in the study, but no GPS or call information was provided. He may regularly go away and leave his phone turned off.

We noted in the introduction that phone calls are only likely to represent a small part of a social network. Individuals who are regularly co-located, for example, will have a significant amount of face-to-face contact and perhaps less need to call. Looking at the number of mutual contacts was interesting in this regard, because it indicates a wider network of connectivity.

Our attention was drawn to participants 17 and 56 whom have an unusually large number of mutual contacts. Since they do not call each other particularly often, they were not easily identifiable within the direct social network view and so we had no particular reason to study their calling behaviour. By brushing on the timeline and map, we established that they tend to visit the same places at the same time. Fig. 8 shows the timeline zoomed to the time they both travelled to Zurich together, arriving in the late evening and leaving in the early hours of the morning. Intriguingly, they appear to be travelling one hour apart on both outward and inward journeys. Investigating the GPS logs revealed that for the first few days of the participant 17's logged data, the

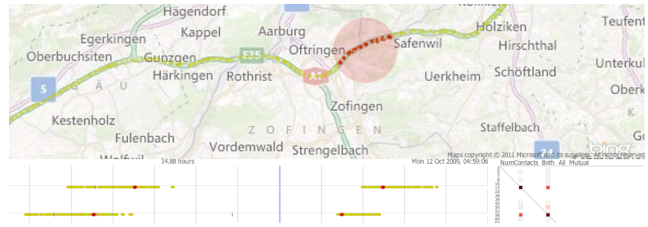

Figure 8: Two participants going to and from Zurich, an hour apart.

timezone was wrong by an hour, before being subsequently corrected. We speculate that this may have been after returning from a trip to a country in a different timezone.

4. FUTURE WORK

We used call and text logs to infer social network information and found that calls to mutual friends was a useful way of associating individual with each other, probably because individuals additionally communicate by other means, particularly when co-located. Co-location was difficult to determine from the patchy GPS data and we would also like try to establish types of social networks with proximity information supplied by the Bluetooth, WiFi and GSM logs which work within buildings.

We would also like to work with a larger and more representative sample in order to establish more general behavioural trends and to investigate the size of sample or sampling strategy that we would require.

As acknowledged in the introduction, we realise that mobile call records represent only a part of a social network. We would like to investigate incorporating other social network information into our spatial and temporal views to help answer our research questions.

5. CONCLUSION

Our aim was to design and use interactive visualisation to characterise the spatial and temporal nature of participants' social networks. Through our iterative prototyping approach, we produced novel and powerful interactive graphics that use coordinated linked views to help reveal spatial and temporal patterns in the social networks of this study. We believe we have succeeded in our aim to a large extent and we have ideas for exploring this further.

Our first research question asked how the spatial and temporal nature of social networks between participants might be characterised. With this relatively small sample of individuals, it was appropriate to provide individual characteristics of participants to help us explain some of the patterns we identified. This is not immediately scalable to larger studies, but was appropriate in this case. However, sorting social network matrices by characteristics of participants did enable us to draw out more general patterns, something that would be more successful with a greater number of users. We only considered three aspects of individuals in this work, but believe that looking at other aspects of individuals is important in this regard.

Our second research question asked whether linking spatial, temporal and call connectivity patterns could help us explain how participants construct their social networks. Coordinated linked views certainly allowed us to identify and describe spatiotemporal aspects of call connectivity and details of participants did help us explain some of the patterns. However, we did not have enough information to explain some of the patterns or artefacts we saw. For example, asymmetries in some of the call matrices, whether the lack of GPS logging was significant and the extent to which certain devices misreported their timestamps. However, we believe that our techniques can help draw attention to such inconsistencies and will be beneficial to those who know the data well.

Finally, we asked to what extent smartphone device logs help us understand social communication behaviour. Since communication does not only occur through mobile phone calls and texts, clearly we cannot use these alone to learn about social communication. Many more people use electronic means to communicate than a decade ago and this is likely to increase. Email and web-based social network contacts could be logged and analysed in this way. The social networks we constructed through mutual contacts is a novel attempt to identify another dimension to the social network and one that we found useful.

This work demonstrates that by combining and facilitating comparison of diverse attribute-rich data, visual analytics can effectively help provide insight into the behaviour of social communication.

6. REFERENCES

[1] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M.-J. Kraak, H. Schumann, and C. Tominski. Space, time and visual analytics. *Int. J. Geogr. Inf. Sci.*, 24(10):1577–1600, Oct. 2010.

[2] J. Bertin. *Semiology of Graphics*. ESRI Press, 2010.

[3] C. Brewer and M. Harrower. Colorbrewer: Color advice for maps. `http://colorbrewer2.org`.

[4] G. Chittaranjan, J. Blom, and D. Gatica-Perez. Who's who with big-five: Analyzing and classifying personality traits with smartphones. In *Proceedings of the 2011 15th Annual International Symposium on Wearable Computers*, ISWC '11, pages 29–36, Washington, DC, USA, 2011. IEEE Computer Society.

[5] J. A. Dykes and D. M. Mountain. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Comput. Stat. Data Anal.*, 43(4):581–603, Aug. 2003.

[6] B. Fry. *Visualizing Data*. Cambridge University Press, 2007.

[7] B. Fry and C. Reas. Processing. `http://processing.org/`.

[8] F. Girardin, J. Blat, and C. Ratti. Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing Magazine*, pages 36–43, 2008.

[9] T. Kim, J. Blom, and J. Stasko. Exploring Complex Mobile Life through Lightweight Visualizations. pages 37–40, Bergen, Norway, 2011. Eurographics Association.

[10] L. C. Koh, A. Slingsby, J. Dykes, and T. S. Kam. Developing and applying a user-centered model for the design and implementation of information visualization tools. In *Proceedings of the 2011 15th International Conference on Information Visualisation*, IV '11, pages 90–95, Washington, DC, USA, 2011. IEEE Computer Society.

[11] M. J. Kraak. The space-time cube revisited from a geovisualization perspective. *Proceedings of the 21st International Cartographic Conference*, 1995, 1988.

[12] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data for mobile computing research. In *Proceedings Mobile Data Challenge by Nokia Workshop, International Conference on Pervasive Computing*, 2012.

[13] G. G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. T. Stasko. Effectiveness of animation in trend visualization. *IEEE Trans. Vis. Comput. Graph.*, 14(6):1325–1332, 2008.

[14] A. Slingsby. Supporting the visual analysis of the behaviour of gulls. `http://bit.ly/HLZcFo`.

[15] M. Theus. Interactive data visualization using Mondrian. *Journal of Statistical Software*, 7(11):1–9, 2003.

[16] J. Thomas and K. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, 2005.

[17] J. Wood, A. Slingsby, and J. Dykes. giCentreUtils. `http://gicentre.org/utils/`.

[18] J. Wood, A. Slingsby, and J. Dykes. Designing visual analytics systems for disease spread and evolution: Vast 2010 mini challenge 2 and 3 award: Good overall design and analysis. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 285 – 286. IEEE, 2010.