

Social Transport with Urban Big Data (STUBD)

Piyushimita (Vonu) Thakuriah, PI, Professor, University of Glasgow, UK

Summary: STUBD is a career integration project which investigates the potential of emerging forms of urban data to support Public, Intermodal, Community, Active and Shared (PICAS) passenger transport. The central focus of the work is on social exclusion and innovative ways to improve mobility outcomes for socially excluded populations in urban areas. STUBD is shaping a critical mass of research focused on understanding the social aspects of urban mobility problems using urban Big Data. The work being undertaken as a part of STUBD provides a strong case for greater exploration of the technological, methodological, epistemological and political economy challenges that arise with using novel forms of data to understand complex urban challenges (Thakuriah, et al, 2017a).

Major research questions: STUBD address two major questions: (1) how do emerging forms of urban data and associated analytics help with a more comprehensive understanding of transportation disadvantage and the multimodal transportation solutions that are needed; and (2) what are the emerging business and collaboration models around big data and urban informatics that support alternatives to the private car?

Main findings: Our key findings are that crime and perceptions of traffic hazards can deter the use of PICAS passenger transport, particularly in disadvantaged neighbourhoods. This contributes to the “last-mile” problem in public transportation use, which is a deterrent to sustainable transport policies (Tilahun, 2016). Barriers to mobility solutions may exist well beyond lack of availability and accessibility of physical services, and may extend to low quality of community relations. For example, lack of trust, language barriers, and contested relationships with law enforcement may be leading to underreporting of minor traffic crashes and related events to authorities, hindering the availability of complete governmental data to fully understand how traffic and crime hazards might be affecting travel behaviour and quality of life, and preventing appropriate governmental planning and investment actions from being taken (Davide-Paule, et al, forthcoming 2018). Instead of reporting to law enforcement, citizens in the same neighbourhoods may be speaking out about these events through social media – we detect this using massive volumes of specially geolocalized social media (Twitter) data (Davide-Paule, et al, 2017).

We also find that there is considerable geographical overlap in where crime occurs and where traffic crashes occur, as many types of traffic deviance leading to crashes are not random, but has a root cause the same social conditions that result in concentrations of crime. Using Bayesian Model Averaging to understand information uncertainty, and Spatial Autoregressive Quantile models to model relationships, we find that the extent of local economic activity, the quality and quantity of public transportation availability and accessibility, and the degree of social media sentiment that are neutral, lead to lower predicted levels of crashes and crimes (Thakuriah and Sun, 2017b). Our conclusion is that having a joined-up approach to crime and crash reduction would lead to better location-based micro-place operational strategies.

Finally, the project found that a large number of non-traditional stakeholders are increasingly involved in providing city services and information, including mobility services (Thakuriah, et al, 2017c). Using machine learning models on unstructured data collected from webpages, our analysis identified “digital urban infomediaries” who are: general-purpose Information and Communication Technology (ICT) providers, urban information service providers, open and civic data infomediaries, and independent and open source developers. We further identified that informal networks consisting of independent developers, data scientists and civic hackers are playing an important role in city and mobility services, with implications for management and governance of these diverse groups.

Work carried out and impact and scientific impact: STUBD has helped advance knowledge relating to acquisition, linkage and curation of complex urban transportation data. One area where the project has had an impact is in using multiple sources of structured and unstructured data to jointly answer complex questions. One particular application has been to link GPS, social media (Twitter and Foursquare) and travel survey data to understand mobility and resource consumption patterns in urban areas.

However, the social media, sensor and related data can be very noisy. Hence considerable effort has gone into developing methods to derive higher value from the noisy data. One example of this is geolocating non-geotagged Tweets with the goal of significantly enhancing the Twitter data that can be used to pinpoint the location of urban events and transport incidents. Locating traffic events extracted from Twitter is crucial for transportation authorities to respond and manage the transportation system. For this reason, geographical information in Twitter data should be exploited in detection techniques. However, only a very small sample of the tweets contain geographical information in its metadata. Part of this project resulted in greater geolocated Twitter samples through location prediction of non-geotagged tweets. Another area has been in understanding information uncertainty when dealing with new forms of data towards inference. There are considerable biases and challenges to representativeness leading to uncertainty about inference regarding what are optimal predictors of traffic crashes and crime and the role of public transportation and area sentiments as detected through social media data. Our approach is to utilise Bayesian Model Averaging when used jointly with a large set of other spatial, behavioural and economic predictors. Finally, we are most interested in “most deprived” areas, in contrast to areas experiencing average levels of deprivation – and for that reason, we have experimented with how all the data generated as a part of the project can be utilized to explain the “upper tail” of the distribution of social hazards facing citizens, while taking into account spatial dependence and other considerations – hence we utilize Spatial Autoregressive Quantile models. Hence the project has addressed a multitude of methodological considerations in urban informatics.

Target groups: The target groups for this work are regional planning agencies, local councils, public transportation agencies, Mobility-as-a-Service community, traffic engineering and traffic safety divisions, crime prevention units, transportation sharing companies, smart cities companies, and companies involved in urban analytics and big data.

Conclusions: STUBD brings together an innovative mix of theory-driven and data-driven approaches to understanding the multiple dimensions of transportation disadvantage, making a substantial contribution towards crystallizing fit-for-purpose PICAS solutions in a wide variety of urban contexts, and better understanding of the civic and business models that are necessary for successful implementation.

References

- Thakuriah, P., N. Tilahun and M. Zellner (2017a). *Big Data and Urban Informatics: Innovations and Challenges to Urban Planning and Knowledge Discovery*. In *Seeing Cities through Big Data: Research, Methods and Applications in Urban Informatics*, Springer, NY, pp. 11-48.
- Tilahun, N., Thakuriah, P., Li, M., and Keita, Y. (2016) *Transit use and the work commute: analyzing the role of last mile issues*. In *Journal of Transport Geography*, Vol 54, pp. 359–368.
- Paule, J. D. G., Y. Sun and P. Thakuriah (forthcoming 2018). *Beyond Geo-Tagged Tweets: Exploring the Geo-Localization of Tweets for Transportation Applications*. In *Transportation Analytics in the Era of Big Data*, Editors: Ukkusuri, Satish, Yang, Chao, to be published by Springer. ISBN 978-3-319-75862-6
- Paule, J. D. G., Y. Moshfeghi, J. Jose and P. Thakuriah (2017). *On Fine-Grained Geo-Localization of Tweets*. Proc ACM SIGIR conference, Amsterdam, Netherlands, 2017 (ICTIR'17).
- Thakuriah, P. and Y. Sun (2017b). Integrating Heterogeneous Sources of Data to Estimate Composite Social Hazards. Paper presented in the Associate Collegiate Schools of Planning (ACSP) Annual Conference, Denver, CO, Oct, 2017. Under revision for submission to journal for review.
- Thakuriah, P., L. Dirks, and Y. Keita Mallon (2017c). *Digital Infomediaries and Civic Hacking in Emerging Urban Data Initiatives*. In *Seeing Cities through Big Data: Research, Methods and Applications in Urban Informatics*, Springer, NY, 189-208.