FINAL TECHNICAL REPORT / RAPPORT TECHNIQUE FINAL

FINAL TECHNICAL REPORT: LEVERAGING MOBILE NETWORK BIG DATA FOR DEVELOPMENTAL POLICY

;

;

IDRC Grant / Subvention du CRDI: 108008-001-Leveraging Mobile Network Big Data for Developmental Policy

Leveraging Mobile Network Big Data for Developmental Policy

IDRC Project Number: 108008-001

Report Type and #: Final Technical Report

Period covered by the report: 13th March 2015 – 31st December 2017

Submitted to IDRC by:

LIRNEasia
Pro-poor. Pro-market.
12 Balcombe Place, Colombo 00800, Sri Lanka
Tel: +94 11 267 1160
Fax: +94 11 267 5212
E-mail: info@lirneasia.net

*LIRNE*asia* is a regional ICT policy and regulation think tank active across the Asia Pacific. It is incorporated as a non-profit company under Sri Lankan law.

Country / Region: Sri Lanka, Bangladesh

Full Name of Research Institution: LIRNEasia

Address of Research Institution: 12 Balcombe Place, Colombo 00800, Sri Lanka

Research team: Helani Galpaya (Chief Executive Operator, LIRNEasia); Rohan Samarajiva, PhD (Chair, LIRNEasia; Project Leader); Sriganesh Lokanathan (Team Leader- Big Data Research, LIRNEasia); Shazna Zuhyle (Research Manager, LIRNEasia); Danaja Maldeniya (Senior Researcher, LIRNEasia); Kaushalya Madhawa (Researcher, LIRNEasia); Madhushi Bandara (Junior Researcher, LIRNEasia); Isuru Jayasooriya (Researcher, LIRNEasia); CD Athuraliya (Researcher, LIRNEasia); Lasantha Fernando (Researcher, LIRNEasia); Keshan De Silva (Researcher, LIRNEasia); Robert Galyean (Researcher, LIRNEasia); Thavisha Perera-Gomez (Research Manager, LIRNEasia); Yudhanjaya Wijeratna (Junior Researcher, LIRNEasia); Laleema Senanayake (Researcher, LIRNEasia); Suthaharan Perampalam (Senior Researcher, LIRNEasia); Moinul Zaber, PhD (Research Fellow, LIRNEasia & Assistant Professor, Department of Computer Science & Engineering, University of Dhaka); Amal Kumarage, PhD (Senior Professor, Department of Transport & Logistics Management, University of Moratuwa); Amal Shehan Perera, PhD (Senior Lecturer, Department of Computer Science and Engineering, Faculty of Engineering, University of Moratuwa); Chathura De Silva (Lecturer, Department of Town & Country Planning, University of Moratuwa); Joshua Blumenstock, PhD (Assistant Professor, School of Information, University of California, Berkeley); Dedunu Dhananjaya (Research Fellow, LIRNEasia); Sandun Siriwardhana (Junior Researcher, LIRNEasia); Ovini Dantanarayana (Junior Researcher, LIRNEasia); Dinusha Amarathunga (Intern); Dushan Anuradha (Intern); Bhagya Jeewanthi (Research Assistant); Gabriel Kreindler (Research Associate, LIRNEasia); Yuhei Miyauchi (Research Associate, LIRNEasia); Dharshana Kasthurirathna, PhD (Research Fellow, LIRNEasia & Senior Lecturer, Sri Lanka Institute of Information Technology); Aparna Surendra (Graduate Research Associate, LIRNEasia); Amin Ahsan Ali, PhD (Associate Professor, University of Dhaka); Abu Seyed (M.Phil Student, University of Dhaka); Samiul Islam (MSc Student, University of Dhaka); Nibras ar Rakib (MSc Student, University of Dhaka); Nahid Newaz (MSc Student, University of Dhaka); Md Samiul Islam (MSc Student, University of Dhaka); Shakil Ahmed (MSc Student, University of Dhaka); Dipanka Chaki (MSc Student, University of Dhaka); Ayumi Arai, PhD (Research Fellow, LIRNEasia & Project Researcher, Center for Spatial Information Science, University of Tokyo, Japan)

Contact Information: info@lirneasia.net (e), 0094112671160 (v), 0094112675212 (f)

Table of Contents

1. Synthesis 6
2. Research Problem and findings 7
   2.1. High Resolution Population Estimates 8
      2.1.1. Findings 8
      2.1.2. Challenges 9
   2.2. Human Mobility Analysis 10
      2.2.1. Understanding atypical mobility behavior: case study of Nallur festival 10
   2.3. Identifying impoverished communities 12
      2.3.1. Findings 13
      2.3.2. Challenges 15
   2.4. Land-use extensions 15
      2.4.1. Findings 16
      2.4.2. Challenges 22
   2.5. Geographical boundaries of human mobility 23
   2.6. Comparing geographical boundaries defined by human mobility and communication 24
   2.7. CCTV Analyses 25
      2.7.1. Findings 26
      2.7.2. Challenges 28
   2.8. Behavioral patterns associated with the opening of the Expressway 29
      2.8.1. Findings 30
      2.8.2. Challenges and planned work 30
   2.9. Forecasting the propagation of Dengue using high resolution human mobility data 30
      2.9.1. Findings 31
      2.9.2. Challenges 33
   2.10 Analyzing Facebook data to understand international connectivity 34
   2.11 Alleviating Harms 37
      2.11.1 Broader issues of MNBD including lawful interception 38
      2.11.2 Challenges 39
3. Project implementation 39
   3.1. Collaborations 39
      3.1.1. University of Moratuwa 39
      3.1.2. University of Sri Jayewardenepura 39

3.1.3. Various Government Departments in Sri Lanka 40
3.1.4. University of Dhaka 40
3.1.5. School of Information, U.C. Berkeley 40
3.1.6. Fields of View 40
3.1.7. University of Tokyo 40
3.2. LIRNEasia Data Analytics Advisory Council 41
3.3. Electricity big data in Dhaka and Colombo 42
3.4. Building trust and consensus in policymaking through open data: an action research study on electoral district demarcations in Sri Lanka 43
3.4.1. Observations 45
3.4.1.1. Inclusiveness 45
3.4.1.2. Trust 45
3.4.1.3. Consensus 45
3.5. Tanzania 46
3.6. Challenges 46
3.6.1. Human resources 46
3.6.2. Setting up Data Analytics in Dhaka, Bangladesh 47
4. Project outputs and dissemination 48
4.1. Reports and publications 48
4.2. Media coverage 50
4.3. Conferences, Meetings, and Forums 54
4.4. Capacity Building 65
5. Impact 66
5.1. Ministry of Megapolis & Western Development, Urban Development Authority, and Sri Lanka Strategic Cities Development Project 66
5.2. Ministry of National Policies and Economic Affairs 67
5.3. Department of Census and Statistics 67
5.4. Epidemiology Unit, Ministry of Health 67
5.5. University of Moratuwa 67
5.6. 2030 Vision of Sri Lanka 68
5.7. Access to Information (A2I) Unit of the Government of Bangladesh 68
5.8. Other 68
6. Recommendations 68

Table of Figures

Figure 1: Process of leveraging mobile network big data for developmental purposes.....7
Figure 2:Comparison with DSD census population.....8
Figure 3:Census population density and resident mobile user density at different resolutions.....9
Figure 4:Commuting patterns to and from the Peliyagoda area.....10
Figure 5:Daily percentage variation of visitors compared to 5-month daily average.....11
Figure 6:Main source location of visitors to Jaffna during the Nallur festival.....11
Figure 7:Visitor percentage in each DSD over the Nallur festival period.....12
Figure 8:Daily visitor percentage in each DSD over the Nallur festival period.....12
Figure 9:Comparison of social CDR features with census attributes.....14
Figure 10: Adjusted R2 of linear regression models built for census features using different CDR feature categories.....15
Figure 11:Land use in the Colombo district in terms of commercial behavior.....16
Figure 12:Geographical distribution of the identified land use behaviors in the Colombo district.....17
Figure 13:Results of clustering based classification for land cover estimation.....18
Figure 14:Final result for reference image tiles from 5 AOI in the order of 1-5 row-wise. Left column is image tiles, middle column is built-up area segmentations with binary morphology, right column is final result after rotated bounding rectangles.....20
Figure 15:Results for different areas of Sri Lanka: Colombo and Gampaha.....21
Figure 16:Estimated functional boundary with connections to neighboring towns.....22
Figure 17:Hierarchical mobility zones.....23
Figure 18:Mobility zones in the Colombo district.....24
Figure 19: (Left) Cohesive regions of overlap; (right) Administrative boundaries and capital cities of Districts.....25
Figure 20:Motion based vehicle detection from CCTV footage.....26
Figure 21:Vehicle detection based on learning approaches.....26
Figure 22:Counting vehicles at each frame.....28
Figure 23:Colombo – Katunayake (E03) Expressway.....29
Figure 24:Prediction of weekly disease incidence in the Colombo MOH area in 2014.....32
Figure 25:Comparison of model performance.....32
Figure 26:Comparison of model performance for final analysis.....33
Figure 27:Gross bilateral migration from 2013 versus friend links between country pairs.....35
Figure 28:Gross bilateral trade from 2014 versus friend links between country pairs.....35
Figure 29:Redistributing DSDs on OpenDemarcations.lk.....44

1. Synthesis

Through this project LIRNEasia extended its pioneering research based on mobile network big data. New sources of data were leveraged in addition to mobile network big data, including Close Circuit Television (CCTV), satellite imagery, electricity data, crowd-sourced data, as well as "small" data from government agencies and others. In addition to extending the work on urban and transportation aspects, additional lines of research were opened. LIRNEasia undertook work to model the spread of infectious diseases that moved forward as a collaboration between LIRNEasia, the University of Moratuwa, and the Epidemiology Unit of the Ministry of Health. In addition LIRNEasia developed more fine-grained and more frequent indicators of economic activity and socio-economic wellbeing of regions. Additional research was conducted to extend some of the research by calibrating both the socio-economic mapping work as well as some aspects of the transportation research through surveys conducted in collaboration with the University of Tokyo. This collaborative work with the University of Tokyo is being funded through a grant from the Global Partnership on Sustainable Development Data (GPSDD). Some preliminary research was also conducted on electricity consumption big data both in Sri Lanka as well as in Bangladesh.

The central problem addressed by this research grant is that of how to realize the potential of big data to better inform public policy and allow for greater use of evidence in the policy making process. LIRNEasia has had recent success in providing evidence drawn from our big data analyses into the policy making process. Principally, the insights developed by LIRNEasia were used by the Western Region Megapolis Planning Project (a major focus of development for the new Sri Lankan government). The engagement has continued during the implementation stages, through adhoc formal and informal requests for insights and/or advise from various government agencies involved in the implementation work. LIRNEasia has spent considerable time in enlightening policy makers and broadening the symbolic environment and now when government agencies consider utilizing new data sources in their work, they do reach out to LIRNEasia for advice as well.

In addition to the generation of actionable insights, LIRNEasia is developing capacity in local universities and among researchers. So far 30 undergraduate students (7 were women) have conducted their thesis through LIRNEasia. Two masters students (1 woman) conducted their research based masters through LIRNEasia. LIRNEasia is also enabling policy makers and government officials to become informed and discerning consumers of big data insights. This engagement is ongoing and will hopefully lead to more significant capacity development amongst the technical staff at relevant government organizations.

LIRNEasia has also extended the policy relevant research to capitalize on a policy window in relation to electoral district demarcations in Sri Lanka. LIRNEasia has developed an interactive tool for stakeholders involved the process to propose electoral boundaries. An open and transparent process in electoral reforms in Sri Lanka can be a beacon of good practice for the rest of Asia. Several small parties utilized the tool when submitting their proposals for provincial council boundaries to the the Delimiting Committee. The tool can also be utilized in relation the parliamentary electorate boundaries. However the Delimiting Committee has yet to call for proposals in relation to the parliamentary elections, which will be held only in 2020. Hence we expect the full impact of this tool will only be realized in the coming years.

In addition LIRNEasia extended the big data research Bangladesh through this grant. The Bangladesh research has had significant challenges working in a difficult political environment This meant there were delays in getting the research off the ground and furthermore required a reorientation from mobile network big data to electricity big data. While the success of the Bangladesh research is limited, it did result LIRNEasia helping to start a data analytics lab at the University of Dhaka, which continues the research work with LIRNEasia's assistance.

LIRNEasia also translated the draft guidelines for mobile operators to responsibly share their data (developed previously) into legal language. LIRNEasia continues to contribute to the global dialogue on leveraging new data sources (and in particular big data) for developmental purposes and is frequently invited both locally as well as globally to share its experiences and thoughts on the opportunities and challenges. LIRNEasia has been recognized for conducting pioneering research (most recently by the ITU). LIRNEasia's Chair has been renewed for a second term in the UN Global Pulse's Privacy Advisory Group. Similarly LIRNEasia's CEO was appointed to the board of the Global Partnership on Sustainable Development Data. LIRNEasia's Team Leader for Big Data Research was appointed to the board of Sri Lanka's first Innovation Lab, and is an Advisor to the Analytics Unit of the Ministry of National Policies and Economic Affairs.

2. Research Problem and findings

Having up-to-date and reliable data is essential for developing economies. Yet extant data collection infrastructures are inadequate. Existing data is often collected through costly infrequent surveys with insights coming well after the data was originally collected. Previous research undertaken by LIRNEasia shows that the analysis of Call Detail Records (CDRs) has the potential to provide valuable insights for urban and transport planning. The ability to extract not just mobility, but also social, and consumption behaviors from citizens' usage of their mobile phones also affords opportunities for areas beyond just urban and transportation planning (see Figure 1).

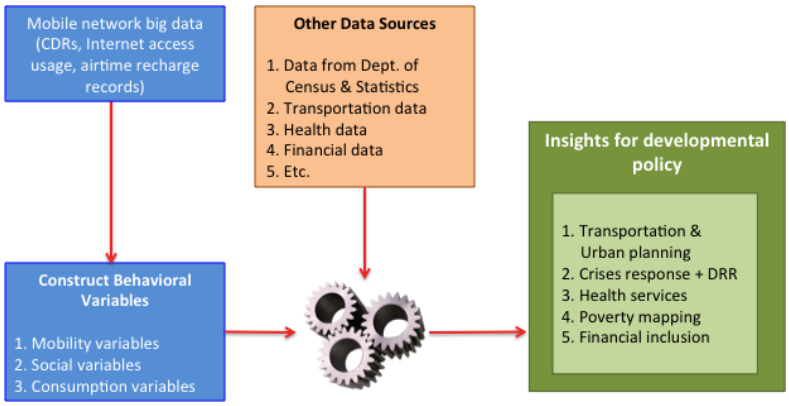


Figure 1: Process of leveraging mobile network big data for developmental purposes

In addition to more detailed research in the area of transportation and urban planning, this research articulate and answers questions in other domains such as health (modeling the spread of diseases) as well as official statistics (mapping poverty, etc.). Furthermore the research

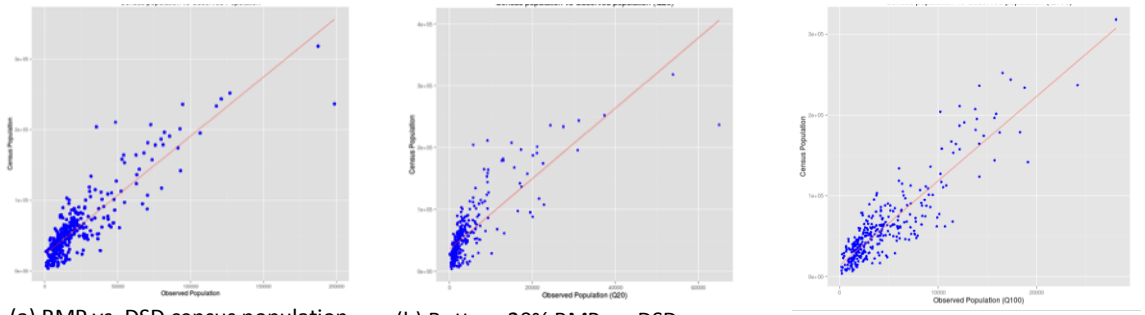
attempts to leverage additional data such as Close Circuit Television (CCTV), satellite imagery, electricity consumption and supply.

2.1. High Resolution Population Estimates

Previous work by LIRNEasia estimated populations at the Divisional Secretariat Division (DSD, 3rd level administrative division) level. The density of base stations particularly in urban areas is sufficiently high to provide significantly higher spatial resolution of populations. We acknowledge that the operational characteristics of mobile networks may introduce an additional degree of uncertainty about the location of a mobile phone user. For example, mobile networks dynamically alter the radius of the cells for load management. As a result, the mobile phone may connect to different base stations although the user location has not changed; i.e. there can be instances where the user is stationary but the active base station has changed leading to uncertainty about the user's real location. Nonetheless, with appropriate methodologies it has been proven to be a sound gauge to develop high-resolution population maps. Our research set out to provide insights into the spatial distribution of populations in Sri Lanka which is a vital pre-requisite in monitoring population growth, transportation planning, and disaster management etc.

2.1.1. Findings

A linear model of the resident mobile population (RMP) and the census population at the DSD level considering all users for whom a home location was identified had an R^2 value of 0.77. In contrast a linear model of the same quantities using the lowest 20 percent of the users considering the confidence had an R^2 of 0.65 while a model using the top 20 percent of the users had an R^2 of 0.8 (Figure 2).



(a) RMP vs. DSD census population

(b) Bottom 20% RMP vs. DSD census population

(c) Top 20% RMP vs. DSD census population

Figure 2:Comparison with DSD census population

The new high-resolution population estimates demonstrate greater alignment with the census figures at the DSD level than the estimates from original approach (R^2 of 0.77 compared to 0.59 with the original method). The proposed reliability measure has been proven to be effective. The use of subscribers with higher reliability for the home location estimate leads to be progressively higher alignment with census population.

These results were further improved by considering the supra-linear relationships between mobile phone usage and population. Previous work (Gomez-Lievano, Youn, & Bettencourt,

2012; Schläpfer et al., 2014) has shown that as the population density of a region increases the mobile phone use increases exponentially. This has implications for predicting population or population density using mobile phone network data. Recent work has introduced a log-linear model that captures the super linearity of the relationship (Deville et al., 2014). This approach was adapted to our data and compared with the predictive capability of a standard linear model. The log-linear model explains 93 percent of the variability of population density at the DSD level while the standard linear model was able to explain only 84 percent. Figure 3 shows comparison between DSD level population densities derived from census population and the log linear model as well as resident mobile phone user densities at the base station coverage region level. The latter shows the increased spatial resolution possible with MNBD.

One particular line of research was to utilize mobile network big data for understanding congregations and 'foot' traffic. These would be beneficial for public policy questions such as the optimal locations for government offices or service locations. Unfortunately our research revealed that the type of mobile network big data we had access to did not provide a good basis for the understanding congregations since the data is sparse at an individual level.

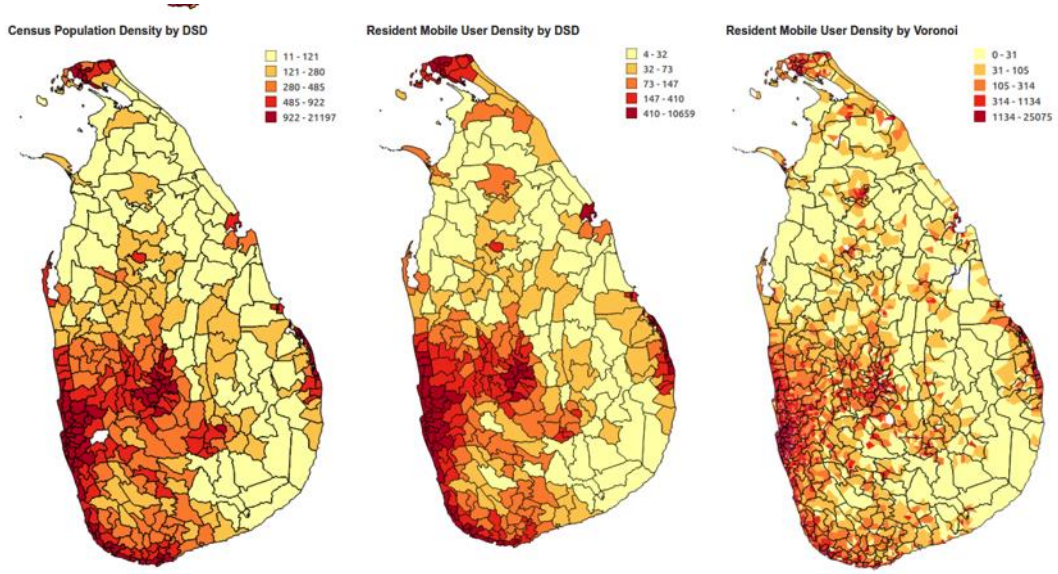


Figure 3:Census population density and resident mobile user density at different resolutions

2.1.2. Challenges

The primary challenge in extending this research beyond the current level was the unavailability, difficulty of access and high cost of additional data that is required. Potential data sources such as high-resolution satellite imagery are expensive while alternative free data sources such as OpenStreetMap building and land use layers are yet to reach a satisfactory level of maturity, standardization or completion in Sri Lanka. Furthermore with the recent public release of high resolution population estimates for a few countries (including Sri Lanka) by Facebook (these were the results of a collaboration between Facebook, Columbia University, and the World Bank), make some aspects of our research redundant.[1] MNBD based estimates would be still of higher frequency, but the resolution is less than that of those produced by Facebook.

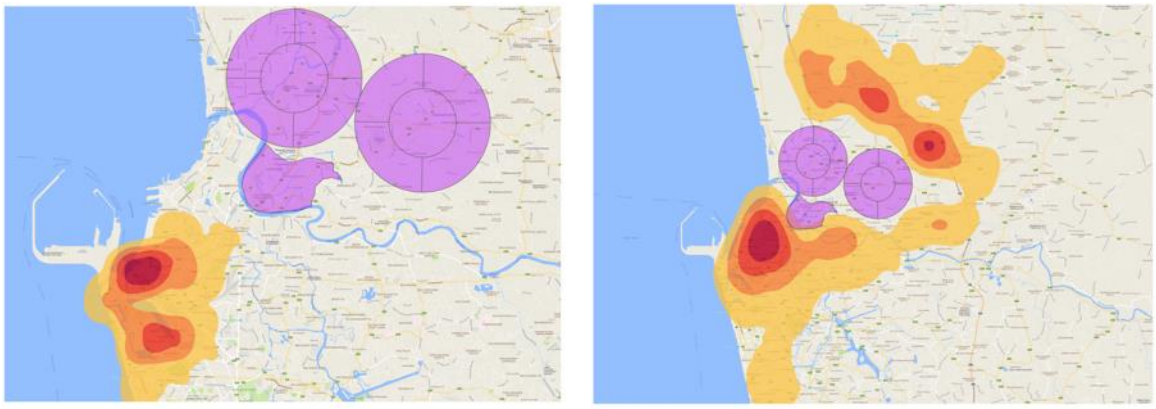
[1] See for example https://code.facebook.com/posts/596471193873876/open-population-datasets-and-open-challenges/

2.2. Human Mobility Analysis

A continuing line of opportunistic research involved identifying commuting patterns under different contexts and for different areas in Sri Lanka. These opportunities arose out of ad hoc requests that come out of our ongoing engagements with the Urban Development Authority (UDA) and the new constituted Western Region Megapolis Ministry.

Commuting patterns of people in the Western province of Sri Lanka were analyzed as requested by the UDA. The target area for this analysis was Peliyagoda and Ragama. The analysis was two-fold, of people who live in these areas vs. people who work in the said areas. The rationale behind this component was to support the planning phase of the Megapolis project where potentially hidden commuting patterns could be identified. Since the use of MNBD is a non-intrusive method that represents natural behaviors, results from such analysis is a good source to generate insights from.

Figure 4 serves as an example of one of the outputs the analysis revealed. The purple regions denote the area subject to the analysis (Peliyagoda and Ragama). The heat map was used to illustrate the main locations of destination for commuters living in the areas analyzed (a) and the residential locations of commuters who travel to these areas (b).



(a) Commuting from Peliyagoda **(b) Commuting into Peliyagoda**

Figure 4:Commuting patterns to and from the Peliyagoda area

The UDA used this information as evidence for their planning process in the next phase of the megapolis project that has been proposed to commence at beginning of 2017.

2.2.1. Understanding atypical mobility behavior: case study of Nallur festival

The Nallur festival takes place annually in Jaffna and is Sri Lanka's longest festival spanning 25 days in August. The influx of visitors for this festival creates a opportunity to understand atypical mobility behavior which is useful for planning purposes. The primary dataset used was a five-month sample of CDRs that covers the festival period as well as two months before and after. Figure 5 represents the visitors to Jaffna during this period, with the x-axis based on the date

range and the y-axis illustrating the percentage of change in comparison to the average count of visitors. The festival period is highlighted in red.

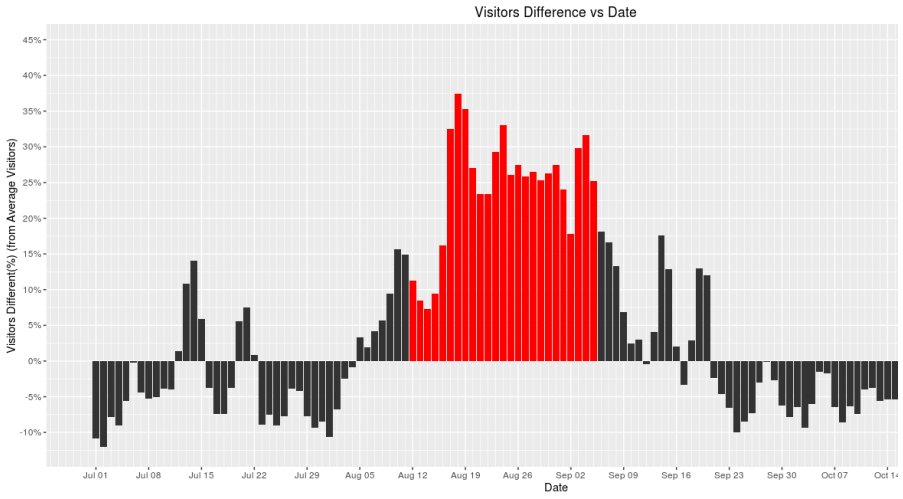


Figure 5:Daily percentage variation of visitors compared to 5-month daily average

As per our analysis there was 25% – 30% increase in visitors during the festival period. We were also able to deduce that the majority of the visitors came from Colombo, Kilinochchi and Vavuniya districts (Figure 6).

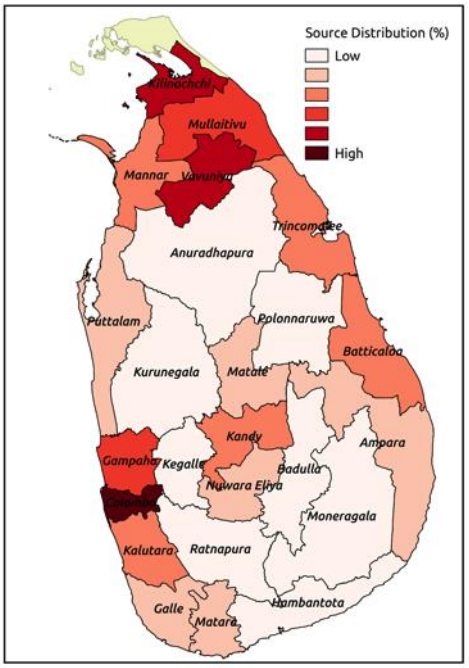


Figure 6:Main source location of visitors to Jaffna during the Nallur festival

We were also able to understand the most popular locations within Jaffna during the Nallur Festival with Jaffna town itself being by far the most popular location over the entire time

period of the festival (Figure 7) but Nallur DSD being the most popular location on the main day of the festival (Figure 8)

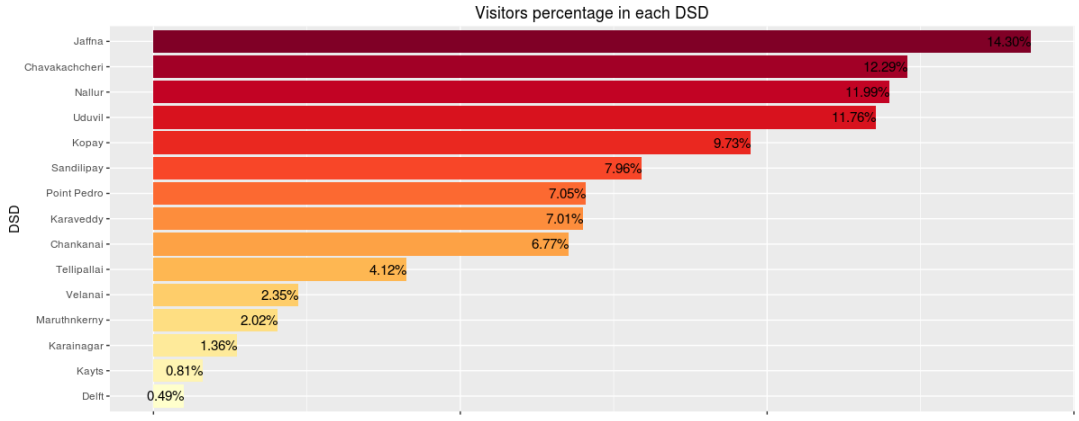


Figure 7:Visitor percentage in each DSD over the Nallur festival period

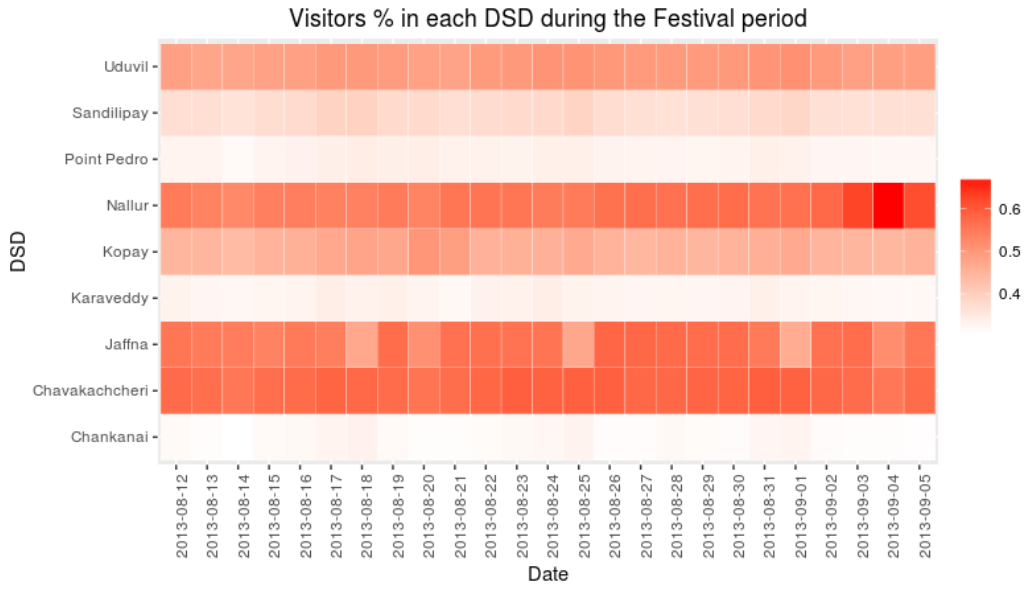


Figure 8:Daily visitor percentage in each DSD over the Nallur festival period

By providing an understanding of the spatio-temporal changes in population density, mobility, as well as the mobility connections during significant (and repeating) events the analyses provides insights of relevance to planning purposes for both local authorities as well as private sector services providers.

2.3. Identifying impoverished communities

Locating impoverished communities is a vital step in providing support swiftly and efficiently. Yet, at present the process is unrefined, involving a great deal of guesswork. Surveys that provide enough detail to accurately assess the situation are costly and can rarely reach a representative portion of the population Academic research has suggested the possibility of using MNBD as a proxy for socio-economic status, allowing fine-grained, real-time sensing of impoverished communities. However, identifying the relationships between users' economic status and usage patterns requires a significant amount of tuning to the Sri Lankan context, as patterns identified elsewhere are obfuscated by peculiarities of the market infrastructure and dynamics in Sri Lanka. For example, it has been hypothesized that high-income users tend to deposit large amounts of money to SIM cards less frequently, while low-income users deposit frequently in small amounts. However, the dominant mode of SIM deposits in SL, the scratch card, is only available in a limited number of denominations and obscures temporal patterns because available data reflects users' redemption of the cards, not their purchases.

2.3.1. Findings

Geographic dispersion of the economic aspects of CDR data were compared to poverty estimates on the DSD level published by the DCS and World Bank. The estimates are derived from the 2012 Census of Population and Housing, and from the 2012/13 Household Income and Expenditure Survey, and thus are only available for this time period. After identifying the strongest indicators in the CDR data, we attempted to formulate a parameter-free algorithm for real-time, high-resolution identification of impoverished areas.

In order to assess the ability of the consumption metrics to model poverty headcount, several regressions of the form:

$$Y_k = \beta + \sum_{\delta=1}^{n} \sum_{i=1}^{m} \beta_{i\delta} X_{ik}^{\delta} + \epsilon$$

Across all components it was seen that the models built on measures aggregated by standard deviation performed better (higher Pearson correlation coefficient) than those built on measures aggregated by mean. However, none of the models performed particularly well. The highest correlation with poverty was seen in the standard deviation of the fourth principal component scores (r = -0.397 for method 1 of mapping poverty to cells and r = -0.393 for method 2), falling well short of the correlation achieved by regressing the cell averages of users' mean machine refill (r = -0.575).

One potential reason for the poor performance of the countrywide models is due to the varying spatial representativity of the CDR dataset. Correcting for representativity bias in some areas however did not result in much improvement. More importantly though that a sole aggregate spatial indicator of poverty (i.e. poverty rate) and that too at a high spatial aggregation was insufficient to develop a robust model). Hence this line of inquiry was temporarily abandoned.

Subsequently a new approach was tried which has yielded promising initial results. The new approach has two principle research questions:

1. Do features generated from Sri Lankan Call Detail Records(CDR) correlate to socio-economic variables from the national census?
2. If so, which analytical models most accurately approximate census variables (from CDR features)?

The post-conflict region of Jaffna (Jaffna Province) was chosen for the initial analyses. Consumption features (number of calls made/ received/ total, duration of calls made/ received/ total, etc.) and social features (number of contacts, rate of contact, physical distance between contacts) were aggregated at the level of a Base Transceiver Station (BTS) and compared to features derived from the Census (which were then mapped to BTS coverage areas). The census features were categorized into different Current work focuses on the post-conflict region of the Northern Province. Early results suggest that social CDR features in particular (Figure 9) show strong statistically significant correlations to attributes in the census. In this current phase 42 census features were selected covering aspects such as education, employment, and house characteristics.

	Contact Count (per user)	Physical Distance (per user)	Contact Rate (per user)
education AL		*** -	*** -
education Degree	*** ++	*** - - -	*** - -
education OL		*** -	
employment Not.Active		*** (-)	
employment Unemployed		*** (-)	
floor_materials Concrete		*** - -	
floor_materials Tile.Granite	*** ++	*** - - -	*** - -
housing_type Improvised	*** -	*** ++	
housing_type Permanent		*** - -	*** -
lighting Electricity		*** - - -	*** -
roof_materials Aluminium.sheet		*** -	
roof_materials Asbestos	*** ++	*** - - -	*** - -
roof_materials Concrete	*** ++	*** -	*** (-)
roof_materials Metal.sheet	*** -	*** +++	*** ++
roof_materials Other		*** ++	*** ++
roof_materials Tile		*** - -	*** -
tenure Encroached	*** +		
tenure Rent.free	*** +	*** - -	*** - -
tenure Rent.Government.owned	*** ++	*** -	
tenure Rent.Private.owned	*** +++	*** - - -	*** - -
type_of_structure Single.1S		*** - -	
type_of_structure Single.2S	*** ++	*** - -	
wall_materials Brick		*** -	
wall_materials Cadjan.Palmyrah	*** - -	*** +++	
wall_materials Cement.block		*** - -	*** - -
wall_materials Plank.Metal.Sheet		*** ++	*** ++
wall_materials Soil.bricks	*** ++	*** - -	

*Table only includes census features with 1+ ANOVA signif. code *** (p=0)*

ANOVA signif. codes:
*** p = 0

Spearman's Rho
(+),(-) < 0.15: very weak
+, - 0.15 - 0.24: weak
++, - - 0.25 - 0.39: moderate
+++, - - - 0.40 - 0.75: strong

Social Features
Contact Count :Number of unique contacts

Contact Rate: Average number of connections made by the user with his/her contacts

Physical Distance: Average physical distance between user and his/her all contacts

X

Figure 9:Comparison of social CDR features with census attributes

Upon further analysis, CDR mobility features were shown to have even more stronger statistically significant correlations with attributes in the census. After considering the correlation between the social, consumption and mobility based CDR features and the attributes of the census, a linear regression model was built for each census feature while varying the CDR features that were used as predictors for the model. Additionally, a linear regression model was built with all of the CDR features(consumption, social, and mobility) as predictors, which gave the best predictive performance. The R^2 coefficient for a selected set of census attributes against different CDR feature categories is given in Fig 10 below.

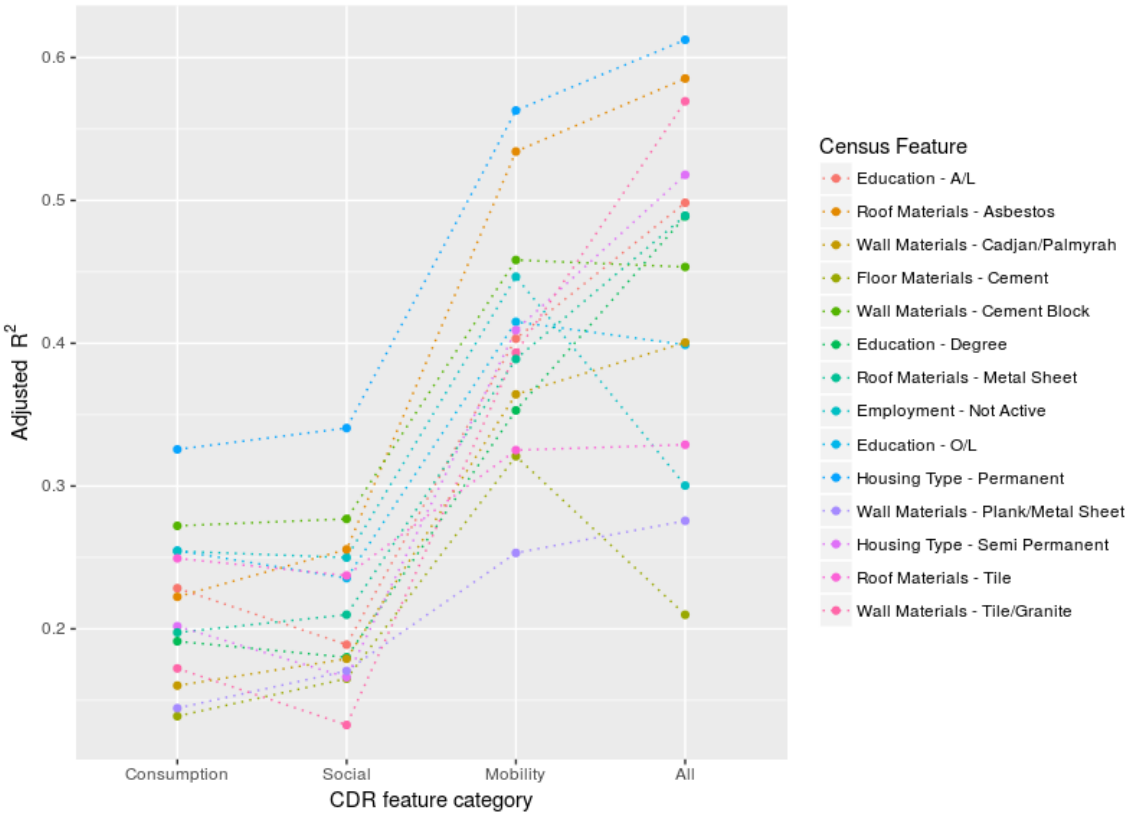


Figure 10: Adjusted R2 of linear regression models built for census features using different CDR feature categories

Currently, we are exploring different machine learning methods instead of linear regression in order to verify whether the predictive performance can be further improved. After that, as a further step, we plan to incorporate additional aspects, such as varying spatial population density, ethnic and religious makeup, etc. Once complete, a workshop is planned for the Census and Statistics Department.

2.3.2. Challenges

The publicly available census data doesn't provide some features at the spatial granularity required for the analyses. Hence the initial analysis doesn't include all the relevant census attributes. Once the analysis work for the current phase is complete, additional features not available in the public domain will be obtained from the Department of Census and Statistics that has supported this research.

2.4. Land-use extensions

In the original land use analysis we leveraged CDR data to understand the land use patterns in the Colombo District of Sri Lanka. We generated a normalized temporal signature of the number of people at each base station and grouped base stations into three clusters (Commercial, Residential and Mixed) based on existing expert knowledge of land use in the Colombo region. However, this discrete representation fails to capture the spectrum of commercial/residential behavior that exists in the Colombo District. Additionally there are aspects of land use behavior such as nightlife and leisure activity that do not easily conform to the somewhat simplistic commercial/residential definitions. Therefore, two extensions to this methodology have been developed. The rationale is to derive more refined methods that will capture a more nuanced

analysis of differential land use behavior. We have extended our data sources to satellite imagery and social media to improve our estimates and capture

2.4.1. Findings

Although previous research was able to identify commercial vs. residential use-based clusters, the methodology was enhanced in order to further understand patterns of land-use in the Colombo district. One approach attempted to understand land-use as spectrum between residential and commercial signatures that had been identified with previous work. Another avenue of research explored the potential of MNBD to extract additional forms of land-use beyond commercial and residential such as nightlife, leisure, administrative, transport etc.

Land use as a spectrum between Commercial and Residential behavior

We developed a linear model that estimated the relative strength of commercial and residential signatures when calibrated using the normalized temporal signature of a base station. The model parameters, error and R^2 was estimated for all base stations in the Colombo District. In general the model captures the variation in the temporal signature of a given base station very well with average R^2 of 0.962.

In reality, there are other land use patterns related to nightlife, leisure etc. The close alignment of the model which is views land use as combination of purely commercial and residential signatures seems to indicate that either,

I. In most base station coverage areas being considered these other behaviors are not sufficiently common to be noticeable among the dominant signatures
II. These activities are not captured well through MNBD due to mobile use being limited at the relevant times

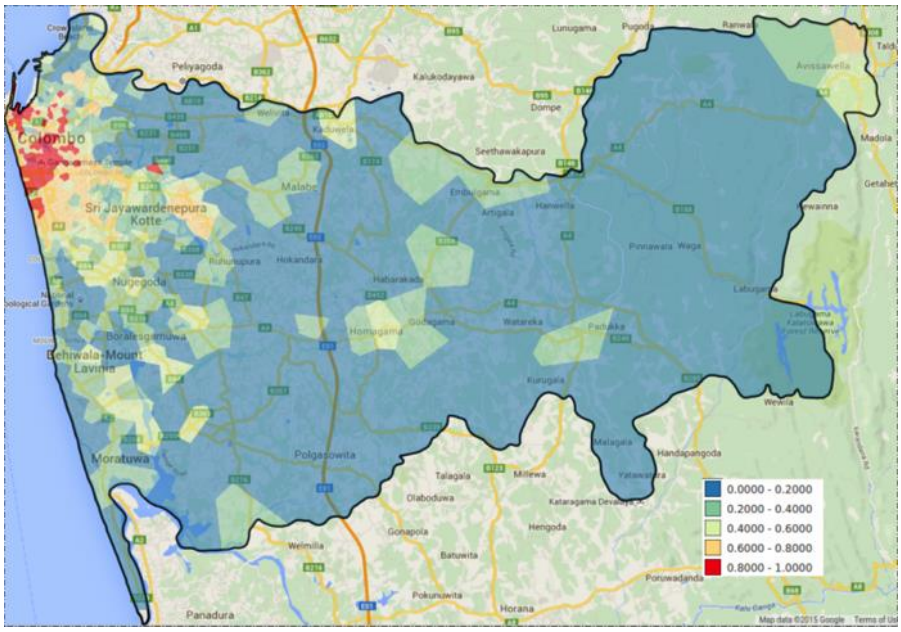


Figure 11:Land use in the Colombo district in terms of commercial behavior

Land use patterns derived from this approach provide a more nuanced perspective of the variation in commercial (residential) activity within the city of Colombo as well as the gradual decline in commercial activity when moving inland compared to the original analysis with discrete land use categories (Figure 10). The comparatively high levels of commercial activity along the main transport corridors in and out of the city are noticeable in amongst the patterns.

Fine-grained land use

Application of GMM with Dirichlet Process detected eight (8) clusters in Colombo district. Visual inspection suggests that not only one but also several clusters can be related to the commercial behavior as found in previous work. One issue with clusters obtained is it is not possible to attribute each cluster to a particular land use only based on the time series signature.

We made effort to infer the potential land use characteristics captured by the clusters by visual inspection of their geographical distribution in the Colombo district (Figure 11).

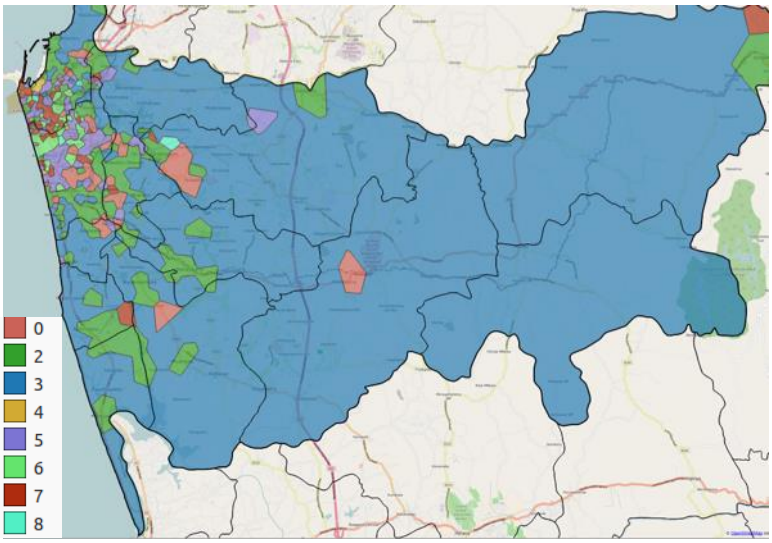


Figure 12:Geographical distribution of the identified land use behaviors in the Colombo district

Cluster 3 is the largest spatial cluster in Colombo district. The daily peak occurs in the evening and it is more consistent with a residential pattern. Clusters 0, 4, 5, 6, and 7 exhibit different gradations of commercial behavior. Cluster 0 consists of BTS in more commercial regions mostly outside Colombo city boundaries. The remaining four clusters consist of BTS showing more commercial patterns within Colombo city limits.

Cluster 8 had unique signature that is significant from the others. This cluster is made up from only 3 base stations in Colombo city. It is difficult to give a meaning to this cluster since it is not possible to interpret the time series signature of this cluster.

We assigned each cluster to a broad land use category depending on visual inspection. With external land use data the accuracy of this methods can be validated and each cluster can be assigned to a more fine-grained land use category.

Land use estimates from heterogeneous data sources

Given the challenges of utilizing MNBD for land use estimations, we have considered data sources such as satellite imagery and social media as we envisioned at the inception of this work. Results from our initial work suggest that we can improve the accuracy and validity of previous results by integrating other data sources into this work. By applying a basic pixel based unsupervised learning classification technique we obtained moderate results for land cover estimates (three clusters) which can be extended for land use estimation with more advanced machine learning techniques (Figure 12). One immediate limitation of extending machine learning techniques would be acquiring labeled data required by supervised learning approaches.

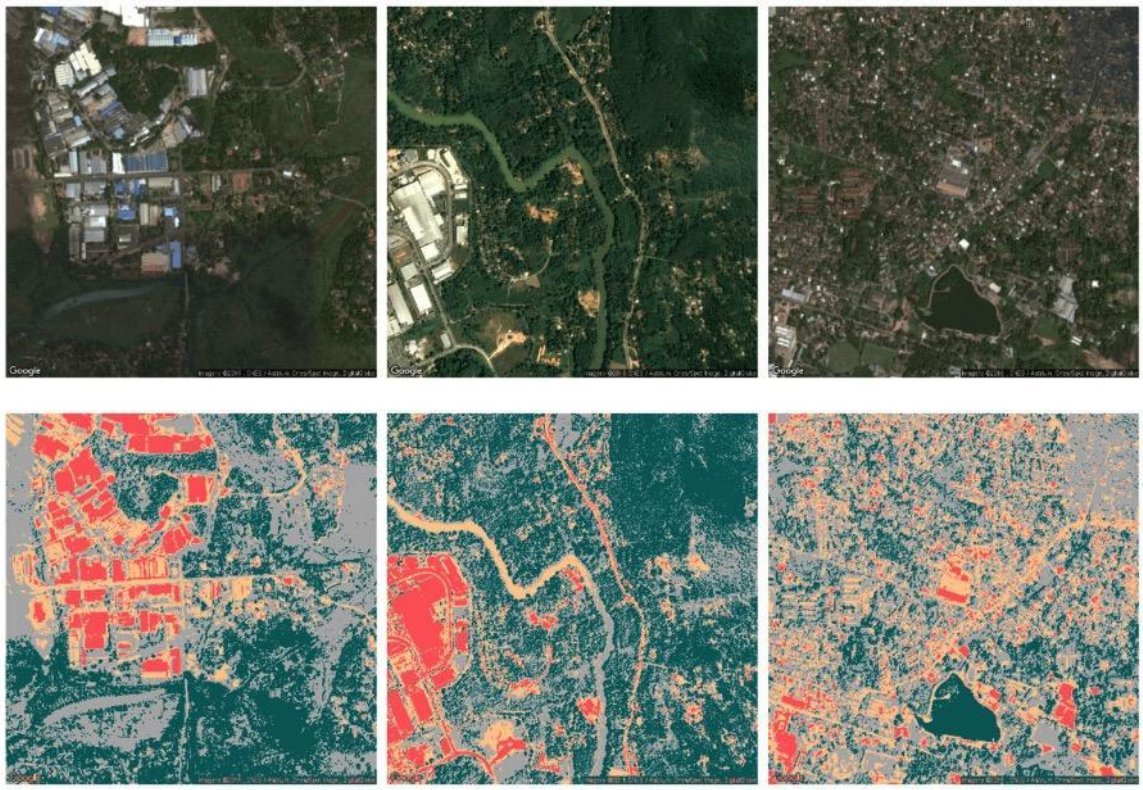


Figure 13:Results of clustering based classification for land cover estimation

Among the many possible social media sources, we opted to work with Foursquare data (as planned initially) due to its focus on point of interest data. The objective of an ongoing student project is to understand the limitations of Foursquare data and formulate a framework to integrate heterogeneous data sources for better land use estimates. A notable limitation in Foursquare data is its bias towards recreational and commercial venues.

Deep learning approach for built-up area estimation

One of the major limitations in previous work that used both supervised and unsupervised machine learning techniques was the low robustness in estimating built-up areas. This lead us to take a step back and trying novel techniques to build more robust models for built-up area estimation. With the recent advancements and successes in applying deep learning (a sub-field of machine learning) for image analysis, many novel techniques have been proposed to use deep learning for remote sensing data analysis. Yet the extensive training data requirement for

deep learning had not been addressed until recent years in remote sensing data domain. Datasets such as SAT-4 and SAT-6 and SpaceNet have reasonably addressed this issue and enabled efficiently applying novel deep learning and machine learning techniques on remote sensing data. Since our interest is on understanding the built-up area distribution we initially developed a deep semantic segmentation model on SpaceNet dataset.

SpaceNet is a publicly available free satellite imagery dataset with pixel level annotations. The dataset includes satellite imagery from five areas of interest (AOI): Rio De Janeiro, Paris, Las Vegas, Shanghai and Khartoum covering 5555 km^2 in the original dataset. Resolution for Rio De Janeiro was 50 cm/pixel and 30 cm/pixel by panchromatic sharpening for other AOI. This dataset is provided through a collaboration between DigitalGlobe, CosmiQ Works and NVIDIA. After removing image tiles with no building footprints to reduce the dataset size and training time, dataset contained 12,831 satellite image titles. After aggregating images from five AOI, we randomly split images as 70% training and 30% test. Widely used VGG16 classification net based Fully Convolutional Network (FCN) was adopted for semantic segmentation of satellite imagery. We also experimented with morphological operations from traditional image processing both as a post-processing step as well as a step within the FCN architecture to observe whether it improves the final result. Applying morphological operations as a post-processing step yielded better results with mean Intersection over Union (IoU) value of 81.4% for test set. IoU is the most commonly used accuracy measure for semantic segmentation. One of the objectives of this work was to explore the possibility of generating more usable built-up area maps from satellite imagery. Typically the results generated by applying state-of-the-art semantic segmentation techniques are coarse and less usable. To improve this we applied contour estimations with rotated bounding boxes to obtain the final built-up area boundaries. The final pipeline that includes deep semantic segmentation and post image processing outputs more robust and usable built-up area estimations from satellite imagery.

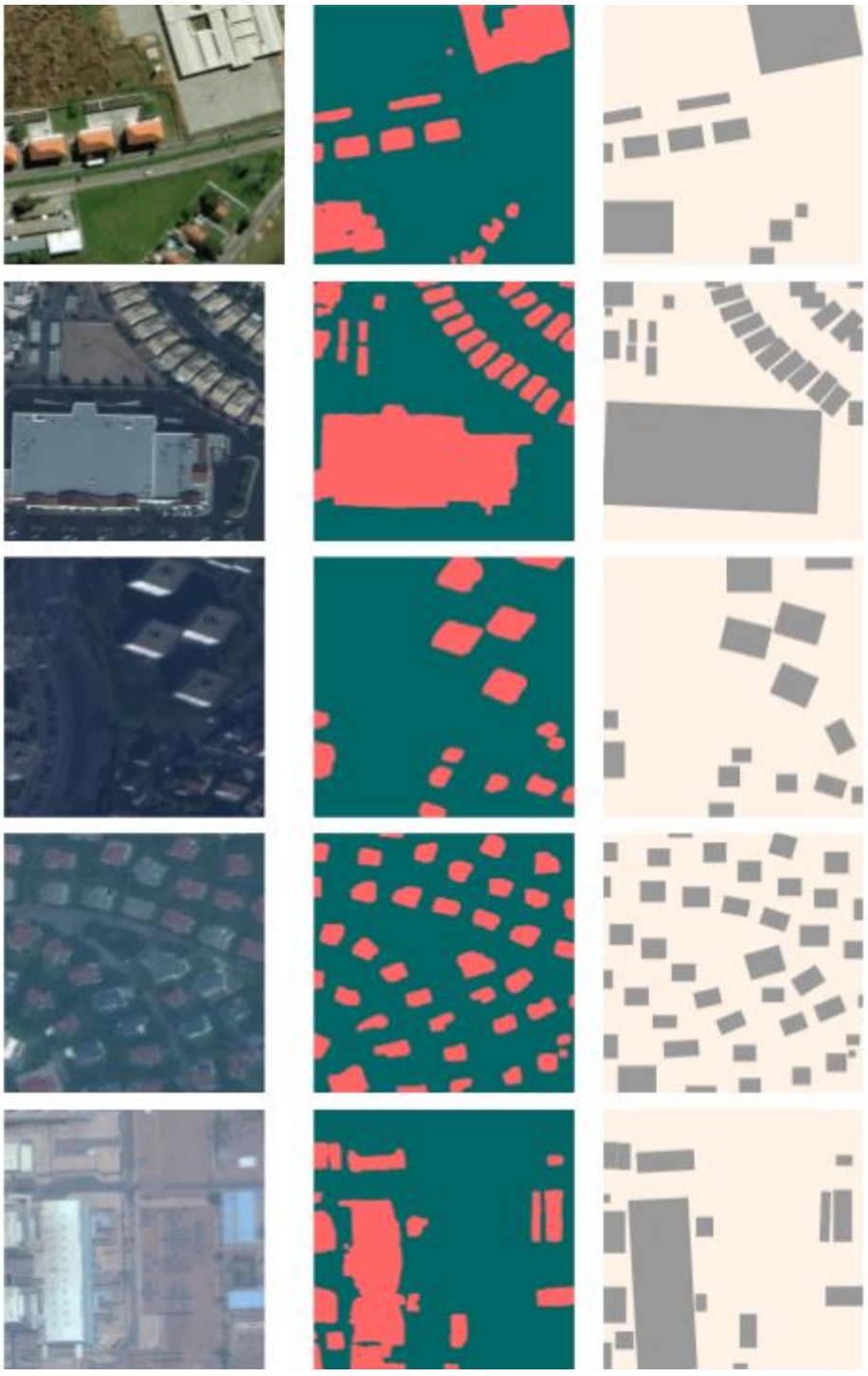


Figure 14:Final result for reference image tiles from 5 AOI in the order of 1-5 row-wise. Left column is image tiles, middle column is built-up area segmentations with binary morphology, right column is final result after rotated bounding rectangles

As a further step, we applied the same pipeline for multiple satellite imagery tiles obtained from different areas of Sri Lanka: Colombo and Gampaha. We retrieved high resolution satellite imagery for these areas from DigitalGlobe Maps API web interface. Resolution was comparable

but lower than SpaceNet imagery. Respective building footprints, obtained from OpenStreetMap were used to validate the results given below qualitatively by visual inspection. We see several potential use cases of this work in socio economic classification and urban planning such as building density as a proxy measure for socio economic level and building distribution for urban area estimates respectively. Same pipeline can be further extended with deep learning and computer vision techniques to estimate other indicators, e.g. roads and rooftop material.

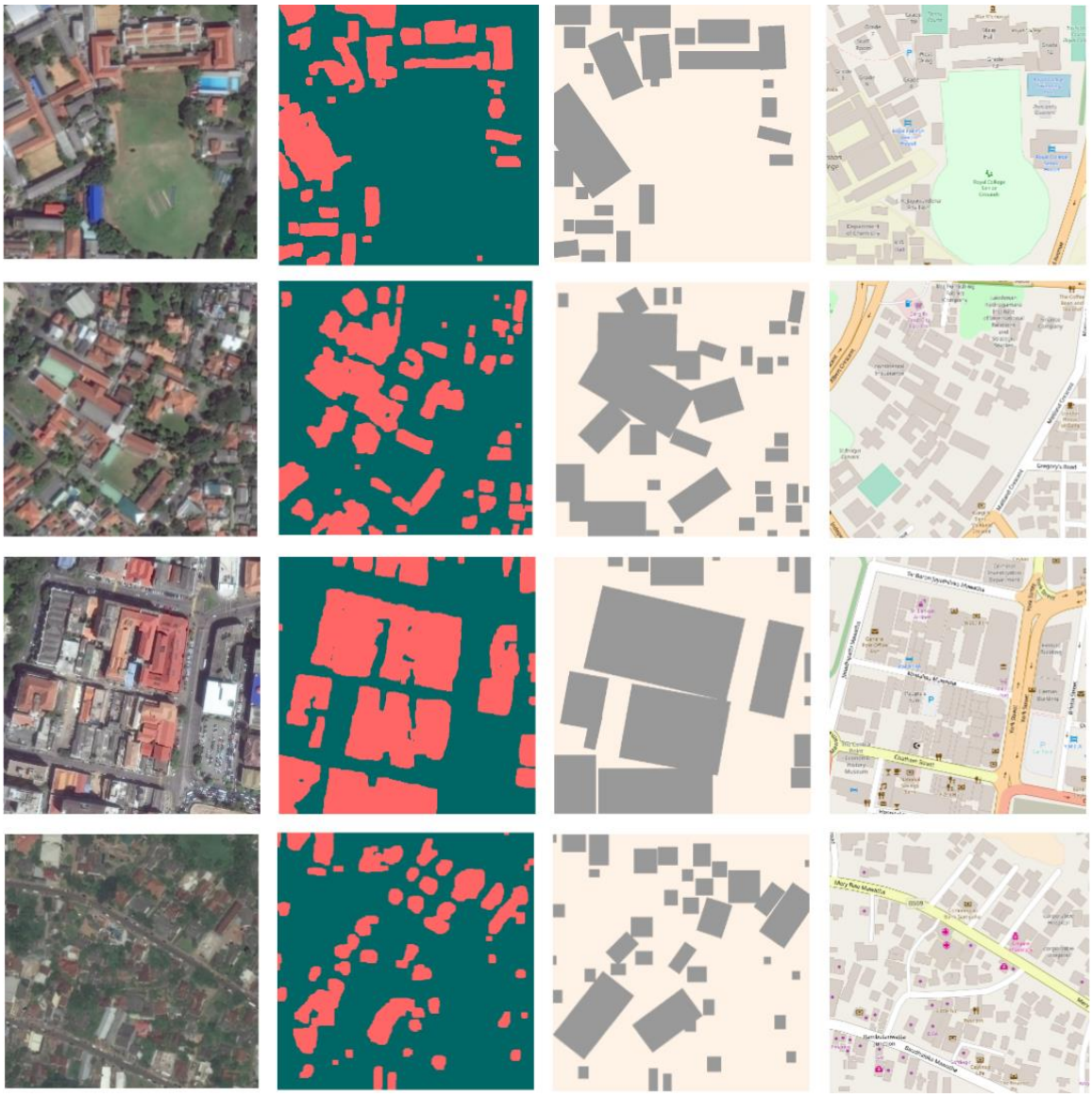


Figure 15:Results for different areas of Sri Lanka: Colombo and Gampaha

Articulating the functional boundaries of urban areas

In urban planning, it is important to delineate the planning boundary. Usually the administrative boundaries of the respective local authority areas are considered as the planning boundary. However, the administrative boundaries do not always fit with functional characteristics of

towns in the planning area. The functionality of a town is depicted by the built up pattern (physical growth of the town) and also by the movement of people and goods. The situation is clearly visible in the urban areas of Asia, where towns have grown naturally with the implementation of transportation infrastructure projects. It is therefore important to devise a mechanism that enables planners and decision makers to better understand a town or region's functional characteristics and its actual spatial extent. A preliminary framework to define *functional boundaries* for commercial land uses was developed by utilizing MNBD.

The study area of Maharagama was chosen due to its urban characteristics and its close proximity to the economic capital of Colombo. The functional boundaries for this region were derived by applying a threshold of distance decay curves of mobility from the Maharagama city center. The derived functional boundary for Maharagama town is visible as a red boundary in Figure 13 below.

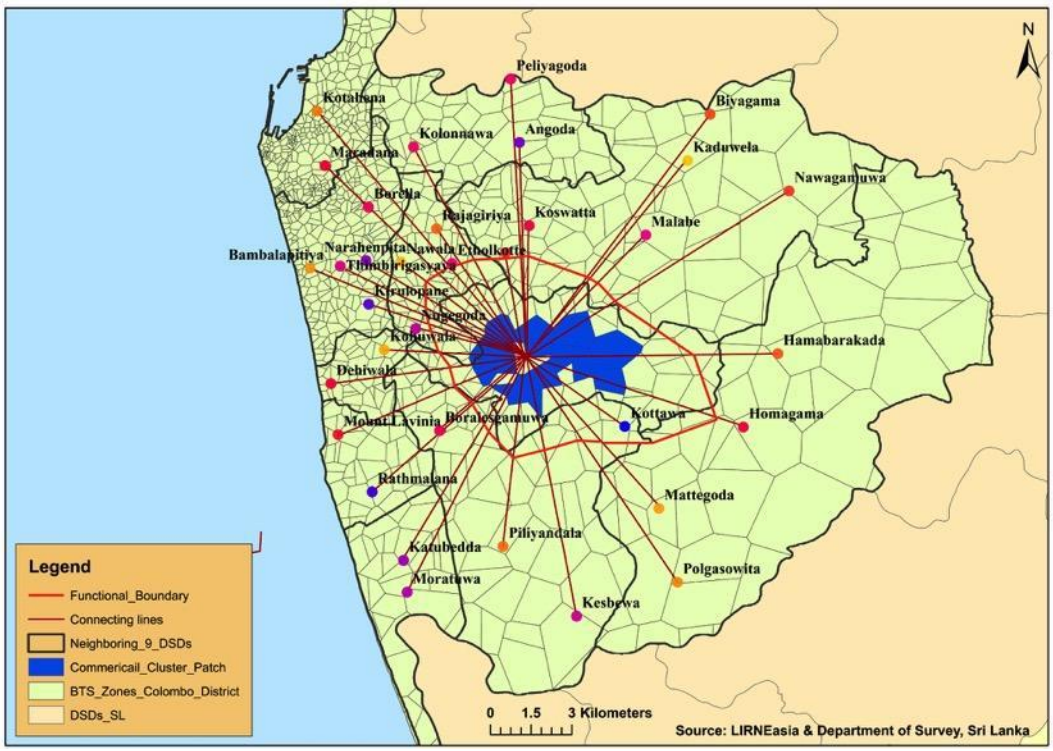


Figure 16:Estimated functional boundary with connections to neighboring towns

2.4.2. Challenges

Further extensions and improvements to land use modeling using MNBD require calibration with additional data such as land use survey maps published by the Urban Development Authority and the Survey Department. Most such data that is available readily is relatively outdated. Different datasets lack standardization, which means additional work is required to aggregate data to generate more comprehensive maps. While we believe that more up to date and high quality data may exist, it has been difficult to develop a comprehensive understanding of the availability of such data as yet. When leveraging additional data sources the main challenge is formulating an integration framework, which can differ between each data source. Ongoing efforts to overcome this challenge will provide us a basis to extend existing results with better estimates. We will continue working on leveraging FourSquare data that can potentially be used in combination with the existing model to identify additional land use behaviors. As this

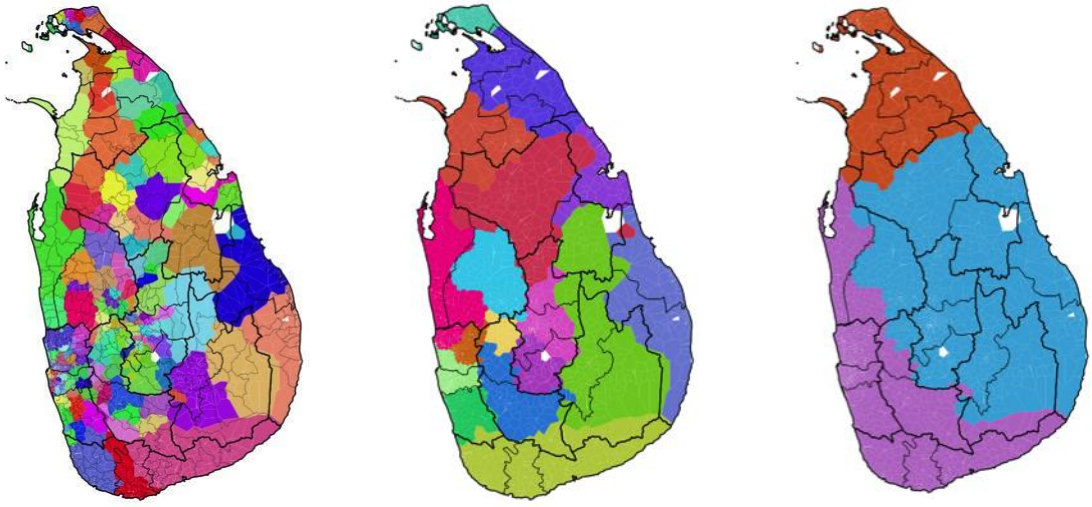
a user-generated dataset with a complicated hierarchy of venue categories substantial manual preprocessing is needed prior to use.

2.5. Geographical boundaries of human mobility

The patterns of aggregate human mobility emerge from the tradeoffs made by people between the availability of opportunities, cost of travel, social and cultural preferences etc. Previous research using CDR data have generated high-resolution estimates of human movement in Sri Lanka. These estimates were analyzed using network analysis techniques to understand the patterns of geospatial clustering present in human mobility behavior in Sri Lanka.

We used an Origin-Destination (O-D) matrix for Sri Lanka that estimated regular travel behavior in this analysis (Maldeniya, Kumarage, Lokanathan, Kreindler, & Madhawa, 2015). An O-D matrix is a representation of the network of flows of people/vehicle between regions. In this analysis the O-D matrix represents the flows of people between base station coverage regions. This representation of mobility as a network of flows makes it possible to employ network analysis techniques traditionally used in identifying community structures in social networks to identify geographical clustering of human movement. Rinzivillo (2012) discusses the application of InfoMap, a hierarchical community detection algorithm to a GPS based mobility network dataset in Italy (Rinzivillo et al., 2012). The algorithm was applied to the Sri Lankan dataset to derive the pattern of geographical clustering in human mobility in Sri Lanka. Human mobility within the regions identified by the algorithm is higher than mobility across these regions.

The algorithm identified 147 small regions with very high intra-region mobility, which it then progressively aggregated to 17 and 3 larger regions as seen in Figure 14.



4th level: 147 zones **3rd Level: 17 zones** **2nd Level: 3 zones**

Figure 17:Hierarchical mobility zones

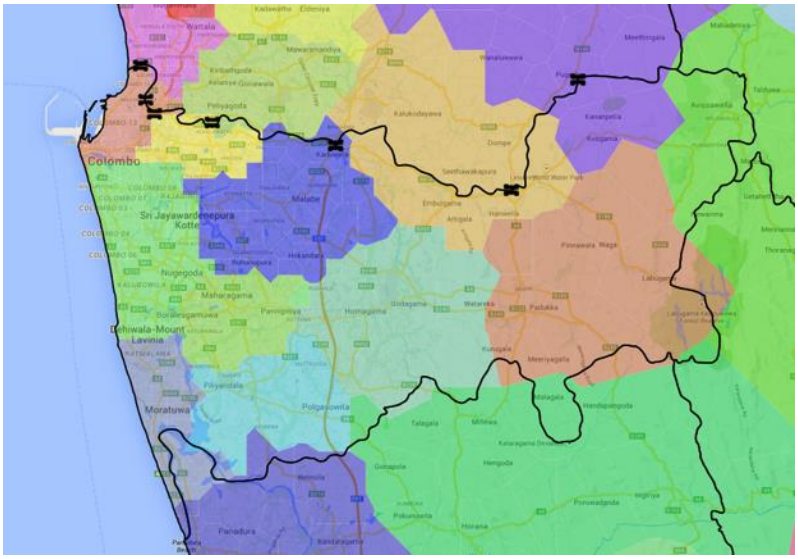


Figure 18:Mobility zones in the Colombo district

Regions identified in this analysis represent strong internal flows of mobility compared to flows linking regions. This has a number of possible implications for urban and transport planning. These regions can provide a new perspective when making decisions on adjustments to existing administrative/institutional boundaries (Figure 15). As the boundaries are defined based on mobility behavior they may be useful as an intuitive alternative frame of reference for transport planning activities. For example the different levels mobility within and across regions in the hierarchy of regions that has been identified may be used as evidence for making decisions on the different modes of travel to be encouraged or developed.

2.6. Comparing geographical boundaries defined by human mobility and communication

Communication and mobility between different regions of a country are a reflection of social, cultural and economic linkages. Utilizing the pseudonymized Call Detail Record (CDR) dataset for Sri Lanka we extracted human calling and travel patterns between different regions in the country at high spatiotemporal resolutions. We created aggregated calling and mobility networks at the base station coverage area level such that links represent, respectively the call volume and the number of trips made by people during the period of observation. This research sought to address two questions. The first being, are there discernible patterns in regional linkages in terms of communication and mobility and do these patterns reveal higher order geographical structures? Such insights have significant policy relevance in terms of defining boundaries for administrative purposes. The second question was whether regional communication and mobility reveal the same patterns? If not where are the linkages reinforced and when are they different between the two networks? We explore the second question by considering the spatial overlap of communities extracted from the two networks. We found that while both calling and mobility networks reduce to small numbers of communities, calling communities are more geographically cohesive. In addition, calling and mobility communities show significant overlap in the west of the country.

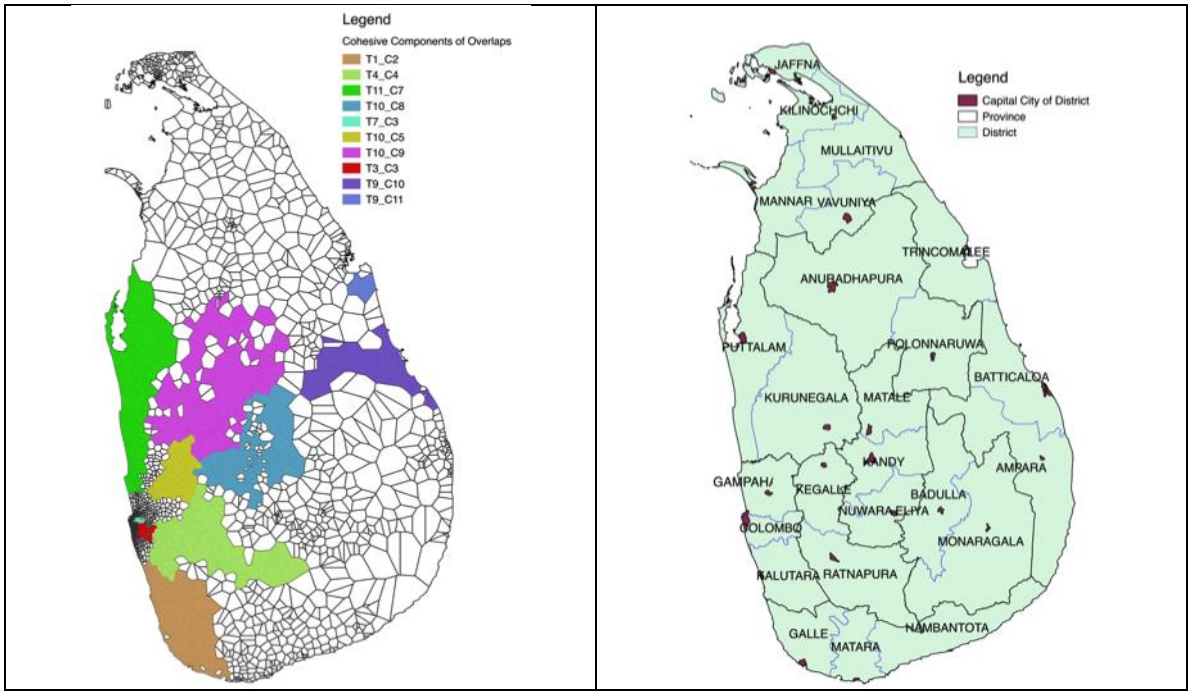


Figure 19: (Left) Cohesive regions of overlap; (right) Administrative boundaries and capital cities of Districts

It is observed that only a few cohesive regions are aligning with the administration boundaries (Figure 16 & 16). Puttalam (the bright green region) is one of them. Part of the reason why the communication-based communities are always geographically co-located, but the mobility based communities are more scattered, could be that the derived communities in each network may consist of different people. This needs to be investigated further. For more information refer to Annex 9: A Comparison of Geographical Boundaries defined by Human Mobility and Communication in Sri Lanka

2.7. CCTV Analyses

Prior and ongoing work leveraging MNBD to gain insights into urban mobility and traffic has highlighted a number of limitations. While it has been very effective at identifying inter-regional travel at good temporal and spatial resolutions MNBD based analytical techniques cannot provide high quality estimates on traffic, travel velocity and travel modes.

There are a number of sensor-based techniques to monitor and generate statistics about road traffic e.g. inductive loops, radar detectors, IR detectors, CCTV. In Sri Lanka, a CCTV system operated by the traffic police has extensive coverage of the greater Colombo area, which has the potential to provide insights at the road level where the usefulness of MNBD is limited. The use of CCTV footage will supplement current MNBD analysis of urban mobility with a particular focus on understanding traffic dynamics.

2.7.1. Findings

The data being used for this work is the property of the City Traffic Department of Sri Lanka Police. We have met with high-level officers of the department regarding access to data and potential outcomes. We have received a limited amount of relatively low-resolution footage from the department for preliminary analysis and for developing a prototype framework. Due to the potentially sensitive nature of the data we have signed a non-disclosure agreement, which bars releasing footage and information about the footage to third parties. We expect to acquire more extensive high-resolution data once the preliminary work is complete.

Two approaches were considered in detecting vehicles from available CCTV footage: a heuristic approach and a learning approach. The heuristic approach performed poorly with the CCTV footage primarily due to the low accuracy of vehicle detection as demonstrated by Figure 18.

Figure 20:Motion based vehicle detection from CCTV footage

Ongoing preliminary work suggests that the learning approach would potentially perform better than the heuristic approach. We are using two learning approaches to understand which performs better. The two approaches are (1) Haar-feature based classification; and (2) Deep learning based classification. Preliminary results are shown in Figure 19.

(a) Haar-feature classification **(b) Deep learning based classifier**

Figure 21:Vehicle detection based on learning approaches

It is important to note that vehicle detection itself is not enough to generate any insight on urban mobility or traffic congestions. Therefore, detected vehicles need to be converted into a format that provides meaningful insights on urban mobility. We leveraged several approaches to do this as we tried to understand the temporal aspect of the vehicle count on roads.

Capturing vehicle count

Even though counting the number of vehicles is a fairly straightforward task for humans, it's a highly complicated task for machines. As the initial approach, we simply counted all vehicles that appeared in each frame of the video. But in practice it is more useful to find the actual vehicle count rather than counting vehicles at each frame. Moreover, counting all vehicles at each frame causes problems, when we have sudden noisy frames or frames with noisy areas. To deal with these kinds of sudden noisy frames, we suggested an approach that is not completely dependent on each single frame. As an alternate solution for partially noisy frames, it is possible to define the less noisy area for detecting vehicles in the given frame.

The second approach suggested for counting vehicles depends on an object tracking method. In this approach, a suitable area for vehicle detection must be defined. It then tracks all vehicles that pass through the defined area. This object tracking approach enabled us to overcome both problems that were outlined above.

Vehicle tracking

We used the Kalman filtering algorithm, which is a framework for predicting a process' state, as the main tracking algorithm. In the Kalman algorithm, the estimation of the state of a process is performed in two steps, namely, prediction and correction. In the prediction step, it predicts future states using prior knowledge about the process. Thereafter, the system asks for feedback about the prediction from the environment and updates that knowledge about the process accordingly.

In our proposed system, the vehicles were initially detected by using a Haar feature based classifier. The system then predicted the next position of the vehicle in the next frame by using the Kalman filter and the predicted output was then corrected by looking at detected positions of vehicles in the next frame (Figure 20). The proposed system continues this process and tracks vehicles until it crosses a given line that is orthogonal to the flowing direction of vehicles. Thus, the count of the vehicle is equivalent to the number of vehicles that cross the given orthogonal line.

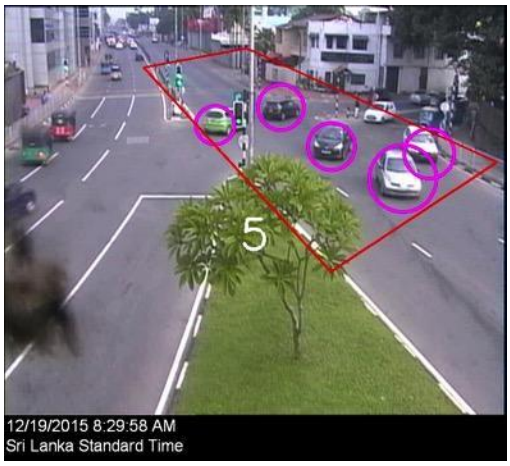


Figure 22:Counting vehicles at each frame

Based on above findings, we continued this work to develop an end-to-end pipeline for vehicle recognition and tracking in road CCTV footage. This pipeline includes both computer vision and deep learning techniques explained above. We use Haar-features, deep learning (Inception network) and kalman filter algorithm for vehicle detection, vehicle type recognition and vehicle tracking respectively. Another important addition to this work was establishing an evaluation framework for these tasks. We had to use multiple accuracy measures from machine learning and computer vision to evaluate different stages of the pipeline such as detection, recognition and tracking. For vehicle detection and recognition we obtained overall test accuracies above 75% and we used multiple object tracking accuracy (MOTA) to evaluate vehicle tracking which we obtained above 90%. One limitation of this evaluation is the dataset sizes. Given the effort required to generate training data from CCTV footage, our datasets included less than 1000 images for each task. Increasing the dataset size is a one major improvement that can be done in future.

We also did preliminary work to detect few basic traffic violation types from CCTV footage. But the lack of standard definitions, evaluation metrics and training data make this task more challenging and less reliable at this stage. Nevertheless qualitative evaluations show the potential of identifying basic traffic violation types such as lane crossing and high speed driving from CCTV footage.

2.7.2. Challenges

The CCTV footage that is used is generated by cameras placed at junctions by the city traffic department of the Sri Lanka police. As at now this footage is monitored and manually analyzed by the police. As a result camera placement has not followed a standard procedure that makes automated analysis simple. The current placement of cameras often result in vehicles moving in oblique angles across the field of vision and multiple roads being captured by the same camera which complicates analysis.

In addition numerous image artifacts due to fluctuating brightness, contrast etc. in the footage pose additional analytical challenges.

- **High congestion scenarios:** When congestion is sufficiently high, vehicles as observed in CCTV footage appear to merge with each other making it difficult to identify individual vehicles. This will affect the accuracy of automatic vehicle detection during high traffic.

More research is need to identify if this issue can be directly remedied or mitigated using modeling the error under different levels of congestion.

- **Manual image annotation:** The high performing learning algorithms used to identify vehicles require a large number of training examples (thousands of images). This represents a significant amount of manual work.

2.8. Behavioral patterns associated with the opening of the Expressway

The primary goal of this work is to develop a methodology for understanding and quantifying the impact of changes in infrastructure on human mobility in the affected region. Specifically the core analysis considered the impact of the opening of Colombo - Katunayaka (E03) expressway that connects Colombo, the commercial capital of Sri Lanka and Katunayaka Airport, the main international airport of the country as well as the nearby city of Negombo (Figure 21).

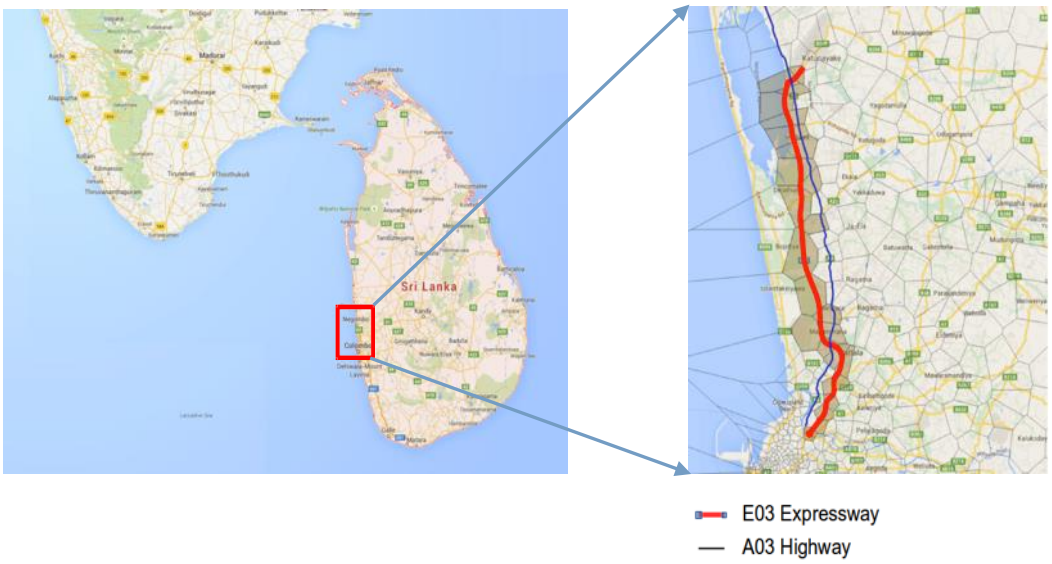


Figure 23:Colombo – Katunayake (E03) Expressway

We have selected a subset from the CDR dataset spanning four months from August to November of 2013 containing all record types except top ups by considering the opening date of the E03 (27th October 2013) and the availability of continuous data.

Two frames of reference were selected in defining the spatial regions of interest when estimating the impact of the opening of the E03.

1. The distance from the E03 expressway
2. The distance from the city centers of Colombo and Negombo

We defined the observation of a “trip” between Colombo and Negombo by looking at the chronologically ordered sequence of location updates of a subscriber on a particular day to capture the travel in mobile network dataset. Then total number of trips were counted for all subscribers for each day of the period of study to generate a time series of travel volume between Colombo and Negombo.

We assume the following when considering the observed travel times to be representative of the actual distribution of travel time.

1. Distortions in travel time due to different effects are unbiased over the population of observed trips.
2. The observed spatial distribution of trips is identical to the distribution of actual trips.

The travel velocity of an observed trip was estimated using origin/destination tower shortest distance and travel time parameters.

2.8.1. Findings

We used the mobility measurements developed based on the mobile phone network data to assess the impact of the Colombo – Katunayaka expressway. The results provided useful insights on travel volume, travel time/velocity and route choice derived from time-series patterns of our analyses. When considering the travel volume between Colombo and Negombo, the change between number trips observed after the opening of the E03 and the observations prior to the opening is insignificant. The E03 expressway seems to have had a limited impact on the volume of travel between these two regions during the first month since opening. In contrast the variation in daily travel time and velocity over the period of study has an abrupt shift, implying that the E03 expressway caused an immediate decrease in the overall travel time between Colombo and Negombo.

2.8.2. Challenges and planned work

To validate the obtained results we need to acquire 3rd party data to confirm the accuracy of the mobility estimates derived based on mobile network dataset. In particular to validate estimates on the number of trips on different routes between Colombo and Negombo before and after the opening of E03, we plan to acquire entry/exit counts for E03 based on toll records and traffic sensor counts for A03 from the Road Development Authority (RDA) of Sri Lanka.

Furthermore we will be estimating time varying mobility variables for subscribers in the study area at a temporal resolution that will allow capturing any changes in behavior during the period of study. The estimated variables will be home/work locations, maximum distance travelled and radius of gyration (ROG), which measures the typical range of an individual's travel trajectory. These mobility estimations will be used to develop a series of high resolution Origin-Destination (OD) matrices that will allow tracking inter and intra city mobility.

2.9. Forecasting the propagation of Dengue using high resolution human mobility data

Dengue fever is one of the most common forms of vector-borne infectious diseases in the world. With a 30-fold increase in infection rate between 1960 and 2010 and a presence in nearly 110 countries Dengue represents a significant threat to global health. In Sri Lanka 30,000-40,000 cases of Dengue have been reported annually in recent years ("Dengue - Sri Lankan Situation," n.d.; Heilman, De Wolff, Beards, & Basden, 2014).

Given the limited spatial range (< 1km) of the Dengue mosquito, human mobility plays a significant role in spreading dengue during seasonal epidemics. Individuals infected by the bite of a mosquito carrying the virus, carry the incubating virus to regions where the virus is not present and pass it on to the local mosquito populations when bitten. Even though dengue has been serologically confirmed in all major towns in Sri Lanka below an elevation of 1200m (WHO Scientific Group on Arthropod-Borne and Rodent-Borne Viral Diseases, 1985), the dengue mosquito density is minimal in high elevation areas of the country (Kusumawathie & Siyambalagoda, 2005). This suggests that the mosquito population might not be enough to survive and pass on the virus to the next generation in high elevation regions, in which case, the disease will not be endemic to that region. Dengue will have to be re-introduced to these regions by infected hosts travelling to these non-endemic regions of the country. Also, number of infected hosts travelling to a particular region will directly influence the magnitude of the outbreak of a particular infectious disease. A model developed for dengue that incorporates mobility can be used in predicting outbreaks of similar diseases such as Zika or any potential epidemic or pandemic threat that we would have to face. Therefore, knowledge of aggregate human mobility patterns in a region is critical to understand and predict the spatial spread of Dengue and many other infectious diseases.

Literature on similar studies done for Dengue and other vector borne diseases also show improved results when using human mobility patterns derived from CDR data (Wesolowski et. al, 2012; Wesolowski et. al, 2015). Also statistical models (Sarzynska, Udiani & Zhang, 2013;Lourenço & Recker, 2014) as well as machine learning techniques such as Neural Networks, Support Vector Machines, Random Forests (Aburas, Cetiner & Sari, 2010; Rachata et al., 2008; Yusof & Mustaffa, 2011; Rehman et al., 2016) have been used in contemporary studies which lead us to focus more on these methodologies during our preliminary work.

Preliminary work was carried out through structured and mentored projects for two undergraduate student group projects, with four final year undergraduate students per group from the University of Moratuwa (UoM) and one full time graduate research student currently doing his Master's degree. On 6th May 2016, LIRNEasia and the Health Informatics Society of Sri Lanka jointly convened a planning meeting on building better models for forecasting the propagation of infectious disease such as dengue in Sri Lanka. The meeting was intended to lay the foundation for a multi-disciplinary collaboration engaging health informatics specialists, epidemiologists, and data scientists to identify research priorities and opportunities. Principally this was our initial engagement with the Epidemiology Unit of the Ministry of Health, with whom we are in the process of building collaboration.

2.9.1. Findings

Primarily two forms of data are required for an analysis that attempts to relate the propagation of Dengue to human mobility patterns. These include dengue incidence data and high-resolution mobility patterns. For the former, weekly Dengue cases at the Ministry of Health (MOH) region level for 2013 and 2014 collected by the Epidemiology Unit were used. Origin-Destination (O-D) matrices representing regular behavior in Sri Lanka generated from Mobile network CDR data were in the initial models that were generated. However, upon further analysis, we came to realize that a different mobility model that captures information about time spent in a region by a given population would be more relevant to disease forecasting, when compared to an O-D

matrix mobility model. This lead us to develop a separate probabilistic mobility model that can then be plugged directly into different models to forecast disease incidence.

The initial modeling work to forecast dengue incidence incorporated the probabilistic mobility model as well as weather (temperature as well as rainfall). The model was calibrated using dengue incidence data from 2013 and weekly predictions were made for disease incidence in 2014 (Figure 22).

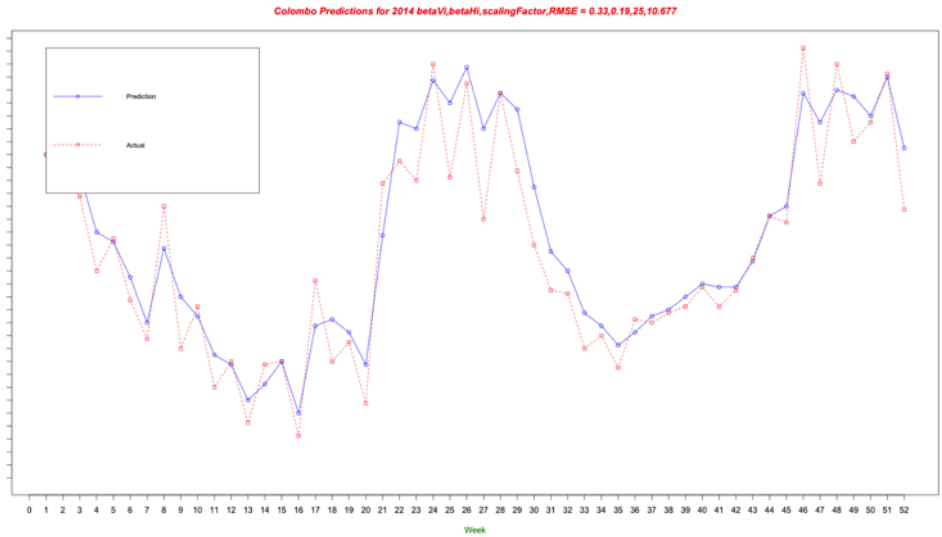


Figure 24:Prediction of weekly disease incidence in the Colombo MOH area in 2014

The preliminary work showed promise and the work was extended utilizing machine learning techniques (Neural Networks, Support Vector Machines, Random Forests, XGBoost). Furthermore a spatial vegetation index was also considered as an additional attribute. As the results below (Figure 23) show, the incorporation of mobility into each of the model improves the results for each technique, suggesting that even with dengue being endemic in most regions of Sri Lanka, human mobility is still a statistically significant contributor to dengue propagation. Amongst the models, the support vector machine (SVR) technique provided the lowest error (i.e. RMSE). However the overall fit (i.e. r^2) of the prediction curve for all the techniques was quite low for the preliminary work.

Model	RMSE		R^2	
	Without Mobility	With Mobility	Without Mobility	With Mobility
Hybrid SEIR model	16.281	14.859	0.203	0.337
Neural Network	10.923	9.867	0.194	0.342
XGBoost	10.023	9.548	0.321	0.384
SVR	6.463	**6.236**	0.351	**0.396**
Random Forests	7.986	N/A	0.010	N/A

Figure 25:Comparison of model performance

Through the additional incorporation of genetic algorithms (GA), we were able to improve the overall fit (i.e. r^2) and thus further reduce the overall error (i.e. RMSE). Annex XX, which was accepted for an oral presentation at the 2017 NetMob conference in Milan, Italy and provides more details on the ongoing work.

We further expanded our analysis to include 20 MOH divisions and incorporated 2 other mobility models as well. A trip-based mobility model was developed that focused on the aggregate value of incoming and outgoing trips of residents of a particular MOH division. Similarly, we developed a risk-based mobility model, which extended upon the probabilistic mobility model developed by us initially, where each CDR was weighted by a risk factor according to the location and time of day. The results of this extended analysis are given below.

Machine Learning Technique	Overall RMSE	Overall R^2
Random Forests	8.258	0.926
Neural Networks	12.154	0.839
XGBoost	7.852	0.933
SVR	8.618	0.919

Figure 26:Comparison of model performance for final analysis

2.9.2. Challenges

- **Availability and access to high resolution geospatial Dengue incidence data:** One of the key issues faced earlier was that we had not been able to locate geographical boundaries that define Ministry of Health regions which are the frame of reference used in the Dengue incidence provided by the Epidemiology Unit. However, this information had been compiled for internal purposes by the Epidemiology unit and we were able to obtain that data through a recent collaboration. In addition, the MOH regions refer to the location of diagnosis and treatment and not the region where the patients contracted the disease. This can potentially add a significant level of error to the analysis given the rather uneven spatial distribution of healthcare infrastructure. We are currently exploring the availability of data on the home locations of Dengue patients at the Epidemiology Unit to the extent of the relevant postal code or a similar resolution.
- **Low temporal resolution and low timespan of Dengue incidence data:** The Dengue incidence data is currently available only at the weekly level. Given the short life span of the Dengue mosquito and the Dengue fever incubation period, a model trained with incidence data at the daily level is likely to perform better. Additionally, it is important to have data for a larger timespan so that the temporal variations that can occur in the disease incidence curve are captured across multiple years. Also, when using Machine Learning techniques, a significant amount of observations is needed to ensure that the training set captures all the possible variations that can be observed in the actual data. Epidemiological Unit, Ministry of Health in Sri Lanka is the main unit that handles Dengue outbreaks. In addition to the epidemiological and entomological expertise of the disease, the epidemiological unit has the

required incidence data for a larger timespan. There is significant dependency placed on the Epidemiological Unit for data and expertise that is vital in order to carry out this component of the project.

- **Sparsity and low resolution of weather data:** Weather conditions can have a significant impact of the Dengue mosquito population in a region and as a result intensity of a Dengue outbreak in a region. Given the limited number of weather stations in Sri Lanka, the spatial resolution of temperature and rainfall data is lower than Dengue incidence and mobility data. In addition a significant number of the weather stations suffer from a non-trivial level of missing data.
- **Limited entomological information about the Dengue mosquito population in Sri Lanka:** Different populations of the same species of mosquito may demonstrate to have significantly different effects and Dengue propagation dynamics based on climate, weather and other conditions. Access to extensive information on the different mosquito populations in Sri Lanka, particularly their population and Dengue propagation dynamics will be useful in calibrating models. Vector surveillance data at spatial and temporal dimensions can also be used to generate a suitable index that can be used to calibrate our predictive models. Vector surveillance data is also available with the Epidemiology Unit, Ministry of Health. We are currently in discussion with them to obtain this dataset.

2.10 Analyzing Facebook data to understand international connectivity

An ongoing study using social network data aims to understand social linkages between countries, and the corresponding links in economic activity and human movement.

In July 2017, Bailey et al published a paper titled Measuring Social Connectedness.
In it, they introduced a new measure of social connectedness between U.S. county-pairs, as well as between U.S. counties and foreign countries. The measure (called the "Social Connectedness Index", or SCI), was based on the number of active Facebook friendship links between geographies as of April 2016, using IP to allocate people to countries and tabulating the number of connections between them. Their study found that the Facebook data, for US counties, correlated strongly with rade, migration and facets of geography. The Facebook study also suggested that the SCI was detecting facets of homophily, showing correlations between friendships links and distance, and shared histories of migration between US counties.

Potent arguments for the existence of such networks in trade and migration are found elsewhere as well: Filip Gariz's research shows the links between homophily and migration, and Başak Bilecena, Markus Gamperb, Miranda J.Lubbers in "*The missing link: Social network analysis in migration and transnationalism*" point out the role that social networks in analyzing migration. *Clusters or networks of economies? A macroeconomy study through GDP fluctuation correlations* by M. Ausloos and R. Lambiotte also introduced methods of correlating GDP fluctuations and proved that a economies can be expressed in terms of interlinked networks and societies. Others, such as Trevor Martin, analyzing such networks in http://www.shorttails.io/are-worldwide-economies-correlated/, found strong correlations between GDP growth over time between certain nations and blocs, indicating network effects that may not be readily visible without specialized political knowledge.

Thus goal of our ongoing study is to examine existing coalitions of countries in the light of this social connectivity, using it to determine qualities such as the interconnectedness of a region and to analyze the social organization of countries within it, providing a replicable metric and methodology for doing so that relies on quantifiable data as opposed to varying social and political insight. A final goal is to attempt to use this data to identify communities of nations based on their social and economic connectivity.

2.10.1 Findings

Our exploratory analysis had revealed that this social network data does strongly correlate to trade and migration between countries, but not to the distance between countries.

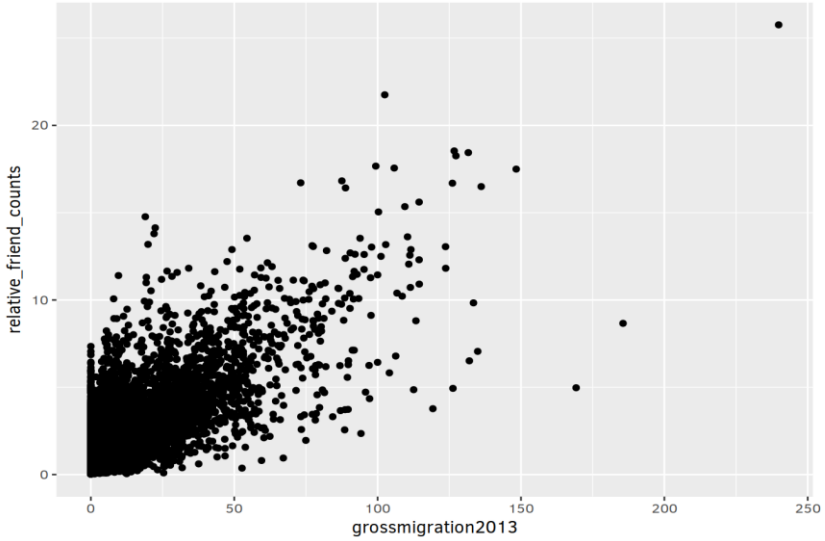


Figure 27:Gross bilateral migration from 2013 versus friend links between country pairs.

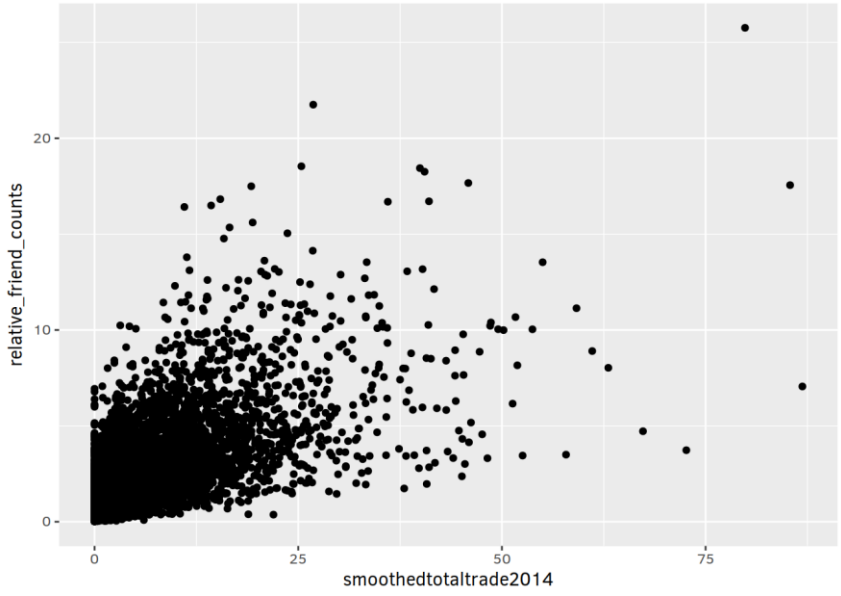
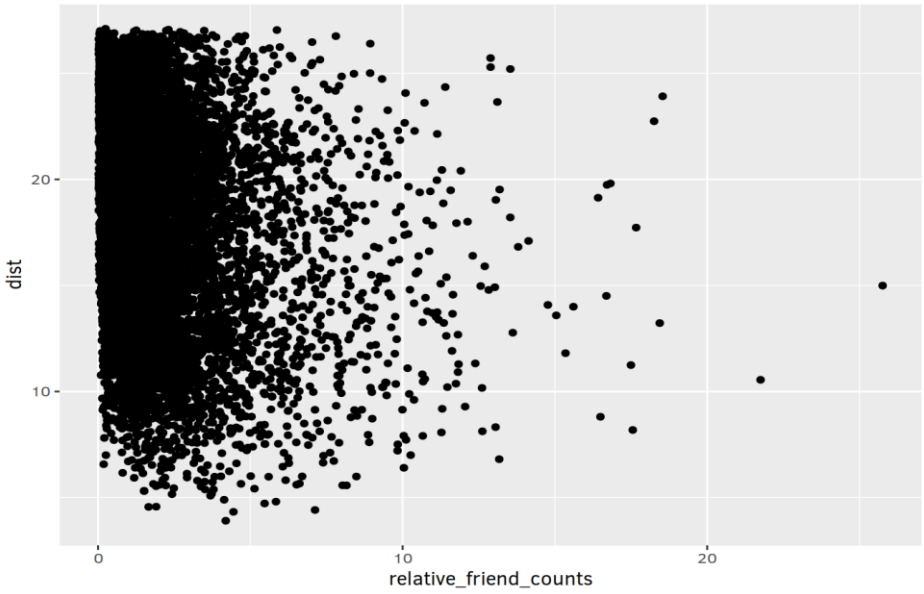


Figure 28:Gross bilateral trade from 2014 versus friend links between country pairs.

Linear regression models built for these three factors - using connectivity versus trade and migration together - show an R-squared value of 0.6443, indicating that Facebook connectivity can be used as an able proxy for a combination of these two factors. No coherent and consistent relationship has been found between distances between nations and their social connectivity.
Thus, social network connectivity between nations potentially serving as a possible way of exploring networks of international communities, human movement and economic activity. This often appears to transcend governmental and political barriers (such as between India and Pakistan, for example).



We wish to explore these relationships further, both by examining homophily (as expressed in the seminal Birds of a Feather paper) at large with this data, and by also analyzing various existing political blocs such as the SAARC, ASEAN, BIMSTEC and others to understand how best this new metric can be put to use. The lack of correlation with geography indicates that this network may not follow the geography of the world, but rather may yield new insight into how countries form socioeconomic communities.

2.10.2 Challenges

- **Facebook data is not accurately representative of regions such as Russia, Japan and China, where alternate networks are popular**
  We will need to come up with a weighting variable based on Facebook penetration statistics, or factor in data from other networks
- **Data pre-processing done by Facebook**
  What we have is not the raw data for the counts of users, but an aggregated matrix of normalized connectivity on a country level. It is difficult to reverse-engineer back to the original population stats. The network could be understood much better with the raw data, but it is unlikely that this can be obtained

- **Alternate patterns of Facebook usage from certain parts of the world**
  One of the underlying assumptions made here is that the Facebook data is representative of Facebook users. This is so in the US, as proven in the original paper, but new research from Lirneasia in Myanmar using a non-representative sample showed users with up to twelve profiles interacting with thousands of others with all such profiles. This may be just an isolated case in Myanmar, with its unique political and cultural situation, or may be a pattern at large across Asia: more representative research will be needed here.

- **MIssing data and lack of precise trade and migration data**
  Trade and migration datasets miss out on illegal economic activity and human movement. Our data is from the World Bank and the Correlates of War project; it would be ideal to have some imputation of expected trade and migration, but this data needs to be found or build manually after studying historical migration and trade worldwide.

2.11 Alleviating Harms

Exchanges with government and industry officials indicated interest in understanding the legal and self-regulatory options pertaining to the broader concept of lawful interception, which is the obtaining of communication network data for the purpose of analysis or as evidence. This generally refers to transaction-generated data but, in some instances includes the content of the communication as well. While the research began with a narrow focus on lawful interception, it soon expanded to explore issues relating to bulk surveillance of communications data.

LIRNEasia, as an expert organization reached out to the Asia Pacific Telecommunity (APT) to use it as a platform for multi-stakeholder discussion with the aim of assisting the holders on MNBD and governments in the countries where the research is ongoing to consider the establishment of a legal or self-regulatory framework for MNBD a pre-condition for third-party use. A discussion document was prepared in non-legal language, a workshop was conducted and draft legal text has also been prepared.

Issues such as privacy and governance were brought out at various conferences. Select conference by Prof. Rohan Samarajiva are included below:

- “Confluence (or lack thereof) of data analytics and law,” SAARC Law Conference 2017, Colombo, 28 October 2017.
- “Big data, development and governance,” presentation at Carnegie India, New Delhi, 19 July 2017.
- “Big data for development research in the Global South: Experiential lessons,” Symposium on Big Data and Human Development, Oxford Internet Institute, 16 September 2016.
- “Public interest research using big data: Opportunities and privacy implications,” presentation at Big Data in the Global South International Workshop, Institute for Technology & Society of Rio de Janeiro, 16-17 November 2015.

2.11.1 Broader issues of MNBD including lawful interception

The acquisition of data in large volumes–bulk data[1]–could include both the communication itself as well as the contents of communication. Bulk data reportedly allow authorities to obtain information not readily accessible using conventional targeted means, for instance identifying behavior patterns and networks (*Operational Case for Bulk Powers*, n.d.) As expressed by a Danish government representative: [2] "It's impossible to know beforehand which data might be relevant in the future for solving criminal cases. Data retention provides investigators with the benefit of hindsight…." (Bowcott, 2016).

Proponents claim that the analysis of bulk data can help authorities detect emerging threats, particularly in situations where they do not have enough information to conduct targeted surveillance. Moreover, after a person of interest (POI) has been identified, an analysis of the communication patterns of individuals the POI has communicated with would help identify a broader network of accomplices. Proponents state that in fast paced investigations, bulk data could provide a sense of directionality enabling prioritization of resources to handle the more serious threats. Once bulk surveillance is used to identify a POI, targeted surveillance techniques could then be deployed to obtain more information on the target. Among other things, the analysis of bulk data can help identify POI and provide insights on the POI, including their social networks, behavior and movement patterns. It can also be used to provide more context on the network environment, and guide decisions on resource prioritization (Anderson, 2016). It is possible to classify the use of bulk data to identify persons of interest and potential targets (places) before an act of terror is committed, and to identify perpetrators after an act of terror. This can be illustrated as follows:

Table 1: Forms of Bulk Surveillance

	1. Person(s)	**2. Place**
1. Pre-event	**1.1 Pre-event – Person**	**1.2 Pre-event – Place**
2. Post-event	**2.1 Post-event - Person**	**2.2 Post-event – Place**

Source: Authors, 2017

However, bulk surveillance has been criticized on numerous grounds including, but not limited to intrusion of privacy of individuals and suppression of freedom of expression. It is useful to approach problems of bulk data and surveillance based on the understanding that the physical and virtual are best seen as a continuum with different qualities predominating in different parts. Virtual space is not radically different from physical space nor is it an appendage of physical space. This would require solutions to problems of law-breaking, deterrence, detection and punishment being derived from common principles and, wherever possible, from laws that are not "medium-specific," by which is meant that what is a transgression should not change depending on whether it occurred in physical or virtual space. The forms of investigation,

evidence gathering, and so on would have to be different, but even here, it is advisable to anchor the practices on principles that are common to the both physical and virtual space.

2.11.2 Challenges

- One of the biggest issues in relation to understanding the issues of bulk surveillance, particularly in relation to MNBD is the veil of secrecy that shrouds many of these activities. If communication data is being used to fight crime, terrorism, strengthen national security and improve public safety, it is in the interest of law enforcement agencies to keep the public remain unaware or unsure of the nature and scope of surveillance being conducted. The rationale would be that it would make the taking of evasive measures more difficult for law-breakers.

- Although LIRNEasia reached out to the APT, there appeared to be little interest in a plurilateral discussion. Many of the other activities (such as assisting government and/or operators to develop legal or self-regulatory frameworks for MNBD) were conditional on this discussion with APT.

3. Project implementation

3.1. Collaborations

Collaborations continue to be key for LIRNEasia for both conducting the research as well as for facilitating the use of big data insights into policy making. To that end collaborations from the previous project have been strengthened and new collaborations have been initiated.

3.1.1. University of Moratuwa

LIRNEasia has been collaborating with a variety of departments within the University of Moratuwa. The linkages are deep and have been developed since the previous research cycle. Currently we collaborate with Dr. Amal Shehan Perera, Senior Lecturer at the Department of Computer Science. Through him we are currently mentoring two undergraduate student research projects and will continue to do so further. Dr. Perera is also a Senior Research Fellow at LIRNEasia and collaborates with LIRNEasia on the development of better models of dengue propagation using mobile network big data. Similarly we collaborate with Prof. Amal Kumarage of the Department of Transport and Logistics, principally on transportation and urban related areas of research. Prof. Kumarage and Sriganesh Lokanathan (LIRNEasia) are joint thesis advisors for a Master's student exploring the use of MNBD in transportation planning. We also initiated collaboration with the Department of Town and Country Planning and are jointly mentoring an undergraduate thesis and internship to develop and forecast new measures of land use interaction.

3.1.2. University of Sri Jayewardenepura

Since late 2016, LIRNEasia has been mentoring two students as part of their undergraduate thesis on extending the state of the art in utilizing CCTV footage for traffic monitoring. This was a follow-up from a lecture given to staff and students at the university in October 2016. The student research projects are just the first step in extending big data research in the development space to the University of Sri Jayewardenepura.

3.1.3. Various Government Departments in Sri Lanka

Multiple engagements have occurred with the Urban Development Authority and the Sri Lanka Strategic Cities Development Project. These have led to multiple requests for insights from our ongoing big data research as well as the exploration of potential collaborations and joint research on a variety of ongoing projects (for example see sections 5.1, 5.3, 5.4). LIRNEasia has continued to engage with the Department of Census and Statistics with a view to building a long-term strategic partnership with the Department of Census and Statistics in the area of big data and official statistics. We are currently in the process of formalizing this engagement. LIRNE*asia* has also initiated collaborations with the Epidemiology Unit to further the research on building better models of dengue propagation. Currently the Sri Lanka Police have provided CCTV footage to explore the potential of applying machine vision techniques to understand the flow of traffic. Upon developing a proof of concept a partnership is envisaged with the police department to leverage their data to improve insights on traffic in Sri Lankan cities.

3.1.4. University of Dhaka

Prof Moinul Zaber is an Associate Professor, Department of Computer Science and Engineering, at the University of Dhaka where he started the Data and Design Lab. He is the principal researcher for the Bangladesh component of this grant, which has focused on the analyses of electricity data. LIRNEasia has helped to develop the lab and also provided training for the researchers and students associated with the lab (Also see section 3.3)

3.1.5. School of Information, U.C. Berkeley

Prof. Joshua Blumenstock is at the forefront in the use of big data for development purposes. LIRNEasia has been collaborating with Prof. Blumenstock, in his capacity of Founder and Co-Director of the Data Science and Analytics Laboratory of the University of Washington, since 2014 and continue to do so principally on areas related to urban agglomeration and the use of mobile network big data to model the effect of new transportation infrastructure and transportation shocks. Since 2016 his affiliations have changed to U.C. Berkley in the capacity of Assistant Professor, School of Information.

3.1.6. Fields of View

Fields of View is a Bangalore based research institute that has collaborated with LIRNE*asia* to develop a simulation of urban transportation dynamics for Colombo. The plan is to develop this into a longer-term collaborative project looking at the unique characteristics of cities in the Global South. Currently, Fields of View is involved in developing a simulation model for a selected area of Colombo by using the MATSim framework. A preliminary simulation is currently being built for the region of Kadawatha (a location close to Colombo and further earmarked to be developed into a transportation hub as per the government's 2030 development plan for the Western Province). Ultimately, insights generated from this proposed simulation model would be useful for transport planners.

3.1.7. University of Tokyo

A research collaboration was formalized with the University of Tokyo in October 2016. In February 2017, the Global Partnership on Sustainable Development Data (GPSDD) competitively awarded the University of Tokyo and LIRNEasia one of the their first ten research grants. The project Dynamic Census (http://lirneasia.net/2017/03/lirneasia-and-university-of-tokyo-receive-competitive-funding-for-dynamic-census/) is intended to extend ongoing research at LIRNEasia both on socio-economic mapping as well as well mobility based insights primarily through the

incorporation of targeted surveys to build richer demographic and economic characteristics to the ongoing research. Overall this collaboration

3.2. LIRNEasia Data Analytics Advisory Council

In September 2015 LIRNEasia constituted an advisory group for LIRNEasia's ongoing big data for development research, consisting of international and local experts from a variety of domains. The inaugural meeting of this advisory group was held in Colombo on 17th January 2016 in Colombo.

Given the multi-disciplinary nature of LIRNEasia big data for development (BD4D) research, the advisory council is intentionally constituted from different disciplines and domains. Members include data science experts, sector specific experts (e.g. transport planning, urban planning), social science experts, and legal experts.

The purpose of the advisory group is multifold. Council members will help hone our research agenda in addition to bringing in new research ideas. Given that the whole field of big data for development is still in its embryonic stages they will serve as important reviewers for our ongoing work and where feasible provide quality assurance on process, methodology, analyses, and results. Additionally they will strengthen our networks and facilitate collaborations as well as help identify funding sources.

The first meeting of the advisory council held in January 2016 served to brief the advisory members on what LIRNE*asia* had accomplished so far with respect to its big data for development research. There was much discussion in relation to sustainability not just in terms of human capital but also financial. Whilst the council was appreciate of LIRNEasia looking at alternate funding models including providing private sector services, more work is needed before any plan can be finalized. The data science experts also provided detailed comments on our existing workflow and internal processes, some of which have been initiated amongst the data science team.

Table 2: LIRNEasia Data Analytics Advisory Council Members

Dr. Linus Bengtsson, PhD Co-Founder and Executive Director Flowminder	Joshua Blumenstock, PhD Assistant Professor, School of Information University of California, Berkeley	Nitesh Chawla, PhD Professor of Computer Science and Engineering Director of The Interdisciplinary Center for Network Science & Applications (iCeNSA) University of Notre Dame
Vanessa Frias-Martinez, PhD Assistant Professor School of Information Studies University of Maryland	Amal Kumarage, PhD Senior Professor at Department of Transport & Logistics Management University of Moratuwa	Ashwin Mahesh, PhD Founder & CEO Mapunity
P.K.S. Mahanama, PhD Professor and Former Dean of Department of Town and Country Planning University of Moratuwa	Wasan Pattara-atikom, PhD Principal Researcher & Head of Intelligent Transportation System Laboratory NECTEC	Amal Shehan Perera, PhD Senior Lecturer, Department of Computer Science and Engineering University of Moratuwa
Louiqa Raschid, PhD Professor, Smith School of Business, Center for Bioinformatics and	Srinath Perera, PhD Vice President, Research WSO2 Inc.	Prabir Sen, PhD Former Chief Data Scientist Infocomm Development Authority of

Computational Biology UMIACS and the Department of Computer Science Robert H. Smith School of Business		Singapore
Hetan Shah Executive Director, Royal Statistical Society United Kingdom	Ryosuke Shibasaki, PhD Professor, Dr.Eng. Center for Spatial Information Science University of Tokyo	Linnet Taylor, PhD Marie Curie Research Fellow, Faculty of Social and Behavioural Sciences University of Amsterdam
Ruvan Weerasinghe, PhD Senior Lecturer University of Colombo - School of Computing	Arj Wignaraja Vice President of Operations Remote Sensing Metrics, LLC	

Subsequent engagements have taken place with some members on specific lines of ongoing work (transport, mobility, poverty mapping, new official statistics), potential new research (around cities). This has lead to one new collaboration with the University of Tokyo, which was formalized in October 2016 (see Section 3.1.7).

3.3. Electricity big data in Dhaka and Colombo

The Public Utility Commission of Sri Lanka (PUCSL) asked LIRNEasia to carry out a study on the use of electricity in Sri Lanka in relation to planned policy outputs and subsequent outcomes. As part of this work, the big data team was requested to use data analytics to identify specific groups of consumers of different social standings. LIRNEasia has been given access to data from the Lanka Electricity Company (Private) Limited (LECO) and some preliminary analysis was conducted. The next stage will be to obtain data from the Ceylon Electricity Board (CEB) and scale our model to the entire population. The aim is to assess consumer behaviors to tariff changes and analyze the migration to different tariff bands over time in response to tariff changes. The Household Income and Expenditure survey and an independent household survey will also be used to construct household energy demand functions. We envision that this will bring to light how consumers who have different energy requirements and income levels respond to changes in electricity tariffs.

Given the various challenges faced with obtaining MNBD in Bangladesh coupled together with the interest and direction from the Prime Ministers Office, the focus was shifted to the electricity sector from the originally intended mobile network big data. The team in Bangladesh headed by Dr. Moinul Zaber collected data from the Dhaka Electric Supply Company Limited (DESCO) and conducted preliminary analysis of the data at hand. Dhaka's electricity needs are supplied by two companies, namely DESCO and DPDC. While there severe ground challenges even for the electricity research (articulated further in Section 3.6.2), some preliminary analyses of the electricity data has been conducted (see Annex 21). LIRNEasia also conducted capacity development for the lab staff including a customized hands-on training program on Hadoop essentials and in developing an analytical pipeline for dealing with large datasets for the Dhaka research team.

3.4. Building trust and consensus in policymaking through open data: an action research study on electoral district demarcations in Sri Lanka

Electoral reforms are a key component of the constitutional changes envisaged by the government elected in August 2015. The Mixed Member Proportional (MMP) method of elections which is currently on the agenda for Parliamentary elections requires the re-demarcation of present 22 electoral districts to 140 or so smaller districts. As is, Parties representing minorities are able to enter into alliances with major Parties, and collect votes across 6 out of 22 districts where minorities are scattered, to return 15 MPs from those districts. If the current 22 districts are divided into smaller ones, minorities returned as first-pass-the post MPs in the nine districts in question would be reduced to almost nothing, unless some sort of affirmative action is taken in the demarcation process or in the overall electoral reform process.

Dr. Sujata Gamage of LIRNEasia has worked with the Forum Minor Parties in the Parliament of Sri Lanka during the last two years doing 'what-if' simulations with data from past elections. She has published her work as several op-ed pieces in popular media and an in depth article on the topic. In recognition of her work Sujata was appointed in 2016 as an advisor on electoral reform representing the Forum of Minor Parties in the Constitutional Assembly of the Parliament of Sri Lanka. In her advisor capacity Sujata has been able to convince the Steering Committee of the Assembly headed by the Prime Minister to release all the data necessary to do mock demarcations by minor Parties and other stakeholders, so that they will be able to understand and respond to the demarcation process more knowledgeably. Going further, LIRNEasia, together with an affiliated team of GIS experts, has created an open online tool that can be used by all stakeholders to do their own demarcation scenarios. The tool (http://www.opendemarcations.lk/) is expected to make the manual and opaque modus operandi of the Delimitation Commission of Sri Lanka more open. Also the tool is expected to build trust in the process and consensus in the outcomes. Democratic processes in developing countries are often hampered by trust and consensus issues often due to ethnic tensions. An open and transparent process in electoral reforms in Sri Lanka can be a beacon of good practice for the rest of Asia.

An initial stakeholder meeting was held on February 1st 2017 at LIRNEasia to understand user needs. Mr. Shan Prabha from the Tamil Progressive Alliance, Mr. Najah Mohamed from National Front for Good Governance, a progressive Islamic front and Mr. Keerthi Tennekoon representing the Campaign for a Free and Fair Elections (CAFFE) served as the core team of stakeholders. This first demo of the software and series of follow-up consultations with these stakeholders led to a user-friendly interface which allowed a user an instant display of population and ethnic data for each change in a boundary of an electorate (Figure 24).

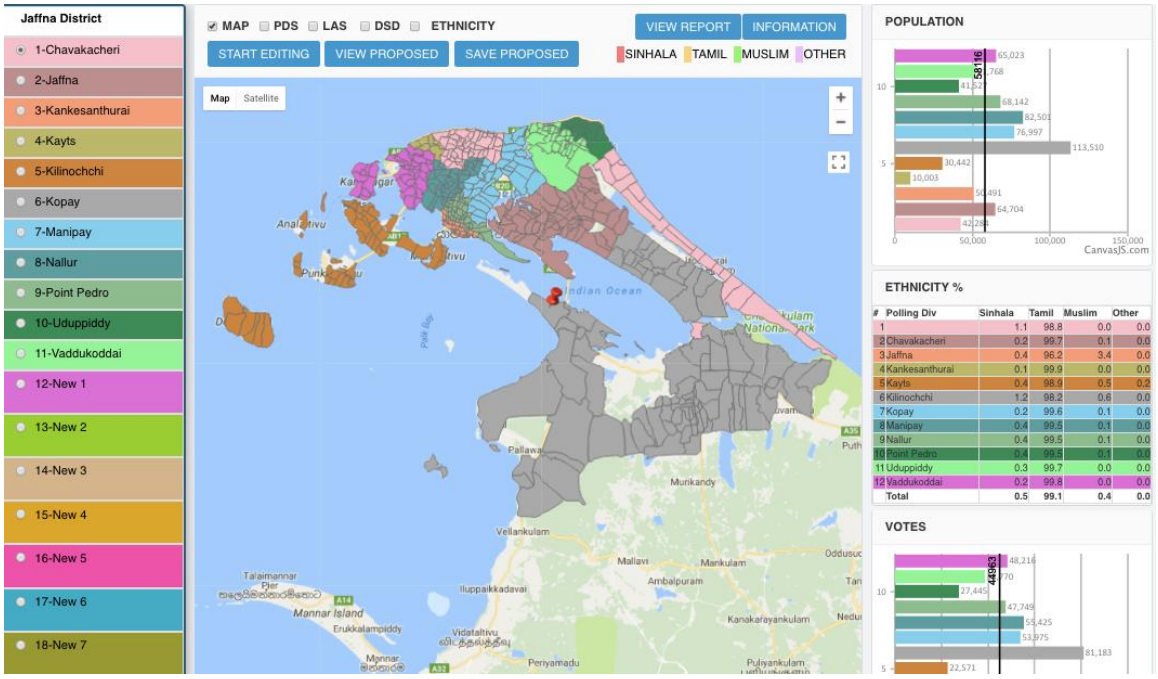


Figure 29:Redistributing DSDs on OpenDemarcations.lk

The tool was officially launched at an interactive session held at SLIDA (Sri Lanka Institute of Development Administration) on 14th September 2017. The launch was well attended with remarks by Ms. Jennifer Hart, Charge de Affairs, High Commission of Canada in Sri Lanka, Hon. Mano Ganesan, Minister for National Co-existence, Dialogue and Official Languages, and Mr. K.Thavalingam, Chairman of the Delimitation Commission providing feedback on the tool.

Developments that have been ongoing since 2014 in the political arena came to a head in 2017 to shift the focus of electoral reforms from the Parliament to Provincial Councils and Local Authorities, the second and third tiers of government, in Sri Lanka. Specifically, a delimitation committee was set up in 2014 to demarcate each of the 340+ local authorities in Sri Lanka into Constituencies or Wards in response to an Act, which presented a Mixed Member Parallel method for local authorities (Act No. 22 of 2012). The Committee completed its work using historical First-past-the-post constituencies, which were in place prior to the change to a proportional representation system in 1978. Perceived malpractices in this demarcation process, perceptions by ethnic minorities in particular, led to the appointment of a new delimitation committee in 2015. As that second committee was completing its mandate, The electoral method for local authorities was further changed to a Mixed Member proportional (MMP) method (Act No. 16 of 2017), and the same MMP method was introduced to Provincial Council elections (Act No. 17of 2017).

will be held on February 12th and appointed another Delimitation Committee to demarcate the Constituencies of Provinces. In this environment it was difficult to engage political Parties on a tool designed for Parliamentary elections. LIRNEAsia quickly repurposed the site to address demarcation issues in provincial councils. In spite of their busy schedules, we were able to convince four registered political Parties and one interest group to work with us to test the system. Observations in this report are based on feedback received from them.

3.4.1. Observations

We monitored the use of the tool by the four political Parties (Akhila Ilankai Tamil Mahasabha, National Front for Good Governance, United National Freedom Alliance, and United Progressive Alliance) and one interest group (All Ceylon Muslim League Youth Fronts) from Sep 14th – December 31st to make the following observations

A total of 93 unique users accessed the system. Since the users were not identifiable we restricted our observations to four Parties and the interest group who worked with us. The largest and the more influential Party of the four did not use the tool. One other Party used the tool as source of information but no further, but the other two Parties and interest group used the tool and submitted their proposals using the graphics and the stats generated by tool.

3.4.1.1. Inclusiveness

Of the four Parties in our stakeholder group, the three small fringe Parties and the fourth interest group were the heaviest users of the system. As they noted, without the tool they would not have been able to access census data and produce the graphics and statistics necessary to present a good proposal to the Delimitation Committee. Therefore, we conclude that one of the major outcomes of a tool of this sort is the inclusiveness it lends to an electoral demarcation process.

3.4.1.2. Trust

We learned that technology itself does not lead to trust in the system. One of the five Parties we worked with was a small but influential Party in the present ruling alliance in Sri Lanka. That Party focused its energy on getting their nominee to the delimitation committee appointed for the purpose. They found it easier to communicate their wishes to their Committee member and trust him/her to present their interests. They would not have trusted an independent commission to look after their interests even if the Commission or the participants were armed with best technology.

3.4.1.3. Consensus

In the present system of demarcation used in Sri Lanka and most Commonwealth countries including India and UK, independent commissions are entrusted with demarcation. They solicit proposals but final determination is theirs. The latest delimitation committee as mandated in Act No. 17 of 2017 is a departure in that the ruling Party gets to formulate a delimitation Committee which included the three existing independent commissioners plus two other government nominees. Further the demarcations have to be approved by a Parliament giving a chance to other political stakeholders to voice their concerns. If the Parliament cannot agree, the matter would be referred to a Committee chaired by the Prime Minister. This method is in fact a mix of three prevailing methods of electoral boundary determinations – (1) independent Commissions as in Commonwealth countries (3) A Commission representative of various

political interests as in New Jersey and several other States in the USA and (3) The practice of incumbent government having the authority to delimit or determine boundaries as in most States in USA.

In the process in place in Sri Lanka, there are three possible stages in the determination of electoral boundaries and there can be sticking points at any of these stages. In anticipation of such a scenario we tested what has come to be known as the "I-cut- you-freeze" method (Pegedne et al)[2]. According to the I-cut-you-choose method, in a state or province with 20 electoral districts, if two Parties A and B are deadlocked, say, Party A would be asked to present a map of its choice to Party B inviting them to freeze one of A's selections and redistrict the other 19. Then Party A would get a chance to freeze a second district and redistrict the remaining 18. According to Pegden et al., after 19 rounds, it is theoretically possible to obtain a demarcation, which is satisfactory to both Parties. An open demarcation tool accessible to both is essential to test the theory. Our preliminary tests with two of small Parties with different views about the demarcation of District of Colombo into 20 electorates showed the utility of opendemarcations.lk tool in such a consensus building exercise.

Open data tools such as opendemarcaitons.lk;

- May not make a significant impact on the trust issue. Because large- or medium-sized parties also seem to prefer to work through persons representing their interests in the Boundary Commissions, making technology somewhat redundant for those parties, BUT,
- Will make the electoral district demarcation processes more inclusive because of access to data and capabilities for data analysis and graphical representation allowed by the tool, AND
- Can be valuable in bringing about consensus between two competing interests by serving as a tool to apply I-cut-you-choose method of consensus building.

3.5. Tanzania

An opportunity arose in Tanzania to help local actors develop capacity to utilize mobile network big data for public purposes. After an initial meeting in Dar Es Saalam in September 2016 with the Ifakara Health Institute and COSTECH, a workshop was planned for early 2017 for academics, researchers, and government officials to learn from our experiences and to then to develop a proposal. The dates have been pushed back at their request and will scheduled during the summer of 2017.

3.6. Challenges

3.6.1. Human resources

One of the principal challenges continues to be human resources. Data scientists are in short supply. LIRNE*asia* has been able to overcome these challenges by building teams with complementary skills: computer science & data mining, statistics, and domain knowledge. However, attracting staff with strong computer science skills remains difficult. Those coming out of Sri Lanka universities are often attracted to higher paying private sector jobs. Applying their skills and knowledge for developmental policy problems is still relatively unknown. LIRNEasia has been engaged in longer-term enlightenment in the local universities. By and large what has

[2] Wesley Pegden, Ariel D. Procaccia, And Dingli Yu (2017). A Partisan Districting Protocol with Provably Nonpartisan Outcomes . Arxiv:1710.08781 [Cs.Gt]

worked well is hiring staff who have a strong prior interest in going for a PhD abroad. For people with such interests, LIRNEasia is the only place that allows them to build a good research profile that can be leveraged to get into good PhD programs abroad. So far four have already left for PhD programs (University of Oregon, Tokyo Institute of Technology, University of Michigan, and The University of New South Wales. As such we do not expect to hold on to good researchers for more than two years, which is roughly the time needed for them to build relevant research experience for graduate school. But also there is a demand for our researchers from the private sector, which is more difficult to manage. LIRNEasia has developed an internal curriculum for new staff hired for the big data research project and this is continuously being refined. This has to a certain extent reduced the effects of the staff churn amongst the big data team.

3.6.2. Setting up Data Analytics in Dhaka, Bangladesh

The Bangladesh research component has been very challenging, not least because considerable manpower had to be (and continues to be) expended to work out data access and conduct data extraction. Initial plans of obtaining data from mobile network operators ran into hurdles partly out of changes in Bangladesh's legal landscape. One such aspect was that subsequent to the start of this research project, the mobile operators in Bangladesh had their licenses renewed. The new licenses included provisions that limited their ability to share data. Multiple interactions with the Prime Minister's Office (PMO) lead to electricity data being the first big data that LIRNEasia analyzed. The PMO facilitated access to data on electricity consumption (i.e. at household level) as well as generation (at sub-station level). We were in the process of negotiating research parameters with the government when the terror attacks occurred in July, that caused further set backs as national security took precedence. Work eventually commenced and a data analytics lab was set up within the University of Dhaka, initially working with household level electricity data. Even with the electricity research the ground situation in Dhaka has been quite challenging. Dhaka's electricity needs are supplied by two companies, namely DESCO and DPDC. There were severe bottlenecks at the zonal office level. Within DESCO there are 16 zones responsible for different geographic locations. The autonomous organization structures have caused a hindrance in the process, in that the heads of the zonal offices do not always see the value and potential of the research and do not wish to allocate the necessary resources required to extract the data. While this was not the case in all zonal offices the problems overall caused severe delays in data collection. Given the highly politicized environment in which the companies operate, we have had to tread with caution so as to not cause any negative effects for our research.

While there was some government funding potentially available towards the early part of 2017, we did not pursue it. Firstly this would have even further delayed getting the lab off the ground. Secondly we wanted to utilize the initial electricity analyses to leverage access to MNBD. Whilst PMO is interested in being part of this broader initiative, we decided to leverage their desire to be officially part of the analytics lab to obtain MNBD data. As of the close of the project we still have not been able to get MNBD. These difficulties despite our steadfast interest and action, does leave some doubt about the future of big data research in Bangladesh. We are continuing to explore avenues including discussing with the government.

4. Project outputs and dissemination

Since inception, the big data team at LIRNEasia has carried out its research effort by consciously documenting and writing papers on the findings and methods used. This section lists a summary of the outputs in terms of reports, publications in journals and conferences, the outcomes in terms of media coverage of our work, contributions made to policy makers or government institutions, interactions with academia and presentations made at different forums. Full-length papers can be found in the Annex as referred to below.

4.1. Reports and publications

The reports and publications from the period March 2015 are listed below:

Authors and Title	Status	Annex #
Lokanathan, S., Kreindler, G., de Silva, N. D., Miyauchi, Y., Dhananjaya, D., & Samarajiva, R. (forthcoming). The Potential of Mobile Network Big Data as a Tool in Colombo's Transportation and Urban Planning.	Published in *Special Issue of Information Technologies & International Development. Vol. 12, No. 2.*	1
Madhawa, K., Lokanathan, S., Maldeniya, D., & Samarajiva, R. (2015). Using mobile network data for land use classification.	Presented at *CPRSouth 2015.*	2
Madhawa, K., Lokanathan, S., Samarajiva, R., & Maldeniya, D. (2015). Understanding communities using mobile network big data.	Presented at *CPRSouth 2015.*	3
Maldeniya, D., Kumarage, A., Lokanathan, S., Kriendler, G., & Madhawa, K. (2015). Where did you come from? Where did you go? Robust policy relevant evidence from mobile network big data.	Presented at *CPRSouth 2015.*	4
Maldeniya, D., Lokanathan, S., & Kumarage, A. (2015). Origin-Destination matrix estimation for Sri Lanka using mobile network big data.	Presented at 13th International Conference on Social Implications of Computers in Developing Countries. Colombo.	5
Lokanathan, S & Gunaratne, R. L. (2015). Mobile Network Big Data for Development: Demystifying the Uses and Challenges.	Published in Communications & Strategies.	6
Samarajiva, R & Lokanathan, S. (2016). Using Behavioral Big Data for Public Purposes: Exploring Frontier Issues of an Emerging Policy Arena.	Published, funded by the New Venture Fund	7
Fernando, L., Perera, A.S., Lokanathan, S., Ghouse, A., Tissera, H. & Samarajiva, R. (2017). Impact of Human Mobility on Spread of Dengue in Sri Lanka	Presented at Netmob 2017, Milan, Italy	8

Bandara, M., Maldeniya, D. & Lokanathan, S. (2017). A Comparison of Geographical Boundaries defined by Human Mobility and Communication in Sri Lanka	Completed. Submitted to Netmob 2017.	9
Kasthurirathna, D., Piraveenan, M., Bandara, M. & Maldeniya, D. (2016). Detecting geographically dispersed overlay communities using community networks	Published in the International Journal of Computer, Electrical, Automation, Control and Information Engineering	10
Anuradha, D., Athuraliya, C.D., De Silva, C., & Lokanathan, S. (2017). Articulating the Functional Boundaries of Urban Areas Leveraging Mobile Network Big Data	Completed. Submitted to Netmob 2017.	11
Fernando, L., Lokanathan, S., Perera, S., Ghouse, A., & Tissera, H. (2017). Improving Disease Outbreak Forecasting Models for Efficient Targeting of Public Health Resources. CPRSouth 2017	Presented at CPR*South* 2017	12
Surendra, A., Lokanathan, S. Fernando, L. & Perera-Gomez, T. (2017). Predicting population-level socio-economic characteristics using Call Detail Records (CDRs) in Sri Lanka. CPRSouth 2017	Presented at CPR*South* 2017`	13
Perera-Gomez, T. & Samarajiva,R. (2017). Bulk Surveillance: Policy Implications	Draft	14
Wijeratne, Y. & Smarajiva, R. (2017). Analyzing Facebook data to understand international connectivity	Complete	15
Athuraliya, C.D. (2017). Automated Traffic Monitoring for Complex Road Conditions	Complete	16
Athuraliya, C.D. (2017). Deep Semantic Segmentation for Built-up Area Extraction and Mapping from Satellite Imagery	Complete	17

Fernando, L., Lokanathan, S., & Perera, A. S. (2017). Forecasting Dengue Outbreaks using Big Data & Machine Learning. 2nd Commonwealth Digital Health Conference & Awards (CDHC&A)	Presented at Commonwealth Digital Health Conference & Awards (CDHC&A)	18
Dharmawardana, K. G. S., Lokuge, J. N., Dassanayake, P. S. B., Sirisena, M. L., Fernando, M. L., Perera, A. S., & Lokanathan, S. (2017). Predictive Model for the Dengue Incidences in Sri Lanka Using Mobile Network Big Data. In Proceedings of 12th IEEE International Conference on Industrial and Information Systems (ICIIS 2017)	Presented at ICIIS 2017	19
Code of Practice for the Secondary Use of Mobile Network Big Data	Complete	20
Zaber, M., Ali, A., Bhyiyan, F., Sayeed, A., Islam, S., & Rakib, N. (2017). Scrutiny of Electricity Billing and Supply Data as a Probable Proxy for Economic Activities: An Analysis of Power Consumption of Dhaka, Bangladesh	Draft	21

4.2. Media coverage

Acknowledgements of research are hard to capture and document. One such opportunity is via mention in media. While we are aware of the fact that this is not all encompassing it provides some visibility to the citations and recognition of our work. The media coverage obtained for the period March 2015 to December 2017 on aspects related to this research project are listed below:

Title	Details
Big data can make South Asian cities smarter	Date: 12th April 2015 Type: Print + Web Name of Newspaper: The Sunday Times Country: Sri Lanka Author: Rohan Samarajiva Link: http://www.sundaytimes.lk/150412/sunday-times-2/big-data-can-make-south-asian-cities-smarter-144237.html
UNDP launches summit on sustainable human development in Sri Lanka	Date: 18th April 2015 Type: Web Name of Site: Colombo Page Country: Sri Lanka Link: http://www.colombopage.com/archive_16A/Apr18_1460987447CH.php
Stakeholders unite to implement SDGs	Date: 19th April 2015 Type: Print Name of Newspaper: Daily FT Country: Sri Lanka Author: Charumini de Silva Link: http://www.ft.lk/article/536999/Stakeholders-unite-to-implement-SDGs

Sri Lanka joins global dialogue to stem looming natural disasters	Date: 23rd April 2015 Type: Print Name of Newspaper: The Island Country: Sri Lanka Author: Steve A Morell Link: http://www.island.lk/index.php?page_cat=article-details&page=article-details&code_title=144078
Promoting foresight and innovation	Date: 24th April 2015 Type: Print Name of Newspaper: The Sunday Observer Country: Sri Lanka Link: http://www.sundayobserver.lk/2016/04/24/spe-hap1.asp
Net neutrality: Over 1 million emails leave Trai in a spot	Date: 24th April 2015 Type: Print + Web Name of Newspaper: Times of India Country: India Author: Sandhya Soman Link: http://timesofindia.indiatimes.com/tech/tech-news/Net-neutrality-Over-1-million-emails-leave-Trai-in-a-spot/articleshow/47032904.cms
Net neutrality: Making sense of a million responses	Date: 24th April 2015 Type: Print + Web Name of Newspaper: Business Standard Country: India Author: N Sundaresha Subramanian Link:http://www.business-standard.com/article/current-affairs/making-sense-of-a-million-responses-115042400044_1.html
Open data culture needed in Sri Lanka – ICT policy expert	Date: 17th May 2015 Type: Print + Web Name of Newspaper: The Sunday Times Country: Sri Lanka Author: Jagdish Hathiramani Link: http://www.sundaytimes.lk/150517/business-times/open-data-culture-needed-in-sri-lanka-ict-policy-expert-148658.html
Big data can make South Asian cities smarter	Date: 24th May 2015 Type: Web Name of Site: SciDev.Net Country: Sri Lanka Author: Nalaka Gunawardene Link:http://www.scidev.net/south-asia/governance/analysis-blog/big-data-can-make-south-asian-cities-smarter.html
Prime Minister Addresses Sri Lanka's First National Summit on Foresight and Innovation	Date: 24th May 2015 Type: Web Name of Site: EarthLanka.net Country: Sri Lanka Author: Saffran Minhar Link: http://www.earthlanka.net/index.php?option=com_content&view=article&id=1491%3Aprime-minister-addresses-sri-lankas-first-national-summit-on-foresight-and-innovation&catid=40%3Afeatures&Itemid=59&lang=en
Sri Lanka's first National Summit on 'Foresight and Innovation for Sustainable Development' kicks off	Date: 25th May 2015 Type: Print Name of Newspaper: Daily FT Country: Sri Lanka Author: Charumini de Silva and Shehana Dain Link: http://www.ft.lk/article/543889/Sri-Lanka-s-first-National-Summit-on--Foresight-and-Innovation-for-Sustainable-Development--kicks-off

Govt. braces to present 4-year economic plan next month	Date: 25th May 2015 Type: Print Name of Newspaper: Daily Mirror Country: Sri Lanka Author: Shabiya Ali Ahlam Link: http://www.dailymirror.lk/109987/Govt-braces-to-present-year-economic-plan-next-month
Foresight and innovation for sustainable human development	Date: 25th May 2015 Type: Print Name of Newspaper: Daily Mirror Country: Sri Lanka Author: Shabiya Ali Ahlam Link: http://www.pressreader.com/sri-lanka/daily-mirror-sri-lanka/20160525/282501477871736
Big data at the heart of smart cities	Date: 15th September 2015 Type: Print + Web Name of Newspaper: Daily Mirror Business Country: Sri Lanka Author: Rohan Samarajiva Link: http://www.dailymirror.lk/87335/big-data-at-the-heart-of-smart-citieshttp://www.dailymirror.lk/87335/big-data-at-the-heart-of-smart-cities
Megapolis Not a License for Megalomania	Date: 1st October 2015 Type: Print Name of print media: Echelon Author: Nalaka Gunawardene
Dengue: Prediction, before Prevention as a Cure	Date: 19th October 2015 Type: Blog Name of site: Design Public Blog Author: Gaurav Chauhan Link: http://designpublic.in/blogs/dengue-prediction-before-prevention-as-a-cure-2/#more-12789http://designpublic.in/blogs/dengue-prediction-before-prevention-as-a-cure-2/ - more-12789
An Open Dialogue on Open Data	Date: 20th October 2015 Type: Web Name of site: ReadMe Author: Mahesh de Andrado Link: http://readme.lk/open-dialogue-open-data/
Lessons from Cellphones on Distributed Wealth	Date: 30th November 2015 Type: Print + Web Name of site: New York Times Author: Sindhya N Bhanoo Link: http://www.nytimes.com/2015/12/01/science/lessons-from-cellphones-on-distribution-of-wealth.html?mabReward=CTM&src=rec&recp=4&_r=0
Mega city to create special zones and new businesses	Date: 1st December 2015 Type: Print Name of print media: Echelon Author: Devan Daniel
Make a call, plan a developing city	Date: 3rd December 2015 Type: Web Name of site: Ozy.com Author: Libby Coleman Link: http://www.ozy.com/fast-forward/make-a-call-plan-a-developing-city/62946http://www.ozy.com/fast-forward/make-a-call-plan-a-developing-city/62946 http://www.ozy.com/fast-forward/make-a-call-plan-a-developing-city/62946

Do we need Big Data to create smart cities?	Date: 21st January 2016 Type: Web Name of Site: Fair Observer Author: Syu-Wen-Chye & Tanvi Mani Link: http://www.fairobserver.com/culture/do-we-need-big-data-to-create-smart-cities-42108/
Government needs master plan for big data	Date:1st February 2016 Type: Print Name of print media: Echelon
LIRNEasia expands knowledge base in big data	Date: 8th March 2016 Type: Web Name of site: Lanka Business Online Link: http://www.lankabusinessonline.com/lirneasia-expands-knowledge-base-in-big-data/
Big data in Sri Lanka: Where numbers meet theory	Date: 12th May 2016 Type: Web Name of site: Techadvisor.lk Author: Navam Niles Link: http://techadvisor.lk/2016/05/12/big-data-sri-lanka-numbers-meet-theory/
VirtusaPolaris sheds light on big data research	Date: 29th May 2016 Type: Web Name of site: Ceylon Business Reporter Link: http://www.lankabusinessonline.com/lirneasia-expands-knowledge-base-in-big-data/
Leveraging Mobile Network Big Data for Developmental Policy: Research papers from the project	Date: 19th September 2016 Type: Web Name of site: Communication Initiative Network Author: Katy Stockton Link: http://www.comminit.com/global/content/leveraging-mobile-network-big-data-developmental-policy-research-papers-project
Big Data can help curb the spread of chikungunya	Date: 22 July 2017 Type: Web Name of Newspaper: benews24 Author: n/a Link: https://opinion.bdnews24.com/2017/07/27/big-data-can-help-curb-the-spread-of-chikungunya/
LIRNEasia CEO Helani Galpaya appointed to GPSDD board	Date: 27th July 2017 Type: Print + Web Name of Newspaper: Daily Mirror Country: Sri Lanka Link: http://www.dailymirror.lk/article/LIRNEasia-CEO-Helani-Galpaya-appointed-to-GPSDD-board-133591.html

Public forum to mark Arthur C. Clarke's birth centenary Sri Lanka 2048: Imagine the Next 30 Years	Date: 08 December 2017 Type: Web Name of site: Ceylon Today Author: n/a Link: http://www.ceylontoday.lk/print20170401CT20170630.php?id=35807
Futuristic event to mark Sir Arthur C Clarke's Birth Centenary	Date: 10 December 2017 Type: Web Name of Newspaper: The Sunday Times Author: n/a Link: http://www.sundaytimes.lk/171210/business-times/futuristic-event-to-mark-sir-arthur-c-clarkes-birth-centenary-271913.html
Sri Lanka 2048: Imagine the Next 30 Years	Date: 12 December 2017 Type: Web Name of Newspaper: Daily Mirror Author: n/a Link: http://www.ft.lk/news/Sri-Lanka-2048--Imagine-the-Next-30-Years/56-645031

4.3. Conferences, Meetings, and Forums

LIRNEasia staff have been invited to participate in multiple forums where our big data research is showcased extensively in the form of presentations and discussions. Whilst we engage with local universities and government agencies regularly it is infeasible to list out all the meetings. Hence this list is summary of selected events/ meetings for the period March 2015 to September 2016

No.	Title	Details
01	**Lessons Learnt Workshop for Countries Designing UN Strategic Development Frameworks (UNDAF)**	Date: 23-27 March 2015 Location: Bangkok, Thailand LIRNE*asia*'s Team Leader – Big Data Research, Sriganesh Lokanathan was invited to brief UN country teams in Asia Pacific on how to leverage big data for the post-2015 development agenda. The series of workshops were organized by the UN Development Group's (UNDG) Asia Pacific Secretariat. Sriganesh showed results from LIRNE*asia*'s ongoing research, as well as articulated the pathways for UN staff (which included country heads of various UN bodies in multiple countries) to start to think about developing data literacy for utilizing new data sources within their respective countries.
02	**Meeting with UNDP**	Date: 20th April 2015 Location: Colombo, Sri Lanka At the request of UNDP, LIRNEasia met with UNDP in Sri Lanka to brief the organization on potential applications of big data research for disaster risk reduction efforts and to identify new sources of data of relevance.

03	**Emerging trends for DRM**	Date: 14-15 May 2015 Location: Ulaanbaatar, Mongolia UNESCAP in partnership with the International Think Tank for Landlocked Developing Countries (ITT-LLDC) held an Expert Workshop on ICT for Promoting Inclusive and Disaster Resilient Development. Sriganesh Lokanathan presented LIRNE*asia*'s ongoing work on disaster risk reduction as well applications of LIRNE*asia*'s big data research in the area of DRR.
04	**Panel sessions at ICTD**	Date: 18th May 2015 Location: Singapore LIRNE*asia* organized a panel session (titled "Practitioner perspectives: Taking big-data evidence to development policy) at ICTD 2015. Rohan Samarajiva moderated the panel that included: LIRNE*asia*'s big data team lead Sriganesh Lokanathan; Data Scientist Prof Joshua Blumenstock from University of California, Berkeley (and a LIRNE*asia* collaborator); Sri Lanka transportation specialist Prof Amal Kumarage (University of Moratuwa and LIRNE*asia* collaborator); and Mr. Ratnapriya Wickramasinghe, Assistant Director at the Sri Lanka Department of Census and Statistics. The panel offered perspectives on the challenges of taking big data evidence into developmental policy.
05	**3rd International Open Data Conference 2015**	Date: 28-29 May 2015 Location: Ottawa, Canada LIRNE*asia*'s Founding Chair, Rohan Samarajiva participated in this two-day conference and spoke on a panel titled "Corporate Data Sharing for the Public Good: Shades of Open". Rohan shared LIRNE*asia* experiences based on our ongoing big data research.
06	**ESRC-DFID-NERC Big Data for Resilience Workshop**	Date: 5th June 2015 Location: London, UK LIRNE*asia*'s Founding Chair, Rohan Samarajiva participated in this workshop presenting LIRNE*asia*'s report on "Leveraging mobile network big data for disaster risk reduction: minimizing harms and facilitating access" to an audience of researchers and academics exploring the use of big data for resilience.
07	**DFID meeting**	Date: 12th June 2015 Location: London, UK LIRNE*asia*'s Founding Chair, Rohan Samarajiva presented LIRNE*asia*'s ongoing Big Data for Development Research to staff at DFID.
08	**Policy-Making in the Big Data Era: Opportunities & Challenges**	Date: 15-17 June 2015 Location: Cambridge, UK LIRNE*asia*'s ongoing big data for development research was presented at this conference by Rohan Samarajiva (Founding Chair, LIRNE*asia*) . The audience included researchers, academics, and policy makers from around the world.

09	**Workshop on Integrated Land Use Transport Modeling Practices in Sri Lanka and around the World**	Date: 11th August 2015 Location: Mount Lavinia, Sri Lanka Sriganesh Lokanathan (Team Leader – Big Data Research) presented LIRNEasia's research related to transport and urban planning at this workshop. Practitioners, consultants, academics and researchers from around the world were present.
10	**Big Data for the analysis of Digital Economy and Society**	Date: 22nd September 2015 Location: Sevilla, Spain The European Union Joint Research Center-Institute of Prospective Technological Studies convened a workshop in Sevilla on "Big Data for the analysis of Digital Economy and Society" on 22 September 2015. Rohan Samarajiva was invited to speak on "Big data for development: New opportunities for emerging markets."
11	**ITU Telecom World**	Date: 14th October 2015 Location: Budapest, Hungary LIRNEasia CEO Helani Galpaya spoke of LIRNEasia's current big data research and LIRNEasia inputs in to the Western region Megapolis plans in Sri Lanka, at the session titled "Data-rich, decision-making-poor: how to use big data for improved government action".
12	**Big Data and Privacy and UN Global Pulse**	Date: 23rd – 24th October 2015 Location: Hague, Netherlands Prof. Rohan Samarajiva participated in a two-day face-to-face meeting of the Privacy Advisory Group of UN Global Pulse. The objective was for the advisory group to consider the risks associated with the misuse of big data along with the risks and harms of non-use of big data.
13	**Presentation at the Internet Governance Forum (IGF)**	Date: 12th November 2015 Location: Joao Pessoa, Brazil Prof. Rohan Samarajiva, founding chair of LIRNEasia was invited to speak at the session titled "Big Data for Development: Privacy, Risks and Opportunities?" organised by SIDA and UN Global Pulse.
14	**Presentation at Institute of Technology and Society (ITS) of Rio de Janeiro**	Date: 16th – 17th November 2015 Location: Joao Pessoa, Brazil Following the IGF, was another event organised by the Institute of Technology and Socienty of Rio de Janeiro from the 16th – 17th November 2015 for which Prof. Rohan Samarajiva was invited to. He made a presentation on public interests of big data that highlighted LIRNEasia research on population density variations of Colombo and the patterns observed from transportation surveys. The majority of the participants had a legal background.

15	**LIRNEasia research at the Urban Development Authority of Sri Lanka**	Date: 4th December 2015 Location: Colombo, Sri Lanka LIRNEasia's ongoing research on leveraging mobile network big data for urban and transportation planning was presented at an event organized by the Urban Development Authority and the Young Planners Association of Sri Lanka.
16	**Big Data for Sustainable Development at ESCAP**	Date: 14th – 15th December 2015 Location: Bangkok, Thailand Prof. Rohan Samarajiva spoke at an event organised by UN ESCAP that focused on the role of big data to measure the progress of the recently finalised sustainable development goals.
17	**Big Data Advisory Council**	Date: 18th January 2016 Location: Colombo, Sri Lanka The first meeting of the data analytics advisory council was held on the 17th January 2016 in Colombo. Data scientists present at the meeting provided insights and commented on the current work flow and future plans. There was also discussion in relation to sustainability in terms of human and financial capital.
18	**Public Lecture on Big Data Application in Government: Digitization for Citizen Services - Scientific Use of Data To Improve Citizens' well-being**	Date: 18th January 2016 Location: Colombo, Sri Lanka LIRNEasia in collaboration with Ministry of Telecommunication and Digital Infrastructure, Sri Lanka Institute of Development Administration (SLIDA) and ICT Agency of Sri Lanka organized a public lecture for government officials titled "Big Data Application in Government: Digitization for Citizen Services - Scientific Use of Data To Improve Citizens' well-being." The lecture was delivered by Dr. Prabir Sen who is a member of LIRNEasia's Data Analytics Advisory Council and who was the former Chief Data Scientist for Singapore's Infocom Development Authority (IDA).
19	**Big Data for Development at Global Affairs, Canada**	Date: 11th March 2016 Location: Ottawa, Canada Sriganesh Lokanathan made a presentation on the responsible use of big data for development. One of the key messages shared was the importance of understanding the local context and developing in-situ capabilities in the global south.
20	**Big Data for Development at IDRC**	Date: 14th March 2016 Location: Ottawa, Canada Sriganesh Lokanathan spoke of LIRNEasia's work in using big data to influence policy largely in the context of developing Sri Lanka. Some outputs of our research were shared. The need to understand local context and develop expertise in the global south was articulated.

21	**Planning meeting on forecasting dengue in Sri Lanka using big data**	Date: 6th May 2016 Location: Colombo, Sri Lanka LIRNEasia and the Health Informatics Society of Sri Lanka jointly convened a planning meeting on building better models for forecasting the propagation of infectious disease such as dengue in Sri Lanka. The meeting was intended to lay the foundation for a multi-disciplinary collaboration engaging health informatics specialists, epidemiologists, and data scientists to identify research priorities and opportunities. (Report annexed)
22	**Foresight & Innovation for Sustainable Human Development**	Date: 24th -25th May 2016 Location: Colombo, Sri Lanka LIRNEasia was a core partner for Sri Lanka's first national summit on "Foresight & Innovation for Sustainable Human Development" that was convened by UNDP and the Ministry of National Policies and Economic Affairs. The summit brought together more than 300 people from government, private sector, and civil society from all over the country. Developing foresight and fostering innovation is a priority for the government and underscored by the Prime Minster's attendance at the event.
23	**21st ITS Biennial Conference**	Date: 27th June 2016 Location: Taipei, Taiwan Rohan Samarajiva gave a talk on "Delineating administrative boundaries using mobile network big data" at a panel on Big Data uses and policies at the 21st Biennial ITS Conference.
24	**Lecture at Department of Census and Statistics**	Date: 14th July 2016 Location: Colombo, Sri Lanka Dr Ayumi Arai from the University of Tokyo's Center for Spatial Information Science and LIRNEasia Research Fellow spoke on her ongoing Dynamic Census research work in Bangladesh which utilizes mobile network big data and official statistics to provide spatio-temporal insights on the socio-economic and demographic characteristics of the population at high granularity and high frequency.. The talk was for the senior staff of the Department of Census and Statistics (DCS) Sri Lanka, as preamble to a longer discussion with the department to collaborate with LIRNEasia and our partners on big data and official statistics in Sri Lanka.
25	**2016 Tech Policy Lab Global Summit**	Date: 2nd – 3rd August 2016 Location: University of Washington LIRNEasia's Founding Chair and Team Leader, Big Data Research attended the 2016 Tech Policy Lab Global Summit to identify grand technology policy challenges. LIRNEasia leveraged the broad expertise and conversations at this invitation-only event to refine our thinking on future research priorities for LIRNEasia's big data research.

26	**International Conference on Big Data for Official Statistics**	Date: 28th August 2016 Location: Dublin, Ireland Organized by the Census and Statistics Organization (CSO) Ireland and the United Nations Statistics Division, the International Conference on Big Data for Official Statistics was held in Dublin from the 30th August to the 1st of September 2016. The conference explored areas of access and partnerships, capacity building for innovation and the role of big data to measure/monitor the SDGs. LIRNEasia Research Manager Thavisha Gomez attended this conference and made some interventions based on our ongoing research. This will most likely be critical event for us in future as our work with the Department of Census and Statistics Sri Lanka scales up.
27	**Talk at Pulse Lab**	Date: 1st September 2016 Location: Jakarta, Indonesia LIRNEasia's Team Leader for Big Data Research was invited by Pulse Lab Jakarta to share our experiences in engaging policy makers, with their staff. Principally we shared our mode of operation and engagement strategy for enlightening policy makers.
28	**Talk at Ifkhara Health Institute**	Date: 11th September 2016 Location: Dar Es Salaam, Tanzania While in Zanzibar for other work, LIRNEasia's Founding Chair and Team Leader, Big Data Research were requested to meet with officials in Dar Es Salaam to advise them on how to leverage mobile network big data for development purposes. LIRNEasia met with officials from the Ifkhara Health Institute and discussed potential research opportunities of relevance to Tanzania. This will be followed up by a workshop at a future date to both familiarize them with how mobile network big data can be utilized for development purposes as well as to brainstorm potential collaborative research.
29	**International Open Data Conference 2016**	Date: 06th October 2016 Location: Madrid, Spain Sriganesh Lokanathan, LIRNE*asia*'s Team Leader for Big Data Research was invited to a panel discussion on data for social good and the opportunities for joint action where he emphasized the need for a perspective shift – from a focus on the type of data used to the problem that needs to be solved.
30	**Self-Organizing Conference on Machine Learning 2016**	Date: 07th-08th October 2016 Location: San Francisco, USA Big Data Researcher, CD Athuraliya was competitively selected to attend the Self-Organizing Conference on Machine Learning 2016 organized by OpenAI where his submission was selected for a poster presentation.

31	**Talk at the Joint Research Center of the European Commission**	Date: 11th October 2016 Location: Seville, Spain LIRNE*asia*'s Team Leader for Big Data Research, Sriganesh Lokanathan was invited to share his insights on the organization's experience in using big data for public purposes with the staff of the Joint Research Centre of the European Commission and other colleagues from Ispra, Italy who joined in via video conference.
32	**Commonwealth Medical Conference**	Date: 15 October 2016 Location: Colombo, Sri Lanka LIRNE*asia*'s Founding Chair, Prof. Rohan Samarajiva was invited to share insights from their ongoing work on using big data to predict the spread of infectious diseases at a session chaired by the head of the Epidemiology unit of Sri Lanka's Ministry of Health.
33	**Lecture at University of Jayewardenepura**	Date: 18th October 2016 Location: Nugegoda, Sri Lanka As part of LIRNE*asia*'s continued effort to engage with local universities for their big data work, LIRNE*asia* team leader for big data research, Sriganesh Lokanathan conducted a lecture for students at the University of Sri Jayewardenepura on LIRNE*asia*'s research in the big data for development space.
34	**Data Mining Validation Workshop**	Date: 09th November 2016 Location: Brussels, Belgium Sriganesh Lokanathan, LIRNE*asia*'s Team Leader for Big Data Research was invited to provide his input on a proposed data mining framework for the European Commission's DG for Research & Innovation.
35	**Sri Lanka Association for the Advancement of Science (SLAAS) Panel on Big Data for Development**	Date: 09th December 2016 Location: Colombo, Sri Lanka The University of Sri Jayewardenepura invited LIRNE*asia* to organize an hour-long panel discussion on big data for development during the 72nd AGM of the Sri Lanka Association for the Advancement of Science. LIRNE*asia* Founding Chair, Prof. Rohan Samarajiva moderated a panel consisting of academics, private sector and civil society. The panelists provided an overview of big data and its role in development, issues around data access and in particular dealing with capacity issues.
35	**UN World Data Forum**	Date: 16th January 2017 Location: Cape Town South Africa Sriganesh Lokanathan, LIRNE*asia*'s Team Leader for Big Data research was given the opportunity to introduce LIRNEasia's big data for development research at the 1st UN data forum where he highlighted partnerships (ongoing or being initiated) with government.

36	**UN World Data Forum**	Date: 17 January 2017 Location: Cape Town, South Africa LIRNE*asia*'s Team Leader for Big Data research, Sriganesh Lokanathan was invited to be a panelist in one of the sessions to speak on the role of big data in supporting the measurement of the progress on the SDG targets.
37	**Expert Stakeholder Meeting, Global Pulse/GSMA**	Date: 17 January 2017 Location: Cape Town, South Africa Sriganesh Lokanathan, LIRNE*asia* Team Leader for Big Data research and Thavisha Gomez, Research Manager were invited to a meeting that sought to obtain feedback on key themes stemming out of a joint study by UN Global Pulse and GSMA Foundation, "State of Mobile Data for Social Impact" as well as validate the recommendations that would be suggested in the report.
38	**UN Data Innovation Lab**	Date: 19th January 2017 Location: Cape Town, South Africa LIRNE*asia* Research Manager, Thavisha Gomez was given the opportunity to share LIRNEasia experience in entering data partnerships and the challenges associated with the same at a workshop, co-hosted by UN Global Pulse, which centred on designing data capacity within the UN system
39	**Lecture at University of Dhaka**	Date: 3-4 February 2017 Location: Dhaka, Bangladesh LIRNE*asia* Team Leader for Big Data Research, Sriganesh Lokanathan shared LIRNEasia's experience in conducting policy relevant research on big data for development in Sri Lanka, with both faculty and students at the University of Dhaka.
40	**International Workshop on Mining for Actionable Insights in Social Networks**	Date: February 2017 Location: Cambridge, UK LIRNE*asia* research fellow, Dharshana Kasthurirathna, Ph.D. presented a paper, 'Detecting Geographically Distributed Communities using Community Networks,' at the International Workshop on Mining for Actionable Insights in Social Networks. The paper was co-authored by three LIRNE*asia* research fellows (Dharshana Kasthurirathna, Madhushi Bandara, Danaja Maldeniya) and Mahendra Piraveenan from the University of Sydney. Based on the presentation, there was an invitation to extend the paper to be submitted to a special issue of the Elsevier Information System's journal, with a draft journal paper due in April 2017.
41	**International Conference on Data Revolution for Policy Makers**	Date: 22nd February 2017 Location: Jakarta, Indonesia LIRNE*asia* Team Leader for Big Data research, Sriganesh Lokanathan was invited to speak on a panel "How might we make better use of the digital data resulting from private sector services?" that focused on the development of strategic partnerships.

42	**Global Festival of Ideas**	Date: 3rd March, 2017 Location: Bonn, Germanyhttps://www.youtube.com/watch?v=BagtubH9pKw LIRNEasia Team Leader for Big Data research, Sriganesh Lokanathan was invited to participate in a panel organized by the DataPop Alliance on using big data for social good. His fellow panelists including Bruno Lepri (MIT Media Lab), Nuria Oliver (Vodophone), and Marte Torskenæs (NORAD). Sriganesh shared LIRNEasia's experience in leveraging big data for social good in Sri Lanka and articulated the challenges of such work especially in the areas of achieving policy impact, accessing data, and developing capacity.

43	**Meeting with the Urban Development Authority of Sri Lanka**	Date: 27th April 2017 Location: Battaramulla, Sri Lanka LIRNEasia Team Leader for Big Data research, Sriganesh Lokanathan was invited to a meeting with the deputy directors of the Urban Development Authority (UDA) of Sri Lanka to discuss the potential operation of a research and innovation unit within the UDA that would explore new analytics and data sources.
44.	**Workshop at the Department of Census and Statistics**	Date: 2nd June 2017 Location: Battaramulla, Sri Lanka LIRNEasia convened a workshop to highlight the applications of big data for development as part of it ongoing engagement with the Department of Census and Statistics (DCS). The objective of the workshop was two-fold: (1) discuss ongoing LIRNEasia projects, outlining the potential applications of new data sources, particularly to measure/monitor the Sustainable Development Goals (SDGs), and (2) encourage government exploration of new data sources. Dr. Ayumi Arai of the University of Tokyo also spoke about their ongoing collaboration with LIRNEasia to derive a dynamic census model using traditional survey data and mobile network big data.
45.	**Multiple Meetings at the National Science Foundation**	Date: Multiple meetings Location: Colombo, Sri Lanka LIRNEasia has been providing input to National Science Foundation-led initiative that seeks to develop research that leverages new data sources.
46.	**Roundtable Talk at Carnegie India**	Date: 19th July Location: New Delhi, India LIRNEasia Chair, Prof. Rohan Samarajiva was invited to a roundtable discussion at Carnegie India where he shared on big data development and governance, highlighting the need for "effective, fit-for-purpose public policy and regulation for big data (including algorithms)".
47.	**Meeting with the Millennium Challenge Corporation**	Date: 28th July 2017 Location: Colombo, Sri Lanka LIRNEasia Team Leader for Big Data research, Sriganesh Lokanathan was invited to share LIRNEasia experience on the use of big data for development with staff from the Millenium Challenge Corporation's Sri Lanka program at an event organized at the Prime Minister's office.

48.	**CPRsouth Conference**	Date: 2nd September 2017 Location: Yangon, Myanmar Three members of the big data team were invited to present their papers at the Communication Policy Research Conference in Yangon. Graduate Research Assistant Aparna Surendra spoke about predicting population-level socio-economic characteristics using Call Detail Records (CDRs) in Sri Lanka. Research Manager, Thavisha Perera-Gomez outlined the potential applications of big data to support the sustainable development goals. Lasantha Fernanda, Researcher, shared findings from a study to improve disease outbreak forecasting models for efficient targeting of public health resources.
49	**2017 International Conference on Sustainable Development Goals Statistics**	Date: 5th October 2017 Location: Manila Peninsula LIRNE*asia* Chair, Prof. Rohan Samarajiva was invited to share LIRNEasia insights on the SDGs being amenable to measurement using big data at a session that dealt with the integrating various data sources to monitor the progress towards the SDGs.
50	**SAARC Law Conference 2017**	Date: 28th October 2017 Location: Colombo Sri Lanka LIRNE*asia* Chair, Prof. Rohan Samarajiva spoke on the topic of 'Confluence (or lack thereof) of data analytics' where he outlined the use of data analytics for public purposes, including the investigation of terrorist acts, and urban planning. And touched on privacy-related harms
51	**APCICT-UNESCAP expert group meeting**	Date: 6-7 December 2017 Location: Manila, Philippines LIRNEasia Team Leader for Big Data research, Sriganesh Lokanathan was invited to provide input into a course module on data driven smart government.
52	**Public Lecture at the University of Philippines**	Date: 08 December 2017 Location: Philippines LIRNEasia Team Leader for Big Data research, Sriganesh Lokanathan was invited to delivering a public lecture on big data for development at the University of Philippines.
53	**2017 Global Technology Summit**	Date: 7-8 December 2017 Location:Bangalore, India LIRNEasia Chair, Prof. Rohan Samarajiva participated in a panel on using technology for governance at the Global Technology Summit - an annual event organized by Carnegie India.

54	**Global Voices Summit 2017**	Date: 2 December 2017 Location: Colombo, Sri Lanka Junior big data researcher Yudhanjaya Wijeratne was invited to participate on the starting panel of the Global Voices 2017, an international conference focused on key issues impacting citizen media and the global Internet. Yudhanjaya shared the panel with public intellectual and science communicator Nalaka Gunawardene, with the founder of Groundviews Sanjana Hattotuwa, and with Verite Research's head of media research, Deepanjalie Abeywickrama
55	**Sri Lanka 2048: Imagining the next 30 years**	Date: 15 December 2017 Location: Colombo, Sri Lanka LIRNEasia Team Leader for Big Data research, Sriganesh Lokanathan was invited to speak on the Future of Data at the public forum organized by the Arthur C Clarke Trust to mark the birth centenary of Arthur C. Clarke.
56	**The Arthur C. Clarke Centennial**	Date: 16 December 2017 Location: Colombo, Sri Lanka LIRNEasia Team Leader for Big Data research, Sriganesh Lokanathan, and junior big data researcher Yudhanjaya Wijeratne spoke at the Arthur C Clarke Centennial in Colombo, examining future conflicts in big data and regulation and the ethics and application of artificial intelligence.

4.4. Capacity Building

Through our collaboration with the University of Moratuwa and now with the University of Sri Jayewardenepura we have and continue to mentor many undergraduate and graduate students principally through research projects that are sponsored by LIRNEasia involving some aspect of our big data for development research. These mentorships often involve significant development of data science skills amongst students. As of March 2017, 23 undergraduate students have been mentored with LIRNEasia staff acting as external mentors on their undergraduate theses (six were women). A further 6 students are in the process of being mentored (two are women) for their undergraduate theses. In addition 2 research masters are being conducted through LIRNEasia (one is a woman). Sriganesh Lokanathan, the Team Leader of the Big Data Research acts as the Chairman of the research review committee for one of the theses. Both the University of Moratuwa as well as the University of Sri Jayewardenepura have organized open lectures for LIRNEasia to share the big data for development research.

But within the sphere of big data for development what is also required are informed and discerning consumers of big data research. As such LIRNEasia has and continues to hold workshops, public lectures, and informal meetings and interactions with officials and experts in various policy domains. Since mid 2015 these are often not supply led but rather demand pull with officials and staff at various government departments often reaching out to LIRNEasia for advice on ongoing work.

5. Impact

The impact of this project needs to be viewed in a continuum, starting with the initial exploratory module in the 2012-2015 research cycle that was also funded by IDRC. Since then LIRNEasia's work in leveraging big data for public purposes has been widely recognized not just in Sri Lanka but also globally. For example, the International Telecommunication Union (ITU) acknowledged LIRNEasia's "pioneering" work in South Asia.[3] LIRNEasia has had considerable success in policy enlightenment with respect the use of the big data for public purposes. This has not just been in Sri Lanka but also globally. In Sri Lanka LIRNEasia is now often the first stop when government departments consider how they may leverage big data. LIRNEasia's emphasis of catalyzing big data research has resulted in the University of Moratuwa writing a successful proposal to start a new center of excellence on data science with a strong focus on the public sector and the use of big data for public purposes. The Data and Design Lab at the University of Dhaka was also a directly result of LIRNEasia's involvement. Some select impacts are listed below.

5.1. Ministry of Megapolis & Western Development, Urban Development Authority, and Sri Lanka Strategic Cities Development Project

The Western Region Megapolis Master plan is a major regional urban and transport plan with a range of objectives focusing on systematic transformation of urban agglomerations, managing congestion, development and transformation of physical and institutional infrastructure etc. principally covering the Western Province of the country. The Western Province is the largest province in Sri Lanka by population and responsible for 45% of the country's GDP. Sriganesh Lokanathan, Team Leader of the LIRNEasia Big Data project was invited to be part of the planning process to provide insights from LIRNEasia's big data for development research. Transportation related insights from LIRNE*asia*'s ongoing research were utilized as inputs to inform the development of the plans. Specifically LIRNEasia insights were used to underscore the need to alleviate daily congestion within Colombo city. LIRNEasia insights were also used to support the case for the development of regions in the periphery of the province with strong linkages (as shown by our analyses) to surrounding regions. The engagements have continued into the implantation phase.

LIRNEasia has been engaging regularly with UDA since 2015. As one of the implementation agencies for the WRMPP, LIRNEasia now fields frequent requests for both insights as well as advice in relation to specific projects at the implementation phase. These include not just mobility insights but also in relation to utilization of CCTV for traffic management and enforcement. Most recently the UDA has also initiated a conversation to obtain LIRNEasia's assistance in developing an in-house research and innovation unit that would be able to apply new techniques and leverage new data sources in order to support their activities.

The development objective of the Strategic Cities Development Project for Sri Lanka is to improve selected urban services and public urban spaces in the participating city regions of Sri

[3] ITU. (2017). Measuring the Information Society Report 2017. International Telecommunication Union. Available at https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2017.aspx

Lanka. LIRNEasia was invited to conduct a presentation for their key officials to share our insights based on our big data research. Specific insights on mobility and land use developed by LIRNEasia for the city of Kandy was presented and which often validated their own findings conducted through costly studies. Since then LIRNEasia has fielded a variety of requests for insights from the teams working on the Kandy specific plans.

5.2. Ministry of National Policies and Economic Affairs

LIRNEasia's team leader for big data research serves as an advisor to the newly formulated Analytics unit at the Ministry of National Policies and Economic Affairs. Amongst others, the unit will serve to provide quick insights for cabinet papers, generate quick analyses as required by the Ministry, and also more broadly provide analytical data support for the development agenda of the Government of Sri Lanka

5.3. Department of Census and Statistics

LIRNEasia has been engaging with the Department of Census and Statistics (DCS) since January 2015 when we were invited to talk about our ongoing big data work to their senior officials. Since then a number of meetings have been held to discuss applications of big data to the work of the department as well as to discuss potential collaborations. LIRNEasia has developed very good relationships with the leadership at the department and has over the course of this project gained significant legitimacy and expertise in the eyes of the department in relation to big data.

5.4. Epidemiology Unit, Ministry of Health

The engagement with the Epidemiology Unit of the Ministry of Health continues as we improve the models for predicting the spatial spread of dengue. However dengue is already endemic to most regions in Sri Lanka. Therefore there is limited scope for the use of this work to help prioritize the deployment of public health resources in the fight against dengue. This was not the case when the research was started. But an equally important aspect of this work has been to put collaborations in place should another infectious disease affect Sri Lanka e.g. Zika or Chikungunya. On this the efforts have been successful with our research showing that the models developed for dengue could be quickly modified for some other arboviruses such as zika or chikungunya.

5.5. University of Moratuwa

LIRNEasia started engaging with various departments at Sri Lanka's premier engineering University, the University of Moratuwa very soon after starting the big data for development research practice. This engagement was for a variety of reasons: to support the development of capacity in data science, enlighten the Sri Lankan academic community of the development use cases from big data, and to have access to top talent for conducting the research. In early 2017, a consortia representing various departments at the University put forth a proposal to start a centre of excellence in data analytics with a focus not just on business use cases but more importantly on development use cases. The new center was approved in August 2017 and much of the design of the research is being modeled around LIRNEasia's pioneering research. This represents a big win for LIRNEasia's aims of catalyzing development-oriented work in Sri Lanka and LIRNEasia is committed to helping build the center as one of its foundational collaborators.

5.6. 2030 Vision of Sri Lanka

LIRNEasia's Founding Chair, Rohan Samarajiva sat on a 41-member experts' committee appointed by the President to draft the 2030 Vision of Sri Lanka, that developed a guiding document for Sri Lanka. Scenarios were developed for 2020, 2025 and 2030. The work was undertaken through three clusters, economic, environmental, and social/governance. In addition, a number of cross-cutting and linking themes were identified. LIRNEasia was also part of the core steering committee for the planning of Sri Lanka's first innovation lab.

Currently the team leader of big data research at LIRNEasia sits on the advisory board of this newly constituted innovation lab, which is a joint venture of the Ministry of Technology and Research and the UNDP.

5.7. Access to Information (A2I) Unit of the Government of Bangladesh

LIRNEasia's founding chair Rohan Samarajiva and Research Fellow Moinul Zaber presented LIRNEasia's ongoing big data research work and its applicability to Bangladesh at a meeting convened by the A2I Unit of the Bangladesh Prime Minister's Office. The meeting was attended by the main officials from the Planning Commission, Bangladesh Bureau of Statistics, Bangladesh Computer Council, and the Bangladesh Telecom Regulatory Commission. This meeting was part of LIRNEasia's efforts to establish big data research in Bangladesh. Since then numerous meetings have been held in relation to the ongoing big data research in Bangladesh.

5.8. Other

LIRNEasia's Founding Chair was renewed for another term as part of UN Global Pulse's Data Privacy Advisory Group. As part of this engagement LIRNEasia has provided inputs into aspects related to data privacy and use of behavioral big data for public purposes, based on our ongoing work in Sri Lanka. LIRNEasia's CEO was also appointed to the board of the Global Partnership on Sustainable Development Data (GPSDD).

6. Recommendations

Human resources to conduct such research continue to be a challenge. Data scientists are in short supply, so sufficient time has to be included for on the job training. The unique proposition of a data scientist is that she brings three different skills to a research project: computer science knowledge, statistics, and domain knowledge. Hence LIRNEasia has worked around the issue of limited data scientists by building teams with different skillsets who work closely with each other. An in-house curriculum was developed for internal training purposes, which continues to be refined. Furthermore we have had to plan for flow-through team. It is difficult to attract computer science researchers to a not-for-profit think tank. We cannot fully match private sector salaries. Where we have been successful is attracting people with PhD ambitions (often with 1-2 years of work experience already under their belt). Hence we have often spent significant time in expanding the horizons of students in our universities through talks and other engagements. Collectively the lessons of developing a research practice should (at the conclusion of this project) be a valuable resource for others who wish to conduct such work.

When LIRNE*asia* originally started conducting big data for development research, there was little to no demand for the insights that we were producing. Significant engagement with policy makers and the symbolic environment of specific policy domains has changed the status quo. Thus now LIRNE*asia* fields occasional requests for insights from government departments. Their needs often require quick turnaround and as such LIRNE*asia* has strived to ensure that it remain nimble so as to be responsive to the government's needs as and when they come in. LIRNEasia now fields requests from various government departments for advice on ongoing work especially in leveraging new data sources.

To some extent the broader question remains as to what model best suits enabling the government to be better consumers of big data insights. If governments and policy making are to make immediate use of big data then there is a need for new types of partnerships between government and others like LIRNEasia and/or universities. Sri Lanka for example can say that they are now utilizing big data insights even though the public sector infrastructure and capabilities are still very much in its nascent stages. While these partnerships are required for conducting analytics and utilizing insights from big data, what is equally needed is a rethink of public-private partnerships specifically in relation to data. Ultimately a confluence of actors, actions, and policy windows are needed for the success. An exploration of this in future would be beneficial for those seeking to replicate LIRNEasia's successes, especially light of the challenges we have and continue to face in Bangladesh.