Understanding social connections

Kaushalya Madhawa, LIRNE*asia*

Responsible use of mobile meta-data to support public purposes

Jetwing Lagoon, Negombo

08 August 2014

LIRNEasia

Pro-poor. Pro-market.

This work was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada.

IDRC | CRDI

International Development Research Centre

Centre de recherches pour le développement international

Canada

UKaid

from the British people

Understanding social connections

- Can we identify communities of people using social connections?
- Do communication patterns reveal the socio-economic status of a population?
- Are people's communication and mobility patterns co-related?

Understanding social connections

- Can we identify communities of people using social connections?
- Do communication patterns reveal the socio-economic status of a population?
- Are people's communication and mobility patterns co-related?

What is a community?

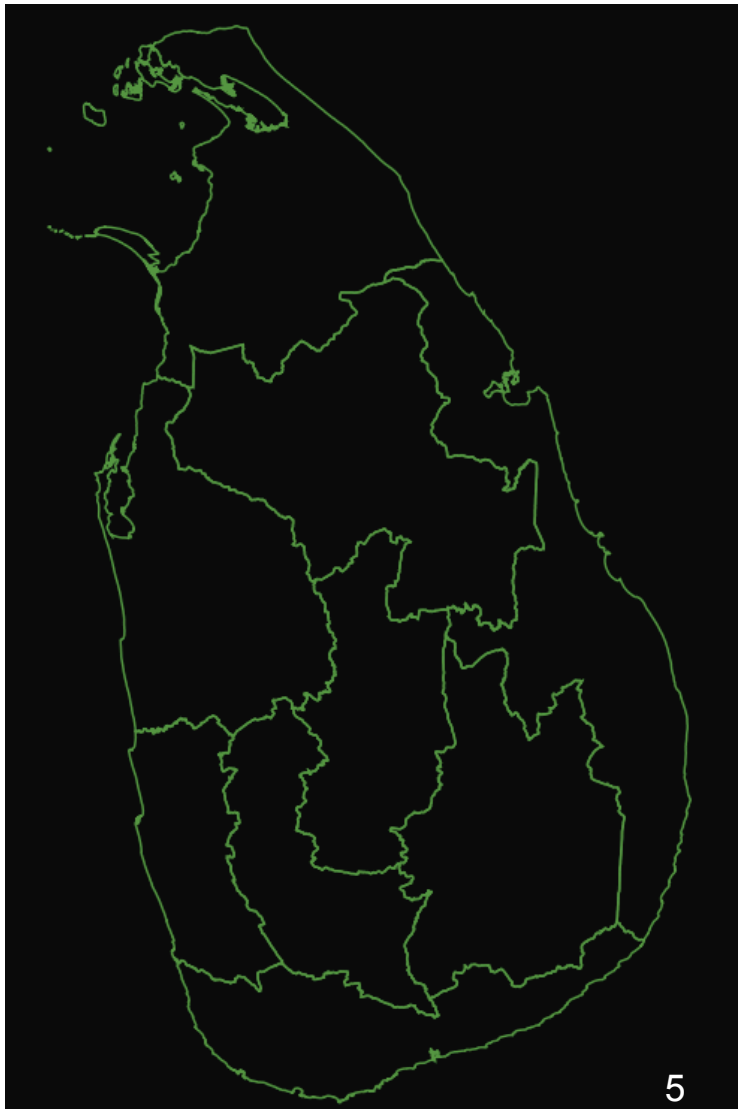
- A community is a group of people who have stronger intra-group connections than inter-group connections

So we asked:

- Do administrative boundaries based on history and geography reflect actual communities in 2014?
- How can mobile network big data be used to identify these actual communities?

What are the administrative boundaries in Sri Lanka?

- 9 Provinces
- 25 Districts
- 331 Divisional Secretariat Divisions (DSD)
- 14021 Grama Niladhari Divisions (GND)



LIRNEasia
Pro-poor. Pro-market.

What are the administrative boundaries in Sri Lanka?

- 9 Provinces
- 25 Districts
- 331 Divisional Secretariat Divisions (DSD)
- 14021 Grama Niladhari Divisions (GND)

LIRNEasia
Pro-poor. Pro-market.

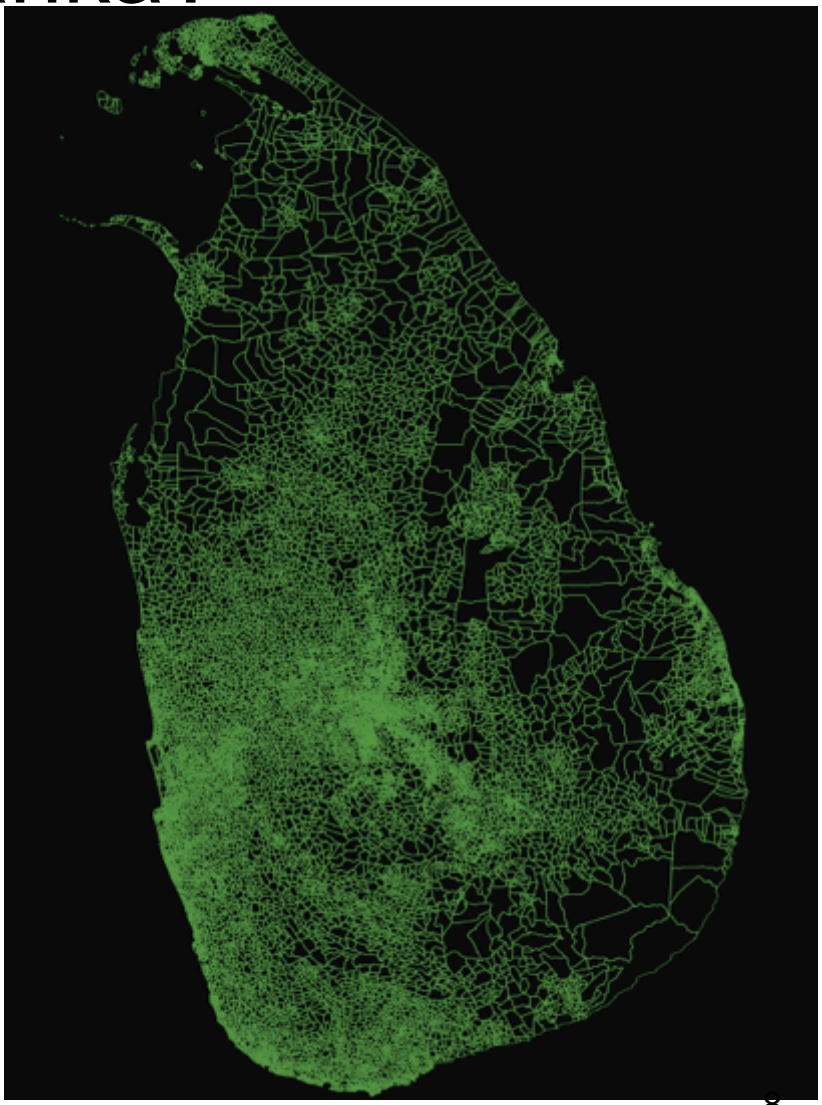
What are the administrative boundaries in Sri Lanka?

- 9 Provinces
- 25 Districts
- 331 Divisional Secretariat Divisions (DSD)
- 14021 Grama Niladhari Divisions (GND)

LIRNEasia
Pro-poor. Pro-market.

What are the administrative boundaries in Sri Lanka?

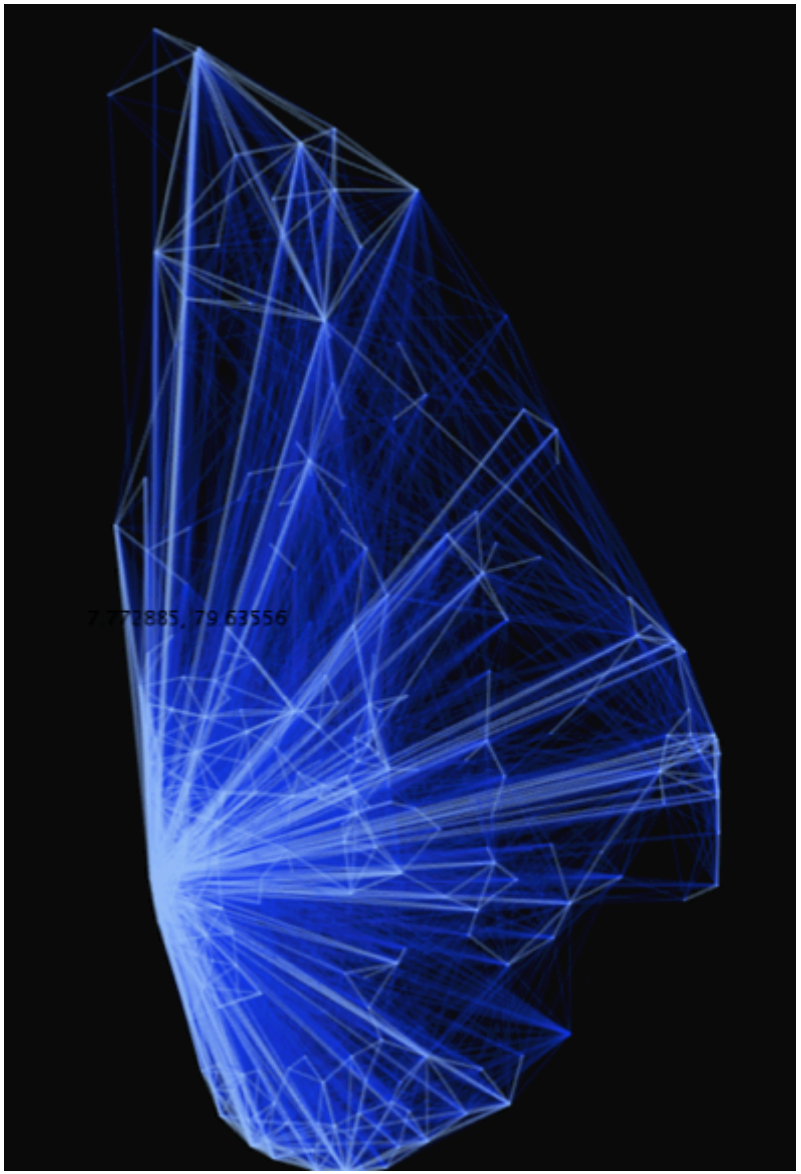
- 9 Provinces
- 25 Districts
- 331 Divisional Secretariat Divisions (DSD)
- 14021 Grama Niladhari Divisions (GND)



LIRNEasia
Pro-poor. Pro-market.

Which DSD talks to which DSD?

- Each link represents the raw number of outgoing and incoming calls between two DSDs



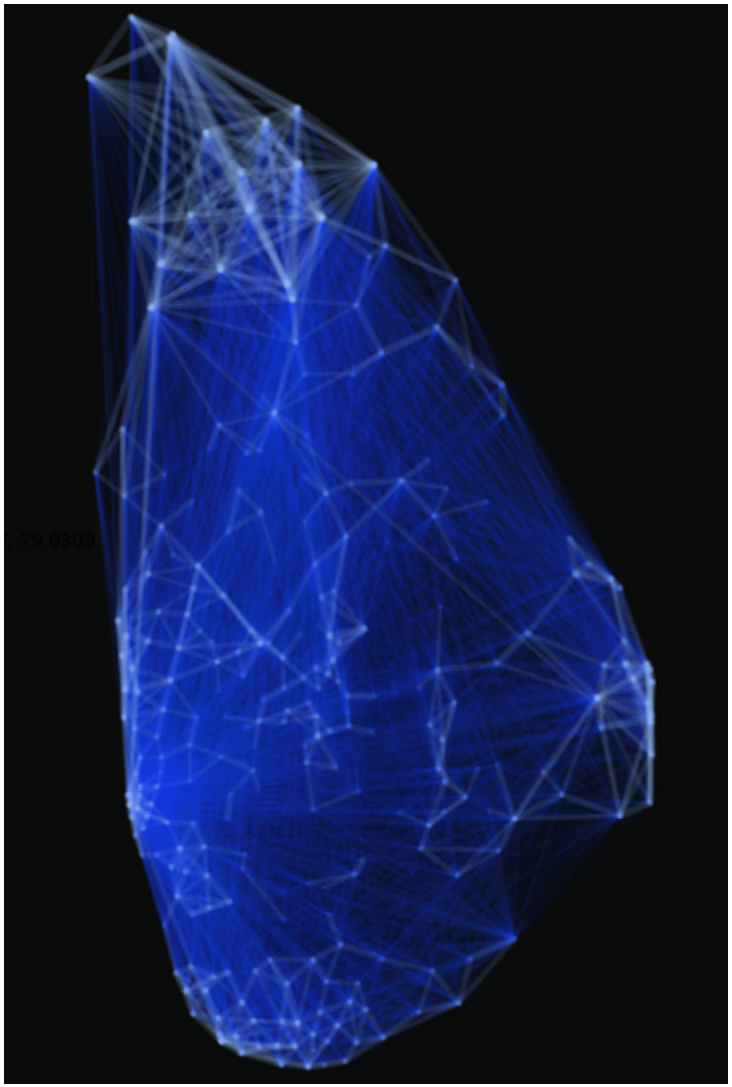
Low → High

No. of calls

A different picture emerges when call volume is normalized by population

$$Normalized\ calls(DS1, DS2) = \frac{No.\ \text{of calls}(\text{DS1}, \text{DS2})}{Pop(DS1) * Pop(DS2)}$$

- Strongly connected components are visible



Low → High

No. of calls

Identifying communities

Identifying communities: methodology

- The social network is segregated such that overlapping connections between communities are minimized.
- Strength of a community is determined by *modularity*
  - Modularity Q = (edges inside the community) – (expected number of edges inside the community)

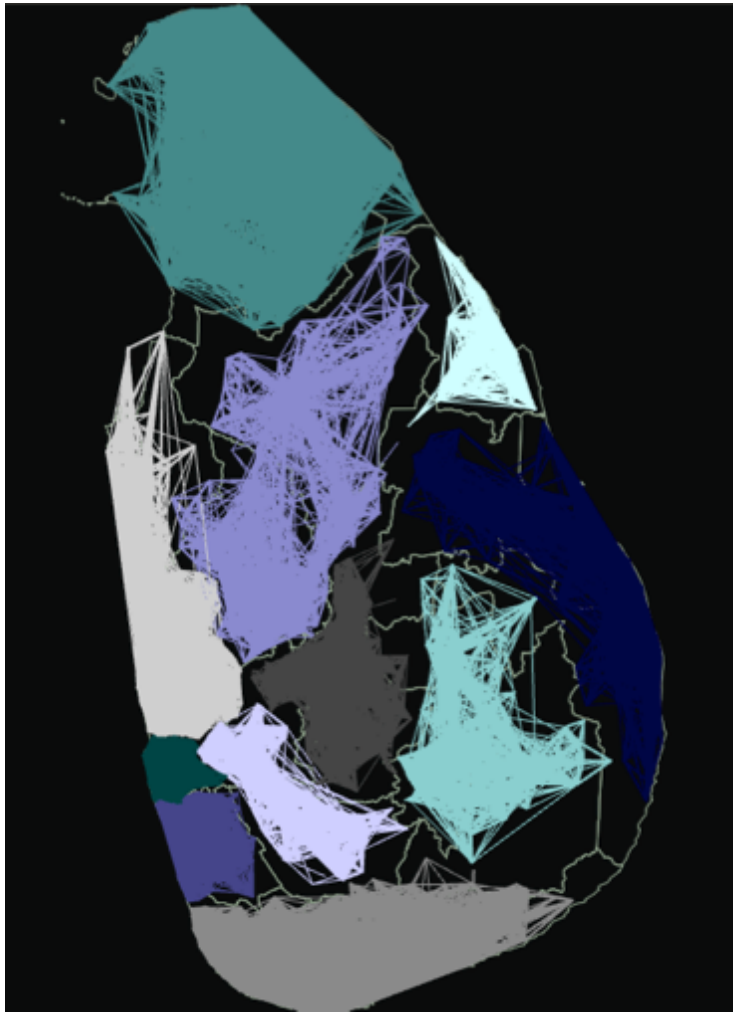
$$Q = \frac{1}{2m} \sum_{a,b} (A_{a,b} - \frac{k_a k_b}{2m}) \delta(c_a, c_b)$$

M. E. J.-Newman, Michele-Girvan, "Finding and evaluating community structure in networks", Physical Review E, APS, Vol. 69, No. 2, p. 1-16, 204.

Actual communities found by the algorithm

- The communities we found are centered around geographic neighbourhoods.

- For Sri Lanka, the optimal number of communities discovered by the algorithm was 11

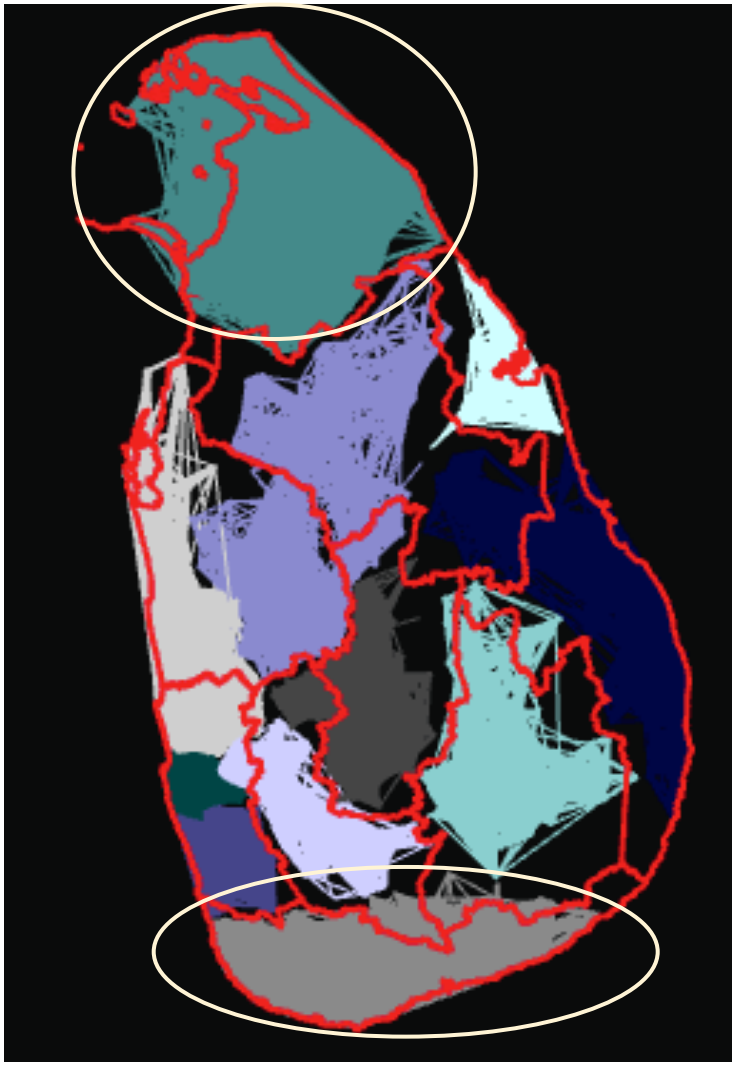


LIRNEasia
Pro-poor. Pro-market.

Actual communities found by the algorithm

- Southern and Northern provinces have the highest similarity to their respective provincial boundaries.

- Colombo district is clustered as a single community and Gampaha is merged with North Western Province

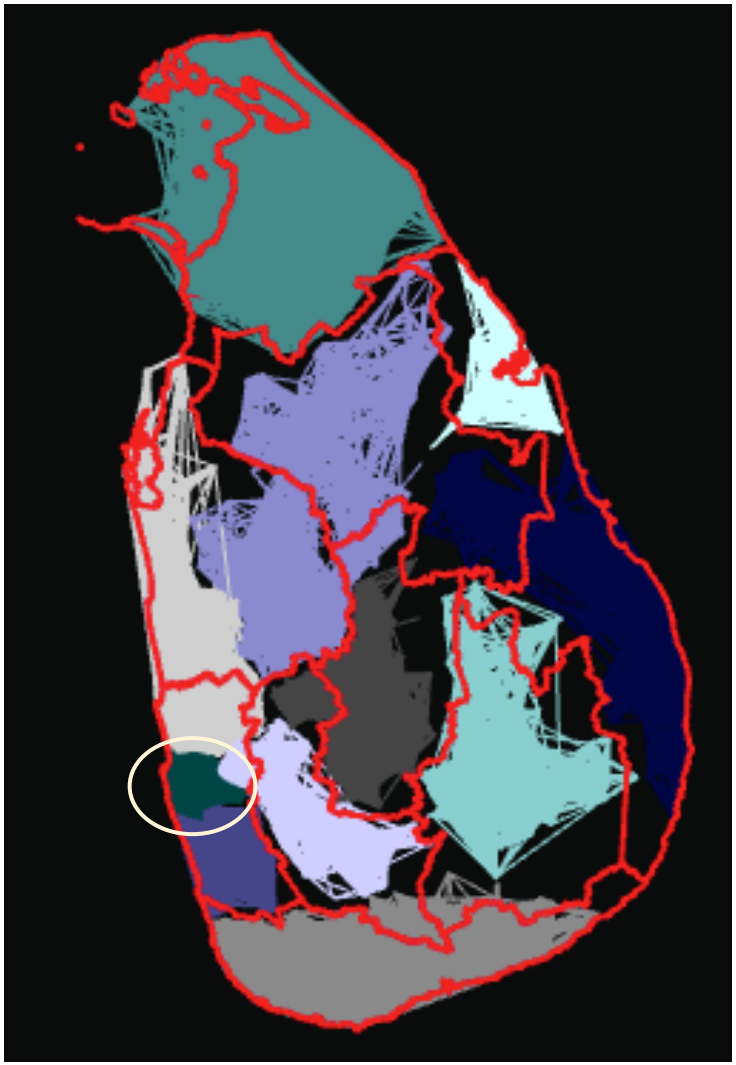


LIRNEasia
Pro-poor. Pro-market.

Actual communities found by the algorithm

- Southern and Northern provinces have the highest similarity to their respective provincial boundaries.

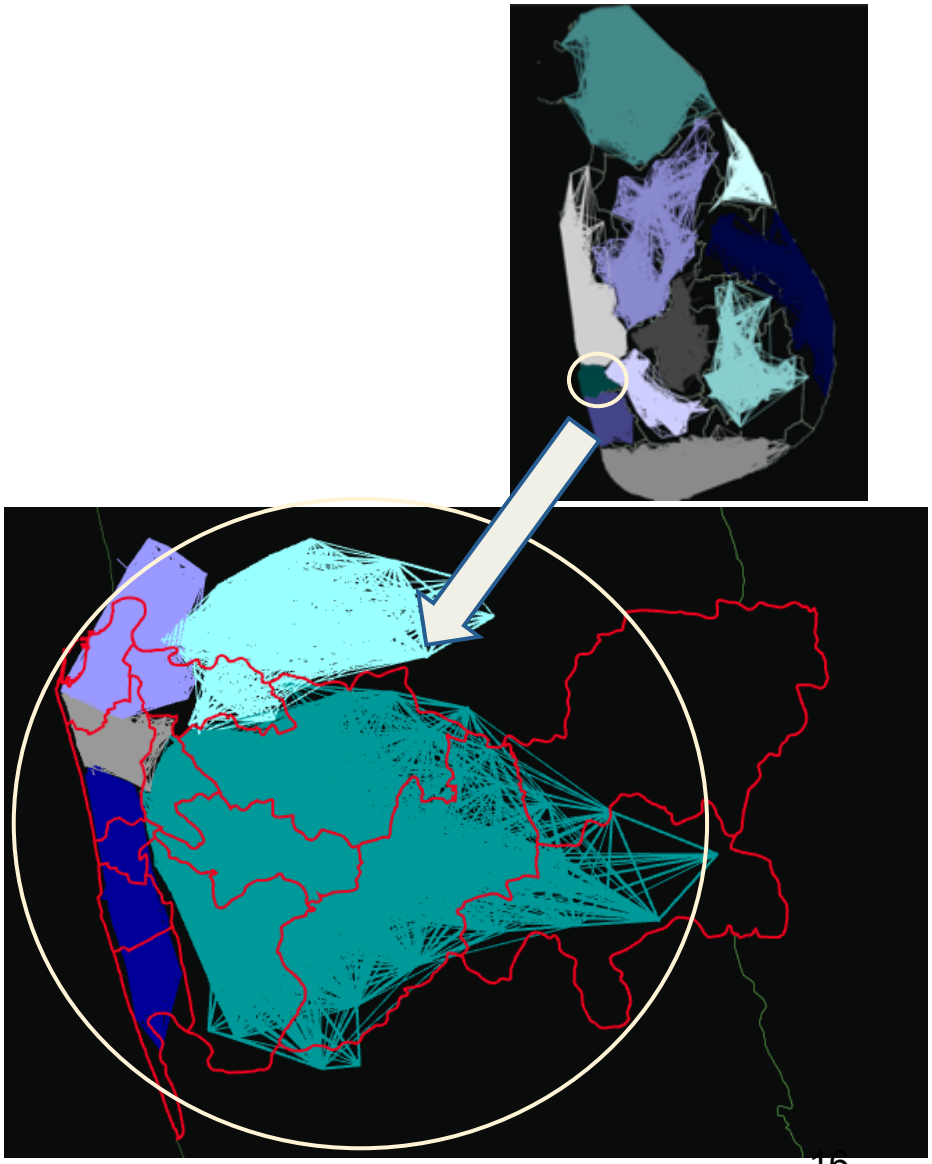
- Colombo district is clustered as a single community and Gampaha is merged with North Western Province



LIRNEasia
Pro-poor. Pro-market.

Zooming into a community

- Community detection is done within identified communities
- These communities are less less related to DSD boundaries in Colombo district



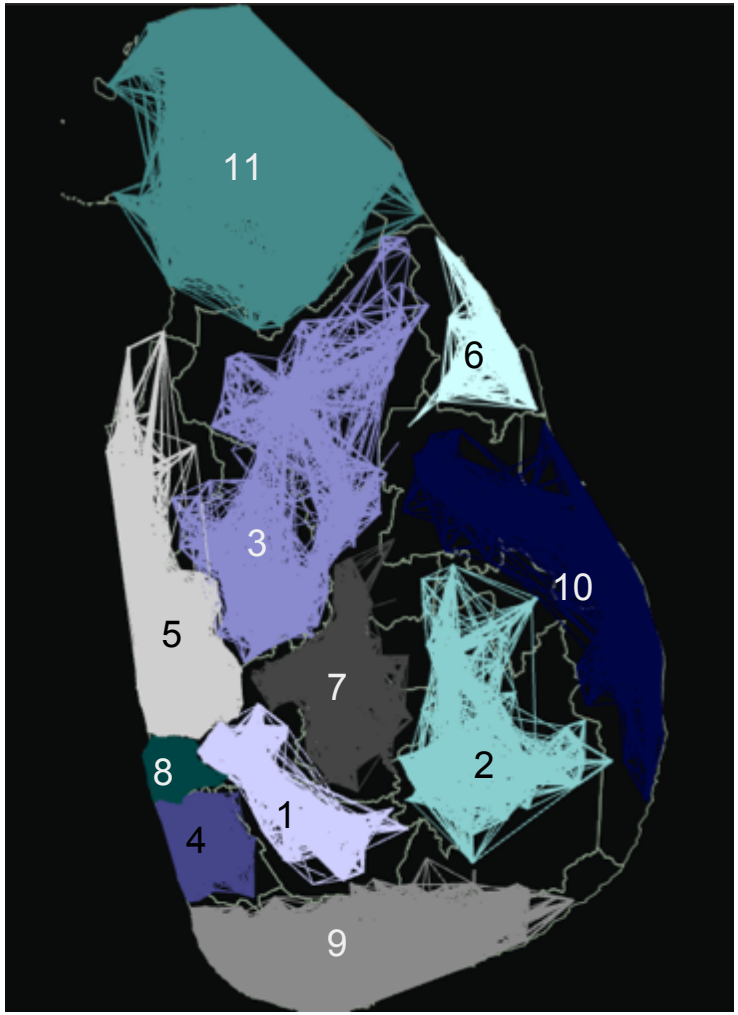
LIRNEasia
Pro-poor. Pro-market.

How communities are connected

$$\text{Connectedness} = \frac{No\ of\ calls\ between\ C\ and\ another\ community}{No\ of\ calls\ involving\ C} \text{x}100$$

- Community consisting of Northern province had the least connectedness with other communities.
- Community consisting of Ratnapura district and a part of Kegalle is the most connected community.

How communities are connected with each other



Region	Connectedness (%)
1. Ratnapura, South Kegalle, North Eastern Colombo	74
2.Badulla, Monaragala	71
3. Anuradhapura, East Kurunegala, North Matale	71
4.Kalutara	70
5. Gampaha, Puttalam, West Kurunegala	65
6.Trincomalee	65
7.Kandy, Nuwara Eliya, North Kegalle, South Matale	63
8.Colombo without North Eastern part belongs that now belongs to 1	63
9. Galle, Matara, Hambantota	62
10. Batticaloa, Polonnaruwa, Ampara	60
11. Jaffna, Kilinochchi, Mannar, Mullaitivu	56

High

connectedness with other communities

Low

LIRNEasia
Pro-poor. Pro-market.

Future work

- Explore the relationship between social network communities and their mobility profiles
- Understand the distribution of socio-economic status within communities