Project Plan

Opeoluwa Flora Fajemirokun

November 2017

**Project Title:** Estimation of health conditions in Cote D’Ivoire and Senegal using Call Detail Records (CDR)

**Supervisor:** Dr Licia Capra

1. Aim

**Aim:** To develop a methodology which produces estimates and models the health conditions (incident of malaria cases, child mortality rate and women’s access to health) in Cote d’Ivoire and Senegal using telecommunication data.

1. Objectives
2. **Pre-processing of DHS data**
   1. Decide what DHS Health data to study (e.g., malaria positivity rates using blood test results and using rapid test results; child mortality rate; HIV rate; women’s access to health; immunity against preventive disease) and collate raw data from DHS website.
   2. Compute spatial aggregates of the above indices, at the various spatial administrative units of analysis available and normalize using population size.
3. **Pre-processing of CDR data**
   1. Decide what CDR metrics to study (e.g., activity, network advantage, introversion, gravity residuals, graph metrics, etc.).
   2. Compute from given raw data, and for the various spatial (administrative) units of analysis available, again normalized by population size.
4. **Build models that use CDR metrics to estimate DHS variables**
   1. Set possible benchmarks (population density, spatial auto correlation)
   2. Use hierarchical stepwise regression models, at each levels of spatial granularity, and measure overall model fit and significance of each metrics.
   3. Study assumptions behind regression models, metrics transformation, normalizations, multi-collinearly issues and tests for heteroscedascity.
5. Deliverables

* A project plan outlining the aims of the proposed project.
* The Interim report; a document detailing the work done completed up to this point and work that is still outstanding. This document will be presented in late January.
* A literature survey that summarises previous work in the specialist area of use of CDR data to estimate variables.
* Fully documented Python scripts written to analysis CDR and DHS data so that data can be reproduced.
* Final report containing a discussion of the results obtained from analysis of CDR and DHS data with statistical analysis and discussion of their significance.

1. Work Plan

**Project start to October 22nd (2 weeks):**

* Literature review on previous D4D Challenge proposals and background readings into how DHS data is gathered and format of contained data.

**October 22nd to October 29th (1 week):**

* Analysis of DHS data for Cote D’Ivoire and selection of 4 metrics to study.
* Identify techniques to accomplish simple spatial aggregates of chosen indices for Cote D’Ivoire.

**October 29th to mid-November (4 weeks):**

* Drafting of Project Plan.
* Selection of survey variables relevant to chosen DHS metrics for Cote D’Ivoire.
* Processing of DHS data in Cote D’Ivoire to determine which survey clusters are present in each spatial administrative unit.
* Computing of simple spatial aggregates of chosen DHS metrics in Cote D’Ivoire at various spatial administrative units and normalize using number of survey participants. This is done assuming survey clusters are points which are fully contained in spatial administrative units.

**mid-November to mid-December (4 weeks):**

* Submission of Project Plan.
* Computing of more complex and spatial aggregates of chosen DHS metrics Voronoi polygons of survey clusters in Cote D’Ivoire to account for survey areas which overlap multiple spatial administrative units.
* Replication of DHS pre-processing for Senegal.

**mid-December to mid-January (4 weeks):**

* Compile Interim Report.
* Compute spatial aggregates using raw CDR data for at various administrative units of analysis available.

**mid-January to mid-February (4 weeks):**

* Submission of Interim Report.
* Build models that use CDR data to estimate DHS data for Senegal and Cote D’Ivoire.

**mid-February to end March (6 weeks):**

* Work on Final Report.