Using spatial features of human settlement to predict epidemic properties

Rahman Sanya

Data Science Africa 2017 Workshop, Jul 20th – 21st, 2017 NM-AIST Arusha Tanzania
the problem
the goal

Understand how settlement characteristics implicate observed spatial epidemic pattern

Features we want to explore

- urbanization, accessibility, mobility flux, connectivity, and population density
In this talk: urbanization

- Look for relationship between urban density & spatial epidemic dynamic
- How best to investigate potential relationship
study area

- Four contiguous LAU 4’s
- Pop size 4,524,073
- Area size 5,114.3 km$^2$
- 58.95 % urban
- Built-up, bare soil, vegetation, water
data sources

Satellite imagery (Landsat 8) from USGS
- 11 bands, cloud < 10%, 30m spatial res

Geospatial dataset (OSM) from Geofabrik
- road network, building blocks

Disease case (DHMIS2)
- TB, HIV/AIDS
general approach

• Map settlements, quantify urban
• Classify settlements into rural or urban
  – urban is 'a place dominated by built environment'
• Model relationship between urban concentration and disease dynamics
Extract built-up footprint
built-up footprint
urban density vs. case distribution
we want to...
relate features to epidemic properties

Using mutual information $I,$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

where

$$H(t) = \frac{-1}{\log V} \sum_j \gamma_j(t) \log \gamma_j(t)$$
extract communities

Using agglomerative hierarchical clustering

\[ \chi_{mn}^o = \frac{J_{(mn)}}{\min(k_m, k_n) + 1 - \Theta(A_{mn})} \]

where

- \( \chi^o \) is topological overlap matrix
- \( J_{(mn)} \) is no. of shared neighbors of nodes \( m, n \)
- \( k \) is node degree
- \( \Theta(A_{mn}) \) is Heaviside step function
relate pattern across nodes

Node pairs expected to have similar epidemic properties if they

- share neighbors
- have direct link between them
- have comparable spatial features
relate pattern across nodes

Epidemic similarity between nodes \( m, n \)

\[
\phi_{mn} = \frac{\rho_m}{\rho_n} = \frac{\rho_n}{\rho_m} \quad \rho_m \approx \rho_n; \vartheta_m \approx \vartheta_n
\]

\[
\rho = \frac{i_m}{\sum_n i_n} \quad \text{and} \quad i(t) = \frac{I(t)}{N}
\]

\[
\phi = 1 \quad \text{if nodes } m, n \text{ have similar epidemic properties}
\]

\[
\phi > 1; \phi < 1 \quad \text{otherwise}
\]
Its early days to conclude

However, our goal is to explore predictive potential of five spatial settlement features for epidemic modeling in coupled human population systems.