

Statistical inference on mobile phone network data

Mobile phone network data is widely considered as one of the gold mines for official statistics. Applications include daytime population, tourism statistics, commuting, and cross-border traffic. These data, known as Call Detail Records or signaling data, contain information about where and when mobile phones are communicating with the network. However, the exact geographic location is often neither measured nor stored. In these cases, the geographic location of mobile phones have to be derived from the location of the connected antennae.

The vast majority of studies on mobile network data use Voronoi tessellation to approximate the geographic location of mobile phones. This method assumes that antennae have a 360 degree coverage and that the coverage areas do not overlap. However, both assumptions do not hold in reality.

We propose a Bayesian method to estimate the probability that a mobile phone is present at a specific location given the antenna by which it is communicating. This method takes overlapping coverage areas into account. Furthermore, we model the signal strength for each antennae using not only their location, but also other properties, such as direction and tilt.

We show how this method can be used to estimate the number of mobile phones over time on a spatial grid, e.g. of 100 by 100 meter cells. These estimates can then be used, together with auxiliary data sources, to estimate the population distribution over time.