Detecting geographical differencing problems in the context of spatial data dissemination

Vianney Costemalle

Methodological Expert, Head of the Spatial Analysis Section

French National Institute of Statistics and Economic Studies (Insee)

www.insee.fr

Abstract

Statistical disclosure control aims neither at revealing the identity of an individual, nor at revealing characteristics of individuals, households or companies that are confidential or personal. Primary statistical secrecy concerns information one can directly assess, whereas secondary statistical secrecy concerns information that a user could deduce indirectly by recombining and crosschecking all the disseminated data.

In the case of spatial data disseminated according to several geographical partitions, it is possible to combine and intersect the geographical areas in order to derive information on new and smaller areas. The differencing technique, which consists in subtracting the value of two overlapping areas, can lead to a breach of confidentiality. When the number of units composing the two geographical partitions become large, as it is the case with grid data or small administrative units, the number of potential differencing issues becomes very important. It then seems impossible to compute every single differences to detect the ones leading to a disclosure.

We have developed a method for dealing with geographical differencing problems by detecting individuals located in small overlapping areas and whose personal information can therefore be disclosed. We modeled the data into a graph structure which enable to focus on relevant geographical regions. The originality of the method resides in reducing the graph size and complexity. We applied the method to French income tax data, which are composed of 27 million households, and which are released on the 150 000 cells of the grid squares and on the 35 000 administrative units. The method appears to work very well and allows to detect all differencing issues, involving in total 10 000 households.