Statistical secrecy aims at not revealing the identity of an individual, nor revealing characteristics of individuals, households or companies that are confidential or personal. Most often in practice, the data protection method consists in not disseminating information resulting from the aggregation of fewer observations than a given threshold. Secondary statistical secrecy concerns information that a user could infer indirectly, by recombining and crosschecking all the information disseminated.

In the case of data disseminated according to several geographical zonings, manipulations are possible to combine and intersect these areas between them in order to derive information on new and smaller areas. The differentiation technique which consist in subtracting the value of two overlapping areas can lead to a breach of confidentiality. When the number of areas of the two zonings become large, as it is the case with grid data and data disseminated over small administrative boundaries, the number of differentiation one can carry out become huge. It then seems impossible to test every single differentiation to detect the ones which lead to a confidentiality issue.

We propose a new way to deal with differentiation problems by modeling the geographical zonings as a graph. We develop a method to reduce the graph size and as a consequence to reduce the number of differentiation to test. When applied to French tax data, this method appears to work very well and allows to detect all differentiation problems that could arise if data were to be disseminated on grid data and on municipalities.