

Yerali Gandica
Université Catholique de Louvain

Does Aggregation Preserve Communities?

The ecological fallacy refers to the statistical bias caused by the aggregation of individuals into categories. In geography, particular form of such fallacy is called the Modifiable Areal Unit Problem (MAUP). MAUP affects results when individual-based measures of spatial phenomena are aggregated, either to a geographical context or according to any individual or social category, for instance age, economical income or the intensity of any kind of social contacts. Some other reasons can be privacy concerns.

We are interested in analyzing the impact of this fallacy. We focus in one of the most common task in network science, community detection. We measure quantitatively the impact of node aggregation on the community structure in networks and we introduce the aggregability index, predicting quantitatively the robustness of the community structure to a graph, in order to place it into a given aggregation class. We show that some community detection methods are more suitable than others when computing communities on aggregated networks.

We illustrate our methodology on a dataset of geolocalized tweets in Belgium, and mobile phone from one provider in Belgium. We show that our proposed index is able to predict that only the phone calls data preserves the community structure of the fine-grain level.