A new method of network clustering based on second-order sensitivity index

Huan Liu, Qiongli Wu*, Yiming Ding
Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences
*wuqiongli@wipm.ac.cn

Networks are organized into communities with dense internal connections, giving rise to high values of the clustering coefficient. In addition, these networks have been observed to be assortative, i.e., highly connected vertices tend to connect to other highly connected vertices, and have broad degree distributions. It means that high interaction between vertices exist within one cluster.

For such an important issue of interaction identification between factors, second-order Sobol's index provides a quantitative way to reveal the case. However, for decades after Sobol's method was proposed, the sensitivity analysis (SA) work using Sobol's indices mostly focuses on the aim of factor priorization (FP) using first-order Sobol's index and factor fixing (FF) using total-order Sobol's index. The application is only qualitatively limited to check which pair of factors' interaction is 'larger' and which pair is 'smaller'. What's more, the performance after these interaction mapping has been ignored. In such case, the potential usefulness of identifying the interaction quantitatively for second-order Sobol's index is not fully tapped.

We have been clear that the interaction reflects the effect strength between factors. It is the intrinsic property of the system. By coincidence, the effect strength between factors is the basis of constructing a network. In a network, communities can be defined as sets of vertices with dense internal connections, such that the inter-community connections are relatively sparse. It is similar to the factor clustering in [1]. Factors are clustered into several modules according to expert modelling experience, then the interaction between modules are detected by Sobol's group analysis. The internal-module interaction is dense and the inter-module interaction is relatively sparse.

To understand the network behaviors, such as the small-world property and high degree of clustering, we need to detect community structure in networks. Network clustering is vastly used for such structure detection. Again, such network clustering needs the quantitative description of the effect strength between factors, and the task can be fulfilled by second-order sobol's index. It means that we can use the interaction information given by second-order sobol's index for network clustering, so that to detect model clustering structure. If such method works, we can even detect the module structure in the work of [1] without expert modelling experience, which problem is also listed in the discussion of the work. The significance of such method is that it can providing a configuration of putting factors into naturally clustering according to the intrinsic interaction property of the model itself, not the subjective knowledge of the modellers. The uncertainty of expert knowledge could be avoid in this way, because if the clustering work by expert knowledge is not validated, the modules analysis result could be misleading.

As such, we propose a new method of network clustering based on second-order sensitivity index in this paper. This method is a combination of second-order Sobol's index matrix and network clustering method.

We use Newman's fast algorithm for detecting community structure in networks. Newman's fast algorithm is an agglomerative algorithm, and its basic idea is from the greedy algorithm. It firstly initializes a network with n vertices, and assumes there are m edges in the network. The fraction of edges e_{ij} is decided by the second-order sensitivity indices of the factors. Then we join communities together in pairs if they have edges connected. At the same time calculate the increase of the modularity Q:

$$\Delta Q = e_{ij} + e_{ji} - 2 \times a_i \times a_j = 2(e_{ij} - a_i \times a_j)$$

Choosing at each step the join that results in the greatest increase (or smallest decrease) in Q. In the process of joining, update the corresponding element e_{ij}, and add the row and columns associated to i, j in the same community in order. Then repeat the step, until the whole network join into a bigger community, then stop.

We present here a test case on a specific function constructed on the basis of G-function:

$$F = g(x_1, x_2, \ldots, x_5) + g(x_6, x_7, \ldots, x_{10}) + g_1(x_1) \cdot g_6(x_6) \cdot g_7(x_7) + g_1(x_1) \cdot g_2(x_2) \cdot g_6(x_6)$$

We can see the main part of F are $g(x_1, x_2, \ldots, x_5)$ and $g(x_6, x_7, \ldots, x_{10})$, and there are also two items constructed by the multiplication of factors from these two items, namely there is connection between the two communities in the network structure, which makes the network clustering not so obvious.

We computed the second-order indices of all the factors, and constructed its network structure as shown in Fig.1(a). Network clustering analysis is shown in Fig.2(b). We can see two groups were separated obviously by our method, which are $(x_1, x_2, \ldots, x_5)$ and $(x_6, x_7, \ldots, x_{10})$. It means that, even with some connections introduced by the multiplication of some vertices, our method can still make a good network clustering.

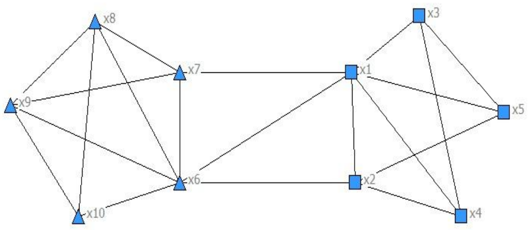

Fig.1.(a)

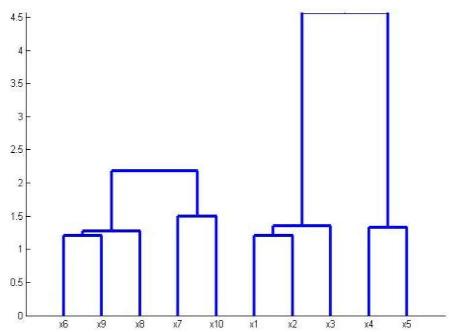

Fig.1.(b)

Our method can also make factor clustering `blindly' to expert knowledge of modelling while revealing the model mechanisms in the case that sensitivity analysis needs to be performed to group of factors. An application to an ecological model will be given in the full paper to present the factor clustering by our method in model analysis processing.

References:

[1] Wu, Q. and Cournède, P.-H., A comprehensive methodology of global sensitivity analysis for complex mechanistic models with an application to plant growth. Ecological complexity,20:219-232.