Review of: "When a Cluster Is a Cluster"

Xun Shi[1]

1 Dartmouth College

Potential competing interests: No potential competing interests to declare.

The author emphasizes the importance of cluster analysis in the study and practice of monitoring, controlling, preventing, and predicting infectious disease epidemics (especially communicable diseases like COVID-19). Particularly, they argue for taking advantage of exact/precise location information in cluster detection and analysis, which is more appropriate for individual-level data (although their case study of COVID-19 is based on city-level data). All these good points are greatly appreciated. For a more systematic discussion of the individualization and spatialization trends in epidemic modeling, one can refer to:

Li, M., Shi, X., and Li, X., 2020, Integration of Spatialization and Individualization: The Future of Epidemic Modelling for Communicable Diseases, Annals of GIS, 26(3): 219-226. DOI: 10.1080/19475683.2020.1768438.

However, a few points raised by the author in this essay are worth further discussion or clarification.

1. What constitutes a cluster in a communicable disease epidemic like COVID-19? This fundamental question, which seems to have an obvious answer, in my opinion, imposes a big challenge to epidemiologists and geographers. The answer to this question is largely determined by the purpose of the particular cluster detection work. Many researchers consciously or unconsciously take the understanding of clusters in environmental health studies, in which cluster detection serves the purpose of suggesting environmental risks, and apply this understanding to communicable diseases. However, within the communicable disease context, the purpose of cluster detection can be much more variable, if not still vague, sometimes related to environmental risks, sometimes not. The purpose of the authors’ case study, for example, seems to be to identify the outbreak location of the epidemic (the location where the epidemic originated). This outbreak location, as well as its identification, may be related to the environment, or may be not. Furthermore, such an outbreak location may not even constitute a cluster in the first place.
2. The author seems to be particularly interested in discerning separate but spatially adjacent clusters. Interestingly, this issue hasn’t received much attention in the large body of literature on spatial cluster analysis of diseases. A seminal work is:

Han, Junhee, et al. "Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics." International journal of health geographics 15 (2016): 1-11.

A consensus seems to be that this is a scale issue and the biggest concern is the modifiable areal unit problem (MAUP) – data at different scales may lead to different results - rather than the actual meaning of the separation (clusters resulting from different spatial sources). This might be because distinguishing two spatially close-by sources (e.g., two pollution sites) may not be that critical in an environmental health study.

3. The author seems to have mixed two types of “clusters”. Many people have pointed out that “cluster” in disease mapping may not be a proper term. The actual meaning of such a cluster is “hotspot”. To call it “cluster” is easy to confuse with another type of clustering analysis for “grouping” or “segmentation”.

4. The TWC method the author recommends is a very interesting approach to point pattern analysis. It may have great potential in studying certain geographical phenomena. However, I am questioning its applicability in modeling a communicable disease epidemic. Essentially, this is a top-down approach trying to characterize the overall pattern of the entire point set. The overall “attraction strength” of a point is determined by its connection with all the other points and, in fact, very often, its “indirect distance” to its closest point is longer than its distance to distant points. This, obviously, may not be realistic, as the infection and spread of a communicable disease is usually a bottom-up process, i.e., the infector usually infects locally, and the general pattern emerges from this kind of local, Markov-chain style process, rather than the other way around.