

Review of: "When a Cluster Is a Cluster"

Lilia Leticia Ramírez Ramírez¹

¹ Centro de Investigación en Matemáticas (CIMAT)

Potential competing interests: No potential competing interests to declare.

In general, the topic of the article is interesting. It is a good article to appear on the internet as it introduces a problem and refers to some studies, but it falls short of being a formal academic survey article. The descriptions of the referred methods are too short and vague as the author tries to describe several characteristics but without a directive idea.

More comments:

The author may consider a more informative title.

The abstract does not fully describe what the article is about.

It would be desirable if the author refers to "the Computerized dot maps" as the cellular automata, the more general spatial agents based model, or some other model.

According to what studies (or institutions) is a cluster defined as "when there are ten or more cases connected through transmission"? Is this definition independent of the absolute time period or independent of infectious agent parameters such as latent and infectious periods? Here you can refer to the work(s) that provide a more precise definition, as mentioned in the first paragraph of Section "Statistical approaches to cluster recognition".

"with more than 400.000 papers" should be "with more than 400,000 papers"

The last statement is according to what source?

"Just to give an idea, in the early phase of COVID-19 spread in China, as the number of cases was increasing, nearly 1000 cluster cases were reported." How is this information connected to the idea in the previous paragraph? Did the government in China facilitate this information, making these types of studies viable, or is it related to having more data due to a more severe outbreak?

In the first paragraph of section "Statistical approaches to cluster recognition," it would be very helpful to provide more information on how the cited authors define the clusters. It is still unclear what the time period for the new cases they consider is. Or maybe the clustering methods also incorporate the onset time.

What are the assumptions that are violated in the usual significance tests?

