

# Beyond central-tendency: If we agree discrete vegetation communities do not exist, should we investigate other methods of clustering?

Mark Tozer<sup>1</sup> and David Keith<sup>2</sup>

<sup>1</sup>University of New South Wales - Kensington Campus

<sup>2</sup>University of New South Wales

September 1, 2023

## Abstract

1) Clustering is indispensable in the quest for robust vegetation classification schemes which aim to partition continua to summarise and communicate pattern. However, clustering solutions are sensitive to methods and data and are therefore unstable, a feature which is usually attributed to noise. Viewed through a central-tendency lens, noise is defined as the degree of departure from type, which is problematic since vegetation types are abstractions of continua and so noise can only be quantified relative to a particular solution to hand. Graph theory models the structure of vegetation data based on the interconnectivity of samples. Through a graph-theoretic lens, the causes of instability can be quantified in absolute terms via the degree of connectivity among objects. 2) We simulated incremental increases in sampling intensity in a dataset over five iterations and assessed classification stability across successive solutions derived using algorithms implementing, respectively, models of central-tendency and interconnectivity. We used logistic regression to model the likelihood of a sample changing groups between iterations as a function of distance to centroid and degree of interconnectivity. 3) Our results show that the degree to which samples are interconnected is a more powerful predictor of instability than the degree to which they deviate from their nearest centroid. The removal of weakly interconnected samples resulted in more stable classifications, although solutions with many clusters were apparently inherently less stable than those with few clusters, and improvements in stability flowing from the removal of outliers declined as the number of clusters increased. 4) Our results reinforce the fact that clusters abstracted from continuous data are inherently unstable, and that the quest for stable, fine-scale classifications from large regional datasets is illusory. Nevertheless, our results show that using models better suited to the analysis of continuous data may yield more stable classifications of the available data.

## Hosted file

Tozer&Keith\_revised\_Main\_document\_clean.docx available at <https://authorea.com/users/505959/articles/662907-beyond-central-tendency-if-we-agree-discrete-vegetation-communities-do-not-exist-should-we-investigate-other-methods-of-clustering>