The impossibility theorem for clustering [report of progress] 2-

Statement of research topic

Based on the axiomatic system proposed by K., we try to relax the consistent theorem, the relaxation is following the idea proposed by 2018.. instead of using the concept of natural number of clusters, we rely on two quantities to measure the degree of perturbation, and taking a further step, we use the redan index as the measurement of the difference between the two partitions. And try to find if there is a general rule which could describe the relation between the degree of perturbation and the changes of partition results for some clustering algorithms, or in general for a type of clustering algorithm. If this relationship exists, then we could rephrase this relation into new consistent theorem.

Summary of progress

Up to now, what we had done is following: With aim of turning the impossibility theorem to possibility theorem, we looking into the several papers to find a proper perspective to relax the theorem, after (finding the consistent theorem is not good?)deciding to relax the consistency theorem, we take some time to find proper condition which could model this kind of anomaly, and

define the three quantities.

Because the relation between the three quantities is not clear, we first rely on the computer program to simulate the perturbation and record the measures of each perturbation as one three dimensional point - (m1,m2,rand index). After numerous data points are generated, plot of these data points can provide some insights for defining the relation between the extend of perturbation and difference between the two partition results. Furthermore, statistical method can help to capture this relation.

At the same time, property of these three metrics are analyzed from the theoretical perspective, f

Plan for rest of the project

Reasoning behind this kind of relation is the foremost part for this project, if the simulation result do confirm the existence of some relation, much work will need to be done： Firstly, we need to compare the simulation results for multiple clustering algorithm, and then summarize these relations into rigorous mathematical statement. Secondly, we need to prove this numerical relation using the properties of clustering algorithm and the definition of three quantities.

On the other hand, if these relation dose not exist, there are two directions to go – The first is to stick to this approach and select other quantities which could measure the degree of perturbation and changes of partition results. The second is to find way to add other constraints to consistency theorem.