There are three approaches to finding agglomerations reproduced by Bosker et al in "Definition Matters": 1) Local labor market approach often attributed to **Duranton** 2015b, 2) Agglomeration Index (**AI**), 3) **Cluster Algorithm**, and 4) Thresholding of nighttime lights (**NTL**) data. Within each, there are threshold parameters to be determined.

The local labor market approach, as implemented by Duranton, is what the authors consider the first best and is an iterative algorithm that builds up an area of agglomeration from the inner-most municipal boundary to the outer most municipal boundary depending on whether the share of households commuting into the inner circle meets a threshold.

The Agglomeration Index (AI) attempts to recreate the Duranton in areas where commuting patterns are not available. An area of agglomeration is one that has a density of at least $150km^2$ and it takes 60 minutes or less to travel to a core area composed of at least 50,000 people.

The cluster algorithm (UC) is what we are using whereby core areas are contiguous $1km^2$ blocks with a minimal density and a minimal total population. The authors find two standards of thresholds in literature: 1) minimum density of 300 people per $1km^2$ and a population of at least 5,000 or 2) a minimum density of at least 1,500 per $1km^2$ and at least 50,000 people.

The thresholding of nighttime lights (NTL) approach is similar to the cluster algorithm but instead of using population density it uses contiguous cells with a minimal light emissions. Depending on the lower bound of light emission per geographic unit, areas of agglomeration can be increased or decreased.

When compared to each other in Indonesia, the Duranton index is used as the ground truth. For the Duranton algorithm, the authors found that a commuting population share threshold of 7% was able to recreate the official metro area delineations. According to the authors, relative to Duranton, the AI and low-density UC algorithm provide implausible results because they tend to over-agglomerate and return very large metro areas with very large populations. Results look more plausible when they used the high-density UC algorithm and a high-threshold (80th quantile) NTL algorithm in that the number of distinct metro areas produced by these more closely aligns with the Duranton approach. More specifically, the NLT and high-density algorithms closely resembled the Duranton approach at a 10% threshold.

When it comes to measuring the effect of agglomeration on wages, the authors find that it is largely a function of the number of metro districts and the average number of metro districts per metro area. The wage premium is highest for the high-density UC algorithm and high-threshold NTL intensity because these methods result in the fewest metro districts. The AI algorithm and NTL at a threshold at the 25th percentile result in the smallest wage premiums because these approaches produce the most number of metro districts. Finally, "...agglomeration wage premium is smallest when using Duranton's algorithm which is characterized by fewer districts per metro area than any of the other approaches." In general, the wage premium seems to be between 6 and 7%.