Node features:

Product score
News sentiment
Report evidence

Cluster 1
Cluster 2
Cluster 3

For example:
- Cluster 1 mean: 3.5
- Cluster 2 mean: -2
- Cluster 3 mean: 0

For regression purpose, one may choose to increase the number of clusters and directly use cluster mean as labels.
- cluster 1 ⇒ label 3.5
- cluster 2 ⇒ label -2
- cluster 3 ⇒ label 0

For classification, one may use the sign function or create distinct ranges for each class.
- cluster 1 ⇒ label +1
- cluster 2 ⇒ label -1
- cluster 3 ⇒ label 0

Goal: Cluster the nodes based on Graph structure + Node features (GCN/GNN modified algorithm)

How to determine the label/final score for companies in each cluster?
1. **(Base)** We must take MSCI's considerations into account and enhance them by examining/adding other factors.
2. **(Additional stuff)** MSCI's score seldom considers report evidence or supply chain-related activities discovered from the knowledge graph.

Solution: In order to incorporate both 1 and 2, we take the mean of each cluster as the label for every company in that cluster