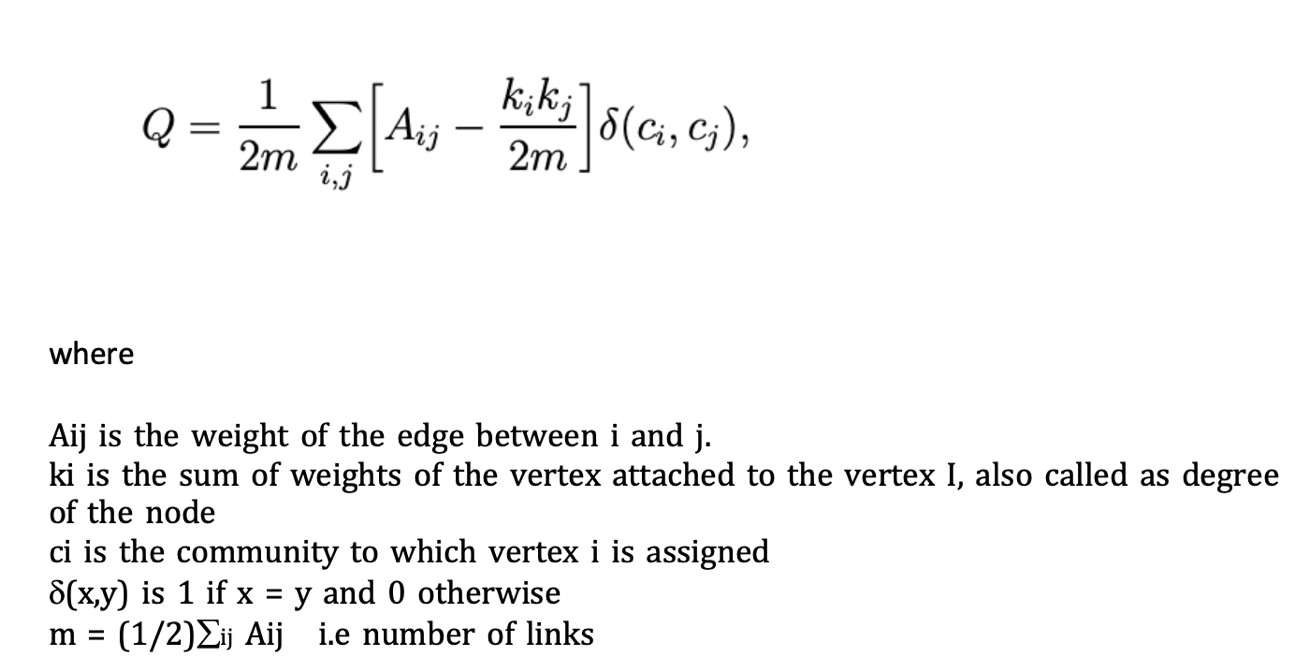
The Louvain Method:

Within a network we have a set of nodes called communities and each community has denser connections internally and sparser connections between the sets of nodes, I.e. The pairs of nodes are highly likely to be connected if they belong to the same set of nodes and less likely connected if they are from different sets.

In order to measure the quality of the partition of nodes into communities, we use the modularity metric that evaluate how densely the nodes are connected within a community compared with random distribution of links between all nodes regardless of communities and it scores partitions highly when most edges fall within the community and penalize partitions with very few or very big parts.

Modularity is an objective function that needs to be maximized for some community detection techniques, it ranges between -1 and 1, The higher the score, the tightly connected the communities detected are.

In the case of weighted connections between the objects, we can define the modularity with the following formula:



The Louvain method developed by Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre in this [paper](https://arxiv.org/pdf/0803.0476.pdf) in 2008, it is a simple and fast algorithm for detecting communities in large networks, even outperform all other known community detection algorithm in terms of computation time and moreover, it is based on a heuristic to maximize the modularity.

Algorithm description:

The algorithm consists of two steps that are repeated until the modularity a maximum modularity is reached and can’t maximize any further.

We start with assigning each node to a different community which means the number of communities is equal to number of nodes N.

The first step:

For each neighbor *j* (j=1,2,…n)of node *i*( i=1,2,…n), we check if the modularity increases by moving node *i* from its community to community *j,* then the node *i* is moved to community *j* which has the highest modularity.

Then we repeat for all the nodes

The second step:

New network is built by grouping the nodes together in the same community.

Remark, the weights of the edges between the two new communities are determined by adding up the weights of edges from every node in one community to another.

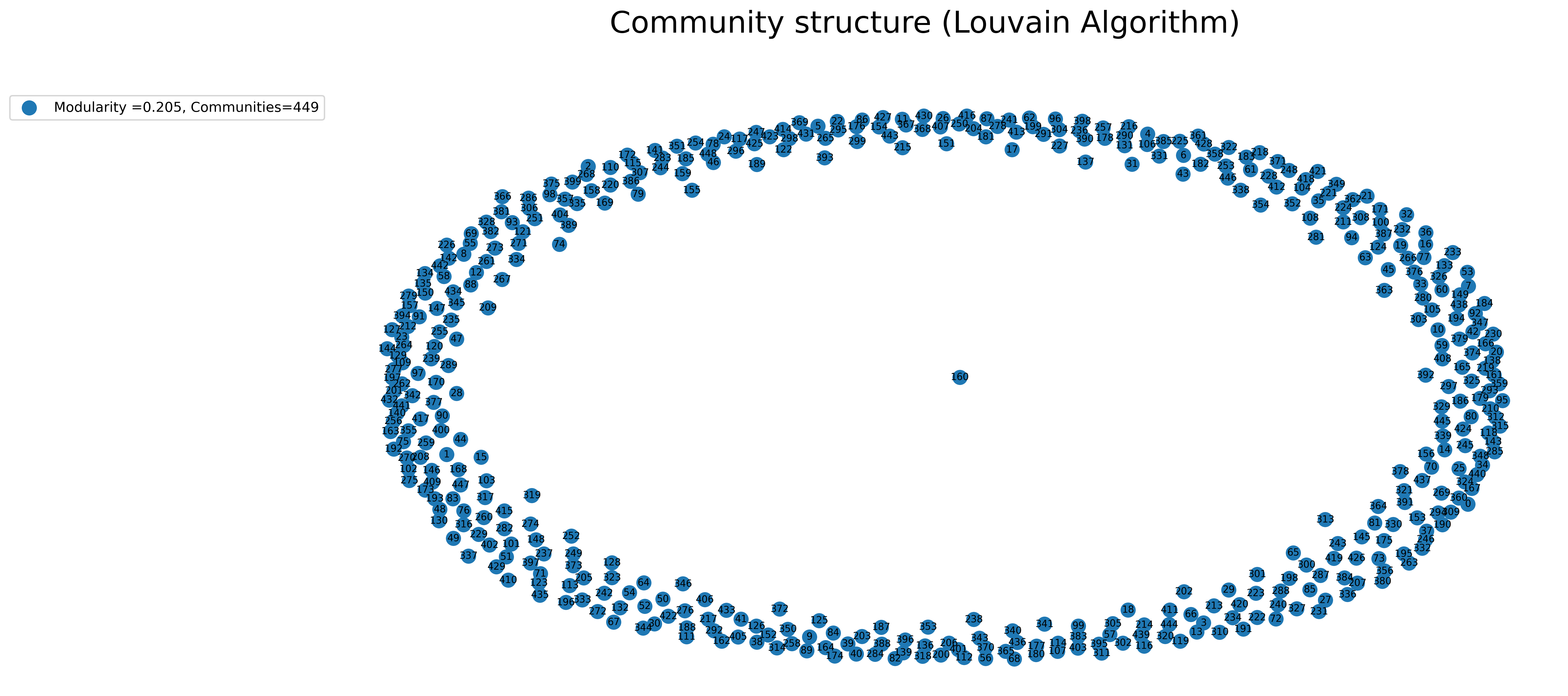
In our coursework we performed the Louvain method on the 3 streams datasets (***BL ,AD & EX*** ) in order to extract the communities as we can see respectively in Figure 1,2 & 3 .

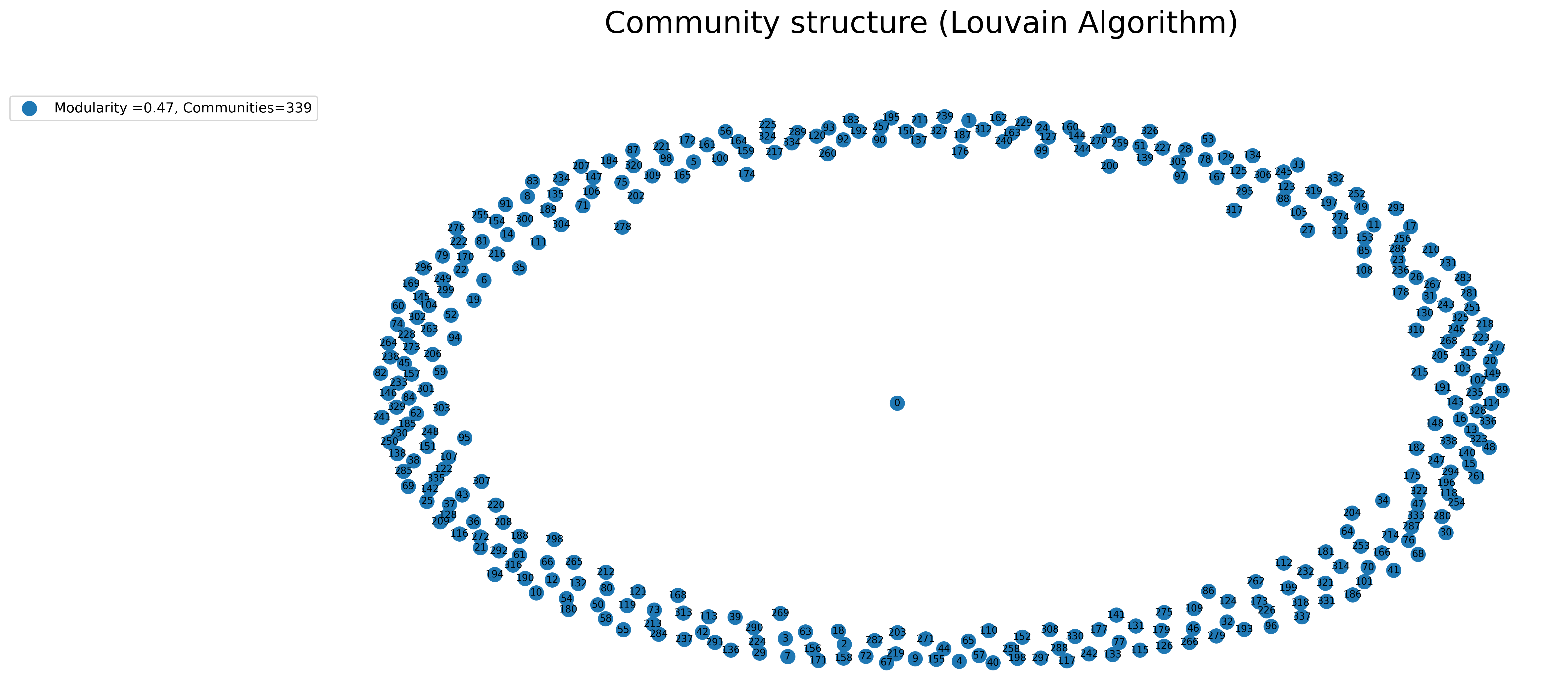
For ***BL*** stream, as we can notice from the fig shown below, we have 330 communities (**330 clusters**) and a modularity measure of **0.353.**

The modularity is a bit low, which indicate that our graph is not well interconnected, and this is certainly due to data; the **Bl** stream is identified by the **baseline crawler** and probably contains more noise and irrelevant tweets.

Also, we see that few communities contain large numbers of hashtags and some of those hashtags are the most influential ones.

For instance, *# Glastonbury* and are all in community 0.





For **AD** stream, we have339 communities but higher modularity 0.45 , the network seems more interconnected compared to BL or Ex, this is mainly due to collecting data using adaptive crawler, AD contains extra content and a higher proportion of event relevant tweets as well as less noise.