# Methodology:

Within the social network, influencers and opinion leaders stand out and gain interest from a marketing perspective because they have the potential to influence the buying behavior in both their first-order contact and in the rest of their network. To address the main problem of this study, the key step is to determine the relevant metrics that would identify these influencers.

SNA provides several methods that can be used to describe and weigh different characteristics of the network in general, the individuals that make it up, and the connections or links between these individuals. However, this study proposes to identify these individuals by adapting the matrix model presented by [Scoponi *et al.* (2016)](https://www.emerald.com/insight/content/doi/10.1108/EJMBE-10-2017-020/full/html#ref052) to classify the actors of a social network in terms of their level of influence through two complementary metrics. Members of the network that simultaneously meet the highest values of both **Betweenness** and **Eigenvector Centrality** should be classified as potential influencers as shown in the figure below.

Diagram

Description automatically generated

This matrix represents a two-dimensional scatter plot in which the individual components of a network are plotted according to their betweenness centrality (x-axis) and their eigenvector centrality (y-axis). Then, according to a relevant criterion to determine thresholds in each dimension, this plane is divided into quadrants, allowing classifying every actor into four different groups:

(1) potential influencers with a high degree of betweenness and eigenvector centrality.

(2) brokers or individuals with high betweenness centrality and low eigenvector centrality.

(3) actors with important connections, their low score in betweenness centrality suggesting a limited outreach to groups outside their local community; and

(4) secondary actors.

For this study, thresholds were defined in the 95th percentile of each one of these two dimensions wherein the top 5% is identified as the potential influencer considering both metrics. However, this threshold can be adjusted according to the purpose of the analysis, the size and characteristics of the network, operational and budgetary constraints of the company.

Although both **Betweenness** and **Eigenvector Centrality** can greatly impact the scale of the marketing campaigns of Huawei, it is important also to know who has the most direct connections or most active individuals in the network. This is where **Degree Centrality** can help as a metric to identify the said individuals who are more likely to have friends as they are popular as reflective in their connections.

For a marketing campaign under budgetary constraints, speed in the transfer of knowledge or word of mouth between consumers is also important. Speed is greatly facilitated by the closeness of the individuals in the network which means information and influence are easily pass on when individuals are directly linked to each other. SNA uses **Closeness Centrality** to identify these individuals who has access to nodes in the network more quickly than anyone else. They have the shortest paths to all others as they are close to everyone else.

In summary, to make this study relevant to the main business case, the order and flow of the analysis will go through respectively using the following metrics:

* Tools:
  + Gephi – evaluation tool
  + Python Networkx – evaluation tool
  + Github – main repository
  + SigmaJS explorer (Gephi Plug-in) – tool for the deployment
* Data Preparation Steps:
  1. Get the dataset
  2. Clean the data
  3. Explore the data
* Evaluation Steps

1. **Degree Centrality**
   * Identify the key players in the network who has the most connections.
   * Choose a range of degree centrality.
   * Filter on the 75% range (214 degree). Use python to generate the filter.
   * Explore and drill-down on both In-degree and Out-degree metrics.
2. **Betweenness –** using same parameters and evaluation steps
3. **Eigenvector Centrality** – using same parameters and evaluation steps
4. **Pagerank (Gephi) –** to validate initial output of the python procedure
5. **Plot Betweenness and Eigenvector together** - 2-dimensional matrix will be used to identify the potential influencers
6. **Closeness Centrality** - to identify individuals who can help speed up awareness of Huawei as a brand
7. **Filter the top influencers considering all criteria –** 
   * Load again 100% of the dataset
   * Create a matrix table to summarize all metrics
   * Create a correlation matrix to validate and finalize top influencer
8. **Explore and compute for each influencer in terms of the following:**
   * Group where they belong
   * Reach of the influencers – how many nodes/connections