Rapport – Projet 2 : Analyse de données

# Point 1 : Le graphe du Karaté Club

Nous avons décidé d’utiliser les librairies networkx (pour la manipulation des graphes) et matplotlib (pour leur affichage). Nous importons le graphe avec la fonction karate\_club\_graph() et l’affichons ensuite.

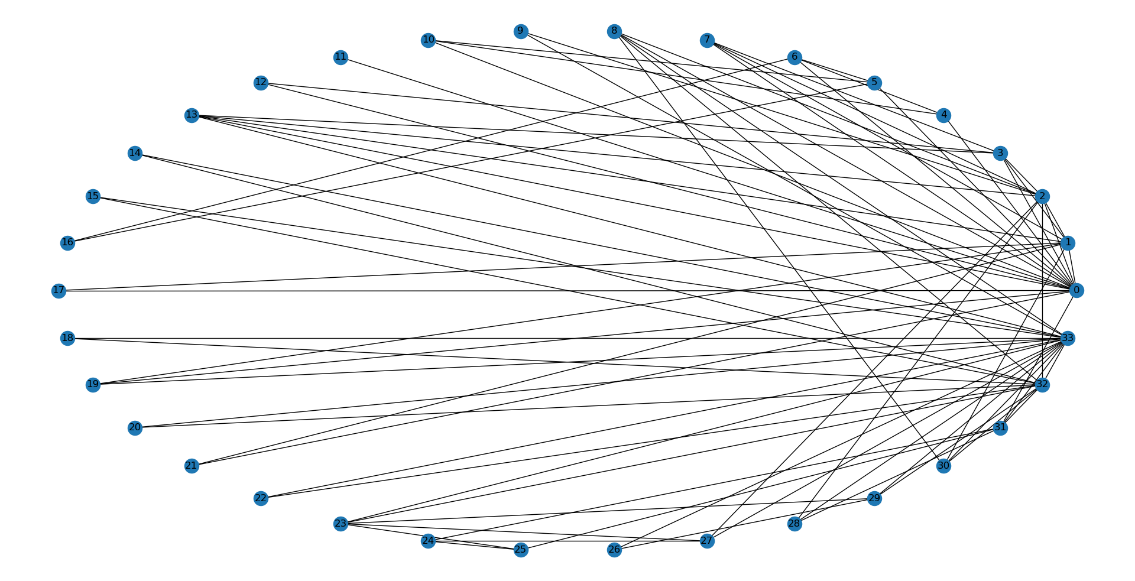


Figure : Graphe du Karaté Club

# Point 2 : La distribution des degrés et de différentes centralités

## La distribution des degrés

Premièrement, nous trions les degrés des nœuds par ordre décroissant. Ensuite, nous stockons cette séquence de degrés dans un objet Counter afin de pouvoir la manipuler plus facilement. A partir de cet objet, nous créons deux listes. L’une contenant l’ensemble des différentes valeurs de degrés du graphe (trié par ordre décroissant) et l’autre, le nombre respectif de degrés de cette valeur présent dans le graphe. C’est-à-dire : (17, 16, 12, 10, 9, 6, 5, 4, 3, 2, 1) et (1, 1, 1, 1, 1, 2, 3, 6, 6, 11, 1). Enfin, nous affichons ces deux listes sous la forme d’un histogramme avec par-dessus le graphe du Karaté Club.

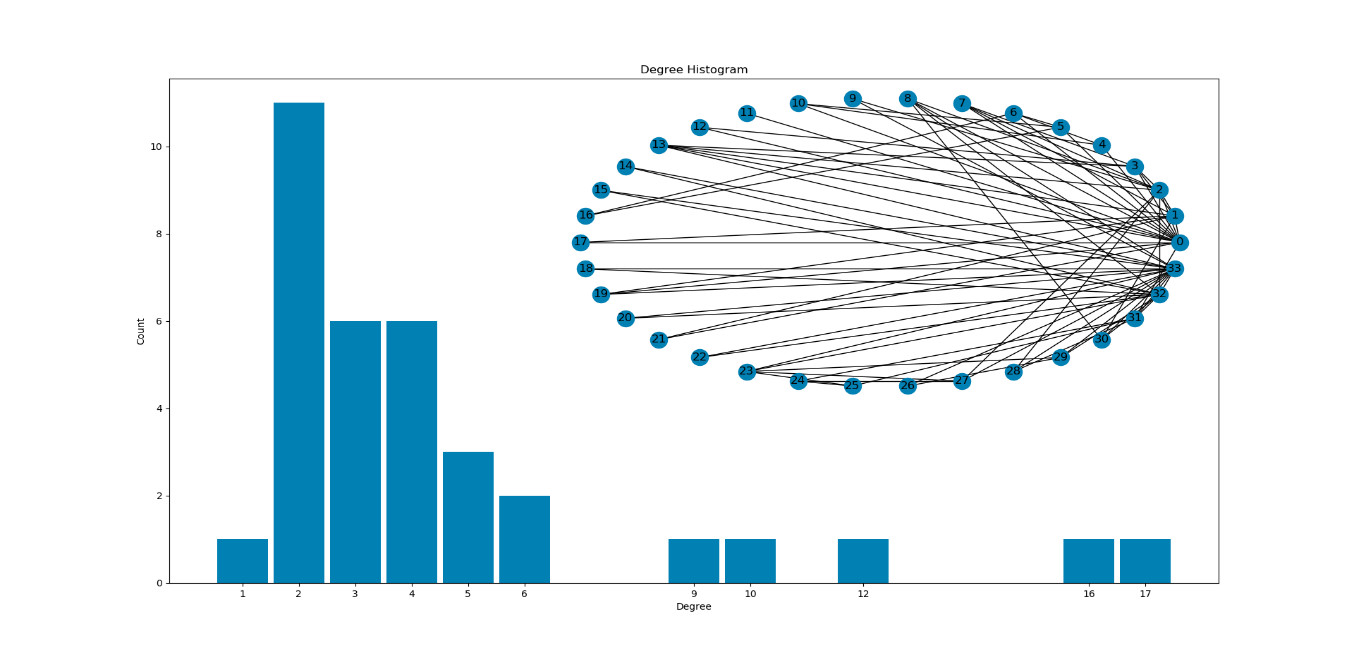


Figure : Distribution de degrés

## La distribution de différentes centralités

Nous calculons les différentes centralités à l’aide des fonctions suivantes closseness\_centrality(), betweenness\_centrality(), katz\_centrality\_numpy(), pagerank(), degree\_centrality().

### Closseness centrality

Cette mesure indique à quel point un nœud est proche des autres. Elle est calculée comme la moyenne des chemins les plus court de ce nœud à tous les autres. Par exemple pour le nœud n°16 la « closseness centrality » est égale à 33 / (2+ 3+ 3+ 3+ 2+ 1+ 1+ 3+ 3+ 4+ 2+ 3+ 3+ 3+ 5+ 5+ 3+ 5+ 3+ 5+ 3+ 5+ 5+ 4+ 4+ 5+ 4+ 4+ 5+ 4+ 3+ 4) = 0.29. Et celle du nœud n°0 est égale à 33 / (1+ 1+ 1+ 1+ 1+ 1+ 1+ 1+ 2+ 1+ 1+ 1+ 1+ 3+ 3+ 2+ 1+ 3+ 1+ 3+ 1+ 3+ 3+ 2+ 2+ 3+ 2+ 2+ 3+ 2+ 1+ 2) = 0.59

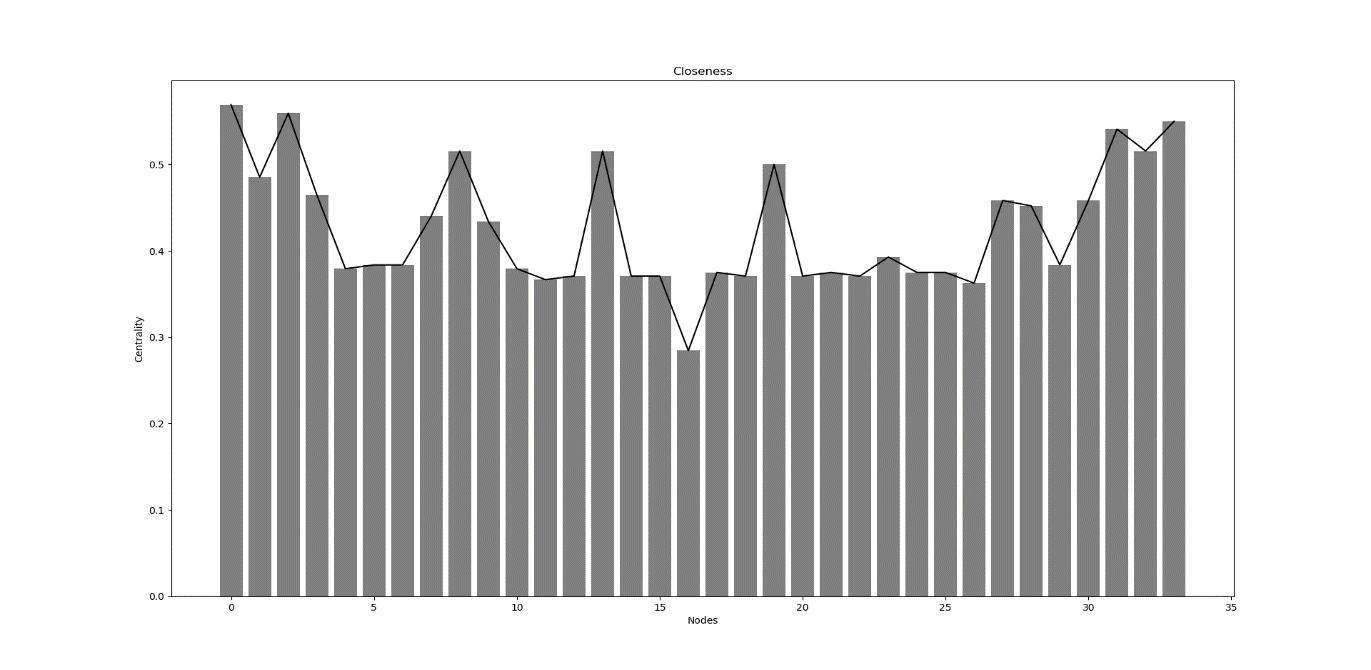


Figure : Closeness centrality

### Betweenness centrality

C’est une autre mesure de la centralité d’un nœud. C’est le nombre de fois qu’un nœud est sur le chemin le plus court entre deux autres nœuds quelconques.

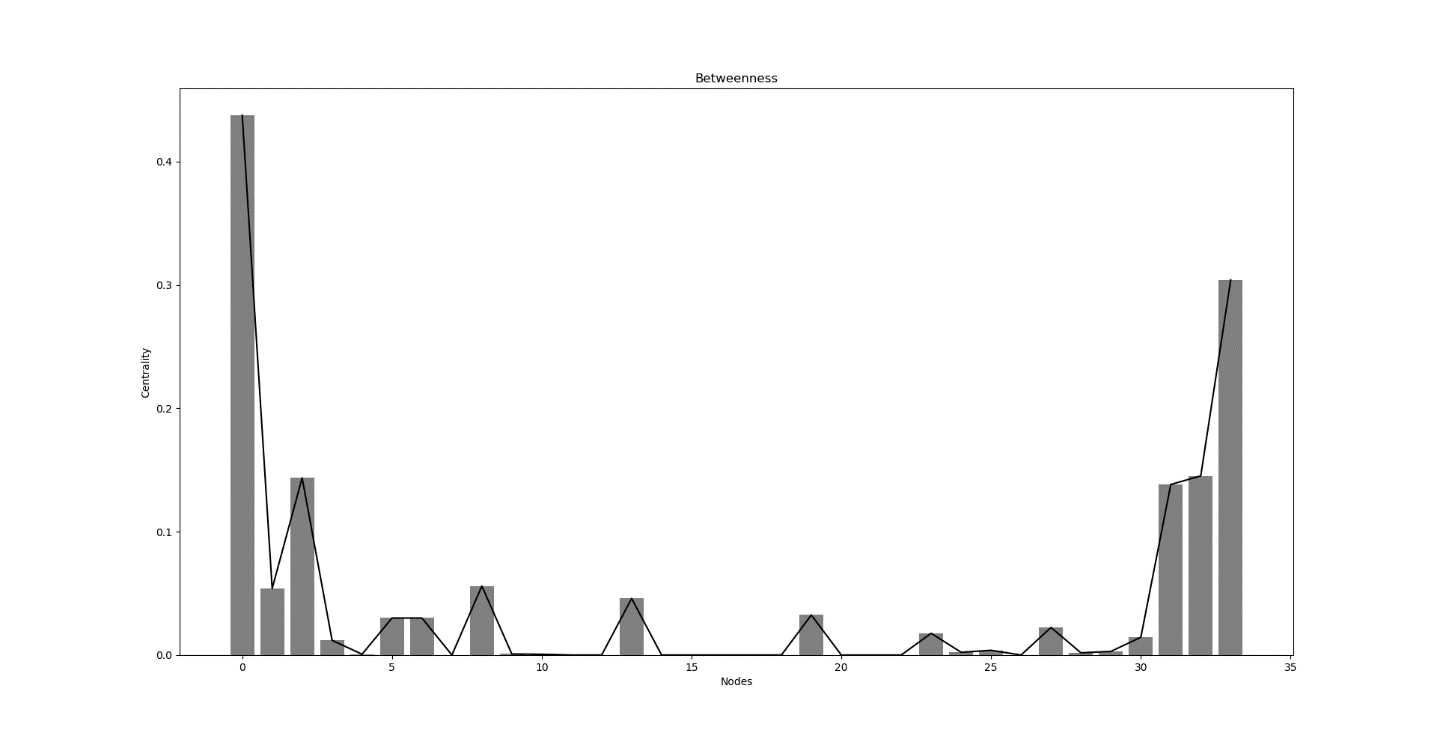


Figure : Betweenness centrality

### Katz centrality

C’est une autre mesure de la centralité d’un nœud, calculée en tenant compte du nombre de voisin direct mais aussi les autres nœuds qui se connecte à ce nœud en passant par les voisins directs.

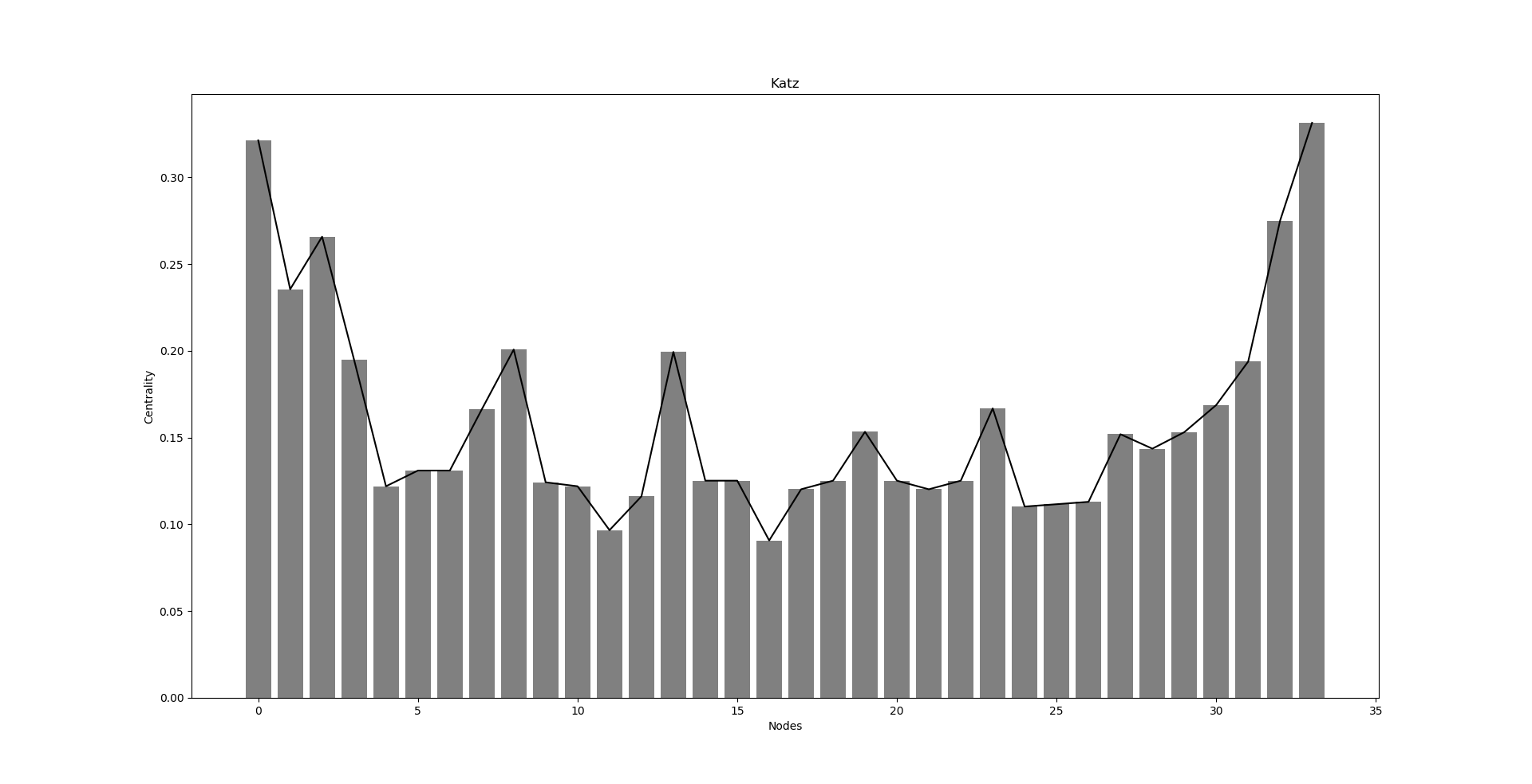


Figure : Katz centrality

### PageRank

PageRank calcule la centralité d’un nœud en prenant en compte, l’importance des nœuds voisin.

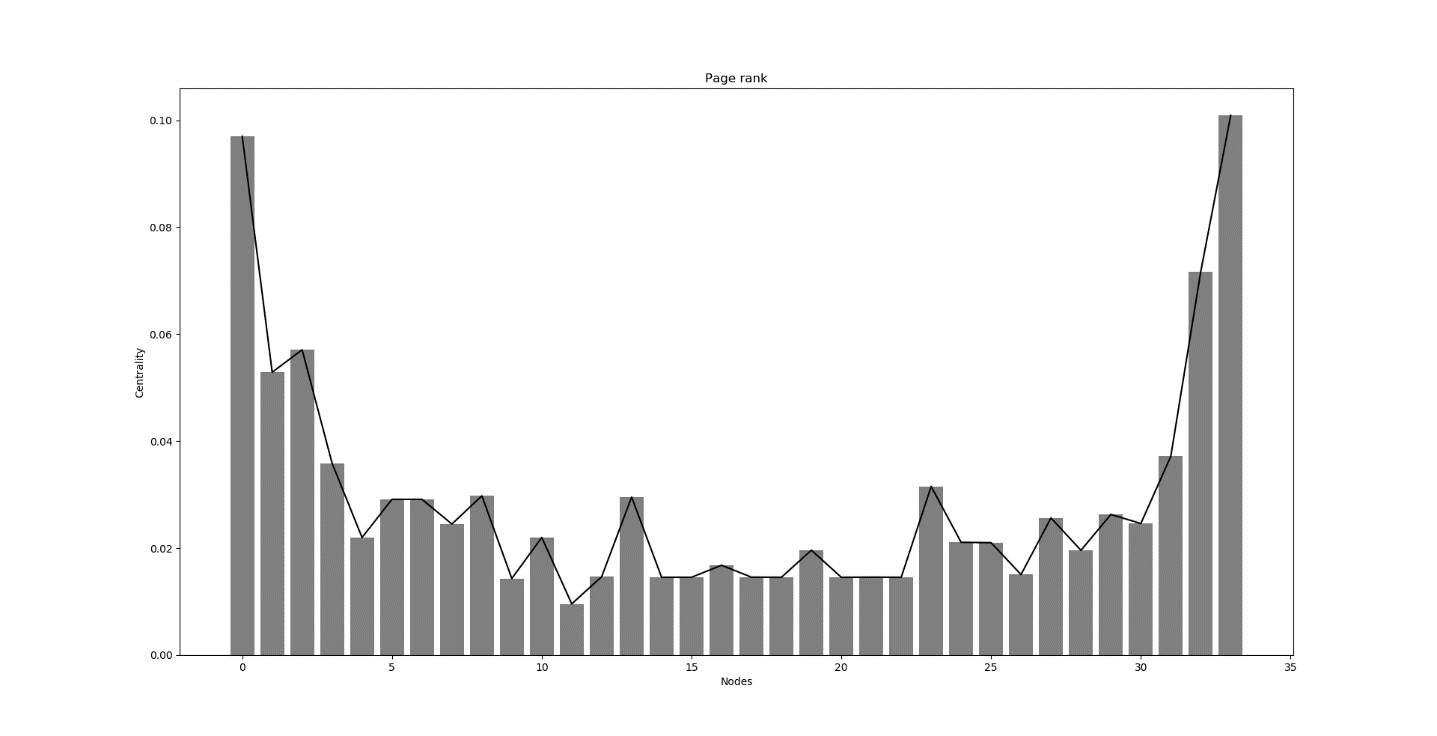


Figure : Page Rank

### Degree centrality

Cette mesure de la centralité est la plus simple car la centralité d’un nœud est simplement égale à la valeur de son degré.

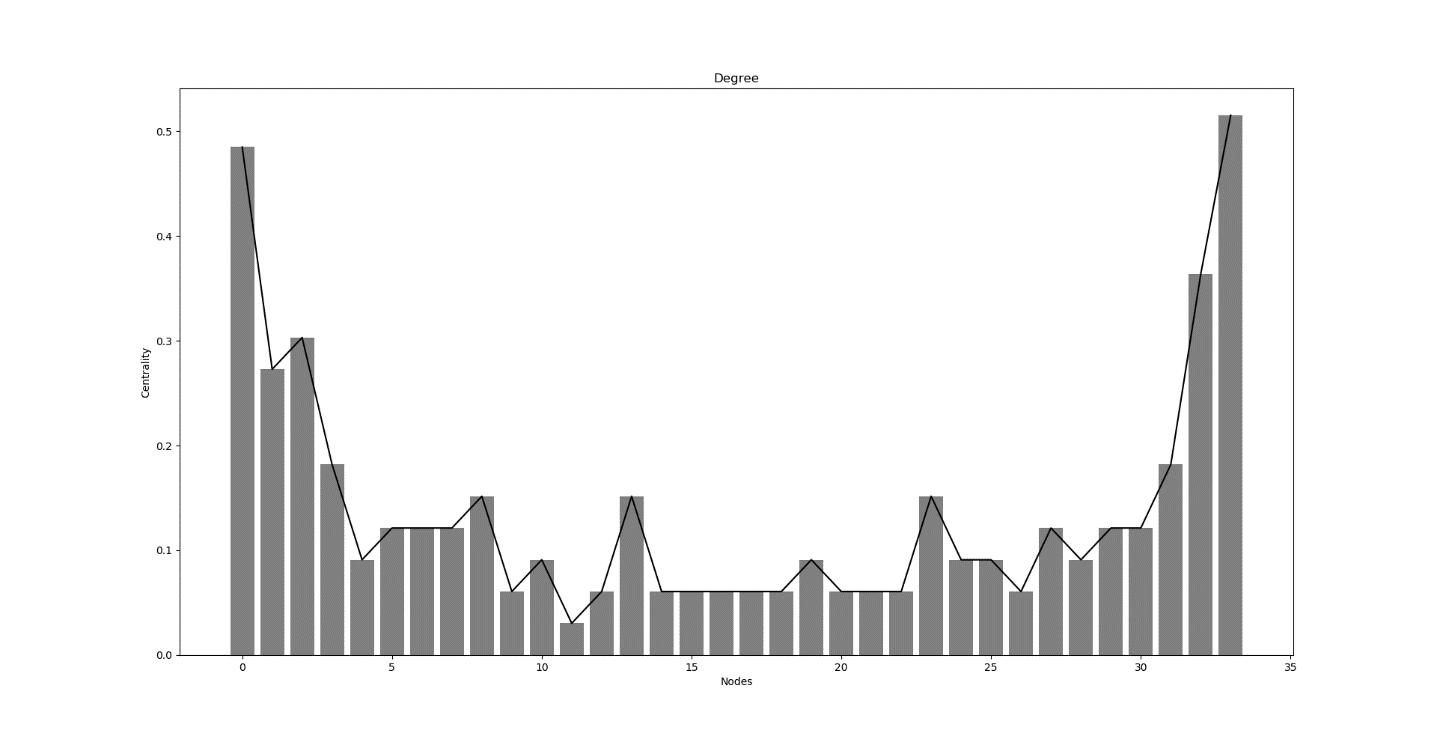


Figure 7 : Degree centrality

# Le clustering coefficient (Global)

Ce coefficient mesure le regroupement des nœuds dans un réseau c’est-à-dire à quel point les nœuds voisins sont connecté entre eux.

Dans le graphe du Karaté Club le clustering coefficient est égal à 0,57.

# Le modèle de configuration de degré

Le modèle de configuration de degré est une méthode qui utilise une séquence de degré donnée afin de générer un réseau aléatoire. Il peut donc exister dans ce réseau des boucles et des liens multiples entre deux nœuds. La fraction de liens à éliminer est de l’ordre de 10%, sur 78 liens entre 6 et 10 sont supprimé.

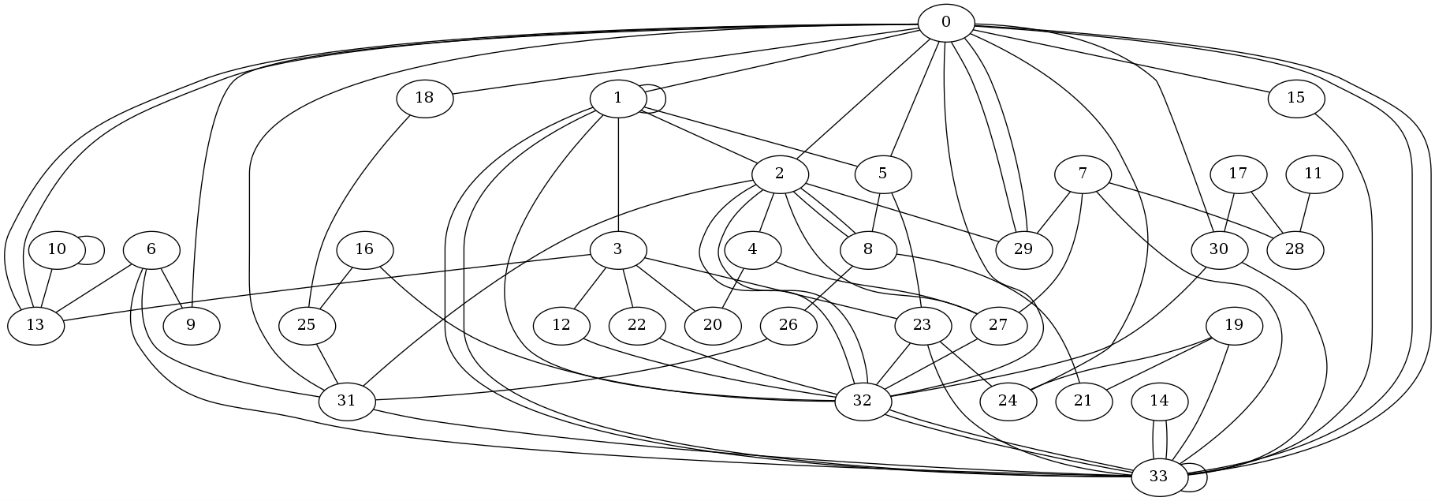


Figure : Graphique généré grâce à un modèle de configuration créé avec les degrés du modèle du karaté club.

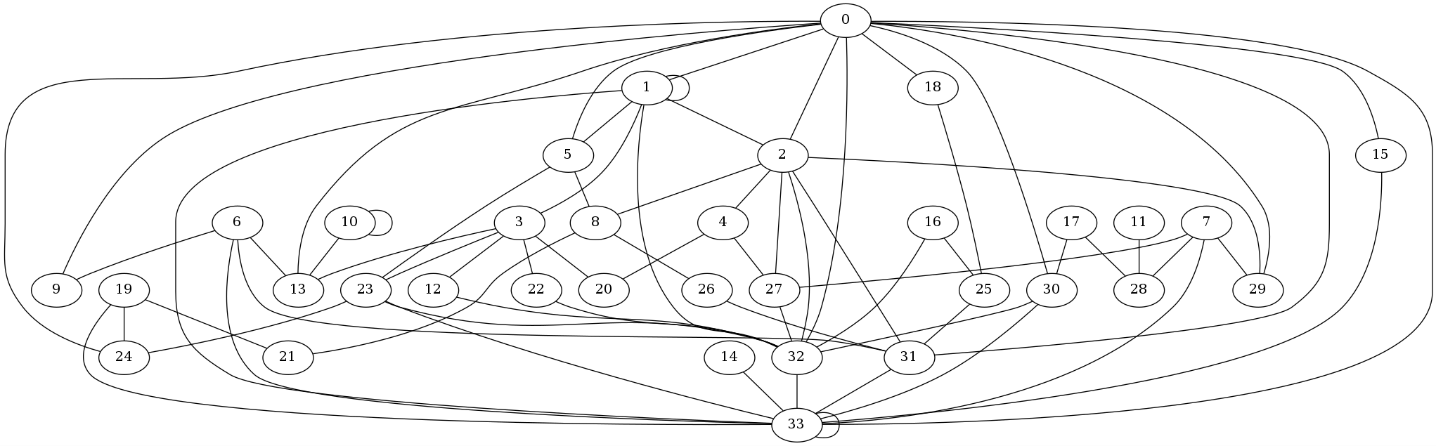


Figure : Le même graphique après élimination des liens multiples

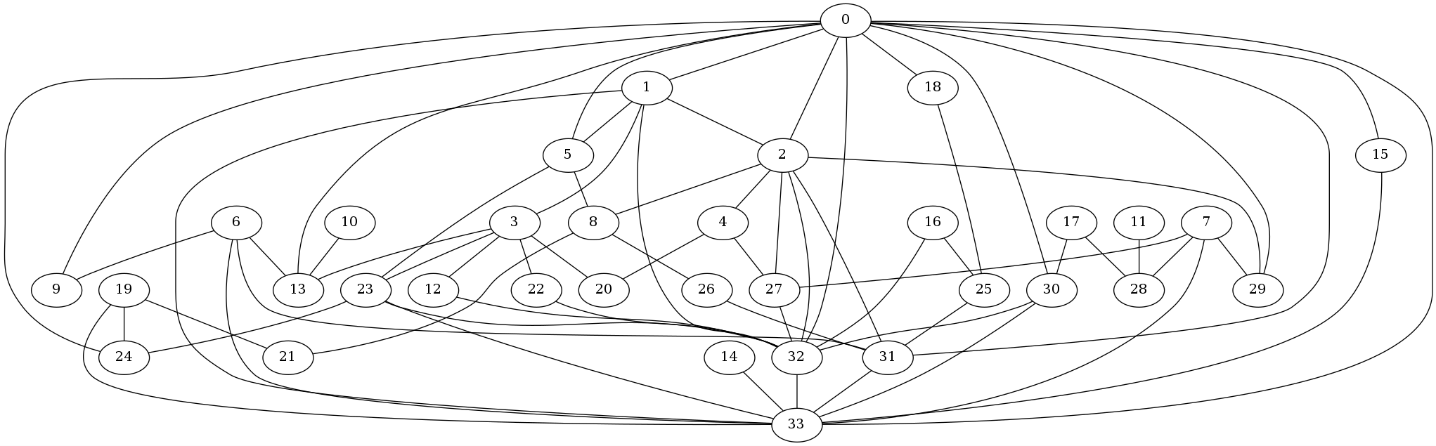


Figure : Le graphique final obtenus après avoir supprimé les boucles

# L’algorithme de Louvain

Cette méthode permet l’extraction de communautés au sein d’un réseau. Nous avons généré 100 graphes aléatoires composé du même nombre de nœuds et de liens que dans le graph du karaté club. Nous avons également fait en sorte qu’il n’y ait pas de boucle, ni des liens multiples dans les graphes aléatoires. On remarque que les graphes ainsi généré ont une modularité comprise entre 0,29 et 0.39. La valeur de modularité du graph du karaté club est-elle égale à 0,42. La modularité est la mesure de la qualité d’un partitionnement donné des nœuds d’un graphe. Les réseaux avec une haute modularité ont des connections plus dense entre les nœuds à l’intérieur d’une même communauté que les nœuds à l’intérieur de communautés différentes.

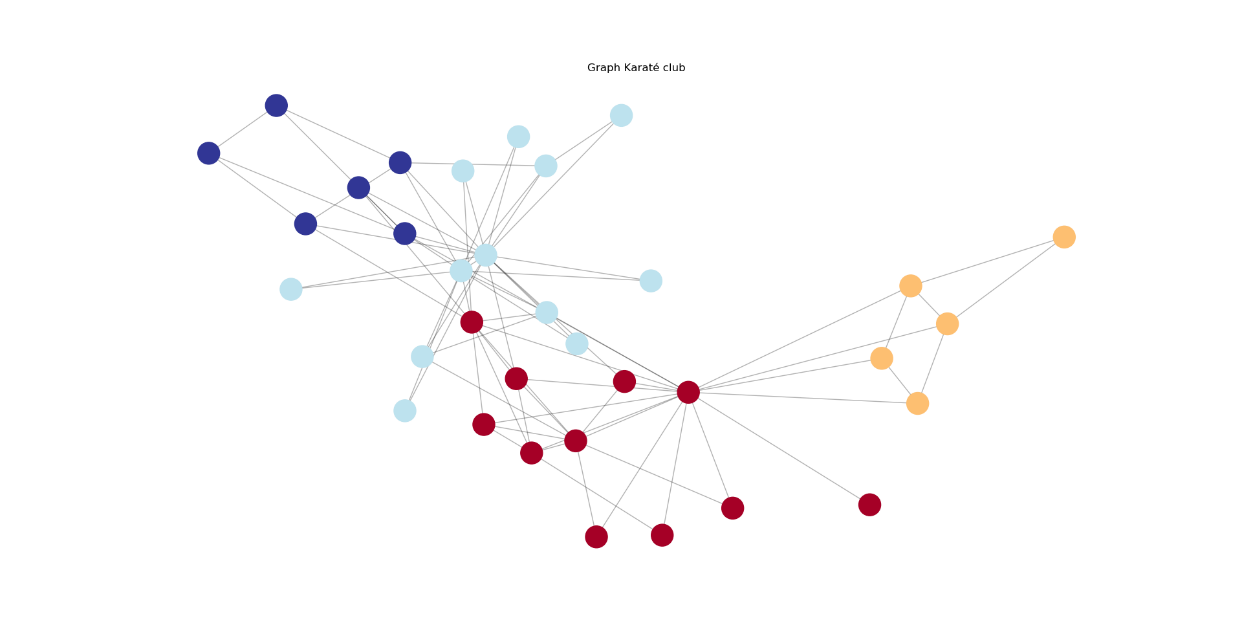


Figure : Représentation des différentes communautés au sein du graph karaté club (modularité : 0,42)

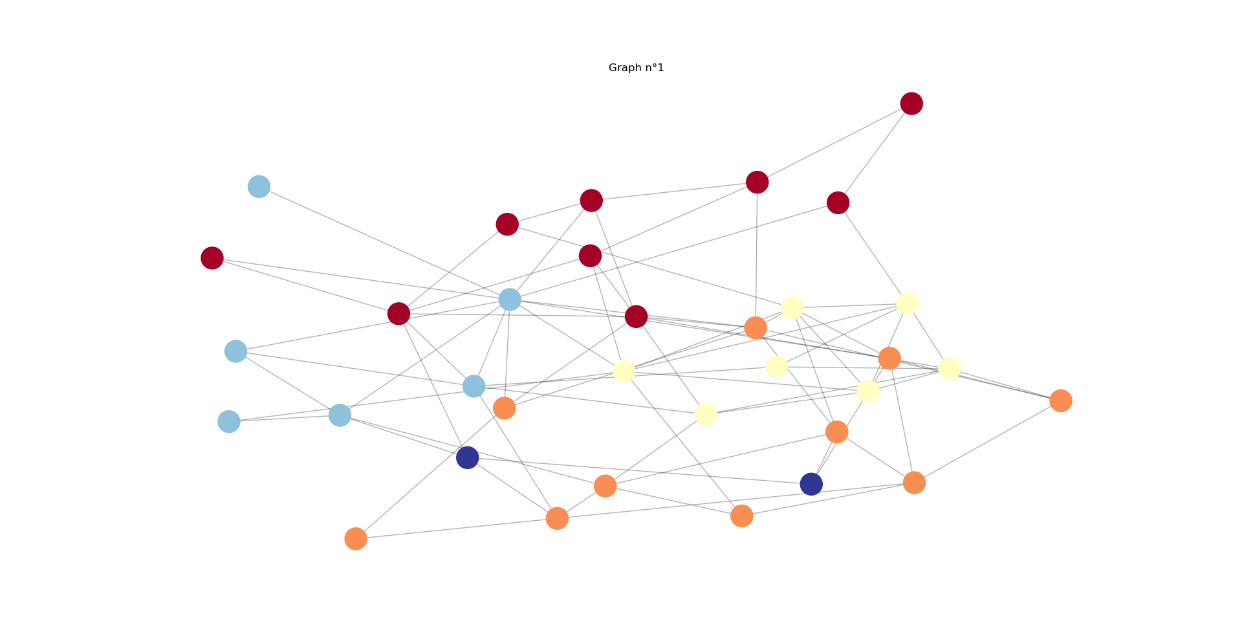


Figure : Représentation des différentes communautés au sein d'un graph aléatoire (modularité : 0,34)