

ANALYSE DES DONNÉES – DEVOIR 2

INTRODUCTION

Nous traiterons dans ce rapport le sujet **116**. Le fichier contient la matrice d'adjacence du réseau étudié. Ce réseau possède des sommets qui ont des noms.

Notre objectif est d'imaginer un contexte puis d'effectuer une analyse de ce réseau. Pour se faire, nous utiliserons les logiciels Excel et Gephi.

I. CONTEXTUALISATION

Le réseau étudié est celui d'un groupe de voyage pour personnes âgées. Il est constitué de 79 sommets qui sont les 79 voyageurs. Il existe un lien entre deux sommets si et seulement si le voyageur A déclare avoir déjà voyager avec le voyageur B et réciproquement si le voyageur B a déclaré avoir voyager avec le voyageur B. L'avantage est que les réponses erronées ne créent pas de liens.

Notre objectif est d'obtenir une visualisation intelligible de ce réseau. Pour se faire, nous ferons tout d'abord une analyse structurelle de base puis une analyse en petit monde du réseau. Nous en dégagerons ensuite les individus influents. Enfin, nous détecterons les communautés.

II. ANALYSE STRUCTURELLE

On voit qu'on a $n = 79$ noeuds et $m = 218$ liens.

Soit une densité $= 2m / n(n-1) = 7,08\% < 10\%$ donc la densité est faible.

Si on prend deux individus dans le réseau on a 7,08% de chance qu'ils soient liés.

Dans Gephi, on exécute *Composantes connexes*.

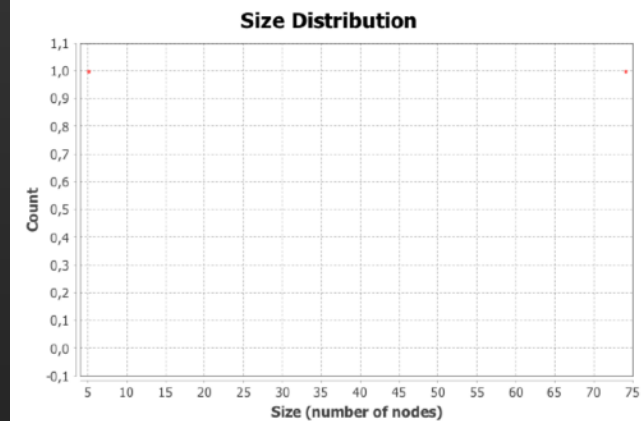
Le logiciel nous indique que le réseau contient 2 composantes connexes respectivement de taille 5 et 74.

$74/79 = 93,67\% > 90\%$

donc il y a une CC géante qui regroupe plus de 90% des noeuds.

Results:

Number of Weakly Connected Components: 2



III. ANALYSE EN PETIT MONDE

- Intermédialités entre 2 individus :

=> **moyenne des plus courts chemins**
en Executant Diamètre = **3,25**

On compare cette moyenne à $\ln(n)/\ln(k)$

avec $k=2m/n...$ $\ln(n)/\ln(k) = 2,46 \leq$ **proche donc ok**

Results:

Diameter: 7

Radius: 1

Average Path length: 3.247879011434895

- Coefficients de Clustering :

Soit C1 la moyenne de la densité du graphe des voisins. On exécute Coefficients de Clustering dans Gephi.

On obtient C1 = 0,515

Results:

Average Clustering Coefficient: 0,534

Total triangles: 193

The Average Clustering Coefficient is the mean value of individual coeff

K	L
degree	dcarré
8	64
10	100
8	64
13	169
5	25
8	64
5	25
9	81
7	49
3	9
6	36
5	25
4	16
...	...
total dcarré =	2898

Soit $C2 = 3 \cdot \text{nb de triangles} / \text{nb de Duet}$

or duet $= 1/2 \cdot \text{somme des } d_i(d_i-1)$

$= 1/2 \cdot \text{somme des } d_i^2 - m$

$= 1/2 \cdot 2898 - 218 = 1231$

On obtient $C2 = 3 \cdot 193 / 1231 = 0,47$

La probabilité que 2 individus soient liés sachant qu'ils en connaissent un même troisième est de 47%. Cette valeur est élevée comparée à la densité qui est de 7,08%.

En conclusion, il y a donc un effet petit-monde car les deux.

IV. INDIVIDUS INFLUENTS

Dans cette partie nous allons chercher à savoir qui a du pouvoir dans le réseau.

1. Dans Gephi, on trie le laboratoire de données par *Degré* décroissant pour savoir qui a le plus de pouvoir en terme de degré.

Id	Label	Interval	Componen...	Eccentricity	Closeness Centr...	Harmonic Closeness Cent...	Betweenness Centr...	Clustering Coeffi...	Number of trian...	Degré	Eigenvector Centr...
Joelle	Joelle		0	4.0	0.45625	0.531963	556.583105	0.179487	14	13	1.0
Francois	Francois		0	4.0	0.459119	0.528539	566.044909	0.19697	13	12	0.980809
Daniel	Daniel		0	4.0	0.447853	0.515982	340.489628	0.236364	13	11	0.909874

Joelle a le plus gros degré (13)

François et Daniel arrivent ensuite avec respectivement un degré de 12 et de 11.

2. On se rend dans les liens du laboratoire de données afin d'explorer la possibilité d'un club-huppé réunissant Joelle, François et Daniel. On constate que François est en lien avec Joelle, Joelle avec Daniel et Daniel avec François. Il existe donc un club-huppé entre ces individus influents.

Joelle	Marie-France
Joelle	Maryse
Joelle	Micheline
Joelle	Monique

Francois	Joel
Francois	Joelle
Francois	Martine
Francois	Maurice
Francois	Monique
Francois	Pierre
Francois	Yves

Daniel	Elisabeth
Daniel	Francois
Daniel	Francoise
Daniel	Jean-Louis
Daniel	Jean-Marie
Daniel	Joelle
Daniel	Joseph
Daniel	Monique
Daniel	Sylvie

3. On trie le laboratoire de données par *Betweenness Centrality* décroissante pour savoir par qui joue le rôle d'intermédiaire dans le réseau.

Id	Label	Interval	Componen...	Eccentricity	Closeness Centr...	Harmonic Closeness Cent...	Betweenness C...	Clustering Coeffi...	Number of trian...	Degré	Eigenvector Centr...
Francois	Francois		0	4.0	0.459119	0.528539	566.044909	0.19697	13	12	0.980809
Joelle	Joelle		0	4.0	0.45625	0.531963	556.583105	0.179487	14	13	1.0

François et Joelle jouent le rôle de relai car ils ont la plus grosse centralité d'intermédialité. De plus ils ont de faibles coefficients de clustering donc leurs liens respectifs sont diversifiés.

4. On trie le laboratoire de données par *Closeness Centrality* (notion de proximité) décroissante pour savoir qui est le plus près de tout le monde en moyenne.

Id	Label	Interval	Componen...	Eccentricity	Closeness C...	Harmonic Closeness Cent...	Betweenness Centr...	Clustering Coeffi...	Number of trian...	Degré	Eigenvector Centr...
Marie	Marie		1	1.0	1.0	1.0	0.333333	0.833333	5	4	0.036962
Marie-Cl...	Marie-Cl...		1	1.0	1.0	1.0	0.333333	0.833333	5	4	0.036962
Patricia	Patricia		1	1.0	1.0	1.0	0.333333	0.833333	5	4	0.036962
Chantal	Chantal		1	2.0	0.8	0.875	0.0	1.0	3	3	0.030269
Marie-th...	Marie-th...		1	2.0	0.8	0.875	0.0	1.0	3	3	0.030269
Francois	Francois		0	4.0	0.459119	0.528539	566.044909	0.19697	13	12	0.980809
Joelle	Joelle		0	4.0	0.45625	0.531963	556.583105	0.179487	14	13	1.0
Daniel	Daniel		0	4.0	0.447853	0.515982	340.489628	0.236364	13	11	0.909874

Marie, Marie-Claude et Patricia ont les plus grosses proximités (Closeness Centrality = 1). Toutefois, elles font partie de la petite CC (cf. component ID), leur forte proximité est donc logique. Il faut étudier la notion de proximité dans la CC géante. **Ce sont François, Joelle et Daniel qui ont les plus grosses proximités dans cette CC géante.**

5. On trie le laboratoire de données par *Eigenvector Centrality* décroissante pour connaître les individus dont les amis ont beaucoup de pouvoir.

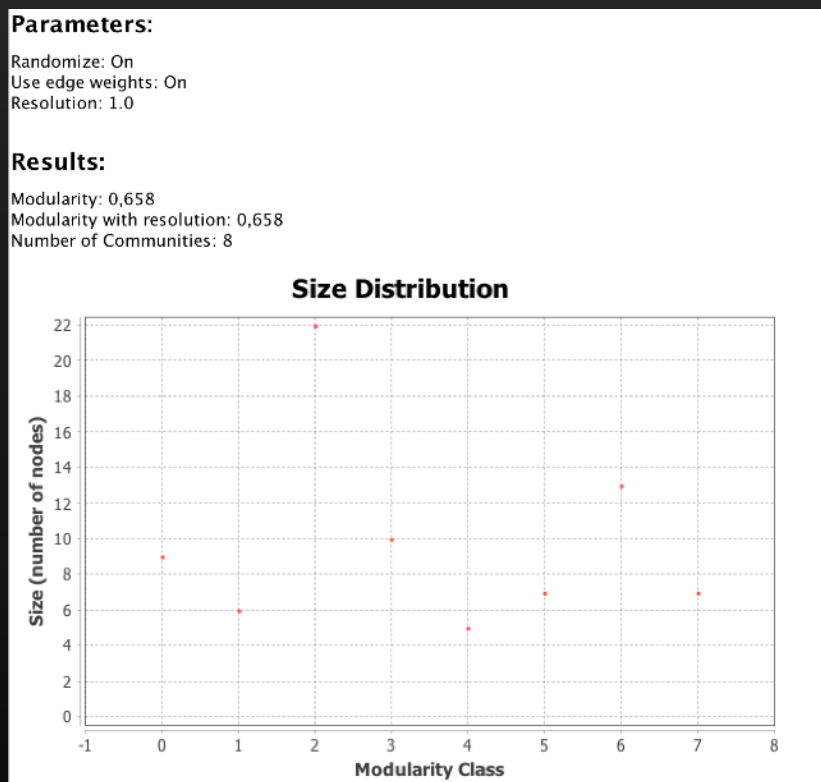
Id	Label	Interval	Componen...	Eccentricity	Closeness Centr...	Harmonic Closeness Cent...	Betweenness Centr...	Clustering Coeffic...	Number of trian...	Degré	Eigenvector C...
Joelle	Joelle	0	4.0	0.45625	0.531963	556.583105	0.179487	14	13	1.0	
Francois	Francois	0	4.0	0.459119	0.528539	566.044909	0.19697	13	12	0.980809	
Daniel	Daniel	0	4.0	0.447853	0.515982	340.489628	0.236364	13	11	0.909874	

Joelle est la personne dont les amis sont les plus influents.

En conclusion, Joelle, François et Daniel constituent un club-huppé. Joelle est la personne la plus influente du réseau : elle a non seulement le pouvoir le plus fort, mais elle est aussi un relai lié aux individus les plus influents du réseau.

V. COMMUNAUTES

Dans Gephi, on exécute Modularité afin de lire les communautés.



On a 7 communautés contenant respectivement 9, 6, 22, 10, 5, 7, 13 et 7 sommets.

V.I CONCLUSIONS

Joelle, François et Daniel sont les personnes âgées les plus influentes du réseau. Elles sont donc celles qui ont voyagé avec le plus de monde. C'est probablement celles qui font circuler l'information et « vendent » la destination aux autres personnes âgées. Il serait donc intéressant pour un organisateur ou une agence de voyage de passer par ce club-huppé pour diffuser leurs programmes de voyage. Il faut vanter les mérites de leurs destinations auprès de Joelle, François et Daniel.