



M2 DSI

LANGAGES POUR LES DONNÉES

Rapport partie 3

Mohammed Walid MATALLAH

Enseignant :
Nassira CHEKAI

Table des matières

1	Première approche : les communautés selon la modularité	2
1.1	Explication du graphe :	2
2	Deuxième approche : les communautés selon l'inférence statistiques	4
2.1	Explication du graphe :	4
3	Conclusion :	5

1 Première approche : les communautés selon la modularité

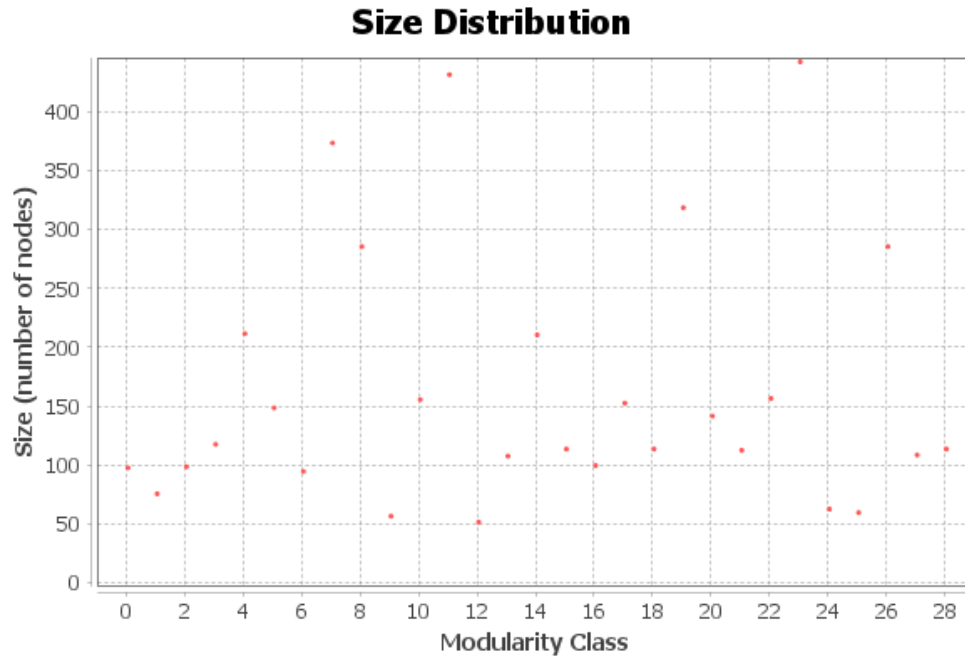


FIGURE 1 – les communautés selon l’approche des modularités

1.1 Explication du graphe :

La modularité est une mesure qui évalue la qualité de la partition d’un graphe en communautés. Elle quantifie à quel point les nœuds sont fortement liés à d’autres nœuds de la même communauté par rapport aux liens entre les nœuds de différentes communautés. Les algorithmes de détection de communautés basés sur la modularité cherchent à maximiser cette mesure pour trouver la partition la plus optimale.

1- Identification des communautés : Le processus de détection de communautés basé sur la modularité a identifié des groupes de nœuds fortement connectés au sein de notre graphe. Chaque communauté représente un sous-ensemble de nœuds qui ont une plus grande similarité ou proximité entre eux par rapport aux autres nœuds du graphe.

2- Structure du graphe : En observant notre graphe de communautés, on peut identifier des sous-graphes ou des clusters distincts qui représentent les différentes communautés détectées.

Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0,409
Modularity with resolution: 0,409
Number of Communities: 29

FIGURE 2 – Rapport généré par le logiciel gephi

2 Deuxième approche : les communautés selon l'inférence statistiques

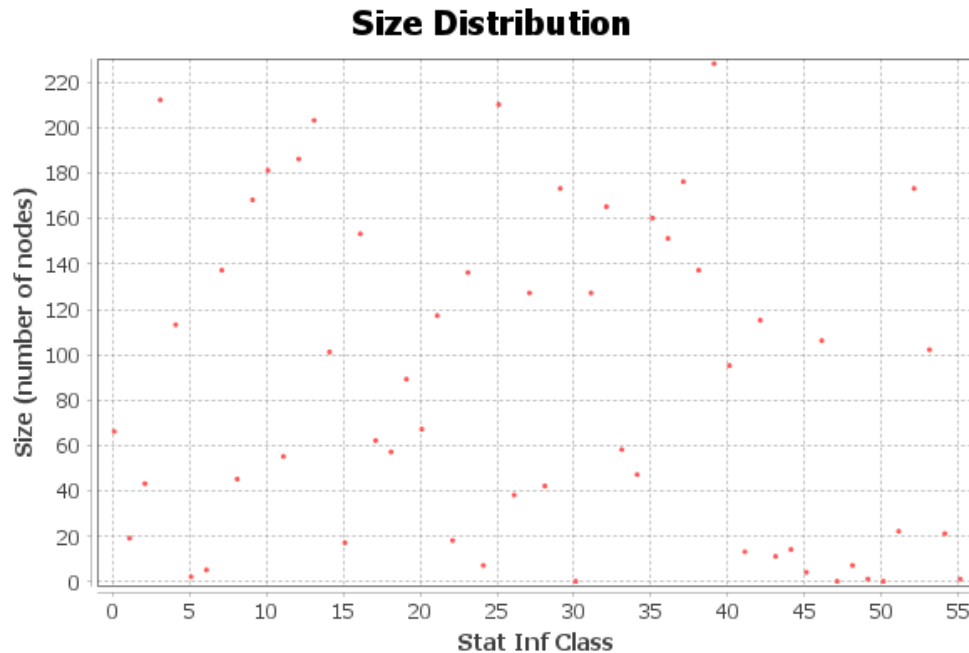


FIGURE 3 – les communautés selon l'approche de l'inférence statistique

2.1 Explication du graphe :

L'inférence statistique utilise des modèles probabilistes pour estimer les communautés dans un graphe. Ces modèles peuvent inclure des méthodes telles que le modèle de mélange, l'allocation de Dirichlet latente, ou d'autres approches statistiques pour estimer les distributions de probabilité des communautés.

Il est important de noter que la modularité est une mesure qui évalue la qualité d'une partition existante, tandis que l'inférence statistique cherche à estimer les communautés à partir des données du graphe.

1- Modèles probabilistes : L'inférence statistique pour la détection de communautés utilise des modèles probabilistes pour représenter les structures de communauté. Ces modèles peuvent inclure des distributions de probabilité telles que les modèles de mélange, les modèles de blocs stochastiques ou d'autres modèles adaptés à notre cas d'étude spécifique.

2- Estimation des communautés : Les algorithmes d'inférence statistique cherchent à estimer les paramètres du modèle probabiliste qui correspondent aux communautés dans notre graphe. Ces estimations sont basées

sur les données du graphe, telles que les liens, les attributs des nœuds ou d'autres informations disponibles.

Statistical Inference Report

Results:

Description Length: 82100,161

Number of Communities: 56



Algorithm:

Statistical inference of assortative community structures

Lizhi Zhang, Tiago P. Peixoto

Phys. Rev. Research 2 043271 (2020)

<https://dx.doi.org/10.1103/PhysRevResearch.2.043271>

Bayesian stochastic blockmodeling

Tiago P. Peixoto

Chapter in “Advances in Network Clustering and Blockmodeling,” edited by

P. Doreian, V. Batagelj, A. Ferligoj (Wiley, 2019)

<https://dx.doi.org/10.1002/9781119483298.ch11>

FIGURE 4 – logiciel généré par le logiciel gephi

3 Conclusion :

En conclusion, tant l'approche de modularité que l'inférence statistique sont des méthodes puissantes pour détecter les communautés dans un graphe. Voici quelques points clés à retenir pour chaque approche :

Approche de modularité : - L'approche de modularité évalue la force des liens internes à chaque communauté

par rapport aux liens externes entre les communautés. - Les communautés sont identifiées en maximisant la modularité du graphe. - La modularité permet de mettre en évidence les communautés qui ont des liens plus forts à l'intérieur de la communauté que vers l'extérieur. - Les communautés détectées peuvent fournir des informations sur les groupes d'entités qui partagent des caractéristiques similaires ou des interactions étroites.

Inférence statistique : - L'inférence statistique utilise des modèles probabilistes pour estimer les communautés dans le graphe. - Les communautés sont estimées en utilisant des modèles probabilistes et des techniques d'estimation des paramètres. - Les modèles probabilistes permettent de représenter la structure des communautés et de fournir des estimations de probabilité pour les communautés détectées. - Les résultats basés sur l'inférence statistique peuvent aider à comprendre les interactions entre les communautés et à identifier les caractéristiques distinctes de chaque communauté.

Dans les deux approches, il est essentiel de considérer le contexte spécifique de notre étude et d'adapter l'interprétation en fonction des données du graphe. L'analyse des communautés dans un graphe peut fournir des informations précieuses sur les structures, les relations et les dynamiques présentes dans le système étudié.