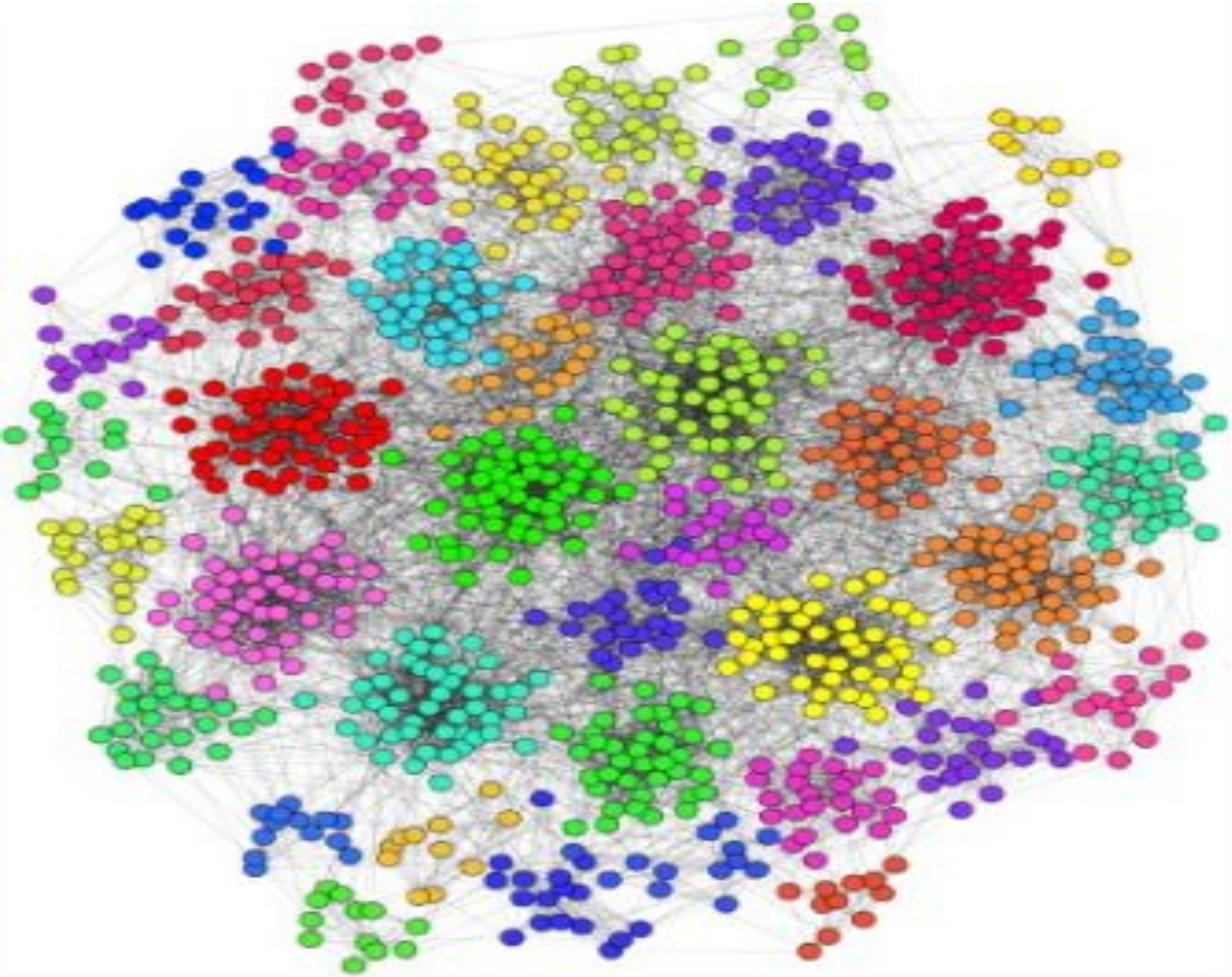
INCREMENTAL-IO

Numerical Linear Algebra Project

Team Members:

- Anna Araslanova
- Sergey Ermakov
- Artem Oborevich
- Bakhyt Zharkynbay

K smallest eigenvalues and graph clustering



Graph Laplacian

$L = S - W$

W is a non-negative weight matrix,

$W_{i,j} \geq 0$, shows the degree of connection between nodes.

$S = diag(s_1, s_2, \ldots s_n)$ where

$s_i = \sum_{j=1}^{n} W_{i,j}$

Normalized Laplacian

$L_N = S^{\frac{-1}{2}} L S^{\frac{-1}{2}} = I - S^{\frac{-1}{2}} W S^{\frac{-1}{2}}$

The Problem - We don’t know optimal number of clusters beorehand

K is generally much smaller than number of data points. Thus, full eigen decompositions are not needed. Instead we only need to compute K leading eigenpairs. One way to compute is Lanczos algorithm.

1. Obtain the K leading eigenpairs $\{t_i, \mathbf{u}_i\}_{i=1}^{K}$ of $\mathbf{T}$. $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_K]$.
2. Residual error $= |\mathbf{T}(Z-1, Z) \cdot \mathbf{U}(Z, K)|$

while Residual error > Tolerence **do**

- 2-1. $Z = Z + Z_{aug}$
- 2-2. Based on $\mathbf{Q}$ and $\mathbf{T}$, compute the next Z_{aug} Lanczos vectors as columns of $\mathbf{Q}_{aug}$ and the augmented tridiagonal matrix $\mathbf{T}_{aug}$
- 2-3. $\mathbf{Q} \leftarrow [\mathbf{Q}\ \mathbf{Q}_{aug}]$ and $\mathbf{T} \leftarrow \begin{bmatrix} \mathbf{T} & \mathbf{O} \\ \mathbf{O} & \mathbf{T}_{aug} \end{bmatrix}$
- 2-4. Go back to step 1

New methodology - incremental computation of eigenpairs

Let $V_k = [v_1(L), v_2(L), \dots v_k(L)]$
$s = \sum_{i=1}^{n} s_i$
$\Lambda_k = diag(s - \lambda_1(L), s - \lambda_2(L), \dots s - \lambda_k(L))$

The eigenpair $\lambda_{k+1}(L), v_{k+1}(L)$ is a leading eigenpair of the matrix:
$\bar{L} = L + V_k \Lambda_k V_k^T + \frac{s}{n} 1_n 1_n^T - sI$

$$L + V_k \Lambda_k V_k^T = \sum_{i=K+1}^{n} \lambda_i(L) v_i(L) v_i^T(L) + \sum_{i=2}^{K} s v_i(L) v_i^T(L)$$

$$\bar{L} = \sum_{i=K+1}^{n} (\lambda_i(L) - s) v_i(L) v_i^T(L)$$

Since $\lambda_{K+1}(L) \leq \lambda_{K+i}(L), |\lambda_{K+1}(L) - s| \geq |\lambda_{K+i}(L) - s|$

Algorithm residuals on a test graph

Unnormalized Laplacian Matrix

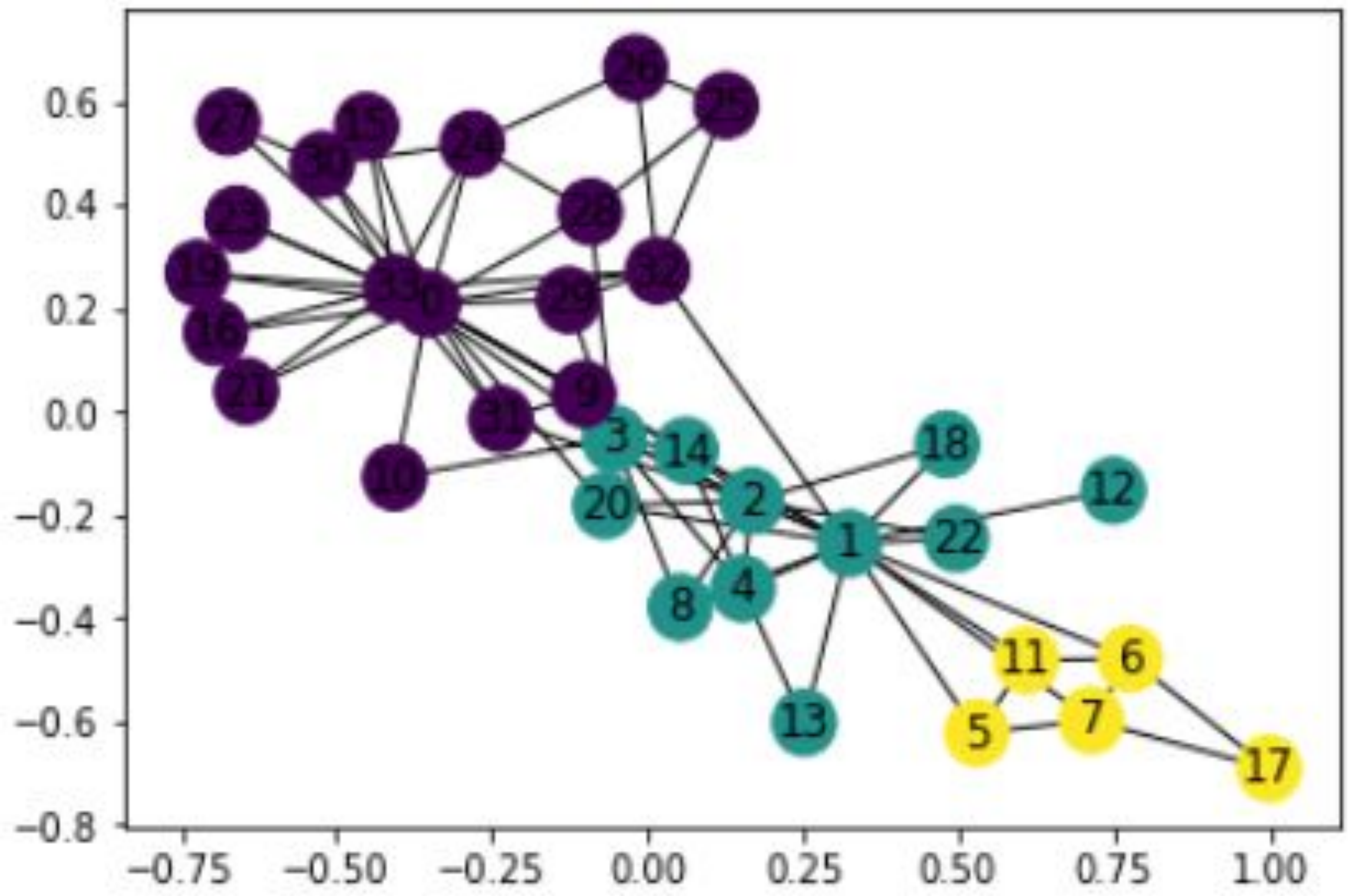
```
Residual 1 : 3.26557197381e-14
Residual 2 : 1.29763397322e-14
Residual 3 : 5.98099251161e-14
Residual 4 : 5.05205108252e-14
Residual 5 : 4.95417387689e-14
Residual 6 : 3.80776131839e-14
Residual 7 : 4.34297459286e-14
Residual 8 : 2.61912193815e-14
Residual 9 : 4.49507744719e-14
Residual 10 : 3.52477658537e-14
```

Normalized Laplacian Matrix

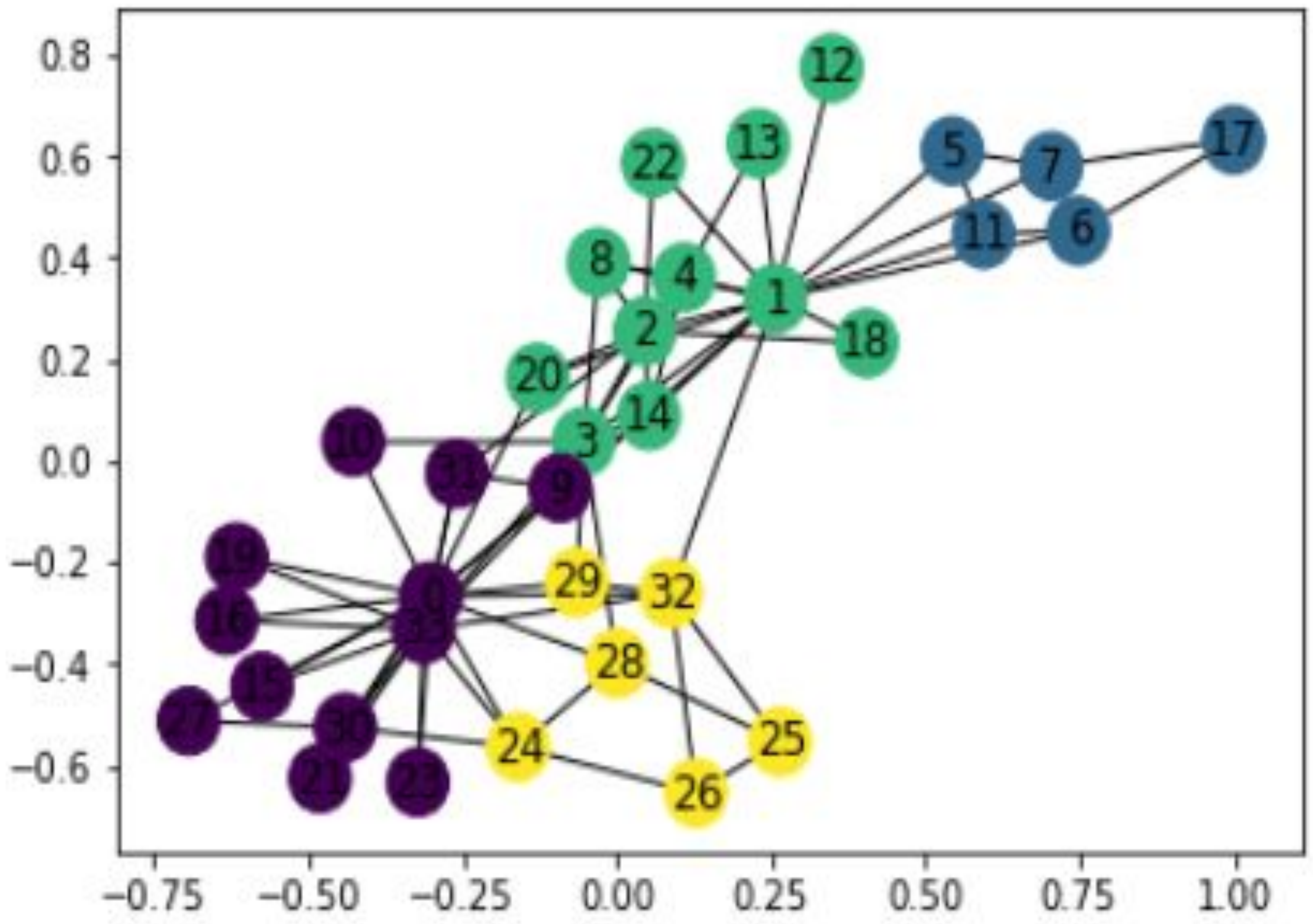
```
Residual 1 : 1.94166845411e-15
Residual 2 : 1.70556513324e-15
Residual 3 : 1.60055810515e-15
Residual 4 : 1.78671960108e-15
Residual 5 : 8.84135222496e-16
Residual 6 : 1.20980391286e-15
Residual 7 : 9.48857666671e-16
Residual 8 : 7.06315835472e-16
Residual 9 : 7.41741874958e-16
Residual 10 : 1.18492980716e-15
```

Clusterings under different metrics

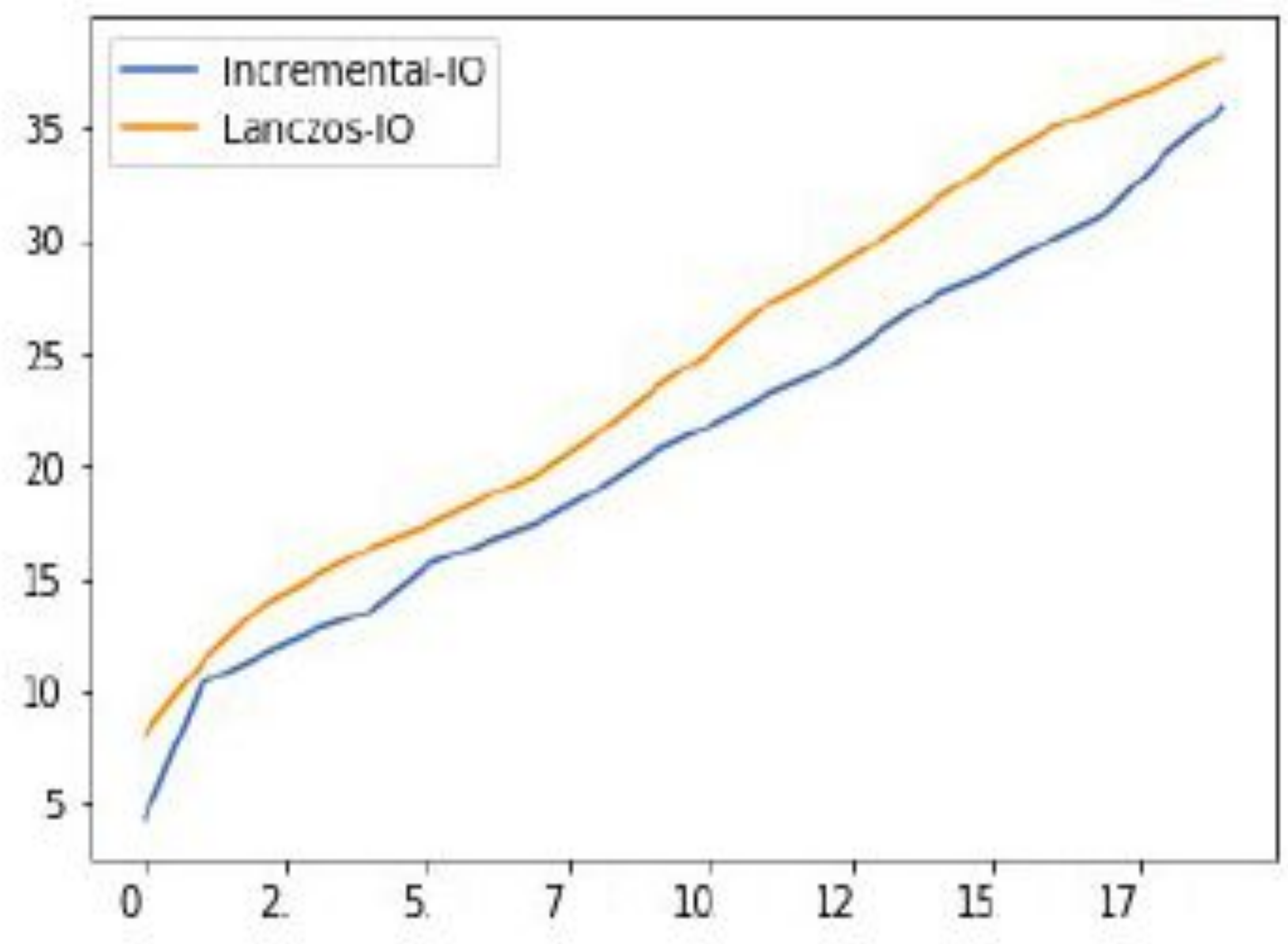
Modularity



Scaled median



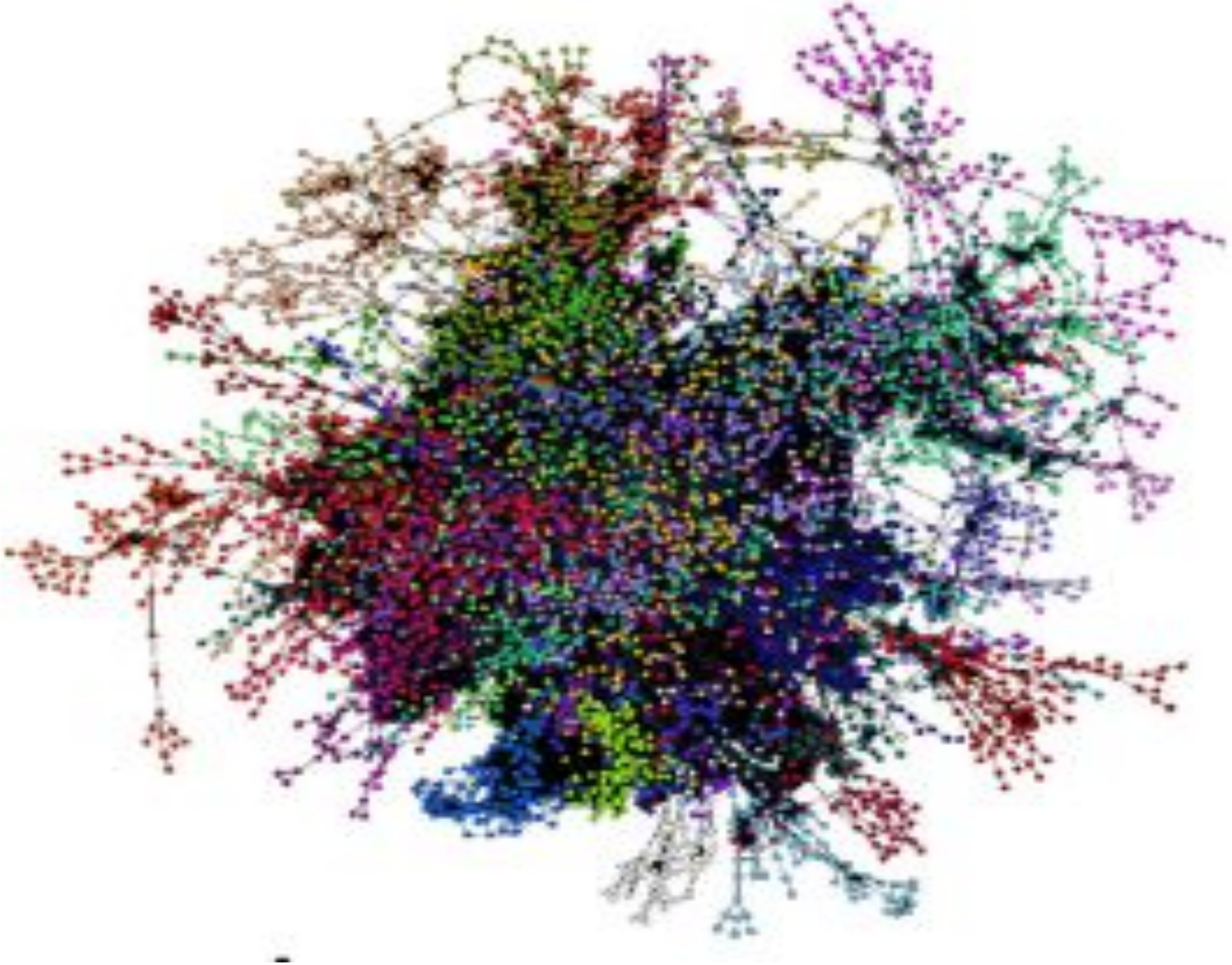
Comparison of Lanczos with Incremental



Iterations

DATASETS WHICH WERE USED

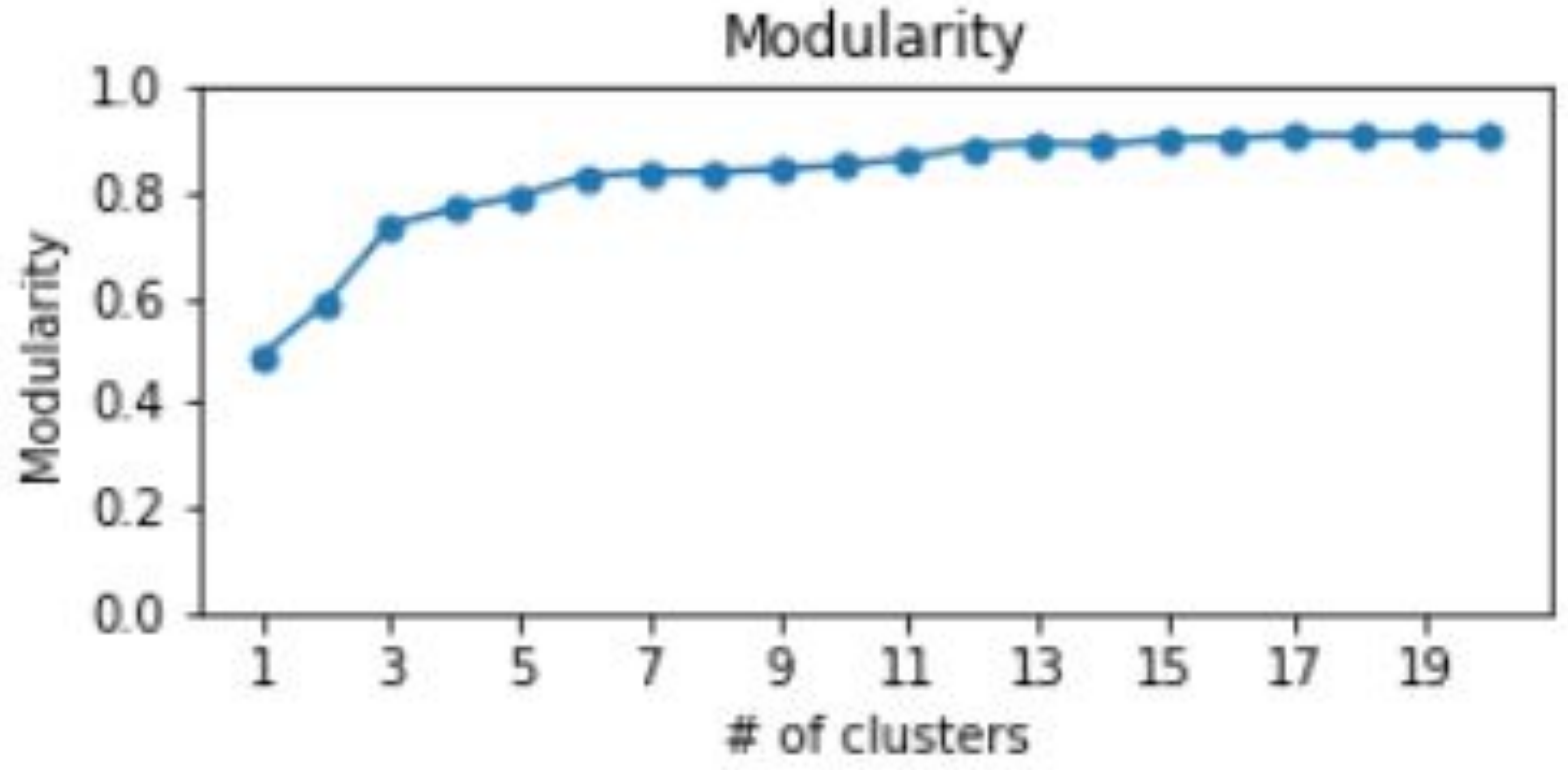
Power grid: An undirected, unweighted network representing the topology of the Western States Power Grid of the United States. Data compiled by D. Watts and S. Strogatz



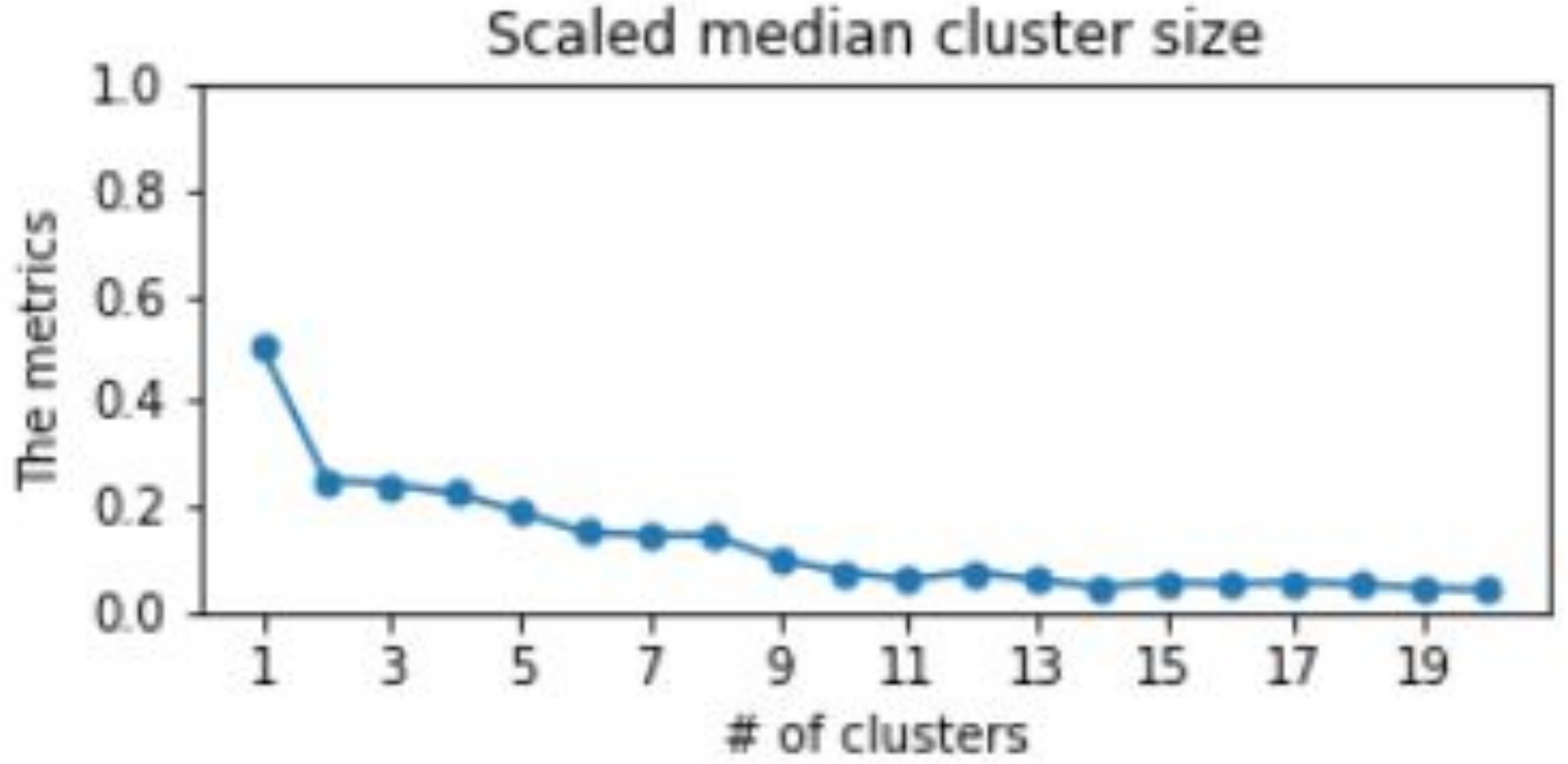
UNNORMALIZED LAPLACIAN MATRIX

Power grid

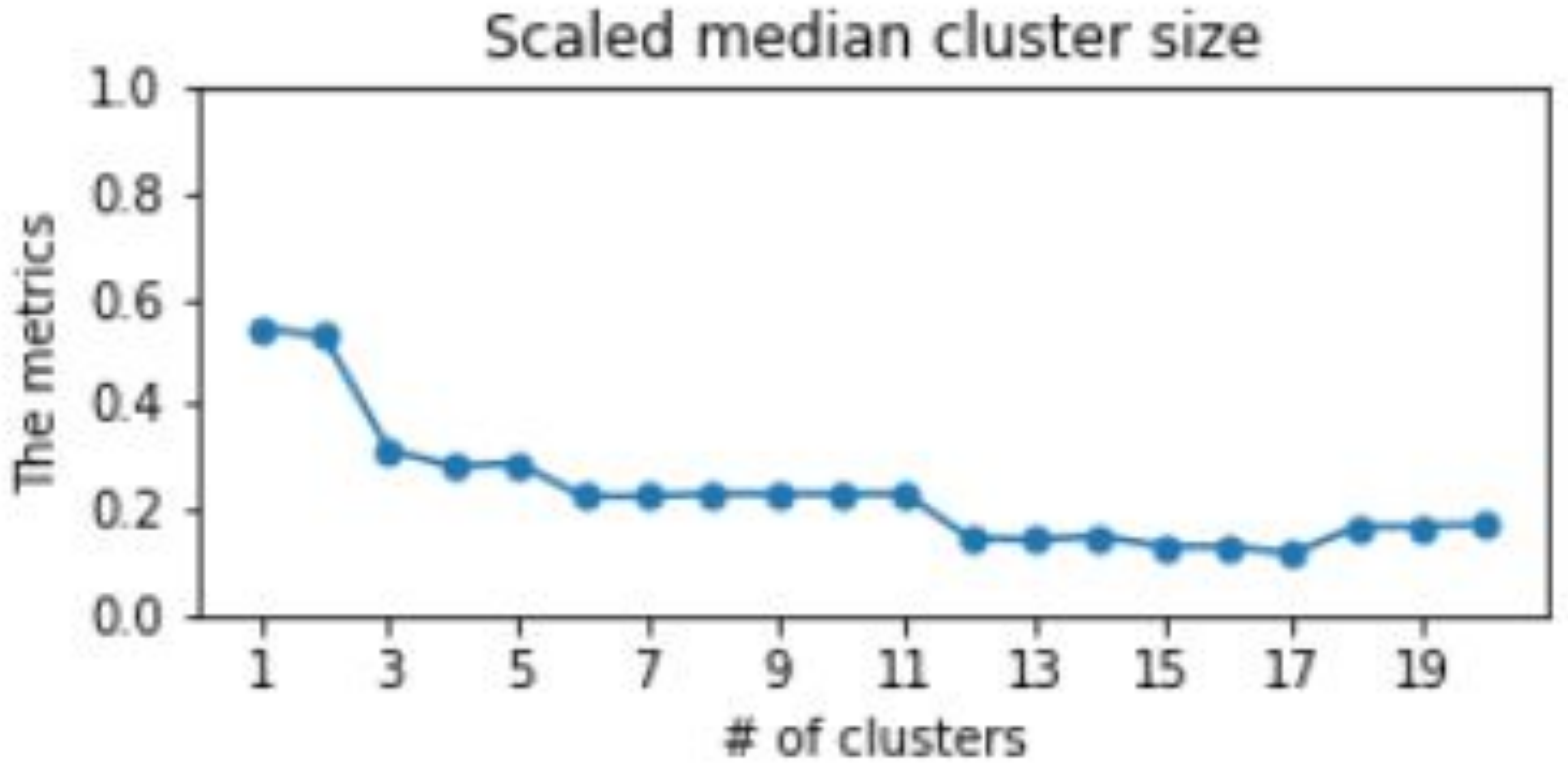
Partition with modularity settings



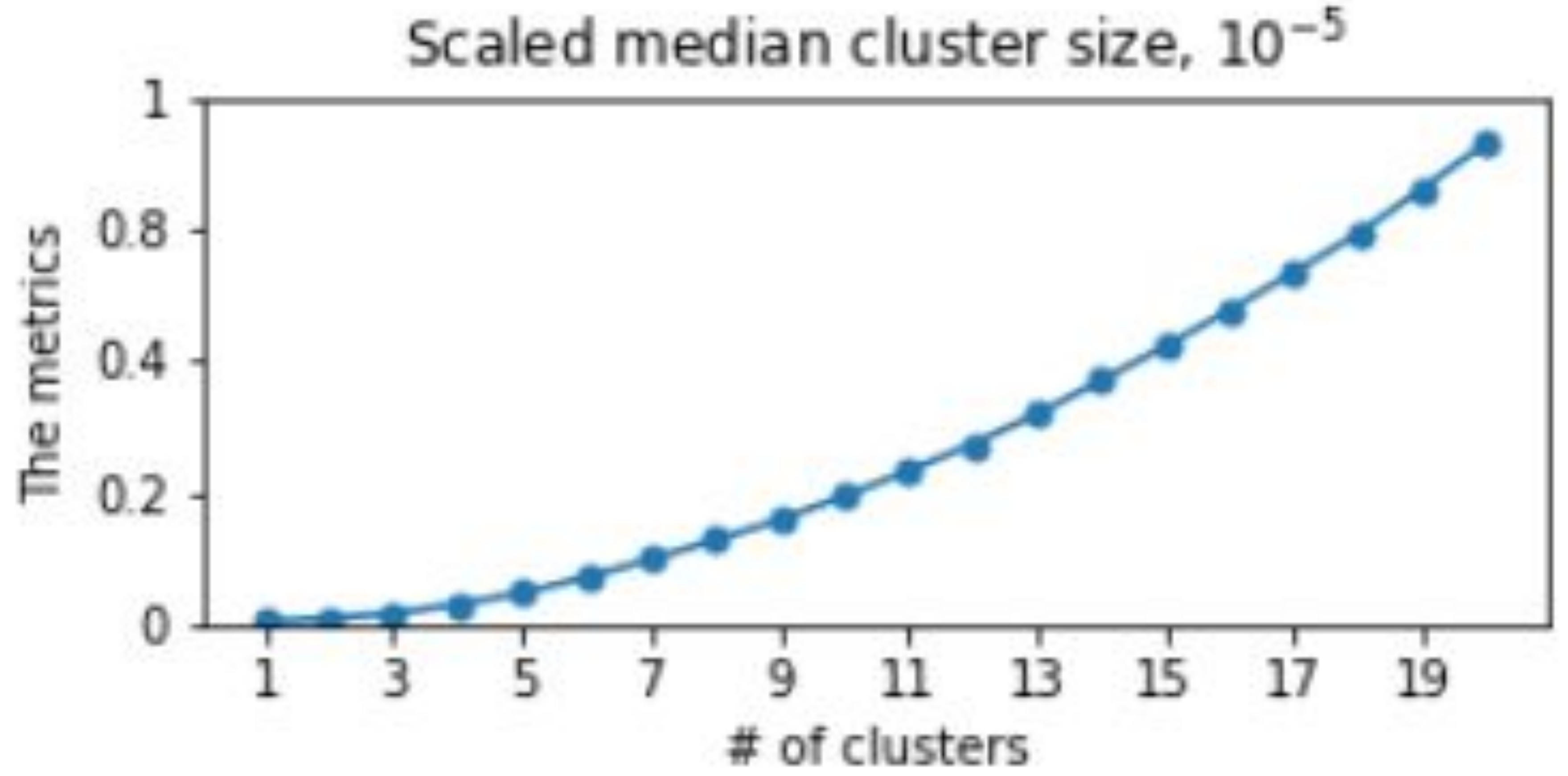
Partition with suitable scaled median cluster size



Partition with suitable scaled maximum cluster size



Partition with suitable scaled spectrum energy



Reference paper

Chen, P. Y., Zhang, B., Hasan, M. A., & Hero, A. O. (2015). Incremental method for spectral clustering of increasing orders. arXiv preprint arXiv:1512.07349. https://arxiv.org/pdf/1512.07349.pdf