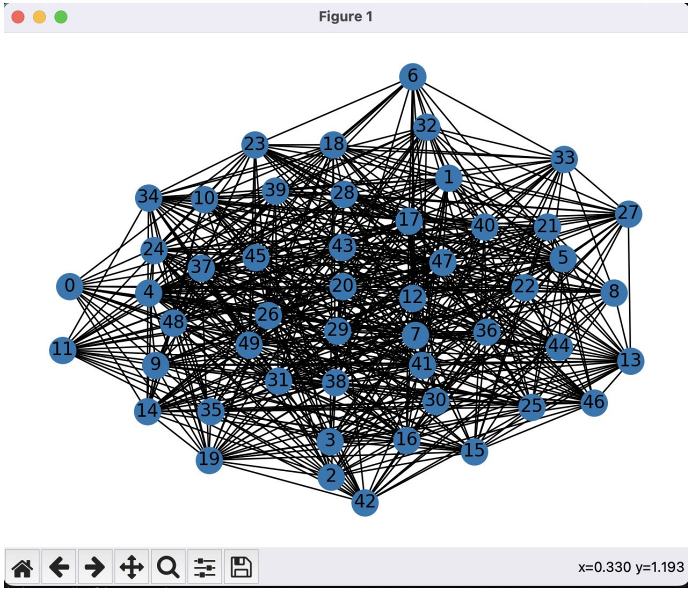
# Spectral Clustering and Cheeger's Inequality

Kaicheng Guo & Zhengyu Zou

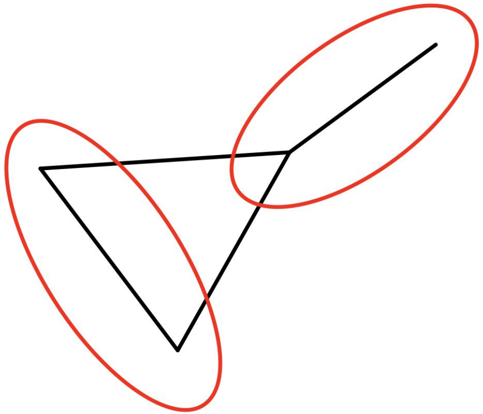
# How do you  find “good” clusters in an extremely large and messy graph ?

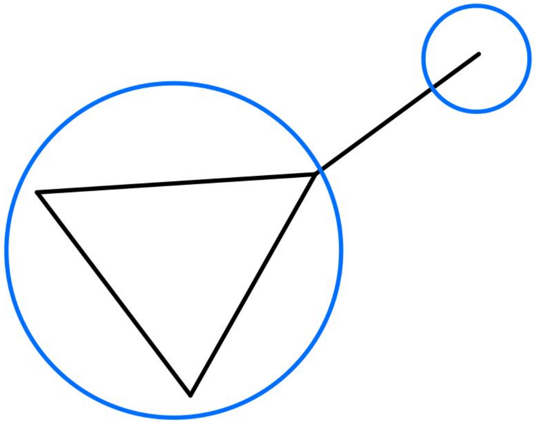



## Which Cut Gives “Good” Clusters?

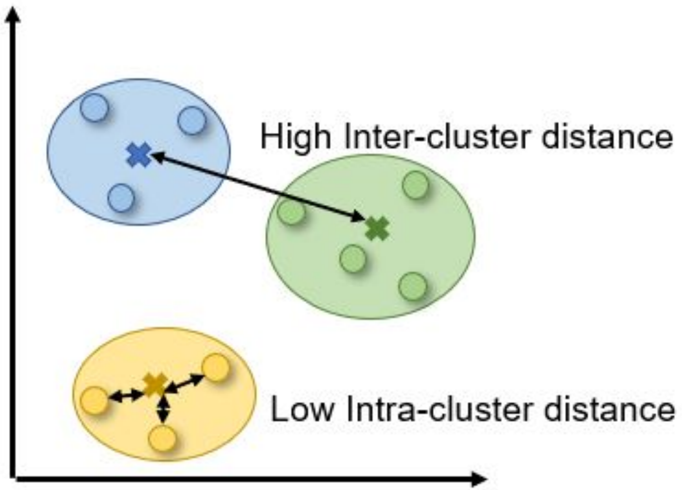
- A good cluster has the following properties:
  - Only need to move a small number of edges
  - The two subgraphs have similar number of vertices / edges. I.e., they are “balanced in sizes”

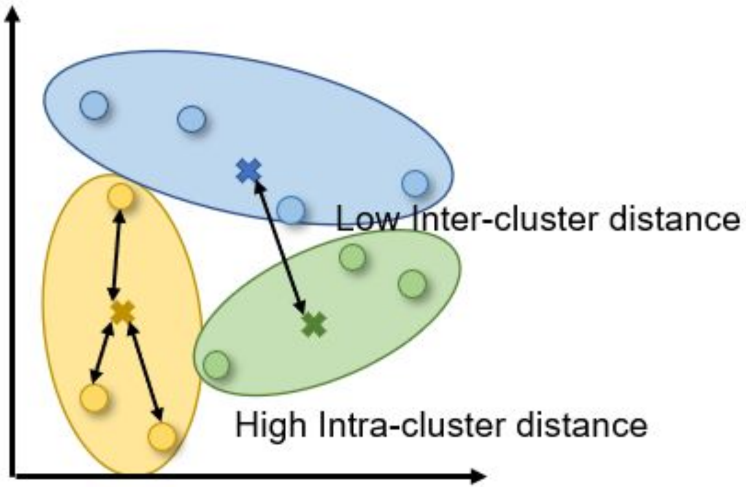
# Examples of clusters :(





# Examples of Clusters :D



# Spectral Approach of a “Good” Cluster

## Volume of a Subgraph

- Volume = sum of the degrees of all vertices in the subgraph
- A “larger” graph has more vertices -> larger volume
- A “denser” graph has more edges -> larger volume
- This is a very good classifier of good clusters: BOTH sides needs to be large and dense (“balanced”)

# Conductance of a Cut

- The conductance is computed by the “edge flux” divided by the minimum volume of the cut and its “complement cut”
- Measures whether a cut is a “good” cut in the following sense:
    - Deleting a small number of edges – Minimizing edge flux
    - “Balanced in size” – maximizing the smaller one of the two volumes

# Conductance of a Graph

- The conductance of a graph is simply given by the minimum conductance of any possible cuts of a graph.
- A “good” cluster yields the minimum conductance

# How to Find “Good” Clusters?

## Naive Way: Consider All Possible Clusters

- Consider every nonempty subset S of V as potential candidates of our "good" cluster
- Calculate the conductance of each candidate
- Return the candidate with the smallest conductance
- Runtime: O(2^{|V|}) :=(

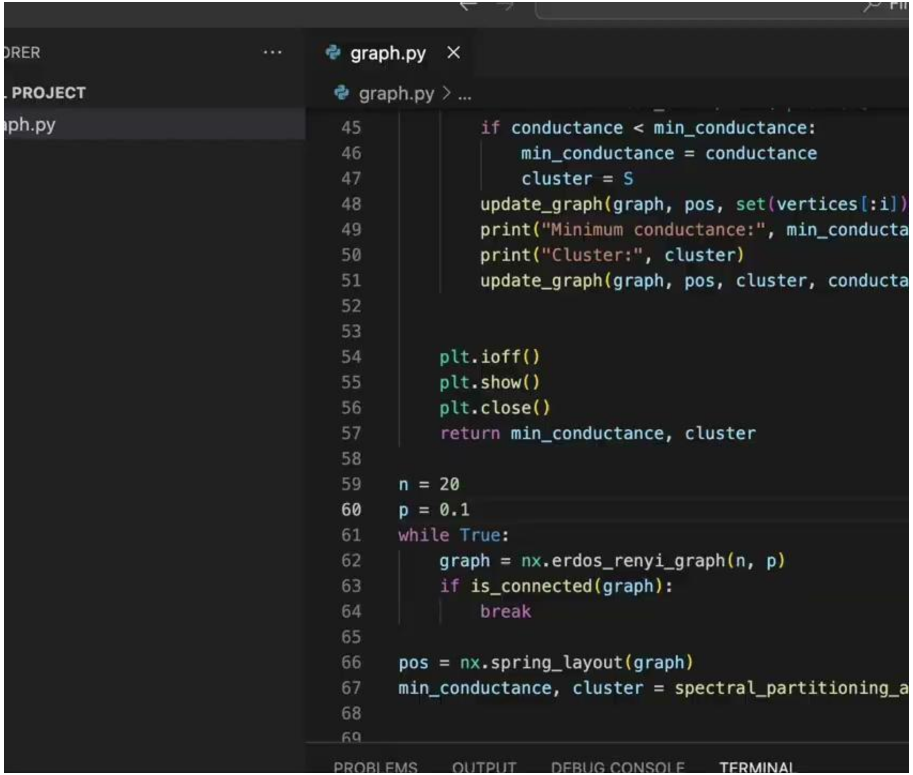
## A Better Way: Spectral Clustering

- Our good old friend: Laplacian Matrix L
- Laplacian matrix gives information about edge flux. Very useful
- Want to find a vector x such that $x^TLx/x^Tx$ is minimized
- Fact: the smallest non-zero eigenvector x gives $\min(x^TLx/x^Tx)$
    - Be careful! L has a 0 eigenvector.

## Algorithm (Spectral Clustering)

- Find a good approximation of the eigenvector $x$ of the second smallest eigenvalue of $L$ (can use the power iteration method)
- Sort the vertices of $V$ according to $x$: $x(v_1)\leq x(v_2)\leq \ldots\leq x(v_n)$
- Iterate over the cuts $S_i$ with each time adding $v_i$ to the cut. For each iteration calculate the conductance of $S_i$
- Return minimum of conductance

# Demonstration of the Algorithm



# Applications

- PageRank: classification of websites in to topics
- Community detection: detect communities or groups of nodes with similar connectivity patterns in social networks, biological networks, and other complex networks.
- Recommendation Systems: Spectral clustering can be used in recommendation systems to group similar items or users based on their preferences, ratings, or behavior.

# References

GPT4 (for the algorithm implementation in python)

Chang, K. C., et al. “THE 1-LAPLACIAN CHEEGER CUT: THEORY AND ALGORITHMS.” Journal of Computational Mathematics, vol. 33, no. 5, 2015, pp. 443–67. JSTOR, http://www.jstor.org/stable/43693873. Accessed 7 Apr. 2023.

Arsić, Branko, et al. “GRAPH SPECTRAL TECHNIQUES IN COMPUTER SCIENCES.” Applicable Analysis and Discrete Mathematics, vol. 6, no. 1, 2012, pp. 1–30. JSTOR, http://www.jstor.org/stable/43666153. Accessed 7 Apr. 2023.

Cvetković, Dragoš, and Slobodan K. Simić. “TOWARDS A SPECTRAL THEORY OF GRAPHS BASED ON THE SIGNLESS LAPLACIAN, III.” Applicable Analysis and Discrete Mathematics, vol. 4, no. 1, 2010, pp. 156–66. JSTOR, http://www.jstor.org/stable/43671298. Accessed 7 Apr. 2023.