Spectral Graph Partitioning: Finding a partition

Mining of Massive Datasets
Leskovec, Rajaraman, and Ullman
Stanford University

λ_2 as optimization problem

- **Fact:** **For symmetric matrix M:**

$$\lambda_2 = \min_x \frac{x^T M\, x}{x^T x}$$

- **What is the meaning of min $x^T L\, x$ on G?**
  - $x^T L\, x = \sum_{i,j=1}^{n} L_{ij}\, x_i x_j = \sum_{i,j=1}^{n} \left(D_{ij} - A_{ij}\right) x_i x_j$
  - $= \sum_i D_{ii} x_i^2 - \sum_{(i,j)\in E} 2x_i x_j$
  - $= \sum_{(i,j)\in E} (\underbrace{x_i^2 + x_j^2} - 2x_i x_j) = \sum_{(i,j)\in E} \left(\boldsymbol{x_i} - \boldsymbol{x_j}\right)^{\mathbf{2}}$

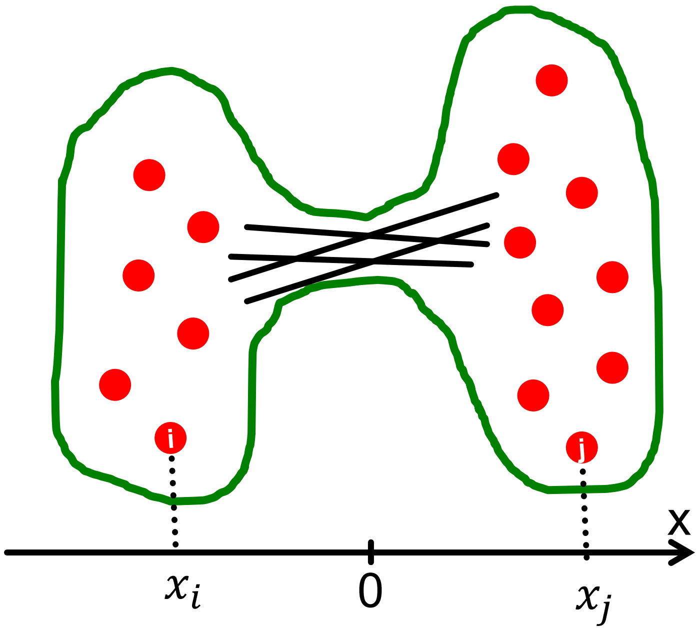
Node i has degree d_i. So, value x_i^2 needs to be summed up d_i times.
But each edge (i, j) has two endpoints so we need $x_i^2 + x_j^2$

λ_2 as optimization problem

- **What else do we know about x?**
  - $\boldsymbol{x}$ is unit vector: $\sum_i \boldsymbol{x}_i^2 = \mathbf{1}$
  - $\boldsymbol{x}$ is orthogonal to **1**st eigenvector $(\mathbf{1}, \dots, \mathbf{1})$ thus: $\sum_i \boldsymbol{x}_i \cdot \mathbf{1} = \sum_i \boldsymbol{x}_i = \mathbf{0}$
- **Remember:**

$$\lambda_2 = \min_{\substack{\text{All labelings} \\ \text{of nodes } i \text{ so} \\ \text{that } \sum x_i = 0}} \frac{\sum_{(i,j)\in E} (x_i - x_j)^2}{\sum_i x_i^2}$$

We want to assign values x_i to nodes i such that few edges cross 0.
(we want x_i and x_j to subtract each other)



Balance to minimize

Find Optimal Cut [Fiedler'73]

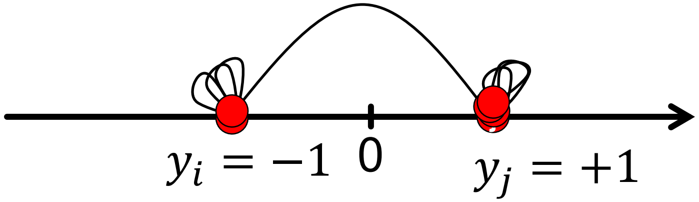
- **Back to finding the optimal cut**
- **Express partition (A,B) as a vector**

$$y_i = \begin{cases} +1 & if\ i \in A \\ -1 & if\ i \in B \end{cases}$$

- We can minimize the cut of the partition by finding a non-trivial vector x that **minimizes**:

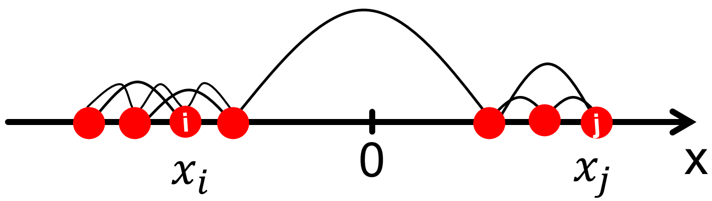
$$\arg\min_{y \in [-1,+1]^n} f(y) = \sum_{(i,j) \in E} (y_i - y_j)^2$$

Can't solve exactly. Let's relax y and allow y to take any real value.



Rayleigh Theorem

$$\min_{y \in \Re^n} f(y) = \sum_{(i,j)\in E} (y_i - y_j)^2 = y^T L y$$



- $\lambda_2 = \min_y f(y)$: The minimum value of $f(y)$ is given by the 2nd smallest eigenvalue λ_2 of the Laplacian matrix L
- $\mathbf{x} = \arg\min_y f(y)$: The optimal solution for y is given by the corresponding eigenvector x, referred as the **Fiedler vector**