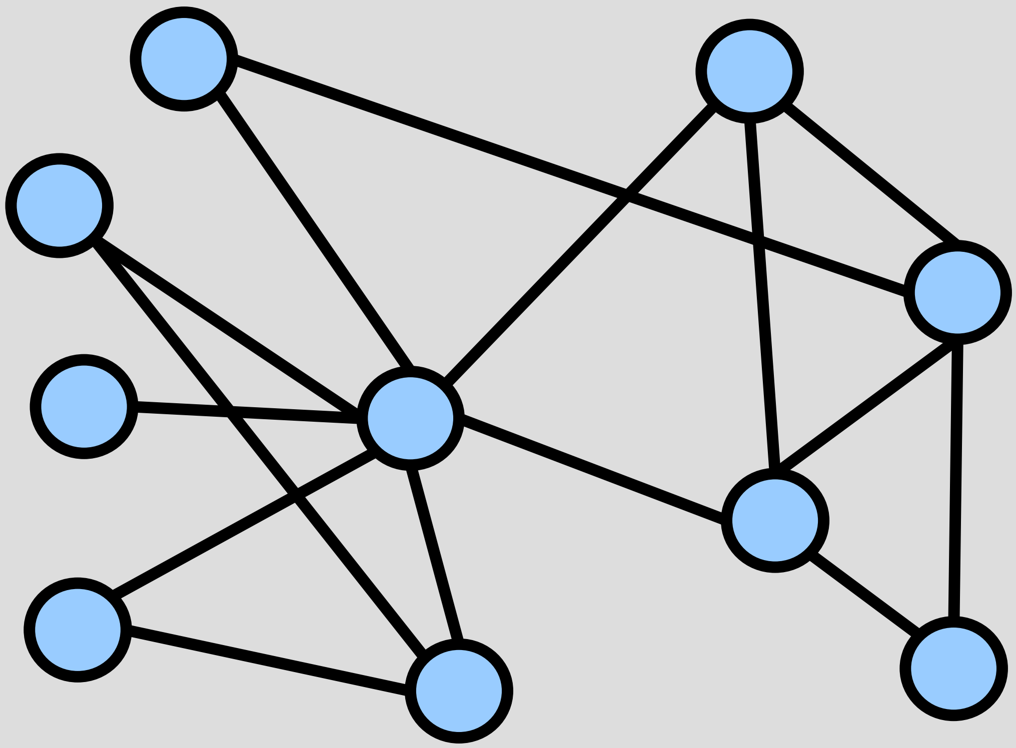
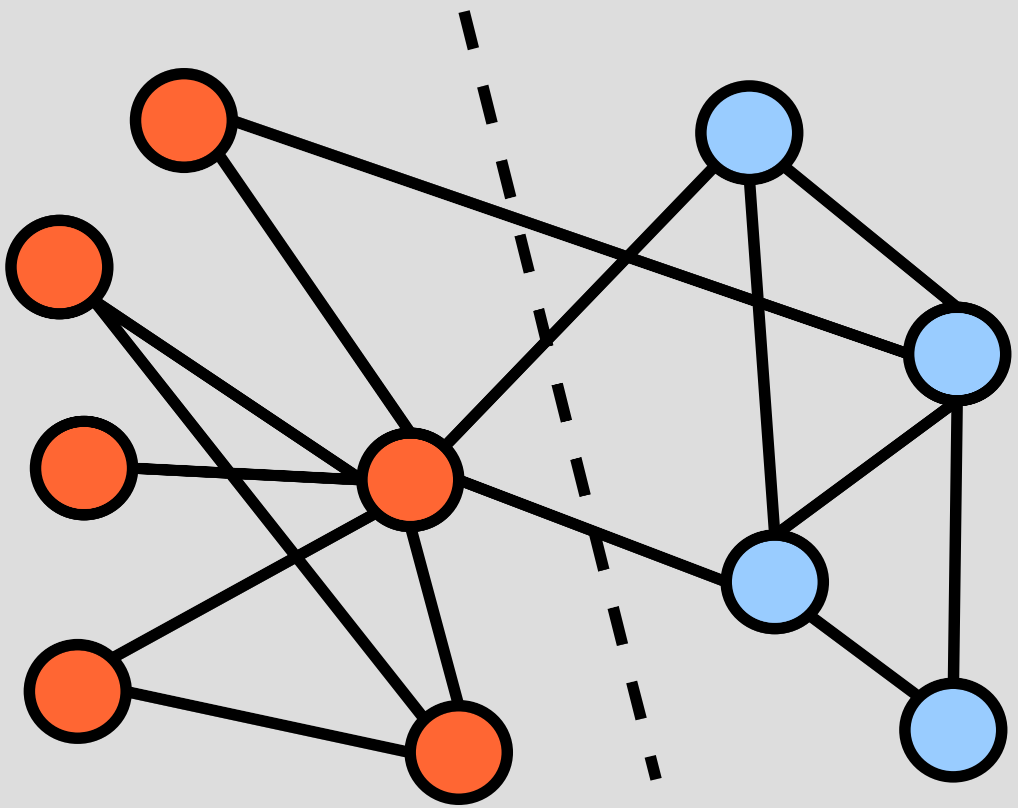
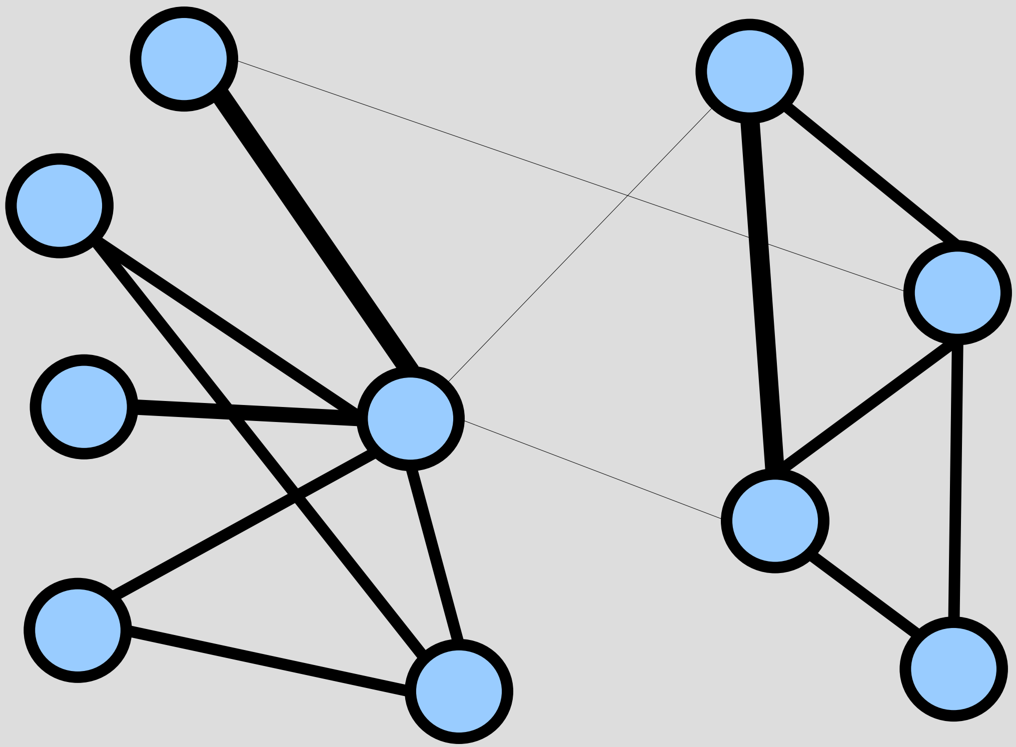
# How to Cluster?



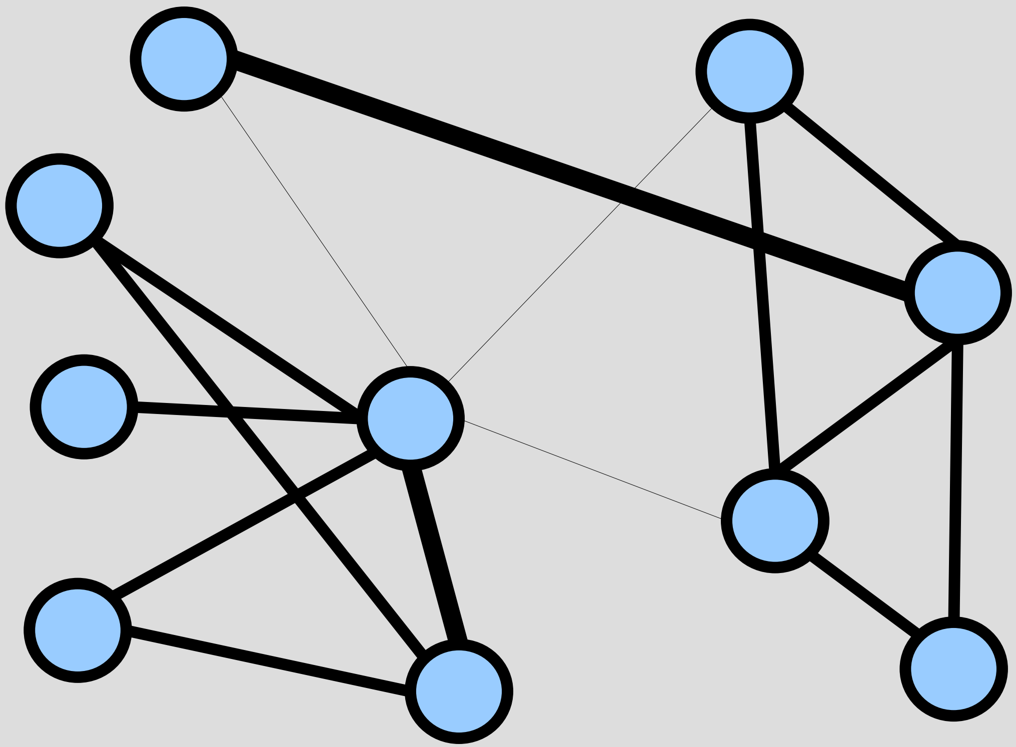
# How to Cluster?



# How to Cluster?



# How to Cluster?

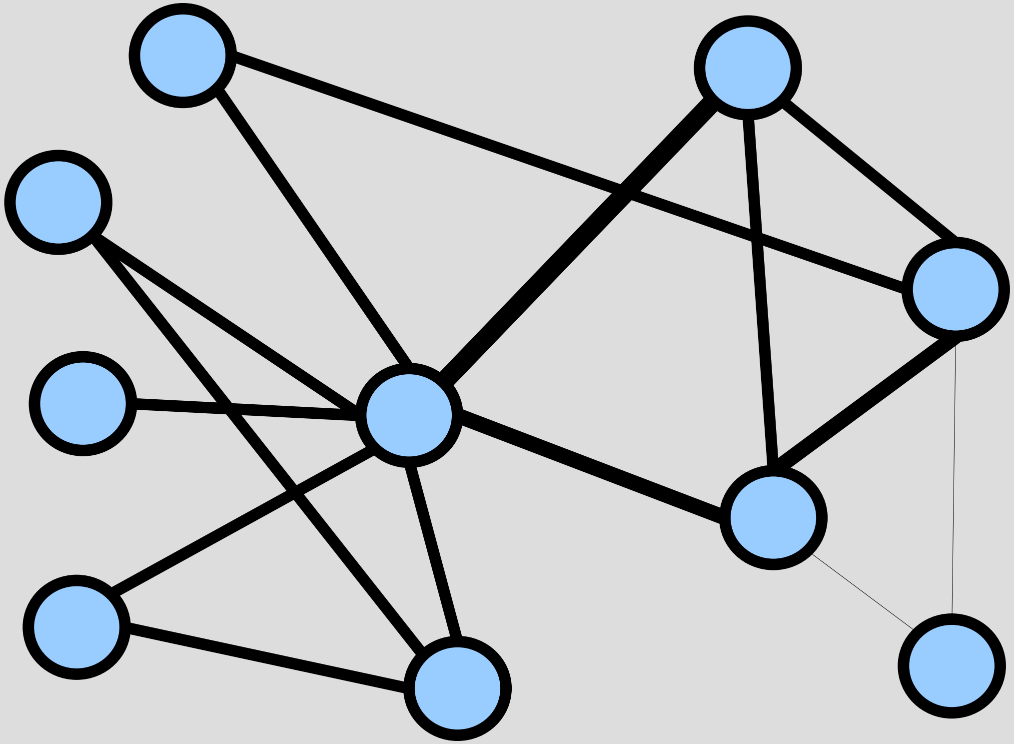


# Minimum "Cut" Criterion

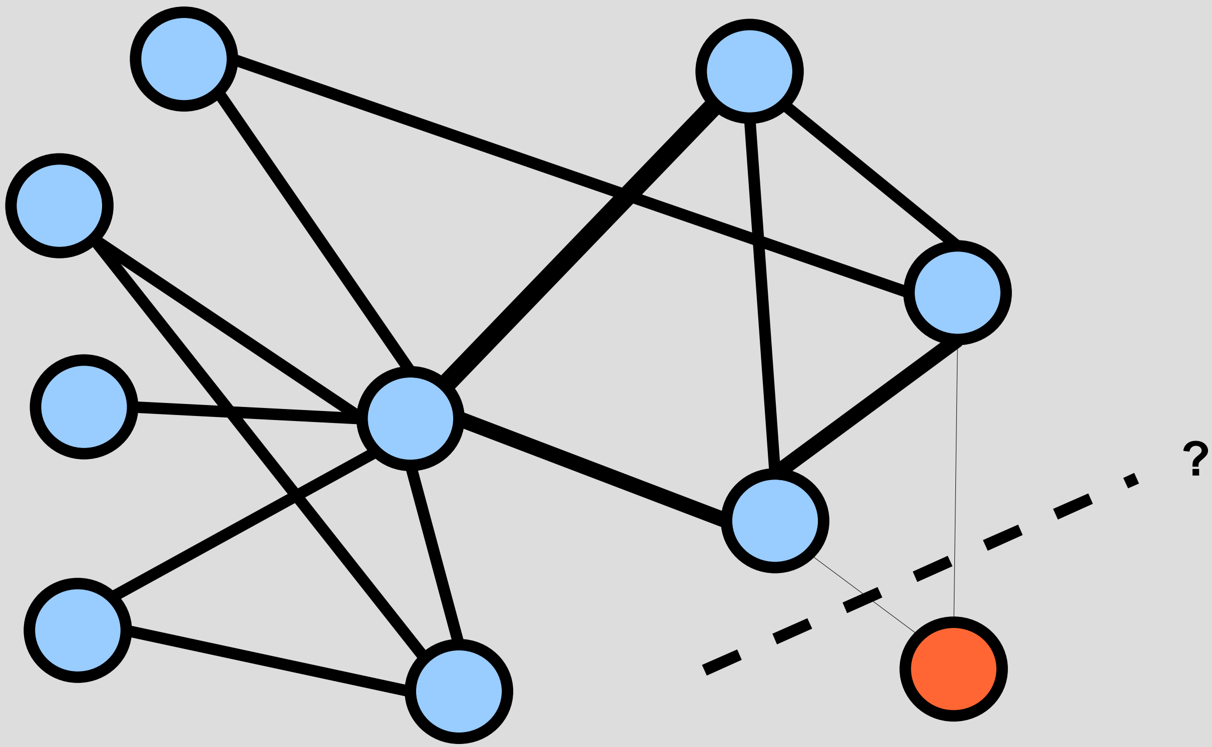
- Graph *G* with nodes $v \in V$, and edge weights $w_{ij} = w_{ji}$ for the edge between nodes *i* and *j*.
- A *cut C* is a set of edges that partitions a graph *G* into two subgraphs with nodes *A* and *A′* ($A \cup A' = V$).
- The *cost of a cut* is the total weight of its edges. So, we want to minimize:

$$min \sum_{i \in A,\, j \in A'} w_{ij} \quad \text{(minimum cut)}$$

# How to Cluster?



# How to Cluster?



# Normalized Cut Criterion

- Look for a low-cost cut. But try to achieve balanced partitions.
- The *degree* of a node is the total weight of its edges:

$$d_i = \sum_{j \in V} w_{ij}$$

- The *volume* of a graph is the sum of the degrees of its nodes:

$$vol(V) = \sum_{i \in V} d_i$$

- Normalized cut:

**minimize** $$NCut(A, A') = \left( \frac{1}{vol(A)} + \frac{1}{vol(A')} \right) \sum_{i \in A,\, j \in A'} w_{ij}$$

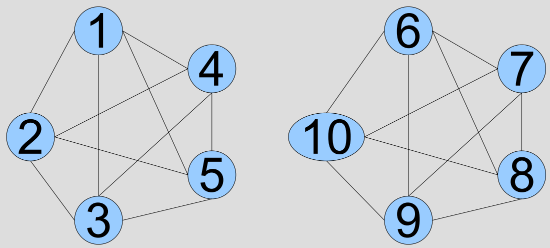
# Spectral Solution

- Computing *NCut* is NP-hard.
- Spectral clustering offers an approximate solution:
  - Similarity (adjacency) matrix of the graph, *S.*
  - Stochastic version of *S*: $P = D^{-1}S$
  - Solve the eigenvector problem: $Px = \lambda.x$
  - Use the second largest eigenvalue $\lambda_2$ and the corresponding eigenvector $x_2$. Cluster the graph based on the values in $x_2$.

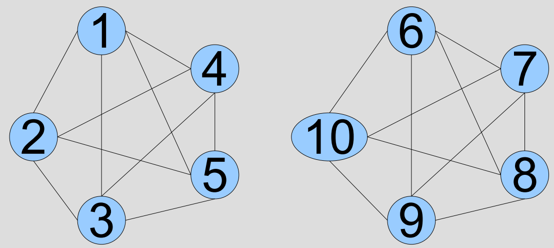
# Spectral Solution

- Computing *NCut* is NP-hard.
- Spectral clustering offers an approximate solution:
  - Similarity (adjacency) matrix of the graph, *S.*
  - Stochastic version of *S*: $P = D^{-1}S$
  - Solve the eigenvector problem: $Px = \lambda.x$
  - Use the second largest eigenvalue $\lambda_2$ and the corresponding eigenvector $x_2$. Cluster the graph based on the values in $x_2$.
  - Shi & Malik (2000) show that $NCut(A,A') = 1 - \lambda_2$

# An Example: Ideal Situation



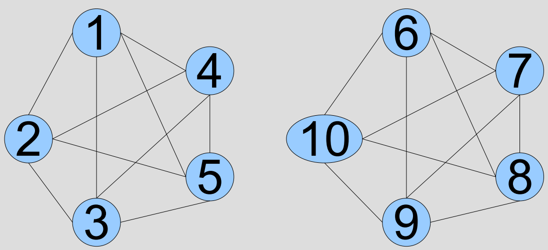
# An Example: Ideal Situation




**S:**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.50 | 0.50 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | 0.50 | 0.00 | 0.20 | 0.50 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **3** | 0.50 | 0.20 | 0.00 | 0.50 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | 0.25 | 0.50 | 0.50 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | 0.25 | 0.30 | 0.50 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **6** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.30 | 0.50 | 0.50 |
| **7** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 | 0.20 | 0.25 | 0.50 |
| **8** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.20 | 0.00 | 0.50 | 0.30 |
| **9** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.25 | 0.50 | 0.00 | 0.20 |
| **10** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.50 | 0.30 | 0.20 | 0.00 |

# An Example: Ideal Situation




S:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.50 | 0.50 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | 0.50 | 0.00 | 0.20 | 0.50 | 0.30 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **3** | 0.50 | 0.20 | 0.00 | 0.50 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | 0.25 | 0.50 | 0.50 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | 0.25 | 0.30 | 0.50 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **6** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.30 | 0.50 | 0.50 |
| **7** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 | 0.20 | 0.25 | 0.50 |
| **8** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.30 | 0.20 | 0.00 | 0.50 | 0.30 |
| **9** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.25 | 0.50 | 0.00 | 0.20 |
| **10** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.50 | 0.30 | 0.20 | 0.00 |

P:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.33 | 0.33 | 0.17 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | 0.33 | 0.00 | 0.13 | 0.33 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **3** | 0.29 | 0.12 | 0.00 | 0.29 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | 0.17 | 0.33 | 0.33 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | 0.19 | 0.23 | 0.38 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **6** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.28 | 0.17 | 0.28 | 0.28 |
| **7** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.00 | 0.14 | 0.17 | 0.34 |
| **8** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | 0.15 | 0.00 | 0.38 | 0.23 |
| **9** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.17 | 0.34 | 0.00 | 0.14 |
| **10** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 0.33 | 0.20 | 0.13 | 0.00 |

# An Example: Ideal Situation

First eigenvector
(same for all
stochastic matrices):

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.33 | 0.33 | 0.17 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | 0.33 | 0.00 | 0.13 | 0.33 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **3** | 0.29 | 0.12 | 0.00 | 0.29 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | 0.17 | 0.33 | 0.33 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | 0.19 | 0.23 | 0.38 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **6** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.28 | 0.17 | 0.28 | 0.28 |
| **7** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.00 | 0.14 | 0.17 | 0.34 |
| **8** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | 0.15 | 0.00 | 0.38 | 0.23 |
| **9** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.17 | 0.34 | 0.00 | 0.14 |
| **10** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 0.33 | 0.20 | 0.13 | 0.00 |

$$\times \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \mathbf{1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

# An Example: Ideal Situation

First eigenvector (same for all stochastic matrices):

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.33 | 0.33 | 0.17 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | 0.33 | 0.00 | 0.13 | 0.33 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **3** | 0.29 | 0.12 | 0.00 | 0.29 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | 0.17 | 0.33 | 0.33 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | 0.19 | 0.23 | 0.38 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **6** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.28 | 0.17 | 0.28 | 0.28 |
| **7** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.00 | 0.14 | 0.17 | 0.34 |
| **8** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | 0.15 | 0.00 | 0.38 | 0.23 |
| **9** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.17 | 0.34 | 0.00 | 0.14 |
| **10** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 0.33 | 0.20 | 0.13 | 0.00 |

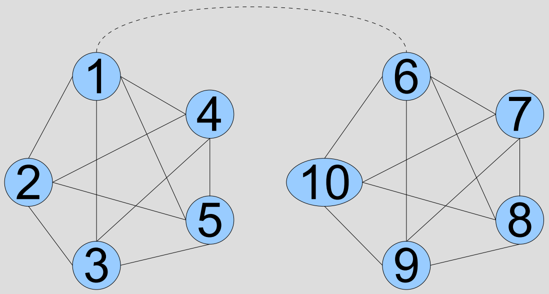
$$\text{X} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \mathbf{1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Second eigenvector:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.33 | 0.33 | 0.17 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | 0.33 | 0.00 | 0.13 | 0.33 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **3** | 0.29 | 0.12 | 0.00 | 0.29 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | 0.17 | 0.33 | 0.33 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | 0.19 | 0.23 | 0.38 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **6** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.28 | 0.17 | 0.28 | 0.28 |
| **7** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.00 | 0.14 | 0.17 | 0.34 |
| **8** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | 0.15 | 0.00 | 0.38 | 0.23 |
| **9** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.17 | 0.34 | 0.00 | 0.14 |
| **10** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 0.33 | 0.20 | 0.13 | 0.00 |

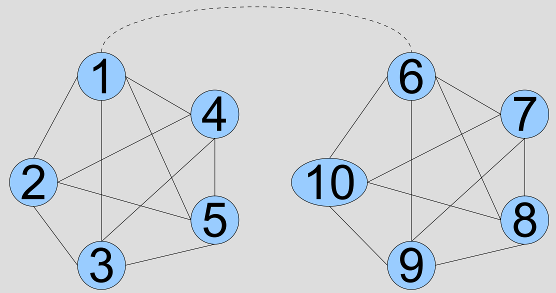
$$\text{X} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

# Slight perturbation




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.17 | 0.33 | 0.17 | 0.17 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | 0.33 | 0.00 | 0.13 | 0.33 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **3** | 0.29 | 0.12 | 0.00 | 0.29 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | 0.17 | 0.33 | 0.33 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | 0.19 | 0.23 | 0.38 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **6** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.28 | 0.17 | 0.28 | 0.28 |
| **7** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.00 | 0.14 | 0.17 | 0.34 |
| **8** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | 0.15 | 0.00 | 0.38 | 0.23 |
| **9** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.17 | 0.34 | 0.00 | 0.14 |
| **10** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 0.33 | 0.20 | 0.13 | 0.00 |

# Slight perturbation




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.00 | 0.17 | 0.33 | 0.17 | 0.17 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 |
| **2** | 0.33 | 0.00 | 0.13 | 0.33 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **3** | 0.29 | 0.12 | 0.00 | 0.29 | 0.29 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **4** | 0.17 | 0.33 | 0.33 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **5** | 0.19 | 0.23 | 0.38 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **6** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.28 | 0.17 | 0.28 | 0.28 |
| **7** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.00 | 0.14 | 0.17 | 0.34 |
| **8** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.23 | 0.15 | 0.00 | 0.38 | 0.23 |
| **9** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.17 | 0.34 | 0.00 | 0.14 |
| **10** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 0.33 | 0.20 | 0.13 | 0.00 |

$$\text{X} \begin{pmatrix} .86 \\ .98 \\ .99 \\ 1.0 \\ 1.0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \mathbf{.966} \begin{pmatrix} .86 \\ .98 \\ .99 \\ 1.0 \\ 1.0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

# Random Walk Interpretation

- Consider the *stationary distribution* of $P$: $P^T.\pi = \pi$
- Remember that $\pi_i$ is the probability that a random walk on the graph will end up at node $i$ "in the long run".
- For a similarity graph $G$ (i.e. symmetric matrix), it turns out that $\pi_i = d_i / vol(G)$. That is, it is proportional to the degrees.
- Given that we start in subgraph $A$, and the random walk start in its stationary distribution, what is the probability of jumping into a node in $A'$?

$$P_{AA'} = \frac{\sum_{i \in A,\, j \in A'} \pi_i P_{ij}}{vol(A)/vol(G)} = \frac{\sum_{i \in A,\, j \in A'} \frac{d_i}{vol(G)} \frac{S_{ij}}{d_i}}{vol(A)/vol(G)} = \sum_{i \in A,\, j \in A'} \frac{S_{ij}}{vol(A)}$$

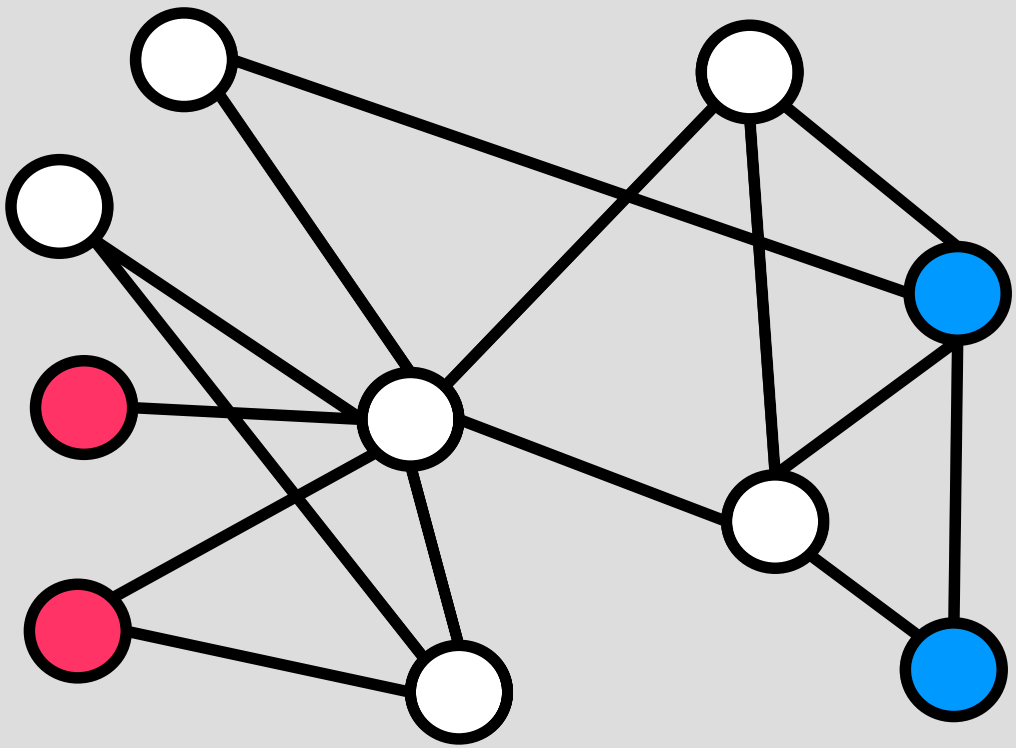
# Random Walk Interpretation

$$NCut(A, A') = P_{AA'} + P_{A'A}$$

- *NCut* is the total probability that the random walk will change partitions in the long run.
- That is, *NCut* partitions the graph into two parts such that the random walk, once in one of the parts, tends to remain in it.
- Remember the value of the second eigenvalue?

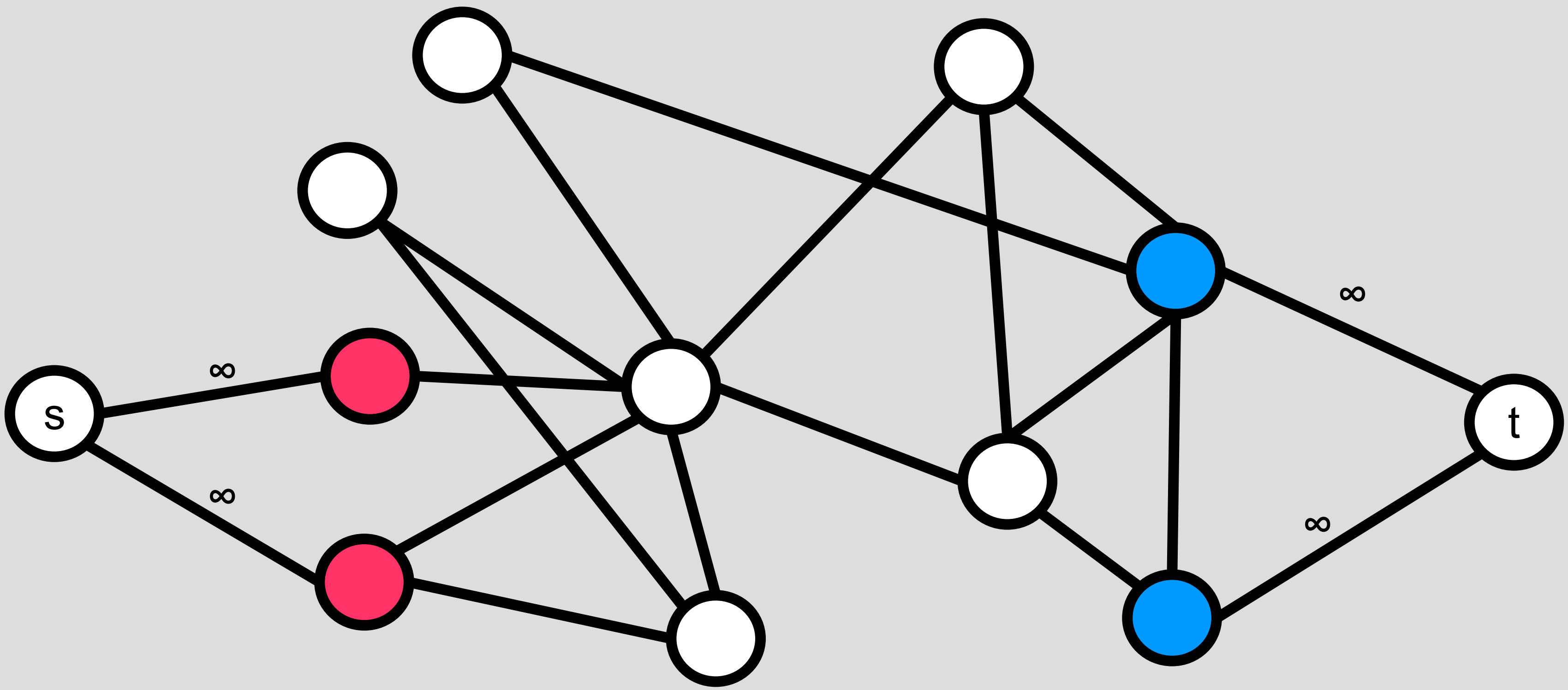
# Classification

- How about when we *know* the label of a few nodes:



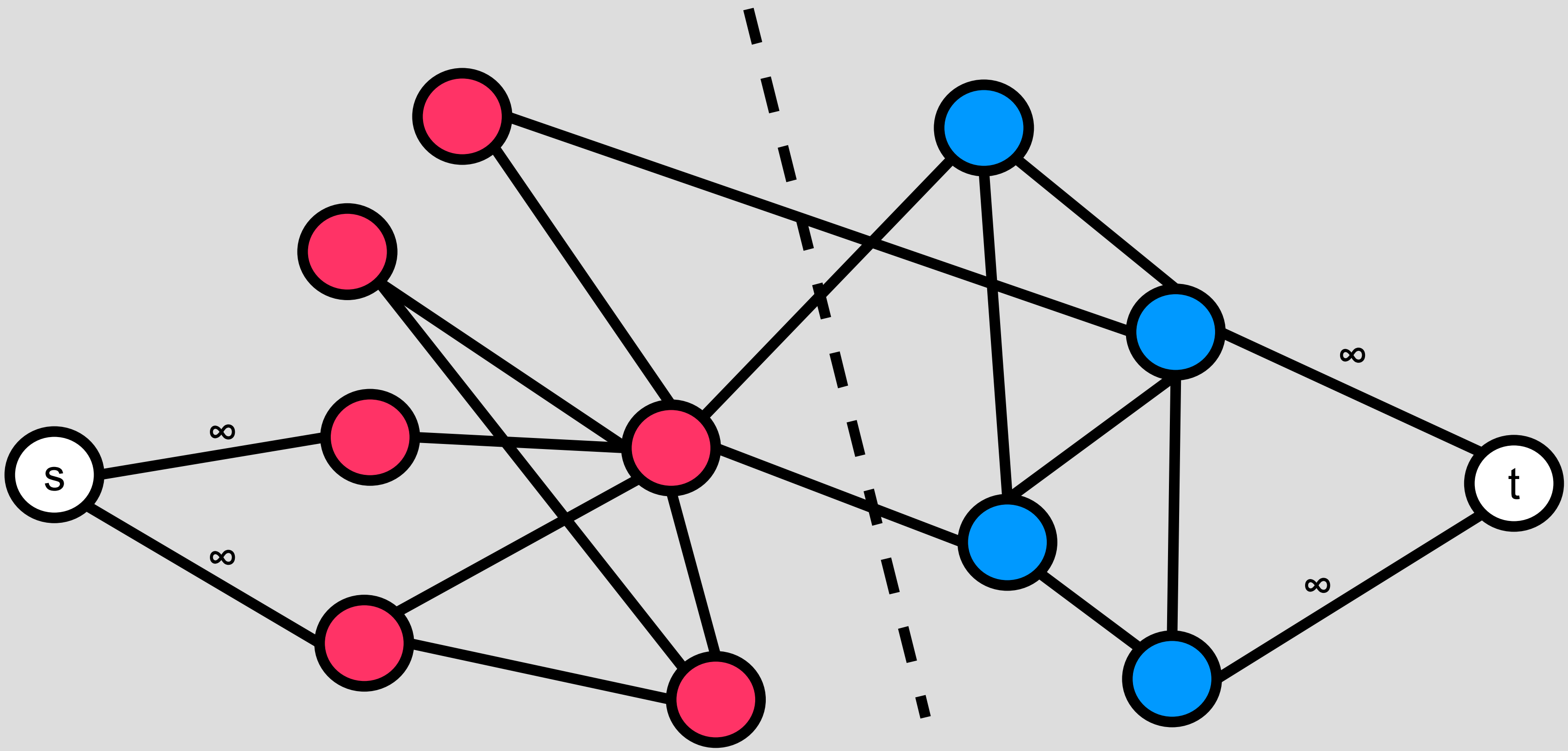
# Classification

- s-t nodes trick:



# Classification

- s-t nodes trick:



# Min-cut Classification

- Menger theorem: *The maximum amount of flow (from “s” to “t”) in a graph is equal to the capacity of a minimal cut.*
- Max-flow algorithm, which is polynomial, can be used to find the min-cut.

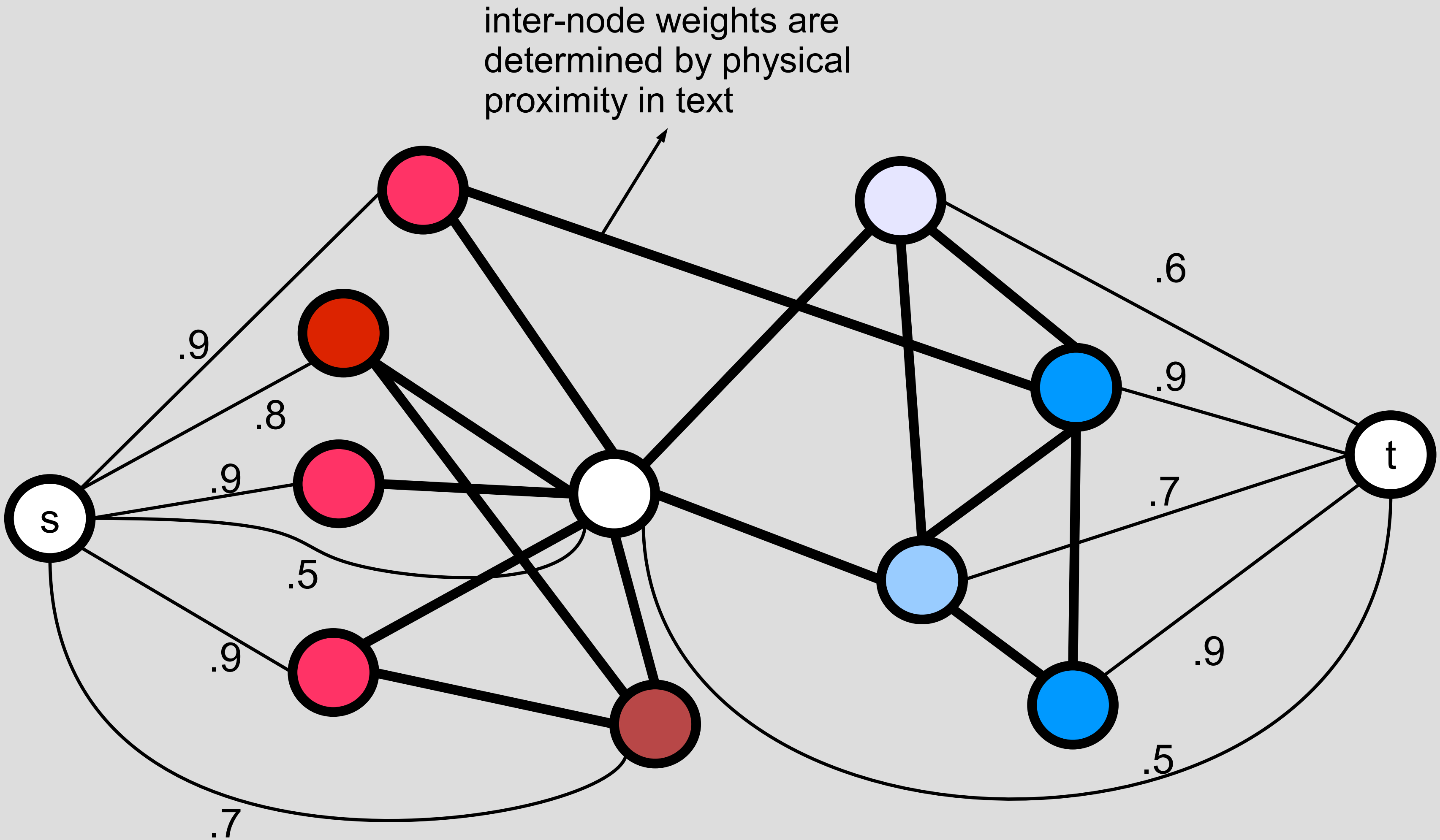
# An Interesting Application: Sentiment Analysis

- Classify movie reviews as *positive* ("thumbs up") or *negative.*

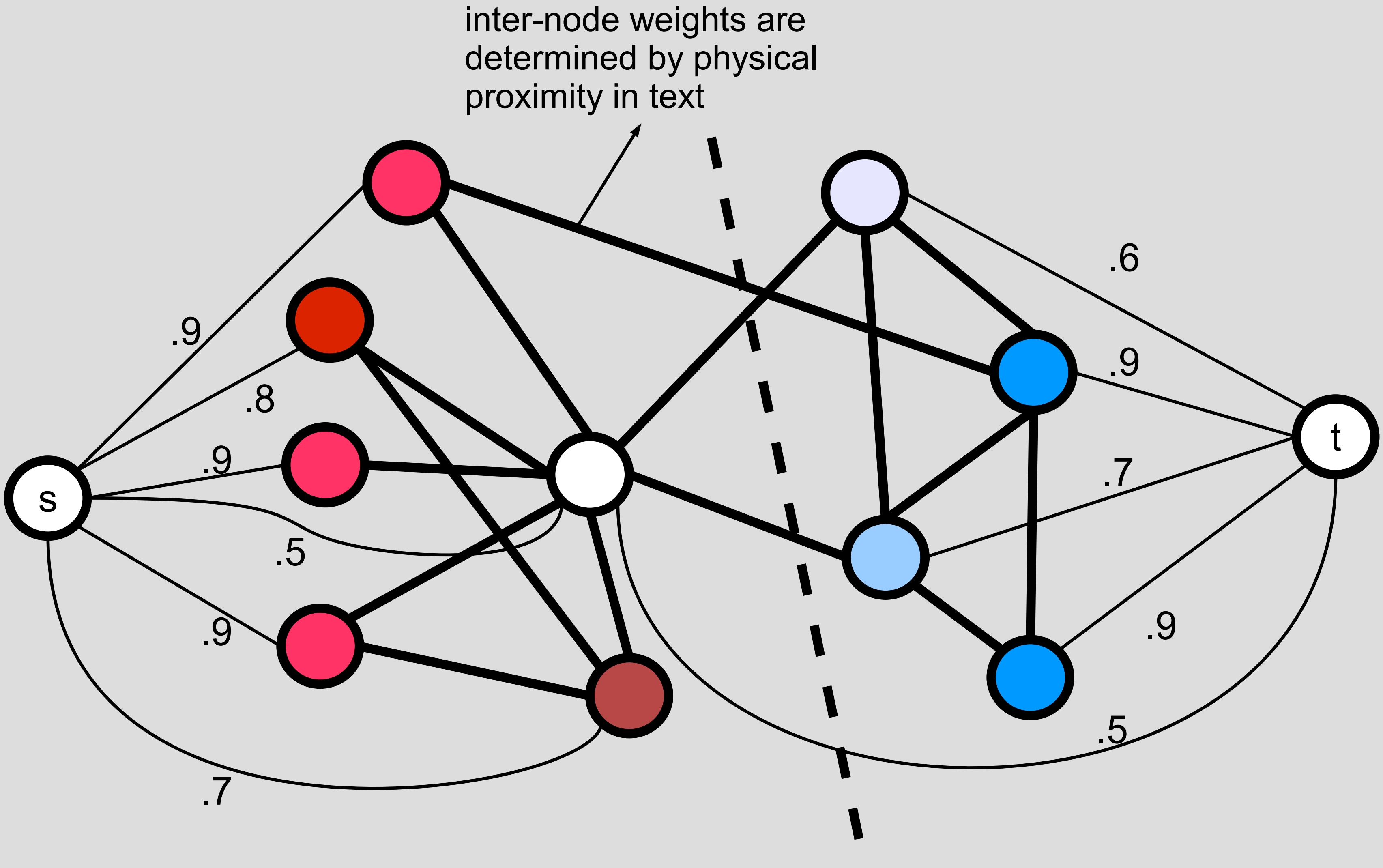
- Suppose we have another learning algorithm that gives us $f(t)$, the probability that sentence $t$ is positive. (or $1\text{-}f(t)$ for being negative).

- We also have a sentence similarity function based on proximity in text.

# An Interesting Application: Sentiment Analysis

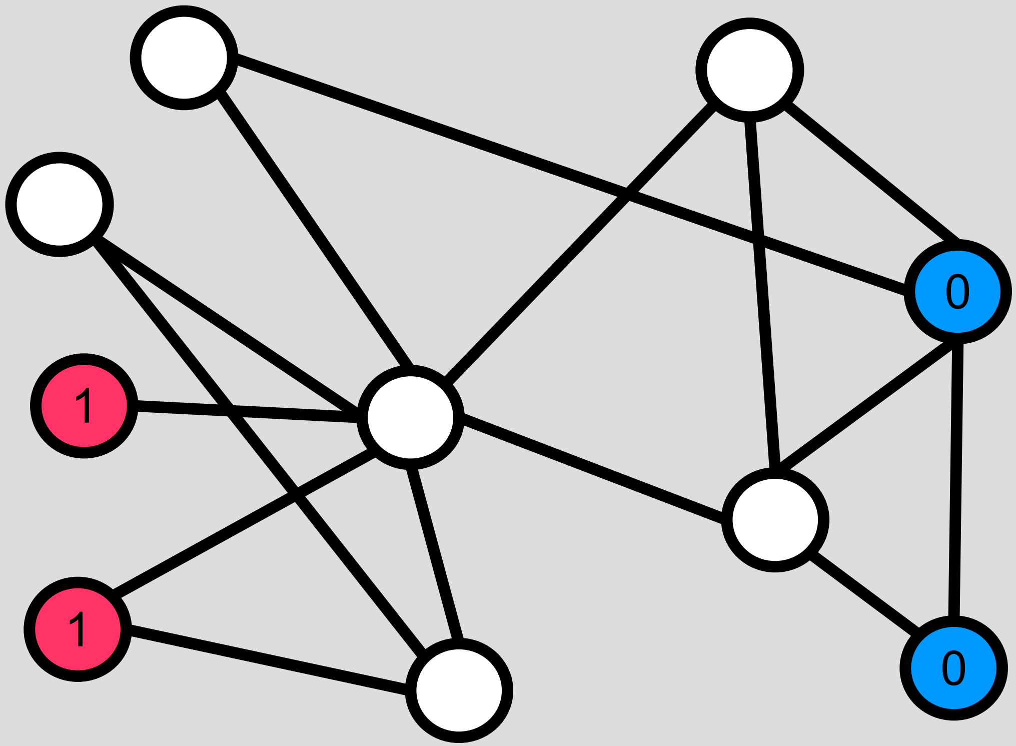


# An Interesting Application: Sentiment Analysis



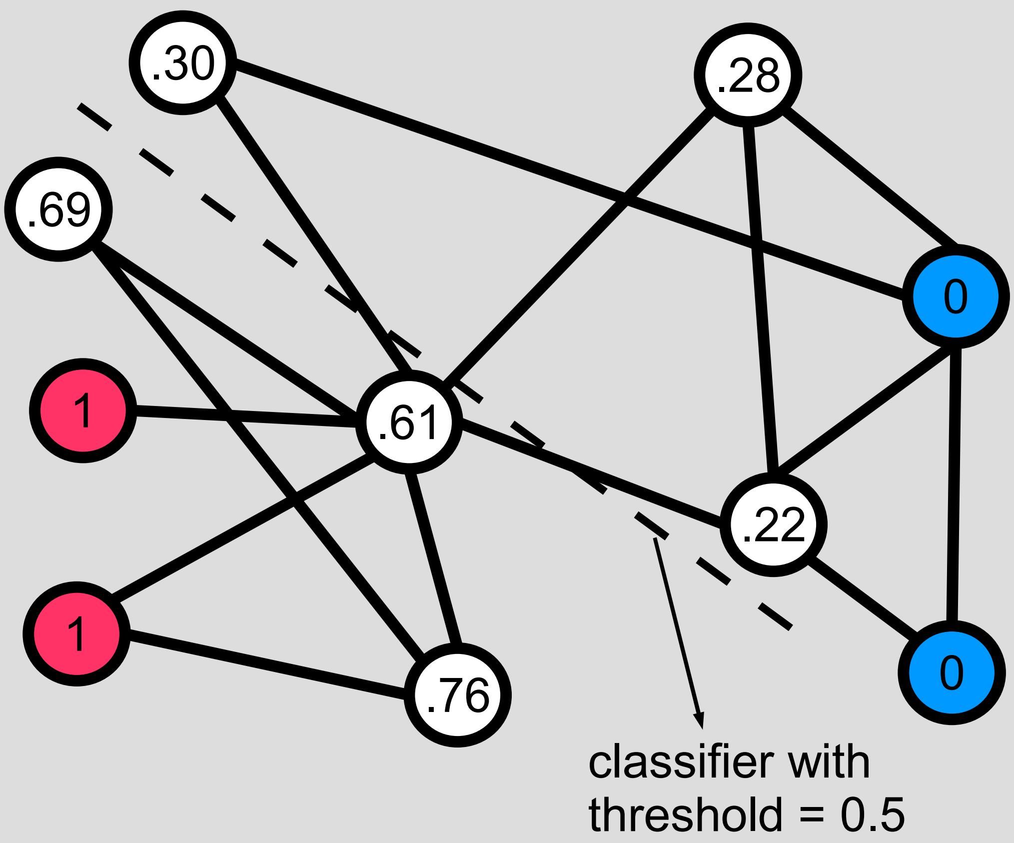
# Harmonic Functions for Classification

- Recall that a harmonic function *averages* values over a graph. This could be used as a classification method:



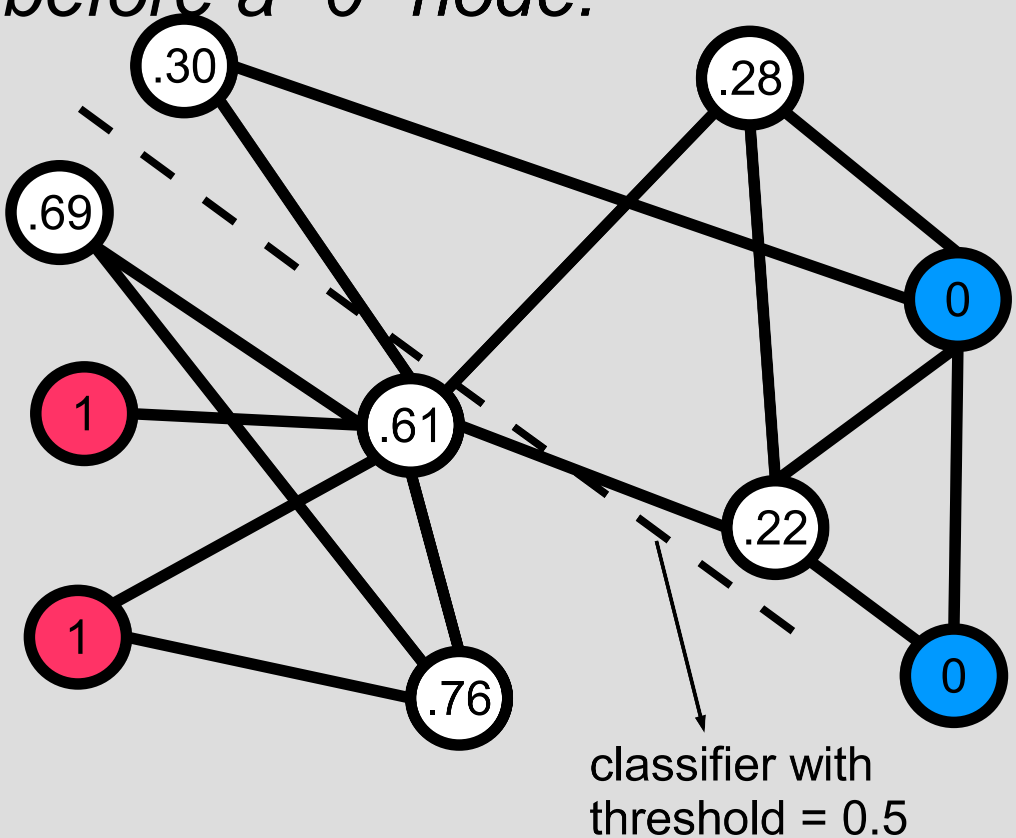
# Harmonic Functions for Classification

- Recall that a harmonic function *averages* values over a graph. This could be used as a classification method:



# Harmonic Functions: Random Walk Interpretation

- The harmonic function value *f(i)* for node *i* is the *probability that a random walk starting at i will hit a “1” node before a “0” node.*



# Energy Minimization Connection

- Harmonic functions minimize:

$$\sum_{i,\,j\in G} w_{ij}(f(i)-f(j))^2 \quad , f\in[0,1]$$

- Min-cuts minimize:

$$\sum_{i,\,j\in G} w_{ij}|f(i)-f(j)| \quad , f\in\{0,1\}$$

- A “cut” is a *discrete* decision/function while the harmonic function tries to achieve the same by fitting a *continuous* function over the graph.