COMPSCI 753

Algorithms for Massive Data

Semester 2, 2020

Tutorial - Graph

Kaiqi Zhao

1 Spam farm

Assume that two spam farmers, each having $p = 1$ support page and 1 target page, agree to link their spam farms. Assume that each target page has page rank $x + \frac{1-\beta}{n}$ without any spam farm. Compute the page rank of the target page when

1. there is a bidirectional link between each target page and its ow

2. there is a bidirectional link between each target page and each support page.

Which case produc

Solution: Denote the rank of each support page as z and the rank of each target page as y. The first case is the same as in our lecture, because each spam farm works on their own:

$$y = x + \beta^2 y + \frac{(\beta p + 1)(1-\beta)}{n},$$

where the number of support page $p = 1$. Then we get $y = \frac{x}{1-\beta^2} + \frac{1}{n}$.

In the second case, for each support page, we have incoming edges from the two target pages, each contributes $\frac{y}{2}$, thus we have:

$$z = \beta y + \frac{(1-\beta)}{n}. \tag{1}$$

Similarly, for each target page, we have incoming edges from the two support pages, each contributes $\frac{z}{2}$, thus we have:

$$y = x + \beta z + \frac{(1-\beta)}{n}. \tag{2}$$

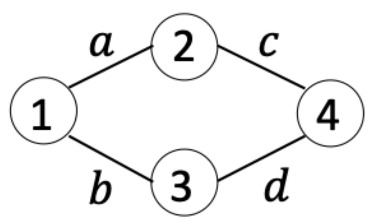
By substituting (1) to (2), w

$$\begin{aligned} y &= x + \beta z + \frac{(1-\beta)}{n} \\ &= x + \beta[\beta y + \frac{(1-\beta)}{n}] + \frac{(1-\beta)}{n} \\ &= x + \beta^2 y + \frac{\beta(1-\beta)}{n} + \frac{(1-\beta)}{n} \end{aligned} \tag{3}$$

Thus the pagerank of a target page y is: $y = \frac{x}{1-\beta^2} + \frac{1}{n}$ get page is exactly the same for

2 Edge betweenness

Compute the edge betweenness of the edges in the followin



Solution:

Use Brandes' algorithm. Notice that using any of the four nodes as root will result in the same hierarchy.

- Using 1 as root, $EB(a) = EB(b) = 1.5$ and $EB(c) = EB(d) = 0.5$
- Using 2 as root, $EB(a) = EB(c) = 1.5$ and $EB(b) = EB(c) = 0.5$
- Using 3 as root, $EB(b) = EB(d) = 1.5$ and $EB(a) = EB(c) = 0.5$

- Using 4 as root, $EB(c) = EB(d) = 1.5$ and $EB(a) = EB(b) = 0.5$

To sum up and divided by 2 (each path was considered twice $u \to ... \to v$ and $v \to ... \to u$), $EB(a) = EB(b) = EB(c) = EB(d) = 2$.

3 Spectral clustering

Given the following adjacency matrix of a graph:

$$\begin{bmatrix} [illegible] \\ 0 & 0 & 0.5 & 1 \end{bmatrix}$$

1. Compute the eigenvalues of the Laplacian. You may need to compute the determinant (refer to `https://en.wikipedia.org/wiki/`
2. What is the best [illegible] cut for bipartitioning the graph? Give the t[illegible] communities.

Solution: First [illegible] en Laplacian matrix $\mathbf{L}$:

$$\mathbf{D} = \begin{bmatrix} 1.8 & 0 & 0 & 0 \\ 0 & 1.8 & 0 & 0 \\ 0 & 0 & 1.5 & 0 \\ 0 & 0 & 0 & 1.5 \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} [illegible] \\ [illegible] \\ [illegible] & 0 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \end{bmatrix}$$

Apply the eigen equation $|\lambda \mathbf{I} - \mathbf{L}| = 0$:

$$
\begin{aligned}
|\lambda \mathbf{I} - \mathbf{L}| = &\begin{vmatrix} \lambda - 0.8 & 0.8 & 0 & 0 \\ 0.8 & \lambda - 0.8 & 0 & 0 \\ 0 & 0 & \lambda - 0.5 & 0.5 \\ 0 & 0 & 0.5 & \lambda - 0.5 \end{vmatrix} \\
=&(\lambda - 0.8) \begin{vmatrix} \lambda - 0.8 & 0 & 0 \\ 0 & \lambda - 0.5 & 0.5 \\ 0 & 0.5 & \lambda - 0.5 \end{vmatrix} - 0.8 \begin{vmatrix} 0.8 & 0 & 0 \\ 0 & \lambda - 0.5 & 0.5 \\ 0 & 0.5 & \lambda - 0.5 \end{vmatrix} \\
=&(\lambda - 0.8)^2((\lambda - 0.5)^2 - 0.25) - 0.8^2((\lambda - 0.5)^2 - 0.25) \\
=&(\lambda^2 - 1.6\lambda)(\lambda^2 - \lambda) \\
=&\lambda^2(\lambda - 1.6)(\lambda - 1) = 0
\end{aligned}
$$

So, we have the four roots in ascending order as $\lambda_1 = \lambda_2 = 0, \lambda_3 = 1, \lambda_4 = 1.6$.
To get the best RatioCut, we need to use $\lambda_2 = 0$. Let $\mathbf{v} = [v_1, v_2, v_3, v_4]^T$ be the corresponding eigenvector: $\mathbf{Lv} = \lambda_2 \mathbf{v} = 0$. Then we have:

$$\begin{cases} 0.8v_1 - 0.8v_2 = 0 \\ -0.8v_1 + 0.8v_2 = 0 \\ 0.5v_3 - 0.5v_4 = 0 \\ 0.5v_3 + 0.5v_4 = 0 \end{cases}$$

According to above, the eig v_4. So, node 1 and 2 should have the same com

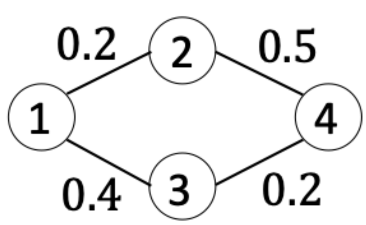
Because we only care about the sign and without loss of generality, we assume $v_1 = v_2 > 0$. Then, we have two cases:

$$\begin{cases} v_1 = v_2 > 0 \text{ and } v \\ v_1 = v_2 > 0 \text{ and } v \end{cases}$$

For the first case, we a at we want as ment es. For the second case, we assig odes to the other. This is the best RatioCut we can do, with cost of zer original graph has two disconnected components with th

4 Influence Spread

In the lectures, we discussed the influence spread on directed graph. But it is very straightforward to be applied on undirected graphs. In the undirected graph below, the number on each edge denotes the probability the edge is live. Let the seed set $S = \{1\}$, consider calculating the influence spread under the independent cascade model using the expectation over the deterministic graphs.



Let $X = \{X_{12}, X_{13}, X_{24}, X_{34}\}$ be the binary label (live=1, blocked=0) for the edges. Fill in the following contingency table with corresponding values. Give the influence spread of the seed set S.

Solution:

$X_{12}, X_{13}, X_{24}, X_{34}$	prob[X]	#nodes reachable from S, $\sigma^X(S)$
0, 0, 0, 0	0.8*0.6*0.5*0.8 = 0.192	1
0, 0, 0, 1	0.8*0.6*0.5*0.2 = 0.048	1
0, 0, 1, 0	0.8*0.6*0.5*0.8 = 0.192	1
0, 0, 1, 1	0.8*0.6*0.5*0.2 = 0.048	1
0, 1, 0, 0		2
0, 1, 0, 1		3
0, 1, 1, 0		2
0, 1, 1, 1	0.8*0.4*0.5*0.2 = 0.032	4
1, 0, 0, 0	0.2*0.6*0.5*0.8 = 0.048	2
1, 0, 0, 1	0.2*0.6*0.5*0.2 = 0.012	2
1, 0, 1, 0	0.2*0.6*0.5*0.8 = 0.048	3
1, 0, 1, 1	0.2*0.6*0.5*0.2 = 0.012	4
1, 1, 0, 0	0.2*0.4*0.5*0.8 = 0.032	3
1, 1, 0, 1	0.2*0.4*0.5*0.2 = 0.008	4
1, 1, 1, 0	0.2*0.4*0.5*0.8 = 0.032	4
1,		4

$\sigma(S) = 0.48 + (0$

5 Submodularity

Given that $f(S)$ and $g(S)$ are two non-negative submodular functions:

1. Let α and β be any non-negative real numbers, is $\sigma_2(S) = \alpha f(S) + \beta g(S)$ submodular? Show your proof if yes, otherwise, give a counter example.

2. Let $A \subset B$ and $B \backslash A$ be the set of elements in B but not in A. Show that $f(B) - f(A) \leq \sum_{v \in B \backslash A} f(v)$. (We use notation $f(v)$ to denote $f(\{v\})$ for convenience.)

Solution:

Let S and T are two sets with $S \subset T$. And v be any node that $v \notin S$ and $v \notin T$.

1. Compute the difference in marginal gains:

$$
\begin{aligned}
&[\sigma_2(S \cup \{v\}) - \sigma_2(S)] - [\sigma_2(T \cup \{v\}) - \sigma_2(T)] \\
=&\alpha[f(S \cup \{v\}) - f(S) - f(T \cup \{v\}) + f(T)] \\
&+ \beta[g(S \cup \{v\}) - g(S) - g(T \cup \{v\}) + g(T)]
\end{aligned}
$$

Note that, because $f(S)$ is submodular, $f(S \cup \{v\}) - f(S) - f(T \cup \{v\}) + f(T) \geq 0$, for the same reason, $g(S \cup \{v\}) - g(S) - g(T \cup \{v\}) + g(T) \geq 0$. Given that $\alpha \geq 0$ and $\beta \geq 0$, then $[\sigma_2(S \cup \{v\}) - \sigma_2(S)] - [\sigma_2(T \cup \{v\}) - \sigma_2(T)] \geq 0$. Therefore, $\sigma_2(S)$ is submodular.

2. We first show that $f(S \cup \{v\}) - f(S) \leq f(v)$ for a set S and any $v \notin S$. By submodularity of f, we have:

$$f(S \cup \{v\}) - f(S) \leq f(\emptyset \cup \{v\}) - f(\emptyset) = f(v) - f(\emptyset) \leq f(v). \tag{4}$$

The last inequity u

Let $B \setminus A = \{v_1, v_2, ..., v_k\}$. Then:

$$f(B) - f(A) = f(A \cup \{v_1, v_2, ..., v_k\}) - f(A)$$

$$= \underbrace{f(A \cup \{v_1, v_2, ..., v_k\}) - f(A \cup}_{\text{Consider } S = A \cup \{v_2, ..., v_k\}, \text{ ap}} \quad v_2, ..., v_k\}) - f(A)$$

$$\underbrace{\qquad\qquad}_{\text{Consider } S = A} \; \underbrace{\qquad\qquad}_{v_3, ..., v_k \text{ , apply (4)}} {}_k\}) + f(A \cup \{v_3, ..., v_k\}) - f(A)$$

$$\leq f(v_1) + f(v_2) + f(A \cup \{v_3$$

$$\leq f(v_1) + f(v_2) + ... + f(v_k$$

$$v \in B \setminus A$$