Spectral clustering and transductive inference for graph data

Dengyong (Denny) Zhou

NEC Laboratories America, Inc.
4 Independence Way, Suite 200
Princeton, NJ 08540, USA

Joint work with Thomas Hofmann, Jiayuan Huang, Bernhard Schölkopf

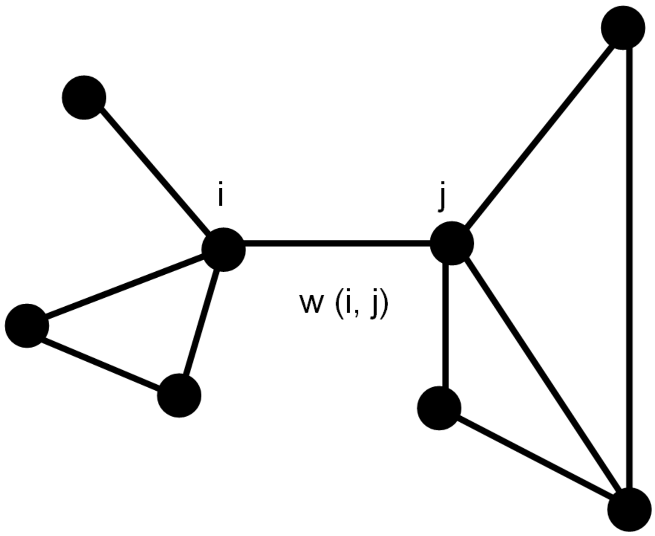
Problem setting of clustering

- Given a set of discrete objects $\mathcal{X} = \{x_1, x_2, \cdots, x_n\}$, our goal is to cluster the object set into two or more clusters in a reasonable way.
- In general, we assume there exist pairwise relationships among objects to be clustered: $w : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}^+$. Thus the object set can be regarded as a graph. In particular, when w is symmetric, i.e. $w(i, j) = w(j, i)$, the corresponding graph is undirected.

What does it mean by "reasonable"?

An undirected graph

Question—How to reasonably cut this graph into two parts?



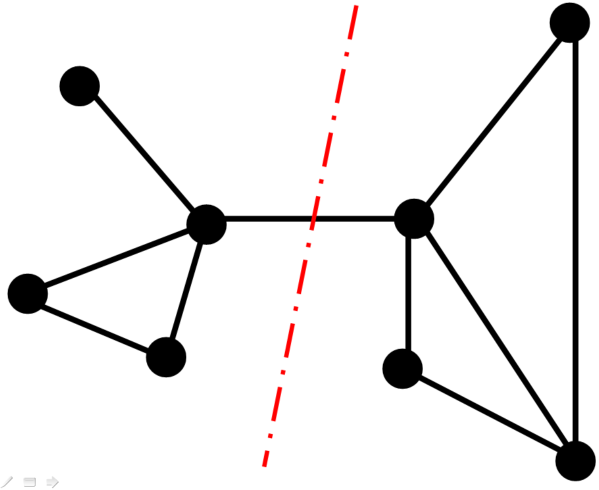
Graph min-cut: formulism

Cluster the object set into the two parts $\mathcal{X} = S \cup S^c$ such that

$$\min_{\emptyset \neq S \subset \mathcal{X}} \sum_{i \in S, j \in S^c} w(i, j)$$

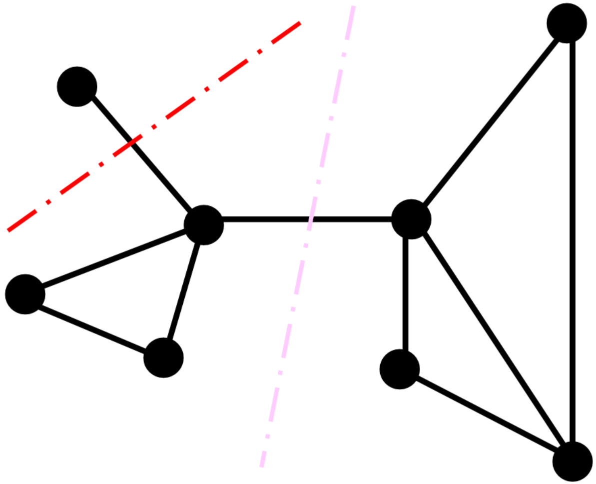
Graph min-cut: toy example

A solution satisfying the min-cut criterion:



Graph min-cut: toy example

The other solution satisfying the min-cut criterion:



Weakness of min-cut: may produce extremely unbalanced clustering!

Normalized cut: basic intuition

Seeking for a clustering such that

1. The connection between two clusters is as weak as possible;
2. The connection among the objects in the same cluster is as strong as possible.

What does it mean by "connection"?
How to measure the strength of connection?

Normalized cut: formalism

(Shi and Malik, 1997; Meilă and Shi, 2001)

- Connection between two clusters:
  $Connection(S, S^c) = \sum_{i \in S, j \in S^c} w(i, j).$
  Note that
  $Connection(S, S^c) = Connection(S^c, S).$

- Connection in a cluster:
  $Connection(S) = \sum_{i \in S} d(i),$
  where
  $d(i) = \sum_{j \sim i} w(i, j).$

Normalized cut: formalism

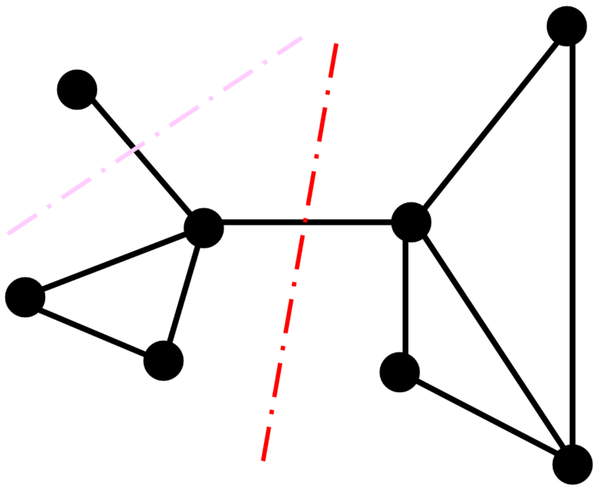
(Shi and Malik, 1997)

Partitioning the object set into two parts by

$$\underset{\emptyset \neq S \subset \mathcal{X}}{\operatorname{argmin}} \sum_{i \in S, j \in S^c} w(i,j) \left(\frac{1}{\sum_{i \in S} d(i)} + \frac{1}{\sum_{i \in S^c} d(i)} \right)$$

Normalized cut: toy example

The unique solution from the normalized cut:



Normalized cut: algorithm

The combinatorial problem is NP-complete. It can be relaxed into the real-valued optimization problem (Shi and Malik, 1997)

$$\operatorname*{argmin}_{f \in \mathbb{R}^n} \sum_{i,j} w(i,j) \left(\frac{f(i)}{\sqrt{d(i)}} - \frac{f(j)}{\sqrt{d(j)}} \right)^2$$

subject to $f^T f = 1, \; f\sqrt{d} = 0.$

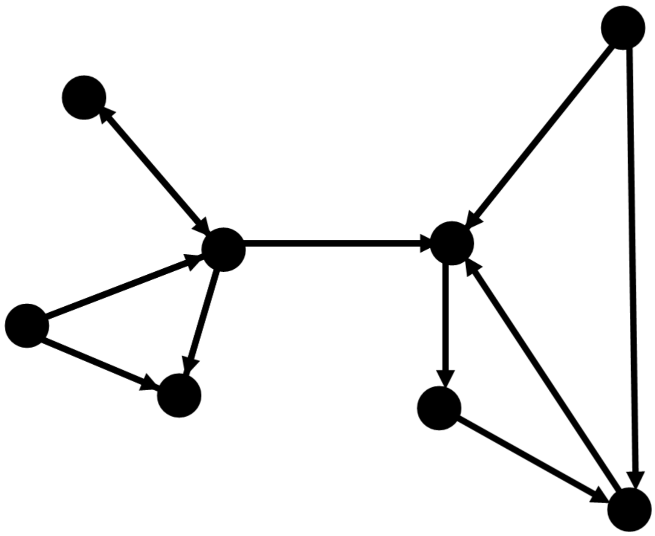
Then $S = \{x_i \in \mathcal{X} | f(i) \leq 0\}$ and $S^c = \{x_i \in \mathcal{X} | f(i) > 0\}$.

Normalized cut: algorithm

Proposition. *The solution of the real-valued optimization problem is an eigenvector of the matrix $D^{-1/2}WD^{-1/2}$ corresponding to the second largest eigenvalue.*

How to partition a directed graph?

A directed version of the previous toy example:



How to partition a directed graph?

A naive generalization:

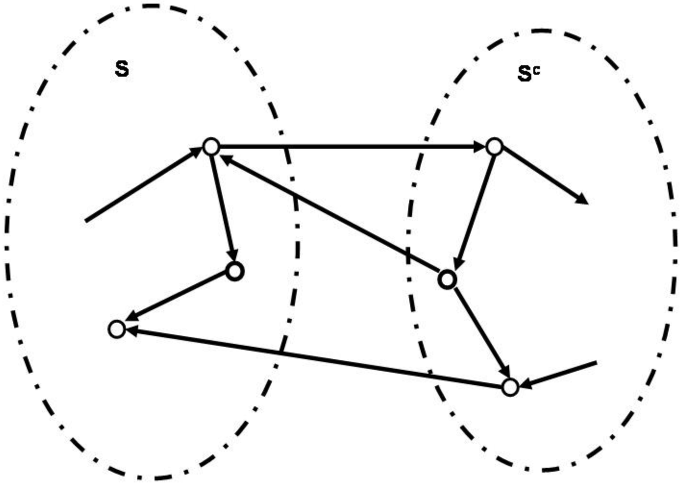
1. Connection between two clusters:
   $Connection(S, S^c) = \sum_{i \in S, j \in S^c} w(i, j)$

2. Connection in a cluster:
   $Connection(S) = \sum_{i \in S} d^-(i)$(in-degree)

However, . . .

How to partition a directed graph?

In general, $Connection(S, S^c) \neq Connection(S, S^c)$!



How to partition a directed graph?

Another naive generalization:

1. Connection between two clusters:
   $Connection(S, S^c) = \sum_{i \in S, j \in S^c} w(i,j) + \sum_{i \in S^c, j \in S} w(i,j)$

2. Connection in a cluster:
   $Connection(S) = \sum_{i \in S} d^-(i) + \sum_{i \in S} d^+(i)$(in-degree + out-degree)

Now we have $Connection(S, S^c) = Connection(S, S^c)$.
However, directionality is ignored! So, . . .

Algorithmic challenges in web search engines

How to generalize the normalized cut based approach to directed graphs was listed as one of six algorithmic challenges in web search engines by the Google labs director (Henzinger, 2003).

Our solution: intuition

We have to face the following two problems as in the case of undirected graphs:

1. How to measure the connection between clusters?
2. How to measure the connection among objects in the same cluster?

Our solution: formulism

(Zhou, Huang and Schölkopf, ICML 05)

Define a random walk over the graph with the transition probabilities $p(i,j)$ and the stationary distribution $\pi(i)$. Then

1. Connection between two clusters:
   $Connection(S, S^c) = \sum_{i \in S, j \in S^c} \pi(i) p(i,j)$

2. Connection in a cluster:
   $Connection(S) = \sum_{i \in S} \pi(i)$

Regarding a directed graph as a Markov chain!

Our solution: formulism

key fact

Proposition. $Connection(S, S^c) = Connection(S^c, S).$

[It follows from the property of stationary distribution, and holds for general Markov chains. The quantity $\pi(i)p(i,j)$ is generally referred to as ergodic flow.]

Recalling the parallel but "obvious" fact in the case of undirected graphs.

Our solution: formulism

Partitioning a directed graph into two parts by

$$\operatorname*{argmin}_{\emptyset \neq S \subset \mathcal{X}} \sum_{i \in S, j \in S^c} \pi(i) p(i,j) \left(\frac{1}{\sum_{i \in S} \pi(i)} + \frac{1}{\sum_{i \in S^c} \pi(i)} \right)$$

An elegant probabilistic explanation: The cut criterion is equivalent to

$$\operatorname*{argmin}_{\emptyset \neq S \subset \mathcal{X}} p(S \to S^c | S) + p(S^c \to S | S^c)$$

Our solution: formulism

In the case of undirected graphs, define a natural random walk with $p(i,j) = w(i,j)/d(i)$. It can be shown that the random walk has a stationary distribution $\pi(i) = d(i)/\sum_j d(j)$. Therefore

$$\sum_{i \in S, j \in S^c} \pi(i) p(i,j) \left(\frac{1}{\sum_{i \in S} \pi(i)} + \frac{1}{\sum_{i \in S^c} \pi(i)} \right)$$

$$= \sum_{i \in S, j \in S^c} w(i,j) \left(\frac{1}{\sum_{i \in S} d(i)} + \frac{1}{\sum_{i \in S^c} d(i)} \right)$$

Recovered the normalized cut for undirected graphs!

Our solution: algorithm

The combinatorial problem can be relaxed into a real-valued optimization problem as in the case of undirected graphs:

$$\operatorname*{argmin}_{f \in \mathbb{R}^n} \sum_{i \to j} \pi(i) p(i,j) \left(\frac{f(i)}{\sqrt{\pi(i)}} - \frac{f(j)}{\sqrt{\pi(j)}} \right)^2$$

subject to $f^T f = 1, \; f \sqrt{\pi} = 0.$

Then $S = \{x_i \in \mathcal{X} | f(i) \leq 0\}$ and $S^c = \{x_i \in \mathcal{X} | f(i) > 0\}$.

Our solution: algorithm

Proposition. *The solution of the real-valued optimization problem is an eigenvector of the matrix* $\Pi^{1/2}P\Pi^{-1/2}+\Pi^{-1/2}P^T\Pi^{1/2}$ *corresponding to the second largest eigenvalue.*

How to define a random walk on a directed graph?

A random walk used by the Google search engine:

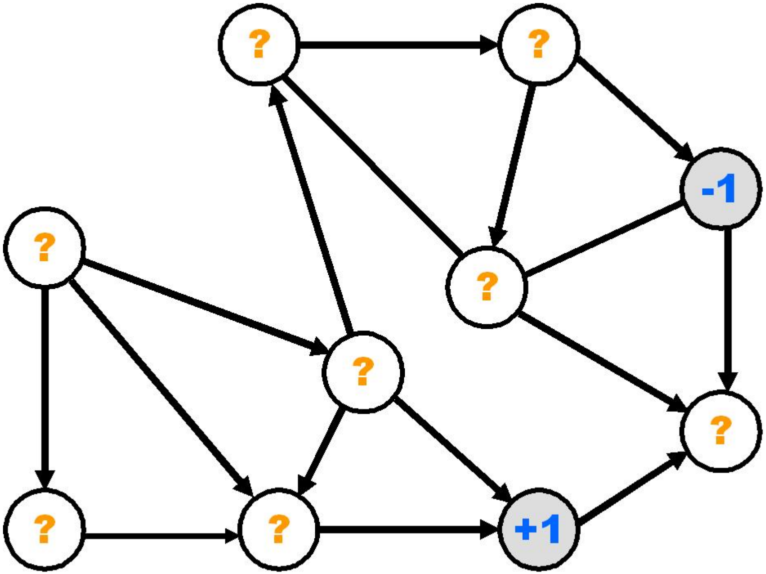
With probability α, follow a link; and, with probability $1 - \alpha$, jump to a randomly chosen vertex.

Nice property:
The random walk can be arbitrarily close to the natural random walk while having a unique positive stationary distribution.

From clustering to classification

If you have some labeled examples, how to utilize them?

Transductive inference



Transductive inference

It is straightforward from spectral clustering to transductive inference:

Given a directed graph $G = (V, E)$, some vertices are labeled. Define a function y on V with $y(v) = 1$ or -1 if vertex v is labeled as 1 or -1, and 0 if v is unlabeled. Then the remaining unlabeled vertices may be classified by using the function

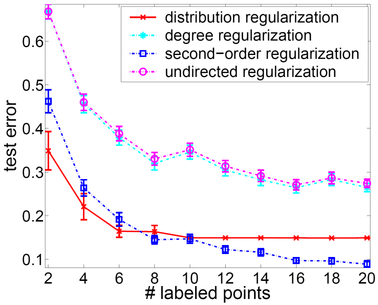
$$\underset{f \in \mathbb{R}^{|V|}}{\operatorname{argmin}} \sum_{i \to j} \pi(i) p(i, j) \left(\frac{f(i)}{\sqrt{\pi(i)}} - \frac{f(j)}{\sqrt{\pi(j)}} \right)^2 + \mu \|f - y\|^2$$

Transductive inference

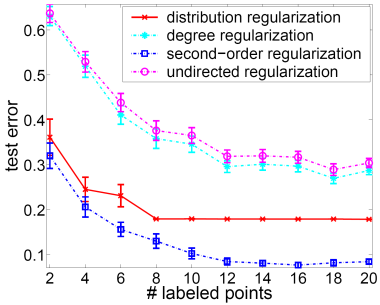
In the case of undirected graphs, the framework reduces to our earlier approach `(Zhou et al, NIPS 03)`:

$$\operatorname*{argmin}_{f \in \mathbb{R}^{|V|}} \sum_{i \sim j} w(i,j) \left(\frac{f(i)}{\sqrt{d(i)}} - \frac{f(j)}{\sqrt{d(j)}} \right)^2 + \mu \|f - y\|^2$$

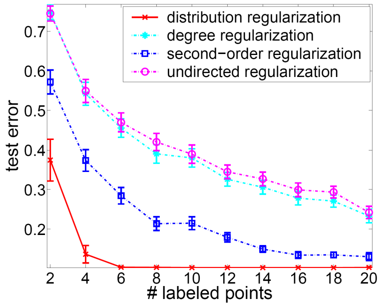
Do you still remember the *two moon* problem in our paper?



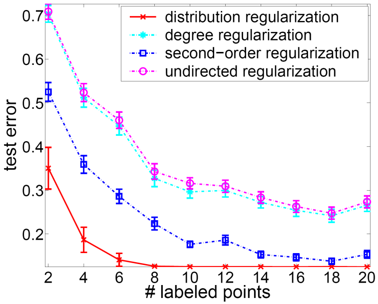
(a) Cornell (student)



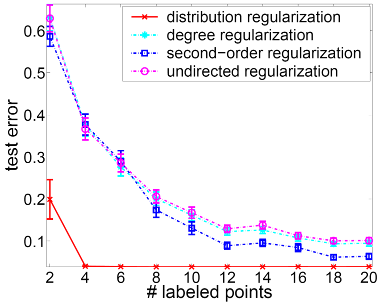
(b) Texas (student)



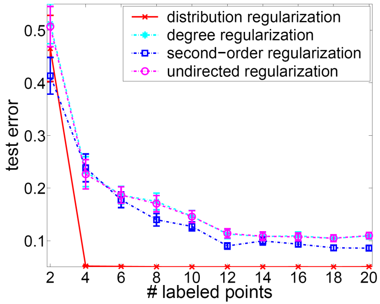
(c) Washington (student)



(d) Wisconsin (student)



(e) Cornell (faculty)



(f) Cornell (course)

Directionality does contain valuable information!

Conclusion

A solid mathematical framework for the web IR

- It is the first time to generalize the spectral clustering approach to directed graphs since it was originated in 1970s;
- A general framework for transductive inference is built on the proposed directed spectral clustering approach.

References:

1. D. Zhou, J. Huang and Schölkopf. *Learning from Labeled and Unlabeled Data on a Directed Graph*. **ICML** 2005.
2. D. Zhou, B. Schölkopf and T. Hofmann. *Semi-supervised Learning on Directed Graphs*. **NIPS** 2004.