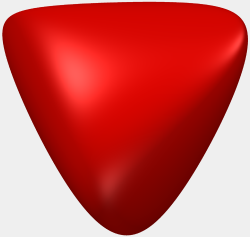
Graph eigenvectors estimation by random walk and diffusion

A study of robustness

Hoang NT

Murata Laboratory
Tokyo Tech



2018/10/24

1. Stability problem
   - Observation
   - Effectiveness of approximation methods
2. Implicit regularizations
   - Problem definition
   - Solution to the regularized SPD
   - Plug the regularizer in!
3. Conclusion
   - Summary
   - Discussion

Stability problem

Consider a "basically expander" graph G, except that it has two poorly connected components:

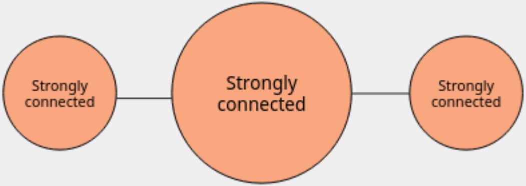
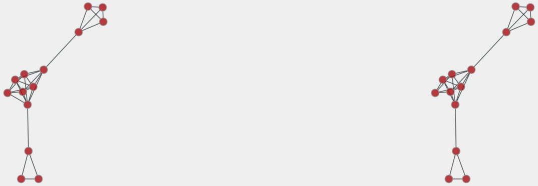


Figure: Almost expander graph.

The problem with this kind of graph is: It is easy to change the edge weights of the central part to make the "mass" of the leading non-trivial eigenvector on the first small component or the second component. **Stability of eigendirection** is clearly an issue.

Example

Graph G having 13 nodes, by removing edge (0,1) we get G':



(a) Original graph (G) **(b)** Perturbed graph (G')

Figure: Almost expander graphs.

The two first non-trivial eigenvectors of these graphs are:

$$v_1^G = \begin{bmatrix} -0.208 & -0.286 & -0.275 & -0.275 & -0.275 & -0.212 & -0.275 & 0.225 & 0.394 & 0.395 \\ 0.126 & 0.221 & 0.221 & 0.221 \end{bmatrix}$$

$$v_1^{G'} = \begin{bmatrix} 0.194 & 0.307 & 0.275 & 0.275 & 0.275 & 0.208 & 0.275 & -0.224 & -0.387 & -0.387 \\ -0.131 & -0.226 & -0.226 & -0.226 \end{bmatrix}$$

Consider an optimization problem:

$$\min_{x \in \mathcal{S}} f(x)$$

This problem might not be "well-posed", so we add a *regularization* term $\lambda g(x)$:

$$\min_{x \in \mathcal{S}} f(x) + \lambda g(x)$$

If we choose $g(x)$ to be strongly convex (σ-strongly convex), we can obtain: increased stability; decreased sensitivity to noise; and avoid overfitting.

In theory, eigenvectors provide a nice way for graph analysis. However, spectral clustering methods are usually worse than heuristic/approximation-based methods [3, 1].

Hence, there must be some form of **implicit** regularization for each of these heuristic algorithms.

Research question

To what extent can one formaliza the idea that performing an approximate computation can implicitly lead to more regular solutions? [2].

Implicit regularizations

Aim

The aim of paper [2] is to connect the solution of a regularized SDP problem to three diffrent commonly used random walk heuristics:

Heat Kernel An operator describing the diffusive spreading of heat on the graph.

$$H_t = \exp(-tL) = \sum_{k=0}^{\inf} \frac{(-t)^k}{k!} L^k$$

PageRank The PageRank vector is given by:

$$\pi(\gamma, s) = \gamma s + (1-\gamma) M \pi(\gamma, s)$$

TLRW $M = AD^{-1}$ is the natural random walk transition matrix, the Truncated α-Lazy Random Walk matrix is given by:

$$W_\alpha = \alpha I + (1-\alpha) M$$

Consider the standard SPECTRAL problem:

$$\begin{aligned} \min \quad & x^T L x \\ \text{s.t.} \quad & x^T x = 1 \\ & x^T D^{1/2} 1 = 0 \end{aligned}$$

This problem is *equivalent* to a SDP:

$$\begin{aligned} \min \quad & L \cdot X \\ \text{s.t.} \quad & Tr(X) = I \cdot X = 1 \\ & X \succeq 0 \end{aligned}$$

The exact solution for SDP is when X is rank-1, i.e. $X = xx^T$. However, when the solution is not rank-1, a simple way to construct a vector x from X is to sample $\xi \backsim N(0, 1/n)$ and build $x = X^{1/2}\xi$.

Regularized SDP

Consider the form:

$$(F,\eta) - SDP \quad \begin{aligned} &\min L \cdot X + \tfrac{1}{\eta}F(X) \\ &\text{s.t. } I \cdot X = 1 \\ &\qquad X \succeq 0 \end{aligned}$$

The authors of [2] connects the regularization term $F(X)$ and approximation heuristic as follow:

- When $F(X)$ is **von Neumann entropy**, the solution X^* can be obtained by the **heat kernel**.
- When $F(X)$ is **log-determinant**, the solution X^* can be obtained by **PageRank**.
- When $F(X)$ is **matrix p-norm**, the solution X^* can be obtained by **truncated lazy random walk**.

Theorem 1 [2]

Let G be a connected, weighted, undirected graph, with normalized Laplacian L. Then, the following conditions are sufficient for X^* to be an optimal solution to $(F, \eta) - SDP$.

1. $X^* = (\nabla F)^{-1}(\eta(\lambda^* I - L))$, for $\lambda^* \in \mathbb{R}$,
2. $I \cdot X^* = 1$,
3. $X^* \succeq 0$.

Proof of theorem 1

The proof is pretty straight-forward. Write the Lagrangian $\mathcal{L}$ as:

$$\mathcal{L} = L \cdot X + \frac{1}{\eta} \cdot F(X) - \lambda \cdot (I \cdot X - 1) - U \cdot X$$

Set the gradient of the Lagrangian w.r.t. X to 0:

$$\nabla\mathcal{L} = L + \frac{1}{\eta}(\nabla F)X - \lambda I - U = 0$$

The dual objective function is minimized when:

$$X = (\nabla F)^{-1}(\eta(-L + \lambda^* I + U))$$

We choose λ^* to satisfy the second condition. By Weak Duality (primal problem solution is always greater than or equal to an associated dual problem solution), X^* (X with appropriate λ^*) is an optimal solution to $(F, \eta) - SDP$.

Generalized Entropy and the Heat Kernel

The generalized entropy function (also von Neumann entropy):

$$F_H(X) = Tr(X \log X) - Tr(X)$$

for which:

$$(\nabla F_H)(X) = \log X$$
$$(\nabla F_H)^{-1}(Y) = \exp Y.$$

Hence, the solution to $(F_H, \eta) - SDP$ is:

$$X_H^* = \exp\left(\eta(\lambda I - L\right)$$

By setting $\lambda = -1/\eta \log\left(Tr(exp(-\eta L))\right)$, we have:

$$X * *_H = \frac{H_\eta}{Tr(H_\eta)}$$

.

Log-determinant and PageRank

The log-determinant function is given by:

$$F_D(X) = -\log\det X$$

Similar to the previous manipulation, lemma 2 of [2] showed that:

$$X_D^* = \frac{D^{-1/2} R_\gamma D^{-1/2}}{Tr(R_\gamma)}$$

The p-norm function is given by:

$$F_p(X) = \frac{1}{p} \parallel X \parallel_p^p = \frac{1}{p} Tr(X^p)$$

Lemma 3 of [2] showed that:

$$X_p^* = \frac{D^{\frac{-(q-1)}{2}} W_\alpha^{q-1} D^{\frac{q-1}{2}}}{Tr(W_\alpha^{q-1})}$$

Thus connecting the p-norm and the heuristic algorithms on TLRW matrix.

Conclusion

Summary

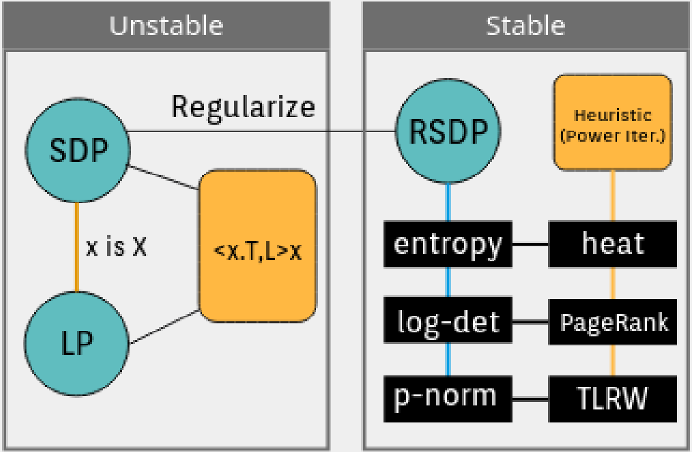


Figure: Graphical summary.

Discussion

The random-walk-based (also difussion-based) view provided several benefits:

- Robustness and stability in computation
- Insight on how heuristic algorithms worked better than spectral methods.

References

Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney.
Statistical properties of community structure in large social and information networks.
In *Proceedings of the 17th international conference on World Wide Web*, pages 695–704. ACM, 2008.

Michael W Mahoney and Lorenzo Orecchia.
Implementing regularization implicitly via approximate eigenvector computation.
arXiv preprint arXiv:1010.0703, 2010.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena.
Deepwalk: Online learning of social representations.
In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.