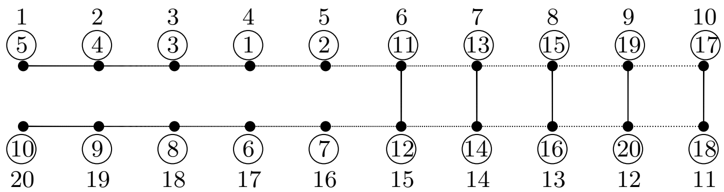
Abstract: Semidefinite Ranking on Graphs

Shankar Vembu[1], Thomas Gärtner[1], and Stefan Wrobel[1,2]

[1] Fraunhofer IAIS, Schloß Birlinghoven, 53754 Sankt Augustin, Germany
[2] Department of Computer Science III, University of Bonn, Germany
{shankar.vembu, thomas.gaertner, stefan.wrobel}@iais.fraunhofer.de

We consider the problem of ranking the vertices of an undirected graph given some preference relation. Without inconsistent preferences in the data, the preferences would form a partial order and we could aim at finding the linear extension that conforms best with the undirected graph. However, in real data there are also inconsistent preferences and hence we have to allow for a few backward edges. This 'ranking on graphs' problem has been tackled before using spectral relaxations. Recently, it has been shown that semidefinite relaxations offer in many cases better solutions than spectral ones for clustering [1] and transductive classification [2]. In this paper, we investigate semidefinite relaxations of ranking on graphs. We incorporate the preferences by fixing certain angles between the metric embedding of the vertices. The final linear extension is obtained by the random projection method. Experiments on real world data sets show the expected improvements over spectral relaxations.

Several special cases of 'ranking on graphs' are well investigated problems like 'topological sort', 'minimum feedback arc set', and 'minimum-length ordering'. The latter problems are known to be NP-hard. The best known approximation algorithms for 'minimum-length ordering' are based on semidefinite relaxations and random projection of the resulting point set.



Spectral relaxation of ranking on graphs has been explored in [3]. Their approach is strongly related to the spectral relaxation made in spectral clustering algorithms. One problem with spectral relaxations that has been found in clustering is that even on simple toy graphs (see the graph above for an example) the spectral solution can be arbitrarily far from the optimal one [4]. In fact, the same happens for spectral and semidefinite relaxations of the minimum-length ordering problem as illustrated by the numbers in the above graph. There, circled numbers indicate the ordering for the semidefinite relaxation (with length of ordering 173), other numbers indicate the ordering for the spectral relaxation (with length of ordering 184).

1 Combinatorial Optimisation Problems

The combinatorial optimisation problem 'ranking on graphs' is given a directed graph (V, D) and an undirected graph (V, E) find:

$$\begin{aligned} &\underset{\pi \in V \to [\![n]\!]}{\operatorname{argmin}} \sum_{(u,v)\in D} \sigma\left(\pi(u) - \pi(v)\right) + \nu \sum_{\{u,v\}\in E} \left[\pi(u) - \pi(v)\right]^2 \\ &\text{subject to: } \pi(u) = \pi(v) \Leftrightarrow u = v \qquad \forall u, v \in V \end{aligned} \tag{1}$$

where σ is the step function and ν a regularisation parameter. In the above optimisation problem, the constraints make sure the function π is bijective and hence a linear embedding (We use $[\![n]\!] = \{1, \ldots, n\}$ and can without loss of generality assume $V = [\![n]\!]$ such that π is just a permutation of $[\![n]\!]$). Several known combinatorial optimisation problems are related to problem (1).

Topological Sort: Here we want to find any permutation of the vertices $\pi : V \to [\![n]\!]$ such that $\forall (u, v) \in D : \pi(u) < \pi(v)$.

Minimum Feedback Arc Set [5]:

$$\underset{\pi : V \to [\![n]\!]}{\operatorname{argmin}} \sum_{(u,v)\in D} \sigma\left(\pi(u) - \pi(v)\right) \quad \text{s.t.: } \forall u, v \in V : \ \pi(u) = \pi(v) \Leftrightarrow u = v$$

Minimum Length Ordering [6]:

$$\underset{\pi : V \to [\![n]\!]}{\operatorname{argmin}} \sum_{\{u,v\}\in E} \left[\pi(u) - \pi(v)\right]^2 \quad \text{s.t.: } \forall u, v \in V : \ \pi(u) = \pi(v) \Leftrightarrow u = v$$

2 Relaxed Ranking on Graphs

To simplify the optimisation problem (1) we take a convex loss function for violated constraints instead of the step function (aka 0/1-loss)

$$\begin{aligned} &\underset{\xi \in \mathbb{R}^D, \pi : V \to [\![n]\!]}{\operatorname{argmin}} \sum_{(u,v)\in D} |\xi_{(u,v)}| + \nu \sum_{\{u,v\}\in E} \left[\pi(u) - \pi(v)\right]^2 \\ &\text{subject to: } \pi(u) < \pi(v) + \xi_{(u,v)} \qquad \forall (u, v) \in D \\ &\qquad\qquad\quad \pi(u) = \pi(v) \Leftrightarrow u = v \qquad \forall u, v \in V \end{aligned}$$

and could obtain a spectral relaxation as

$$\begin{aligned} &\underset{\xi \in \mathbb{R}^D, f : V \to \mathbb{R}}{\operatorname{argmin}} \sum_{(u,v)\in D} |\xi_{(u,v)}| + \nu \sum_{\{u,v\}\in E} \left[f(u) - f(v)\right]^2 \\ &\text{subject to: } f(u) < f(v) - 1 + \xi_{(u,v)} \qquad \forall (u, v) \in D \ . \end{aligned}$$

The latter optimisation problem is considered in [3]. While spectral relaxations have been and still are very popular in the machine learning community, current reseach results show that semidefinite relaxations are more powerful than spectral ones. The conceptual difference is that rather than searching for an embedding of the vertices onto the real line ($f : V \to \mathbb{R}$) we are now looking for a metric embedding into a Euclidean space ($f : V \to \mathbb{R}^d$). While SDP relaxations for the related problem of minimum length ordering are known [6], how to best incorporate the preference constraints is an open question.

Our approach is motivated by the random projection method typically employed to obtain an order from the metric embedding. A typical approach for semidefinite relaxations is to ensure that the embedded vertices lie indeed in a low-dimensional subspace and are equispaced using spreading metric constraints. It can be shown that then the probability of three points being in a particular order is related to the angle between them [6]. Hence, we pursue the approach of encoding each preference $(u, v) \in D$ as a constraint on the angle such that $\langle f(z) - f(u), f(v) - f(u)\rangle = -\|f(z) - f(u)\|\|f(v) - f(u)\|$ where $f(z) \in \mathbb{R}^d$ is a auxiliary vector that is used for all preferences. Unfortunately, it is not a semidefinite constraint and hence we resort to

$$\begin{aligned} &\underset{f:V\cup\{z\}\to\mathbb{R}^d}{\operatorname{argmin}} \sum_{(u,v)\in D} \langle f(z) - f(u), f(v) - f(u)\rangle + \nu \sum_{\{u,v\}\in E} \|f(u) - f(v)\|^2 \\ &\text{subject to: } \|f(z)\|^2 + \textstyle\sum_{u\in V} \|f(u)\|^2 = n \end{aligned} \tag{2}$$

which can directly be encoded in a SDP by a change of variables from $f : V \to \mathbb{R}^d$ to $X \in \mathbb{R}^{(V\cup\{z\})\times(V\cup\{z\})}$ with $X_{uv} = \langle f(u), f(v)\rangle$. The embedding can then be obtained by an incomplete Cholesky factorisation of X. From the embedding we obtain a ranking by projecting onto a random vector [6].

3 Experimental results

We compared the performance of our algorithm with spectral relaxation for ranking on graphs with a benchmark collection of nine metric regression data sets. The data sets were transformed into ordinal regression data sets by discretising the target values into equal-length bins where the bins correspond to categories. The similarity graphs were were computed using a Gaussian kernel. We generated preference constraints by forming a complete bipartite graph between training instances of successive categories. We used an inverse 5-fold cross validation in all our experiments. In all but the *Diabetes* data set, we used five bins corresponding to five ordinal values. For the *Diabetes* data set, we used only two bins because the number of training instances in each (inverse) fold would be very low if five bins were used. To evaluate the quality of the rankings, we used the Kendall tau rank correlation coefficient τ which measures the difference in the number of concordant and discordant pairs as a fraction of the total number of pairs of instances. τ varies from -1 to $+1$, where $+1$ indicates perfect correlation. To verify statistical significance of the observed differences, we use a one-sided Wilcoxon signed-ranks test with the null hypothesis that semidefinite ranking on graphs does not outperform spectral ranking on graphs. The critical value of the one-sided Wilcoxon signed ranks test for 8 samples on a 5% significance level is 5. The test statistic for comparing semidefinite ranking on graphs

Table 1. Results of semidefinite and spectral relaxations for ranking on ordinal regression data sets. The results are the values of τ averaged over all folds along with their standard deviation. The total number of instances is given in *Ins* column and the number of preferences used in each fold is given in the *Pref* column

Dataset	Ins	Pref	Bins	$(\tau+1)/2$, SDP	$(\tau+1)/2$, Spectral
Diabetes	43	15	2	0.60 ± 0.03	0.58 ± 0.06
Pyrimidines	74	99	5	0.71 ± 0.05	0.63 ± 0.09
Triazines	186	284	5	0.58 ± 0.04	0.52 ± 0.01
Wisconsin	194	212	5	0.59 ± 0.03	0.54 ± 0.03
Machine	209	164	5	0.76 ± 0.02	0.73 ± 0.02
Auto	392	1230	5	0.78 ± 0.02	0.82 ± 0.01
Housing	506	2043	5	0.66 ± 0.02	0.61 ± 0.02
Stocks	950	6736	5	0.78 ± 0.02	0.81 ± 0.04

with spectral ranking on graphs is 4.5. Hence on the 5% significance level we can reject the null hypothesis. We conclude that semidefinite relaxation significantly outperforms spectral relaxation for ranking on graphs.

4 Conclusion

We proposed a semidefinite relaxation for ranking on graphs and empirically evaluated it on several real world ordinal regression data sets. Our formulation is inspired by the approximation algorithm for minimum-length ordering. It searches for a metric embedding such that projections of the points on random vectors respect the preference constraints with high probability and such that the embedding is smooth in the sense that points corresponding to neighboring vertices should not have a large distance. The empirical results show clear improvements of the semidefinite relaxation over the spectral one. In future, we will investigate the effect of using spreading metric constraints as well as alternative encodings for the preference constraints. Last but not least, we aim at an implementation that easily scales to huge networks with millions of vertices.

References

1. Lang, K.: Fixing two weaknesses of the spectral method. In: NIPS. (2005)
2. Bie, T.D., Cristianini, N.: Fast SDP relaxations of graph cut clustering, transduction, and other combinatorial problems. JMLR **7** (2006) 1409–1436
3. Agarwal, S.: Ranking on graph data. In: ICML. (2006)
4. von Luxburg, U.: A tutorial on spectral clustering. Technical report, Max Planck Institute for Biological Cybernetics (2006)
5. Even, Naor, Schieber, Sudan: Approximating minimum feedback sets and multi-cuts in directed graphs. In: IPCO. (1995)
6. Blum, A., Konjevod, G., Ravi, R., Vempala, S.: Semi-definite relaxations for minimum bandwidth and other vertex-ordering problems. TCS **235**(1) (2000) 25–42