Semidefinite Ranking on Graphs
— Shankar Vembu, Thomas Gärtner, and Stefan Wrobel —

Abstract
In this paper we tackle the problem of ranking the vertices in an undirected graph given some preference relation. We view the preferences as a preorder and aim at finding a linear order which conforms best with the undirected graph while at the same time violating as few as possible preference constraints. This problem has been tackled before using spectral relaxations. For many other problems, including combinatorial optimisation problems as well as clustering and classification problems, it has been observed that semidefinite relaxations offer several advantages over spectral ones. To solve ranking problems using semidefinite programming, we incorporate the preferences by enforcing certain angles between the embedding of the vertices. The final ordering is then obtained by the random projection method. Experiments on benchmark data sets show the expected improvements over spectral relaxations.

1 Introduction
We consider the problem of ranking the vertices in an undirected graph given some preference relation. In a noise free setting, i.e., with only consistent 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.

To formalise this ‘ranking on graphs’ problem, we represent the preferences between the training vertices on an arbitrary vertex set $V$ by a directed graph $(V, D)$ with $D \subseteq V \times V$. The intended meaning of each edge $(u, v) \in D$ of the directed graph is that $u$ should be ranked before $v$. In addition to these preferences we assume prior knowledge about the smoothness of the ordering given in the form of an undirected graph $(V, E)$ with $E \subseteq \{(u, v) \mid u, v \in V\}$. The intended meaning of the undirected graph is that we prefer orderings in which vertices that are connected by an edge are close together in the ordering. The combinatorial optimisation problem ‘ranking on graphs’ is then defined as:

$$\arg\min_{\pi \in V \rightarrow [n]} \sum_{(u,v) \in D} \sigma(\pi(u) - \pi(v)) + \nu \sum_{\{u,v\} \in E} [\pi(u) - \pi(v)]^2$$

subject to: $\pi(u) = \pi(v) \iff u = v$ \quad $\forall u, v \in V$

(1)

where $\sigma$ is the step function ($\sigma(a) = 1 \iff a \geq 0$ and $\sigma(a) = 0 \iff a < 0$) and $\nu$ a parameter.

The first term in the objective function counts the number of violated preference constraints and we will, hence, refer to it as the empirical error term. The second term in the objective function sums the lengths of the undirected edges and, hence, takes care of our preference for smooth embeddings. We will refer to the second term as the regularisation term and to $\nu$ as the regularisation parameter.

The constraints of the optimisation problem make sure that the function $\pi$ is bijective and hence a linear ordering. We use $[n] = \{1, \ldots, n\}$ and can without loss of generality assume $V = [n]$. In this case the above constraint implies that $\pi$ is just a permutation of $[n]$. 

— 1 —
The above defined ‘ranking on graphs’ problem has been tackled before using spectral relaxations in (Agarwal, 2006). In the machine learning community as well as in the combinatorial optimisation community, it has recently been noticed that semidefinite relaxations offer advantages over spectral ones. For machine learning, this has been shown in clustering (Lang, 2004, 2005) as well as classification (Bie and Cristianini, 2006) problems where semidefinite relaxations lead to stronger learning algorithms. For combinatorial optimisation, it has been shown that with semidefinite relaxations often better approximation guarantees can be obtained than with spectral relaxations. This holds in particular also for many vertex ordering problems (sometimes called (linear) layout problems) like minimum linear arrangement (Charikar et al., 2006), minimum bandwidth (Blum et al., 2000), minimum length ordering (Blum et al., 2000), and many more. Motivated by these findings, 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 ordering 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 (Agarwal, 2006). 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 in Figure 1 for an example) the spectral solution can be arbitrarily far from the optimal one (von Luxburg, 2006). In fact, the same happens for spectral and semidefinite relaxations of the minimum-length ordering problem as illustrated in Figure 1. There, circled numbers indicate the ordering for the semidefinite relaxation (with total length 173), other numbers indicate the ordering for the spectral relaxation (with total length 184). The total length is the sum of the squared differences of the images of adjacent vertices.

![Figure 1: The cockroach graph along with the linear orderings found from a spectral relaxation as well as from a semi-definite relaxation of the minimum length ordering problem. Circled numbers indicate the ordering for the semidefinite relaxation (with total length 173), other numbers indicate the ordering for the spectral relaxation (with total length 184).](image)

This paper is organised as follows: Section 2 describes optimisation problems related to (1), among them several well known combinatorial optimisation problems that can be obtained as special cases of (1) by dropping one or the other term. Section 3 describes how semidefinite relaxations of the empirical error term as well as of the regularisation term can be obtained, gives the SDP formulation, and summarises the proposed learning algorithm.
After we describe experimental results on several benchmark data sets in Section 4, we discuss open question and future work in Section 5. Finally, Section 6 concludes.

2 Related Work

2.1 Combinatorial Optimisation Problems

Several known combinatorial optimisation problems are related to the ranking on graphs problem that we defined above in Equation (1):

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

**Minimum Feedback Arc Set**
\[
\arg\min_{\pi : V \rightarrow [n]} \sum_{(u,v) \in D} \sigma(\pi(u) - \pi(v)) \quad \text{s.t.:} \quad \forall u, v \in V : \pi(u) = \pi(v) \Leftrightarrow u = v
\]

**Minimum Length Ordering**
\[
\arg\min_{\pi : V \rightarrow [n]} \sum_{\{u,v\} \in E} [\pi(u) - \pi(v)]^2 \quad \text{s.t.:} \quad \forall u, v \in V : \pi(u) = \pi(v) \Leftrightarrow u = v
\]

While topological sort is a rather simple problem, minimum feedback arc set as well as minimum length ordering are NP-hard optimisation problems. Approximate solutions to the minimum feedback arc set problem have, for instance, been obtained using linear programming relaxations as described in (Even et al., 1995) and have later been improved by introducing spreading constraints in (Even et al., 2000). Approximate solutions to the minimum length ordering problem have been obtained, for instance, in (Blum et al., 2000) using a semidefinite relaxation with spreading constraints. For related optimisation problems (though not special cases of the ranking on graphs problem), similar semidefinite relaxations have been proposed, e.g., for minimum bandwidth in (Blum et al., 2000), and in some cases even give the best approximation guarantees known so far, e.g., for minimum linear arrangement (Charikar et al., 2006).

While the minimum feedback arc set set corresponds to the unregularised error term of (1), the minimum length ordering corresponds to the unsupervised special case of (1). To the best of our knowledge no semidefinite relaxations of the minimum feedback arc set are known.

2.2 Spectral 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)
\[
\arg\min_{\xi \in \mathbb{R}^D, \pi : V \rightarrow [n]} \sum_{(u,v) \in D} |\xi_{u,v}|^p + \nu \sum_{\{u,v\} \in E} [\pi(u) - \pi(v)]^2 \quad \text{s.t.:} \quad \forall (u, v) \in D : \pi(u) < \pi(v) + \xi_{u,v} \\
\pi(u) = \pi(v) \Leftrightarrow u = v \quad \forall u, v \in V
\]
with parameter $p \in \{1, 2\}$ and could obtain a spectral relaxation\footnote{We use the term spectral relaxation here as the Fiedler vector serves as a good solution to the unsupervised problem.} as

$$\argmin_{\xi \in \mathbb{R}^D, f : V \to \mathbb{R}} \sum_{(u,v) \in D} |\xi_{(u,v)}|^p + \nu \sum_{\{u,v\} \in E} [f(u) - f(v)]^2$$

subject to: $f(u) < f(v) - 1 + \xi_{(u,v)} \forall (u, v) \in D$.

For $p = 1$, this is the optimisation problem considered in (Agarwal, 2006). In the empirical evaluation this relaxation with $p = 2$ will serve as the baseline to our semidefinite relaxation. To the best of our knowledge (Agarwal, 2006) and (Agarwal et al., 2006) are the only transductive ranking approaches. In (Agarwal, 2006) it has been shown that transductive ranking—as expected—well outperforms supervised ranking (in this case represented by (Freund et al., 2003)). The relation between (Agarwal, 2006) and (Agarwal et al., 2006) has been discussed in (Agarwal and Chakrabarti, 2007).

### 3 Semidefinite Relaxation

While spectral relaxations have led to several successful machine learning algorithms, current research results show that semidefinite relaxations are more powerful than spectral ones. The conceptual difference to spectral relaxations 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 an embedding into a Euclidean space ($f : V \to \mathbb{R}^d$). While searching for an optimal embedding often is a non-convex optimisation problem, a change of variables from the embedding to the inner products between the embedded vertices is often sufficient to obtain a convex optimisation problem in the form of a semidefinite program. From the solution matrix of the semidefinite program, an embedding can be recovered by factorising the matrix. From this embedding we can obtain a ranking by projecting onto a random vector. This is also the method used, for instance, when approximating the minimum length ordering problem using an SDP relaxation (Blum et al., 2000).

The algorithm consists hence of three main steps:

1. Solve the semidefinite relaxation of ranking on graphs.
2. Factorise the Gram matrix using the incomplete Cholesky method.
3. Choose a set of random vectors, project the embedded points on each of them, and choose the projections that results in the best objective function value for the ranking on graphs problem.

While steps two and three are straightforward, it remains to give the details of the semidefinite program. In the remaining parts of this section, we investigate the error term and the regularisation term of (1) separately. While the regularisation term corresponds to the minimum length ordering problem, for which SDP relaxations are known, how to best incorporate the preferences used in the error term is an open question. Section 3.1 will introduce our relaxation of the error term, starting from a set of constraints that—in the noise free setting—enforces all preference constraints, which however, is not convex. The result of Section 3.1 is a convex relaxation of the empirical error term that will be used in the objective function.
Section 3.2 discusses the relaxation of our regularisation term as it is used in the approximation algorithm for the minimum length ordering problem as well as in relaxations of similar combinatorial optimisation problems. We argue that the spreading constraints, which are essential for obtaining good approximation guarantees in the unsupervised case (minimum length ordering), can, in fact, be ignored for ranking on graphs.

3.1 Empirical Error Term

To see how to relax the empirical error term of (1) we note that the probability of three points being in a particular order after a random projection is proportional to the angle between them (Blum et al., 2000). In a noise free learning problem (the set of preferences forms a partial order) we could simply enforce that for each preference \((u, v) \in D\) the vectors \(f(z) - f(u)\) and \(f(v) - f(u)\) are anti-parallel, i.e., the following constraint:

\[
\langle f(z) - f(u), f(v) - f(u) \rangle = -\|f(z) - f(u)\| \|f(v) - f(u)\|
\]

where \(f(z) \in \mathbb{R}^d\) is a auxiliary vector that is used in all preferences and introduced for convenience only.

As the preference constraints might contain inconsistencies due to noise, we have to allow for some violation of the constraints. To this end, our approach is to aim at an embedding such that \(\forall (u, v) \in D\) the angle between \(f(z) - f(u)\) and \(f(v) - f(u)\) is as large as possible. Hence, we would like to use the term

\[
\sum_{(u,v) \in D} \langle f(z) - f(u), f(v) - f(u) \rangle + \sum_{(u,v) \in D} \|f(z) - f(u)\| \|f(v) - f(u)\|
\]

in the objective function as our relaxation of the error term. Note, however, that neither \(\langle f(z) - f(u), f(v) - f(u) \rangle\) nor \(\|f(z) - f(u)\| \|f(v) - f(u)\|\) are convex functions (in fact, they are not even quasiconvex). This holds already for a single edge. To see non-convexity, consider \(g(a,b) = (0-a)(b-a) = a^2 - ab\) with Hessian \(
\begin{pmatrix} 2 & -1 \\ -1 & 0 \end{pmatrix}
\). This shows that for \((1,3)^\top\) the function has negative curvature and for \((1,-3)^\top\) the function has positive curvature. To see non-quasiconvexity (i.e., that not all the sublevel sets are convex) we need to find two points \(x, y \in \mathbb{R}^2\) \((x = (a,b)^\top, y = (a’,b’)\top)\) such that \(y^\top \nabla g(x) = 0 \land y^\top \nabla^2 g(x)y < 0\). Such points can be found as \(x = (0,0)^\top\) and \(y = (1,3)^\top\). This means that we cannot guarantee that the overall function that we are trying to minimise is convex. In order to be able to solve the problem efficiently we, hence, need to relax the problem further such that we obtain a convex problem.

Consider first the inner product \(\sum_{(u,v) \in D} \langle f(z) - f(u), f(v) - f(u) \rangle\). Despite the non-convexity of this term, we can directly transform it into a convex problem by a change of variables from vectors \(v \mapsto f(u) \in \mathbb{R}^d\) to inner products between vectors \(u, v \mapsto \langle f(u), f(v) \rangle \in \mathbb{R}\). That is, instead of optimising over the embedding \(F \in \mathbb{R}^{(V \cup \{z\}) \times d}\) with \(F^\top_v = f(v)\) directly, we optimise over the Gram matrix \(X = FF^\top\), subject to the constraint that \(X\) is a valid Gram matrix, i.e., it is positive definite. In fact, in terms of the entries of the Gram matrix we get

\[
\sum_{(u,v) \in D} \langle f(z) - f(u), f(v) - f(u) \rangle = X_{zz} - X_{zu} - X_{v} + X_{uv}
\]

and hence

\[
\sum_{(u,v) \in D} \langle f(z) - f(u), f(v) - f(u) \rangle = \sum_{(u,v) \in D} [X_{zz} - X_{zu} - X_{v} + X_{uv}]
\]

becomes a linear function! This implies that we can optimise over the Gram matrix using semidefinite program-
ming. The embedding can then be obtained from the resulting Gram matrix by, e.g., an incomplete Cholesky factorisation.

Consider now the norm-product term \( \sum_{(u,v) \in D} \|f(z) - f(u)\| \|f(v) - f(u)\| \), which is neither convex nor allows for a similar neat trick to obtain an equivalent convex problem. For most of this paper, we will hence ignore this term. In Section 5 we will briefly describe alternative approaches, including an attempt to minimise the non-convex problem. We will see there that we cannot achieve better orderings by including the norm-product term.

3.2 Regularisation Term

The regularisation term and its semidefinite relaxation correspond directly to the minimum length ordering problem and its semidefinite relaxation, respectively. In either case, the semidefinite relaxation is given by

\[ \sum_{\{u,v\} \in E} \|f(u) - f(v)\|^2 = \sum_{\{u,v\} \in E} [X_{uu} - 2X_{uv} + X_{vv}] \]

and can, hence, readily be computed from the Gram matrix.

To obtain good approximation guarantees for linear or semidefinite relaxations of combinatorial optimisation problems, the most common approach is to ensure that the embedded vertices lie in a low-dimensional subspace and are equispaced. This can be achieved, for instance, with the use of spreading constraints (Even et al., 2000) of the form

\[ \forall v \in V, S \subseteq V \setminus \{v\} : \sum_{u \in S} \|f(u) - f(v)\|^2 \geq 2 \left( 1^2 + 2^2 + 3^2 + \cdots + \frac{|S|^2}{4} \right). \]

These constraints are convex and indeed ensure that the points \( \{f(v) \mid v \in V\} \) lie in a low-dimensional subspace forming approximately a line. It can indeed be seen that if these constraints are satisfied then the point set has a low fractal dimension. For minimum length ordering (Blum et al., 2000) used an additional set of constraints on the area of the triangle formed by triples of vectors. While there is an exponential number of spreading constraints \( |V|^2 |V|^{-1} \), semidefinite optimisation problems with exponentially many constraints can be solved in polynomial time using the ellipsoid method (see, e.g., (Korte and Vygen, 2002)) if there is a polynomial time (weak) separation oracle. For spreading constraints such a separation oracle is well known (Even, 1999). Despite the polynomial complexity of this approach, its practical runtime requirements are too large to allow for solving large scale problems.

In this paper we will hence exploit the presence of the preference constraints and investigate if we can achieve good empirical results without spreading constraints. Note while spreading constraints cannot be ignored for minimum length ordering, they can potentially be ignored for ranking on graphs. The difference is that for the semidefinite relaxation of minimum length ordering the spreading constraints are also needed to ensure that the SDP solution does not map all vertices onto the same point. For ranking on graphs this gets readily taken care of by the error term, which (implicitly) enforces the training vertices to be mapped to different points in the embedding. This, in turn, carries over to the test instances ‘between’ some training instances.
3.3 Semidefinite Program

We are now ready to summarise the main part of our approach, the semidefinite program. In the objective function we have the relaxed error and regularisation term. The only constraint we use is to make sure that the sum of the norms of the embeddings is fixed.

Let $Z \in \mathbb{R}^{(V \cup \{z\}) \times D}$, $P \in \mathbb{R}^{(V \cup \{z\}) \times D}$ with $\forall (u, v) \in D : Z_{z,(u,v)} = +1, Z_{u,(u,v)} = -1, P_{v,(u,v)} = +1, P_{u,(u,v)} = -1$ and let $L$ be the Laplacian of the graph $(V, E)$. For an embedding $F \in \mathbb{R}^{V \cup \{z\} \times d}$ the error term is $\text{tr} F^\top P Z^\top F = \frac{1}{2} \text{tr} \left[ (P Z^\top + Z P^\top) F F^\top \right]$ and the regularisation term is $\text{tr} F^\top L F = \text{tr} L F F^\top$. Then, the final SDP optimising over $X = F F^\top$ is:

$$\arg\min_{X \in \mathbb{R}^{V \cup \{z\} \times V \cup \{z\}}} \text{tr} \left[ \left( \frac{Z P^\top + P Z^\top}{2} + \nu L \right) X \right]$$

subject to: $\text{tr} [X] = |V|$

where $\text{tr}$ denotes the matrix trace.

4 Experimental results

We compared the performance of our algorithm with spectral ranking on a benchmark collection of nine metric regression data sets. These are the same data sets used in (Chu and Ghahramani, 2005) for ordinal regression. The problem of ordinal regression is indeed subsumed under the more general problem of ranking. Here the goal is to ensure that categories of high ordinal values are ranked higher than those with lower values. The benchmark 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 data was standardised to have zero mean and unit variance and the similarity graphs were computed using a Gaussian kernel. We generated preference constraints by forming a complete bipartite graph between training instances of successive categories. The number of instances (vertices in the undirected graph) and the number of preferences (edges in the preference graph) are shown in Table 1.

We used an inverse $k$-fold (with $k = 5$) cross validation in all our experiments. Inverse $k$-fold cross-validation is commonly used to estimate the generalisation error of transductive learning algorithms, where instances in a single fold are used for training and the rest from the $k-1$ folds are used for testing. 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. The number of random projections was fixed to 100. We used DSDP (Benson et al., 2000) and L-BFGS-B (Zhu et al., 1997) as the underlying solvers in our implementations of semidefinite and spectral ranking algorithms. The run times reported in Table 1 indicate the average time spent in each call to the corresponding solver.

To evaluate the quality of the rankings, we used the Kendall tau rank correlation coefficient $\tau$ which measures the difference in the number of concordant and discordant pairs as a fraction of the total number of pairs of instances. $\tau$ varies from $-1$ to $+1$, where $+1$ indicates perfect correlation. Table 1 shows the Kendall tau for the semidefinite as well as the spectral relaxation of ranking on graphs.

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 out-
Table 1: Results of semidefinite and spectral relaxations for ranking on ordinal regression data sets. The results are the values of $\tau'$ averaged over all folds along with their standard deviation. The total number of instances is given in the $\text{Ins}$ column and the number of preferences used in each fold is given in the $\text{Pref}$ column. Run time in seconds is also included.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$\text{Ins}$</th>
<th>$\text{Pref}$</th>
<th>Bins</th>
<th>$\tau'$, SDP</th>
<th>Time (in s)</th>
<th>$\tau'$, Spectral</th>
<th>Time (in s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>43</td>
<td>15</td>
<td>2</td>
<td>0.60 ± 0.03</td>
<td>0.05</td>
<td>0.58 ± 0.06</td>
<td>0.26</td>
</tr>
<tr>
<td>Pyrimidines</td>
<td>74</td>
<td>99</td>
<td>5</td>
<td>0.71 ± 0.05</td>
<td>0.50</td>
<td>0.63 ± 0.09</td>
<td>0.26</td>
</tr>
<tr>
<td>Triazines</td>
<td>186</td>
<td>284</td>
<td>5</td>
<td>0.58 ± 0.04</td>
<td>3.70</td>
<td>0.52 ± 0.01</td>
<td>0.75</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>194</td>
<td>212</td>
<td>5</td>
<td>0.59 ± 0.03</td>
<td>10.02</td>
<td>0.54 ± 0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>Machine</td>
<td>209</td>
<td>164</td>
<td>5</td>
<td>0.76 ± 0.02</td>
<td>3.59</td>
<td>0.73 ± 0.02</td>
<td>3.12</td>
</tr>
<tr>
<td>Auto</td>
<td>392</td>
<td>1230</td>
<td>5</td>
<td>0.78 ± 0.02</td>
<td>15.59</td>
<td>0.82 ± 0.01</td>
<td>21.48</td>
</tr>
<tr>
<td>Housing</td>
<td>506</td>
<td>2043</td>
<td>5</td>
<td>0.66 ± 0.02</td>
<td>67.69</td>
<td>0.61 ± 0.02</td>
<td>23.83</td>
</tr>
<tr>
<td>Stocks</td>
<td>950</td>
<td>6736</td>
<td>5</td>
<td>0.78 ± 0.02</td>
<td>204.18</td>
<td>0.81 ± 0.04</td>
<td>187.51</td>
</tr>
</tbody>
</table>

perform spectral ranking on graphs. Using the Wilcoxon signed-ranks test to compare learning algorithms across a range of data sets has recently been proposed in (Demšar, 2006). The Wilcoxon signed ranks test is a nonparametric test to detect shifts in populations given a number of paired samples. The underlying idea is that under the null hypothesis the distribution of differences between the two populations is symmetric about 0. It proceeds as follows: (i) compute the differences between the pairs, (ii) determine the ranking of the absolute differences, and (iii) sum over all ranks with positive and negative difference to obtain $W_+$ and $W_-$, respectively. The null hypothesis can be rejected if $W_+$ is located in the tail of the null distribution which has sufficiently small probability.

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

5 Discussion and Future Work

While we were able to show that for ranking on graphs our semidefinite relaxation obtains significantly better vertex orderings than the spectral relaxation, one major shortcoming of semidefinite formulations is the rather bad scalability with respect to larger data sets. One approach proposed recently to overcome this problem (Burer and Monteiro, 2003) is to optimise over the low-rank factorisation of the Gram matrix, i.e., over the embedding, rather than over the Gram matrix itself. In general and—as shown above—for our specific formulation, rank restricted optimisation is a non-convex optimisation problem. Still, very good results have been reported for clustering (Lang, 2004) using the low rank SDP solver SDP-LR (Burer and Monteiro, 2003).

To investigate whether similar runtime improvements can also be achieved in semidefinite ranking on graphs, we tried two different approaches. On the one hand, we replaced the SDP solver DSDP (Benson et al., 2000) used in our implementation by the SDP-LR solver. On
the other hand, as low-rank SDP optimisation is anyhow a non-convex problem, we aimed at solving the non-convex problem—now including the norm-product term which had been removed above because of its non-convexity—using the iterative solver L-BFGS-B (Zhu et al., 1997).

For SDP-LR we used the standard settings and for L-BFGS-B we aimed at finding a good local optimum by performing random restarts close to the spectral solution. Quite surprisingly, we were not able to observe results similar to those observed in (Lang, 2004) for clustering. In fact, the solutions found by L-BFGS-B and SDP-LR were not better than the spectral solutions. That both low-rank approaches to ranking on graphs fail might indicate some inherent difficulty posed by our ranking on graphs formulation to low rank SDP solvers.

As one goal of our future work, we will hence investigate whether there is a fundamental reason why low rank SDP solvers fail to find good solutions for the ranking on graphs problem or whether there is an alternative formulation that low rank SDP solvers can deal with more easily than the one proposed in this paper. As an alternative approach to improve the scalability, we will also investigate spectral bundle methods (Helmberg and Rendl, 1999) for ranking on graphs.

As another goal of future work, we will investigate whether the predictions can further be improved by, for instance, making use of spreading constraints. We have argued above that—in contrast to unsupervised ordering problems—spreading constraints are not essential for supervised ordering problems. However, intuitively spreading constraints might serve us well by introducing a kind of margin between the unlabelled points.

6 Conclusion

The ranking on graphs problem studied in the machine learning community generalises well-known combinatorial optimisations problems. In this paper 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 work we will investigate alternative SDP formulations for ranking on graphs. In particular, 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


Steven J. Benson, Yinyu Ye, and Xiong Zhang. Solving large-scale sparse semidefinite programs for combinatorial optimization. SIAM Journal on Optimization, 10(2):443–461, 2000. 7, 8


J. Demsar. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 7(1), 2006. 8


Guy Even. Fast approximate graph partitioning algorithms. SIAM Journal Computing, 28(6), 1999. 6


