Convex Optimization: Project #5

Due on Wednesday 13nd December, 2023 at 23:59pm

Jacopo Peroni

Problem 1

From the first constraint in the QCQP we get that

$$x_i\,(x_i - 1) = 0 \Longrightarrow x_i \in \{0,1\}$$

that is the second constraint of the IP.
The same condition from the QCQP let us prove that the two objectives are the same, in fact

$$\mathbf{x}^T\mathbf{x} = \sum_{i\in V} x_i^2 = \sum_{i\in V} x_i$$

To conclude we can see that, given $x_i \in \{0,1\}$, for all $(i,j) \in E$, the possible combinations are

$$(0,0)\ (1,0)\ (0,1)\ (1,1)$$

but only the first three follow the constraint $x_i x_j = 0$. These combinations are also the only ones that follow the condition $x_i + x_j \leq 1$.
This proves that the two problems are equivalent.

Problem 2

If we define $\mathbf{X} = \mathbf{x}\mathbf{x}^T$, we have that $\mathbf{X}_{i,j} = \mathbf{x}_i\mathbf{x}_j$, thus $\mathrm{Tr}(\mathbf{X}) = \mathbf{x}^T\mathbf{x}$. This tells us that the objectives in the QCQP and in the SDP relaxation are the same.
If, for every $i \in V$, we set

$$(Q_i)_{h,k} = \begin{cases} 1 & h = k = i \\ 0 & \text{otherwise} \end{cases}$$

so that

$$(Q_i\mathbf{X})_{h,k} = \begin{cases} x_i^2 & h = k = i \\ 0 & \text{otherwise} \end{cases}$$

and we define e_i as the versor in the i-th euclidian direction.
This gives us the equivalence between the condition $x_i^2 = x_i$ and $\mathrm{Tr}(Q_i\mathbf{X}) - e_i^T\mathbf{x} = 0$.
By construction of the matrix $\mathbf{X}$ we have that the constraint $\mathbf{x}_i\mathbf{x}_j = 0$ for all $(i,j) \in E$ can be rewritten as $\mathbf{X}_{i,j} = 0$.
So the QCQP becomes

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{R}^n,\mathbf{X}\in\mathbb{S}^n}{\text{maximize}}\ \mathrm{Tr}\,(\mathbf{X}) \\ &\text{subject to}\ \begin{array}{ll} \mathrm{Tr}(Q_i\mathbf{X}) - e_i^T\mathbf{x} = 0 & \forall i \in V \\ \mathbf{X}_{i,j} = 0 & \forall (i,j) \in E \\ \mathbf{X} = \mathbf{x}\mathbf{x}^T & \end{array} \end{aligned}$$

From the lecture notes on slide 220, we know that

$$\mathbf{X} = \mathbf{x}\mathbf{x}^T \Longleftrightarrow \begin{pmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{pmatrix} \succcurlyeq 0 \text{ and } \mathrm{rank}\begin{pmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{pmatrix} = 1$$

From Schur's lemma we know that

$$\begin{pmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{pmatrix} \succcurlyeq 0 \Longleftrightarrow \mathbf{X} - \mathbf{x}\mathbf{x}^T \succcurlyeq 0 \Longleftrightarrow \begin{pmatrix} 1 & \mathbf{x}^T \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \succcurlyeq 0$$

Thus the problem can be rewritten as

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n, \mathbf{X}\in\mathbb{S}^n}{\text{maximize}} \; & \mathrm{Tr}\left(\mathbf{X}\right) \\ \text{subject to} \; & \mathrm{Tr}(Q_i\mathbf{X}) - e_i^T\mathbf{x} = 0 \quad \forall i \in V \\ & \mathbf{X}_{i,j} = 0 \quad \forall (i,j) \in E \\ & \begin{pmatrix} 1 & \mathbf{x}^T \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \succcurlyeq 0 \\ & \mathrm{rank}\begin{pmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{x}^T & 1 \end{pmatrix} = 1 \end{aligned}$$

And lifting the rank condition we obtain the SDP relaxation

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n, \mathbf{X}\in\mathbb{S}^n}{\text{maximize}} \; & \mathrm{Tr}\left(\mathbf{X}\right) \\ \text{subject to} \; & \mathrm{Tr}(Q_i\mathbf{X}) - e_i^T\mathbf{x} = 0 \quad \forall i \in V \\ & \mathbf{X}_{i,j} = 0 \quad \forall (i,j) \in E \\ & \begin{pmatrix} 1 & \mathbf{x}^T \\ \mathbf{x} & \mathbf{X} \end{pmatrix} \succcurlyeq 0 \end{aligned}$$

Problem 3

Using Schur's Lemma we can reformulate the problem in this form

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^n, \mathbf{X}\in\mathbb{S}^n}{\text{minimize}} \; & -\mathrm{Tr}\left(\mathbf{X}\right) \\ \text{subject to} \; & \mathrm{Tr}(Q_i\mathbf{X}) - e_i^T\mathbf{x} = 0 \quad \forall i \in V \\ & \mathbf{X}_{i,j} = 0 \quad \forall (i,j) \in E \\ & \mathbf{X} - \mathbf{x}\mathbf{x}^T \succcurlyeq 0 \end{aligned}$$

Let's write the Lagrangian of the problem, which in our original form will be a map like this

$$L : \mathbb{R}^n \times \mathbb{S}^n \times \mathbb{R}^n \times \mathbb{R}^{|E|} \times \mathbb{S}^n \longrightarrow \mathbb{R}$$

We use an abuse of notation to define the element of the vector in $\mathbb{R}^{|E|}$ with the indices $(i,j) \in E$

$$L\left(\mathbf{x}, \mathbf{X}, \tilde{\lambda}, k, \mathbf{K}\right) =$$

$$= -\mathrm{Tr}\left(\mathbf{X}\right) + \sum_{i \in V} \tilde{\lambda}_i \left(\mathrm{Tr}\left(Q_i \mathbf{X}\right)\right) - \sum_{i \in V} \tilde{\lambda}_i e_i^T \mathbf{x} + \sum_{(i,j) \in E} k_{i,j} \mathbf{X}_{i,j} + \mathrm{Tr}\left(\mathbf{K}\mathbf{x}\mathbf{x}^T\right) - \mathrm{Tr}\left(\mathbf{K}\mathbf{X}\right)$$

If we focus on the k term we can see that, given that $\mathbf{X} \in \mathbb{S}^n$, we can define a matrix $\tilde{\mathbf{W}} \in \mathbb{S}^n$ like this

$$(W)_{i,j} = \begin{cases} k_{i,j} & (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

and we have that

$$\sum_{(i,j) \in E} k_{i,j} \mathbf{X}_{i,j} = \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{\mathbf{W}}_{i,j} \mathbf{X}_{i,j} = \mathrm{Tr}\left(\tilde{\mathbf{W}} \mathbf{X}\right)$$

Another noticeable fact is that

$$\sum_{i \in V} \tilde{\lambda}_i \left(\mathrm{Tr}\left(Q_i \mathbf{X}\right)\right) = \sum_{i \in N} \tilde{\lambda}_i \mathbf{X}_{i,i} = \mathrm{Tr}\left(\mathrm{Diag}\left(\tilde{\lambda}\right) \mathbf{X}\right)$$

And that

$$\sum_{i \in V} \tilde{\lambda}_i e_i^T \mathbf{x} = \tilde{\lambda}^T \mathbf{x}$$

So the map L is now of this type

$$L : \mathbb{R}^n \times \mathbb{S}^n \times \mathbb{R}^n \times \mathbb{S}^n \times \mathbb{S}^n \longrightarrow \mathbb{R}$$

and it can be written like this

$$L\left(\mathbf{x}, \mathbf{X}, \tilde{\lambda}, \tilde{\mathbf{W}}, \mathbf{K}\right) =$$

$$= -\mathrm{Tr}\left(\mathbf{X}\right) + \mathrm{Tr}\left(\mathrm{Diag}(\tilde{\lambda})\mathbf{X}\right) - \tilde{\lambda}^T \mathbf{x} + \mathrm{Tr}\left(\tilde{\mathbf{W}}\mathbf{X}\right) + \mathrm{Tr}\left(\mathbf{K}\mathbf{x}\mathbf{x}^T\right) - \mathrm{Tr}\left(\mathbf{K}\mathbf{X}\right)$$

The strategy now is to compute the $g(\tilde{\lambda}, \tilde{\mathbf{W}}, \mathbf{K}) = \inf_{\mathbf{x} \in \mathbb{R}^n, \mathbf{X} \in \mathbb{S}^n} L$ and define the dual problem as a maximization of the function g.

$$\nabla_{\mathbf{x}} L = -\tilde{\lambda} + 2\mathbf{K}\mathbf{x} = 0 \Longrightarrow \mathbf{x} = \frac{1}{2}\mathbf{K}^{-1}\tilde{\lambda}$$

Thus the Lagrangian becomes

$$L\left(\mathbf{X}, \tilde{\lambda}, \tilde{\mathbf{W}}, \mathbf{K}\right) =$$

$$= -\mathrm{Tr}\left(\mathbf{X}\right) + \mathrm{Tr}\left(\mathrm{Diag}(\tilde{\lambda})\mathbf{X}\right) - \frac{1}{2}\tilde{\lambda}^T \mathbf{K}^{-1}\tilde{\lambda} + \mathrm{Tr}\left(\tilde{\mathbf{W}}\mathbf{X}\right) + \frac{1}{4}\tilde{\lambda}^T \mathbf{K}^{-1}\mathbf{K}\mathbf{K}^{-1}\tilde{\lambda} - \mathrm{Tr}\left(\mathbf{K}\mathbf{X}\right) =$$

$$= -\mathrm{Tr}\left(\mathbf{X}\right) + \mathrm{Tr}\left(\mathrm{Diag}(\tilde{\lambda})\mathbf{X}\right) - \frac{1}{4}\tilde{\lambda}^T \mathbf{K}^{-1}\tilde{\lambda} + \mathrm{Tr}\left(\tilde{\mathbf{W}}\mathbf{X}\right) - \mathrm{Tr}\left(\mathbf{K}\mathbf{X}\right)$$

We can look a minimization in $\mathbf{X}$

$$\nabla_{\mathbf{X}} L = -I + \mathrm{Diag}\left(\tilde{\lambda}\right) + \tilde{\mathbf{W}} - \mathbf{K} = 0 \Longrightarrow \mathbf{K} = \mathrm{Diag}\left(\tilde{\lambda}\right) + \tilde{\mathbf{W}} - I$$

Thus the inf of L is

$$g\left(\tilde{\lambda}, \tilde{\mathbf{W}}\right) =$$

$$-\mathrm{Tr}\left(\mathbf{X}\right) + \mathrm{Tr}\left(\mathrm{Diag}(\tilde{\lambda})\mathbf{X}\right) - \frac{1}{4}\tilde{\lambda}^T\left(\mathrm{Diag}\left(\tilde{\lambda}\right) + \tilde{\mathbf{W}} - I\right)^{-1}\tilde{\lambda} + \mathrm{Tr}\left(\tilde{\mathbf{W}}\mathbf{X}\right) +$$
$$+\mathrm{Tr}\left(\mathbf{X}\right) - \mathrm{Tr}\left(\mathrm{Diag}\left(\tilde{\lambda}\right)\mathbf{X}\right) - \mathrm{Tr}\left(\tilde{\mathbf{W}}\mathbf{X}\right) =$$
$$= -\frac{1}{4}\tilde{\lambda}^T\left(\mathrm{Diag}\left(\tilde{\lambda}\right) + \tilde{\mathbf{W}} - I\right)^{-1}\tilde{\lambda}$$

Hence the dual problem is

$$\begin{aligned} \underset{\tilde{\lambda}\in\mathbb{R}^n, \tilde{\mathbf{W}}\in\mathbb{S}^n}{\text{maximize}} \; & -\frac{1}{4}\tilde{\lambda}^T\left(\mathrm{Diag}\left(\tilde{\lambda}\right) + \tilde{\mathbf{W}} - I\right)^{-1}\tilde{\lambda} \\ \text{subject to} \; & \tilde{\mathbf{W}}_{i,j} = 0 \quad \forall(i,j) \notin E \end{aligned}$$

If we define two variables

$$\mathbf{W} = -\tilde{\mathbf{W}} \text{ and } \lambda = -\tilde{\lambda}$$

noticing that the problem remains equivalent and the constraint doesn't change.
Transforming the problem in a minimization, we get

$$\begin{aligned} \underset{\lambda\in\mathbb{R}^n, \mathbf{W}\in\mathbb{S}^n}{\text{minimize}} \; & -\frac{1}{4}\lambda^T\left(\mathrm{Diag}\left(\lambda\right) + \mathbf{W} + I\right)^{-1}\lambda \\ \text{subject to} \; & \mathbf{W}_{i,j} = 0 \quad \forall(i,j) \notin E \end{aligned}$$

Adding an epigraphical variable $-\mu$ the problem becomes

$$\begin{aligned} \underset{\lambda\in\mathbb{R}^n, \mathbf{W}\in\mathbb{S}^n, \mu\in\mathbb{R}}{\text{minimize}} \; & -\mu \\ \text{subject to} \; & \mathbf{W}_{i,j} = 0 \qquad \forall(i,j) \notin E \\ & -\tfrac{1}{4}\lambda^T\left(\mathrm{Diag}\left(\lambda\right) + \mathbf{W} + I\right)^{-1}\lambda \leq -\mu \end{aligned}$$

Where the last condition, with Schur's Lemma is

$$\mu - \frac{1}{4}\lambda^T\left(\mathrm{Diag}\left(\lambda\right) + \mathbf{W} + I\right)^{-1}\lambda \leq 0 \iff \begin{pmatrix} \mu & \frac{1}{2}\lambda^T \\ \frac{1}{2}\lambda & \mathrm{Diag}\left(\lambda\right) + \mathbf{W} + I \end{pmatrix} \preccurlyeq 0$$

The strong duality holds because the Slater's condition is satisfied by the couple

$$\mathbf{x} = \left[\frac{1}{2}\;\frac{1}{4}\cdots\frac{1}{2^n}\right]$$
$$\mathbf{X} = \mathrm{Diag}\left(\mathbf{x}\right)$$

because by construction $\mathrm{Tr}(Q_i\mathbf{X}) - \mathbf{x}e_i = 0 \iff \mathbf{X}_{i,i} - \mathbf{x}_i = 0$.
For what concerns the second equality condition, without information on E, the best we can do is have $\mathbf{X}$ not-zero only on the diagonal, knowing that $(i,i) \notin E$. In this way for every possible E the condition is satisfied.

We can prove by induction that the following matrix is positive definite using Sylvester's criterion

$$\begin{pmatrix} 1 & \mathbf{x}^T \\ \mathbf{x} & \mathbf{X} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{4} & \cdots & \frac{1}{2^n} \\ \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2^n} & 0 & 0 & \cdots & \frac{1}{2^n} \end{pmatrix}$$

Let's call A_k the top-left $k \times k$ sub-matrix.
$A_1 = 1$ is positive, thus A_1 is positive definite.
If we fix that A_{k-1} is positive definite, we can compute the determinant of A_k with Laplace's theorem applied on the last column

$$\det A_k = (-1)^{n+1}\frac{1}{2^k}\det\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{2^{k-1}} \\ \frac{1}{2^n} & 0 & 0 & \cdots & 0 \end{pmatrix} + (-1)^{2n}\frac{1}{2^k}\det A_{k-1} =$$

$$= (-1)^{n+1}\frac{1}{2^k}(-1)^{n+1}\frac{1}{2^k}\prod_{i=1}^{k-1}\frac{1}{2^i} + \frac{1}{2^k}\det A_{k-1} = \frac{1}{2^k}\frac{1}{2^k}\prod_{i=1}^{k-1}\frac{1}{2^i} + \frac{1}{2^k}\det A_{k-1} > 0$$

Problem 4

If we set $\mathbf{X} \in \mathbb{S}^n$ and $k \in \mathbb{R}^d$ with $d = |\neg E|$ that is the number of 2-elements combinations not in E.

$$L(\mu, \lambda, \mathbf{W}, k, \mathbf{X}) = -\mu + \sum_{(i,j)\in\neg E} k_{i,j}\mathbf{W}_{i,j} + \mathrm{Tr}\left(\mathbf{X}\left(\mathrm{diag}(\lambda) + I + \mathbf{W} - \frac{1}{4\mu}\lambda\lambda^T\right)\right)$$

If we define

$$(\mathbf{\Lambda})_{i,j} = \begin{cases} k_{i,j} & (i,j) \in \neg E \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 0 & (i,j) \in E \\ k_{i,j} & \text{otherwise} \end{cases}$$

The Lagrangian becomes

$$L(\mu, \lambda, \mathbf{W}, \mathbf{X}, \mathbf{\Lambda}) = -\mu + \mathrm{Tr}(\mathbf{\Lambda}\mathbf{W}) + \mathrm{Tr}\left(\mathbf{X}\left(\mathrm{diag}(\lambda) + I + \mathbf{W} - \frac{1}{4\mu}\lambda\lambda^T\right)\right) =$$

$$= -\mu + \mathrm{Tr}(\mathbf{\Lambda}\mathbf{W}) + \mathrm{Tr}(\mathbf{X}\mathrm{diag}(\lambda)) + \mathrm{Tr}(\mathbf{X}) + \mathrm{Tr}(\mathbf{X}\mathbf{W}) - \frac{1}{4\mu}\mathrm{Tr}\left(\mathbf{X}\lambda\lambda^T\right) =$$

If we compute the gradient respect to $\mathbf{W}$, we get

$$\nabla_{\mathbf{W}} L = \mathbf{\Lambda} + \mathbf{X} = 0 \Longleftrightarrow \mathbf{\Lambda} = -\mathbf{X}$$

This condition makes the Lagrangian

$$= -\mu - \mathrm{Tr}\left(\mathbf{X}\mathbf{W}\right) + \mathrm{Tr}\left(\mathbf{X}\mathrm{diag}\left(\lambda\right)\right) + \mathrm{Tr}\left(\mathbf{X}\right) + \mathrm{Tr}\left(\mathbf{X}\mathbf{W}\right) - \frac{1}{4\mu}\mathrm{Tr}\left(\mathbf{X}\lambda\lambda^T\right) =$$

$$= -\mu + \mathrm{Tr}\left(\mathbf{X}\mathrm{diag}\left(\lambda\right)\right) + \mathrm{Tr}\left(\mathbf{X}\right) - \frac{1}{4\mu}\mathrm{Tr}\left(\mathbf{X}\lambda\lambda^T\right) =$$

$$= -\mu + \mathrm{Tr}\left(\mathbf{X}\mathrm{diag}\left(\lambda\right)\right) + \mathrm{Tr}\left(\mathbf{X}\right) - \frac{1}{4\mu}\lambda^T\mathbf{X}\lambda =$$

If we fix the same notation from before of Q_i and e_i, we get that

$$\mathrm{Tr}\left(\mathbf{X}\mathrm{diag}\left(\lambda\right)\right) = \sum_{i\in V}\lambda_i \mathrm{Tr}\left(Q_i\mathbf{X}\right)$$

Thus, if we compute the gradient respect to λ of L

$$\nabla_\lambda L = \sum_{i\in V}\mathrm{Tr}\left(Q_i\mathbf{X}\right)e_i - \frac{1}{2\mu}\mathbf{X}\lambda = 0 \Longleftrightarrow \sum_{i\in V}\mathrm{Tr}\left(Q_i\mathbf{X}\right)e_i = \frac{1}{2\mu}\mathbf{X}\lambda$$

If now we set $\mathbf{x} = \frac{1}{2\mu}\mathbf{X}\lambda$, we get that the previous condition becomes $\forall i \in V$

$$\mathrm{Tr}\left(Q_i\mathbf{X}\right) = e_i^T\mathbf{x}$$

The lagrangian can be rewritten as follows

$$= -\mu + \mathrm{Tr}\left(\mathbf{X}\right) + \lambda^T\left(\sum_{i\in V}\lambda_i\mathrm{Tr}\left(Q_i\mathbf{X}\right)e_i - \frac{1}{4\mu}\mathbf{X}\lambda\right) = -\mu + \mathrm{Tr}\left(\mathbf{X}\right) + \frac{1}{4\mu}\lambda^T\mathbf{X}\lambda =$$

$$= \mathrm{Tr}\left(\mathbf{X}\right) - \mu\left(1 - \frac{1}{4\mu^2}\lambda^T\mathbf{X}\lambda\right) =$$

Using the definition of $\mathbf{x}$ given before, we get

$$= \mathrm{Tr}\left(\mathbf{X}\right) - \mu\left(1 - \mathbf{x}^T\mathbf{X}^{-1}\mathbf{x}\right)$$

If now we compute the $\inf_{\mu\in\mathbb{R}}$ of this expression we get

$$\inf_{\mu\in\mathbb{R}} L = \begin{cases} 0 & 1 - \mathbf{x}^T\mathbf{X}^{-1}\mathbf{x} = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Thus the dual problem of the D-QCQP is

$$\begin{array}{ll} & \underset{\mathbf{X}\in\mathbb{S}^n,\mathbf{x}\in\mathbb{R}^n}{\text{maximize}}\ \mathrm{Tr}\left(\mathbf{X}\right) \\ & \mathrm{Tr}\left(Q_i\mathbf{X}\right) - e_i^T\mathbf{x} = 0 \quad \forall i \in V \\ \text{subject to} & \mathbf{X}_{i,j} = 0 \qquad\qquad\qquad \forall (i,j)\in E \\ & 1 - \mathbf{x}^T\mathbf{X}^{-1}\mathbf{x} = 0 \end{array}$$

If we had the condition $1 - \mathbf{x}^T\mathbf{X}^{-1}\mathbf{x} \geq 0$, we would have finished through Schur's lemma, the problem is that with this constraint the $\inf_{\mu\in\mathbb{R}}$ would be $-\infty$.

I don't know how to weaken the condition to get the desired one.

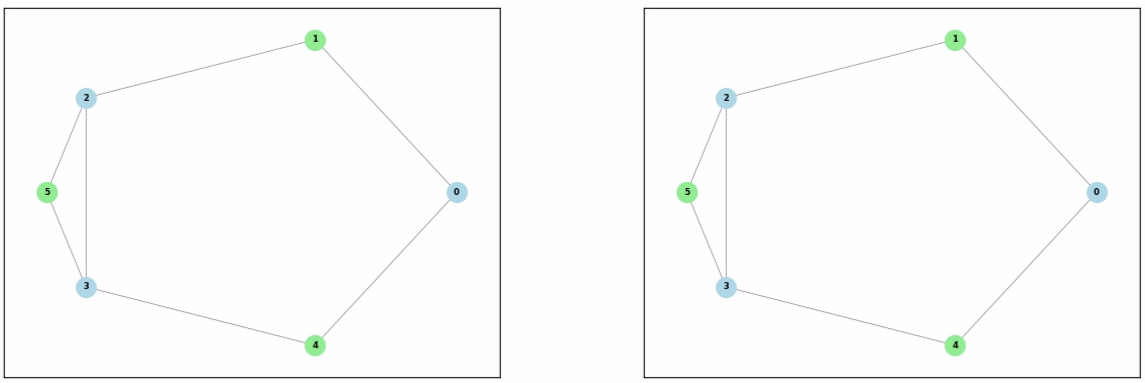
We can see that strong duality holds because given that:

- D-QCQP is the dual problem of SDP
- We proved that strong duality holds for SDP
- We proved that the dual problem of D-QCQP is SDP

We get that the duality gap for the D-QCQP is the same of the SDP one, so strong duality holds for D-QCQP.

Problem 5

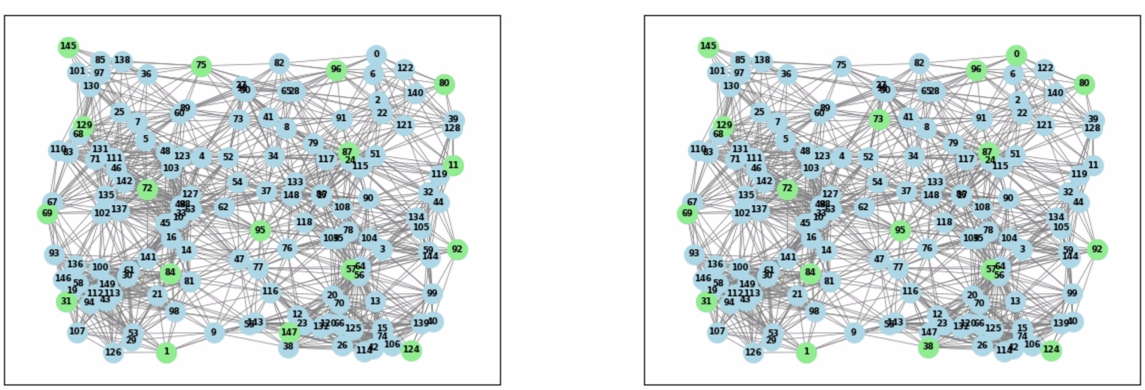
The computation of the first set of point gave back the following result.



On the left the rounding heuristic and on the right the IP solution.
Numerically the algorithm output these values in these times:

1. Optimal value of IP is between 3 and 3 with Total time SDP: 0.07722783088684082 s
2. Optimal IP value 3 with Total time IP: 0.060965776443481445 s

The computation of the second set of point gave back the following result.



On the left the rounding heuristic and on the right the IP solution.
Numerically the algorithm output these values in these times:

1. Optimal value of IP is between 17 and 17 with Total time SDP: 56.19373297691345 s
2. Optimal IP value 17 with Total time IP: 402.842796087265 s