

1 Introduction and relevant background

1.1 DAGs and their algebraic properties

1.2 Tensoring Graphical models

Let $\mathcal{G}_1, \dots, \mathcal{G}_n$ be DAGs of vertex sizes m_1, \dots, m_n respectively. Take

$$\mathcal{M}_{\mathcal{G}_i} := \{(\mathbf{1} - \Lambda)^\top \Omega^{-1} (\mathbf{1} - \Lambda) : \Lambda, \Omega \in \mathbb{R}^{m_1 \times m_i}, \Lambda_{ij} \neq 0 \implies j \rightarrow i, \Omega \text{ diagonal and PD}\}.$$

Note that taking $\Omega = \mathbf{1}, \Lambda = \mathbf{0}$ forces $\mathbf{1} \in \mathcal{M}_{\mathcal{G}_i}$. Look at the model (concentration matrices) $M := \{\Psi_1 \otimes \dots \otimes \Psi_n : \Psi_i \in \mathcal{M}_{\mathcal{G}_i}\}$ where \otimes stands for the Kronecker product for matrices (which is associative). Also for a group G of matrices, define $\mathcal{M}_G := \{g^\top g : g \in G\}$.

The question we want to ask is: Is $M = \mathcal{M}_G$ for some group G ? Trying to answer this question is not easy, but thinking about it led me to writing up this article.

To begin with, define the group for the tensor normal model for a collection of DAGs \mathcal{G}_i as $G(\mathcal{G}_1, \dots, \mathcal{G}_n) := \{\Psi_1 \otimes \dots \otimes \Psi_n : \Psi_i \in G(\mathcal{G}_i)\} = \bigotimes_{i=1}^n G(\mathcal{G}_i)$. Recall that the set of matrices corresponding to a single DAG \mathcal{G} is

$$G(\mathcal{G}) = \{X \in GL_{|V(\mathcal{G})|} : X_{ij} = 0 \text{ for } i \neq j \text{ with } j \not\rightarrow i \text{ in } \mathcal{G}\}.$$

Lemma 1

$G(\mathcal{G}_1, \mathcal{G}_2)$ is a group iff both \mathcal{G}_i are TDAGs. If both \mathcal{G}_i 's are TDAGs, the model $M = \{\Psi_1 \otimes \Psi_2 : \Psi_i \in \mathcal{M}_{\mathcal{G}_i}\}$ is exactly $\mathcal{M}_{G(\mathcal{G}_1, \mathcal{G}_2)}$.

Proof. Suppose \mathcal{G}_1 is not a TDAG. Then (after reordering), $3 \rightarrow 2, 2 \rightarrow 1$ but $3 \not\rightarrow 1$ in \mathcal{G}_1 . Take the elementary matrix E_{12} which has a 1 on the diagonal entries and at the $(1, 2)$ position, 0 everywhere else. Also look at E_{23} . Consider $g = E_{12} \otimes \mathbf{1}, h = E_{23} \otimes \mathbf{1}$. Then $gh = E_{12}E_{23} \otimes \mathbf{1}$. This matrix has a 1 in the $(1, 2m_2 + 1)$ position. If this were in $G(\mathcal{G}_1 \otimes \mathcal{G}_2)$ then there are matrices $X \in G(\mathcal{G}_1), Y \in G(\mathcal{G}_2)$ such that $x_{13}y_{11} = 1$. This is impossible because $3 \not\rightarrow 1$ in \mathcal{G}_1 whence $x_{13} = 0$. So $g, h \in G(\mathcal{G}_1 \otimes \mathcal{G}_2)$ but $gh \notin G(\mathcal{G}_1 \otimes \mathcal{G}_2)$. A similar argument can be used if \mathcal{G}_2 is not a TDAG.

Suppose \mathcal{G}_i are both TDAG's. Then we know that $\mathcal{M}_{\mathcal{G}_i}$ are both groups given by $\mathcal{M}_{\mathcal{G}_i} = \mathcal{M}_{G(\mathcal{G}_i)}$ where $G_i := G(\mathcal{G}_i) = \{X \in GL_{m_i} : X_{pq} = 0 \text{ for } p \neq q \text{ with } q \not\rightarrow p \text{ in } \mathcal{G}_i\}$. So let $G = \{X_1 \otimes X_2 : X_i \in G_i\}$. This is clearly a group. We will show that $M = \mathcal{M}_G$ for this group G . Indeed, if $X \in M$ then $X = \Psi_1 \otimes \Psi_2$ for some $\Psi_i \in \mathcal{M}_{\mathcal{G}_i}$. But then $\Psi_i = X_i^\top X_i$ for some $X_i \in G_i$. This means $X = \Psi_1 \otimes \Psi_2 = (X_1^\top X_1) \otimes (X_2^\top X_2) = (X_1 \otimes X_2)^\top (X_1 \otimes X_2) \in \mathcal{M}_G$ because $X_i \in G_i$. Note that these steps are reversible. This shows that $M = \mathcal{M}_G$. \blacksquare

2 (Trying to) Reduce to single DAG case

While thinking about questions on $G(\mathcal{G}_1) \otimes G(\mathcal{G}_2)$, I got ‘lazy’ and tried to figure out the following question in order to make my life easy.

Question 1 (answered)

Given DAGs $\mathcal{G}_1, \mathcal{G}_2$, is there a DAG \mathcal{G} such that $G(\mathcal{G}_1) \otimes G(\mathcal{G}_2) = G(\mathcal{G})$? What if $\mathcal{G}_1, \mathcal{G}_2$ are transitive DAGs?

Let’s start thinking about what happens if the answer to the question is ‘YES’ in the hope that the nonzero matrix components in the RHS set will give us the edges of \mathcal{G} . For every $G(\mathcal{G})$, there is a distinguished element $\mathfrak{M}(\mathcal{G})$ where $\mathfrak{M}(\mathcal{G})_{ij} = \mathbf{1}_{j \xrightarrow{\equiv} i}$ where $\xrightarrow{\equiv}$ is defined later in definition 2. This element gives all the edges of the DAG – these edges precisely correspond to the off-diagonal entries. So tensoring $\mathfrak{M}(\mathcal{G}_1)$ and $\mathfrak{M}(\mathcal{G}_2)$ will give the off-edges of \mathcal{G} . Denote $m_i = |V(\mathcal{G}_i)|$ and $e_i = |E(\mathcal{G}_i)|$.

The construction has been made explicit in section 3 and it can be seen that such \mathcal{G} does not exist. (Suggestion: look at the next section and come back to this argument). Let’s just focus on the TDAG case now. It is clear that if such \mathcal{G} exists, then it has to have $m_1 m_2$ vertices and at least $e_1 e_2$ edges. There is a map $G(\mathcal{G}_1) \otimes G(\mathcal{G}_2) \rightarrow G(\mathcal{G})$, which is not surjective. That is because the dimension of the right-hand side group is $(m_1 + e_1)(m_2 + e_2)$ whereas the image has dimension $m_1 + e_1 + m_2 + e_2 - 1$ which is much smaller for nontrivial DAGs \mathcal{G}_i .

This motivates the question

Question 2

Given TDAGs $\mathcal{G}_1, \mathcal{G}_2$, is there a TDAG \mathcal{G} such that $G(\mathcal{G}_1) \otimes G(\mathcal{G}_2) \cong G(\mathcal{G})$, where the isomorphism is in the category of groups?

3 Tensor of two DAGs

3.1 Algebraic definition

Start with a DAG \mathcal{G} and label the vertices so that $i \not\rightarrow j$ if $i < j$. In literature, this is referred to as a topological ordering, and every DAG has one. Consider the special matrix $\chi = \mathfrak{M}_m(\mathcal{G})$ given by $\chi_{ii} = 1, \chi_{ij} = \mathbf{1}_{j \rightarrow i}$ for $i \neq j$, that is, it’s 1 if $j \rightarrow i$ is an edge, and otherwise 0. Note that this is an upper triangular matrix because $j < i \implies \chi_{ij} = \mathbf{1}_{j \rightarrow i} = 0$. So $\det \chi = 1$. Further if $i \neq j$ and $j \not\rightarrow i$ in \mathcal{G} then $\chi_{ij} = 0$. This shows that $\chi \in \mathcal{M}_{\mathcal{G}}$.

Note, given an $m \times m$ upper triangular matrix χ with entries in $\{0, 1\}$ and diagonal entries 1, we can define a graph $\mathcal{G} = \mathfrak{G}_m(\chi)$ whose vertex set is $\{1, \dots, m\}$ and there is an edge $i \rightarrow j$ iff $\chi_{ji} = 1$. \mathcal{G} has no directed cycles. Indeed, if there were

a directed cycle $i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_k \rightarrow i_1$, then by the upper triangularity of χ and construction of \mathcal{G} , we get that $i_1 > i_2 > \cdots > i_k > i_1$ which is impossible. This means \mathcal{G} is a DAG.

Thus we've established a 1 – 1 association between the set of DAGs of size m and the set of $m \times m$ upper triangular matrices with entries in $\{0, 1\}$ and diagonal entries 1.

$$\{\text{Topologically ordered DAGs of size } m\} \xrightleftharpoons[\mathfrak{G}_m]{\mathfrak{M}_m} \left\{ \begin{array}{l} m \times m \text{ upper triangular} \\ \text{matrices with entries in } \{0, 1\} \\ \text{and diagonal entries 1} \end{array} \right\}$$

When the size m of the DAG or the matrix is understood we suppress the subscript and simply write $\mathfrak{G}, \mathfrak{M}$. Now start with two topologically ordered DAGs $\mathcal{G}_1, \mathcal{G}_2$. Define $\mathcal{G}_1 \otimes \mathcal{G}_2 := \mathfrak{G}(\mathfrak{M}(\mathcal{G}_1) \otimes \mathfrak{M}(\mathcal{G}_2))$ where \otimes refers to the Kronecker product of matrices. Now one might ask the question: is $G(\mathcal{G}_1) \otimes G(\mathcal{G}_2) = G(\mathcal{G}_1 \otimes \mathcal{G}_2)$? The answer is negative, though one of them is a subset of the other. The proof is presented quite elegantly with a combinatorial description.

3.2 A purely combinatorial description

Recall that the Kronecker product of matrices A, B are given by $(A \otimes B)_{(i,j),(k,l)} = A_{ik}B_{jl}$. In our above description $\mathfrak{M}(\mathcal{G}_1 \otimes \mathcal{G}_2) = \mathfrak{M}(\mathcal{G}_1) \otimes \mathfrak{M}(\mathcal{G}_2)$. The rows of $\mathfrak{M}(\mathcal{G}_1 \otimes \mathcal{G}_2)$ are indexed by (i_1, i_2) and its columns by (j_1, j_2) where $i_k, j_k \in V(\mathcal{G}_k)$. This suggests that the vertex set of $\mathcal{G}_1 \otimes \mathcal{G}_2$ should be $V(\mathcal{G}_1) \times V(\mathcal{G}_2)$. To determine whether there is an edge $(j_1, j_2) \rightarrow (i_1, i_2) \iff \mathfrak{M}(\mathcal{G}_1)_{i_1 j_1} = \mathfrak{M}(\mathcal{G}_2)_{i_2 j_2} = 1$ and $(i_1, i_2) \neq (j_1, j_2) \iff$ one of the following holds:

- $j_1 \rightarrow i_1$ is an edge in \mathcal{G}_1 and $j_2 = i_2$ in \mathcal{G}_2 .
- $j_1 = i_1$ in \mathcal{G}_1 and $j_2 \rightarrow i_2$ is an edge in \mathcal{G}_2 .
- $j_1 \rightarrow i_1$ is an edge in \mathcal{G}_1 and $j_2 \rightarrow i_2$ is an edge in \mathcal{G}_2 .

We will introduce a notation.

Definition 2

Let $\bar{\rightarrow}$ be the union of $=$ and \rightarrow in a DAG \mathcal{G} .

In other words $i \bar{\rightarrow} j$ means that either $i = j$ or $i \rightarrow j$. Note that if \mathcal{G} is a transitive DAG then $i \bar{\rightarrow} j \bar{\rightarrow} k \implies i \bar{\rightarrow} k$. In this new notation, $(u, v) \rightarrow (u', v')$ is an edge in $\mathcal{G}_1 \otimes \mathcal{G}_2$ iff $u \bar{\rightarrow} v$ in \mathcal{G}_1 and $u' \bar{\rightarrow} v'$ in \mathcal{G}_2 but $(u, v) \neq (u', v')$.

With this notation and description we can more elegantly prove that

Lemma 3

$$G(\mathcal{G}_1) \otimes G(\mathcal{G}_2) \subseteq G(\mathcal{G}_1 \otimes \mathcal{G}_2)$$

Proof. Take any $\Psi_1 \otimes \Psi_2 \in G(\mathcal{G}_1) \otimes G(\mathcal{G}_2)$. Suppose $(i_1, i_2) \neq (j_1, j_2)$ are distinct vertices in $\mathcal{G}_1 \otimes \mathcal{G}_2$ such that $(\Psi_1 \otimes \Psi_2)_{(i_1, i_2), (j_1, j_2)} \neq 0$. But $0 \neq (\Psi_1 \otimes \Psi_2)_{(i_1, i_2), (j_1, j_2)} = (\Psi_1)_{i_1 j_1} (\Psi_2)_{i_2 j_2} \implies (\Psi_1)_{i_1 j_1} \neq 0 \neq (\Psi_2)_{i_2 j_2} \implies j_1 \xrightarrow{=} i_1$ and $j_2 \xrightarrow{=} i_2 \implies (j_1, j_2) \rightarrow (i_1, i_2) \implies \Psi_1 \otimes \Psi_2 \in G(\mathcal{G}_1 \otimes \mathcal{G}_2)$. ■

Remark 4

Note that in the above lemma, it is not necessary that the collections of matrices described are groups. The inclusion is merely an inclusion of sets.

Lemma 5

\mathcal{G}_1 and \mathcal{G}_2 are TDAGs iff $\mathcal{G}_1 \otimes \mathcal{G}_2$ is a TDAG.

Proof. Say \mathcal{G}_1 and \mathcal{G}_2 are TDAGs. Say there are edges $(i_1, j_1) \rightarrow (i_2, j_2) \rightarrow (i_3, j_3)$ in $\mathcal{G}_1 \otimes \mathcal{G}_2$. Then $i_1 \xrightarrow{=} i_2 \xrightarrow{=} i_3$ and $j_1 \xrightarrow{=} j_2 \xrightarrow{=} j_3$. Since $\mathcal{G}_1, \mathcal{G}_2$ are TDAGs, $i_1 \xrightarrow{=} i_3$ and $j_1 \xrightarrow{=} j_3$. This implies that $(i_1, j_1) \xrightarrow{=} (i_3, j_3)$. We will show that these two are unequal. If $i_1 = i_3$ then $i_1 \rightarrow i_2$ can't be an edge (if it were, then we would have a cycle $i_1 \rightarrow i_2 \rightarrow i_3 = i_1$), so $i_1 = i_2 = i_3$, and in this case we need $j_1 \rightarrow j_2 \rightarrow j_3$ to be real edges which gives the real edge $j_1 \rightarrow j_3$.

Conversely, assume $\mathcal{G}_1 \otimes \mathcal{G}_2$ is a TDAG. Suppose $i_1 \rightarrow i_2 \rightarrow i_3$ are edges in \mathcal{G}_1 . Pick a vertex j in \mathcal{G}_2 . Then there are edges $(i_1, j) \rightarrow (i_2, j) \rightarrow (i_3, j)$ in $\mathcal{G}_1 \otimes \mathcal{G}_2$. Since $\mathcal{G}_1 \otimes \mathcal{G}_2$ is a TDAG, we have an edge $(i_1, j) \rightarrow (i_3, j)$. By definition, $i_1 \rightarrow i_3$ is an edge. Similar argument proves that \mathcal{G}_2 is a TDAG. ■

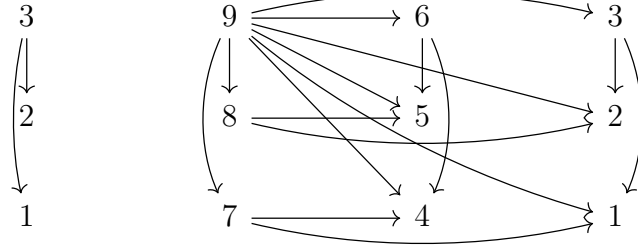
Corollary 6

$G(\mathcal{G}_1) \otimes G(\mathcal{G}_2)$ is a group iff $G(\mathcal{G}_1 \otimes \mathcal{G}_2)$ is a group.

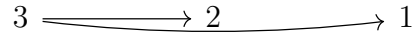
Proof. Use the fact that $G(\mathcal{G})$ is a group iff \mathcal{G} is a TDAG along with lemma 1 and lemma 5. ■

Example 7

\mathcal{G}_2



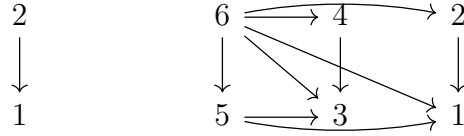
\otimes



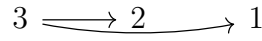
\mathcal{G}_1

Example 8

\mathcal{G}_2



\otimes



\mathcal{G}_1

While thinking about “when are $G(\mathcal{G}_1) \otimes G(\mathcal{G}_2)$ Gaussian group models?”, I asked

Question 3

Is it possible to find a group G such that $\mathcal{M}_{\mathcal{G}} = \mathcal{M}_G$ if \mathcal{G} is a DAG but not TDAG?

That is, I was trying to characterize all DAGs such that their models are Gaussian group models.

3.3 Properties

Define the function $\delta(\mathcal{G}) = \text{in-deg}(\mathcal{G}) + 1$ where $\text{in-deg}(\mathcal{G})$ means $\max_{i \in I} (\#\{\text{parents of } i\})$. It has been proven that $\text{mlt}(\mathcal{G}) = \text{in-deg}(\mathcal{G}) + 1 = \delta(\mathcal{G})$. We will show that

Lemma 9

$$\delta(\mathcal{G}_1 \otimes \mathcal{G}_2) = \delta(\mathcal{G}_1)\delta(\mathcal{G}_2).$$

Proof. Denote by $p(v)$ the number of parents of vertex v . We will abuse notation and use the same p for $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_1 \otimes \mathcal{G}_2$.

It is not hard to see that $p((i, j)) = p(i) + p(j) + p(i)p(j) \implies p((i, j)) + 1 = (p(i) + 1)(p(j) + 1)$. This directly proves what we had wanted. \blacksquare

It easily follows that $\text{mlt}(\mathcal{G}_1 \otimes \mathcal{G}_2) = \delta(\mathcal{G}_1 \otimes \mathcal{G}_2) = \delta(\mathcal{G}_1)\delta(\mathcal{G}_2) = \text{mlt}(\mathcal{G}_1)\text{mlt}(\mathcal{G}_2)$.

Next, we can easily see that $\mathcal{G}_1 \otimes \mathcal{G}_2$ always has an unshielded collider as long as each \mathcal{G}_t has an edge $i_t \rightarrow j_t$. This is because $(i_1, j_2) \rightarrow (j_1, j_2) \leftarrow (j_1, i_2)$ but no edge between (i_1, j_2) and (j_1, i_2) .

Corollary 10

Let $\mathcal{G}_1, \dots, \mathcal{G}_m$ be TDAGs and $\mathcal{G} = \mathcal{G}_1 \otimes \dots \otimes \mathcal{G}_m$.

1. Consider the action of $\left(\prod G(\mathcal{G}_i)_{\text{SL}}\right)^+$ on on tuples of $n \geq \prod \delta(\mathcal{G}_i)$ samples. Then the null cone is not Zariski closed.
2. The undirected graphical model obtained from \mathcal{G} is not a Gaussian group model.

4 A computational problem

We focus on the object $\mathcal{G}_1 \otimes \mathcal{G}_2$ for this section. Given $\mathcal{G}_1, \mathcal{G}_2$ one can easily construct $\mathcal{G}_1 \otimes \mathcal{G}_2$ in $\mathcal{O}((m_1 + e_1)(m_2 + e_2))$ time. We ask the inverse question:

Question 4

Given a DAG \mathcal{G} , is it possible to decompose it as $\mathcal{G}_1 \otimes \mathcal{G}_2$ for nontrivial \mathcal{G}_i ?

This problem is clearly at least as hard as integer factoring. Indeed, if we could find such a decomposition, then clearly $|V(\mathcal{G})| = |V(\mathcal{G}_1)| \cdot |V(\mathcal{G}_2)|$.

Going the other way, suppose we are given a graph \mathcal{G} and we want to determine whether it is a tensor of two graphs. Find all nontrivial factors of $|V(\mathcal{G})|$, call this set S . Do the following for each $d \in S$. Find all sub-blocks of size $d \times d$ with top-left corner $(1 + id, 1 + jd)$ ($i, j \in \mathbb{Z}_{\geq 0}$) in the matrix $\mathcal{M}(\mathcal{G})$ where at least one entry is non-zero, and check if all such blocks are same. If one such d is found, then \mathcal{G} can be decomposed as a tensor product.

5 An approximation problem

We want to ask how “dense” is the class of tensored DAGs. That is given two integers m_1, m_2 and a graph G with $m_1 m_2$ vertices, we want to solve the optimization problem

$$\inf\{d(G, \mathcal{G}_1 \otimes \mathcal{G}_2) : |V(\mathcal{G}_i)| = m_i\}$$

that is, finding the closest approximation of G by tensor of two graphs of the given sizes. Of course, what the distance d means must be made clear.

As a first attempt, one such distance could be defined to be $d(G, H) = \|A(G) - A(H)\|$ for some choice of norm, where $A(\cdot)$ stands for the adjacency matrix. After playing around a bit, one immediately realizes that this depends on the labelling of vertices. But we would obviously want the metric to be invariant under graph isomorphisms. For example the Frobenius norm d_F (under our definition) of the graphs

$$1 \longrightarrow 2 \qquad 2 \longrightarrow 1$$

is $\sqrt{2}$, but we would really want them to be 0. So one way to correct this would be to redefine the metric as

$$d_{\|\cdot\|}(G, H) := \inf_{\substack{\text{relabellings} \\ \text{of } G}} \|A(G) - A(H)\|$$

which is the same as

$$d_{\|\cdot\|}(G, H) = \inf_{\substack{\text{permutation} \\ \text{matrices } P}} \|PA(G)P^{-1} - A(H)\|$$

where $\|\cdot\|$ is a norm invariant under conjugation by permutation matrices, that is, $\|X\|$ is such that $\|PXP^{-1}\| = \|X\|$ for every matrix X and every permutation matrix P . Such a norm $\|\cdot\| : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}_{\geq 0}$ will be referred to as *symmetry invariant* (abbr.: SI). Moreover, we will call a norm *strongly left-symmetry invariant* (*strongly right-symmetry invariant* resp.) if $\|PX\| = \|X\|$ ($\|XP\| = \|X\|$ resp.) for all matrices X and permutation matrices P . Abbreviate them as SLSI and SRSI respectively. Note that a norm which is both SLSI and SRSI is clearly SI.

To get a large class of examples, we note that

Proposition 11

$\|\cdot\|_p$ is SLSI and SRSI $\forall p \in [1, \infty]$, and thus SI.

Proof. It's well known that $\|XY\|_p \leq \|X\|_p \|Y\|_p$. Let Q be a permutation matrix. It is known that $\|Q\|_p = 1$.

Then $\|X\|_p \leq \|XQ\|_p \cdot \|Q^{-1}\|_p = \|XQ\|_p \leq \|XQQ^{-1}\|_p \|Q\|_p = \|X\|_p$.

Similarly $\|X\|_p \leq \|Q^{-1}X\|_p \cdot \|Q\|_p = \|Q^{-1}X\|_p \leq \|Q^{-1}QX\|_p \|Q\|_p = \|X\|_p$. ■

Question 5

Find some non-SI matrix norm $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}$.

One might also note that the Frobenius norm, defined as $\|X\|_F = \sqrt{\sum_{i,j} X_{i,j}^2}$, is also both SLSI and SRSI, hence SI. This means that all $d_x := d_{\|\cdot\|_x}$, for $x \in [1, \infty] \cup \{F\}$, are valid metrics on graphs.

Fix an SI norm $\|\cdot\|$. We are yet to prove a few things. For example, it's not clear that $d_{\|\cdot\|}$ is invariant under isomorphism of the graph in the second component. From this, one might feel that

$$d'_{\|\cdot\|}(G, H) := \inf_{\substack{\text{permutation} \\ \text{matrices } P, Q}} \|PA(G)P^{-1} - QA(H)Q^{-1}\|$$

'seems' to be more fit to our situation better as opposed to how the metric is really defined.

To resolve these issues, observe that

$$\begin{aligned} \|PA(G)P^{-1} - Q^{-1}A(H)Q\| &= \|Q^{-1}((QP)A(G)(QP)^{-1} - A(H))Q\| \\ &= \|(QP)A(G)(QP)^{-1} - A(H)\|. \end{aligned}$$

This proves that if H' is a graph isomorphic to H , that is $A(H') = Q^{-1}A(H)Q$ for some permutation matrix Q , then

$$\begin{aligned} d_{\|\cdot\|}(G, H') &= \inf_{\substack{\text{permutation} \\ \text{matrices } P}} \|PA(G)P^{-1} - Q^{-1}A(H)Q\| \\ &= \inf_{\substack{\text{permutation} \\ \text{matrices } P}} \|(QP)A(G)(QP)^{-1} - A(H)\| \\ &= d_{\|\cdot\|}(G, H). \end{aligned}$$

The same observation shows that $d_{\|\cdot\|} = d'_{\|\cdot\|}$ and also proves symmetry of $d_{\|\cdot\|}$.

6 Continuous constrained optimization

The distance we're trying to find above involves minimizing a function as the inputs vary over a *discrete* space. Explicitly, this is the space $\mathbb{D}_{m_1} \times \mathbb{D}_{m_2}$ where \mathbb{D}_n is the set of all DAGs with n vertices. It's a huge space and thus computationally expensive. Let's look at examples which shows us how considering a continuous version of a problem reduces the search space.

6.1 A high-school math problem

We want to consider the problem of finding $\sup_{n \in \mathbb{N}} n^{1/n}$. Well, the search space here is not just ‘large’, it’s infinite. Maximizing $n^{1/n}$ is same as maximizing $\frac{\ln n}{n}$ and intuition tells us that n dominates $\ln n$ for large n , so we could as well reduce the search space to something like $n \leq 100$. But that’s not a good justification. Also this ‘dominating’ feature is a more familiar feature of the functions x and $\ln x$ for $x \in \mathbb{R}_{>0}$ rather than $x \in \mathbb{N}$. So let’s just try to optimize the function $\frac{\ln x}{x}$ where the search space is $x \in \mathbb{R}_{>0}$. This is now a continuous problem. In fact this function is differentiable with derivative $\frac{1-\ln x}{x^2}$. Clearly this derivative is positive for $x < e$ and negative for $x > e$, and has a flat tangent at $x = e$, which is clearly our maxima. But the problem is not yet over. We wanted to optimize over \mathbb{N} , not $\mathbb{R}_{>0}$. This is easy because we know that the function decreases as we go left from e and also decreases as we go right from e . So we simply need to compare $2^{1/2}$ and $3^{1/3}$ because $2 < e < 3$. That gives our answer to be $n = 3$ for which the supremum is $3^{1/3}$.

Note that in the above example, we reduced the search space from \mathbb{N} to $\{2, 3\}$.

6.2 A max-flow problem

Consider a directed graph $G = (V, E)$ with n vertices and m edges. Fix two distinct arbitrary vertices $s \neq t \in V$ which are to be thought of as *source* and *target* vertices respectively. Additionally s, t satisfy that there is no edge going into s and no edge going out of t . An $s - t$ flow is an assignment $\mathbf{f} : E \rightarrow \mathbb{R}$ of weights to each edge satisfying $|\mathbf{f}_e| \leq 1 \forall e \in E$ (that is, we set *capacity* of each edge to be 1) and Kirchoff’s law:

$$\sum_{\{j \in V: j \rightarrow v\}} \mathbf{f}_{j \rightarrow v} = \sum_{\{j \in V: v \rightarrow j\}} \mathbf{f}_{v \rightarrow j} \text{ for each } v \in V \setminus \{s, t\},$$

that is, the incoming flow is equal to outgoing flow at each vertex i . Consider its incidence matrix $H \in M_{n \times m}(\mathbb{R})$ where the $(i, e)^{\text{th}}$ entry is 1 if e goes into i , -1 if e comes out of i and 0 otherwise. One can think of \mathbf{f} as an element of $\mathbb{R}^m \cong \mathbb{R}^E$, so that the capacity condition is equivalent to $\|\mathbf{f}\|_\infty \leq 1$ and Kirchoff’s law is equivalent to $\langle \mathbf{e}_i, H\mathbf{f} \rangle = 0 \forall i$ where $\mathbf{e}_i \in \mathbb{R}^n$ is the i^{th} standard basis vector for $1 \leq i \leq n$. The problem is to maximize $\sum_{\{j \in V: s \rightarrow j\}} \mathbf{f}_{s \rightarrow j} = \langle \mathbf{e}_s, H\mathbf{f} \rangle = -\langle \mathbf{e}_t, H\mathbf{f} \rangle$,

the flow out of s , given the constraints (capacity constraint and Kirchoff’s law). On a side note, a special case of this problem is the problem of finding a perfect matching in a bipartite graph.

Suppose we are given a flow value F , note that an $s - t$ flow with flow value F (that is, the total flow coming out of s) looks like a vector $\mathbf{f} \in \mathbb{R}^E$ satisfying $H\mathbf{f} = F(\mathbf{e}_s - \mathbf{e}_t)$. For the flow to respect the capacity, we additionally need $|\mathbf{f}_i| \leq 1$. So a *feasible flow* is a vector in the intersection of the two sets $X_F :=$

$\{\mathbf{f} \in \mathbb{R}^E : H\mathbf{f} = F(\mathbf{e}_s - \mathbf{e}_t)\}$ and $B_\infty := B_\infty(0, 1) = \{\mathbf{f} \in \mathbb{R}^E : \|\mathbf{f}\|_\infty \leq 1\}$. It is easy to see that both of these are convex sets. Indeed B_∞ is a unit ball and a polytope, and if $\mathbf{f}, \pm \mathbf{f}' \in X_F$ then any convex linear combination (say weights λ, μ) satisfy $H(\lambda\mathbf{f} + \mu\mathbf{f}') = (\lambda + \mu)F(\mathbf{e}_s - \mathbf{e}_t) = F(\mathbf{e}_s - \mathbf{e}_t)$ whence $\lambda\mathbf{f} + \mu\mathbf{f}' \in X_F$. Now this problem about finding the intersection of two sets (embedded in Euclidean space) can be converted into a convex optimization problem. Let's discuss this a bit in general framework.

Let $S, T \subseteq \mathbb{R}^d$ be convex and closed. We want to determine a point in $S \cap T$, or whether it's empty. A different way to think about this is to find a point $\mathbf{x} \in S$ that minimizes $d(\mathbf{x}, T) = \inf_{\mathbf{y} \in T} d(\mathbf{x}, \mathbf{y})$.

Now, we want to actually find points for which this minimum distance is attained, that is, we want existence of $\mathbf{x} \in S, \mathbf{y} \in T$ so that the above minimum is $d(\mathbf{x}, \mathbf{y})$. So why does that not work? Indeed, $S = \mathbb{R} \times \{0\}, T = \{(u, v) \in \mathbb{R}^2 : v \geq 1/u, u > 0\}$ is an example of closed convex sets such $d(S, T) = 0$ but $S \cap T = \varnothing$. However, if one of S, T is compact, we can guarantee that $d(S, T) = 0 \implies S \cap T \neq \varnothing$. Let's prove this in the following lemma:

Lemma 12

Let X be a metric space and $A, B \subseteq X$ be disjoint subsets such that A is closed and B is compact. Then $d(A, B) = \inf_{x \in A, y \in B} d(x, y) > 0$.

Proof. For the sake of contradiction, assume there are sequences (x_n) in A and (y_n) in B such that $\lim_{n \rightarrow \infty} d(x_n, y_n) = 0$. B is compact so (y_n) has a convergent subsequence (y_{k_n}) , and say it converges to $y \in B$. Then $d(x_{k_n}, y) \leq d(x_{k_n}, y_{k_n}) + d(y_{k_n}, y)$ and both of the terms on the RHS can be made arbitrarily small. So $x_{k_n} \rightarrow y$, whence y is a limit point of the sequence (x_{k_n}) . But this is a sequence in A , whence $y \in A$. So $y \in A \cap B$, a contradiction. ■

So we now also assume that T is compact. This guarantees that the minimum distance is actually attained, and not merely an infimum.

So the problem becomes

$$\begin{aligned} & \min d(\mathbf{x}, T)^2 \\ & \text{subject to } \mathbf{x} \in S. \end{aligned}$$

We are using the squared (Euclidean) distance because it's smooth. The fact that S, T are convex sets and $d(\mathbf{x}, T)^2$ is a convex function makes this problem a convex program. We could, as well, exchange the roles of S, T and have the problem

$$\begin{aligned} & \min d(\mathbf{x}, S)^2 \\ & \text{subject to } \mathbf{x} \in T. \end{aligned}$$

7 Tensoring coloured DAGs

7.1 Introduction

A *colouring of a DAG* assigns colours to the vertices and edges. A *coloured DAG* is a tuple (\mathcal{G}, c) , where $\mathcal{G} = (I, E)$ is a DAG on vertices I and edges E , and $c : I \cup E \rightarrow C$ is a colouring of the vertices and edges. Vertex $i \in I$ has colour $c(i) \in C$, and edge $j \rightarrow i$ has colour $c(ij) \in C$. We sometimes denote the vertex colour $c(i)$ by $c(ii)$, with no ambiguity because a DAG cannot have loops. Corresponding to such coloured DAGs, there are RDAGs (Restricted DAGs), where the model on (\mathcal{G}, c) is parameterized by

$$A(\mathcal{G}, c) = \left\{ X \in \text{GL}_{|I|} \mid \begin{array}{l} j \not\rightarrow i \implies X_{ij} = 0 \text{ for } i \neq j \\ c(ij) = c(kl) \implies X_{ij} = X_{kl} \end{array} \right\}.$$

We will only focus on “good” colourings, where this means .

Definition 13

A colouring c of a directed graph \mathcal{G} is *compatible*, if:

- (a) $c(I) \cap c(E) = \emptyset$
- (b) If edges $j \rightarrow i$ and $l \rightarrow k$ have the same colour, then the child vertices i and k also have the same colour, i.e. $c(ij) = c(kl) \implies c(i) = c(k)$.

It has been proven that

Proposition 14

Fix a coloured DAG (\mathcal{G}, c) . The RDAG model on (\mathcal{G}, c) is equal to $\mathcal{M}_{A(\mathcal{G}, c)}$ if and only if the colouring c is compatible.

7.2 Tensoring

Given two coloured DAGs $(\mathcal{G}_1, c_1), (\mathcal{G}_2, c_2)$ we already have a way to find an underlying graph $\mathcal{G}_1 \otimes \mathcal{G}_2$ such they behave well wrt their specialized matrices, that is, $\mathfrak{M}(\mathcal{G}_1 \otimes \mathcal{G}_2) = \mathfrak{M}(\mathcal{G}_1) \otimes \mathfrak{M}(\mathcal{G}_2)$. In the coloured case, we will associate a different specialized matrix to every coloured DAG where the entries will be formal (non-commuting) variables.

Fix a colored DAG $(\mathcal{G} = (I, E), c)$. Let $C = c(I) \cup c(E)$. Our specialized matrix for this graph will be the matrix $\mathfrak{M}((\mathcal{G}, c))$ whose $(i, j)^{\text{th}}$ entry is $\mathbf{1}_{j \rightarrow i} \cdot c(ji)$, where the $c(ji)$ are non-commuting variables (with possible equalities). So our matrix is really an element of $k \langle C \rangle$. Now we define $(\mathcal{G}, c) = (\mathcal{G}_1, c_1) \otimes (\mathcal{G}_2, c_2)$ to be such that $\mathcal{M}((\mathcal{G}, c)) = \mathcal{M}(\mathcal{G}_1, c_1) \otimes \mathcal{M}(\mathcal{G}_2, c_2)$. Combinatorially, (\mathcal{G}, c) is described as follows:

- Vertices of \mathcal{G} are $I(\mathcal{G}) = I(\mathcal{G}_1) \times I(\mathcal{G}_2)$,

- There is an edge $(u, v) \rightarrow (u', v')$ iff $u \xrightarrow{=} u', v \xrightarrow{=} v'$ and $(u, v) \neq (u', v')$.
- The colour function is $c((u'v')(u, v)) = c_1(u'u) \cdot c_2(v'v)$.

Denote $\mathcal{G} = \mathcal{G}_1 \otimes \mathcal{G}_2, c = c_1 \otimes c_2$.

This is the analog of TDAGs for coloured DAGs because

Proposition 15

Let $(\mathcal{G}_1, c_1), (\mathcal{G}_2, c_2)$ be two nontrivial coloured DAGs with their colouring sets disjoint from each other. $c := c_1 \otimes c_2$ is a compatible colouring for $\mathcal{G} := \mathcal{G}_1 \otimes \mathcal{G}_2$ iff c_i is a compatible colouring for \mathcal{G}_i , for each $i = 1, 2$.

Proof. Note $c((i_1, i_2)(j_1, j_2)) = c((k_1, k_2)(l_1, l_2)) \iff c_1(i_1j_1)c_2(i_2j_2) = c_1(k_1l_1)c_2(k_2l_2) \iff c_1(i_1j_1) = c_1(k_1l_1) \text{ and } c_2(i_2j_2) = c_2(k_2l_2) \implies c_t(i_t) = c_t(k_t) \forall t = 1, 2$.

Suppose c is a compatible colouring for \mathcal{G} . Suppose $c_1(ij) = c_1(kl)$ in \mathcal{G}_1 . Let $y \rightarrow x$ be an edge in \mathcal{G}_2 . Then there are edges $(j, y) \rightarrow (i, x)$ and $(l, y) \rightarrow (k, x)$ in \mathcal{G} . Taking $i_1 = i, j_1 = j, k_1 = k, l_1 = l, i_2 = k_2 = x, j_2 = l_2 = y$ gives $c((i, x)(j, y)) = c((k, x)(l, y)) \implies c((i, x)) = c((k, x)) \implies c_1(i) = c_1(k)$. This shows that c_1 is compatible for \mathcal{G}_1 . A similar argument shows that c_2 is compatible for \mathcal{G}_2 .

Suppose c_t is compatible for \mathcal{G}_t , for each $t = 1, 2$. Say there are edges (or vertices, if both coordinates equal) $(j_1, j_2) \xrightarrow{=} (i_1, i_2)$ and $(l_1, l_2) \xrightarrow{=} (k_1, k_2)$ which have the same colour. Using the above calculation, we get $c_t(i_t) = c_t(k_t)$ for $t = 1, 2$. This means $c((i_1, i_2)) = c((k_1, k_2))$. ■

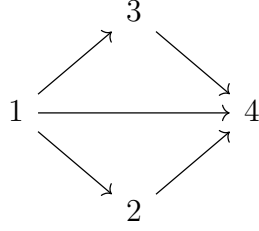
Remark 16

Similar questions can be asked for coloured DAGs, as we had asked for the non-coloured case.

Remark 17 (Places I see quivers arising (maybe forced thinking))1. Let \mathcal{G} be a DAG such that for any vertices i, j there is at most one path from i to j . Take the transitive closure $\overline{\mathcal{G}}$ of \mathcal{G} , that is, if there is a path $i_1 \rightarrow \dots \rightarrow i_k$, then connect i_1, i_k by an edge. Then $G(\overline{\mathcal{G}})$ is precisely the path algebra of the quiver \mathcal{G} .

I was thinking if one can "detransitivise" a TDAG \mathcal{G} (that is, find a quiver Q such that there is at most one path between any two vertices and $\overline{Q} = \mathcal{G}$) then we bring in quiver theory. But one can construct \mathcal{G} for which such Q does not

exist. In fact consider the following quiver \mathcal{G} :



But one cannot find a Q with the aforementioned property. One has to have $1 \rightarrow 3, 3 \rightarrow 4, 1 \rightarrow 2$ and $2 \rightarrow 4$ because absence of any of these in Q would result in no path from the respective sources to the targets in \overline{Q} . But then this forces two paths $1 \rightsquigarrow 4$, one through 2 and 3 each.

2. While constructing the colour map for tensor of two coloured DAGs, we had to invoke the algebra $k\langle X_1, \dots, X_n \rangle$. This is the path algebra of the quiver with one vertex and n loops on it.