TOTALLY NONNEGATIVE LOWER TRIANGULAR MATRICES AND THEIR PLANAR NETWORKS

ADRIAN PĂCURAR

Contents

1. Introduction

Definition 1. For an $n \times n$ matrix A, $I \subseteq \{1, 2, \ldots, n\}$, and $J \subseteq \{1, 2, \ldots, n\}$, let $A[I, J]$ be the submatrix of A with row set indexed by I and column set indexed by J.

Definition 2. The (I, J)-minor of matrix A, denoted by $A_{I,J}$, is the determinant

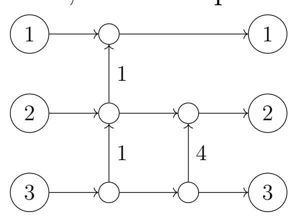
$$A_{I,J} = \det A[I, J]$$

where $|I| = |J|$.

Definition 3. A matrix A is totally nonnegative if minor of A is nonnegative. For the following matrix, one can easily verify that every minor is nonnegative:

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 5 & 1 \end{pmatrix}$$

Any $n \times n$ nonnegative matrix can be associated with a planar network with n sources and n sinks, where the (i, j)-entry of our matrix counts the number of weighted paths from source i to sink j. For the matrix above, one such planar network is



and one can easily verify that the number of paths from source i (the nodes on the left) to sink j (the nodes on the right) is given by the (i, j)-entry of our matrix.

Definition 4. For any planar network G, define the path matrix $(p_{i,j})$ of G by

$$p_{i,j} = \#\text{ of (weighted) paths from source } i \text{ to sink } j$$

where the number of rows is equal to the number of sources of G, and the number of columns is equal to the number of sinks of G.

Definition 5. A k-Path is a collection of k paths from a set of k sources to a set of k sinks.

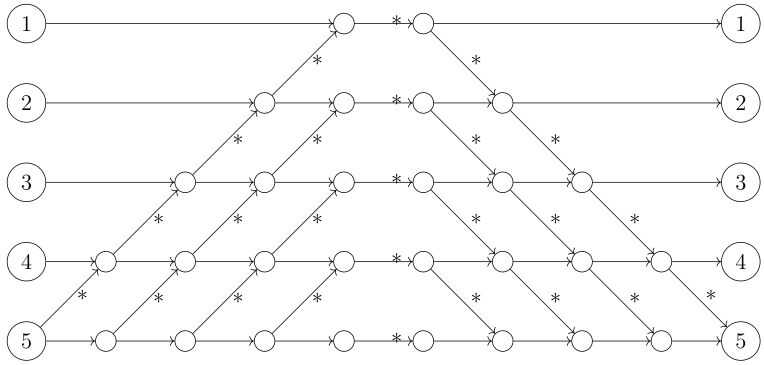
Theorem 1. (Lindström's Lemma) The path matrix A of any planar network is totally nonnegative. In particular, each minor $A_{I,J}$ is equal to the number of families of nonintersecting paths from sources indexed by I to sinks indexed by J. In other words, $A_{I,J}$ is the number of $|I|$-Paths that are non-intersecting from sources indexed by I to sinks indexed by J.

Theorem 2. The weighted path matrix A of any planar network is totally nonnegative. In particular, each minor $A_{I,J}$ is equal to the weighted sum of families of nonintersecting paths from sources indexed by I to sinks indexed by J. (Note: Theorem 1 is a special case of Theorem 2, with all edges having weight one).

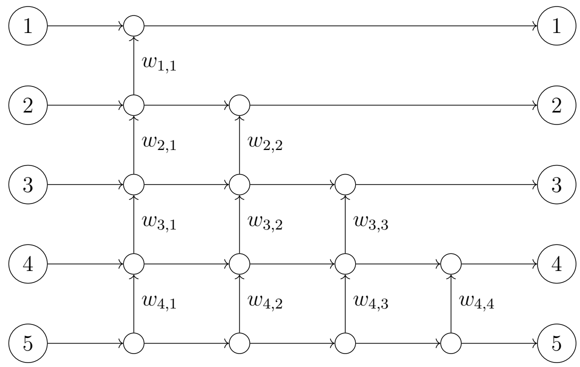
It turns out that the converse of Lindström's Lemma is also true.

Theorem 3. Every totally nonnegative matrix can be achieved as the weighted path matrix of some planar network.

In general, this network is not unique, but if A is an invertible matrix, one can find a planar network of the form (picture shown for 5 by 5 matrix):



where edges labeled with $*$ are the only weights that need to be determined (all other weights are one). The picture becomes even simpler for *invertible lower triangular matrices* with ones along the main diagonal:

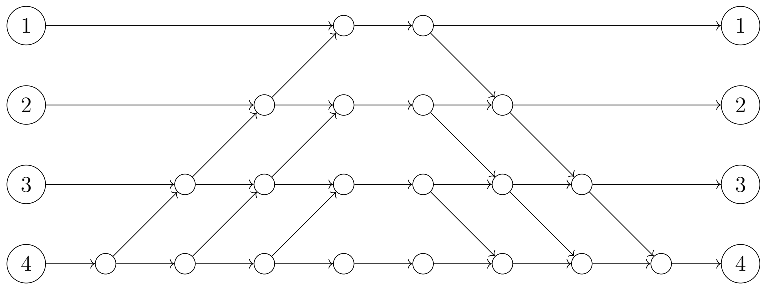


Theorem 4. Given totally nonnegative matrices A and B with associated planar networks G_A and G_B, the planar network of the matrix product AB is $G_{AB} = G_A * G_B$, a concatenation of G_A with G_B. More precisely, sink i of G_A connects to source i of G_B, the sources of G_A become the sources of G_{AB}, and the sinks of G_B become the sinks G_{AB}.

Proof. Consider the matrix product AB whose (n, k)-entry is given by $\sum_i A_{n,i}B_{i,k}$. Here i denotes the index of the sink of A (and source of B) that the paths from source n to sink k pass through. $\square$

2. The Permanent of a Path Matrix

Consider the planar network



with path matrix given by

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{pmatrix}$$

Question: In how many ways can sources 1,2,3,4 simultaneously reach different sinks (so necessarily the target sinks are 1,2,3,4). In other words, how many 4-Paths are there?

We begin with a simpler scenario. If we are looking for the number of 4-Paths such that each source i reaches sink i (crossing is allowed), the number if ways to do so is

$$p_{1,1} \cdot p_{2,2} \cdot p_{3,3} \cdot p_{4,4} = 1 \cdot 2 \cdot 6 \cdot 20 = 240$$

To see why, recall that 4-Paths is a collection of 4 different paths, so for each $i = 1, 2, 3, 4$, choose one path going from source i to sink i (the number of choices for one such path is equal to entry $p_{i,i}$). Since we don't care about whether or not the paths are crossing, we have exactly $p_{1,1}\ p_{2,2}\ p_{3,3}\ p_{4,4}$ ways to do so.

In general, if we want for a permutation π of $[4]$ to count the 4-Paths where each source i goes to sink $\pi(i)$, then we get

$$p_{1,\pi(1)} \cdot p_{2,\pi(2)} \cdot p_{3,\pi(3)} \cdot p_{4,\pi(4)}$$

ways to do so. The answer to our question is given by summing up over all permutations π in S_4,

$$\sum_{\pi \in S_4} p_{1,\pi(1)} \cdot p_{2,\pi(2)} \cdot p_{3,\pi(3)} \cdot p_{4,\pi(4)}$$

This is equal to the **permanent** of our matrix.

3. The Determinant of a Path Matrix

A natural question to ask is the following: what happens if we impose the restriction that the paths discussed in the previous question are nonintersecting?

Question: How many ways can sources reach different sinks, with no two intersecting paths?

Observe that source k must end up at sink k, otherwise the paths would have to cross at some point. For the path matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{pmatrix}$$

there were 240 intersecting and nonintersecting 4-Paths, but most of them will intersect at some point. The answer is given by the following theorem:

Theorem 5. Let A be a TN n by n matrix, i.e. a path matrix for some planar network G. Then the number of n-Paths from sources $1, \dots, n$ to sinks $1, \dots n$ that do not intersect is equal to the determinant of A,

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \dots a_{n,\sigma(3)}$$

where $\operatorname{sgn}(\sigma) = 1$ if σ is an even permutation (expressible as a product of an even number of transpositions), and $\operatorname{sgn}(\sigma) = -1$ if σ is odd.

Proof. See *Counting on determinants* by Arthur Benjamin and Naiomi Cameron. □

Theorem 6. Let G be an acyclic graph with n sources and n sinks. Let A be the path matrix of the graph (so entry (i, j) is the number of paths from source i to sink j). Then

a) The # of n-Paths is the permanent of A.
b) If G is non-permutable, the # of nonintersecting n-Paths is the determinant of A.
c) If G is NOT non-permutable, $\det A = \text{Even}(G) - \text{Odd}(G)$, where

$$\text{Even}(G) = \# \text{ of nonintersecting paths corresponding to even permutations}$$
$$\text{Odd}(G) = \# \text{ of nonintersecting paths corresponding to odd permutations}$$

4. Testing for TP and TN

Given an $m \times n$ matrix A, how can one determine if A is TP or TN? From the definition, one would have to check

$$\sum_{k=1}^{\min\{m,n\}} \binom{m}{k}\binom{n}{k}$$

minors, which is very time consuming even for small examples. However, it turns out this is possible to do in polynomial time (Fallat and Johnson, *Totally Nonnegative Matrices*, page 74).

Permuting the rows/columns generally alters TP, so we introduce the following definition to measure how spread out the index sets I and J are relative to $\{1, 2, \dots, n\}$.

Definition 6. For $I \subseteq \{1, 2, \dots, n\}$, $I = \{i_1 < i_2 < \cdots < i_k\}$, the **dispersion** of the index set I is given by

$$d(I) = i_k - i_1 - k + 1$$

Whenever $d(I) = 0$, we say that I is a **contiguous index set**.

Example 1. For $n = 8$, we have $d(\{1, 2, 4\}) = 4 - 1 - 3 + 1 = 1$. Similarly $d(\{1, 2, 3\}) = d(\{4, 5, 6, 7\}) = d(\{2, 3, 4, 5, 6\}) = 0$. In other words, the contiguous sets are the subsets of $\{1, 2, \dots, n\}$ containing consecutive elements.

Definition 7. If I and J are two contiguous sets of the same size k, then the submatrix $A[I, J]$ is called a **contiguous submatrix** of A, and the corresponding minor $A_{I,J}$ is called a **contiguous minor**.

Definition 8. A contiguous minor in which I or J is of the form $\{1, 2, \dots k\}$ (where $k \leq n$) is called **initial**.

Example 2. Consider the 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

The initial minors in this case are a_{11} (corresponding to $I = \{1\}$ and $J = \{1\}$), a_{12} (corresponding to $I = \{1\}$ and $J = \{2\}$), a_{21} (corresponding to $I = \{2\}$ and $J = \{1\}$), and $\det A$ (corresponding to $I = J = \{1, 2\}$).

Theorem 7. If all contiguous minors of an $m \times n$ matrix A are positive, then A is $\text{TP}_{\min\{m,n\}}$.

Theorem 8. If all contiguous minors of an $m \times n$ matrix A are positive, then A is TP.

Definition 9. For an $m \times n$ matrix A, an upper right (lower left) minor of A is one of the form $A_{I,J}$ where I consists of the first k (last k) and J consists of the last k (first k) indeces ($k \leq \min\{m, n\}$). A **corner minor** of A is one that is either a lower left minor or an upper right minor.

Note: there are exactly $2n - 1$ distinct corner minors for an $n \times n$ matrix.

Theorem 9. Suppose that $A \in M_n$ is TN. Then A is TP if and only if all corner minors of A are positive.

Now that we have some idea on how much we need for TP, let's see what set of minors is sufficient for TN. It turns out that for given index sets I and J, it is possible to find a matrix

$A \in M_{m,n}$ such that every minor of A is nonnegative, except for $A_{I,J} < 0$. If the rank of A is known, some reduction of the collection of minors we need to test for TN is possible:

Theorem 10. If $A \in M_{m,n}$ and $r = \text{rank}(A)$, then A is TN if and only if for all index sets I, J with $|I| = |J|$ and $d(J) \leq n - r$, the minor $A_{I,J} \geq 0$.

In our case, we are working with lower triangular matrices with 1's along the main diagonal, so $r = n$. Then it suffices to test only the minors corresponding to I, J with $d(J) = 0$. The number of minors to test is

$$\sum_{k=1}^{n} \binom{n}{k}(n-k+1)$$

since for $|I| = |J| = k$, there are $\binom{n}{k}$ choices for the rows and the number of contiguous index sets J of length k is $n - k + 1$. For example, when $k = 1$, we have n possibilities for J (the singleton sets containing 1,2, etc). When $k = n$, we have 1 choice for J (the entire set).

Definition 10. A minor $A_{I,J}$ is **quasi-initial** if either $I = \{1, 2, \ldots, k\}$ and $|J| = k$ is arbitrary, or $|I| = k$ is arbirary, while $J = \{1, 2, \ldots, k\}$, where $k = 1, 2, \ldots, n$.

In the case of triangular matrices, the following holds:

Theorem 11. Suppose $L \in M_n$ is invertible and lower triangular. Then L is InTN (invertible and TN) if and only if

$$L_{I,\{1,2,\ldots,k\}} \geq 0$$

for all $I \subseteq \{1, 2, \ldots, n\}$. Here $k = 1, 2, \ldots, n$.

5. The Pascal Matrix P_n

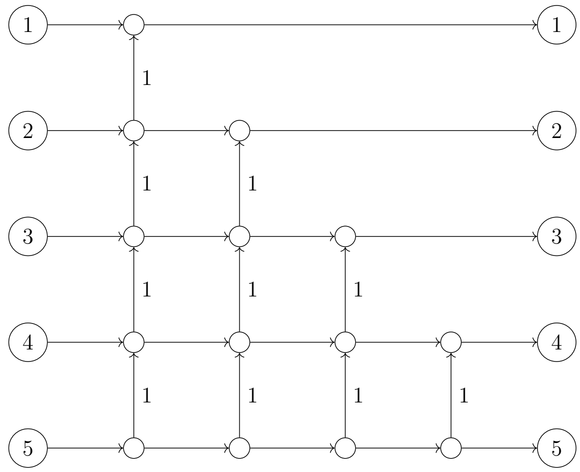
The Pascal triangle is given by the well-known recurrence

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

and so the Pascal matrix is defined to be the matrix whose (n,k)-entry is the binomial coefficient $\binom{n-1}{k-1}$, so the entry in the first row and first column is $\binom{0}{0} = 1$. For example, the 5 by 5 Pascal matrix P_5 is

$$P_5 = \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 2 & 1 & & \\ 1 & 3 & 3 & 1 & \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix}$$

It is an easy exercise to check that the planar network for the Pascal matrix is of the form



6. Stirling numbers of the second kind

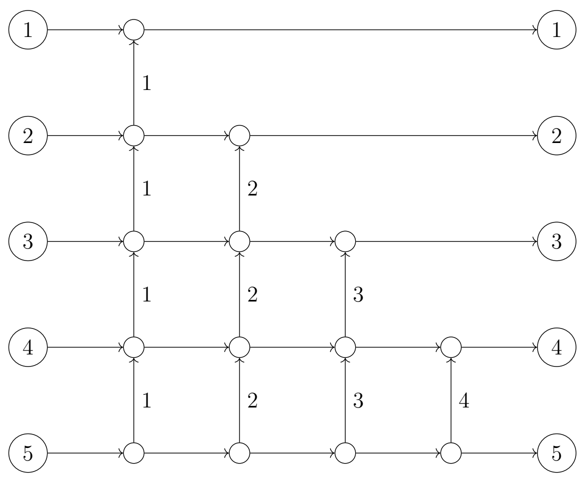
The Stirling numbers of the second kind are given by the recurrence

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \cdot \left\{ {n-1 \atop k} \right\}$$

and so the Stirling 2 matrix S is defined to be the matrix whose (n,k)-entry is the number $\left\{ {n \atop k} \right\}$. For example, the 5 by 5 Stirling matrix of the second kind S_5 is

$$S_5 = \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 3 & 1 & & \\ 1 & 7 & 6 & 1 & \\ 1 & 15 & 25 & 10 & 1 \end{pmatrix}$$

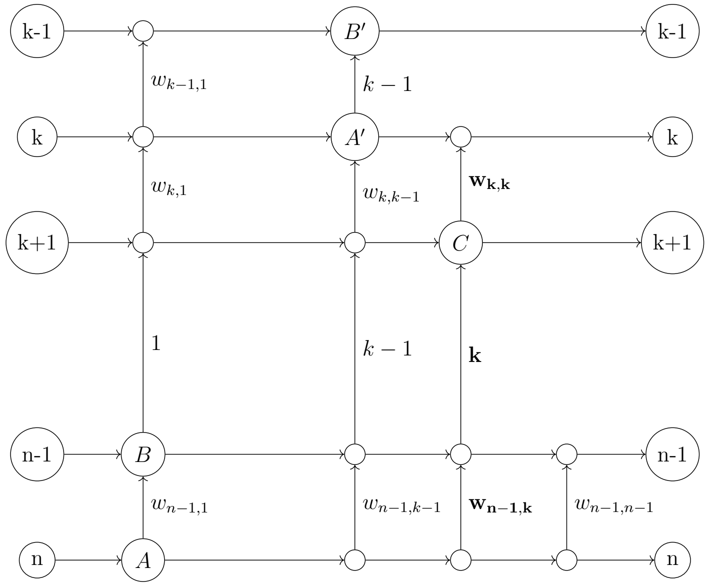
with planar network



Theorem 12. In general, the planar network of the S_n matrix will have weight matrix W_S of dimension $(n-1) \times (n-1)$ with weights

$$w_{n,k} = k$$

Proof. Consider the simplified picture of a planar network with rows k, $n-1$, and k shown below.



with the boldface edges $w_{n-1,k} = \cdots = w_{k,k} = k$ being the last column of edges that one can take on a path from source n to sink k.

The paths from source n to sink k can be partitions over those paths which avoid the edge with weight $w_{k,k} = k$, and those that don't. First, it is useful to observe that for any path connecting source n to sink k, it is possible to shift it down (while keeping its shape) and connect $n+1$ to $k+1$ without changing the weights (this follows from the vertical invariance of the proposed weight matrix - the weights do not depend on n, only on k).

For the first category, there are precisely $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}$ such weighted paths. To see why, observe that $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}$ counts the weighted paths from node B to node B', but due to the vertical invariance of the weights, this is the same as the weighted paths from A to A' (and these are the paths from n to k which avoid edge $w_{k,k}$).

For the second category, there are precisely $k \cdot \begin{bmatrix} n-1 \\ k \end{bmatrix}$ such weighted paths. To see why, consider a path from $n-1$ to k. Shifted down, this will be a path from n to node C, to which we can append edge $w_{k,k}$ to obtain a path from n to k which travels through $w_{k,k}$. This is shifting and appending the last weighted edge is a bijection.

We showed that for the proposed weight matrix $w_{n,k} = k$, the path matrix satisfies the recurrence for the Stirling numbers of the second kind. This concludes our proof.

□

7. Inverses of nonnegative matrices

Definition 11. Consider a matrix M with entries $m_{n,k}$. Define the row flip matrix M_{-1} to be the matrix whose (n,k)-entry is $m_{n,n-k+1}$. For example

$$\begin{pmatrix} 1 & & & \\ 1 & 2 & & \\ 1 & 2 & 3 & \\ 4 & 5 & 6 & 1 \end{pmatrix}_{-1} = \begin{pmatrix} 1 & & & \\ 2 & 1 & & \\ 3 & 2 & 1 & \\ 1 & 6 & 5 & 4 \end{pmatrix}$$

Another way to think about this is that M_{-1} has column k equal to the k-th diagonal of M (so the main diagonal becomes the first column, the entries below the main diagonal become the second column, etc.).

Definition 12. For a matrix M, define its dual $\overline{M}$ to be the matrix whose (n,k)-entry is $(-1)^{n+k}m_{n,k}$. Notice that for a non-negative matrix, this introduces the alternating sign pattern.

Theorem 13. (Dual Matrix Theorem) Suppose A is a nonnegative matrix with planar network G_A, and G_A has weight matrix W_A. Then

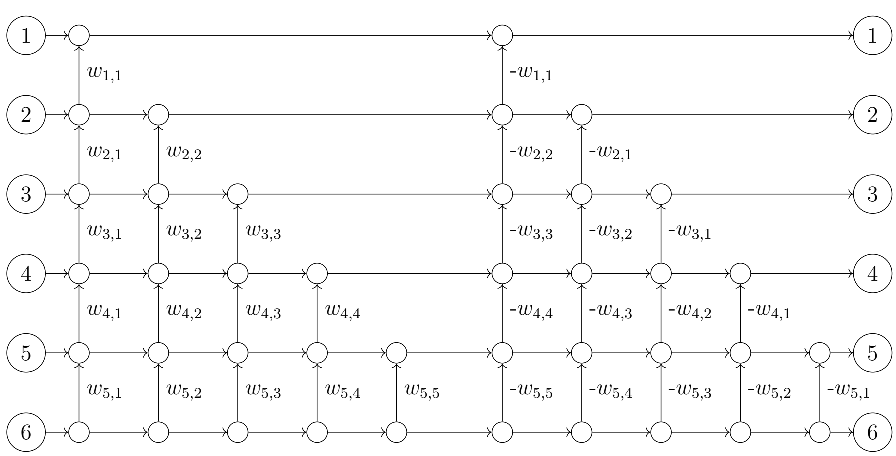
$$W_{\overline{A}} = -W_A \tag{1}$$

Proof. The result is easy to see: the paths going from n to k must have $n-k$ up steps. If $n-k$ is odd, so is $n+k$, so the entry will have the sign $(-1)^{n+k} = -1$, as desired. If $n-k$ is even, the opposite takes place. In either case, the magnitude of the weighted sum is not changed. $\square$

Theorem 14. (Inverse Matrix Theorem) Suppose A is a nonnegative matrix with planar network G_A, and G_A has weight matrix W_A. Then

$$W_{A^{-1}} = -(W_A)_{-1} \tag{2}$$

Proof. This is harder to prove. Our strategy is to look at the concatenation of G_A and G_B (where $B = A^{-1}$), and show that the resulting network is equivalent to the network for the identity matrix. Consider the concatenation $G_A * G_B$ shown below:



Our goal is to show that the sum of weighted paths from n to k is either zero (when $n \neq k$) or one (when $n = k$). It is trivial to see why for $n = k$, as there is only one (horizontal) path with weight one in this case.

For the case $n \neq k$, the idea is to match every path p from node n to node k, to a similar path $\overline{p}$ from n to k with a weight equal in magnitude, but opposite in sign, so that the sum of the weighted paths will be zero. We first illustrate that it is possible to do this with some examples:

a) Consider the path p from $n = 6$ to $k = 2$ determined (uniquely) by the up steps $(w_{5,1})(w_{4,3})(w_{3,3})(-w_{2,2})$, which takes its last up step on the A side of the network $(w_{3,3})$, travels horizontally to enter the B side, then takes one final up step on the B side $(-w_{2,2})$ before reaching sink 2. There is another path from 6 to 2 with up steps

$$(w_{5,1})(w_{4,3})(-w_{3,3})(-w_{2,2})$$

which travels horizontally after $w_{4,3}$, enters the B side of the network one level earlier, takes the up step $-w_{3,3}$, then continues with the up steps of the original path before reaching sink 2.

b) Consider the path from $n = 6$ to $k = 2$ with the up steps $(w_{5,1})(w_{4,2})(-w_{3,2})(-w_{2,1})$. One might think that a similar strategy can be used: enter the B side of the network one level sooner, via $(w_{5,1})(-w_{4,2})(-w_{3,2})(-w_{2,1})$, but this is not a valid path since there is no way to get to sink 2 via edge $-w_{4,2}$. Instead, this time we enter the B network one level later, and obtain the path path $(w_{5,1})(w_{4,2})(w_{3,2})(-w_{2,1})$.

c) $(w_{5,2})(w_{4,3})(-w_{3,2})(-w_{2,1}) \quad \longleftrightarrow \quad (w_{5,2})(-w_{4,3})(-w_{3,2})(-w_{2,1})$

d) $(w_{5,1})(w_{4,1})(w_{3,2})(-w_{2,1}) \quad \longleftrightarrow \quad (w_{5,1})(w_{4,1})(-w_{3,2})(-w_{2,1})$

e) $(w_{5,2})(w_{4,3})(-w_{3,3})(-w_{2,2}) \quad \longleftrightarrow \quad (w_{5,2})(w_{4,3})(w_{3,3})(-w_{2,2})$

f) $(w_{5,3})(w_{4,4})(-w_{3,3})(-w_{2,1}) \quad \longleftrightarrow \quad (w_{5,3})(-w_{4,4})(-w_{3,3})(-w_{2,1})$

g) $(w_{5,1})(w_{4,2})(w_{3,2})(-w_{2,1}) \quad \longleftrightarrow \quad (w_{5,1})(w_{4,2})(-w_{3,2})(-w_{2,1})$

h) $(w_{5,1})(w_{4,1})(w_{3,1})(-w_{2,1}) \quad \longleftrightarrow \quad (w_{5,1})(w_{4,1})(w_{3,1})(w_{2,1})$

In the general case, each weighted path from source n to sink k is uniquely determined by the up steps, so for such a path p we may write:

$$p = (w_{n,*}) \cdots (w_{r+1,s})(w_{r,a})(-w_{r-1,b})(-w_{r-2,t}) \cdots (-w_{k,*})$$

It is important to note that for a valid path (i.e. a path that follows the arrow flow), we must always have $b > t$ and $s \leq a$ (this is easily seen by examining the network diagram).

The above examples makes it clear that there are two cases to consider, depending on the last up step taken on the A side of the network (call it $w_{r,a}$), and the first up step taken on the B side (call it $w_{r-1,b}$). We have the following two cases:

1) If $a \leq b$, then associate the path p to

$$\overline{p} = (w_{n,*}) \cdots (w_{r+1,s})(w_{r,a})(w_{r-1,b})(-w_{r-2,t}) \cdots (-w_{k,*})$$

2) If $a > b$, then associate the path p to

$$\overline{p} = (w_{n,*}) \cdots (w_{r+1,s})(-w_{r,a})(-w_{r-1,b})(-w_{r-2,t}) \cdots (-w_{k,*})$$

Under this mapping, the image is always going to be a valid path. To complete the proof, observe that this mapping is an involution. $\square$

8. Stirling numbers of the first kind

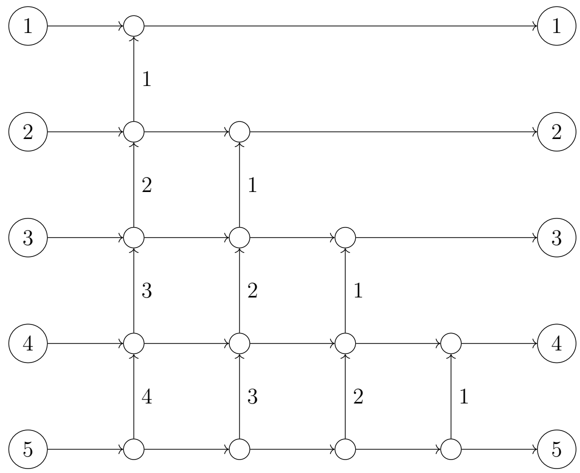
The Stirling numbers of the first kind are given by the recurrence

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + (n-1) \cdot \left\{ {n-1 \atop k} \right\}$$

and so the Stirling 1 matrix s is defined to be the matrix whose (n,k)-entry is $\left\{ {n \atop k} \right\}$.

$$s_5 = \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ 2 & 3 & 1 & & \\ 6 & 11 & 6 & 1 & \\ 24 & 50 & 35 & 10 & 1 \end{pmatrix}$$

One can verify that the following planar network yields the above matrix:



Theorem 15. In general, the planar network of the s_n matrix will have weight matrix W_s of dimension $(n-1) \times (n-1)$ with weights

$$w_{n,k} = n - k + 1$$

Proof. Follows directly from the inverse matrix theorem, but there is also a combinatorial proof for this. The paths from source n to sink k can be partitioned depending on whether or not they travel through the edge $w_{n-1,1} = n-1$ (consult previous images to identify this edge).

If the paths avoid that edge, notice that a diagonal shift creates a bijection between those paths avoiding $w_{n-1,1}$ and the paths from $n-1$ to $k-1$: taking ANY path and shifting it down and to the right will not affect its weight (similar to Stirling 2 networks, but this time we have diagonal invariance). If the paths travel through $w_{n-1,1}$, then their weight is equal to $w_{n-1,1}$ times the weight of a path from $n-1$ to k. Hence the path matrix of the proposed network must satisfy the recurrence, as desired.

□