Symbolic method and directed graph enumeration

Élie de Panafieu
Bell Labs France, Nokia

Sergey Dovgal
LIPN, Institut Galilée Université Paris 13

Abstract

We introduce the arrow product, a new generating function technique for directed graph enumeration. It provides new short proofs for previous results of Gessel on the number of directed acyclic graphs and of Liskovets, Robinson and Wright on the number of strongly connected directed graphs. We also obtain new enumerative results on directed graphs with given numbers of strongly connected components and source-like components using this new technique.

keywords. directed graph, digraph, analytic combinatorics, generating functions

1 Introduction

The enumeration of two important digraph families, the Directed Acyclic Graphs (DAGs) and the strongly connected digraphs, has been successfully approached at least since 1969. Apparently, it was Liskovets [Lis69a; Lis70] who first deduced a recurrence for the number of strongly connected digraphs and also introduced and studied the concept of initially connected digraph, a helpful tool for their enumeration. Subsequently, Wright [Wri71] derived a simpler recurrence for strongly connected digraphs and Liskovets [Lis73] extended his techniques to the unlabeled case. Stanley counted labeled DAGs in [Sta73], and Robinson, in his paper [Rob77b], counted unlabeled DAGs with a given number of sources, which was the culmination of a series of publications he started in 1970 independently of Stanley. In the unlabeled case, his approach is very much related to the Species Theory [BLL98] which systematises the usage of cycle index series. Robinson also announced [Rob77a] a simple combinatorial explanation for the generating function of strongly connected digraphs in terms of the cycle index function. Publications on the exact enumeration of digraphs slowed down, until Gessel [Ges95], in 1995, returned to the problem with a new approach, based on graphic generating functions. It allowed him to enumerate initially connected components of a digraph and to immediately extend the analysis of DAGs by marking sources and sinks [Ges96]. The first English version paper we found containing the elegant expression for the generating function of strongly connected digraphs recalled in Theorem 9 is [Lis00]. It points to an earlier publication [Lis73] in Russian, which contains the proof.

The symbolic method [BLL98; FS09] is a dictionary that translates combinatorial operations into generating function relations. In particular, it allows to manipulate the generating functions without going into the coefficient level. Our contribution is twofold. Firstly, we describe a new operation, the arrow product, which enriches the symbolic method. Secondly, we propose simple proofs for the generating functions of directed acyclic digraphs (DAGs) and strongly connected graphs (SCCs), and obtain a new enumerative result using this technique: the number of digraphs with given numbers of SCCs and source-like components.

Those techniques enabled precise description of simple graphs phase transition (see e.g. [Jan+93]), so the techniques developed here might enable the study of digraphs phase transition [LS09; Luc90].

In this paper, we consider directed graphs (digraphs) with labeled vertices without loops and multiple edges. Two vertices \( u, v \) can be simultaneously linked by both edges \( u \rightarrow v \) and \( v \rightarrow u \). We also consider simple graphs which are undirected graphs with neither multiple edges nor loops.

*Élie de Panafieu is a member of Lincs www.lincs.fr and Sergey Dovgal is supported by the French ANR project MetACOne, ANR-15-CE40-0014. The order of the authors is alphabetical.
2 The symbolic approach

Definitions. Consider a sequence $(a_n(w))_{n=0}^\infty$. Define the exponential generating function (EGF) and the graphic generating function (GGF) (introduced in [Ges95]) of the sequence $(a_n(w))_{n=0}^\infty$ as

$$A(z, w) := \sum_{n \geq 0} a_n(w) \frac{z^n}{n!} \quad \text{and} \quad A(z, w) := \sum_{n \geq 0} a_n(w) \frac{z^n}{(1 + w)^{(\ell)} n!}.$$

To distinguish EGF from GGF, the latter are written in bold characters. The $n$th coefficient of a series $A(z)$ with respect to the variable $z$ is denoted by $[z^n]A(z)$, so $A(z) = \sum_{n \geq 0} [z^n]A(z) z^n$.

The exponential Hadamard product of two series $A(z) = \sum_{n \geq 0} a_n(z) \frac{z^n}{n!}$ and $B(z) = \sum_{n \geq 0} b_n(z) \frac{z^n}{n!}$ is denoted by and defined as

$$A(z) \odot B(z) := \left( \sum_{n \geq 0} a_n(z) \frac{z^n}{n!} \right) \odot \left( \sum_{n \geq 0} b_n(z) \frac{z^n}{n!} \right) := \sum_{n \geq 0} a_n b_n \frac{z^n}{n!}.$$

All Hadamard products are taken with respect to the variable $z$. The Hadamard product can be used to convert between EGF and GGF (see Corollary 4). The exponential Hadamard product should not be confused with the ordinary Hadamard product $\sum_{n \geq 0} (\lfloor z^n \rfloor A(z))(\lfloor z^n \rfloor B(z)) z^n$.

If $A$ is a certain family of digraphs or graphs, we can associate to it a sequence of series $(a_n(w))_{n=0}^\infty$, such that $[w^m]a_n(w)$ is equal to the number of elements in $A$ with $n$ vertices and $m$ directed edges. Consequently, we can associate both EGF and GGF to the same family of digraphs or graphs.

An advantage of the symbolic method is its ability to keep track of a collection of parameters in combinatorial objects. The two default parameters are the numbers of vertices and edges, and the arguments $z$ and $w$ of a generating function $F(z, w)$ correspond to these parameters. As a generalization, we consider multivariate generating functions

$$A(z, w, u) := \sum_{n,p} a_{n,p}(w) u^p \frac{z^n}{n!} \quad \text{and} \quad A(z, w, u) := \sum_{n,p} a_{n,p}(w) u^p \frac{z^n}{(1 + w)^{(\ell)} n!},$$

where $u = (u_1, \ldots, u_d)$ is the vector of variables, $p = (p_1, \ldots, p_d)$ denotes a vector of parameters, and the notation $u^p := \prod_k u_k^{p_k}$ is used. We say that the variable $u_k$ marks its corresponding parameter $p_k$.

Combinatorial operations. The next proposition recalls classic operations on EGFs (see [FS09]), which extend naturally to GGFs.

Proposition 1. Consider two digraph (or graph) families $A$ and $B$. The EGF and GGF of the disjoint union of $A$ and $B$ are $A(z,w) + B(z,w)$ and $A(z, w) + B(z, w)$. The EGF and GGF of the digraphs from $A$ where one vertex is distinguished are $z \partial_z A(z)$ and $z \partial_z A(z, w)$. The EGF of sets of digraphs from $A$ is $e^{A(z,w)}$. The EGF of pairs of digraphs $(a, b)$ with $a \in A$ and $b \in B$ (relabeled so that the vertex labels of $a$ and $b$ are disjoint, see [FS09]) is $A(z, w) B(z, w)$. If a variable $u$ marks the number of specific items in the EGF $A(z, w, u)$ or the GGF $A(z, w, u)$ of the family $A$, then the EGF and GGF for the objects $a \in A$ which have a distinguished subset of these specific items are $A(z, w, u + 1)$ and $A(z, w, u + 1)$.

The next proposition presents our new combinatorial interpretation of the product of GGFs. It was implicitly used by Gessel in several proofs (e.g. [Ges96]) at coefficient level, but we have not found it expressed at the generating function level. However, a combinatorial interpretation of the exponential of GGFs can be found in [Ges96; GS96].

Proposition 2. We define the arrow product of $A$ and $B$ as the family $C$ of pairs $(a, b)$, with $a \in A$, $b \in B$ (relabeled so that $a$ and $b$ have disjoint labels), where an arbitrary number of edges oriented from vertices of $a$ to vertices of $b$ are added (see Figure 1). The GGF of $C$ is equal to $A(z, w) B(z, w)$.

Proof. Consider two digraph families $A$ and $B$, with associated sequences $(a_n(w))$, $(b_n(w))$. Then the sequence associated to the GGF $A(z, w) B(z, w)$ is

$$c_n(w) = (1 + w)^{(\ell)} n! [z^n] \left( \sum_k \frac{a_k(w)}{1 + w} \frac{z^k}{k!} \right) \left( \sum_{\ell} \frac{b_\ell(w)}{(1 + w)^{(\ell)} n!} \right) = \binom{n}{k} \sum_{k+\ell=n} (1 + w)^{k\ell} a_k(w) b_\ell(w).$$

2
This series $c_n(w)$ has the following combinatorial interpretation: it is the generating function (the variable $w$ marks the edges) of digraphs with $n$ vertices, obtained by

- choosing a digraph $a$ of size $k$ in $A$, and a digraph $b$ of size $\ell$ in $B$, such that $k + \ell = n$,
- choosing a subset of $\{1, \ldots, n\}$ for the labels of $a$ (and $b$ receives the complementary set for its labels),
- each oriented edge $(u, v)$ with $u$ vertex from $a$, and $v$ vertex from $b$, is or not added.

Hence, $(c_n(w))$ is the sequence associated to the arrow product of $A$ and $B$.

### 3 Generating functions from the symbolic method

We start by defining the building bricks for the symbolic method of the directed graphs.

**Proposition 3.** The EGF of graphs $G(z, w)$, GGF of digraphs $D(z, w)$, and GGF of sets $\text{Set}(z, w)$ (graphs that contain no edge) are

$$G(z, w) = D(z, w) = \sum_{n \geq 0} (1 + w)^{\binom{n}{2}} \frac{z^n}{n!} \quad \text{and} \quad \text{Set}(z, w) = \sum_{n \geq 0} \frac{1}{(1 + w)^{(n^2)}} \frac{z^n}{n!}.$$

**Proof.** The sequences associated to the families of graphs, digraphs and sets are

$$g_n(w) = (1 + w)^{\binom{n}{2}}, \quad d_n(w) = (1 + w)^{n(n-1)}, \quad \text{and} \quad \text{set}_n(w) = 1.$$

The expressions of the EGFs and GGFs follow.

**Corollary 4.** The EGF and GGF of a family $A$ are linked by the relations

$$A(z, w) = G(z) \circ A(z, w) \quad \text{and} \quad A(z) = \text{Set}(z, w) \circ A(z, w).$$

**Directed acyclic graphs.** The next proposition illustrates the power of the symbolic method. The first result and its proof are classic. The second comes from [Sta73; Rob77b; Ges96].

**Proposition 5.** The EGF $C(z, w)$ of connected graphs and the GGF $\text{DAG}(z, w, u)$ of directed acyclic graphs (DAGs) with an additional variable $u$ marking the sources (i.e. there are no oriented edge pointing to those vertices) are

$$C(z, w) = \log(G(z, w)) \quad \text{and} \quad \text{DAG}(z, w, u) = \frac{\text{Set}((u - 1)z, w)}{\text{Set}(-z, w)}.$$

**Proof.** Since a graph is a set of connected graphs, we have, applying the symbolic method (Proposition 1),

$$G(z, w) = e^{C(z, w)}, \quad \text{so} \quad C(z, w) = \log(G(z, w)).$$

The GGF of DAGs where each source is either marked, or left unmarked by the variable $u$, is $\text{DAG}(z, w, u+1)$ (see Proposition 1). Such a DAG is decomposed as the arrow product of a set (the marked sources) with a digraph (Figure 2), so

$$\text{DAG}(z, w, u+1) = \text{Set}(zu, w)DAG(z, w).$$

Taking $u = -1$ gives $1 = \text{Set}(-z, w)\text{DAG}(z, w)$, so $\text{DAG}(z, w) = 1/\text{Set}(-z, w)$. Replacing $u$ with $u - 1$ gives $\text{DAG}(z, w, u) = \text{Set}((u - 1)z, w)/\text{Set}(-z, w)$. This second proof also illustrates the translation into the generating function world of the inclusion-exclusion principle. 

3
**Strongly connected graphs.** Let us recall that the *condensation* of a digraph is the directed acyclic graph (DAG) obtained from it by contracting each strongly connected component (SCC) to a vertex. The SCCs of the digraph corresponding to sources of the condensation are called *source-like SCCs*.

**Lemma 6.** The GGF $D(z, w, u, v)$ of digraphs where $u$ marks the number of source-like SCCs, and $v$ the sum of their sizes, is equal to

$$D(z, w, u, v) = \left( \text{Set}(z, w) \odot e^{u \text{SCC}(z, w) - \text{SCC}(z, w)} \right) G(z, w).$$

**Proof.** Let $W$ denote the family of sets of strongly connected components, each of them either

- marked by $u$, and where each vertex is either marked by $v$ or left unmarked,
- not marked by $u$, but containing at least one vertex marked by $v$.

This construction translates into the relation

$$W(z, w, u, v) = e^{u \text{SCC}(z, w+1, w)} \cdot e^{\text{SCC}(z, w+1, w) - \text{SCC}(z, w)}.$$

By substituting $u \mapsto u + 1$ and $v \mapsto v + 1$ in $D(z, w, u, v)$, we obtain the generating function of digraphs with variables $u$ marking distinguished source-like components, and $v$ marking distinguished vertices in source-like components (see Proposition 1). As illustrated by Figure 3, such digraphs can be represented as an arrow product of the family $W$ and the family of all digraphs $D$. On the level of generating functions, it writes as

$$D(z, w, u + 1, v + 1) = W(z, w, u, v)D(z, w) = (\text{Set}(z, w) \odot W(z, w, u, v)) D(z, w),$$

using Corollary 4. According to Proposition 3, the digraphs GGF $D(z, w)$ is equal to the graphs EGF $G(z, w)$. Finally, $u$ is replaced by $u - 1$ and $v$ by $v - 1$ to obtain the claimed result. \qed

The *initially connected digraphs* are defined as digraphs where any vertex is reachable from the vertex with label 1 via an oriented path. Such digraphs have exactly one source-like SCC.

**Lemma 7.** The GGFs $IC(z, w)$ and $[u^1]D(z, w, u, v)$ of initially connected digraphs and, respectively, digraphs containing only one source-like SCC, whose size is marked by $v$, are linked by the relation

$$\partial_{v=1}[u^1]D(z, w, u, v) = z \partial_v IC(z, w).$$

**Proof.** As noted in [BLL98], the derivative $\partial_v IC(z, w)$ corresponds to the class of initially connected digraphs with omitted vertex of label 1. Multiplying by $z$ implies inserting a node with arbitrary label in place of it. We thus obtain digraphs reachable from a distinguished node. On the other hand, the operation $v \partial_v$ distinguishes a vertex in a source-like SCC. Taking $[u^1]v \partial_v D(z, w, u, v)|_{v=1}$ brings us to the case where there is exactly one source-like SCC. It implies that every vertex is reachable from the distinguished one. Therefore, the two generating functions enumerate the same family. \qed

Our motivation for introducing initially connected digraphs is that there is a relation between their enumeration and the number of connected graphs ([Lis69b], proof also available in the conclusion of [Jan+93]).

**Lemma 8.** The GGF of initially connected digraphs is equal to the EGF of connected graphs

$$IC(z, w) = C(z, w) = \log(G(z, w)).$$

Combining the previous results, we finally provide our new proof for the EGF of strongly connected digraphs (original result from [Lis00; Lis73]).

**Theorem 9.** The exponential generating function of strongly connected digraphs is equal to

$$\text{SCC}(z, w) = -\log \left( G(z, w) \odot \frac{1}{G(z, w)} \right), \quad \text{where} \quad G(z, w) = \sum_{n \geq 0} (1 + w)^{[n]} \frac{z^n}{n!}.$$
Proof. Combining Lemma 7 and Lemma 8, we obtain

\[ \partial_{v=1}[u^1]D(z, w, u, v) = z\partial_z IC(z, w) = z\partial_z C(z, w). \]

Since \( C(z, w) = \log(G(z, w)) \) (see Proposition 5), the right hand-side is equal to \( z(\partial_z G(z, w))/G(z, w) \). Injecting the expression of \( D(z, w, u, v) \) from Lemma 6 gives

\[ \partial_{v=1}[u^1] \left( \text{Set}(z, w) \circ e^{u \text{SCC}(zv, w) - \text{SCC}(z, w)} \right) G(z, w) = \frac{z\partial_z G(z, w)}{G(z, w)}. \]

Since

\[ \partial_{v=1}[u^1] e^{u \text{SCC}(zv, w) - \text{SCC}(z, w)} = (z\partial_z \text{SCC}(z, w)) e^{-\text{SCC}(z, w)} = -z\partial_z e^{-\text{SCC}(z, w)}, \]

this relation becomes, after dividing both sides by \( G(z, w) \),

\[ \text{Set}(z, w) \circ z\partial_z e^{-\text{SCC}(z, w)} = \frac{z\partial_z G(z)}{G(z, w)^2} = z\partial_z \frac{1}{G(z, w)}. \]

By construction, the product \( z\partial_z A(z) \circ B(z) \) is always equal to \( z\partial_z (A(z) \circ B(z)) \). Thus, the series \( \text{Set}(z, w) \circ e^{-\text{SCC}(z, w)} \) and \( 1/G(z, w) \) have equal derivatives and constant term 1, so they are equal. The neutral element for the exponential Hadamard product is the exponential function (i.e. for any series \( A(z) \), we have \( A(z) \circ e^z = A(z) \)), so the inverse of \( \text{Set}(z, w) \), with respect to the Hadamard product, is \( G(z, w) \), implying

\[ e^{-\text{SCC}(z, w)} = G(z, w) \circ \frac{1}{G(z, w)}. \]

Applying the logarithm to both sides concludes the proof. \( \square \)

This formula enables fast computation of the numbers of strongly connected digraphs: \( O(nm \log(n + m)) \) arithmetic operations to compute the array of SCCs with at most \( n \) vertices and at most \( m \) edges, \( O(n \log(n)) \) for the SCCs with at most \( n \) vertices without edge constraint. The following theorem is an original result. It could be a key element to the investigation of the structure of critical digraphs.

Theorem 10. The GGF of digraphs where the numbers of source-like SCCs and all SCCs are marked by the variables \( u \) and \( s \) is equal to

\[ \frac{\text{Set}(z, w) \circ e^{(u-1)s \text{SCC}(z, w)}}{\text{Set}(z, w) \circ e^{-s \text{SCC}(z, w)}} \]

where \( \text{SCC}(z, w) \) has been expressed in Theorem 9.

Proof. In this proof, let \( D(z, w, u, s) \) denote the GGF of the digraph family expressed in the theorem. Then \( D(z, w, u + 1, s) \) is the GGF of digraphs where each source-like SCC is either marked by \( u \) or left unmarked. Such a digraph is decomposed as the arrow product of a set of marked source-like SCCs with a digraph, so

\[ D(z, w, u + 1, s) = \left( \text{Set}(z, w) \circ e^{us \text{SCC}(z, w)} \right) D(z, w, 1, s). \]

Replacing \( u \) with \(-1\) gives

\[ 1 = \left( \text{Set}(z, w) \circ e^{-s \text{SCC}(z, w)} \right) D(z, w, 1, s) \quad \text{so} \quad D(z, w, 1, s) = \frac{1}{\text{Set}(z, w) \circ e^{-s \text{SCC}(z, w)}}. \]

Injecting this in the previous expression, with \( u \) replaced by \( u - 1 \), finishes the proof. \( \square \)

Conclusion. Many other digraph families could be enumerated using the same technique: symbolic method enriched with the arrow product, translation from EGF to GGF using exponential Hadamard product. The next challenge is the extraction of the coefficient asymptotics of such series. This would give access to a precise analysis of the phase transition of digraphs, following [Jan+93].

Acknowledgements. We would like to thank Cyril Banderier for pointing out the reference [Lis00], and Vlady Ravelomanana for introducing us to the topic and for many fruitful discussions.
References

[BLL98] François Bergeron, Gilbert Labelle, and Pierre Leroux. Combinatorial species and tree-like structures. Vol. 67. Cambridge University Press, 1998.

[FS09] Philippe Flajolet and Robert Sedgewick. Analytic Combinatorics. Cambridge University Press, 2009.

[Ges95] Ira Martin Gessel. “Enumerative applications of a decomposition for graphs and digraphs”. In: Discrete Mathematics 139.1 (1995), pp. 257–271. issn: 0012-365X.

[Ges96] Ira Martin Gessel. “Counting acyclic digraphs by sources and sinks”. In: Discrete Mathematics 160.1–3 (1996), pp. 253–258.

[GS96] Ira Martin Gessel and Bruce Eli Sagan. “The Tutte polynomial of a graph, depth-first search, and simplicial complex partitions”. In: Electron. J. Combin 3.2 (1996), R9.

[Jan+93] Svante Janson, Donald Ervin Knuth, Tomasz Łuczak, and Boris Pittel. “The Birth of the Giant Component”. In: Random Structures and Algorithms 4.3 (1993), pp. 233–358.

[Lis00] Valery Anisimovich Liskovets. “Some easily derivable integer sequences”. In: Journal of Integer Sequences 3.2 (2000), p. 3.

[Lis69a] Valery Anisimovich Liskovets. “Лисковец Валерий Анисимович. Об одном рекуррентном методе подсчета графов с отмеченными вершинами”. Russian. In: Доклады Академии наук. Vol. 184. 6. [On one recurrent method of counting graphs with marked vertices, Doklady Akademii Nauk]. 1969, pp. 1284–1287.

[Lis69b] Valery Anisimovich Liskovets. “Лисковец Валерий Анисимович. Подсчет корневых инициально связных ориентированных графов”. Russian. In: Известия АН БССР. 5. [Enumeration of rooted initially connected oriented graphs, Izv. Akad. Nauk BSSR]. 1969, pp. 23–32.

[Lis70] Valery Anisimovich Liskovets. “The number of strongly connected directed graphs”. In: Mathematical notes of the Academy of Sciences of the USSR 8.6 (1970), pp. 877–882.

[Lis73] Valery Anisimovich Liskovets. “Лисковец Валерий Анисимович. К перечислению сильно связных ориентированных графов.” In: ДАН БССР 17 (1973). [A contribution to the enumeration of strongly connected digraphs, Dokl. Akad. Nauk BSSR], pp. 1077–1080.

[LS09] Tomasz Łuczak and Taral Guldahl Seierstad. “The critical behavior of random digraphs”. In: Random Structures & Algorithms 35.3 (2009), pp. 271–293.

[Luc90] Tomasz Łuczak. “The phase transition in the evolution of random digraphs”. In: Journal of Graph Theory 14.2 (1990), pp. 217–223.

[Rob77a] Robert W. Robinson. “Counting strong digraphs”. In: Journal of Graph Theory 1.2 (1977), pp. 189–190.

[Rob77b] Robert W Robinson. “Counting unlabeled acyclic digraphs”. In: Combinatorial mathematics V. Springer, 1977, pp. 28–43.

[Sta73] Richard P Stanley. “Acyclic orientations of graphs”. In: Discrete Mathematics 5.2 (1973), pp. 171–178.

[Wri71] Edward Maitland Wright. “The number of strong digraphs”. In: Bulletin of the London Mathematical Society 3.3 (1971), pp. 348–350.
A Appendix

Proof of Proposition 3. Consider a graph with \( n \) vertices. Each unordered pair of distinct vertices is either linked by an edge, or not. Thus, the sequence of series associated to the family of graphs is

\[
g_n(w) = (1 + w)^{\binom{n}{2}},
\]
and its EGF is, by definition,

\[
G(z, w) = \sum_{n\geq 0} g_n(w) \frac{z^n}{n!} = \sum_{n\geq 0} (1 + w)^{\binom{n}{2}} \frac{z^n}{n!}.
\]

In a digraph with \( n \) vertices, each ordered pair of distinct vertices is either linked by an oriented edge, or not. So the sequence of series associated to the family of digraphs is

\[
d_n(w) = (1 + w)^{n(n-1)},
\]
and its GGF is

\[
D(z, w) = \sum_{n\geq 0} d_n(w) \frac{z^n}{n!} = \sum_{n\geq 0} (1 + w)^{\binom{n}{2}} \frac{z^n}{n!}.
\]

There is exactly one graph without any edge, so the sequence of series associated to the set family is

\[
set_n(w) = 1,
\]
and its GGF is

\[
Set(z, w) = \sum_{n\geq 0} \frac{1}{(1 + w)^{\binom{n}{2}}} \frac{z^n}{n!}.
\]

Proof of Corollary 4. Consider a family \( A \) with sequence of series \( (a_n(w)) \). By definition of the EGF, GGF and exponential Hadamard product, we have

\[
G(z) \odot A(z) = \left( \sum_n (1 + w)^{\binom{n}{2}} \frac{z^n}{n!} \right) \odot \sum_n \frac{a_n(w)}{(1 + w)^{\binom{n}{2}}} \frac{z^n}{n!} = \sum_n a_n(w) \frac{z^n}{n!} = A(z),
\]
and similarly

\[
Set(z) \odot A(z) = \left( \sum_n \frac{1}{(1 + w)^{\binom{n}{2}}} \frac{z^n}{n!} \right) \odot \sum_n \frac{a_n(w)}{(1 + w)^{\binom{n}{2}}} \frac{z^n}{n!} = \sum_n \frac{a_n(w)}{(1 + w)^{\binom{n}{2}}} \frac{z^n}{n!} = A(z).
\]

Proof of Lemma 6. The proof relies on the inclusion-exclusion principle. Consider the digraph family where each source-like SCC is either marked, using the variable \( u \), or left unmarked, and where each vertex belonging to a source-like SCC is either marked, using the variable \( v \), or left unmarked. Its GGF is then \( D(z, w, u + 1, v + 1) \). Any such digraph can be decomposed into three parts (see Figure 3):

- a set of SCCs marked by \( u \), where each vertex is or is not marked by \( v \),
- a set of SCCs not marked by \( u \), but containing at least one vertex marked by \( v \), where each vertex is either marked or not by \( v \),
- the rest, which can be any arbitrary digraph.

There are no edges between the two first parts, because they contain only source-like SCCs, and there are arbitrary edges from them to the third part. Applying the symbolic method (Proposition 1), the EGF of the first and second parts are \( e^{u \text{SCC}(z(v+1),w)} \) and \( e^{v \text{SCC}(z(v+1),w) - \text{SCC}(z,w)} \). Since there are no edge between those parts, the EGF of their combination \( A \) is the product of their EGF

\[
A(z, w) = e^{(u+1) \text{SCC}(z(v+1),w) - \text{SCC}(z,w)}.
\]
Corollary 4 is applied to turn this EGF into a GGF
\[ A(z, w) = \text{Set}(z, w) \odot e^{(u+1) \text{SCC}(z(v+1), w) - \text{SCC}(z, w)}. \]

The GGF of the third part is the GGF of digraphs \( D(z, w) \), and the final digraph is the arrow product of the combination of the first two parts with the third, so its GGF is
\[ D(z, w, u + 1, v + 1) = \left( \text{Set}(z, w) \odot e^{(u+1) \text{SCC}(z(v+1), w) - \text{SCC}(z, w)} \right) D(z, w). \]

According to Proposition 3, the digraphs GGF \( D(z) \) is equal to the graphs EGF \( G(z, w) \). Finally, \( u \) is replaced by \( u - 1 \) and \( v \) by \( v - 1 \) to obtain the claimed result.

Proof of Lemma 8. Since a graph is a set of connected graphs, the EGF of graphs and connected graphs are linked by the relation
\[ G(z, w) = e^{C(z, w)}, \quad \text{so} \quad C(z, w) = \log(G(z, w)). \]

Injecting the expression of \( G(z, w) \) from Proposition 3 provides the second equation of the lemma. Let \( (ic_n(w)) \) and \( (c_n(w)) \) denote the sequences of series associated with initially connected digraphs and connected graphs. Extracting the coefficient \( n![z^n] \), the first equation is equivalent with
\[ \frac{ic_n(w)}{(1 + w)^{(2)}} = c_n(w) \quad \text{for all } n \geq 0. \]

Let \( (g_n(w)) := ((1 + w)^{(2)}) \) denote the sequence of series associated to (non-oriented) graphs. The last relation is rewritten
\[ ic_n(w) = g_n(w)c_n(w) \quad \text{for all } n \geq 0. \]

To prove this relation, let us decompose the oriented edges of an initially connected digraph \( I \) on \( n \) vertices into two sets of non-oriented edges, which will form non-oriented graphs on \( n \) vertices \( C \) and \( G \). Consider two vertices \( i \) and \( j \) from \( I \). Let us assume that either \( i \) is strictly closer to the vertex 1 than \( j \), or that they are at the same distance from 1, but \( i < j \). If \( I \) contains an oriented edge from \( i \) to \( j \), then the non-oriented edge \( \{i, j\} \) is added to \( C \). If \( I \) contains an oriented edge from \( j \) to \( i \), then \( \{i, j\} \) is added to \( G \). Since any vertex is reachable from 1 in \( I \), the graph \( C \) must be connected. Reversing the construction, any pair \( C, G \), of connected graph and graph, leads to an initially connected digraph. Thus, this construction provides a bijection between initially connected digraphs of size \( n \) and pairs (connected graph of size \( n \), graph of size \( n \)), which implies
\[ ic_n(w) = g_n(w)c_n(w). \]

\[ \square \]