# A combinatorial identity

Tony Dorlas,* Alexei Rebenko,† Baptiste Savoie*,‡

**Abstract**

We give an elementary proof of an interesting combinatorial identity which is of particular interest in graph theory and its applications.

The aim of this note is to give a new, elementary proof of the combinatorial identity (1) in Theorem 1. Done by induction, the proof we give is simple in that it only requires the use of the binomial formula along with the derivation operator. This identity finds an interest in graph theory (for enumeration of forests) and its applications. For instance, we came across (1) when evaluating, within the framework of rigorous statistical mechanics, contributions from forest graphs to a cluster expansion for classical gas correlation functions in [1]. The combinatorial identity (1) provides a means to directly derive a closed-form expression for the number of distinct forests on a collection of sets of vertices, see formula (2) in Theorem 2. We point out that the proof of formula (2) is self-contained and does not rely on graph theory related techniques as, e.g., in [2, Chap. 5].

## 1 The combinatorial identity.

**Theorem 1.** *Given $m, p \in \mathbb{N}$ with $1 \leq p \leq m$, and a collection $(x_i)_{i=1}^m$ of (complex) numbers $x_i \in \mathbb{C}$, the following identity holds*

$$\sum_{\{I_1,\dots,I_p\} \in \Pi_p(\{1,\dots,m\})} \prod_{j=1}^{p} \left( \sum_{i \in I_j} x_i \right)^{|I_j|-1} = \binom{m-1}{p-1} \left( \sum_{i=1}^{m} x_i \right)^{m-p}, \tag{1}$$

*where the sum on the left-hand side is over the set $\Pi_p(\{1,\dots,m\})$ of all partitions $(I_j)_{j=1}^p$ of $\{1,\dots,m\}$ into $p$ non-empty subsets.*

The proof of Theorem 1 is postponed to Sec. 3.

## 2 An application to enumeration of forests.

A *forest* is defined on a collection $(V_i)_{i=1}^m$ of sets of vertices $V_i$ as a graph on $V = V_1 \cup \dots \cup V_m$ with connected components which are *trees*, such that there are no lines between vertices of any individual $V_j$, and moreover, if each set $V_j$ is reduced to a single point, then the graph reduces to a single (connected) tree. Further, if in the reduced tree, $V_j$ $(j \geq 2)$ is connected to $V_i$ in the path from $V_j$ to $V_1$, then all lines between $V_i$ and $V_j$ eminate from a single vertex in $V_i$ (a *root*). Deriving a closed-form expression for the number of (distinct) forests as described above becomes straightforward by using Theorem 1

*Dublin Institute for Advanced Studies, School of Theoretical Physics, Dublin, Ireland
†Institute of Mathematics, Ukrainian National Academy of Sciences, Kyiv, Ukraine
‡Corresponding author - e-mail: baptiste.savoie@gmail.com

**Theorem 2.** *Given $m \in \mathbb{N}$, the number $N_m$ of distinct forests on a collection of sets of vertices $(V_i)_{i=1}^m$ is given by*

$$N_m = |V_1| \left( \sum_{i=1}^m |V_i| \right)^{m-2} \prod_{i=2}^m \left( 2^{|V_i|} - 1 \right). \tag{2}$$

*Proof.* For $m = 2$ we can choose the root in $|V_1|$ ways and connect it to $k = 1, \ldots, |V_2|$ points in $V_2$. This results in $|V_1|(2^{|V_2|} - 1)$ different forests (in fact they are trees). Suppose now that the statement is true for $m$. Adding an additional set of vertices $V_{m+1}$, it can be connected to $p = 1, \ldots, m$ other sets $V_j$ in the reduced tree on $V_1, \ldots, V_m$. Omitting the connections to $V_{m+1}$ we obtain $p$ separate forests given by subsets $I_1, \ldots, I_p$ of $\{1, \ldots, m\}$. If $I_1$ contains 1, then this yields a factor

$$|V_1| \left( \prod_{i \in I_1 \setminus \{1\}} \left( 2^{|V_i|} - 1 \right) \right) \left( \sum_{i \in I_1} |V_i| \right)^{|I_1|-2} \left( \sum_{i \in I_1} |V_i| \right) \left( 2^{|V_{m+1}|} - 1 \right),$$

the additional factor $\left( \sum_{i \in I_1} |V_i| \right)$ being due to the choice of root for the connection to $V_{m+1}$. Similarly, the other branches yield factors

$$|V_{m+1}| \left( \sum_{i \in I_r} |V_i| \right)^{|I_r|-1} \prod_{i \in I_r} \left( 2^{|V_i|} - 1 \right), \quad r = 2, \ldots, m.$$

The resulting expression for $N_{m+1}$ is

$$|V_1| \sum_{p=1}^m |V_{m+1}|^{p-1} \sum_{\{I_1, \ldots, I_p\} \in \Pi_p(\{1, \ldots, m\})} \prod_{r=1}^p \left( \left( \sum_{i \in I_r} |V_i| \right)^{|I_r|-1} \prod_{\substack{i \in I_r \\ i \neq 1}} \left( 2^{|V_i|} - 1 \right) \right).$$

Identity (1) yields

$$\begin{aligned} N_{m+1} &= |V_1| \left( \sum_{p=1}^m \binom{m-1}{p-1} |V_{m+1}|^{p-1} \left( \sum_{i=1}^m |V_i| \right)^{m-p} \right) \prod_{i=2}^{m+1} \left( 2^{|V_i|} - 1 \right) \\ &= |V_1| \left( \sum_{i=1}^{m+1} |V_i| \right)^{m-1} \prod_{i=2}^{m+1} \left( 2^{|V_i|} - 1 \right). \end{aligned}$$

This concludes the proof of Theorem 2. □

# 3 Proof of Theorem 1.

The proof is done by induction on $m$ and $p$. Note that, for $m = p$ both sides are equal to 1 and for $p = 1$ both sides are equal to $\left( \sum_{i=1}^m x_i \right)^{m-1}$. Now assume that identity (1) holds true for a given $m \geq 1$ and all $p \leq m$. Let $\mathcal{L}_m$ and $\mathcal{R}_m$ denote the left-hand side and right-hand side of (1) respectively. We may assume that $1 \in I_1$ and expand the factor $\left( \sum_{i \in I_1} x_i \right)^{|I_1|-1}$ by the binomial formula in powers of $x_1$

$$\left( \sum_{i \in I_1} x_i \right)^{|I_1|-1} = \sum_{n=0}^{|I_1|-1} \binom{|I_1|-1}{n} x_1^n \left( \sum_{i \in I_1 \setminus \{1\}} x_i \right)^{|I_1|-1-n}.$$

Inserting this into $\mathcal{L}_{m+1}$ and denoting $\tilde{I}_1 := I_1 \setminus \{1\}$, we have

$$\mathcal{L}_{m+1} = \sum_{n=0}^{m+1-p} x_1^n \sum_{\substack{\{I_1,\ldots,I_p\}\in\Pi_p(\{1,\ldots,m+1\})\\ 1\in I_1,\,|I_1|\geq n+1}} \binom{|\tilde{I}_1|}{n} \left(\sum_{i\in\tilde{I}_1} x_i\right)^{|\tilde{I}_1|-n} \prod_{j=2}^{p} \left(\sum_{i\in I_j} x_i\right)^{|I_j|-1}.$$

If $n=0$, we separate out $\tilde{I}_1=\emptyset$, for which $\{I_2,\ldots,I_p\}\in\Pi_{p-1}(\{2,\ldots,m+1\})$. If $\tilde{I}_1\neq\emptyset$ then $\{\tilde{I}_1,I_2,\ldots,I_p\}\in\Pi_p(\{2,\ldots,m+1\})$. Conversely, given a partition $\{\tilde{I}_1,\tilde{I}_2,\ldots,\tilde{I}_p\}\in\Pi_p(\{2,\ldots,m+1\})$ we obtain a unique partition of $\{1,\ldots,m+1\}$ by adding 1 to any of the sets $\tilde{I}_l$ with $l\in\{1,\ldots,p\}$. We can therefore write

$$\begin{aligned}\mathcal{L}_{m+1} = &\sum_{\{I_2,\ldots,I_p\}\in\Pi_{p-1}(\{2,\ldots,m+1\})} \prod_{j=2}^{p}\left(\sum_{i\in I_j} x_i\right)^{|I_j|-1}\\ &+\sum_{n=0}^{m+1-p} x_1^n \sum_{l=1}^{p} \sum_{\substack{\{\tilde{I}_1,\ldots,\tilde{I}_p\}\in\Pi_p(\{2,\ldots,m+1\})\\ |\tilde{I}_l|\geq n}} \binom{|\tilde{I}_l|}{n}\left(\sum_{i\in\tilde{I}_l} x_i\right)^{|\tilde{I}_l|-n} \prod_{\substack{j=1\\ j\neq l}}^{p}\left(\sum_{i\in\tilde{I}_j} x_i\right)^{|\tilde{I}_j|-1}.\end{aligned}$$

Note that in the second term on the right-hand side, $|\tilde{I}_l|\neq 0$ since $\{\tilde{I}_1,\ldots,\tilde{I}_p\}\in\Pi_p(\{2,\ldots,m+1\})$. Expanding the quantity $\mathcal{R}_{m+1}$ (the right-hand side of (1) but with $m+1$) in powers of $x_1$ and then replacing $\{2,\ldots,m+1\}$ by $\{1,\ldots,m\}$, it follows that it suffices to prove the equivalent identities

$$\begin{aligned}&\sum_{\{I_1,\ldots,I_{p-1}\}\in\Pi_{p-1}(\{1,\ldots,m\})} \prod_{j=1}^{p-1}\left(\sum_{i\in I_j} x_i\right)^{|I_j|-1}\\ &+\sum_{l=1}^{p} \sum_{\{\tilde{I}_1,\ldots,\tilde{I}_p\}\in\Pi_p(\{1,\ldots,m\})} \left(\sum_{i\in\tilde{I}_l} x_i\right)^{|\tilde{I}_l|} \prod_{\substack{j=1\\ j\neq l}}^{p}\left(\sum_{i\in\tilde{I}_j} x_i\right)^{|\tilde{I}_j|-1} = \binom{m}{p-1}\left(\sum_{i=1}^{m} x_i\right)^{m+1-p} \qquad (3)\end{aligned}$$

for $n=0$, and

$$\begin{aligned}&\sum_{l=1}^{p} \sum_{\substack{\{\tilde{I}_1,\ldots,\tilde{I}_p\}\in\Pi_p(\{1,\ldots,m\})\\ |\tilde{I}_l|\geq n}} \binom{|\tilde{I}_l|}{n}\left(\sum_{i\in\tilde{I}_l} x_i\right)^{|\tilde{I}_l|-n} \prod_{\substack{j=1\\ j\neq l}}^{p}\left(\sum_{i\in\tilde{I}_j} x_i\right)^{|\tilde{I}_j|-1}\\ &\qquad = \binom{m}{p-1}\binom{m+1-p}{n}\left(\sum_{i=1}^{m} x_i\right)^{m+1-p-n} \qquad (4)\end{aligned}$$

for all $1\leq n\leq m+1-p$. We start with the case of $n=0$. By the induction hypothesis, the first term on the left-hand side of (3) is equal to

$$\binom{m-1}{p-2}\left(\sum_{i=1}^{m} x_i\right)^{m+1-p}.$$

The second term on the left-hand side of (3) can be rewritten as

$$\sum_{\{\tilde{I}_1,\ldots,\tilde{I}_p\}\in\Pi_p(\{1,\ldots,m\})} \prod_{j=1}^{p}\left(\sum_{i\in\tilde{I}_j} x_i\right)^{|\tilde{I}_j|-1} \sum_{l=1}^{p}\left(\sum_{i\in\tilde{I}_l} x_i\right) = \binom{m-1}{p-1}\left(\sum_{i=1}^{m} x_i\right)^{m+1-p},$$

again by the induction hypothesis. Adding the two contributions for $n = 0$ yields the right-hand side of (3). For the case $n = 1$, the left-hand side of (4) can simply be rewritten as

$$\sum_{\{\tilde{I}_1,\ldots,\tilde{I}_p\}\in\Pi_p(\{1,\ldots,m\})} \left(\sum_{l=1}^{p} |\tilde{I}_l|\right) \prod_{j=1}^{p} \left(\sum_{i\in\tilde{I}_j} x_i\right)^{|\tilde{I}_j|-1} = m\binom{m-1}{p-1}\left(\sum_{i=1}^{m} x_i\right)^{m-p}$$
$$= \binom{m}{p-1}(m+1-p)\left(\sum_{i=1}^{m} x_i\right)^{m-p}.$$

Next, we prove the case where $2 \le n \le m+1-p$. The key idea is to apply the derivation operator $\sum_{k=1}^{m} \frac{\partial^{n-1}}{\partial x_k^{n-1}}$ to the left-hand side of (1). This gives

$$\sum_{k=1}^{m} \frac{\partial^{n-1}}{\partial x_k^{n-1}} \sum_{\{I_1,\ldots,I_p\}\in\Pi_p(\{1,\ldots,m\})} \prod_{j=1}^{p} \left(\sum_{i\in I_j} x_i\right)^{|I_j|-1}$$
$$= \sum_{j'=1}^{p} \sum_{\{I_1,\ldots,I_p\}\in\Pi_p(\{1,\ldots,m\})} \prod_{\substack{j=1\\ j\neq j'}}^{p} \left(\sum_{i\in I_j} x_i\right)^{|I_j|-1} \sum_{k\in I_{j'}} \frac{\partial^{n-1}}{\partial x_k^{n-1}} \left(\sum_{i\in I_{j'}} x_i\right)^{|I_{j'}|-1}. \quad (5)$$

The $j'$ term on the right-hand side of (5) is equal to zero unless $|I_{j'}| \ge n$. In that case,

$$\sum_{k\in I_{j'}} \frac{\partial^{n-1}}{\partial x_k^{n-1}} \left(\sum_{i\in I_{j'}} x_i\right)^{|I_{j'}|-1} = |I_{j'}| \prod_{r=1}^{n-1} (|I_{j'}|-r) \left(\sum_{i\in I_{j'}} x_i\right)^{|I_{j'}|-n},$$

independently of $k \in I_{j'}$. Hence, the left-hand side of (5) can be rewritten as

$$\sum_{j'=1}^{p} \sum_{\substack{\{I_1,\ldots,I_p\}\in\Pi_p(\{1,\ldots,m\})\\ |I_{j'}|\ge n}} \prod_{r=0}^{n-1} (|I_{j'}|-r) \left(\sum_{i\in I_{j'}} x_i\right)^{|I_{j'}|-n} \prod_{\substack{j=1\\ j\neq j'}}^{p} \left(\sum_{i\in I_j} x_i\right)^{|I_j|-1},$$

which is nothing but $n!$ times the left-hand side of (4). On the other hand, applying the derivation operator $\sum_{k=1}^{m} \frac{\partial^{n-1}}{\partial x_k^{n-1}}$ to the right-hand side of (1) gives

$$m\binom{m-1}{p-1} \prod_{r=0}^{n-2} (m-p-r) \left(\sum_{i=1}^{m} x_i\right)^{m+1-p-n} = n!\binom{m}{p-1}\binom{m+1-p}{n}\left(\sum_{i=1}^{m} x_i\right)^{m+1-p-n},$$

which is just $n!$ times the right-hand side of (4). This concludes the proof of (4), and hence also the proof of the theorem. □

## Acknowledgment.

The authors thank Adrien Kassel for having brought to their attention reference [2]. The second author gratefully acknowledges the financial support of the Ukrainian Scientific Project "III-12-16 Research of models of mathematical physics describing deterministic and stochastic processes in complex systems of natural science".

## References

[1] T.C. Dorlas, A.L. Rebenko, B. Savoie, Correlation of clusters: Partially truncated correlation functions and their decay, e-print arXiv:1811.12342

[2] R.P. Stanley, *Enumerative Combinatorics, Volume 2*, Cambridge Studies in Advanced Mathematics, Series 62, Cambridge University Press, Cambridge, 1999