TI 2016-011/II
Tinbergen Institute Discussion Paper

Generalization of Binomial Coefficients to Numbers on the Nodes of Graphs

Anna Khmelnitskaya[a]
Gerard van der Laan[b]
Dolf Talman[c]

[a] *Saint-Petersburg State University, Russia;*
[b] *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute, the Netherlands;*
[c] *Tilburg University, Tilburg, the Netherlands.*

Generalization of binomial coefficients to numbers on the nodes of graphs.[1]

Anna Khmelnitskaya[2] Gerard van der Laan[3] Dolf Talman[4]

February 8, 2016

[1]The authors would like to thank Dinard van der Laan and Rene van den Brink for their helpful remarks. The research of Anna Khmelnitskaya was supported by RFBR (Russian Foundation for Basic Research) grant #16-01-00713 and NWO (Dutch Organization of Scientific Research) grant #040.11.516. Her research was done partially during her stay at Vrije Universiteit Amsterdam and the University of Twente, the hospitality of both universities is highly appreciated.

[2]A.B. Khmelnitskaya, Saint-Petersburg State University, Faculty of Applied Mathematics, Universitetskii prospekt 35, 198504, Petergof, SPb, Russia, e-mail: a.b.khmelnitskaya@utwente.nl

[3]G. van der Laan, Department of Econometrics and Tinbergen Institute, VU University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, e-mail: g.vander.laan@vu.nl

[4]A.J.J. Talman, CentER, Department of Econometrics & Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, e-mail: talman@tilburguniversity.edu

Abstract

The triangular array of binomial coefficients, or Pascal's triangle, is formed by starting with an apex of 1. Every row of Pascal's triangle can be seen as a line-graph, to each node of which the corresponding binomial coefficient is assigned. We show that the binomial coefficient of a node is equal to the number of ways the line-graph can be constructed when starting with this node and adding subsequently neighboring nodes one by one. Using this interpretation we generalize the sequences of binomial coefficients on each row of Pascal's triangle to so-called Pascal graph numbers assigned to the nodes of an arbitrary (connected) graph. We show that on the class of connected cycle-free graphs the Pascal graph numbers have properties that are very similar to the properties of binomial coefficients. We also show that for a given connected cycle-free graph the Pascal graph numbers, when normalized to sum up to one, are equal to the steady state probabilities of some Markov process on the nodes. Properties of the Pascal graph numbers for arbitrary connected graphs are also discussed. Because the Pascal graph number of a node in a connected graph is defined as the number of ways the graph can be constructed by a sequence of increasing connected subgraphs starting from this node, the Pascal graph numbers can be seen as a measure of centrality in the graph.

Keywords: binomial coefficient, Pascal's triangle, graph, Markov process, centrality measure.

JEL Classification Number: C00
AMS/MSC Classification: 05A10, 11B65, 60J10

1 Introduction

In mathematics, for any two integers $n \geq 0$ and $0 \leq k \leq n$, the number of combinations of k distinct elements of a given set composed by n different objects is conventionally denoted by C_n^k or $\binom{n}{k}$. This number occurs in many different contexts, in particular it appears as a coefficient in binomial expansions, wherefrom it gets its name, binomial coefficient. Arranging the binomial coefficients $C_n^0, \ldots, C_n^n$ from left to right in a row for successive values of n, we obtain a triangular array called Pascal's triangle. This triangle has as apex the number 1 in row 0 and has the property that every entry in row n of the triangle for $n \geq 1$ is the sum of the two entries in row $n-1$ to its left and to its right, i.e., $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$ for every $k = 0, \ldots, n$, with positions outside the triangle being assigned zero. It is well-known that in each row of Pascal's triangle the ratio of two consecutive binomial coefficients C_n^k and C_n^{k+1} is equal to the ratio of the number $k+1$ of positions $0, \ldots, k$ in that row from the position k to the left and the number $n-k$ of positions $k+1, \ldots, n$ in that row from the position $k+1$ to the right. Moreover, if $n-1$ is a prime number, then all binomial coefficients in row n, except those at the two ends which are equal to 1, are divisible by this prime. Also, it is well-known that any binomial coefficient C_n^k is equal to the number of different paths in Pascal's triangle starting at the apex and terminating at position k on row n, when at every step a path moves either to the left or to the right to the nearest position at the next row.

In this paper we first revisit the binomial coefficients within the framework of line-graphs (chains). For each integer $n \geq 0$, row n of Pascal's triangle corresponds to a line-graph with nodes $0, \ldots, n$ and edges between the nodes k and $k+1$ for $k = 0, \ldots, n-1$. To node k of this line-graph we assign the binomial coefficient C_n^k. Using the above properties for binomial coefficients we show that the binomial coefficient C_n^k assigned to node k is equal to the number of ways the line-graph can be constructed starting with the single node k and adding subsequently neighboring nodes one by one. We further show that the binomial coefficient of a node in a line-graph is equal to the sum of binomial coefficients of this node in the two subgraphs obtained by deleting precisely one of the extreme (end) nodes 0 and n. Moreover, it appears that the binomial coefficients, when normalized to sum up to one, are the steady state probabilities of a Markov process in which at every node the process moves to any of its neighbors in the line-graph with a probability proportional to the number of nodes connected to this node through the corresponding neighboring node.

We generalize the binomial coefficients assigned to the nodes of line-graphs to numbers assigned to the nodes of arbitrary connected graphs by defining the number of a node as the number of ways the graph can be constructed when starting with this node and adding subsequently neighboring nodes one by one. This is equivalent to say that the number of a node is equal to the number of ways that extreme nodes, the nodes for which the subgraph on the set of the remaining nodes is connected, can be removed from the graph one by one until the considered node is left. We call these numbers Pascal graph numbers. On the class of connected cycle-free graphs we prove that the Pascal graph number of a node is determined by the Pascal graph numbers of its neighbors in the subgraphs obtained by deleting the edges adjacent to the node. From this it immediately follows that, similar as for the binomial coefficients in any row of Pascal's triangle, in a cycle-free graph the ratio between the Pascal graph numbers of any two neighboring nodes is equal to the ratio of the numbers of nodes in the two subgraphs resulting from deleting the edge connecting these two neighbors. Moreover, if the number of nodes in a graph is a prime plus one, then the Pascal graph number of every node not being an extreme node is divisible by this prime. Further, we prove that, similar to binomial coefficients, the Pascal graph number of a node is equal to the sum of the Pascal graph numbers of this node in all subgraphs obtained by deleting precisely one of the extreme nodes of the graph, with the convention that the Pascal graph number of a node outside the subgraph is zero. It also holds that the Pascal graph numbers being normalized to sum up to one are the steady state probabilities of a Markov process in which at any node the process moves to a neighboring node with a probability proportional to the number of nodes connected to the node through this neighboring node. We also discuss properties of the Pascal graph numbers for arbitrary connected graphs. Because the Pascal graph number of a node in a connected graph is defined as the number of ways the graph can be constructed by a sequence of increasing connected subgraphs starting from the singleton subgraph on this node, the Pascal graph numbers can be seen as a measure of centrality in the graph.

The structure of this paper is as follows. In Section 2 we recall some well-known properties of binomial coefficients. The related properties when the rows of Pascal's triangle are considered as line-graphs are discussed in Section 3. In Section 4 the notion of Pascal graph numbers is introduced and we show that on the class of connected cycle-free graphs these numbers have properties that on the class of line-graphs reduce to the properties of binomial coefficients discussed in Sections 2 and 3. In Section 5 we show that the Pascal graph numbers being normalized to sum up to one are the steady state probabilities of a specific Markov process on the nodes of a graph. Section 6 is devoted to consideration of the Pascal graph numbers as a centrality measure for nodes in connected graphs. Properties of the Pascal graph numbers for arbitrary connected graphs are discussed in Section 7.

2 The binomial coefficients

For any two integers $n \geq 0$ and $0 \leq k \leq n$, the binomial coefficient C_n^k is given by

$$C_n^k = \binom{n}{k} = \frac{n!}{(n-k)!k!}. \tag{2.1}$$

Binomial coefficients have the property that for any two integers $n \geq 1$ and $0 \leq k \leq n$ it holds that

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k, \tag{2.2}$$

with the convention that $C_{n-1}^{k-1} = 0$ if $k = 0$ and $C_{n-1}^k = 0$ if $k = n$. This property is illustrated in Pascal's triangle, a triangular array in which for any integer $n \geq 0$ the row n consists of $n+1$ positions, $0, \ldots, n$, from left to right, with binomial coefficient C_n^k assigned to position $k = 0, \ldots, n$ on row n, and where this position is located to the right below position $k-1$ on row $n-1$ and to the left below position k on row $n-1$. The first eight rows of Pascal's triangle, corresponding to $n = 0, \ldots, 7$, together with the numbers C_n^k are depicted in Figure 1. Equation (2.2) says that for any integer $n \geq 1$ the binomial coefficient at position k on row n is equal to the sum of the binomial coefficients on the two positions on row $n-1$ diagonally to the left and to the right, where the coefficient is considered to be zero if the position is not in the triangle. From (2.1) immediately follows the well-known prime number property of the binomial coefficients that if n is a prime, then for $k = 1, \ldots, n-1$ the binomial coefficient C_n^k is divisible by this prime. This property is also illustrated in Figure 1 for $n = 2$, 3, 5, and 7. Moreover, from (2.1) it follows that for each integer $n \geq 1$ it holds that

$$\frac{C_n^k}{C_n^{k+1}} = \frac{k+1}{n-k}, \; k = 0, \ldots, n-1, \tag{2.3}$$

saying that for any two consecutive binomial coefficients C_n^k and C_n^{k+1} in row n of Pascal's triangle it holds that their ratio is equal to the ratio of the number $k+1$ of the positions $0, \ldots, k$ in that row from the position k to the left and the number $n-k$ of the positions $k+1, \ldots, n$ in that row from the position $k+1$ to the right. For example, for $n = 6$ and $k = 1$, we have $\frac{C_6^1}{C_6^2} = \frac{6}{15} = \frac{2}{5} = \frac{1+1}{6-1}$.

In the sequel, position k on row n in Pascal's triangle is denoted (n, k). It is well-known that for any integers $n \geq 0$ and $0 \leq k \leq n$ the binomial coefficient C_n^k can also be interpreted as the number of different paths in Pascal's triangle that start at the apex $(0, 0)$ and terminate at position (n, k), where at every step a path moves diagonally downwards to the next row either to the left or to the right. For instance, there are $C_7^3 = 35$ such paths from $(0, 0)$ to $(7, 3)$. Two of these paths are indicated in Figure 2 by the numbers with a star and with a plus correspondingly. Indeed, for any integer $n \geq 1$ the number

$$
\begin{array}{ccccccccccccccc}
&&&&&&&1&&&&&&& \\
&&&&&&1&&1&&&&&& \\
&&&&&1&&2&&1&&&&& \\
&&&&1&&3&&3&&1&&&& \\
&&&1&&4&&6&&4&&1&&& \\
&&1&&5&&10&&10&&5&&1&& \\
&1&&6&&15&&20&&15&&6&&1& \\
1&&7&&21&&35&&35&&21&&7&&1
\end{array}
$$

Figure 1: The first eight rows of Pascal's triangle.

of paths from the apex $(0, 0)$ to position (n, k) is equal to the total number of paths from $(0, 0)$ to the positions $(n-1, k-1)$ and $(n-1, k)$, where the number to a position is zero if the position is not in the triangle. From this it follows that the number of paths from $(0, 0)$ to position (n, k) meets condition (2.2) and, therefore, is equal to C_n^k. Obviously, the number of paths in Pascal's triangle that start at any given position (n, k) and move at every step diagonally upwards either to the left or to the right until the apex $(0, 0)$ is reached, is also equal to C_n^k.

$$
\begin{array}{ccccccccccccccc}
&&&&&&&1^{*+}&&&&&&& \\
&&&&&&1^{*}&&1^{+}&&&&&& \\
&&&&&1&&2^{*}&&1^{+}&&&&& \\
&&&&1&&3^{*}&&3&&1^{+}&&&& \\
&&&1&&4^{*}&&6&&4^{+}&&1&&& \\
&&1&&5^{*}&&10&&10^{+}&&5&&1&& \\
&1&&6&&15^{*}&&20^{+}&&15&&6&&1& \\
1&&7&&21&&35^{*+}&&35&&21&&7&&1
\end{array}
$$

Figure 2: Two of the paths from the apex $(0, 0)$ to position $(7, 3)$.

3 The binomial coefficients revisited on line-graphs

In this section we discuss the binomial coefficients within the framework of line-graphs. For a given finite nonempty set N, a graph on N is a pair (N, E) with N the set of nodes and $E \subseteq \{\{i, j\} \mid i, j \in N, j \neq i\}$ a set of edges between nodes. A graph (N, E) is connected if for each pair $i, j \in N$, $i \neq j$, there is a path from i to j in (N, E), i.e., for some $k \geq 1$ there exists a set of edges $\{i_h, i_{h+1}\}$, $h = 1, \ldots, k$, in E such that $i_1 = i$ and $i_{k+1} = j$. For a graph (N, E) and a subset $N' \subseteq N$, $E|_{N'}$ is the subset of edges of E on N' and $(N', E|_{N'})$ is the subgraph of (N, E) on N'. A subset of nodes $N' \subseteq N$ is connected in the graph (N, E) if the subgraph $(N', E|_{N'})$ is connected. A node $k \in N$ is an extreme node of a connected graph (N, E) if $|N| = 1$ or $N \backslash \{k\}$ is connected in (N, E). We denote the set of extreme nodes of a connected graph (N, E) by $S(N, E)$. If $\{i, j\} \in E$, then node j is a neighbor of node i in the graph (N, E). For a connected cycle-free graph (N, E) and node $k \in N$, $B^E_k = \{i \in N \mid \{i, k\} \in E\}$ is the set of neighbors of node k in (N, E) and the number of neighbors of node k in (N, E), denoted by $d_k(N, E)$, is degree of node k in (N, E). A connected graph (N, E) is a line-graph, or chain, if every node has at most two neighbors and $|E| = |N| - 1$, where $|A|$ is the cardinality of a finite set A. In the sequel, for a graph (N, E) and node $i \in N$, we denote $N \backslash \{i\}$ by N_{-i} and $E|_{N_{-i}}$ by E_{-i}.

Given a finite set N, $\Pi(N)$ denotes the set of linear orderings on N. For a connected graph (N, E) and node $k \in N$, a linear ordering $\pi \in \Pi(N)$, $\pi = (\pi_1, \ldots, \pi_{|N|})$, is feasible with respect to k in (N, E) if $\pi_1 = k$ and for $j = 2, \ldots, |N|$ the set of nodes $\{\pi_1, \ldots, \pi_j\}$ is connected in (N, E). The subset of all linear orderings feasible with respect to k in (N, E) is denoted by $\Pi^E_k(N)$ and its cardinality is denoted by $c_k(N, E)$.

For given integer $n \geq 0$, we may consider the $n + 1$ positions on row n of Pascal's triangle as nodes on the line-graph (N, E) with $N = \{0, \ldots, n\}$ as the set of nodes and $E = \{\{k, k+1\} \mid k = 0, \ldots, n-1\}$ as the set of edges, where to every node k, $k = 0, \ldots, n$, the binomial coefficient C^k_n is assigned. For row $n = 7$ this is illustrated in Figure 3, where the numbers below the line-graph indicate the nodes and the number above node k, $k = 0, \ldots, 7$, is the corresponding binomial coefficient C^k_7. Within this framework we make several observations.

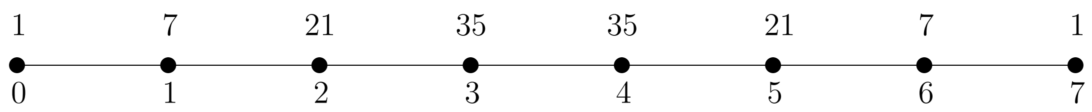

Figure 3: The binomial coefficients on the line-graph induced by row 7 of Pascal's triangle.

First, we give a new interpretation of the binomial coefficient C^k_n as the number of paths in Pascal's triangle from position (n, k) to apex $(0, 0)$. Given the above line-graph (N, E), for each linear ordering $\pi \in \Pi^E_k(N)$ it holds that for every $j = 2, \ldots, n + 1$ node π_j is the neighbor of the node either on the left end or on the right end of the connected set $\{\pi_1, \ldots, \pi_{j-1}\}$. From this it follows immediately that every $\pi \in \Pi^E_k(N)$ corresponds one-to-one to a path in Pascal's triangle from position (n, k) to $(0, 0)$, namely to the path along which we move upwards to the left (right), when π_j is the neighbor of the extreme node on the left (right) end of $\{\pi_1, \dots, \pi_{j-1}\}$, in total k times to the left and $n-k$ times to the right. For example, for $n = 7$ the linear ordering $\pi = (3, 2, 1, 4, 5, 6, 0, 7)$ is feasible with respect to node 3 and corresponds to the path from $(7, 3)$ to $(0, 0)$ depicted in Figure 2 by the numbers with a star. First, starting with node 3, node 2 is added to the left of node 3, corresponding to a move upwards to the left from position $(7, 3)$ to $(6, 2)$. Next, node 1 is added to the left of node 2, corresponding to a move upwards to the left from position $(6, 2)$ to $(5, 1)$. Then, node 4 is added to the right of node 3, corresponding to a move upwards to the right from position $(5, 1)$ to $(4, 1)$, and so on. From this interpretation it follows that for the line-graph (N, E) with $N = \{0, \dots, n\}$ and $E = \{\{k, k+1\} \mid k = 0, \dots, n-1\}$, for every $k \in N$ it holds that $c_k(N, E) = C^k_n$, i.e., the number of linear orderings on N that are feasible with respect to k in (N, E) is equal to the binomial coefficient C^k_n. This yields the following theorem.

Theorem 3.1 *For any two integers $n \geq 0$ and $0 \leq k \leq n$ it holds that*

$$C^k_n = |\{\pi \in \Pi(N) \mid \pi_1 = k, \{\pi_1, \dots, \pi_j\} \text{ is connected in } (N, E),\ j = 2, \dots, n+1\}|,$$

where (N, E) is the line-graph on $N = \{0, \dots, n\}$ with $E = \{\{k, k+1\} \mid k = 0, \dots, n-1\}$.

The theorem implies that the binomial coefficient C^k_n is equal to the number of ways the line-graph (N, E) can be constructed by starting with node k and adding at each step a node that is connected to one of the nodes that already have been added. Equivalently, C^k_n is the total number of ways that extreme nodes can be removed one by one from the graph until only the node k remains.

Second, we reconsider formula (2.2) within the framework of line-graphs. For the line-graph (N, E) defined above, consider the two line-subgraphs (N_{-0}, E_{-0}) and (N_{-n}, E_{-n}), both of which have n nodes and therefore correspond to row $n-1$ of Pascal's triangle. For every $k \in N_{-n} = \{0, \dots, n-1\}$ it holds that $c_k(N_{-n}, E_{-n}) = |\Pi^{E_{-n}}_k(N_{-n})| = C^k_{n-1}$, while for every $k \in N_{-0} = \{1, \dots, n\}$ it holds that $c_k(N_{-0}, E_{-0}) = |\Pi^{E_{-0}}_{k-1}(N_{-0})| = C^{k-1}_{n-1}$. Further, define $c_n(N_{-n}, E_{-n}) = 0$ and $c_0(N_{-0}, E_{-0}) = 0$. Since $c_k(N, E) = |\Pi^E_k(N)| = C^k_n$, the next result follows straightforwardly from equation (2.2).

Theorem 3.2 *Let (N, E) be the line-graph with $N = \{0, \dots, n\}$ and $E = \{\{k, k+1\} | k = 0, \dots, n-1\}$ for some integer $n \geq 1$. Then for any integer $0 \leq k \leq n$ it holds that*

$$c_k(N, E) = c_k(N_{-0}, E_{-0}) + c_k(N_{-n}, E_{-n}). \tag{3.4}$$

The theorem implies that the number of linear orderings that are feasible with respect to a node in the line-graph (N, E) is equal to the number of linear orderings that are feasible with respect to this node in the subgraph (N_{-0}, E_{-0}) without the extreme node

0, plus the number of linear orderings that are feasible with respect to this node in the subgraph (N_{-n}, E_{-n}) without the other extreme node n. The property is illustrated for $n = 7$ in Figure 4, where the numbers above the upper, middle, and lower line-graphs are the binomial coefficients assigned to the nodes in the graphs (N_{-0}, E_{-0}), (N_{-n}, E_{-n}), and (N, E), respectively, and the numbers below the lower line-graph indicate the nodes. For each node on the lower line-graph the binomial coefficient is equal to the sum of the numbers of that node in the upper and middle line-graphs.

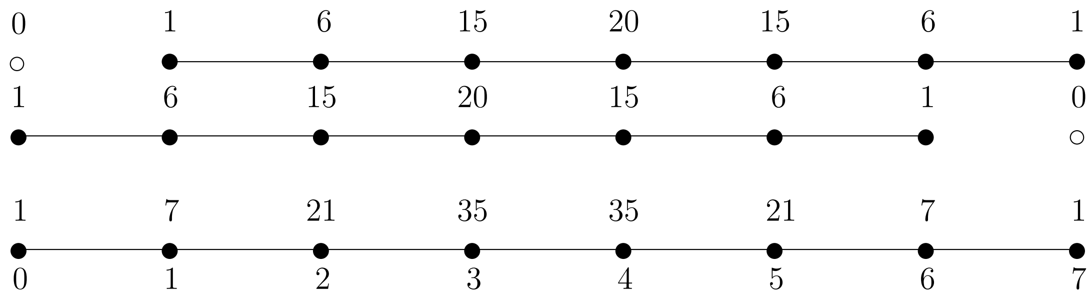

Figure 4: Illustration for $n = 7$ that the binomial coefficient of a node of a line-graph is equal to the sum of the binomial coefficients of this node in the two line-subgraphs without one of the extreme nodes.

4 Pascal graph numbers

In the previous section we defined for a connected graph (N, E) and node $k \in N$ the number $c_k(N, E) = |\Pi_k^E(N)|$ as the number of linear orderings π on N such that $\pi_1 = k$ and for $j = 2, \ldots, |N|$ the set $\{\pi_1, \ldots, \pi_j\}$ is connected in (N, E). We have seen that these numbers are the binomial coefficients on row $|N| - 1$ in Pascal's triangle when (N, E) is a line-graph. Therefore in the sequel we call these numbers Pascal graph numbers.

Definition 4.1 For a connected graph (N, E), the *Pascal graph number* of node $k \in N$ is the number $c_k(N, E)$.

For an arbitrary connected graph (N, E), the Pascal graph number of a node $k \in N$ is equal to the number of ways the graph can be constructed by starting with this node and adding at each step a node that is connected to one of the nodes that already have been added, or equivalently, it is the number of ways extreme nodes can be removed from the graph one by one until only the node k remains.

In this section we consider the class of connected cycle-free graphs and show that the properties of binomial coefficients discussed in the previous section for line-graphs generalize to similar properties of Pascal graph numbers for this class of graphs and therefore these numbers can be seen as a generalization of binomial coefficients.

We first introduce some notions with respect to connected cycle-free graphs. A connected graph (N, E) is cycle-free if for any pair $i, j \in N$, $i \neq j$, there is precisely one path from node i to node j. Note that in a connected cycle-free graph (N, E) there are precisely $|N| - 1$ edges. Given a connected cycle-free graph $(N.E)$ and node $k \in N$, for every neighbor $h \in B_k^E$, N_{kh}^E is the set of nodes $i \in N$ for which the unique path from node k to node i in (N, E) contains node h. We call N_{kh}^E a satellite of node k in (N, E). Each neighbor of k in (N, E) induces one satellite of k, so the number of satellites of k is equal to the number of neighbors of k in (N, E). For every $k \in N$ the satellites of node k in (N, E) form a partition of N_{-k} and, therefore, $\sum_{h \in B_k^E} |N_{kh}^E| = |N| - 1$. For any $k \in N$ and $h \in B_k^E$, we denote by (N_{kh}^E, E_{kh}) the subgraph of (N, E) on N_{kh}^E, where $E_{kh} = E|_{N_{kh}^E}$. Each of these subgraphs is connected and cycle-free.

Example 4.2 Throughout this section we take as example the graph (N, E) with eight nodes depicted in Figure 5. For node 4 of this graph it holds that $B_4^E = \{3, 5, 8\}$ and that $N_{43}^E = \{1, 2, 3, 7\}$, $N_{45}^E = \{5, 6\}$, and $N_{48}^E = \{8\}$ are its satellites. The three satellites of node 4 are depicted in Figure 6.

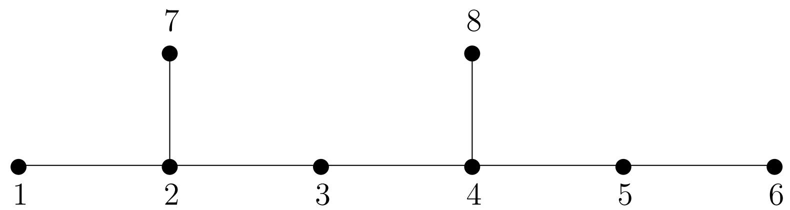

Figure 5: The graph (N, E) of Example 4.2.

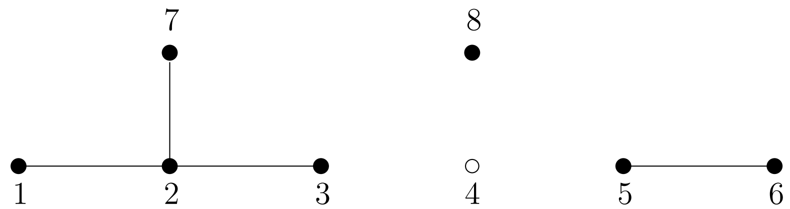

Figure 6: The three satellites of node 4 of the graph in Figure 5.

A connected cycle-free graph (N, E) with $|N| \geq 2$ has at least two extreme nodes and, moreover, a node is an extreme node of (N, E) if and only if it has precisely one neighbor in (N,E). For example, in Figure 3 it holds that $S(N,E) = \{0,7\}$ and in Figure 5 it holds that $S(N,E) = \{1,6,7,8\}$ and all these extreme nodes have just one neighbor. In general, a subgraph of a connected cycle-free graph (N,E) may not be connected, but for an extreme node h of (N,E) it holds by definition that the set N_{-h} is the unique satellite of node h in (N,E) and therefore the subgraph (N_{-h}, E_{-h}) is a connected cycle-free graph on N_{-h} with $|N_{-h}| = |N| - 1$ nodes.

We are now ready to consider the properties of the Pascal graph numbers on the class of connected cycle-free graphs. When the number of nodes is small, it is easy to calculate the Pascal graph number of a node by counting the number of linear orderings which are feasible with respect to that node.

Example 4.2 (continued) Consider node 4 in the subgraph on the set $N' = \{3,4,5,6,8\}$ of the graph in Figure 5. For any linear ordering π feasible with respect to node 4 in graph $(N', E|_{N'})$ we have $\pi_1 = 4$ and there are 12 feasible ways to place nodes 3, 5, 6 and 8 after node 4, because the positions of nodes 3, 5 and 8 can be chosen independently from each other and node 6 is chosen after node 5, but not necessarily directly after node 5. Hence, $c_4(N', E|_{N'}) = 12$. For any feasible ordering with respect to node 3 in the subgraph on N' we have that $\pi_1 = 3$, $\pi_2 = 4$, nodes 5 and 8 can be chosen independently from each other, and then node 6 after node 5. This yields $c_3(N', E|_{N'}) = 3$.

The next theorem generalizes (2.1) and relates the Pascal graph number of a node with the Pascal graph numbers of the neighboring nodes in the subgraphs on the satellites these nodes induce. For positive integers n_h, $h = 1, \ldots, k$, with sum equal to n, the multinomial coefficient $\binom{n}{n_1,\ldots,n_k}$ is given by

$$\binom{n}{n_1, \ldots, n_k} = \frac{n!}{\prod_{h=1}^{k} n_h!}.$$

Recall that for a connected cycle-free graph (N,E) and $k \in N$ it holds that $\sum_{h \in B_k^E} |N_{kh}^E| = |N| - 1$ and therefore the multinomial coefficient

$$\binom{|N|-1}{|N_{kh}^E|,\ h \in B_k^E} = \frac{(|N|-1)!}{\prod_{h \in B_k^E} |N_{kh}^E|!}$$

is well defined.

Theorem 4.3 *For any connected cycle-free graph (N,E) it holds that for every $k \in N$*

$$c_k(N,E) = \begin{cases} 1, & |N| = 1, \\ \binom{|N|-1}{|N_{kh}^E|,\ h \in B_k^E} \prod_{h \in B_k^E} c_h(N_{kh}^E, E_{kh}), & |N| \geq 2. \end{cases}$$

Proof. Clearly, $c_k(N,E) = 1$ if $N = \{k\}$. Suppose $|N| \geq 2$ and let $k \in N$. Since (N,E) is a connected graph on at least two nodes, k has at least one neighbor in (N,E) and therefore B_k^E is not empty. Moreover, since (N,E) is cycle-free, k is only connected to node h in the satellite N_{kh}^E, $h \in B_k^E$. Therefore, a linear ordering $\pi \in \Pi(N)$ is feasible with respect to k in (N,E) if and only if $\pi_1 = k$ and for every $h \in B_k^E$ the linear subordering of π on N_{kh}^E is feasible with respect to h in the subgraph (N_{kh}^E, E_{kh}). Hence, $c_k(N,E)$ is equal to the number of linear orderings $\pi \in \Pi(N)$ satisfying that $\pi_1 = k$ and for every $h \in B_k^E$ the linear subordering of π on N_{kh}^E is feasible with respect to h in (N_{kh}^E, E_{kh}). For each $h \in B_k^E$, there are $c_h(N_{kh}^E, E_{kh})$ linear orderings π^h on N_{kh}^E that are feasible with respect to h in the subgraph (N_{kh}^E, E_{kh}), which yields a total number of $\prod_{h \in B_k^E} c_h(N_{kh}^E, E_{kh})$ different linear orderings $(\pi^h, h \in B_k^E)$ on the satellites of k in (N,E). Since the satellites of a node in (N,E) are not connected to each other, the nodes of different satellites are unordered concerning feasibility with respect to k. Therefore, if for every $h \in B_k^E$ the linear ordering π^h on N_{kh}^E is feasible with respect to h in (N_{kh}^E, E_{kh}), then the number of linear orderings on N for which $\pi_1 = k$ and for all $h \in B_k^E$ its subordering on N_{kh}^E is π^h, is equal to in how many ways, for each $h \in B_k^E$, $|N_{kh}^E|$ nodes can be selected from $\sum_{h \in B_k^E} |N_{kh}^E| = |N| - 1$ nodes. This is precisely the multinomial coefficient $\binom{|N|-1}{|N_{kh}^E|,\ h \in B_k^E}$. Hence, the product of this latter multinomial coefficient and the number $\prod_{h \in B_k^E} c_h(N_{kh}^E, E_{kh})$ is equal to the number $c_k(N,E)$ of linear orderings on N that are feasible with respect to node k in the graph (N,E). □

The theorem says that in a connected cycle-free graph the Pascal graph number of a node is equal to the multinomial coefficient of the sizes of all its satellites times the product of the Pascal graph numbers of each of its neighbors in the subgraph on the satellite containing this neighbor. In case of a line-graph all these multinomials are binomials, because for every node there are (at most) two satellites, and moreover, for any node the Pascal graph number of each neighbor in the subgraph on the satellite containing this neighbor is equal to 1. This yields precisely (2.1).

From Theorem 4.3 we obtain straightforwardly the following two corollaries. The first one states that the Pascal graph number of an extreme node k in a connected cycle-free graph (N,E) with $|N| \geq 2$ is equal to the Pascal graph number that his unique neighbor h has in the subgraph (N_{-k}, E_{-k}). Recall that when k is an extreme node, then for his unique neighbor h it holds that $N_{kh}^E = N_{-k}$, and therefore $|N_{kh}^E| = |N| - 1$ and $\binom{|N|-1}{|N_{kh}^E|,\ h \in B_k^E} = 1$.

Corollary 4.4 *If $k \in N$ is an extreme node of a connected cycle-free graph (N,E) and $\{k,h\} \in E$, then $c_k(N,E) = c_h(N_{-k}, E_{-k})$.*

The second corollary shows that similar to binomial coefficients the Pascal graph numbers meet the prime number property.

Corollary 4.5 *If $|N| - 1$ is a prime number, then the Pascal graph number of any node of a connected cycle-free graph (N, E) other than an extreme node of the graph is divisible by this prime. Moreover, the Pascal graph number of any extreme node of this graph is not divisible by this prime.*

Theorem 4.3 provides an iterative procedure to find the Pascal graph numbers for connected cycle-free graphs. It shows that the Pascal graph number of a node can be calculated from the Pascal graph numbers of the neighboring nodes in the smaller subgraphs of the satellites. Clearly, for the latter numbers the same procedure can be applied and so on, making the satellite subgraphs smaller and smaller. For small enough subgraphs the number of feasible linear orderings is easy to compute, in particular it holds that eventually all satellites become line-graphs, on which the Pascal graph numbers are binomial coefficients.

Example 4.2 (continued) For node 2 in Figure 5 we obtain by Theorem 4.3 that

$$c_2(N, E) = \frac{7!}{1!\,1!\,5!}\; c_1(\{1\}, E|_{\{1\}}) \cdot c_7(\{7\}, E|_{\{7\}}) \cdot c_3(N', E_{N'}),$$

where $N' = \{3, 4, 5, 6, 8\}$. Clearly, $c_1(\{1\}, E|_{\{1\}}) = c_7(\{7\}, E|_{\{7\}}) = 1$, and applying Corollary 4.4 we find that $c_3(N', E|_{N'}) = c_4(N'_{-3}, E|_{N'_{-3}}) = 3$, because the subgraph on $N'_{-3} = \{4, 5, 6, 8\}$ is a line-graph. Hence,

$$c_2(N, E) = \frac{7!}{1!\,1!\,5!} \cdot 1 \cdot 1 \cdot 3 = 42 \cdot 3 = 126.$$

Similar, it holds that

$$c_4(N, E) = \frac{7!}{4!\,2!\,1!} \cdot c_3(N'', E|_{N''}) \cdot c_5(\{5, 6\}, E|_{\{5,6\}}) \cdot c_8(\{8\}, E|_{\{8\}}),$$

where $N'' = \{1, 2, 3, 7\}$. Clearly, $c_8(\{8\}, E|_{\{8\}}) = c_5(\{5, 6\}, E|_{\{5,6\}}) = 1$, and, again by Corollary 4.4, we find that $c_3(N'', E|_{N''}) = c_2(N''_{-3}, E|_{N''_{-3}}) = 2$. Hence,

$$c_4(N, E) = \frac{7!}{4!\,2!\,1!} \cdot 2 \cdot 1 \cdot 1 = 210.$$

Note that according to Corollary 4.5 both $c_2(N, E) = 126$ and $c_4(N, E) = 210$ are divisible by the prime number $|N| - 1 = 7$. For the extreme node 1 we have that

$$c_1(N, E) = c_2(N_{-1}, E_{-1}) = \frac{6!}{1!5!} c_7(\{7\}, E|_{\{7\}}) \cdot c_3(N', E|_{N'}) = 6 \cdot 1 \cdot 3 = 18,$$

which is according to Corollary 4.5 not divisible by 7.

Example 4.6 Let (N, E) be the star graph given by $N = \{0, \ldots, n\}$ and $E = \{\{0, h\} \mid h = 1, \ldots, n\}$, in which each node $k \neq 0$ is connected to the hub at node 0. From Theorem 4.3 it follows that

$$c_0(N, E) = n!,$$

because $|N| - 1 = n$, $B_0^E = \{1, \ldots, n\}$ and for $h = 1, \ldots, n$ it holds that $c_h(N_{0h}^E, E_{0h}) = 1$ since $N_{0h}^E = \{h\}$. Further, because each node h, $h = 1, \ldots, n$, is an extreme node connected to only node 0 and the subgraph on its unique satellite $N_{h0}^E = N_{-h}$ is also a star graph with hub node 0, but having in total n nodes, it follows from Corollary 4.4 that for all $h = 1, \ldots, n$,

$$c_h(N, E) = c_0(N_{-h}, E_{-h}) = (n-1)!.$$

Note that $c_0(N, E) = nc_h(N, E)$ for all $h \in N_{-0}$. So, in a star graph the Pascal graph number of the hub is equal to the sum of the Pascal graph numbers of all other nodes.

Next, let (N, E) be a generalized star graph given by $N = \{0, \ldots, n\}$ with the hub at node 0 having as neighbors nodes $m_1, \ldots, m_k$, that is the graph (N, E) for which for every $h = 1, \ldots, k$ the subgraph on the satellite $N_{0m_h}^E$ of node 0 is a line-graph with n_h nodes having node m_h as an extreme node. Hence, $c_{m_h}(N_{0m_h}^E, E_{0m_h}) = 1$ for $h = 1, \ldots, k$ and $\sum_{h=1}^k n_h = n$. Then, from Theorem 4.3 it follows that

$$c_0(N, E) = \binom{n}{n_1, \ldots, n_k}.$$

Therefore, in a generalized star graph the Pascal graph number of the hub is equal to the multinomial coefficient for the numbers of nodes in each of the satellites of the hub.

The next theorem is a consequence of Theorem 4.3. The theorem states that for a connected cycle-free graph (N, E) the ratio between the Pascal graph numbers of any two neighbors in the graph is equal to the ratio of the numbers of nodes in the two subgraphs that result from deleting the edge between these two nodes. For the line-graph (N, E) with $N = \{0, \ldots, n\}$ and $E = \{(k, k+1)|\ k = 0, \ldots, n-1\}$ this result reduces to (2.3).

Theorem 4.7 *For any connected cycle-free graph (N, E) and $\{k, h\} \in E$ it holds that*

$$\frac{c_k(N, E)}{c_h(N, E)} = \frac{|N_{hk}^E|}{|N_{kh}^E|}.$$

Proof. According to Theorem 4.3 and since $|N| \geq 2$, it holds that

$$c_k(N, E) = \binom{|N| - 1}{|N_{k\ell}^E|,\ \ell \in B_k^E} \prod_{\ell \in B_k^E} c_\ell(N_{k\ell}^E, E_{k\ell}) \tag{4.5}$$

and

$$c_h(N,E) = \binom{|N|-1}{|N^E_{h\ell}|,\ \ell \in B^E_h} \prod_{\ell \in B^E_h} c_\ell(N^E_{h\ell}, E_{h\ell}). \tag{4.6}$$

Since $\{k,h\} \in E$, it holds that both $h \in B^E_k$ and $k \in B^E_h$. Moreover, since the graph (N,E) is cycle-free, the sets $N^E_{h\ell}$, $\ell \in B^E_h\backslash\{k\}$, are the satellites of node h in the subgraph (N^E_{kh}, E_{kh}), while the sets $N^E_{k\ell}$, $\ell \in B^E_k\backslash\{h\}$, are the satellites of node k in the subgraph (N^E_{hk}, E_{hk}). Therefore, using again Theorem 4.3,

$$c_h(N^E_{kh}, E_{kh}) = \begin{cases} 1, & |N^E_{kh}| = 1, \\ \binom{|N^E_{kh}|-1}{|N^E_{h\ell}|,\ \ell \in B^E_h\backslash\{k\}} \prod_{\ell \in B^E_h\backslash\{k\}} c_\ell(N^E_{h\ell}, E_{h\ell}), & |N^E_{kh}| \geq 2, \end{cases} \tag{4.7}$$

and

$$c_k(N^E_{hk}, E_{hk}) = \begin{cases} 1, & |N^E_{hk}| = 1, \\ \binom{|N^E_{hk}|-1}{|N^E_{k\ell}|,\ \ell \in B^E_k\backslash\{h\}} \prod_{\ell \in B^E_k\backslash\{h\}} c_\ell(N^E_{k\ell}, E_{k\ell}), & |N^E_{hk}| \geq 2. \end{cases} \tag{4.8}$$

Substituting (4.7) into (4.5) and (4.8) into (4.6) yields

$$\frac{c_k(N,E)}{c_h(N,E)} = \frac{(|N^E_{kh}|-1)!/|N^E_{kh}|!}{(|N^E_{hk}|-1)!/|N^E_{hk}|!} = \frac{|N^E_{hk}|}{|N^E_{kh}|}.$$

□

This theorem implies that if the Pascal graph number of one node is known, the Pascal graph numbers of the other nodes can be calculated by successive application of the ratio property. Starting from the node for which the Pascal graph number is known, the Pascal graph numbers of the other nodes follow in any linear ordering which is feasible with respect to the initial node. The next result immediately follows from Theorem 4.7.

Corollary 4.8 *If in a connected cycle-free graph the deletion of an edge splits the graph in two subgraphs having the same number of nodes, then irrespective to the structure of the two subgraphs obtained, the two nodes adjacent to that edge have equal Pascal graph numbers. Moreover, the Pascal graph number of any other node is smaller.*

Note that in Pascal's triangle indeed $C^{k-1}_n = C^k_n$ holds for $k = \frac{1}{2}(n+1)$ when n is odd.

Example 4.2 (continued) For the graph in Figure 5 we found above that $c_4(N,E) = 210$. Since the deletion of the edge $\{3,4\}$ yields two subgraphs with four nodes in each, due to Corollary 4.8 we obtain that $c_3(N,E) = c_4(N,E) = 210$. Next, by Theorem 4.7, we get

$$c_2(N,E) = \frac{3}{5}c_3(N,E) = 126,$$

which we also found above. Continuing this way we find

$$c_1(N,E) = c_7(N,E) = \frac{1}{7}c_2(N,E) = 18$$

and

$$c_5(N,E) = \frac{2}{6}c_4(N,E) = 70,\;\; c_6(N,E) = \frac{1}{7}c_5(N,E) = 10,\;\; c_8(N,E) = \frac{1}{7}c_4(N,E) = 30.$$

To summarize, the nodes 3 and 4 have equal and maximal Pascal graph numbers and the Pascal graph numbers of the extreme nodes are not divisible by 7, whereas these numbers for all the other nodes are divisible by 7. All Pascal graph numbers are given in Figure 7.

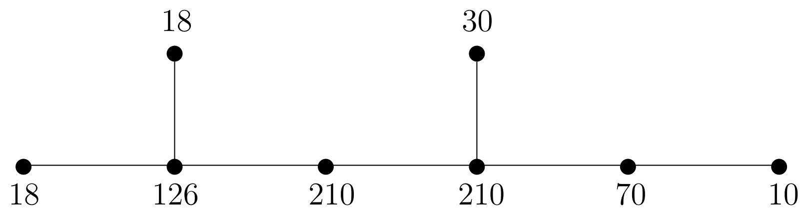

Figure 7: The Pascal graph numbers for the graph in Figure 5.

The next theorem generalizes formula (3.4) and states that the Pascal graph number of a node in a connected cycle-free graph is equal to the sum of the Pascal graph numbers of that node in all subgraphs obtained by deleting one of the extreme nodes from the graph.

Theorem 4.9 *For any connected cycle-free graph* (N,E) *it holds that for every* $k \in N$

$$c_k(N,E) = \begin{cases} 1, & |N| = 1, \\ \sum_{h \in S(N,E)} c_k(N_{-h}, E_{-h}), & |N| \geq 2, \end{cases}$$

where $c_h(N_{-h}, E_{-h}) = 0$ *for all* $h \in S(N,E)$.

Proof. Clearly, $c_k(N,E) = 1$ if $N = \{k\}$. Suppose $n = |N| \geq 2$ and let $k \in N$. Since (N,E) is a connected cycle-free graph on at least two nodes, for every linear ordering $\pi \in \Pi_k^E(N)$ it holds that $\pi_n \in S(N,E)$. Clearly, for any $h \in S(N,E)$, the linear ordering $\pi = (\pi_1, \ldots, \pi_{n-1}, h)$ on N is feasible with respect to node k in (N,E) if and only if the linear ordering $(\pi_1, \ldots, \pi_{n-1})$ on N_{-h} is feasible with respect to node k in the subgraph (N_{-h}, E_{-h}). Hence, for every $h \in S(N,E)$, $h \neq k$, $c_k(N_{-h}, E_{-h})$ is precisely the number of linear orderings π in $\Pi_k^E(N)$ satisfying $\pi_n = h$, which proves the theorem. $\square$

Theorem 4.9 gives a third iterative procedure for finding the Pascal graph numbers by starting with the calculation of the Pascal graph numbers of the nodes in the subgraphs of small size and increasing successively their sizes. When $|N| = 1$, the Pascal graph number of the unique node is 1. For the (unique) graph for $|N| = 2$ from Theorem 4.9 we get that the Pascal graph number of any of the two nodes is equal to the sum of their Pascal graph numbers in the two subgraphs with one of the nodes (with a number equal to zero when the node is not in the graph), which gives Pascal graph number 1 for both nodes. For $|N| = 3$ the unique cycle-free graph is (still) a line-graph, with the Pascal graph numbers equal to the sum of the numbers in the two (line-)subgraphs with two nodes obtained by leaving out one of the two extreme nodes. There are two types of cycle-free graphs for $|N| = 4$. Type 1 is the line-graph with four nodes and we have again that the Pascal graph numbers are the sum of the numbers in the two line-graphs with three nodes obtained by leaving out one of the two extreme nodes. Type 2 is a star graph with one hub node and three extreme nodes. From Example 4.6 we then know that the Pascal graph number of the hub node is equal to 6 and each extreme node has Pascal graph number 2. This also follows by applying Theorem 4.9. Assuming that node 1 is the hub and the nodes 2, 3 and 4 are the extreme nodes, it follows from (4.9) that for any node $k = 1, 2, 3, 4$ it holds that

$$c_k(N, E) = \sum_{h \in \{2,3,4\}} c_k(N_{-h}, E_{-h}).$$

Since every of the three subgraphs in this summation is a line-graph with three nodes and corresponding Pascal graph numbers 1, 2 and 1, with in any of the three graphs the number 2 for the hub 1, and the Pascal graph number is zero if $h = k$, it follows that $c_1(N, E) = 6$ and $c_h(N, E) = 2$ for $h = 2, 3, 4$. Continuing in this way we can find for any graph (N, E), the Pascal graph numbers for any connected cycle-free subgraph with $|N| - 1$ nodes. Then the Pascal graph numbers for the nodes in (N, E) follow from adding up their Pascal graph numbers on all these subgraphs, where a number is zero if the node is not in the subgraph.

Example 4.2 (continued) The graph in Figure 5 has four extreme nodes, nodes 1, 6, 7, and 8. The Pascal graph numbers on the four subgraphs obtained by deleting precisely one of these extreme nodes are given in Figure 8, with number zero for the node deleted from the graph. By applying Theorem 4.9, for each node the sum of the Pascal graph numbers in the four subgraphs in Figure 8 is equal to the Pascal graph number of this node in the graph, as depicted in Figure 7.

5 The Pascal graph numbers and steady state probabilities

In this section we show that when normalizing the sum of the Pascal graph numbers of the nodes of a cycle-free connected graph to one we get the steady state probabilities of a

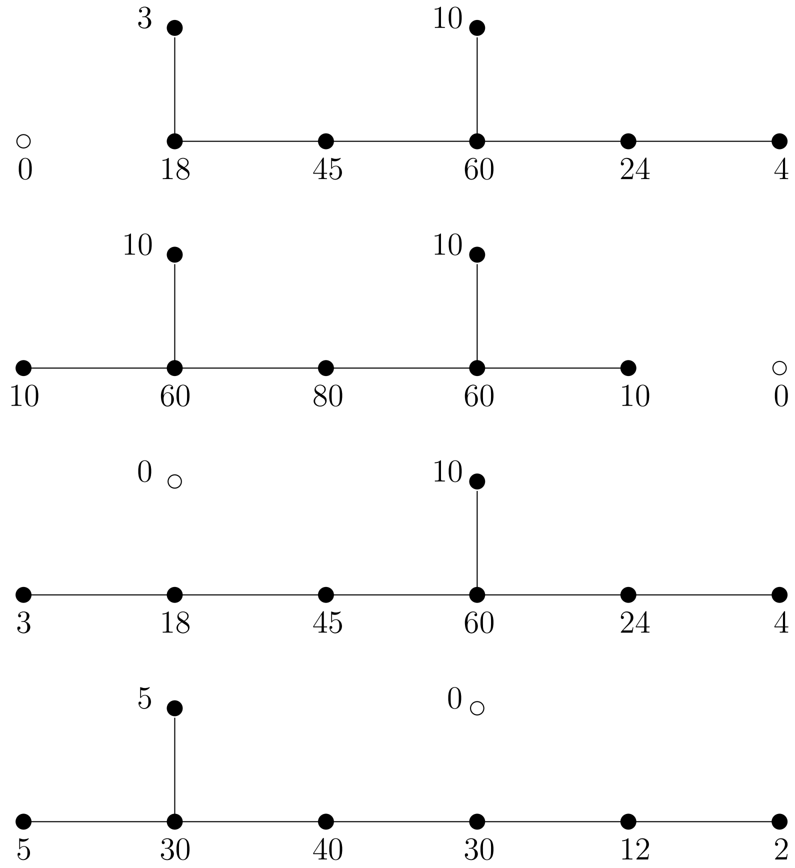

Figure 8: The Pascal graph numbers on the four subgraphs (N_{-h}, E_{-h}), $h = 1, 6, 7, 8$, of the graph in Figure 5.

Markov chain with the set of nodes as the states. For an integer $n \geq 1$, we first consider row n of Pascal's triangle represented by the line-graph (N, E) with $N = \{0, \ldots, n\}$ and $E = \{\{k, k+1\} \mid k = 0, \ldots, n-1\}$. Let $s^n \in \mathbb{R}^N$ be the row vector of corresponding binomial coefficients, i.e., $s^n_k = C^k_n$ for $k = 0, \ldots, n$, and let P^n be the $(n+1) \times (n+1)$ transition matrix defined by

$$P^n = \begin{bmatrix} 0 & 1 & 0 & . & . & . & . & 0 \\ \frac{1}{n} & 0 & \frac{n-1}{n} & 0 & . & . & . & 0 \\ 0 & \frac{2}{n} & 0 & \frac{n-2}{n} & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & . & . & . & 0 & \frac{n-1}{n} & 0 & \frac{1}{n} \\ 0 & . & . & . & . & 0 & 1 & 0 \end{bmatrix},$$

i.e., for $k, j = 0, \ldots, n$ the (k, j)th element of the matrix P^n is given by $p^n_{kj} = \frac{k}{n}$ if $j = k-1$, $p^n_{kj} = \frac{n-k}{n}$ if $j = k+1$, and $p^n_{kj} = 0$ otherwise. Since in the line-graph (N, E) induced by the row n of Pascal's triangle there are k nodes to the left of node k, those with index smaller than k, and $n-k$ nodes to the right of node k, those with index larger than k, the transition probability from node k to any of its neighbors is proportional to the number of nodes that are connected to k through that neighbor. The next theorem follows from straightforward calculations.

Theorem 5.1 *For any integer $n \geq 1$ it holds that $s^n P^n = s^n$, i.e., for every $k \in N$ the normalized binomial coefficient $C^k_n/2^n$ is the steady state probability that the Markov process with transition matrix P^n is in state k.*

The theorem implies that for every positive integer n the row vector s^n of binomial coefficients C^k_n, $k = 0, \ldots, n$, is a left eigenvector of the matrix P^n and therefore the vector $s^n/2^n$ gives the stationary state distribution of the Markov chain with transition matrix P^n. Thus, the binomial coefficients yield the relative probabilities that the Markov process is in each state.

Next, we show that also for a connected cycle-free graph the Pascal graph numbers determine the stationary distribution of a Markov process with the set of nodes as the set of states. For a connected cycle-free graph (N, E) with $|N| \geq 2$, let $s^E \in \mathbb{R}^N$ be the $|N|$-dimensional row vector of Pascal graph numbers with $s^E_k = c_k(N, E)$ the Pascal graph number of node $k \in N$. Further, let P^E be the $|N| \times |N|$ transition matrix with for $k, h \in N$ the (k, h)th element given by

$$p^E_{kh} = \begin{cases} \dfrac{|N^E_{kh}|}{|N| - 1}, & \{k, h\} \in E, \\ 0, & \text{otherwise.} \end{cases}$$

When being in state (node) $k \in N$, the process goes with probability $\frac{|N^E_{kh}|}{|N|-1}$ to neighboring state $h \in N$ and with zero probability to any non-neighbor, i.e., the probabilities p^E_{kh}, $h \in B^E_k$, are proportional to the number of nodes that are connected to node k in (N, E) through node h. Then we have the following theorem, which generalizes Theorem 5.1 to the class of all connected cycle-free graphs.

Theorem 5.2 *For any connected cycle-free graph (N, E) with $|N| \geq 2$ it holds that $s^E P^E = s^E$, i.e., for every $k \in N$ the normalized Pascal graph number $c_k(N, E)/\sum_{h \in N} c_h(N, E)$ is the steady state probability that the Markov process with transition matrix P^E is in state k.*

Proof. Since (N, E) is a connected cycle-free graph with $|N| \geq 2$, for every $k \in N$ it holds that $\sum_{h \in B_k^E} |N_{kh}^E| = |N|-1$. From Theorem 4.7 it follows that $|N_{kh}^E| = |N_{hk}^E| c_h(N, E)/c_k(N, E)$ for all $h \in B_k^E$ and $k \in N$. Therefore, for every $k \in N$,

$$\sum_{h \in B_k^E} |N_{hk}^E| c_h(N, E) = (|N| - 1) c_k(N, E).$$

Dividing both sides by $|N| - 1$ yields for every $k \in N$

$$\sum_{h \in B_k^E} s_h^E p_{hk}^E = s_k^E.$$

□

The theorem states that for a connected cycle-free graph (N, E) with $|N| \geq 2$ the row vector s^E of Pascal graph numbers is a left eigenvector of the transition matrix P^E and therefore the vector $s^E / \sum_{k \in N} s_k^E$ gives the stationary state distribution of the Markov chain. Thus when normalizing the Pascal graph numbers to sum up to one, for any $k \in N$ the normalized Pascal graph number $c_k(N, E)/ \sum_{h \in N} c_h(N, E)$ of node k is the long-term probability that the process is in state k.

It is well-known that the degrees, when normalized to sum equal to one, are the steady state probabilities of the Markov process that in any node moves with equal probability to each of its neighbors. This property also holds for connected graphs that are not cycle-free.[1]

6 Pascal graph numbers as a centrality measure

Each linear ordering feasible with respect to some fixed node in a connected graph induces a way to construct the graph by a sequence of increasing connected subgraphs starting from the singleton subgraph determined by this node. This gives grounds to consider the Pascal graph number of a node in a given connected graph as a measure of centrality of the node in the graph. A centrality measure answers the question which nodes in a graph under consideration are important. In fact, it gives a complete or partial ordering of the nodes with respect to importance, cohesiveness, or influence. Formally, let $\mathcal{G}$ be the collection of all connected graphs. Then a centrality measure is a function f on $\mathcal{G}$ which assigns to each connected graph $(N, E) \in \mathcal{G}$ a vector $f(N, E) \in \mathbb{R}^N$ with entries $f_i(N, E)$, $i \in N$. The entry $f_i(N, E)$ measures the centrality of node i in graph (N, E). The higher $f_i(N, E)$ is, the higher the influence of node i within the graph. A well-known centrality measure is the degree measure which assigns to any graph the vector of degrees of its nodes.

[1]It is easy to verify that for any connected graph (N, E) it holds that $dP^E = d$, where d is the vector of degrees $d_k(N, E)$, $k \in N$, and P^E is a transition matrix with elements $p_{kh}^E = 1/d_k(N, E)$ when h is a neighbor of k, and zero otherwise.

We define the *connectivity centrality measure* as the mapping c on $\mathcal{G}$ that assigns to each connected graph $(N, E) \in \mathcal{G}$ the vector $c(N, E) \in \mathbb{R}^N$ of the Pascal graph numbers of its nodes. For each node in a given connected graph it measures in how many ways the graph can be generated when starting with this node and adding one by one the other nodes which are connected to at least one node that already has been added. The connectivity centrality measure is illustrated by the binomial coefficients on the line-graph in Figure 3 and the Pascal graph numbers on the connected cycle-free graph in Figure 7. Also, as shown in Example 4.6, for a star graph with $n + 1$ nodes, the connectivity centrality of the hub is n times as large as the connectivity centrality of each of the n extreme nodes and is therefore equal to the sum of the connectivity centrality of all other nodes. For a star graph this property holds for many centrality measures, for instance, also for the degree measure. Example 4.6 also shows that in a generalized star graph the connectivity centrality of the hub is equal to the multinomial coefficient of the sizes of the subgraphs connected to the hub.

In the literature it is quite standard to characterize centrality measures by a number of their properties (axioms). We show below that the connectivity centrality measure on the subclass of cycle-free connected graphs, denoted by $\widehat{\mathcal{G}}$, can be characterized by the three following properties.

Single node normalization A centrality measure f on $\mathcal{G}$ satisfies single node normalization if $f_k(N, E) = 1$ when $N = \{k\}$.

Ratio property A centrality measure f on $\widehat{\mathcal{G}}$ satisfies the ratio property if for every $(N, E) \in \widehat{\mathcal{G}}$ and edge $\{k, h\} \in E$ it holds that $\frac{f_k(N,E)}{f_h(N,E)} = \frac{|N^E_{hk}|}{|N^E_{kh}|}$.

Extreme node consistency A centrality measure f on $\widehat{\mathcal{G}}$ satisfies extreme node consistency if for every $(N, E) \in \widehat{\mathcal{G}}$ with $|N| \geq 2$ and extreme node $k \in S(N, E)$ it holds that $f_k(N, E) = f_h(N_{-k}, E_{-k})$, where h is the unique neighbor of node k in (N, E).

Because in a singleton connected graph there is just one node, only this node is important to measure centrality in the graph. This importance is normalized to be equal to one.[2] To the best of our knowledge the ratio property does not hold for any centrality measure known in the literature, nevertheless it seems to be rather natural. It states that the ratio of centralities of two neighboring nodes k and h is equal to the number of nodes including node h for which node h is on the path to node k divided by the number of nodes including

[2]Note that the degree measure does not satisfy single node normalization, because the degree of a node in a graph with one node is equal to zero, saying that the importance of a node in a singleton connected graph is zero. It seems to be more natural to define this value to be positive.

node k for which node k is on the path to node h. Consistency properties are quite usual in the literature on characterization of functions, for instance, in the theory of solutions for cooperative games. Here it states that if in a graph a node is connected to only one other node, then its centrality is the same as the centrality of this connected node in the subgraph without the node. We now have the following result.

Theorem 6.1 *A centrality measure f on the class of cycle-free connected graphs $\widehat{\mathcal{G}}$ satisfies single node normalization, the ratio property, and extreme node consistency if and only if it is the connectivity centrality measure.*

Proof. We prove by induction that the centrality measure f determined by the three properties is unique and yields the Pascal graph numbers on each connected cycle-free graph, i.e., $f(N,E) = c(N,E)$ for all $(N,E) \in \widehat{\mathcal{G}}$. First, the single node normalization uniquely determines the Pascal graph numbers on (N,E) when $|N| = 1$. Next, suppose the three properties uniquely determine the Pascal graph numbers on each $(N,E) \in \widehat{\mathcal{G}}$, i.e., $f(N,E) = c(N,E)$, when $|N| \leq n-1$ for some $n \geq 2$. Take some graph $(N,E) \in \widehat{\mathcal{G}}$ with $|N| = n$. Since (N,E) is connected and cycle-free, it has at least one extreme node. Let k be any extreme node of (N,E) and let node h be the unique neighbor of k in (N,E). Since $|N| \geq 2$ this unique neighbor exists. By extreme node consistency it holds that $f_k(N,E) = c_h(N_{-k}, E_{-k})$. From Corollary 4.4 it follows that $f_k(N,E)$ is the Pascal graph number of node k in (N,E). By repeated application of the ratio property for some linear ordering that is feasible with respect to k in (N,E) we obtain numbers $f_j(N,E)$ for every $j \neq k$. Since $f_k(N,E)$ is the Pascal graph number of node k in (N,E), it follows from Theorem 4.7 that for every $j \neq k$ the number $f_j(N,E)$ is the Pascal graph number of the node j in (N,E). Therefore, $f(N,E) = c(N,E)$. $\square$

Note that in the proof the determination of the numbers $f_j(N,E)$ of every node $j \in N$ is independent of the choice of the extreme node k in (N,E).

7 Pascal graph numbers on arbitrary connected graphs

In Section 4 the Pascal graph numbers are defined on the class of all connected graphs and on the subclass of connected cycle-free graphs certain properties of the binomial coefficients are generalized to similar properties of the Pascal graph numbers. In this section we discuss whether or not these properties can be generalized also to properties of the Pascal graph numbers on the class of all connected graphs.

We first reconsider Theorem 4.3. For a connected graph (N,E) and subset $N' \subseteq N$ with $E' = E|_{N'}$, N'/E' denotes the collection of maximal connected subsets of N' in (N,E), called components of N' in (N,E). For a cycle-free graph (N,E) and node $k \in N$, the components of N_{-k} are the satellites of node k in (N,E).

For an arbitrary connected graph (N,E), $k \in N$ and $C \in N_{-k}/E_{-k}$, the extended subgraph of (N,E) on C with respect to node k is the graph (C, E^k_C) on C with

$$E^k_C = E|_C \cup \{\{i,j\} \subseteq C \mid i \neq j,\ \{i,k\} \in E \text{ and } \{j,k\} \in E\}.$$

So, when two different nodes i and j in C do not form an edge in (N,E) but both form an edge with node k, then edge $\{i,j\}$ is added to the subgraph $(C, E|_C)$. Now, Theorem 4.3 generalizes as follows. The proof follows straightforwardly along the lines of the proof of Theorem 4.3 and is therefore omitted.

Theorem 7.1 *For any connected graph (N,E) it holds that for every $k \in N$*

$$c_k(N,E) = \begin{cases} 1, & |N| = 1, \\ \binom{|N|-1}{|C|,\ C \in N_{-k}/E_{-k}} \prod_{C \in N_{-k}/E_{-k}} \sum_{h \in B^E_k \cap C} c_h(C, E^k_C), & |N| \geq 2. \end{cases}$$

In case $k \in N$ is an extreme node of (N,E) and thus the collection of components N_{-k}/E_{-k} only contains N_{-k} as its unique element, the expression for $c_k(N,E)$ reduces to the following generalization of Corollary 4.4.

Corollary 7.2 *If k is an extreme node of a connected graph (N,E) with $|N| \geq 2$, then*

$$c_k(N,E) = \sum_{h \in B^E_k} c_h(N_{-k}, E^k_{N_{-k}}).$$

In case $k \in N$ is not an extreme node of (N,E), node k is an extreme node of the set $C_{+k} = C \cup \{k\}$ for any component C of N_{-k} in (N,E). From Theorem 7.1 and the previous corollary we obtain that for that case $c_k(N,E)$ can be expressed as follows.

Corollary 7.3 *If k is not an extreme node of a connected graph (N,E) with $|N| \geq 2$, then*

$$c_k(N,E) = \binom{|N|-1}{|C|,\ C \in N_{-k}/E_{-k}} \prod_{C \in N_{-k}/E_{-k}} c_k(C_{+k}, E|_{C_{+k}}).$$

The latter expression can also be used to express the Pascal graph number of a node that is not an extreme node of a cycle-free connected graph. In that case a satellite C of k in (N,E) is equal to N^E_{kh} with $h \in B^E_k$ being the unique node in C connected to node k, i.e., $C_{+k} = N^E_{kh} \cup \{k\}$, and therefore $c_k(C_{+k}, E|_{C_{+k}}) = c_h(N^E_{kh}, E_{kh})$.

From the last corollary it follows that the first part of Corollary 4.5 still holds. When $|N| - 1$ is a prime number, then the Pascal graph number of any node that is not an extreme node of a connected graph (N,E) on N is divisible by this prime. In case the graph contains cycles, however, it might be that the Pascal graph number of an extreme node is divisible by this prime. For example, if (N, E) is the complete graph, then every node is an extreme node and its Pascal graph number is equal to $(|N| - 1)!$.

When (N, E) is cycle-free, then for any edge $\{k, h\} \in E$ the graph $(N, E \setminus \{\{k, h\}\})$ consists of the two components N^E_{hk} and N^E_{kh} and the ratio property of Theorem 4.7 applies. When (N, E) contains cycles, the ratio property still holds for any edge $\{k, h\} \in E$ which is a bridge in (N, E), i.e., deleting the edge $\{k, h\}$ from E splits the remaining graph in two disconnected subgraphs, (N^E_{kh}, E_{kh}) containing h as a node and (N^E_{hk}, E_{hk}) containing k as a node.

Theorem 7.4 *For any connected graph (N, E) and bridge $\{k, h\} \in E$, it holds that*

$$\frac{c_k(N, E)}{c_h(N, E)} = \frac{|N^E_{hk}|}{|N^E_{kh}|}.$$

Note that in a cycle-free connected graph every edge is a bridge. If the graph (N, E) contains cycles and the edge $\{k, h\} \in E$ is not a bridge, then the graph $(N, E\setminus\{\{k, h\}\})$ is still connected and the ratio property does not apply. Since the ratio property may not hold in case of graphs with cycles, Theorems 5.2 and 6.1 cannot be generalized to the class of connected graphs.

Finally, Theorem 4.9 holds for any connected graph. The proof goes along the same lines of the proof of Theorem 4.9, because for any linear ordering $\pi \in \Pi(N)$ that is feasible with respect to a node $k \in N$ in a connected graph (N, E) with $|N| \geq 2$ it holds that $\pi_{|N|}$ is an extreme node of (N, E).

Theorem 7.5 *For any connected graph (N, E) it holds that for every $k \in N$*

$$c_k(N, E) = \begin{cases} 1, & |N| = 1, \\ \sum\limits_{h \in S(N,E)} c_k(N_{-h}, E_{-h}), & |N| \geq 2, \end{cases}$$

where $c_h(N_{-h}, E_{-h}) = 0$ for all $h \in S(N, E)$.

In Figure 9 we illustrate this theorem by a connected graph with four nodes and a cycle on three of the nodes.

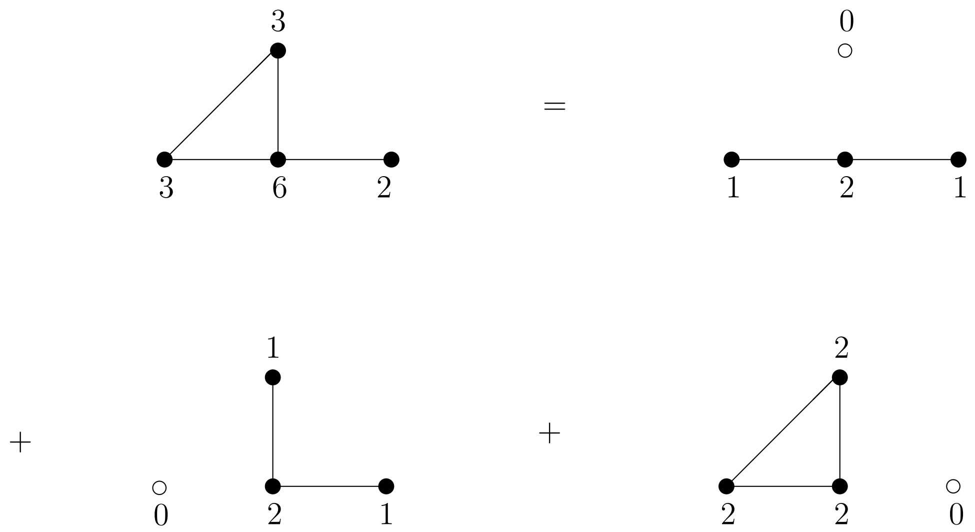

Figure 9: Illustration of Theorem 7.5.