Revista de Matemática: Teoría y Aplicaciones 2016 **23**(2) : 409–420
cimpa – ucr   issn: 1409-2433 (Print), 2215-3373 (Online)

A COMBINATORIAL PROBLEM ON A DIRECTED GRAPH

UN PROBLEMA COMBINATORIO BASADO EN UN GRAFO ORIENTADO

Osvaldo Marrero*   Paul C. Pasles†

Received: 13 May 2015; Revised: 14 Apr 2016;
Accepted: 22 Apr 2016

---

*Department of Mathematics and Statistics, Villanova University, Villanova, Pennsylvania, U.S.A. E-Mail: Osvaldo.Marrero@villanova.edu
†Misma dirección que/*Same address as*: O. Marrero. E-Mail: Paul.Pasles@villanova.edu

Abstract

We consider two options for a particle's entire journey through a certain directed graph. Both options involve a random assignment to the journey route to be followed. We are interested in the option that offers, on average, the shortest route. Therefore, we determine the average journey length for each of the two options. As part of our analysis, we prove some combinatorial identities that appear to be new. Some suggestions for further work are given.

Keywords: directed graphs; games on graphs; combinatorial identities; combinatorial probability.

Resumen

Se consideran dos opciones para la jornada total de una partícula que se desplaza a través de un cierto grafo orientado. Bajo ambas opciones, la ruta de la jornada es asignada aleatoriamente. Nos interesa saber la opción bajo la cual uno espera la ruta más corta. Por eso, para cada opción, determinamos la esperanza matemática del largo de la ruta. Al parecer novedosas, algunas identidades combinatorias son demostradas como parte de nuestro análisis. Para concluir, mencionamos varias oportunidades para futuros estudios.

Palabras clave: grafos orientados; juegos en grafos; identidades combinatorias; probabilidad combinatoria.

Mathematics Subject Classification: 05C20, 05C57, 05A19, 60C05.

1 The problem

Along the edges of a certain directed graph, a particle moves from an *initial vertex* v_0 to a *terminal vertex* v_T, which is different from v_0. In the directed graph there are m different paths that have lengths $x_1, \ldots, x_m$. Each path is made up of n equal-length edges. For each $i = 1, \ldots, m$, we set $y_i := x_i/n$; that is, y_i is the length of each of the n equal-length edges that make up the path of length x_i. All having the same direction, the paths originate at v_0 and end at v_T. Moreover, between v_0 and v_T, the equal-length edges meet at an additional $n-1$ distinct *intermediate vertices*, which serve as interchanges between paths. There are no more vertices and no other edges; thus, the graph consists of $n+1$ vertices and mn edges. An example with $m = 4$ and $n = 3$ is shown in Figure 1, where we have written the edge length y_i next to the corresponding edge.

For the particle's journey, we only wish to consider two possible choices: *stay-on-path* or *variable-path*.

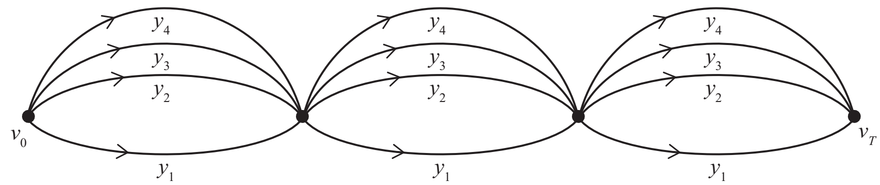

Figure 1: The directed graph when $m = 4$ and $n = 3$.

- Under the stay-on-path choice, the m paths are equally likely, and, at random, the particle is assigned to one path for the entire journey. For example, with respect to Figure 1, we have $m = 4$ possible paths having lengths $x_1, \ldots, x_4$. Moreover, each path is divided into $n = 3$ equal-length edges; therefore, in this case, for each $i = 1, \ldots, 4$, $x_i = 3y_i$. Thus, if the particle is assigned to the path with length x_2, then the journey length is $x_2 = 3y_2$.

- Under the variable-path choice, we regard the equal-length edges as being ordered, with the first being the edge closest to the initial vertex, and so on. Then the particle's journey proceeds through an ordered sequence of n edges whose lengths may be different. Note that changing between paths is only possible at the intermediate vertices, and any such change does not add any length to the path actually followed by the particle. Thus, the total journey length is the sum of the lengths of the edges that make up the actual journey path. For notation, for example, with respect to Figure 1, when $m = 4$ and $n = 3$, we write (y_3, y_1, y_1) to signify the following: the first journey edge is the first edge of the path having length x_3; the second journey edge is the second edge of the path having length x_1; the third journey edge is the third edge of the path having length x_1; therefore, in this case, the journey length is $2y_1 + y_3$. In general, there are m^n equiprobable possible journey paths, and the particle is assigned to one of these at random. In Figure 1, for example, some possible assignments are $(y_2, y_1, y_3), (y_3, y_3, y_3), (y_4, y_1, y_4)$, etc.

The problem is this: Between the two choices, stay-on-path or variable-path, we wish to choose the one that minimizes the particle's journey length. Therefore, we ask the following question: On average, which of the two choices has the smaller journey length? We invite the reader to pause here for a moment, and to guess the answer.

The paper is virtually self-contained. Nevertheless, we give two references: [1] for graph theory and [2] for probability.

2 The solution

To answer our question, we set out to compute the mean or expected journey distance for each of the two choices. Along the way, we prove some combinatorial identities that seem to be new.

Let the random variables $D_{\mathrm{SOP}(m,n)}$ and $D_{\mathrm{VP}(m,n)}$ denote the total journey distance to be traversed, respectively, under the stay-on-path choice and under the variable-path choice. We want to compute the expected values $E[D_{\mathrm{SOP}(m,n)}]$ and $E[D_{\mathrm{VP}(m,n)}]$ for each of the two choices.

2.1 Stay-on-path

Under the stay-on-path choice, $D_{\mathrm{SOP}(m,n)}$ is a discrete uniform random variable whose possible values are $x_1, \ldots, x_m$, each having probability $1/m$. Therefore, in this case, the expected value

$$\begin{aligned} E[D_{\mathrm{SOP}(m,n)}] &= \left(\frac{1}{m}\right) x_1 + \cdots + \left(\frac{1}{m}\right) x_m \\ &= \left(\frac{1}{m}\right)(ny_1 + \cdots + ny_m) \\ &= \left(\frac{n}{m}\right)(y_1 + \cdots + y_m). \end{aligned}$$

2.2 Variable-path

Under the variable-path choice, an actual journey route is an ordered list of n edges, the first being the one closest to the initial vertex, the second being the next closest, and so on, through the nth edge which ends at the terminal vertex. Thus, in this case, the circumstances are somewhat more complicated than under the stay-on-path choice. Therefore, we introduce some notation to help facilitate the discussion.

We regard a possible value for the random variable $D_{\mathrm{VP}(m,n)}$ as the result of the composition of two functions, `lengths()` followed by `sum()`. The value of the function `lengths()` is the ordered list of the n edge lengths that correspond to the similarly ordered list of edges that make up an actual journey route; and then `sum()` adds up the lengths of the pertinent journey edges. Thus, `lengths()` is a vector-valued function

$$\texttt{lengths}\colon \{y_1, \ldots, y_m\} \to \{y_1, \ldots, y_m\}^n$$

$$\texttt{lengths}\left(\{y_1, \ldots, y_m\}\right) := \left(y_{j_1}, \ldots, y_{j_n}\right),$$

where $y_{j_1}, \ldots, y_{j_n} \in \{y_1, \ldots, y_m\}$, and y_{j_i} is the length of the ith edge in the journey route.

To simplify the notation, we let

$$\mathcal{L} := \{\texttt{lengths}\,(\{y_1, \ldots, y_m\})\} = \{y_1, \ldots, y_m\}^n,$$

the set of possible values or codomain of the function `lengths`(). Then `sum`() is a real-valued function

$$\texttt{sum}\colon \mathcal{L} \to \mathbb{R}.$$

Thus, if $\texttt{lengths}\,(\{y_1, \ldots, y_m\}) = (y_{j_1}, \ldots, y_{j_n})$, the corresponding value of the random variable $D_{\mathrm{VP}(m,n)}$ is

$$\begin{aligned} D_{\mathrm{VP}(m,n)} &= \texttt{sum}\,(\texttt{lengths}\,(\{y_1, \ldots, y_m\})) \\ &= \texttt{sum}\,((y_{j_1}, \ldots, y_{j_n})) := y_{j_1} + \cdots + y_{j_n}. \end{aligned}$$

To illustrate, when $m = 2$ and $n = 3$, the possible values for `lengths`(), `sum`(), and $D_{\mathrm{VP}(m=2,n=3)}$ are given in Table 1.

Table 1: Example when $m = 2$ and $n = 3$.

`lengths`()	`sum`() = value of $D_{\mathrm{VP}(m=2,n=3)}$
(y_1, y_1, y_1)	$3y_1$
(y_1, y_1, y_2)	$2y_1 + y_2$
(y_1, y_2, y_1)	$2y_1 + y_2$
(y_2, y_1, y_1)	$2y_1 + y_2$
(y_1, y_2, y_2)	$y_1 + 2y_2$
(y_2, y_1, y_2)	$y_1 + 2y_2$
(y_2, y_2, y_1)	$y_1 + 2y_2$
(y_2, y_2, y_2)	$3y_2$

In general, we have m possible different paths, each divided into n equal-length edges, and so it is clear that there are m^n possible, pairwise distinct values for the function `lengths`(); that is, $\mathcal{L}$ has cardinality $\operatorname{card}(\mathcal{L}) = m^n$. Therefore, there are m^n possible values for `sum`(), but these values are not necessarily pairwise distinct, as shown, for example, in Table 1 when $m = 2$ and $n = 3$. Because the m^n possible routes are equiprobable, it follows that, under the variable-path choice, $D_{\mathrm{VP}(m,n)}$ is a discrete random variable whose m^n equiprobable values are the possible, but not necessarily distinct, values of the function `sum`(). Thus, we see that

$$E[D_{\mathrm{VP}(m,n)}] = \sum_{(y_{j_1}, \ldots, y_{j_n}) \in \mathcal{L}} \left(\frac{1}{m^n}\right) \texttt{sum}\,((y_{j_1}, \ldots, y_{j_n}))\,,$$

an expression that is correct, but not as informative or explicit as we would like. Hence, we now proceed to clarify the value of $E[D_{\mathrm{VP}(m,n)}]$. A value of `sum()` involves each of $y_1, \ldots, y_m$ a nonnegative number of times. Therefore, we can write

$$E[D_{\mathrm{VP}(m,n)}] = \left(\frac{1}{m^n}\right) \{f(y_1)y_1 + \cdots + f(y_m)y_m\}\,, \tag{1}$$

where, for each $j = 1, \ldots, m$, $f(y_j)$ is the number of times that y_j appears in the expression (1) for $E[D_{\mathrm{VP}(m,n)}]$; that is, $f(y_j)$ is the total number of times that y_j appears throughout all the elements $(y_{j_1}, \ldots, y_{j_n}) \in \mathcal{L}$. Our goal now is to determine the value of $f(y_j)$ for each $j = 1, \ldots, m$, and we do so by means of the Counting Algorithm below.

Table 2: First of two tables pertinent to the counting algorithm.

COLUMN 1 number of times y_j appears in the ordered list $(y_{j_1}, \ldots, y_{j_n})$	COLUMN 2 corresponding number of possible different ordered lists $(y_{j_1}, \ldots, y_{j_n})$ containing that many y_js	COLUMN 3 number of places not containing y_j that remain in the ordered list $(y_{j_1}, \ldots, y_{j_n})$
n	$\binom{n}{n}$	0
$n-1$	$\binom{n}{n-1}$	1
$n-2$	$\binom{n}{n-2}$	2
$n-3$	$\binom{n}{n-3}$	3
$\vdots$	$\vdots$	$\vdots$
1	$\binom{n}{n-(n-1)} = \binom{n}{1}$	$n-1$
0	$\binom{n}{n-n} = \binom{n}{0}$	n

Table 3: Second of two tables pertinent to the counting algorithm.

COLUMN 1	COLUMN 2	COLUMN 3	COLUMN 4
number of places not containing y_j that remain in the ordered list $(y_{j_1}, \ldots, y_{j_n})$	number of times y_j appears in the ordered list $(y_{j_1}, \ldots, y_{j_n})$	corresponding total number of possible different ordered lists $(y_{j_1}, \ldots, y_{j_n})$ containing that many y_js	corresponding sum total of the number of y_js throughout all such ordered lists $(y_{j_1}, \ldots, y_{j_n})$
0	n	$\binom{n}{n}(m-1)^0$	$(n-0)\binom{n}{n}(m-1)^0$
1	$n-1$	$\binom{n}{n-1}(m-1)^1$	$(n-1)\binom{n}{n-1}(m-1)^1$
2	$n-2$	$\binom{n}{n-2}(m-1)^2$	$(n-2)\binom{n}{n-2}(m-1)^2$
3	$n-3$	$\binom{n}{n-3}(m-1)^3$	$(n-3)\binom{n}{n-3}(m-1)^3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n-1$	1	$\binom{n}{n-(n-1)}(m-1)^{n-1}$	$1\binom{n}{n-(n-1)}(m-1)^{n-1}$
n	0	$\binom{n}{n-n}(m-1)^n$	$0\binom{n}{n-n}(m-1)^n$

Begin: Counting Algorithm

Objective: For each $j = 1, \ldots, m$, count how many times y_j appears in $\mathcal{L}$, the set of all possible ordered lists $(y_{j_1}, \ldots, y_{j_n})$, where $y_{j_1}, \ldots, y_{j_n} \in \{y_1, \ldots, y_m\}$; that is, the objective is to determine the value of $f(y_j)$.

Step 1 Fix both a value $j \in \{1, \ldots, m\}$ and the corresponding y_j, and consider an ordered list $(y_{j_1}, \ldots, y_{j_n}) \in \mathcal{L}$. At this moment, we regard the n places in that ordered list as being “empty".

Step 2 Note the possible number of times that y_j can appear in $(y_{j_1}, \ldots, y_{j_n})$. Each such possible number is given in column 1 of Table 2, and defines a row in that table. At this point, we think of $(y_{j_1}, \ldots, y_{j_n})$ as containing only y_js, and the remaining places, if any, are “empty".

Step 3 For each entry in column 1 of Table 2, count the corresponding number of possible different ordered lists $(y_{j_1}, \ldots, y_{j_n}) \in \mathcal{L}$ that contain that many y_js. That number is given in column 2 of

Table 2. In column 3 of that table we note the corresponding number of "empty" places in $(y_{j_1}, \ldots, y_{j_n})$ yet to be filled in with elements from $\{y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_m\}$.

Step 4 We now pay attention to the ways in which the "empty places" in $(y_{j_1}, \ldots, y_{j_n})$ can be filled in with the $m-1$ elements in $\{y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_m\}$. In each row of columns 1 and 2 of Table 3 we note, respectively, the corresponding number of "empty places" and the number of y_js already in $(y_{j_1}, \ldots, y_{j_n})$. In column 3 of Table 3 we note the corresponding total number of different ordered lists $(y_{j_1}, \ldots, y_{j_n})$ that exist. Then in column 4 of Table 3, we record the corresponding total frequency of y_js throughout all such ordered lists $(y_{j_1}, \ldots, y_{j_n})$; that frequency is, of course, the product of the corresponding entries in columns 2 and 3 of that table.

Step 5 To obtain the overall total number of times that y_j appears in $\mathcal{L}$, we now add all the entries in column 4 of Table 3. The result is the value of $f(y_j)$.

End: Counting Algorithm

Thus, it follows from the counting algorithm that, for each $j = 1, \ldots, m$,

$$f(y_j) = \sum_{k=0}^{n} (n-k) \binom{n}{n-k} (m-1)^k \,, \tag{2}$$

which we now proceed to simplify. In Equation (2), the last term in the sum is equal to zero, and, therefore, it may be ignored. Ignoring that term, we have

$$\begin{aligned}
\sum_{k=0}^{n-1} (n-k) \binom{n}{n-k} (m-1)^k &= \sum_{k=0}^{n-1} (n-k) \frac{n!}{(n-k)!k!} (m-1)^k \\
&= \sum_{k=0}^{n-1} n \frac{(n-1)!}{(n-1-k)!k!} (m-1)^k \\
&= n \sum_{k=0}^{n-1} \binom{n-1}{k} (m-1)^k \\
&= n \sum_{k=0}^{n-1} \binom{n-1}{k} (m-1)^k (1)^{n-1-k} \\
&= n \left[\{(m-1)+1\}^{n-1} \right] \\
&= n \left(m^{n-1} \right);
\end{aligned} \tag{3}$$

therefore, from Equation (2), we now obtain, for $j = 1, \ldots, n$,

$$f(y_j) = n\left(m^{n-1}\right); \tag{4}$$

also, from (3), we have the identities

$$\begin{aligned}\sum_{k=0}^{n}(n-k)\binom{n}{n-k}(m-1)^k &= \sum_{k=0}^{n-1}(n-k)\binom{n}{n-k}(m-1)^k \\ &= n\left(m^{n-1}\right). \end{aligned} \tag{5}$$

Returning to Table 3, we see that, when we add all the entries in column 3 of that table, we are counting the number of ordered lists in $\mathcal{L}$; as noted earlier, this number is m^n, of course. Thus,

$$\sum_{k=0}^{n}\binom{n}{n-k}(m-1)^k = m^n; \tag{6}$$

this identity can also be obtained by using techniques similar to those used above in Display (3).

Therefore, from (5) and (6), we have:

Theorem 1. *If $m > 1$ and $n > 1$ are integers, then*

$$\sum_{k=0}^{n}(n-k)\binom{n}{n-k}(m-1)^k = \sum_{k=0}^{n-1}(n-k)\binom{n}{n-k}(m-1)^k = n\left(m^{n-1}\right)$$

and $\displaystyle\sum_{k=0}^{n}\binom{n}{n-k}(m-1)^k = m^n.$ □

When $m = n$, then $n\left(m^{n-1}\right) = n^n$. This gives us the following result.

Corollary. *For each integer $n > 1$, we have the following combinatorial identities:*

$$\begin{aligned}\sum_{k=0}^{n}(n-k)\binom{n}{n-k}(n-1)^k &= \sum_{k=0}^{n-1}(n-k)\binom{n}{n-k}(n-1)^k \\ &= \sum_{k=0}^{n}\binom{n}{n-k}(n-1)^k = n^n. \end{aligned}$$ □

Substituting (4) in (1), we see that

$$
\begin{aligned}
E[D_{\mathrm{VP}(m,n)}] &= \left(\frac{1}{m^n}\right)\left\{n\left(m^{n-1}\right)y_1 + \cdots + n\left(m^{n-1}\right)y_m\right\} \\
&= \left(\frac{n}{m}\right)(y_1 + \cdots + y_m),
\end{aligned}
$$

as is the case under the stay-on-path choice. We can now answer our question thus:

Theorem 2. *On average, both choices have the same journey distance; explicitly, for either journey choice, the expected values* $E[D_{\mathrm{SOP}(m,n)}] = E[D_{\mathrm{VP}(m,n)}] = \left(\frac{n}{m}\right)(y_1 + \cdots + y_m)$. □

3 Concluding remarks

Usually an important distribution parameter, the mean generally does not determine a probability distribution. In particular, in our problem, the random variables $D_{\mathrm{SOP}(m,n)}$ and $D_{\mathrm{VP}(m,n)}$ have the same mean, but they actually follow different distributions. For instance, when $m = 2$ and $n = 3$, the distributions of $D_{\mathrm{SOP}(m=2,n=3)}$ and $D_{\mathrm{VP}(m=2,n=3)}$ are completely determined by the following tables:

$D_{\mathrm{SOP}(m=2,n=3)}$	$3y_1$	$3y_2$
prob	$\frac{1}{2}$	$\frac{1}{2}$

and

$D_{\mathrm{VP}(m=2,n=3)}$	$3y_1$	$2y_1 + y_2$	$y_1 + 2y_2$	$3y_2$
prob	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

.

4 Future work

For further study, we are content with briefly discussing just four possibilities. It's clear that much more could be done.

4.1 Same directed graph and journey options; further examine the distributions of the two total journey distances

We have the same directed graph and the same two options for the particle's journey. For each of $D_{\mathrm{SOP}(m,n)}$ and $D_{\mathrm{VP}(m,n)}$, we only wanted to determine the mean. To continue our investigation, one would naturally want to evaluate other distribution parameters. In particular, it would be interesting to know the variances. To gain a little insight, we used the distribution tables in Section 3 to compute

$$\mathrm{Var}\left[D_{\mathrm{SOP}(m=2,n=3)}\right] = \left(\frac{3}{2}\right)^2 (y_1 - y_2)^2 \qquad \text{and}$$
$$\mathrm{Var}\left[D_{\mathrm{VP}(m=2,n=3)}\right] = \frac{3}{4}(y_1 - y_2)^2 .$$

Thus, the variance of $D_{\mathrm{SOP}(m=2,n=3)}$ is exactly three-times greater than that of $D_{\mathrm{VP}(m=2,n=3)}$. Based on this minuscule evidence, we are tempted to conjecture that

$$\mathrm{Var}\left[D_{\mathrm{SOP}(m,n)}\right] = n\,\mathrm{Var}\left[D_{\mathrm{VP}(m,n)}\right] ;$$

is this conjecture true? Moreover, we ask for explicit, closed-form expressions for each of $\mathrm{Var}\left[D_{\mathrm{SOP}(m,n)}\right]$ and $\mathrm{Var}\left[D_{\mathrm{VP}(m,n)}\right]$.

4.2 Same directed graph; always switch at an intermediate vertex

We have the same directed graph, but the options for the journey are different. Upon arrival at an intermediate vertex, the particle must change path. Then, for example, we could study and compare two choices: (1) the particle is not allowed to return to any edge of a path where it has already been and (2) the particle may return to an edge of a previously traversed path, but only after it has been on another path.

4.3 Same directed graph; paths divided differently

We have the same directed graph, but each path is divided into edges of unequal length. There are, of course, many ways to effect such path division. A number of choices for the particle's journey may then be specified.

4.4 Different directed graph

Our directed graph can be modified, for example, so that not all paths meet at every intermediate vertex. Consequently, a path may then be divided into edges of unequal length. Of course, several options may then be prescribed for the particle's changing between paths.

Acknowledgement

The authors thank the referees for their comments.

References

[1] Chartrand, G.; Lesniak, L.; Zhang, P. (2010) *Graphs & Digraphs*, 5th edition. CRC Press, Boca Raton FL.

[2] Grimmett, G.; Welsh, D. (2014) *Probability: An Introduction*, 2d edition. Oxford University Press, Oxford, United Kingdom.