Liet. mat. rink. LMD darbai, **48/49**, 2008, 338–342

An empirical study of the structure of the shortest path tree

Mindaugas BLOZNELIS, Irmantas RADAVIČIUS (VU)
e-mail: mindaugas.bloznelis@mif.vu.lt, irmantas.radavicius@mif.vu.lt

Abstract. We consider a complete graph on n vertices. Edges of the graph are prescribed random positive weights $X_1, X_2, \dots, X_m$. Here $m = \binom{n}{2}$. We assume that these random variables are independent and have the common probability distribution with density function $f(x)$, $x \geqslant 0$. Given a vertex v let T denote the shortest path tree with root v. Let $T_1, T_2, \dots \subset T$ denote the trees that are obtained from T after removal of the root v. Let $N_1 \geqslant N_2 \geqslant N_3 \geqslant \dots$ denote (ordered) sequence of sizes of these trees. We study statistical properties of this sequence for various densities f and large n.

Keywords: shortest path tree, weighted graph, random weights.

1. Introduction

1. How does a typical output of Dijkstra's algorithm look like? We are interested in the structure of the typical shortest path tree. We report results of the simulation study of several classes of weighted graphs.

We consider a weighted complete graph with the vertex set $V = \{v_1, \dots, v_n\}$ and the edge set $E = \{e_1, \dots, e_m\}$, where every edge e_j is prescribed a positive weight (length) X_j, $1 \leqslant j \leqslant m$. Here $m = \binom{n}{2}$.

We say that the path consisting of edges $e_{j_1}, \dots, e_{j_k}$ has length $X_{j_1} + \dots + X_{j_k}$. Given a vertex $v \in V$, let $T = T(v)$ denote the shortest path tree with the root v (i.e., T is a spanning tree with the property that for every vertex $u \in V \setminus \{v\}$ the length of the path of the tree T connecting u and v is not greater than the length of any other path of the graph connecting u and v). Let $v_1^*, v_2^*, \dots$ denote the neighbors of v in the tree T numbered according to their distances to v, i.e., v_1^* is the closest vertex, v_2^* is the second closest etc. In the tree T, the vertices that can be reached from v through (intermediate vertex) v_i^* make a tree with the root v_i^*. We denote this tree $T_i = T_i(v)$. In particular, $T - v$ splits in to the trees $T_1, T_2, \dots$ with the roots $v_1^*, v_2^*, \dots$ respectively. Let $N_i := |T_i|$ denote the number of vertices of T_i. It is reasonable to expect that $N_1 \geqslant N_2 \geqslant \dots$.

We are interested in statistical (typical) properties of the sequence $(N_1, N_2, \dots)$.

We assume that the weights $X_1, X_2, \dots$ are drawn independently at random from the common distribution with a density function $f(x)$, $x \geqslant 0$. The corresponding weighted complete graph is denoted $K_n(f)$. Choosing various densities f we study distributions of values of parameters $N_1, N_2, \dots$.

2. The shortest path tree with exponential weights ($f(x) = e^{-x}$, $x \geqslant 0$) is studied in [1] and [2]. It was shown that given i, asymptotically as $n \to \infty$ the average size of the tree T_i is approximately $n/2^i$, for every fixed i.

One may guess that for large n the statistical properties of the numbers $N_1, N_2, \dots$ depend heavily on the behavior of the density function $f(x)$ in a neighborhood of 0, say $|x| < \varepsilon$, whereas the behavior of $f(x)$ outside this neighborhood ($|x| > \varepsilon$) has little influence. Indeed the $n-1$ edges of the tree T are among those receiving the smallest weights X_j. On the other hand, it is known that the statistical properties of the smallest sample values $X_{1:m} \leqslant X_{2:m} \leqslant \dots \leqslant X_{i:m} \leqslant \dots$ are described by the behavior of $f(x)$ in a neighborhood of the point $x = 0$.

3. We are interested whether and how the behavior of (continuous) density function $f(x)$ in a vicinity of $x = 0$ affects the distribution of values of N_i. It is convenient to consider the cases (i) $\lim_{x\to 0} f(x) > 0$; (ii) $\lim_{x\to 0} f(x) = 0$; (iii) $\lim_{x\to 0} f(x) = \infty$.

Simulation study suggests that in the case (i) for large n we have $\mathbf{E}N_j \approx n/2^j$, $j = 1, 2, \dots$.

The asymptotic distribution of N_j in the case (ii) depends heavily on the rate of decay of $f(x) \to 0$ as $x \to 0$. We consider densities of the form $f_{(\alpha)}(x) = a_\alpha x^\alpha$, $0 < x < b_\alpha$, where $\alpha > 0$.

Similarly, the asymptotic distribution of N_j in the case (iii) depends heavily on the rate of decay of $1/f(x) \to 0$ as $x \to 0$. We consider densities of the form $f_{(-\alpha)}(x) = a_\alpha x^{-\alpha}$, $0 < x < b_\alpha$, where $0 < \alpha < 1$.

Tables and graphs describing the results of simulation are given in Section 2.

2. Simulation results

We generate 1000 independent copies of the complete graph $K_n(f)$ with random edge weights $X_1, \dots, X_m$. Here $n = 1000$ denotes the number of nodes and $f(x)$ denotes the density function of the probability distribution of iid weights $X_1, \dots, X_m$.

Let $K_n^{(j)}(f)$ denote the j-th copy of $K_n(f)$. Therefore, $K_n^{(1)}(f), \dots, K_n^{(1000)}(f)$ are independent and identically distributed weighted complete graphs with random edge weights.

Fix $v \in V$. For $j = 1, \dots, 1000$, let $N_i^{(j)}$ denote the size of the tree $T_i(v)$ in $K_n^{(j)}(f)$.

Table A gives average values of observed values $N_i = (N_i^{(1)} + \dots + N_i^{(1000)})/1000$ for $i = 1, 2, \dots, 5$.

Graph B shows the histogram of the observed values $N_1^{(1)}, \dots, N_1^{(1000)}$.

Table A

i	N_i/n
1	0.5243
2	0.2386
3	0.1163
4	0.0600
5	0.0313

Graph B

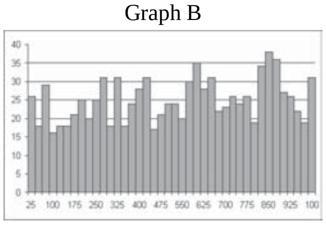

Fig. 1. $f(x) = x + \frac{1}{2}$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.5116
2	0.2455
3	0.1296
4	0.0538
5	0.0305

Graph B

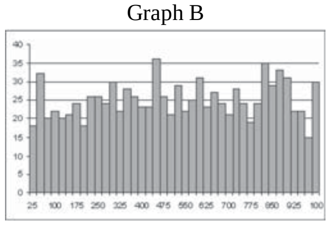

Fig. 2. $f(x) = 1$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.4944
2	0.2561
3	0.1285
4	0.0599
5	0.0298

Graph B

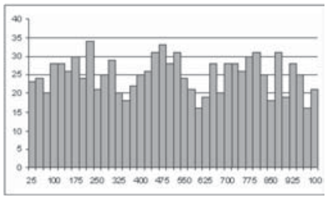

Fig. 3. $f(x) = -x + \frac{3}{2}$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.3370
2	0.1868
3	0.1153
4	0.0796
5	0.0586

Graph B

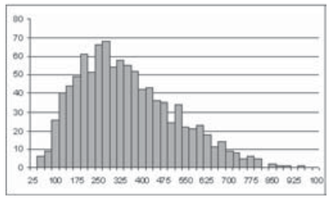

Fig. 4. $f(x) = \frac{2}{9}x$, $x \in [0; 3)$.

Table A

i	N_i/n
1	0.3282
2	0.1806
3	0.1149
4	0.0830
5	0.0589

Graph B

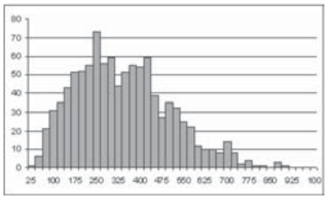

Fig. 5. $f(x) = \frac{1}{2}x$, $x \in [0; 2)$.

Table A

i	N_i/n
1	0.3290
2	0.1848
3	0.1197
4	0.0783
5	0.0595

Graph B

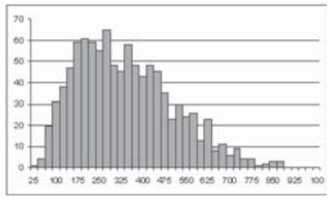

Fig. 6. $f(x) = 2x$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.3271
2	0.1849
3	0.1194
4	0.0826
5	0.0581

Graph B

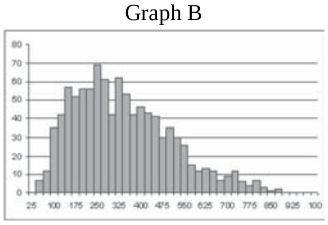

Fig. 7. $f(x) = \frac{9}{2}x$, $x \in [0; \frac{2}{3})$.

Table A

i	N_i/n
1	0.7031
2	0.2259
3	0.0601
4	0.0087
5	0.0022

Graph B

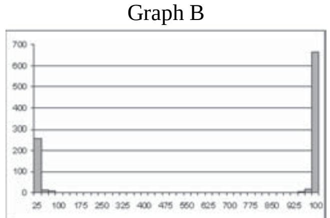

Fig. 8. $f(x) = \frac{1}{32}x^{-\frac{31}{32}}$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.6088
2	0.2618
3	0.0931
4	0.0272
5	0.0058

Graph B

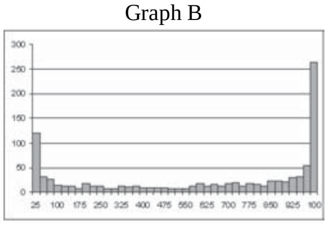

Fig. 9. $f(x) = \frac{1}{2}x^{-\frac{1}{2}}$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.5152
2	0.2530
3	0.1177
4	0.0593
5	0.0274

Graph B

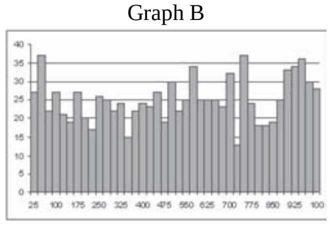

Fig. 10. $f(x) = \frac{31}{32}x^{-\frac{1}{32}}$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.5009
2	0.2534
3	0.1238
4	0.0606
5	0.0321

Graph B

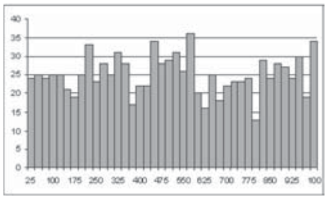

Fig. 11. $f(x) = 1$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.4627
2	0.2489
3	0.1306
4	0.0713
5	0.0370

Graph B

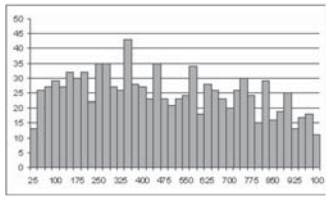

Fig. 12. $f(x) = \frac{9}{8}x^{\frac{1}{8}}$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.4086
2	0.2149
3	0.1312
4	0.0806
5	0.0511

Graph B

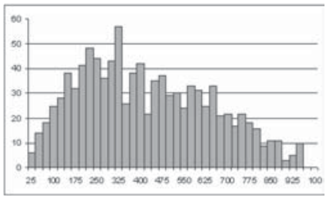

Fig. 13. $f(x) = \frac{3}{2}x^{\frac{1}{2}}$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.3271
2	0.1835
3	0.1175
4	0.0800
5	0.0592

Graph B

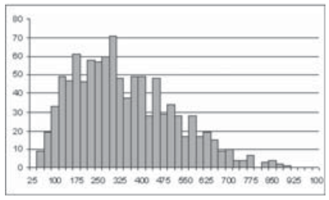

Fig. 14. $f(x) = 2x$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.2198
2	0.1251
3	0.0878
4	0.0673
5	0.0539

Graph B

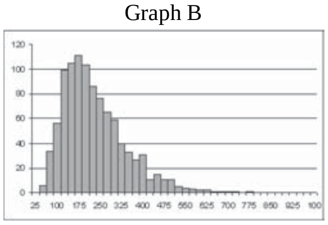

Fig. 15. $f(x) = 3x^2$, $x \in [0; 1)$.

Table A

i	N_i/n
1	0.0043
2	0.0021
3	0.0016
4	0.0014
5	0.0013

Graph B

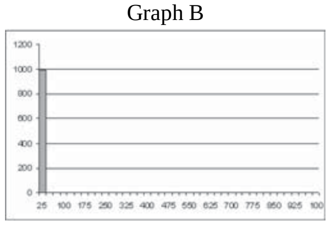

Fig. 16. $f(x) = 9x^8$, $x \in [0; 1)$.

3. Discussion

1. Simulation results reported in Figs. 1–3 correspond to bounded probability density functions f which do not vanish at point $x = 0$. All three cases considered show similar average values N_i/n of subtree sizes (Table A) and similar empirical distribution of random variable N_i (Graph B).

Simulation results reported in Figs. 4–7 correspond to bounded probability density functions f satisfying $f(0) = 0$ and $f'(x) = c$, $c > 0$. Note that examples in Figs. 4–7 are almost equivalent up to a change of scale of the random weights $X_1, \ldots, X_m$.

It is clear that distribution of N_1 and average values of $N_1, N_2, \ldots, N_5$ are different in the cases (i) and (ii).

2. Simulation results reported in Figs. 8–16 correspond to probability density functions $f(x) = a_\alpha x^\alpha$, $x \to 0$, for various $\alpha > -1$.

3. One could expect that, for $K_n(f)$ with f satisfying $f(x) \sim x^\alpha$, $x \to 0$, the asymptotic distribution of values of N_i (as $n \to \infty$) depends only on the parameter $\alpha > -1$. It would be interesting to find a theoretical argument proving or disproving this claim as well as to identify asymptotic distributions.

Given $v, u \in V$ let $v = v_{i_1}, \ldots, v_{i_{k+1}} = u$ denote the vertices of the shortest path connecting v and u in $K_n(f)$. We encode the path by the sequence of integers $(y_1, y_2, \ldots, y_k)$. Here $y_j = r$ whenever $u \in T_r(v_{i_j})$. The code shows relative weights of the edges of the path. It would be interesting to learn about statistical properties of such "codes".

References

1. R. Van der Hofstad, G. Hooghiemstra, P. Van Mieghem, The weight of the shortest path tree, *Random Structures and Algorithms*, **30**, 359–379 (2007).
2. P. Van Mieghem, *Performance Analysis of Communications Networks and Systems*, Cambridge university Press, Cambridge (2006).

REZIUMĖ

M. Bloznelis, I. Radavičius. Svorinio grafo trumpiausiųjų kelių medžio empirinė analizė

Pilnojo grafo, turinčio n viršūnių briaunoms $e_1, \ldots, e_m$ priskiriame neneigiamus svorius $X_1 = X(e_1), \ldots, X_m = X(e_m)$. Čia $m = \binom{n}{2}$. Trajektorijos $e_{i_1}, e_{i_2}, \ldots, e_{i_k}$ ilgiu vadiname šią trajektoriją sudarančių briaunų svorių sumą $X_{i_1} + \cdots + X_{i_k}$. Trajektoriją, jungiančią viršūnes u ir v, vadiname trumpiausiu keliu tarp u ir v, jei nėra kitos trajektorijos, kurios ilgis būtų mažesnis.

Darbe tiriamas medis, sudarytas iš trumpiausių kelių, vedančių iš fiksuotos viršūnės u į visas kitas grafo viršūnes. Tokį medį gauname pritaikę, pvz., Dijkstros algoritmą.

Darbe nagrinėjamos statistinės trumpiausių kelių medžio savybės, kai svoriai parenkami atsitiktinai ir nepriklausomai vienas nuo kito. Tiriame medžio struktūros priklausomybę nuo atsitiktinių svorių tikimybinio skirstinio parametrų.

Raktiniai žodžiai: trumpiausių kelių medis, atsitiktinis svorinis grafas.