Information Processing Letters 46 (1993) 135–141
Elsevier

11 June 1993

The expected length of a shortest path

Robert Davis and Armand Prieditis

Department of Computer Science, University of California, Davis, CA 95616, USA

Communicated by D. Gries
Received 26 January 1993

Abstract

Davis, R. and A. Prieditis, The expected length of a shortest path, Information Processing Letters 46 (1993) 135–141.

We derive an exact summation formula and a closed-form approximation for the expected length of a shortest path for a complete graph where the arc lengths are independent and exponentially distributed random variables. Experimental data validates both results. The property of completeness allows us to exploit certain symmetries to derive these results, which would otherwise require computing an exponential number of recursive equations. We have also found that this formula is a close approximation for the expected length of a shortest path in complete graphs with uniformly distributed arc lengths.

Keywords: Analysis of algorithms; shortest path, expected length

1. Introduction and motivation

A shortest path problem involves finding a path of shortest length between two nodes in a graph. Such problems are perhaps the most common and fundamental of all transportation and communication network problems. Although most shortest path problems involve arc lengths with fixed values, many practical situations dictate arc lengths that are random variables with certain probability distributions. For example the driving time from one location to another is typically not fixed, but follows some probability distribution. Kulkarni [4] has developed an analytical method for the computation of the expected length of a shortest path for networks with independent and exponentially distributed arc lengths – such graphs are important in communication and queuing problems. The analytic method can then be used to compute the probability that a given path is the shortest path and the conditional distribution of the length of a path given that it is the shortest path. Unfortunately, Kulkarni's method involves computing a recursive function that requires solving an exponential number (of the graph size) of recursive equations. We derive an easily-computable summation formula and a closed-form approximation for the expected length of a shortest path by additionally assuming that the graphs are complete.

Other approaches to computing the expected length of a shortest path have included numerical evaluation of multiple integrals that represent the probability distribution [2,5,7,8]. However, such numerical evaluation is feasible only for small networks. Consequently, several Monte Carlo simulation techniques have been developed to estimate the probability distributions from the integrals [1,9]. In contrast, Mirchandiani derives a *bound* on the expected length of a shortest path and avoids numerical evaluation of the multiple integrals by assuming that the arc lengths are discrete random variables [6]. Zemel and Hassin

Correspondence to: A. Prieditis, Department of Computer Science, University of California, Davis, CA 95616-8562, USA.

derive a bound for the expected length of a shortest path for uniform distributions with a fixed mean [3].

The rest of this paper is organized as follows. Section 2 briefly presents Kulkarni's framework, which we will use throughout the paper. For brevity, Kulkarni's theorems are stated without proof – these proofs can be found in his paper. Next, Section 3 derives the expected length of a shortest path for complete graphs. Section 4 then experimentally validates this result. Also presented in this section are some preliminary experimental results for uniform distributions. Finally, Section 5 summarizes our results and discusses several promising directions for future research.

2. Kulkarni's analytical framework

Kulkarni's key idea is to treat the graph of nodes as a communication network where the time taken for a message to travel from one node to another is the arc length between the two nodes. The process starts when the source node receives the message and ends when the sink node receives the message. As soon as a node receives an incoming message, it transmits along all its outgoing arcs and then disables itself and all nodes without a path to the sink node containing only active nodes from receiving or transmitting any future messages. The time taken for the message to travel from the source node to the sink node is the length of the shortest path. Using this idea, Kulkarni constructs a recursive formula that yields the various moments of the shortest path.

More formally, let $G = (V, A)$ be a directed network, where V is the set of nodes and A is the set of arcs. Let $L(u, v)$ be the length of arc $(u, v) \in A$. The objective is to find the expected shortest path length between any two different nodes. The source node will be denoted by s; the sink node, by t. Each node sends and receives messages travelling at unit speed: the time for a message to travel between nodes u and v is $L(u, v)$.

To formalize the message transmission process. Kulkarni defines the following functions:

$X(t)$ = the set of all disabled nodes at time t (i.e. the state of the *system*),

$X(t-)$ = the state of the system immediately *before* time t,

$R(X) = \{v \in V$: $\exists$ path from v to t with nodes not in $X\}$, for any $X \subset V$ such that $s \in X$ and $t \in V - X$,

$S(X) = V - R(X)$,

$\Omega = \{X \subset V$: $s \in X$, $t \in \bar{X}$, $X = S(X)\}$,

$\Omega^* = \Omega \cup \{V\}$,

$C(X, \bar{X}) = \{(u, v) \in A$: $u \in X$, $v \in \bar{X}\}$, for all $X \subset V$,

$Y(t) = \{(u, v) \in A$: the arc (u, v) is carrying the message at time $t\}$.

At time 0, the source node receives the message and $X(0-) = \emptyset$. When any node $u \in V$ receives a message at time t:

1. The message is transmitted from u to all $v \in V$ such that $(u, v) \in A$.
2. All nodes in $S(X(t-) \cup \{u\})$ are disabled.
3. All messages heading for nodes in $S(X(t-) \cup \{u\})$ are aborted.

The process terminates when the sink node receives the message and all the nodes are disabled.

Using these definitions, Kulkarni proves that:

1. $X(t) \in \Omega^*$ for all $t \geqslant 0$.
2. There is a unique minimal cut, $C(X)$, contained in $C(X, \bar{X})$ if $X \in \Omega$.
3. If $X(t) \neq V$, then $Y(t) = C(X(t))$.

Now we come to Kulkarni's central theorem:

Theorem 1. *If for all $(u, v) \in A$ $L(u, v)$ are independent random variables exponentially distributed with mean $1/\mu(u, v)$, then $\{X(t)$: $t \geqslant 0\}$ is a Continuous Time Markov Chain (CTMC) with state space Ω^* and infinitesimal generator matrix $Q = [q(D, B)]$ $(D, B \in \Omega^*)$ given by*

$$q(D, B)\begin{cases} \sum_{(u,v)\in C_v(D)} \mu(u, v) & \\ \quad \text{if } (\exists v \in \bar{D})\ B = S(D \cup \{v\}), & \\ -\sum_{(u,v)\in C_v(D)} \mu(u, v) & \text{if } B = D, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where $C_v(D) = \{(u, v) \in C(D)\}$ and $C(V) = \emptyset$.

To simplify notation, let $N = |\Omega^*|$, and let the states in Ω^* be labeled from 1 to N. Whenever $X(t)$ changes to another state, the number of nodes in $X(t)$ is increased by at least one. Therefore, if the elements in Ω^* are ordered by non-decreasing cardinality, then the generator matrix Q will be upper triangular. Since $q(D, D) \neq 0$ for $D \in \Omega$, all states in Ω are transient. The state V, however, is absorbing, since $q(V, V) = 0$. Using these facts, Kulkarni derives a set of differential equations describing the distribution of the length of the shortest path and a recursive formula for finding the moments of the shortest path.

In order to compute the moments of the shortest path, he defines

$$T_i = \min\{t \geqslant 0: X(t) = N | X(0) = i\}, \quad 1 \leqslant i \leqslant N,$$
$$\tau_i(k) = E(T_i^k), \quad k > 0, \tag{2}$$

where $\tau_i(0) = 1$ for all $1 \leqslant i \leqslant N$ and for $k \geqslant 1$, $\tau_N(k) = 0$, where N is the number of states in the CTMC. Kulkarni proves (by induction) that:

$$\tau_i(k) = \frac{k\tau_i(k-1) + \sum_{j>i} q_{ij}\tau_j(k)}{-q_{ii}}. \tag{3}$$

Thus, to compute $E(T_1^k)$, one needs to compute $\tau_i(r)$ for $r = 1, 2, \ldots, k$, $i = N, N-1, \ldots, 1$ in that order. Hence, to compute the expected value of the shortest path, $E(T_1) = \tau_1(1)$, one needs to compute $\tau_i(1)$ for $i = N, N-1, \ldots, 1$.

3. The expected length of a shortest path

For a complete graph, the number of states in the continuous time Markov chain is $2^{n-2} + 1$. Therefore to determine the expected length of a shortest path would require solving $2^{n-2} + 1$ equations – roughly one equation per possible system state. Clearly this is not feasible for large n and is more computationally expensive than experimentally computing the expected length of a shortest path. However, if we restrict our consideration to the case when all the arc lengths are exponentially distributed with parameter μ, symmetries in the graph allow these equations to be reduced to a simple closed-form formula. This section derives that formula. For the remainder of the paper, we will assume that the graph is complete and that edge weights are independent and exponentially distributed with parameter μ. First we must determine the values of the entries in the infinitesimal generator matrix Q of the CTMC.

Lemma 2. *$q_{ij} > 0$ if and only if either state $j = N$ or state j contains one more node than state i.*

Proof. q_{ij} is positive if and only if there is a path in the network from state i to state j. Since every node in G has an edge to the sink node, every state will have a path to state N and q_{iN} has a positive value for all i. Now, consider the case when $j < N$. State changes occur when a message reaches a new node and it is disabled. For a complete graph, all nodes have a direct path to the sink node so only the node which received the message is disabled. Therefore, state changes occur only by moving to states with one more node if $j < N$. □

Theorem 3. *If $q_{ij} > 0$ and state i contains l nodes, then $q_{ij} = l\mu$.*

Proof. From Lemma 2, $q_{ij} > 0$ only if state j contains one more node than state i or $j = N$. If $j < N$ call the new node v, else let $v = t$. q_{ij} is equal to the rate at which messages travel from the nodes in state i to the node v. For each path from a node in i to v, the rate is $1/(1/\mu) = \mu$, because each edge is exponentially distributed with mean $1/\mu$. Since the graph is fully connected, each of the l nodes in i have a path to v. Therefore, the rate at which the messages travel from nodes in i to the node v is the sum of the rates along all these paths, $l\mu$. □

Theorem 4. *Let l be the number of nodes in state i. Then the value of q_{ii} is given by $-(n-l)l\mu$.*

Proof. From Lemma 2, the only positive values in Q are those q_{ij} such that state j has one more node than state i and q_{iN} for $i < N$. The number of states containing one more node than state i is $n - 1 - l$. This follows because a state containing one more node is formed by adding any node not in state i except the sink node to the nodes

already in i. Therefore, row i of Q contains $n-l$ positive values. Since the sum of each row of Q must be 0 and the diagonal element is the only negative element, q_{ii} must have a negative value to offset the $n-l$ positive values. From Theorem 3, each of the positive elements in a row has value $l\mu$. Hence, the value of $q_{ii} = -(n-l)l\mu$. □

The following theorem relates the value of $\tau_i(k)$ for different states containing the same number of nodes.

Theorem 5. *For any states i and j, such that the number of nodes in state i equals the number of nodes in state j, $\tau_i(k) = \tau_j(k)$.*

Proof. $\tau_l(k)$ is the expected value of the shortest distance from any of the nodes in state l to the sink node t. Since the network is complete and all arc lengths have the same distribution, the expected value of the distance from any node to t is the same. Therefore, the expected value of the distance from any two sets of nodes containing the same number of elements is the same. □

We can now simplify the formula for the expected length of the shortest path by defining $\zeta_i = \tau_j(1)$, where $i = 1, 2, \ldots, n$, and j is a state containing i nodes. Note that by Theorem 5 all states containing the same number of nodes have the same value, so this definition assigns only one value to each ζ_i even though one can use any state with i nodes to determine this value. $\tau_1(1) = \zeta_1$ is the expected length of the shortest path.

Theorem 6.

$$\zeta_i = \begin{cases} 0 & \text{if } i = n, \\ \dfrac{1 + i(n-i-1)\mu\zeta_{i+1}}{(n-i)i\mu} & \text{otherwise.} \end{cases} \tag{4}$$

Proof. Since state N contains n nodes, $\zeta_n = \tau_N(1) = 0$. When $i < n$ we have from equation (3):

$$\tau_m(1) = \left[\tau_m(0) + \sum_{j>m} q_{mj}\tau_j(1)\right] / -q_{mm}.$$

From Theorem 4, $q_{mm} = -(n-i)i\mu$ where i is the number of nodes in state m. From the definition of τ_m, $\tau_m(0) = 1$. Also, from Lemma 2, q_{mj} is nonzero only if state j contains one more node than state m or is the final state, so the sum can simply be over those nodes j containing one more node than state m and state N. However, $\tau_N(1) = 0$. From Theorem 3, $q_{mj} = i\mu$ for the remaining states. From the proof of Theorem 4, we showed that the number of states containing one more node than state m is $n-1-i$. Therefore, we have

$$\zeta_i = [1 + i(n-i-1)\mu\zeta_{i+1}]/(n-i)i\mu. \qquad \square$$

As before this formula is recursive, requiring one to compute $\zeta_n, \zeta_{n-1}, \ldots, \zeta_1$ to compute the expected length of a shortest path, but now only n steps will be taken instead of the original $2^{n-2} + 1$ steps. We will now reduce the formula for ζ_i to a summation from which we can derive a closed form approximation.

Theorem 7. *For $i > 0$,*

$$\zeta_{n-i} = \frac{1}{i\mu} \sum_{k=n-i}^{n-1} \frac{1}{k}.$$

Proof. (By induction on i) For the base case $i = 1$ and equation (4) gives us

$$\zeta_{n-1} = \frac{1 + (n-1)0\mu\zeta_{i+1}}{(1)(n-1)\mu} = \frac{1}{(n-1)\mu},$$

which agrees with Theorem 7. For the induction step, we will assume Theorem 7 holds for i and show this implies that it holds for $i+1$. From equation (4):

$$\begin{aligned} \zeta_{n-(i+1)} = \zeta_{n-i-1} &= \frac{1 + (n-i-1)i\mu\zeta_{n-i}}{(i+1)(n-i-1)\mu} \\ &= \frac{1 + (n-i-1)i\mu\left(\dfrac{1}{i\mu}\displaystyle\sum_{k=n-i}^{n-1}\frac{1}{k}\right)}{(i+1)(n-i-1)\mu} \\ &= \frac{\dfrac{1}{(n-i-1)} + \left(\displaystyle\sum_{k=n-i}^{n-1}\frac{1}{k}\right)}{(i+1)\mu} \\ &= \frac{1}{(i+1)\mu}\left(\sum_{k=n-(i+1)}^{n-1}\frac{1}{k}\right). \end{aligned} \tag{5}$$

Therefore by induction, Theorem 7 is true. □

Using Theorem 7, the expected value of the shortest path is $(1/(n-1)\mu)\sum_{k=1}^{n-1} 1/k$.

Theorem 8. *The closed-form function*

$$\frac{\ln(n-1)}{(n-1)\mu}$$

approximates $(1/(n-1)\mu)\sum_{k=1}^{n-1} 1/k$ *within* $1/(n-1)\mu$.

Proof.

$$\sum_{k=1}^{n-1} \frac{1}{k} - \int_1^{n-1} \frac{1}{k} \leqslant 1$$

$$\Rightarrow \sum_{k=1}^{n-1} \frac{1}{k} - \ln(n-1) \leqslant 1$$

$$\Rightarrow \frac{1}{(n-1)\mu} \sum_{k=1}^{n-1} \frac{1}{k} - \frac{\ln(n-1)}{(n-1)\mu}$$

$$\leqslant \frac{1}{(n-1)\mu}. \quad \square \tag{6}$$

Theorem 8 implies several interesting facts as graph size increases while mean arc length remains fixed. First, the difference between the approximate and the theoretical length of the shortest path shrinks to 0. Second, the expected length of a shortest path also shrinks to 0. For small graphs, the expected length of the shortest path can be better approximated by adding in polynomial error terms to the closed-form formula. The theorem also implies that the expected length of a shortest path is a *linear* function of the mean arc length for fixed graph sizes.

4. Experimental results

To validate our analytical results we ran three sets of experiments. The first set was on small graphs (from 2–20 nodes in increments of 1), the second on mid-size (from 30–100 in increments of 10), and the last on large graphs (from 200–1000 in increments of 100). For smaller graph sizes we were able to run many more experiments. The distribution of arc lengths is exponential with a mean arc length of 10. For each experiment we computed the theoretical and approximate shortest path length and compared it to the experimental (actual) one. Figure 1 summarizes the results for small graphs. As the figure shows, the difference between the predicted (theoretical) and actual (experimental) is negligible. Moreover, as graph size grows, the approximate shortest path length gets closer to the theoretical and experimental. For the mid-size to large graphs, the difference between the theoretical, approximate, and experimental shortest path lengths shrinks to zero. (These results are not shown because the curves are essentially the same.) These results validate Theorem 7.

Our experimental results also validate Theorem 8. For the smallest graph size (2), the absolute error between the approximate and theoretical expected shortest path length is at its maximum, which is the mean arc length (in our case 10). This error quickly begins to shrink and even for graphs of size 20, the error is only 0.33. The error continues shrinking for mid-size and large graphs until it shrinks to zero for large graphs. These results show that using the closed-form approximation for the expected length of a shortest path is appropriate in mid-size to large graphs. To get less error for small graphs would require the addition of error terms to the closed form. Although the error terms are easy to compute, their addition makes the form of the approximation function appear more cumbersome and therefore less easily remembered.

Figure 2 shows that the expected length of a shortest path is a linear function of the mean for fixed graph sizes. This result also validates Theorem 8.

Does Theorem 8 hold for graphs with non-exponentially distributed arc lengths? To test the hypothesis that the theorem does indeed hold, we ran the same set of experiments except with uniformly distributed arc lengths. The uniform distribution that we used starts at 0 and is positive up to the mean of the exponentially distributed arcs. For small graphs, the results were nearly identical; for large graphs, the difference shrank to zero. Although we have not theoretically verified this result, it is consistent with the upper bound

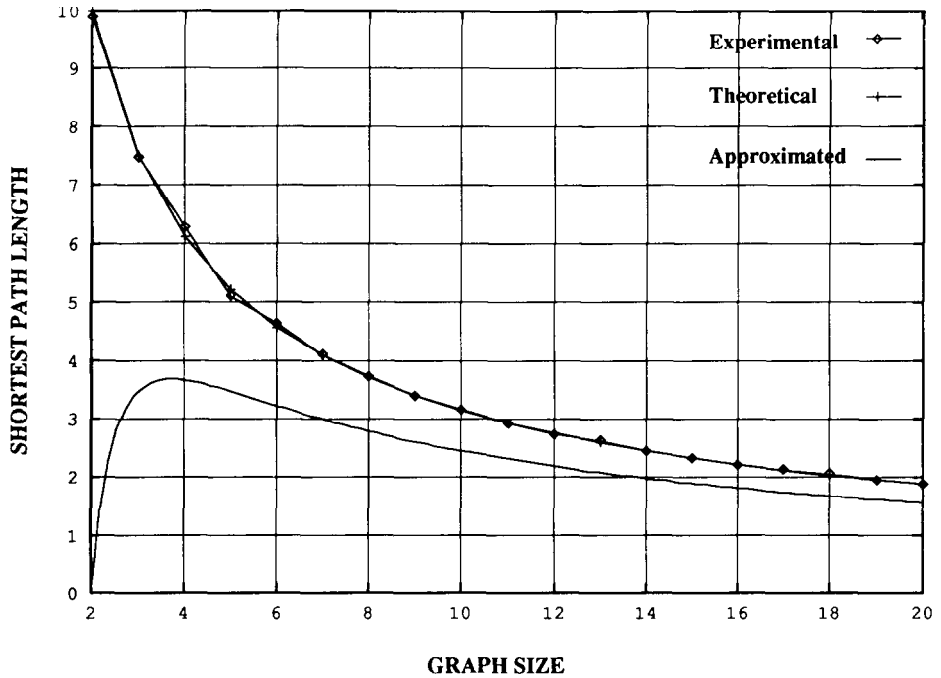

Fig. 1. Graph size versus shortest path length for small graphs (averaged over 2000 trials).

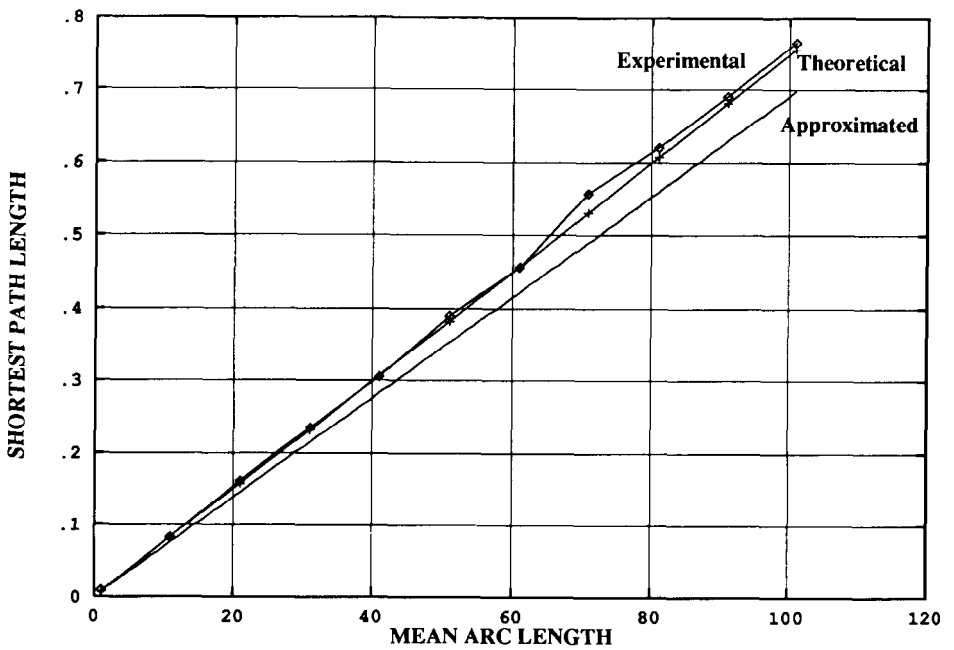

Fig. 2. Mean arc length versus average shortest path length for a 1000 node graph (averaged over 80 trials).

obtained by Hassin and Zemel [3] for such uniformly distributed arcs lengths.

5. Conclusions and future work

Using analytical methods developed by Kulkarni [4], this paper derived a closed-form formula for the expected length of a shortest path in complete graphs where the arc lengths are independent and exponentially distributed random variables. The property of completeness allows us to exploit certain symmetries to derive this closed-form, which would otherwise require numerically solving an exponential number of recursive equations. We are currently extending our results to other probability distributions and graphs of a given sparsity, and deriving a closed-form approximation for the probability distribution function for shortest path lengths. This probability distribution function will be useful in predicting which nodes are most likely to be on a shortest path from one node to another.

Acknowledgment

This research is supported by the National Science Foundation under grant number IRI-9109796.

References

[1] V. Adlakha and G Fishman, An improved conditional Monte Carlo technique for the stochastic shortest route problem, Tech. Rept. 84-9, Curriculum in Operations Research and Systems Analysis, University of North Carolina, Chapel Hill, NC 27514, 1984.

[2] H. Frank, Shortest paths in probabilistic graphs, *Oper. Res.* **17** (1969) 583–599.

[3] R. Hassin and E. Zemel, On shortest paths in graphs with random weights, *Math. Oper. Res.* **10** (1985) 557–564.

[4] V.G. Kulkarni, Shortest paths in networks with exponentially distributed arc lengths, *Networks* **16** (1986) 255–274.

[5] J. Martin, Distribution of the time through a directed acyclic network. *Oper. Res.* **13** (1965) 46–66.

[6] P.B. Mirchandani, Shortest distance and reliability of probabilistic networks, *Comput. Oper. Res.* **3** (1976) 347–356.

[7] A. Pritsker, Application of multi-channel queing results to the analysis of conveyor systems, *J. Industrial Engrg.* **17** (1966) 14–21.

[8] C. Sigal, A. Pritsker and J. Solberg, The use of cutsets in Monte Carlo analysis of stochastic networks, *Math. Comp. Simulation* **21** (1979) 376–384.

[9] C. Sigal, A. Pritsker and J. Solberg, The stochastic shortest route problem, *Oper. Res.* **28** (1980) 1122–1129.