Optimal determination of source-destination connectivity in random graphs

Roger Fu

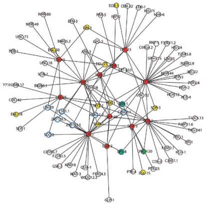
June 3rd, 2016

Outline

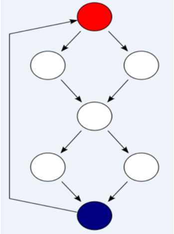
Outline

Outline

Graphs Are Everywhere

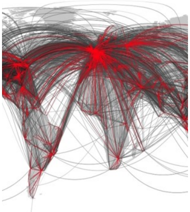


Biological Network

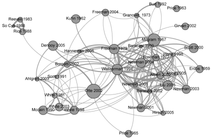


Program Flow

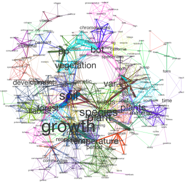
Social Network



Transportation Network



Citation Network



Topic Network

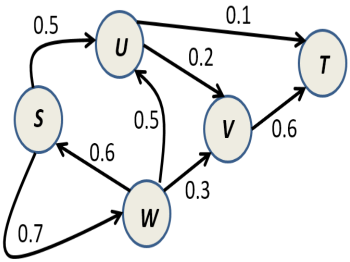
Uncertainty Is Prevalent

"... the real world is always certain; it is our knowledge of it that is sometimes uncertain." [1]

[1] Amihai Motro [Management of Uncertainty in Database Systems]

Uncertainty Is Prevalent

"... the real world is always certain; it is our knowledge of it that is sometimes uncertain. " [1]



Uncertain Graph
(Edge Uncertainty)

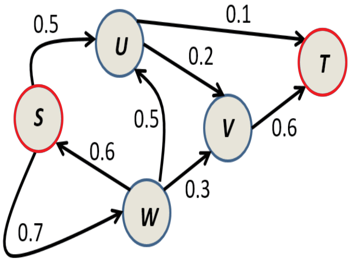
- Communication Networks
- Citation Networks
- Topic Networks

[1] Amihai Motro [Management of Uncertainty in Database Systems]

Outline

Determination of Genuine Relation

- Fundamental problem of determining s-t connectivity.



Uncertain Graph
(Edge Uncertainty)

- Communication Networks
- Citation Networks
- Topic Networks

Outline

Model

Definition (Problem Formulation)

Given a random graph $\mathcal{G}(V, E)$, a probability vector $p = (p_1, p_2, ..., p_{|E|})$, a cost vector $c = (c_1, c_2, ..., c_{|E|})$ and two nodes $s, t \in V$, find an adaptive testing strategy to determine the $s - t$ connectivity while incurring minimum expected cost.

Model

Definition (Problem Formulation)

Given a random graph $\mathcal{G}(V, E)$, a probability vector $p = (p_1, p_2, ..., p_{|E|})$, a cost vector $c = (c_1, c_2, ..., c_{|E|})$ and two nodes $s, t \in V$, find an adaptive testing strategy to determine the $s - t$ connectivity while incurring minimum expected cost.

- The true underlying graph has a product distribution and the expectation of cost is taken over all possible realizations of $\mathcal{G}$.

Adaptive Testing Strategy

- Define the set of states of $\mathcal{G}$ as $S = \{0, 1, *\}^{|E|}$.
- Formally, an adaptive testing strategy is a mapping $S \mapsto E$.
- A strategy terminates by verifying the existence of an $s - t$ path or an $s - t$ cut.

Outline

NP Hardness of the Decision Version

Given a random graph $\mathcal{G}(V, E)$, a probability vector $p = (p_1, p_2, ..., p_{|E|})$, a cost vector $c = (c_1, c_2, ..., c_{|E|})$ and two nodes $s, t \in V$, is there an adaptive testing strategy with expected cost less than k.

A Stronger Result

Theorem

Computing the expected cost of the optimal strategy is #P-hard [2].

Proof

- By reduction from network reliability problem.
- Inspired by Papadimitriou, Christos H., and Mihalis Yannakakis. "Shortest paths without a map."Theoretical Computer Science 84.1 (1991): 127-150.

[2]Valiant, Leslie G. "The complexity of enumeration and reliability problems."

More Words on Class #P

Theorem

Computing the expected cost of the optimal strategy is #P-hard [3].

Leslie Valiant

- Complexity of Counting
- PAC Leaning Model
- Professor at Harvard University
- ACM Turing Award Winner

[3] Valiant, Leslie G. "The complexity of enumeration and reliability problems."

An Even Stronger Result

Theorem

Deciding the next edge to test is NP-hard.

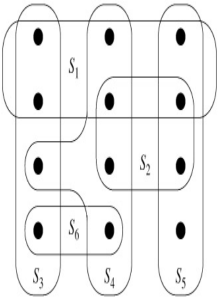
- *By reduction from set cover.*

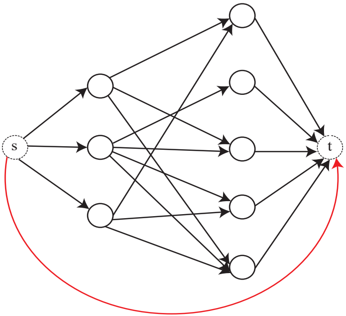
An Even Stronger Result

Theorem

Deciding the next edge to test is NP-hard.

- *By reduction from set cover.*





Outline

Outline

Markov Decision Process Framework

Definition (Markov Decision Process)

A mathematical model for modeling decision making under uncertain situations. A Markov Decision Process (MDP) model contains:

- A set of possible world states S
- A set of possible actions A
- A reward function $R(s, a)$
- A description T of each action's effect in each state

Markov Property: the effects of an action taken in a state depend only on that state and not one the prior history.

Exact Algorithm Based on Dynamic Programming

Input: Random graph $\mathcal{G}(V, E)$, probability vector $\mathbf{p}$, node s and t
Output: The optimal testing strategy $\mathcal{D}$, that is, a mapping from each partial graph to the next edge to test.
Precompute the elements in the corresponding Finite Horizon Markov Decision Process.
Initialize: $u(s) = 0$, for all $s \in S_N$
for $i = N$ *to* 0 **do**
  for *All* s *in* S_i **do**
    $t_{e^*} = \arg\min_{t_e \in A_s}\{c_e + p_e \times u(S \cdot e) + (1 - p_e) \times u(S \backslash e)\}$
    $u(s) = c_{e^*} + p_{e^*} \times u(S \cdot e^*) + (1 - p_{e^*}) \times u(S \backslash e^*)$
    $\mathcal{D}(s) = t_{e^*}$
  end
end

Exact Algorithm with Exponential Running Time

Outline

A Simple Greedy Approach

Theorem

A strategy that tests the edges following the ascending order of costs is an $O(|E|)$ approximation.

Adaptive Submodular Framework

- An approximation algorithm based on adaptive submodular framework.
- $O(\ln(Q_p Q_c))$ approximation, where Q_p and Q_c are the number of $s-t$ paths and $s-t$ cuts in the graph.
- Q_p and Q_c are polynomial to $|E|$ for most graphs, that is, the algorithm yields a logarithmic approximation in most cases.

A More Sophisticated Approximation Algorithm

Input: Random graph $\mathcal{G}(V, E)$, probability vector $\mathbf{p}$, cost vector $\mathbf{c}$, node s and t
Output: An approximate sequential testing strategy
Initialize: Partial realization ψ, Set of tested edges E_π as empty sets.
Repeat until $g(\psi) = Q_p Q_c$
$e = \arg\max_{e \in E \setminus E_\pi} \{\Delta(e|\psi)\}$.
Set E_π as $E_\pi \cup \{e\}$, test e and observe the outcome.
if $e = 1$ **then**
| $\psi = \psi \cup (e, 1)$
else
| $\psi = \psi \cup (e, 0)$
end

Approximation Algorithm under Adaptive Submodularity Framework

- Fu, Luoyi, Xinbing Wang, and P. R. Kumar. "Optimal determination of source-destination connectivity in random graphs."Mobihoc 2014.
- Arijit Khan and Lei Chen "On Uncertain Graphs Modeling and Queries"
- Valiant, Leslie G. "The complexity of enumeration and reliability problems."SIAM Journal on Computing 8.3 (1979): 410-421.
- Bob Givan and Ron Parr "An introduction to Markov Decision Process"
- Golovin, Daniel, and Andreas Krause. "Adaptive submodularity: Theory and applications in active learning and stochastic optimization."Journal of Artificial Intelligence Research (2011): 427-486.

Thank You!