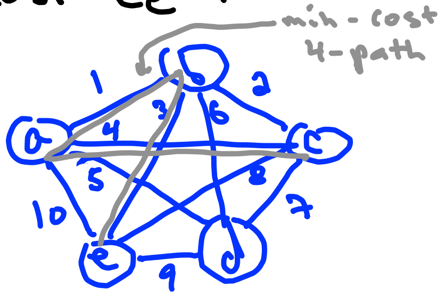
Section 21.2: Long Paths via Color Coding

The Minimum-Cost *k*-Path Problem

Motivation: protein-protein interaction (PPI) networks.

Goal: identify interesting linear pathways.

Input: undirected graph $G=(V,E)$, real-valued cost c_e for each edge e, positive integer k.



Output: A k-path P of G with min-possible total cost $\sum_{e \in P} c_e$.

[if G has no k-paths, report this]

Fact: NP-hard.
(generalizes TSP)

Quiz #1

Subproblems (first attempt): compute $C_{S,v}$, the minimum cost of a cycle-free path ending at v & visiting precisely the vertices in S.

[for each $S \subseteq V$ with $|S| \le k$ and $v \in S$]

Question: How many subproblems. (Choose the strongest true statement.)

(a) $O(n^k)$

(b) $O(kn^k)$

(c) $O(n^{k+1})$

(d) $O(2^n)$

(e) $O(k2^n)$

Quiz #2

What is the running time of a straightforward implementation of exhaustive search, as a function of n and k? (Choose the strongest true statement.)

(a) $O(n \cdot k!)$

(b) $O(kn^k)$

(c) $O(n^{k+1})$

(d) $O(k2^n)$

(e) $O(k \cdot n!)$

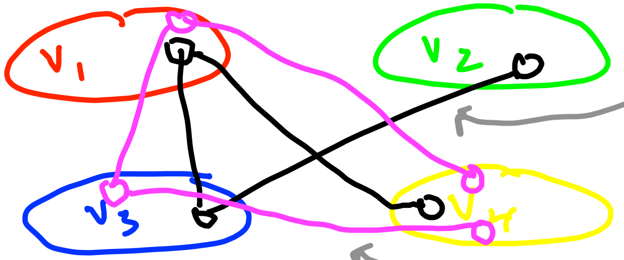
Color Coding

"panchromatic"

Step 1: partition V into $V_1, V_2, \ldots, V_k$ so that at least one min-cost k-path has exactly one vertex per group.

Step 2: Compute min-cost k-path with exactly one vertex per group.

panchromatic path



k-path but not panchromatic

Min-Cost Panchromatic Path: Subproblems

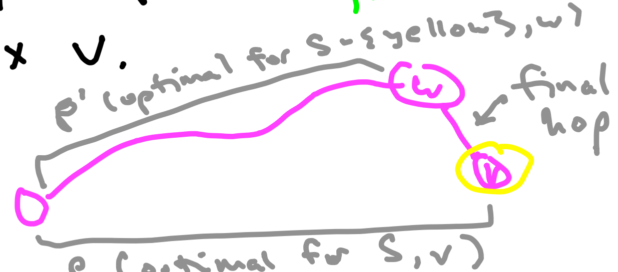
Goal: given partition/coloring $V = V_1 \cup V_2 \cup \cdots \cup V_k$, compute min-cost panchromatic path.

Intuition: only need to track colors of vertices, not vertices themselves.

Defn: for $S \subseteq \{1,2,\ldots,k\}$, an S-path visits exactly one vertex with each color of S, no other vertices. (k-path = $\{1,2,\ldots,k\}$-path)

Subproblems: for each $S \subseteq \{1,2,\ldots,k\}$ and $v \in V$, compute $C_{S,v} :=$ min cost of an S-path ending at vertex v.

Recurrence: $C_{S,v} = \min_{(w,v) \in E} \left(C_{S-\{\text{color of } v\},\, w} + c_{wv} \right)$



Min-Cost Panchromatic Path: Algorithm

Panchromatic Path [$\sigma(v)$ = color/group of v]

$A := (2^k - 1) \times |V|$ 2-D array [subproblem solutions]

For all $i \in \{1, 2, \ldots, k\}$ and $v \in V$: [base cases]

– if $\sigma(v) = i$ then $\underbrace{A[\{i\}][v] := 0}_{\text{empty path}}$ else $\underbrace{A[\{i\}][v] = +\infty}_{\text{no such path}}$

For $s = 2$ to k: [s = subproblem size]

For $S \subseteq \{1, 2, \ldots, k\}$ with $|S| = s$:

For $v \in V$:

$$A[S][v] := \min_{(w,v) \in E} \left(A[S - \{\sigma(v)\}][w] + c_{wv} \right)$$

[invoke recurrence]

return $\min_{v \in V} A[\{1, 2, \ldots, k\}][v]$

Min-Cost Panchromatic Path: Properties

Correctness: by induction on subproblem size. (standard)

Reconstruction: $O(k)$ post processing. (standard)

Running time: (a la Bellman-Ford algorithm)

- to compute $A[S][v]$: $O(\deg(v))$ time
- to compute $A[S][v]$ for all v: $O(\sum_{v \in V} \deg(v)) = O(m)$

$\Rightarrow$ $O(2^k m)$ over all

Quiz #3

Suppose each vertex of G assigned a color from $\{1, 2, \ldots, k\}$ independently and uniformly at random (each color equally likely).

Let P = k-path of G.

What is the probability that P winds up panchromatic?

(a) $1/k$

(b) $1/k^2$

(c) $1/k!$

(d) $\frac{k!}{k^k}$

(e) $\frac{1}{k^k}$

Random Colorings Are Good Enough

Notation: $p := \frac{k!}{k^k}$.

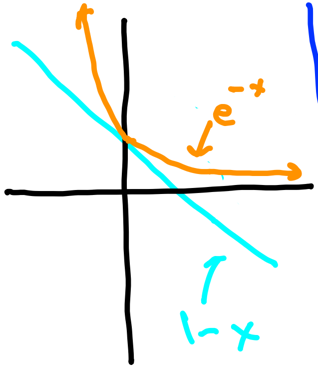
$\overset{\text{(Stirling)}}{\Longrightarrow}$ $p \approx \frac{1}{\cancel{k^k}} \cdot \sqrt{2\pi k} \left(\frac{\cancel{k}}{e}\right)^k$

$= \frac{\sqrt{2\pi k}}{e^k}$] < 1% for $k=7$

$\Longrightarrow$ setting T so that $e^{-pT} = \delta$ ← e.g., 0.01

$\Longrightarrow$ $T \geq \underbrace{\frac{1}{p}\ln\frac{1}{\delta}}$ trials suffice

$= \frac{e^k}{\sqrt{2\pi k}} \cdot \ln\frac{1}{\delta}$



Stirling's Approximation

$n! \approx \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n$

Q: how many trials needed to succeed at least once w/ probability $\geq 99\%$?

Answer:

- 1 trial fails w/ prob $= 1-p$
- T trials fail w/ prob $= (1-p)^T$
- success in ≥ 1 trial $= 1 - (1-p)^T$ ← $\leq e^{-pT}$
  $\geq 1 - e^{-pT}$

The Final Algorithm

Color Coding

- $T := \frac{e^k}{\sqrt{2\pi k}} \ln(\frac{1}{\delta})$ [rounded up to nearest integer]
- for $t = 1$ to T [independent random trials]
  - choose colors $\sigma_t(v)$ for all $v \in V$, independently + uniformly at random
  - compute min-cost panchromatic path for coloring σ_t
- return cheapest of the T paths found

The Final Running Time

Correctness: with probability $\geq 1-\delta$, computes a minimum-cost k-path of G. (if a min-cost k-path turns panchromatic, subroutine will find it)

Running time: $\underbrace{(\text{\# of trials})}_{= O(e^k \ln \frac{1}{\delta})} \times \underbrace{(\text{time per trial})}_{= O(2^k \cdot m)}$

$$= O\left((2e)^k \cdot m \cdot \ln \frac{1}{\delta}\right).$$

"fixed-parameter algorithm"