Cognoms, Nom Data

Algorísmia QT 2019–2020

Examen final **13 de Gener de 2020**

Durada: 3h

Instruccions generals:

- L'exercici 1 s'ha de resoldre fent servir l'espai reservat per a cada resposta.
- Heu d'argumentar la correctesa i l'eficiència dels algorismes que proposeu. Per això podeu donar una descripció d'alt nivell de l'algorisme amb les explicacions i aclariments oportuns que permetin concloure que l'algorisme és correcte i té el cost indicat.
- Heu de justificar totes les vostres afirmacions, en cas contrari la nota de la pregunta serà 0.
- Podeu fer crides a algorismes que s'han vist a classe, però si la solució és una variació n'haureu de donar els detalls.
- Es valorarà especialment la claredat i concisió de la presentació.
- Entregueu per separat les vostres solucions de cada bloc d'exercicis (Ex 1, Ex 2, Ex 3 i Ex 4).
- La puntuació total d'aquest examen és de **10 punts**.

Exercici 1 (4 punts)

(a) **(0.5 punts)** Antoni escull un enter entre 1 i 10^6, Bartomeu ha d'endevinar quin és l'enter escollit fent el mínim nombre de preguntes que tinguin com a resposta (sí/no). Doneu una fita superior a aquest mínim nombre de preguntes.

(b) **(0.5 punts)** Computar la mediana de n elements utilitzant comparacions té un cost $\Omega(n \lg n)$?

(c) **(0.5 punts)** Hem vist a classe que, en general, el problema de la planificació de tasques amb pesos per obtenir un conjunt compatible amb màxim pes, no té sempre solució voraç. És cert que, quan els pesos de les tasques son únicament valors en $\{1, 2\}$, es pot resoldre el problema utilitzant l'algorisme voraç explicat a classe per a la versió del problema sense pesos?

(d) **(0.5 punts)** A l'algorisme de programació dinàmica per a trobar les distàncies més curtes entre dos vèrtexs qualsevol d'un graf amb pesos, però sense cicles negatius, la recurrència ve donada per

$$d_{ij}^{(k)} = \min\{d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}\}.$$

És cert que en aquesta recurrència $d_{ij}^{(k)}$ representa la longitud del camí més curt del vèrtex i al vèrtex j que conté com a màxim k arestes?

(e) **(1 punt)** Considereu l'alfabet $\Sigma = \{a, b, c, d, e\}$, amb les següents freqüències: $(a, 1/2)$, $(b, 1/4)$, $(c, 1/8)$, $(d, 1/16)$, $(e = 1/16)$.

i. (0.3 punts) Dibuixeu l'arbre de prefixos resultant d'aplicar l'algorisme de Huffman i doneu un codi Huffman per als símbols de Σ.

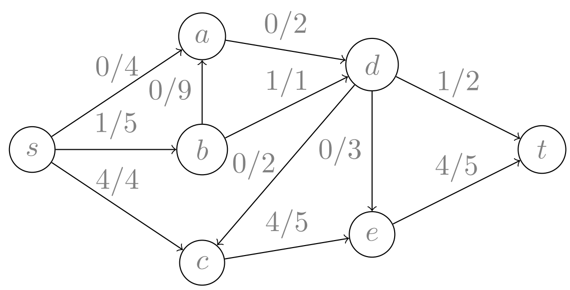
ii. (0.4 punts) L'entropia de Σ és la suma, per tots els símbols a Σ, de la freqüència de cada símbol multiplicada per la longitud de la seva representació en un codi Huffman. Quina és l'entropia de Σ en aquest cas d'acord amb l'apartat (a)?

iii. (0.3 punts) L'entropia de Σ ens proporciona el nombre esperat de bits per símbol. Si tenim un text de 10^6 caràcters de Σ i el comprimim utilitzant el codi Huffman dissenyat a l'apartat (a), quin és el nombre esperat de bits que utilitzarem?

(f) **(1 punt)** Considereu la xarxa de flux $\mathcal{N}$ de la figura de sota (A), que té un flux inicial f. El propòsit d'aquest exercici és que implementeu una iteració de l'algorisme d'Edmonds-Karp. En particular,

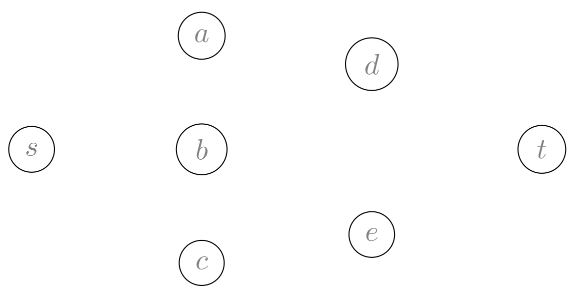
i. Dibuixeu a la figura (B) el graf residual,
ii. Dibuixeu a la figura (C) el graf de nivells.
iii. Expliqueu per què es necessita utilitzar el graf de nivells.
iv. Digueu quin seria l'increment del flux, resultant d'aquesta iteració d'Edmonds- Karp.

(A)



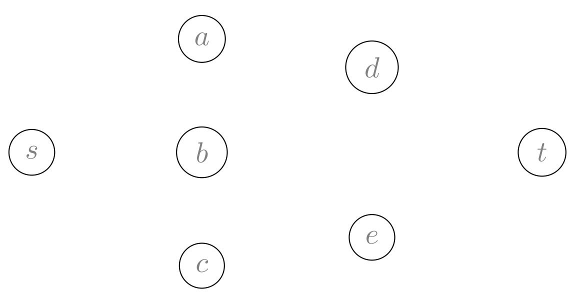
$\mathcal{N}, f$

(B)



G_f

(C)



L

Exercici 2 (2 punts) (Joc).

Tenim un graf connex no dirigit $G = (V, E)$ on cada node $v \in V$ té associat un valor $\ell(v) \geq 0$; considerem el següent joc unipersonal:

(a) Els nodes inicialment no estan marcats i la puntuació del jugador és 0.

(b) El jugador selecciona un node $u \in V$ no marcat. Sigui $M(u)$ el conjunt de veïns de u a G que ja han sigut marcats. Aleshores, s'afegeix a la puntuació del jugador el valor $\sum_{v \in M(u)} \ell(v)$ i marquem u.

(c) El joc es repeteix fins que tots els nodes siguin marcats o el jugador decideixi finalitzar la partida, deixant possiblement alguns nodes sense marcar.

Per exemple, suposeu que el graf té tres nodes A, B, C on A està connectat a B i B amb C, amb $\ell(A) = 3$, $\ell(B) = 2$, $\ell(C) = 3$. En aquest cas, una estratègia òptima seria marcar primer A, després C i finalment B. Aquest ordre dona al jugador una puntuació total de 6.

(a) És possible obtenir una puntuació millor deixant algun dels nodes sense marcar? Justifiqueu la vostra resposta.

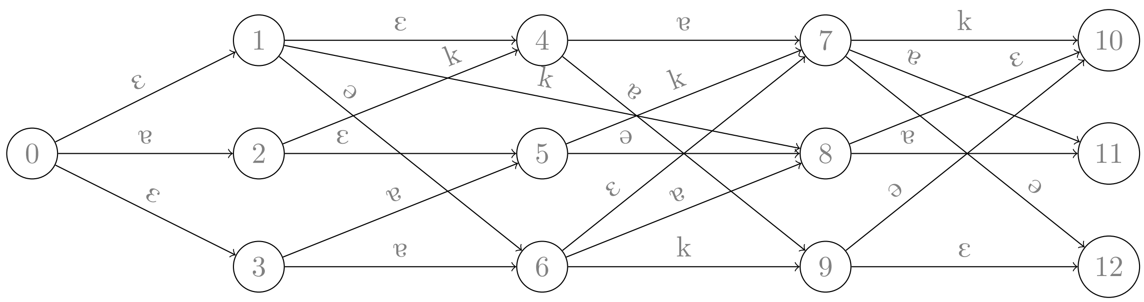
(b) Dissenyeu un algorisme voraç per tal d'obtenir la millor puntuació possible. Justifiqueu la seva correctesa i doneu-ne el cost.

(c) Suposeu ara que $\ell(v)$ pugui ser negatiu. Continua el vostre algorisme proporcionant la puntuació màxima possible?

(d) Considereu la següent modificació del joc per aquest cas en què els nodes poguessin tenir valors negatius: eliminem primerament de G tots els nodes $v \in V$ amb $\ell(v) < 0$, per tal de seguidament executar el vostre algorisme sobre el graf resultant d'aquesta eliminació. Doneu un exemple on aquesta variació no proporcioni la màxima puntuació possible.

Exercici 3 (2 punts) En aquest problema heu de dissenyar un algorisme pel reconeixement de la parla.

Tenim un llenguatge que consisteix en un conjunt finit de sons (fonemes) Σ on $|\Sigma| = m$, per tant des de el punt de vista lingüístic, el llenguatge parlat és força restringit. Considerem el següent model per definir la parla d'una persona en aquest llenguatge. El model està format per un digraf $G = (V, E), |V| = n$, amb un vèrtex distingit $v_0 \in V$. Cada aresta $(u, v) \in E$ està etiquetada amb un so $\sigma(u, v) \in \Sigma$. En aquest model cada camí a G que comença a v_0, correspon a una possible seqüencia de sons a Σ. L'etiqueta associada a un camí és la concatenació (seguint el camí) de les etiquetes de les arestes del camí. Un exemple de model de parla el teniu a la següent figura per l'alfabet $\Sigma = \{$ɛ, ɐ, k, ə$\}$ i $v_0 = 0$.



Una seqüència de fonemes de Σ, $s = (\sigma_1, \sigma_2, \ldots, \sigma_k)$, és vàlida al model (G, σ, v_0) si existeix un camí a G que comença a v_0 amb etiqueta $\sigma_1 \cdot \sigma_2 \cdots \sigma_k$. Per exemple, $s =$ (ɛ, ɐ, k, ə) és vàlida al nostre model ja que es correspon amb l'etiqueta del camí $0 \to 3 \to 5 \to 7 \to 12$. La seqüència $s' =$ (ə, ɛ, ɐ) no és vàlida.

(a) Dissenyeu un algorisme eficient, basat en PD, tal que donats un model (G, σ, v_0) i una seqüència de fonemes s, determini si s és una seqüència vàlida al model. Quina és la complexitat del vostre algorisme?

Suposem ara, que modifiquem el model de manera que a cada aresta $(u, v) \in E$ li assignem un pes $p(u, v)$, $0 \leq p(u, v) \leq 1$. Aquest pes representa la probabilitat que en agafar l'aresta (u, v) es produeixi el so $\sigma(u, v)$. Definim la probabilitat d'un camí com el producte de les probabilitats de les arestes del camí. D'aquesta forma, tota seqüència vàlida, que correspon a camins etiquetats al graf que comencen a v_0, tindrà associades probabilitats per a cadascun d'aquests camins.

Per exemple, la seqüència $s =$ (ɛ, ɐ, k, ə) és l'etiqueta del camí $0 \to 3 \to 5 \to 7 \to 12$ i del camí $0 \to 3 \to 6 \to 9 \to 10$. Si, $p(0, 3) = 0.1$, $p(3, 5) = 0.3$, $p(3, 6) = 0.2$, $p(5, 7) = 0.02$, $p(6, 9) = 0.1$, $p(7, 12) = 0.4$ i $p(9, 10) = 0.2$, aleshores el primer camí té probabilitat 0.00024 $(= 0.1 \cdot 0.3 \cdot 0.02 \cdot 0.4)$ de pronunciar s i el segon probabilitat 0.0004 $(= 0.1 \cdot 0.2 \cdot 0.1 \cdot 0.2)$.

(b) Modifiqueu l'algorisme de l'apartat **(a)** de manera que, donat un model (G, σ, p, v_0) i una seqüència de fonemes s, determini si s és vàlida i, en cas de que ho sigui, retorni un camí etiquetat amb s que tingui la probabilitat més gran de produir s.

Exercici 4 (2 punts) (Discurs). Davant el gran nombre de intervencions sobre la utilitat de l'ensenyament universitari, el rector de la UPC ha decidit encarregar a UPCnet un mecanisme per redactar aquest discursos.

UPCnet ha desenvolupat un model consistent en un graf dirigit $G = (V, E, \omega)$, on cada aresta és etiquetada amb una paraula, i un vèrtex distingit $t \in V$. Segons aquest model, un camí al graf que finalitza a t representa una frase amb sentit que es pot incloure al discurs. La frase es correspon amb la seqüència ordenada de mots que apareixien al camí, seguint l'ordre establert pel propi camí. Abans de preparar un discurs el rector decidirà els punts d'inici del seu discurs, tot indicant una seqüència de nodes del graf on han de començar les frases que el formaran.

Per tal d'eliminar redundàncies al discurs, l'equip rectoral aconsella que el discurs mai faci servir més d'una vegada una aresta del graf.

(a) Dissenyeu un algorisme per escriure un discurs que acompleixi les especificacions del rector i segueixi els consells de l'equip rectoral. En cas que no fos possible, s'ha de presentar una versió reduïda del discurs on també s'especifiquin els nodes d'inici de frase que no s'han pogut completar.

(b) Indiqueu com modificaríeu l'algorisme proposat a l'apartat previ si l'equip rectoral relaxa la condició prèvia de redundància i permet un nombre màxim c d'aparicions de cada aresta al discurs.