Πιθανοτικοί Αλγόριθμοι

Διδάσκοντες: **Σ. Ζάχος, Δ. Φωτάκης**

Επιμέλεια διαφανειών: **Δ. Φωτάκης**

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Εθνικό Μετσόβιο Πολυτεχνείο

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδεια Χρήσης

Πιθανοτικοί Αλγόριθμοι

- **Πιθανοτικός αλγόριθμος** κάνει τυχαίες επιλογές και εξαρτά εξέλιξή του από αυτές.
  - Κατανομή πιθανότητας πάνω σε ντετερμινιστικούς αλγόριθμους.
- Πλεονεκτήματα πιθανοτικών αλγόριθμων:
  - Απλότητα και κομψότητα (π.χ. quickselect, primality).
  - Συνήθως ταχύτεροι από ντετερμινιστικούς.
  - Όταν έχουμε μερική γνώση, περιορισμένη μνήμη, κλπ., πρακτικά αποτελούν μόνη αποδοτική λύση.
- Μειονεκτήματα:
  - Λάθος απάντηση (με μικρή πιθανότητα).
  - Κυμαινόμενος χρόνος εκτέλεσης.
  - Δύσκολο debugging.

Πώς τα Καταφέρνουν;

- Εκμεταλλεύονται «εργαλεία» της πιθανότητας.
- «Αδυνατίζει» (και γίνεται πιο ρεαλιστική) η χειρότερη περίπτωση (π.χ. quicksort).
- Τυχαία δειγματοληψία: αντιπροσωπευτικό δείγμα και λύση (π.χ. clustering, sublinear algs).
- Ικανό πλήθος πιστοποιητικών (βλ. property testing).
- Τυχαία μοιρασιά εργασιών: ισορροπημένη και με ελάχιστο κόστος (υπολογιστικό, επικοινωνιακό).
- Fingerprinting και hashing.
- «Σπάσιμο» συμμετρίας (π.χ. Ethernet, leader election).
- Προσομοίωση διαδικασιών και rapid mixing.

Γινόμενο Πολυωνύμων

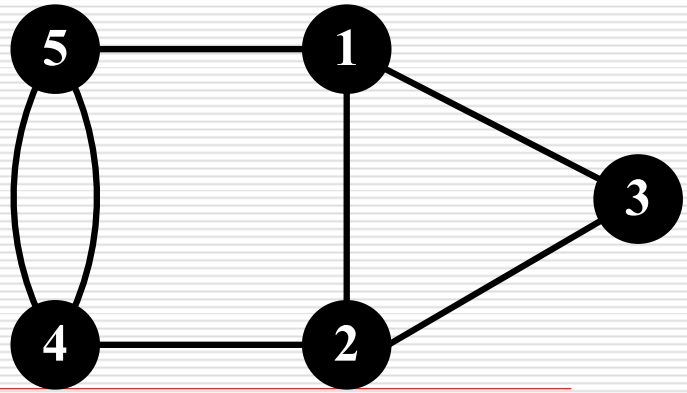
- Πολυώνυμα $P_1(x)$, $P_2(x)$, $P_3(x)$ ορισμένα σε field F.
- Έλεγχος αν $P_1(x) \times P_2(x) = P_3(x)$
  - … σε χρόνο (σημαντικά) μικρότερο του πολλαπλασιασμού;
- Ελέγχουμε αν $Q(x) = P_1(x) \times P_2(x) - P_3(x)$ είναι (ταυτ.) 0.
  - Έστω Q(x) βαθμού d και όχι (ταυτοτικά) 0.
    Για κάθε $S \subseteq F$, $Pr_{r \in S}[Q(r) = 0] \leq d/|S|$.
  - Για $|S| = 100d$ και 3 ανεξ. δείγματα, πιθαν. λάθους $\leq 10^{-6}$.
  - Χρόνος πολ/μού: $\Theta(d^2)$. Χρόνος ελέγχου: $\Theta(d)$.
- Επεκτείνεται σε πολυώνυμα πολλών μεταβλητών, όπου αντίστοιχη πιθανότητα ορίζεται με συνολικό βαθμό.
  - Θεώρημα Schwartz-Zippel.

Γινόμενο Πινάκων

- Δίνονται A, B, C πίνακες $n \times n$.
  - Έλεγχος αν $AB = C$ σε χρόνο $O(n^2)$.
- Τυχαίο διάνυσμα $r \in \{0, 1\}^n$. Απαντ. NAI ανν $A(Br) = Cr$.
  - Ισοδύναμα ανν $Dr = 0$, όπου $D = (AB - C)$.
  - Αν $D \neq 0$, D έχει μη μηδενικά στοιχεία.
    Χβτγ., κάποια στην 1η γραμμή του D, ένα στην 1η στήλη.
  - Για κάθε επιλογή των $r_2, \ldots, r_n$,
    υπάρχει μια (το πολύ) επιλογή για το r_1 τ.ω. $\sum_{j=1}^{n} D_{1j} r_j = 0$
  - Άρα πιθανότητα λάθους $\leq 1/2$.
  - Με π.χ. 30 ανεξάρτητες επαναλήψεις, πιθ. λάθους $< 10^{-6}$.

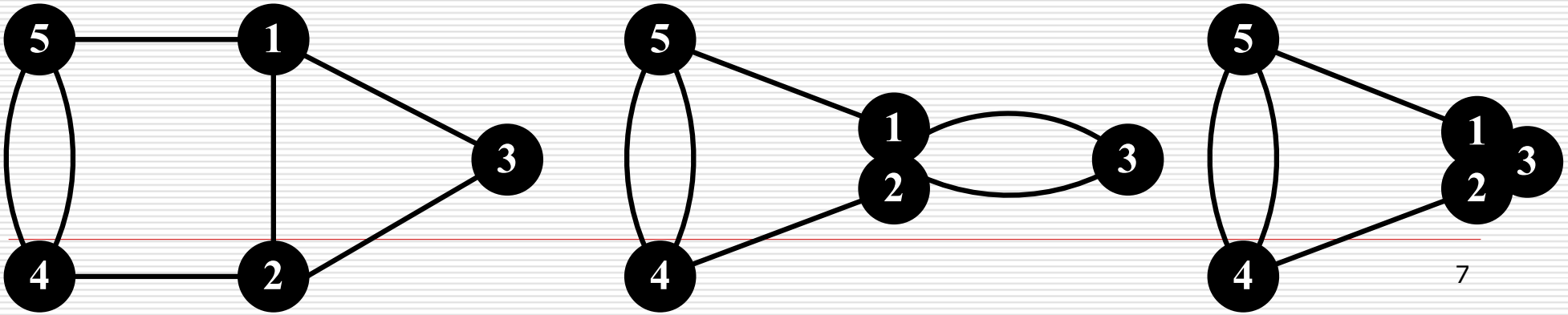
Ελάχιστη Τομή

- Μη κατευθυνόμενο συνεκτικό πολυγράφημα G(V, E).
  - Πολλαπλές ακμές, όχι χωρητικότητες / βάρη.
- Τομή: διαμέριση κορυφών (S, V \ S) με $\emptyset \neq S \subset V$.
  - Σύνολο ακμών που αφαίρεσή τους δημιουργεί τουλ. 2 συνεκτικές συνιστώσες.
  - Μέγεθος τομής $b(S, V \setminus S) = |\{\{u, v\} \in E : u \in S, v \notin S\}|$
- Πρόβλημα: υπολογισμός μιας ελάχιστης τομής.
  - Λύνεται σε χρόνο $O(n^4)$ με διαδοχικές εφαρμογές αλγόριθμου μέγιστης ροής.
  - Υπάρχουν εξειδικευμένοι αλγόριθμοι με χρόνο $O(n^3)$.



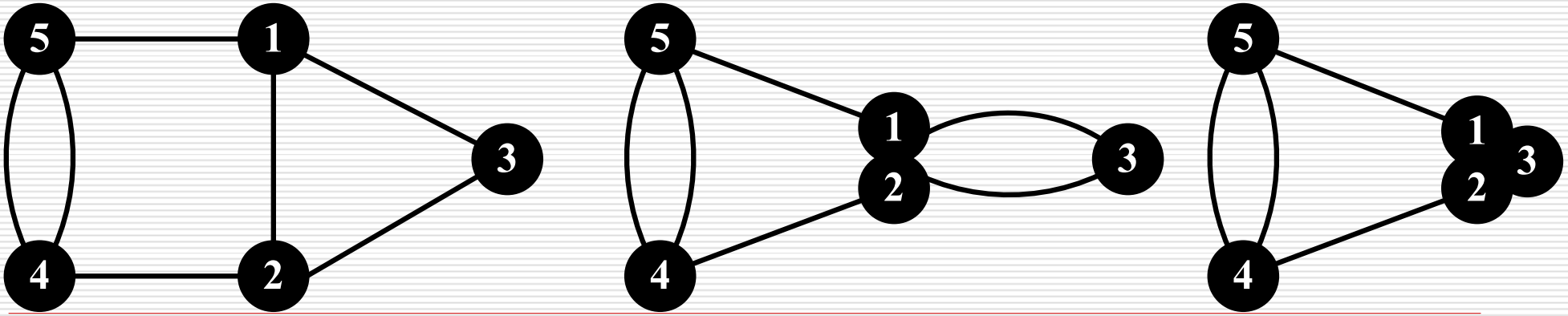
Σύμπτυξη Κορυφών

- **Σύμπτυξη** κορυφών u και v:
  - Αντικατάσταση u, v από μία νέα κορυφή uv.
  - Κάθε ακμή {x, u} / {x, v} αντικαθίσταται από ακμή {x, uv}.
  - Ακμές {u, v} παραλείπονται.
  - Διαδοχικές συμπτύξεις κορυφών 1, 2 και 12, 3.
- Τομή σε γράφημα μετά από διαδοχικές συμπτύξεις αντιστοιχεί σε τομή σε αρχικό γράφημα.
  - Λειτουργία σύμπτυξης **δεν** μειώνει ελάχιστη τομή.



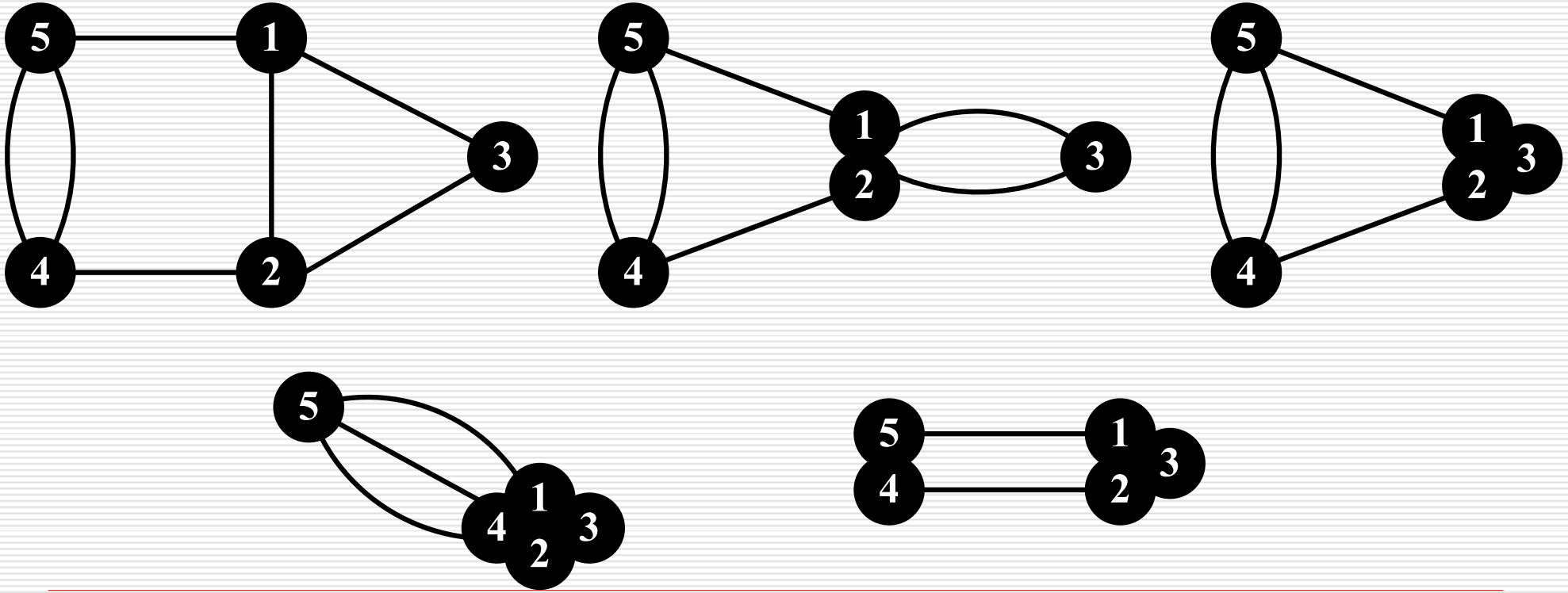
Πιθανοτικός Αλγόριθμος [Karger, 93]

- Ενόσω το γράφημα που απομένει έχει > 2 κορυφές:
  - Διάλεξε μια τυχαία **ακμή** {u, v}.
  - Αντικατέστησε γράφημα με αυτό που προκύπτει από σύμπτυξη κορυφών u και v.
- Ακμές τομής αυτές μεταξύ 2 κορυφών που απομένουν.
- Τομή ορίζεται από κορυφές που συμπτύχθηκαν στις 2 κορυφές που απομένουν.



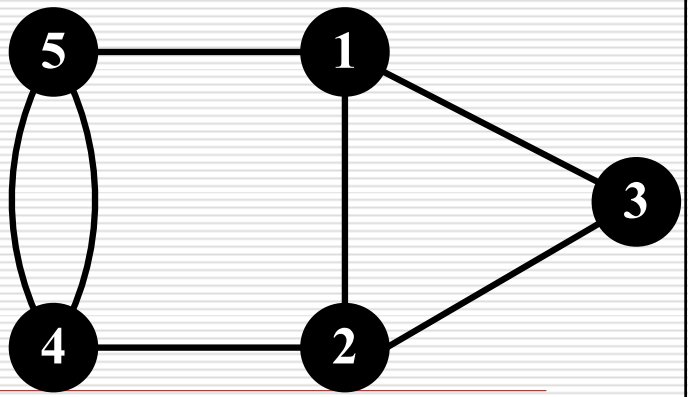
Παράδειγμα

- Αρχικές συμπτύξεις 1, 2, και 12, 3.
  - Σύμπτυξη 123, 4.
  - Σύμπτυξη 5, 4.



Πιθανοτικός Αλγόριθμος [Karger, 93]

- Βασικές ιδιότητες:
  - Πάντα τερματίζει έπειτα από $n - 2$ συμπτύξεις.
  - Υπολογίζει μία τομή, μπορεί **όχι** ελάχιστη.
  - Ποιά πιθανότητα p να καταλήξει σε ελάχιστη τομή;
  - Αν p όχι αμελητέα, μεγαλώνει γρήγορα με επαναλήψεις.
  - Αν $p \geq 2/n^2$, πιθανότητα τουλ. μία από $n^2 \ln n$ επαναλήψεις να καταλήξει σε ελάχιστη τομή $\geq 1 - 1/n^2$.
- Έστω ελάχιστη τομή $C = \{e_1, \ldots, e_k\}$ μεγέθους k.
  - Αλγ. επιστρέφει C ανν καμία από ακμές C δεν επιλεγεί για σύμπτυξη.



Πιθανότητα Επιτυχίας

- Συγκεκριμένη ελάχιστη τομή $C = \{e_1, \ldots, e_k\}$ μεγέθους k.
  - Πιθανότητα καμία από ακμές C **δεν** επιλέγεται για σύμπτυξη.
  - Ελάχιστος βαθμός κορυφής ≥ ελάχιστη τομή.
  - G(V, E) έχει ελάχιστο βαθμό κορυφής ≥ k.
    - G έχει #ακμών ≥ $nk/2$.
    - Πιθανότητα δεν επιλέγεται ακμή του C στην 1η σύμπτυξη: $p_1 \geq \frac{\frac{nk}{2} - k}{\frac{nk}{2}} = \frac{n-2}{n}$
  - Μετά από t συμπτύξεις, γράφημα έχει ελάχιστο βαθμό ≥ k.
    - #ακμών ≥ $(n-t)k/2$.
    - Πιθανότητα δεν επιλέγεται ακμή C του ούτε στην $(t+1)$η σύμπτυξη: $p_{t+1} \geq \frac{\frac{(n-t)k}{2} - k}{\frac{(n-t)k}{2}} = \frac{n-t-2}{n-t}$

Πιθανότητα Επιτυχίας

- Συγκεκριμένη ελάχιστη τομή $C = \{e_1, \ldots, e_k\}$ μεγέθους k.
  - Πιθανότητα καμία από ακμές C δεν επιλέγεται για σύμπτυξη:

$$p = p_1 \cdot p_2 \cdots p_{n-2} \geq \frac{n-2}{n} \cdot \frac{n-3}{n-1} \cdot \frac{n-4}{n-2} \cdots \frac{2}{4} \cdot \frac{1}{3} = \frac{2}{n(n-1)}$$

- Άρα $p \geq 2/n^2$, και πιθανότητα τουλ. μία από $n^2 \log n$ επαναλήψεις να καταλήξει σε **ελάχιστη τομή** $\geq 1 - 1/n^2$.
  - Χρόνος εκτέλεσης $O(n^2)$ / επανάληψη.
  - Συνολικός χρόνος $O(n^4 \log n)$.

Χρόνος Εκτέλεσης

- Όμως (σχετικά) μικρή πιθανότητα αποτυχίας στις πρώτες μισές συμπτύξεις!
  - Π.χ. πιθανότητα να μην συμπτυχθεί καμία ακμή C στις πρώτες (n–3)/2 συμπτύξεις $\geq 1/4$.
  - «Ακριβές» συμπτύξεις είναι «επιτυχημένες».
- Αναδρομική υλοποίηση σε φάσεις:
  - Εκτέλεση βασικού αλγόριθμου για n/2 συμπτύξεις 4 φορές.
  - Συνεχίσουμε αναδρομικά για καθένα από τα αποτελέσματα.
- Χρόνος εκτέλεσης $O(n^2 \log^3 n)$ για πιθανότητα επιτυχίας $= 1 - O(1/n)$.

Μέγιστη Τομή

- Μη κατευθυνόμενο γράφημα G(V, E) με m ακμές.
  - Αλγόριθμος γενικεύεται και όταν ακμές έχουν βάρη.
- Τομή: διαμέριση κορυφών (S, V \ S) με $\emptyset \neq S \subset V$.
  - Σύνολο ακμών που αφαίρεσή τους δημιουργεί τουλ. 2 συνεκτικές συνιστώσες.
  - Μέγεθος τομής $b(S, V \setminus S) = |\{\{u, v\} \in E : u \in S, v \notin S\}|$
- Πρόβλημα: υπολογισμός μιας μέγιστης τομής.
  - NP-complete, αλγόριθμος με λόγο προσέγγισης 0.878 [Goemans, Williamson, 94], τυχαία στρογγυλοποίηση.

Μέγιστη Τομή

- (Απλός) αλγόριθμος: κάθε κορυφή v εντάσσεται στο S ανεξάρτητα με πιθανότητα 1/2 (διαφορετικά στο V \ S).
  - X πλήθος ακμών στην τομή (S, V \ S) (τυχαία μεταβλητή).
  - Ακμή e «διασχίζει» τομή (S, V \ S) με πιθανότητα 1/2.
  - Αναμενόμενο πλήθος ακμών στην τομή (S, V \ S): E[X] = m/2 (γραμμικότητα μέσης τιμής).
  - Pr[X ≥ m/2] ≥ 2/(m+2)

$$\frac{m}{2} = \mathbb{E}[X] \leq (1 - \mathbb{P}\mathrm{r}[X \geq \tfrac{m}{2}])(\tfrac{m}{2} - 1) + \mathbb{P}\mathrm{r}[X \geq \tfrac{m}{2}]\, m$$

  - Πιθανότητα τουλ. μία από (m+2)ln n επαναλήψεις να καταλήξει σε τομή με τουλ. m/2 ακμες ≥ $1/n^2$.
  - Βέλτιστη λύση ≤ m.
  - Λόγος προσέγγισης 1/2, με πιθανότητα ≥ $1 - 1/n^2$.

Monte Carlo vs Las Vegas

- Monte Carlo αλγόριθμοι (π.χ. min-cut, max-cut):
  - Μπορεί να δώσουν λάθος απάντηση (με μικρή πιθανότητα), χρόνος εκτέλεσης ντετερμινιστικός (συνήθως!).
  - Πιθανότητα λάθους μπορεί να γίνει πολύ-πολύ μικρή με ανεξάρτητες επαναλήψεις.
  - Προβλήματα απόφασης: one-sided error και two-sided error.
  - Πολυωνυμικοί one-sided error αλγόριθμοι: RP και coRP.
  - Πολυωνυμικοί two-sided error αλγόριθμοι: BPP.
- Las Vegas αλγόριθμοι (π.χ. quicksort, quickselect):
  - Πάντα σωστή απάντηση, χρόνος εκτέλεσης τυχαία μεταβλητή.
  - Πολυωνυμικοί αλγόριθμοι: ZPP.

Χρηματοδότηση

Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα» του ΕΜΠ έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του υλικού.
Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.