A számítástudomány néhány trendje

Szubjektív és időkorlátos válogatás

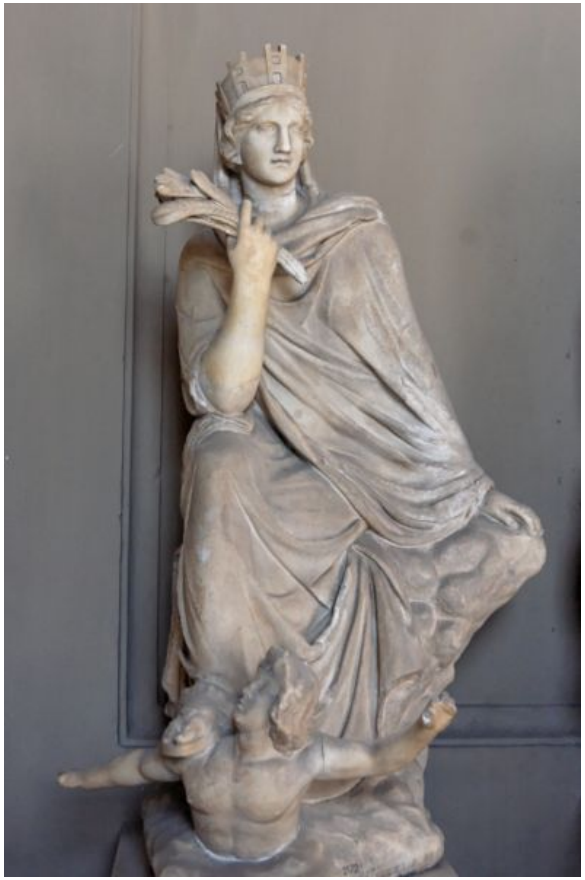
Rónyai Lajos
MTA SZTAKI, BME

Szegedi Egyetem, 2009. május 12.

Tartalom

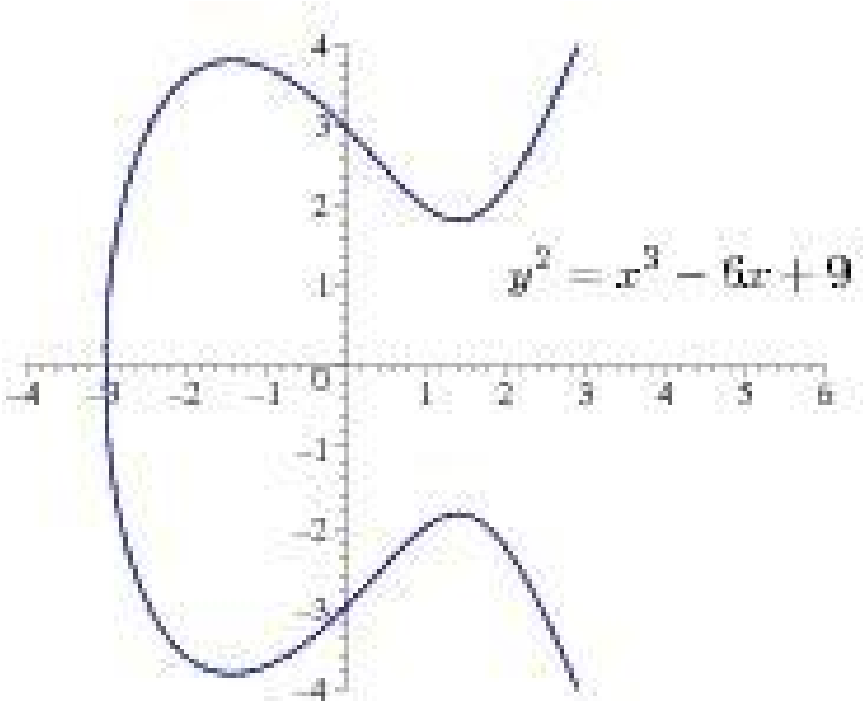
- Derandomizálás
- PCP-elmélet
- Meglepő kollektív alkalmazások a Világhálón
- Adatfolyamok kezelése
- Elérhetőség gráfokban
- Kvantumszámítások
- Genetikai számítások
- Szövegek matematikája
- Zárszó

Tud-e nagyot segíteni a véletlen?



Tud-e csodát tenni Tükhé érméje az algoritmusok világában? Levihető-e polinomiálisra ennél nagyobb számítási igény, ha véletlent használhatunk: igaz-e, hogy $BPP > P$?

Szemléletes válasz: igen



Olyan pontot keresünk, ami *nincs rajta* az adott görbén (felületen, stb.)

Nevezetes eszköz: a Schwartz–Zippel-lemma

Legyen $f(x_1, \ldots, x_n) \in \mathbb{R}[x_1, x_2, \ldots, x_n]$. Tegyük fel, hogy f nem azonosan 0 és a foka $\leq d$. Legyenek $a_1, \ldots, a_n$ egyenletes eloszlású teljesen független elemek az $\{1, 2, \ldots, N\}$ számhalmazból. Ekkor

$$\mathrm{Prob}(f(a_1, a_2, \ldots, a_n) = 0) \leq \frac{d}{N}.$$

Pl. ha $N = 4d$, akkor a valószínűség legfeljebb 1/4.

Szokásos szóhasználat: nyelv:= igen–nem-feladat

$P :=$ determinisztikus polinom időben felismerhető nyelvek

$BPP :=$ véletlennel polinom időben felismerhető nyelvek

$EXP := \cup_{k>0} TIME(2^{n^k})$ exponenciális idő

$E := \cup_{\delta>0} TIME(2^{\delta n})$ szelíd EXP

Különös-talányos kapcsolatok a nehéz és a véletlen fogalmai között

Nehéz és véletlen

- R. Impagliazzo, A. Wigderson, V. Kabanets és mások nyomán:
  - Vagy E-ben nincs exponenciális hálózati bonyolultságú feladat, vagy $BPP = P$
  - Vagy $EXP = BPP$, vagy a BPP-feladatokhoz van olyan determinisztikus, szubexponenciális algo., amely majdnem minden inputra jó. Emiatt nem lehet, hogy a véletlen egyszer exponenciálisat segít, máskor meg nem
  - Ha az azonosság-ellenőrzés (Schwartz-Zippel) derandomizálható, akkor bizonyos kemény szuperpolinomiális alsó korlátok érvényesek. Vagyis $BPP = P$-t bizonyítani (ha igaz) nem lesz sétagalopp

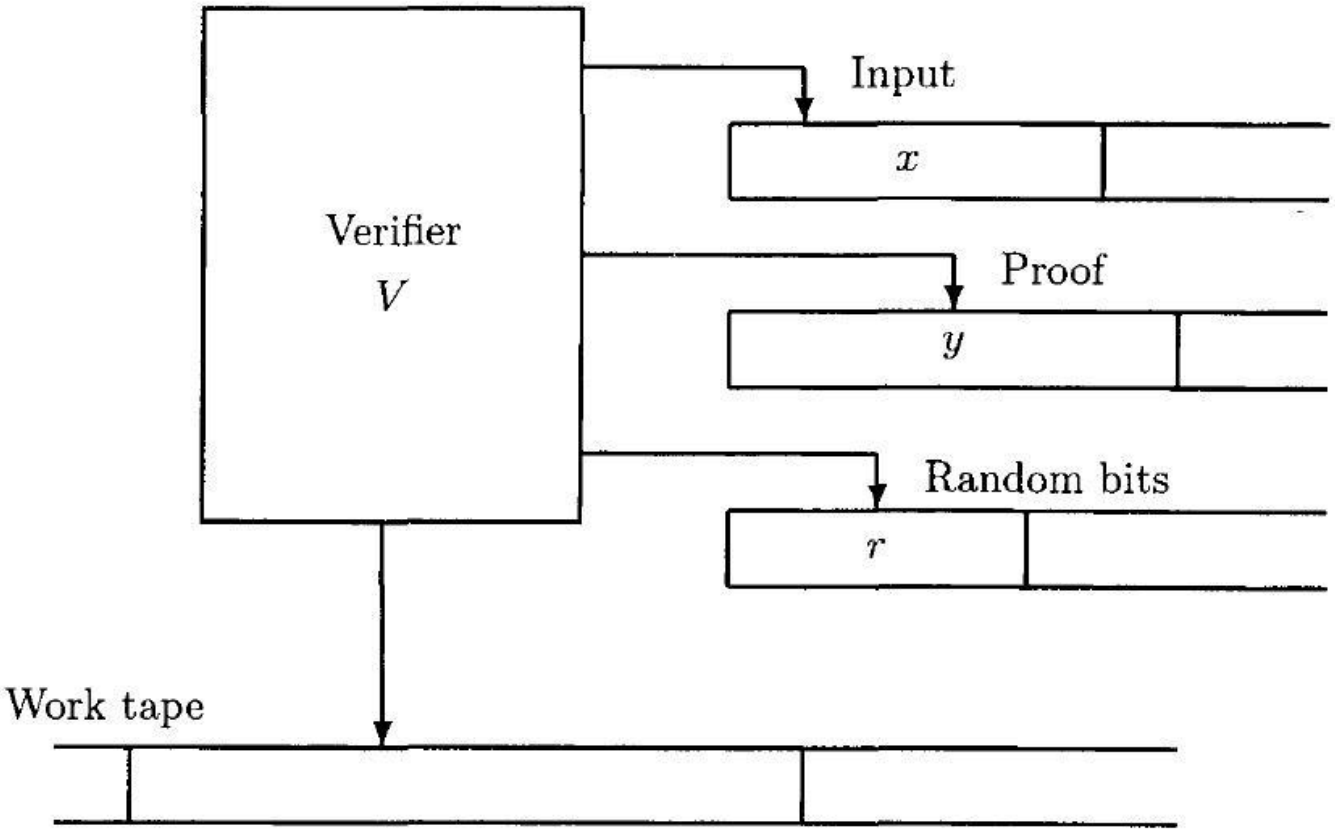
A. W.: Randomness is in the eye of the beholder

Véletlennel ellenőrizhető bizonyítások (PCP-elmélet)

PCP-Tétel. $NP = PCP(\log n, 1)$.

Másként mondva: minden $L \in NP$ nyelvhez létezik randomizált polinom idejű V TG a következő tulajdonságokkal.
1. Ha $x \in L$, akkor van olyan y bizonyíték, hogy V biztosan elfogadja (x, y)-t.
2. Ha $x \notin L$ akkor tetszőleges y esetén a V csak legfeljebb $1/2$ véggel fogadja el (x, y)-t.
3. V csak $O(\log n)$ véletlen bitet használhat.
4. V az y-ból csak $O(1)$ bitet olvashat el.

A V Turing-gép

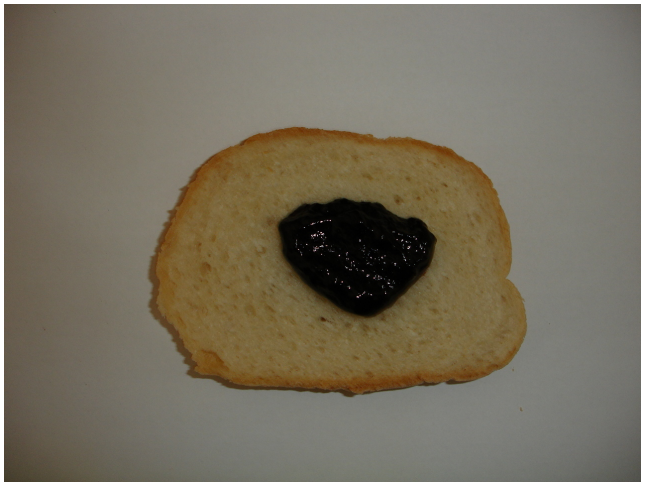


PCP-elmélet II.

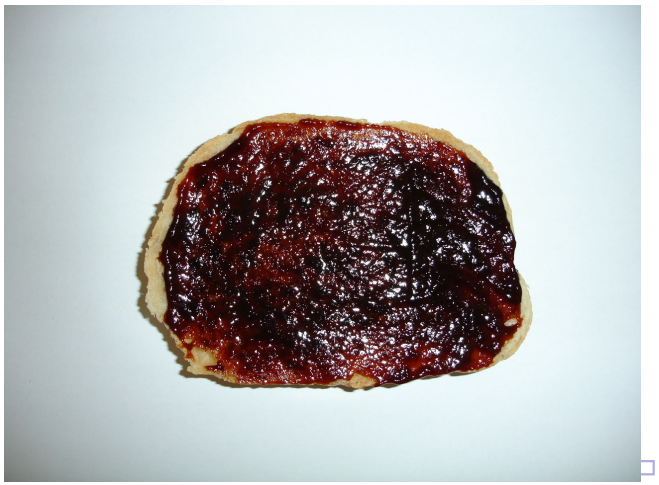
Meseszerűen: egy n hosszú y formális bizonyításból készíthető egy $poly(n)$ hosszú y^* bizonyítás, aminek a helyessége nagy biztonsággal tesztelhető egy olyan randomizált módszerrel, amely legfeljebb K (konstans sok) bitet vizsgál meg (szúrópróbaszerűen) y^*-ból.

A lekváros kenyér-hasonlat (I. Dinur nyomán)

Eredeti (hibás) bizonyítás:



Az új bizonyítás, szétkent hibával:



A PCP-tétel eredeti gondolatmenete igen nehéz és hosszú

Irit Dinur (2005): fogyasztható, tanítható gondolatmenet

`http://www.cs.princeton.edu/~chazelle/pubs/nature06.pdf`

A PCP-elmélet alkalmazásai

- PCP-tétel következménye: bizonyos NP-teljes feladatok közelítő változatai is nehezek (hardness of approximation). Pl. MAXKLIKK nem közelíthető hatékony algoritmussal $\frac{1}{n^{1-\epsilon}}$ szorzóval, kivéve, ha $P = NP$. ($\frac{1}{n}$ szorzóval triviálisan közelíthető).
- Lokálisan (szublineáris időben) tesztelhető/dekódolható kódok
- Privát információkérés
- Újfajta bizonyítást hordozó kódok (PCCs)??

A PCP-elmélet nagy nevei

Az 1993-as Gödel-díjasok: Babai László, Shafi Goldwasser, Silvio Micali, Shlomo Moran, Charles Rackoff az interaktív bizonyítások fogalmáért.

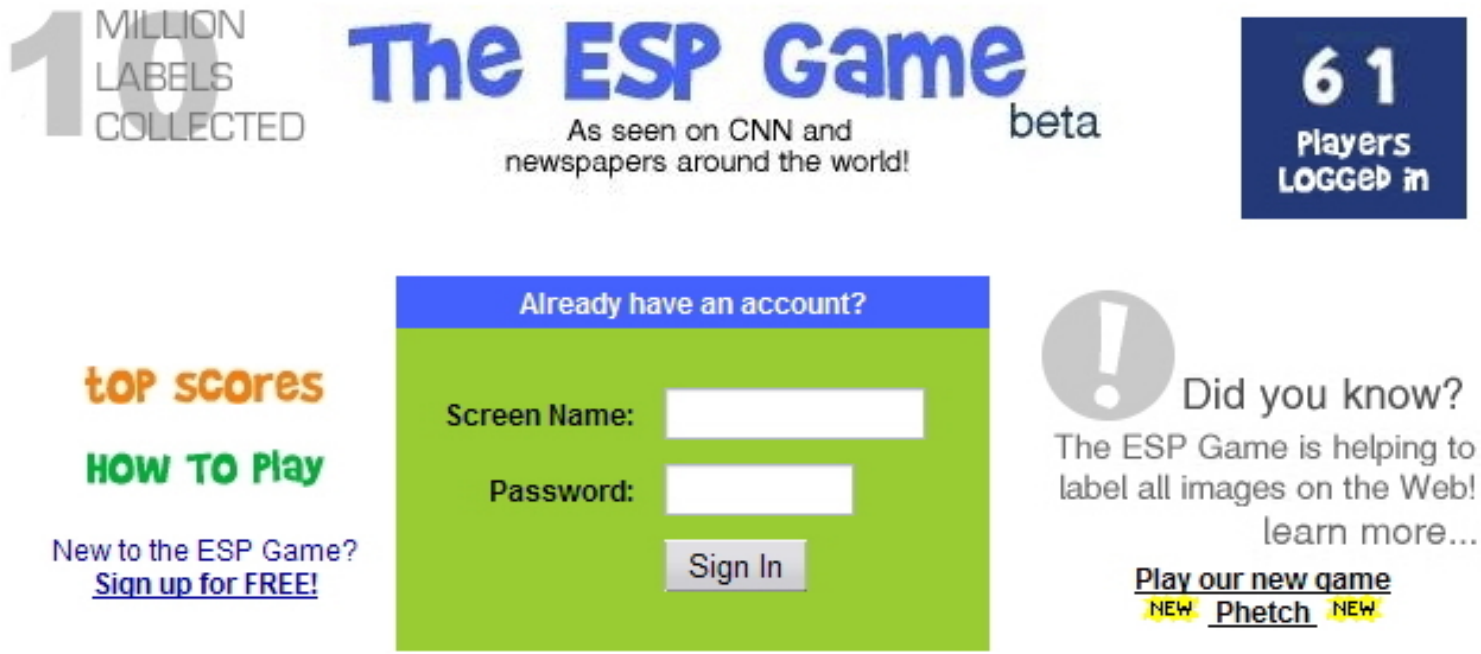
A 2001-es Gödel-díj kitüntetettjei: Sanjeev Arora, Uriel Feige, Shafi Goldwasser, Carsten Lund, Lovász László, Rajeev Motwani, Shmuel Safra, Madhu Sudan, Szegedy Márió a PCP-tételért, és az approximációs algoritmusok nehézségével kapcsolatos alkalmazásaiért.

Céltudatos játékok – L. von Ahn nyomán

- Milliárdnyi órát játszunk az interneten. Lehet-e ezt hasznosítani?
- Algoritmusok emberekre (human computation, social computation) szilikonmorzsák helyett
- Nagyméretű, elosztottan támadható, emberi intelligenciát igénylő feladatok, játékká édesítve

Példa: ESP Game (www.espgame.org): képek felcímkézése

- Két egymástól izolált játékos ugyanazt a képet látja, címkéket (jó kulcsszavakat) ír hozzá, ezzel megpróbálja kitalálni, hogy a másik mivel címkézte a képet
- A játék véget ér, ha van közös címke. Jó, ha gyorsan találnak ilyet...
- Potenciális képaláírás: gyakran (és gyorsan) kapott közös címkék


10 MILLION LABELS COLLECTED
The ESP Game beta
As seen on CNN and newspapers around the world!
61 Players LOGGED in
top scores
HOW TO Play
New to the ESP Game?
Sign up for FREE!
Already have an account?
Screen Name:
Password:
Sign In
Did you know?
The ESP Game is helping to label all images on the Web!
learn more...
Play our new game
NEW Phetch NEW
Terms of Service | FAQ | ESP Image Search | Contact Us | Credits
Funded in part by the National Science Foundation (NSF).
© 2005 Carnegie Mellon University, all rights reserved. Patent Pending.

The ESP Game beta

As seen on CNN and newspapers around the world!

27,282,200

Labels collected since October 5, 2003

(This number is updated every 12 hours)

Labeling an image means associating word descriptions to it, as shown below. Computer programs can't yet determine the contents of arbitrary images, but the ESP game provides a novel method of labeling them: players get to have fun as they help us determine their contents. If the ESP game is played as much as other popular online games, we estimate that all the images on the Web can be labeled in a matter of weeks!

guys
sitting
music
picture

Having proper labels associated to each image on the Internet would allow for very accurate image search, would improve the accessibility of the Web (by providing word descriptions of all images to visually impaired individuals), and would help users block inappropriate (e.g., pornographic) images from their computers.

close window

Kérdést/választ ösztönző hálózatok (Kleinberg-Raghavan)

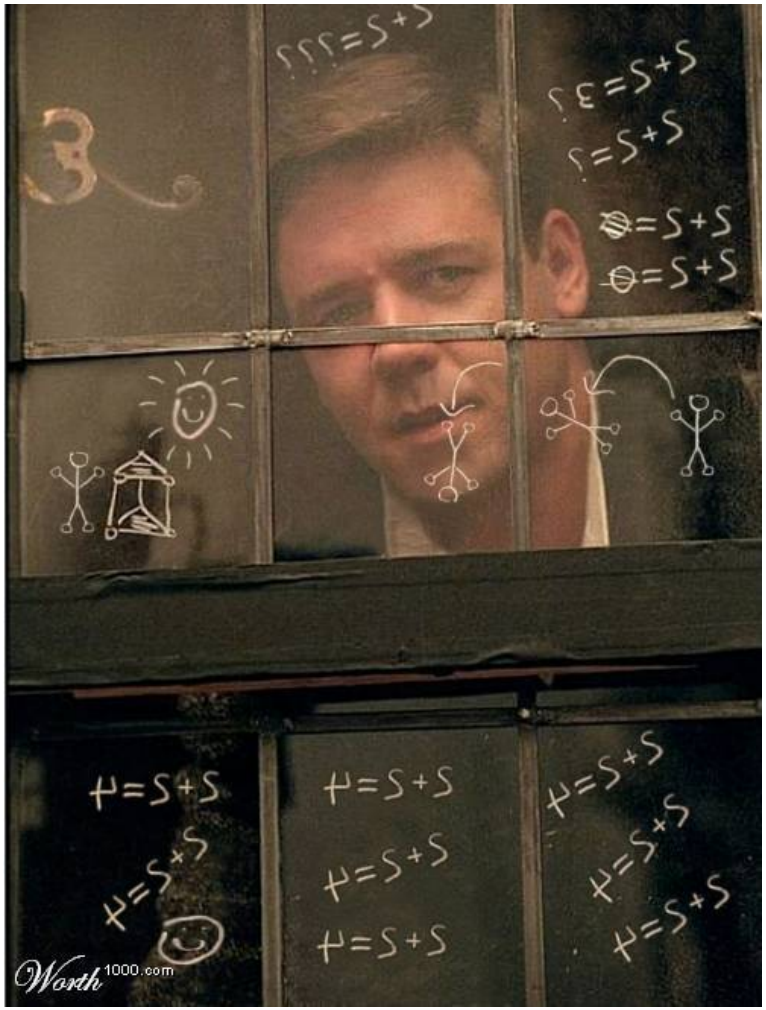
Bizonyos on-line közösségek játékelméleti modellje (pl. decentralizált file-cserélők, iwiw-típusú közösségi hálózatok)

Alaphelyzet: kérdést intézünk ismerőseinkhez, ezt ők továbbítják ismerőseiknek ...
Fizetünk érte a szomszédainknak, ebből ők továbbadnak valamennyit...

Modell: játék, aminek a csúcsok a résztvevői

Természetes Nash-egyensúlyi helyzete van, ez jól elemezhető

Nash–Crowe



Érdekes paraméterek: a válasz *ritkasága*, ami n, ha átlagosan n csúcs közül tudja valaki a választ

A hálózat *igazi elágazása*: a kezdőpontból indított mélységi bejárás során a csúcsok átlagos újszomszéd-száma

Kritikus viselkedés, ha az utóbbi érték 2

Fehérjék otthon

Napi csúcs: 5 PFLOPS (2009. február 18.), zöme GPU, PS3

Szóló szuperszámítógép (DOE, Roadrunner) napi csúcsa 2008-ban: 1,48 PFLOPS

Egy budapesti képernyő

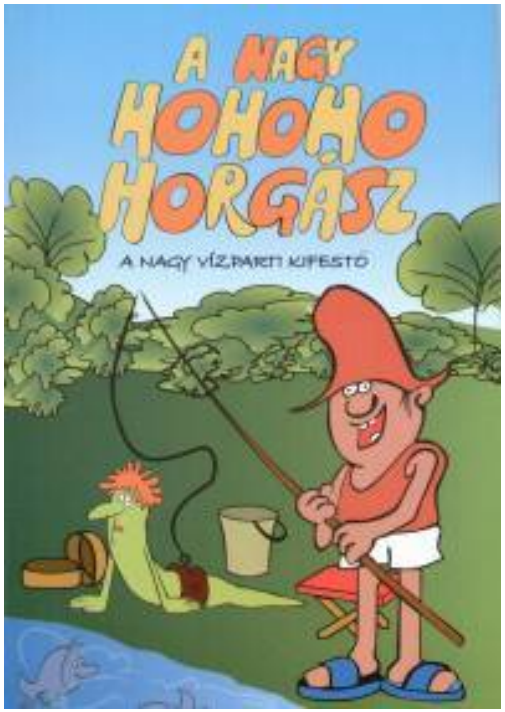
Adatfolyamkezelés (R. Motwani nyomán)

- **Hagyományos adatbázisok** véges, strukturált, perzisztens adathalmazok
- **Adatfolyamok** nem korlátos, elosztott, strukturálatlan, folyamatos, zajos, nagy sebességű, erős időfüggést mutató
- **Adatfolyamkezelő rendszerek (DSMS)** seregnyi fontos alkalmazás
  - Hálózatfelügyelet, hálózati forgalomtervezés
  - Telekom hívási adatok
  - Hálózatbiztonság
  - Pénzügyi adatfolyamok
  - Érzékelő-hálózatok
  - Weblogok
  - Gyártási folyamatok adatsorai
  - Gigantikus adathalmazok

Algoritmikus kérdések

- Modell
  - Folyamatos adatáram (vagy sok adat külső táron)
  - Adatmennyiséghez képest kicsi belső memória
- Technikák
  - Új/régi paradigmák
  - Negatív eredmények (alsó becslések)
  - Közelítés
  - Véletlen módszerek
- Bonyolultsági mértékek
  - Memóriahasználat
  - Idő/(érkező adategység) (gyakran valós idő)
  - Feldolgozó menetek száma

Illusztráció: Paul horgászik...



Paul horgászik, és az éppen fogott halakról üzenetet küld Carole-nak, aki ezt a folyamatosan érkező információt célirányosan feldolgozza

- Paul minden percben fog max. egy halat, H a fogható halak halmaza, $|H| = u$
- Carole percenként kapja a hírt, hogy Paul mit fogott éppen
- A $h \in H$ hal *ritka* a t időpontban, ha t-ig pont egyet fogott belőle Paul
- Legyen $\rho(t) :=$(a ritka halak száma t-ig)$/u$
- Carole célja: $\rho(t)$ számítása/karbantartása kis helyen, azaz $o(u)$ tárban

Tétel A feladat nem oldható meg (kicsi, vagyis $o(u)$ tárral)

Mert legyen $S \subseteq H$ tetszőleges.
Tfh. t_0-ig nincs kapás, utána t_1-ig pont az S-beli halak akadnak horogra, mindegyik egyszer.
A t_1 után Carole-nál teljes információ van S-ről.
Ugyanis, ha most $h \in H$ jön, akkor $h \in S$ pontosan akkor, ha h után a ρ csökken.
Ellentmondás: a H részhalmazainak leírásához legalább u bit kell.

Egy jó közelítés

- Carole választ k véletlen halat, egymástól teljesen ftlenül, $1/u$ véséggel: $H^* = \{h_1, h_2, \ldots, h_k\}$
- Csak a H^*-beli halakat számolja, legyen közülük a ritkák száma t-ig $r(t)$
- az $r(t)/k$ jó közelítése lesz $\rho(t)$-nek, ha pl. $\rho(t) \geq 1/k$

A ρ túl kicsi...

Gyakran az u túl nagy, ezért $\rho(t)$ túl kicsi, ami nagy k-t jelent...

Jobb mértéket kapunk, ha az eddigi ritka halak számát az eddig fogott fajták számával osztjuk: $\gamma(t)$

$\gamma(t)$-t is lehet gyorsan, kis helyen közelíteni

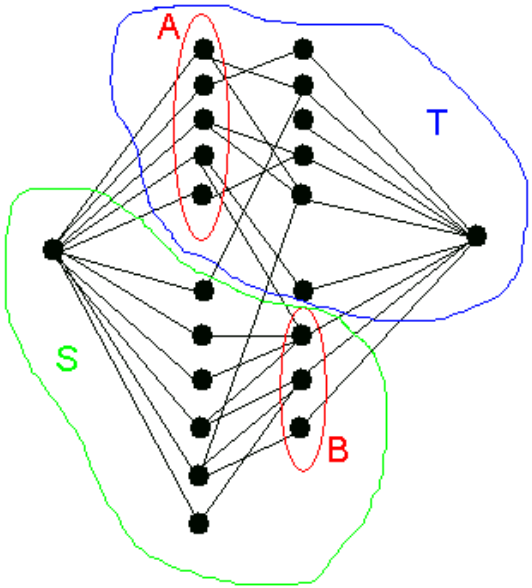
A megoldás itt is randomizált

Algebrai hátterű különleges hash-függvényeket használ

Elérhetőség gráfokban

USTCON: Legyen G egy adott, n pontú irányítatlan gráf, ennek S, T két kitüntetett csúcsa

Kérdés: van-e út S-ből T-be?

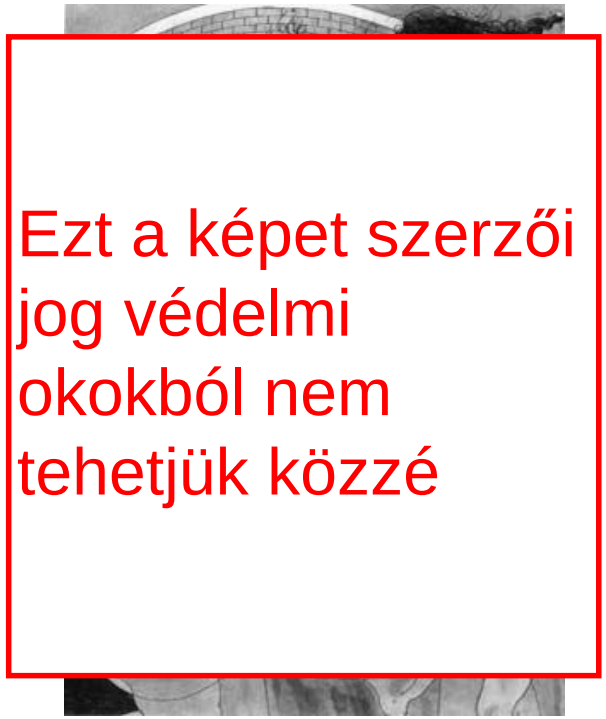


USTCON Lineáris időben megoldható, pl. BFS, DFS

Mekkora tárral oldható meg? Különösen érdekes, mert a feladat teljes az *SL* (szimmetrikus log tár) osztályban

Híres módszer 1979-ből (Aleliunas, Karp, Lipton, Lovász, Rackoff): randomizálva $O(\log n)$ tár elég
SŐT: a véletlen séta (egyenletesből választjuk a következő élet) várhatóan $\leq n^3$ lépésben elér S-ből T-be

Minótaurosz, Thészeusz, Ariadné és a Labirintus (H. Erni, 1997)



Thészeusznak nem kell a fonal, egy jó dobókocka is elég...

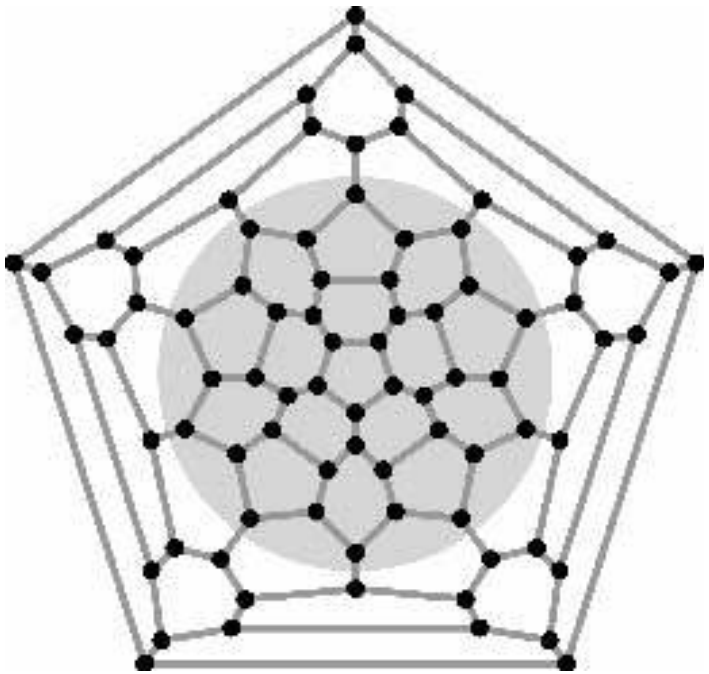
Omer Reingold (2005)

Tétel: $USTCON \in SPACE(\log n)$

- Best paper in ACM STOC 2005
- Hosszú fejlődés eredménye; egy fontos előzmény Ajtai–Komlós–Szemerédi (1987): ha L felismerhető $O(s)$ tárral, $O(s^2/\log s)$ véletlen bittel, akkor $L \in SPACE(s)$
- Expander-konstrukciók nagy szerepet játszanak a munkában
- Kicsi tár és derandomizálás

Talán nincs messze $RL = L$

Egy Ramanujan-gráf



80 csúcsú, harmadfokú reguláris gráf, a szürke rész expanziója 1/4

(Forrás: Peter Sarnak, Notices of AMS, 2004)

Kvantumszámítások

Richard P. Feynman javaslata

Hagyományos kétbites regiszter lehetséges tartalmai: 00, 01, 10, 11
Két qubites kvantumregiszter egy állapota

$$|\phi\rangle = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle,$$

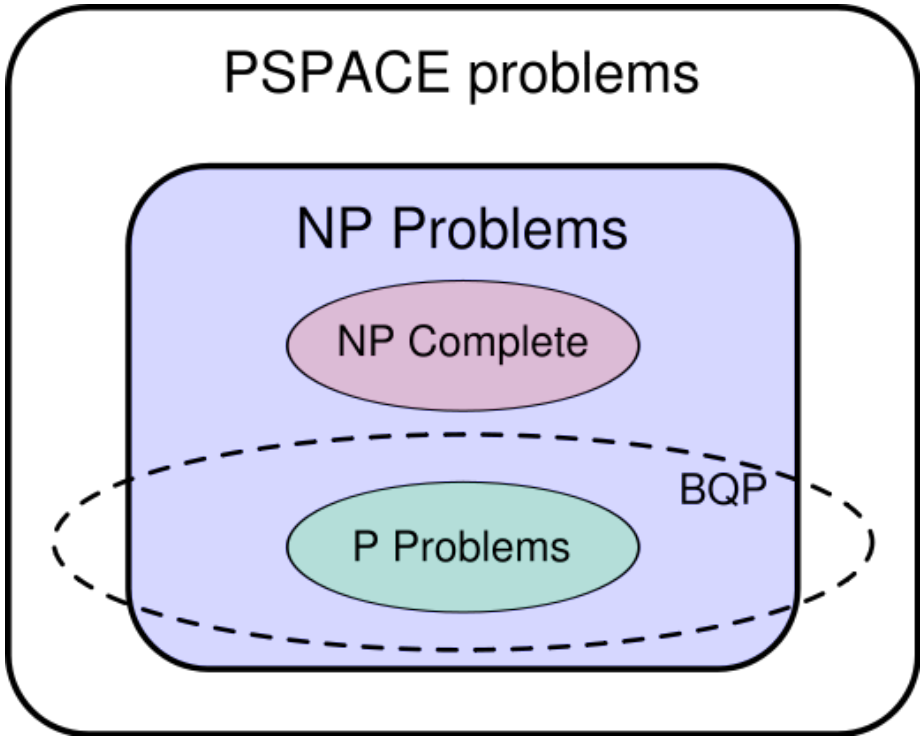
ahol a $v = (a, b, c, d)$ komplex vektor hossza 1

Számolás: lokális unitérekkel, ezek 1 és 2 qubiten ható transzformációk kiterjesztései
Mérés: a $|\psi\rangle$ mérésekor egy bázisállapotot kapunk, pl. $|11\rangle$ valószínűsége $|d|^2$

Színe és visszája

- Prímfelbontás, diszkrét logaritmus (DLOG) polinomidőben (P. Shor, 1994)
- Keresés n-elemű strukturálatlan táblában $O(\sqrt{n})$ időben (L. K. Grover, 1996)
- DE: nem világos, hogy lehet-e sok qubites kvantumgépet építeni

A kvantumosan kezelhető feladatok a térképen



Forrás: M. Nielsen, I. Chuang (2000). Quantum Computation and Quantum Information. Cambridge University Press

- Intenzíven kutatott terület világszerte
  - Új algoritmusok
  - Nagyobb gép
  - Kvantumkriptográfia (ketyere kapható!)
  - Szkepszis
- Több helyen is felbukkan itthon: BME, Rényi Intézet, SZTAKI
- Két érdekes dolgozat:
  Friedl, K; Ivanyos, G; Sántha, M: *Efficient testing of groups* ACM STOC 2005, 157-166.
  Ivanyos, G: *Deciding universality of quantum gates,* Journal of Algebra, 2007.

Rejtett részcsoportok

Adott: G csoport, és egy rajta értelmezett f függvény. Az f *rejt* egy (ismeretlen) H részcsoportot abban az értelemben, hogy $f(x) = f(y)$ pont akkor igaz, ha $x^{-1}y \in H$
Cél: találjuk meg H-t

Nevezetes speciális esetek: DLOG, gráfok izomorfizmusa

Polinomidejű kvantumalgoritmus ismert több érdekes speciális esetre

Ivanyos Gábor előadása az Akadémián
`http://www.math.bme.hu/akademia/ivanyos_gabor.pdf`

Genetikai kódsorozatok elemzése

- Igen erős trend a jelenben és a közeli jövőben is (biológia évszázada)
- Szövegkezelés giga-skálán (pl. teljes genom illesztés)
- Gyors gépek, gyors algoritmusok, csiszolt adatszerkezetek

Jeles részfeladat: leghosszabb növő sorozat keresése

$$\mathbf{3}, 9, 11, \mathbf{6}, \mathbf{7}, \mathbf{8}, 5, \mathbf{13}, 2, 4, \mathbf{17}$$

Erre $O(n \log n)$ idejű módszert használnak
Robinson–Schensted-algoritmus: egykor tiszta (magas) matematika...

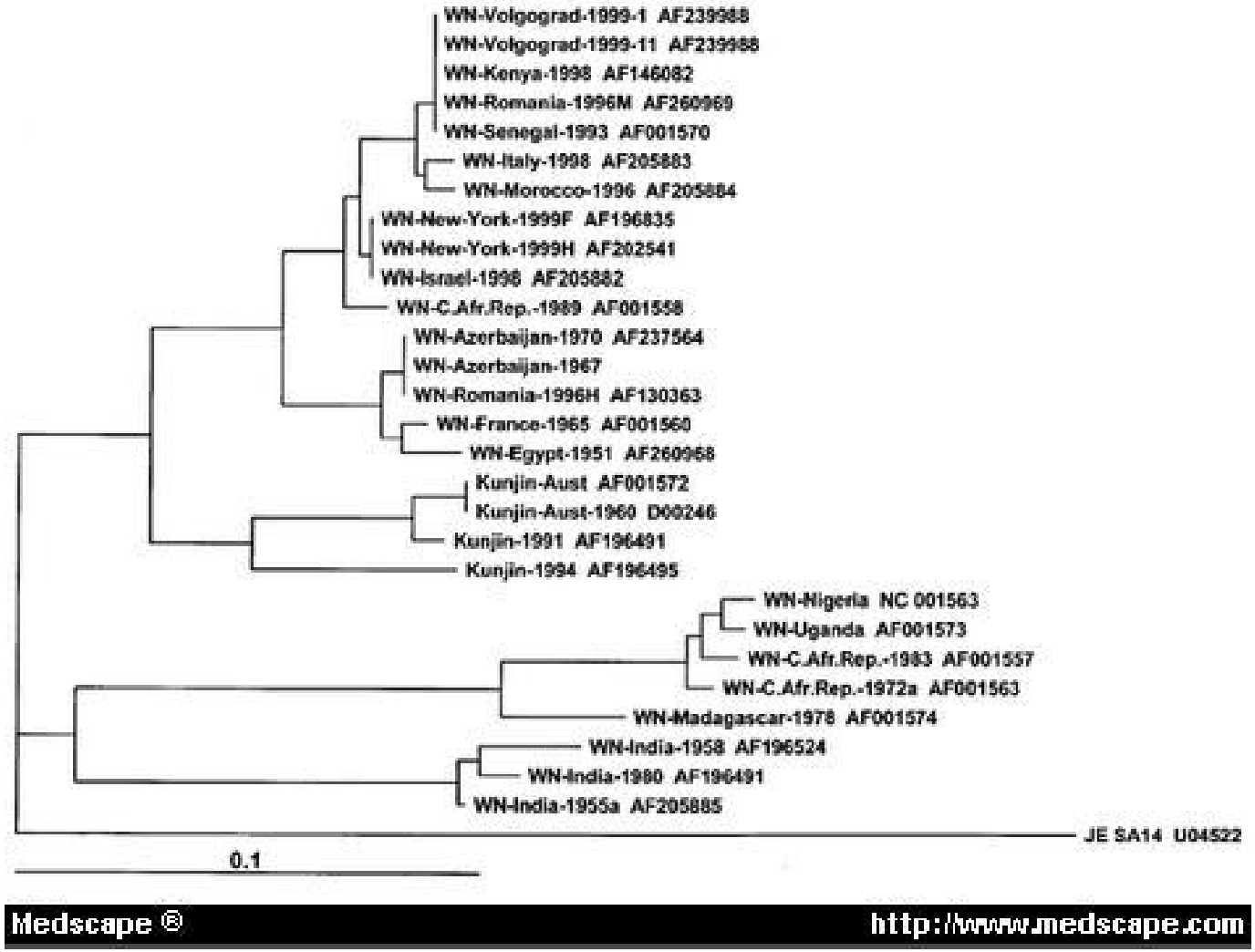
Vezérfa, avagy bátraké a cseresznye

Érdekes-különös matematikai/számítási problémák születnek Saitou–Nei-módszer vezérfa konstrukciójára:
Vezérfa: egy F fa(gráf) élein pozitív súlyokkal. A súlyok metrikát adnak a csúcsokon, $d(x, y)$ az x, y csúcsok távolsága

Cseresznye: két levél közös szomszéddal

Feladat: Pusztán a $d(u, v)$ értékek ismeretében (u, v levél) találjunk F-beli cseresznyét

Vezérfa a West Nile vírus variánsaival



Forrás: *www.medscape.com/viewarticle/414341_2*

Saitou N. és Nei M. – távolságokból jól számolható (lineáris) $\delta(u, v)$ függvény: legyen X az F fa leveleinek halmaza, és $v \in X$ esetén

$$T_v = \sum_{u \in X} d(v, u)$$

Ezután, ha $u, v \in X$, akkor

$$\delta(u, v) := d(u, v) - \frac{T_u + T_v}{r - 2}, \text{ ahol } |X| = r$$

Tétel: *Ha $x \neq y$ levelek, és ezek között $\delta(x, y)$ minimális, akkor x, y cseresznye*

Rengeteg alkalmazás, $\geq 10^4$ hivatkozás

Egy érdekes cikk a sajtóból

BusinessWeek online

TOP NEWS | BW MAGAZINE | INVESTING | ASIA | EUROPE

JANUARY 23, 2006

COVER STORY

Math Will Rock Your World

A generation ago, quants turned finance upside down. Now they're mapping out ad campaigns and building new businesses from mountains of personal data

COVER STORY PODCAST

Neal Goldman is a math entrepreneur. He works on Wall Street, where numbers rule. But he's focusing his analytic tools on a different realm altogether: the world of words.

-

Matematika a szövegek világában

- Statisztikai módszerek és gépi tanulás a fordításban
  Igen kevés nyelvészeti tudást használnak (mint a gyerekek?)
- Szövegek geometriai reprezentációja
  A geometria gráciája: árnyaltabb közelség, távolság, hasonlóság; csoportok képezhetők
  Egy sor másutt bevált elemzési/számítási eszköz alkalmazható

Példa: egy szó–dokumentum mátrix

- *1. dokumentum:*
  „Vagy vagy, vagy nem vagy”
- *2. dokumentum:*
  „Vagy vagy a Vagy, vagy vagy a Nem Vagy”
- *3. dokumentum:*
  „Ha vagy a Vagy, nem vagy a Nem Vagy, ha vagy a Nem Vagy, nem vagy a Vagy”

	a	ha	nem	vagy
1. dok.	0	0	1	4
2. dok.	2	0	1	6
3. dok.	4	2	4	8

Két szöveg hasonló, ha vektoruk közel egyirányba mutat

LSI: szinonimák, asszociatív kapcsolatok ábrázolása

- Téma: *zsivány*.
- Ismerjük fel, hogy a *bandita*, *betyár*, *útonálló*, *haramia*, *lator*, *rabló*, *Sobri*, *viszkis* is a témáról szól
- Akkor is, ha *zsivány* nem szerepel benne verbatim!

Mögöttes szemantikájú indexelés (LSI): Dumais, Deerwester, Berry, Landauer:

- Nyelvi kifejezés sokfélesége mint *zaj*, *bizonytalanság*
- Szó-dokumentum mátrixban indokoltnál több szabadsági fok
- Dokumentumkezelés rang-redukált téren: keresünk a mátrixhoz közeli, nagyon kicsi rangú mátrixot (SVD, Eckart–Young-tétel)

Végezetül

- Rengeteg minden kimaradt (pl. kriptográfia, szimuláció, szimbolikus számítások, adatmodellezés...)
- Színpompás tarka szőttes
- Felgyorsult innováció a fejlett világban
- Itthon: törekvések akadémiai/egyetemi kutatóhelyeken
- Európai remények