

Az SZTE Kutatóegyetemi Kiválósági Központ tudásbázisának  
kiszélesítése és hosszú távú szakmai fenntarthatóságának megalapozása  
a kiváló tudományos utánpótlás biztosításával”



# Eötvös Loránd Kollégium Matematika Műhely

## A borkóstolas matematikája, avagy hogyan mérjük a borok értékelőinek szakértelmét és ezáltal a borok minőségét?

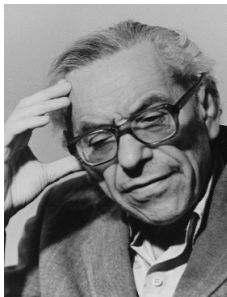
London András

2013. április 8.





## Erdős Pál



1913. március 26. – 1996. szeptember 20.





Több, széles körben elfogadott módszer létezik borok minősítésére

- **20 pontos értékelés:** *szín* (fehér: max. 2, vörös: max. 4), *tisztaság* (max. 2), *illat* (max. 4), *íz, összbenyomás* (fehér: max. 12, vörös: max. 10)
- **100 pontos bírálat** (Robert M. Parker): *megjelenés, tisztaság* (5, 4, 3, 2, 1), *szín* (10,8,6,4,2), *illat-tisztaság* (6,5,4,3,2), *intenzitás* (8,7,6,4,2), *minőség* (16,14,12,10,8), *zamat-tisztaság* (6,5,4,3,2), *intenzitás* (8,7,6,4,2), *minőség* (22,19,16,13,10), *hosszúság* (8,7,6,5,4), *összbenyomás* (11,10,9,8,7).





## Értékelés: borkóstolók, borversenyek

A hagyományos kiértékelés az, hogy 1-1 borkóstoló csoport pontjaiból a kilógó értékeket elhagyják (ha vannak), és a maradékot átlagolják.

Például: International Wine Challenge

- **Arany:** 95-100
- **Ezüst:** 90-94
- **Bronz:** 85-89
- **Ajánlott:** 80-84





Néhány további meghatározó forrás: „Eichelmann deutsche Weine” borkalauz, „Gault Millau WeinGuide”, Heiner Lobenberg: „Gute Weine”, „Mosel Fine Wines”

**Magyarországon:** Pannon Bormustra (legjobb vétel, fehér csúcsborok, vörös csúcsborok, édes csúcsborok, különdíjak)





## DE hogyan értékeljük a borkóstolók „teljesítményét”?

- Kik azok, akik tényleg „értenek” a borokhoz?
- Mennyire következetes a borok pontozása?
- Van-e olyan pontozó, aki (valamilyen értelemben) „csal” / manipulál?





**Cél:** olyan algoritmus fejlesztése, amely konzisztensen képes rangsorolni a borkóstolókat a borokra leadott pontjaik alapján.

*A borkóstoló értékelése azért fontos, mert ezáltal szeretnénk tisztábban látni a borok világában.*

Ha széles körben elfogadott módszert találunk: webes alkalmazás fejlesztése, élőpontrendszer.



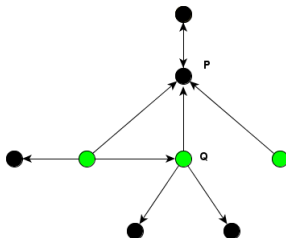


- Egy elismert borszakértő által adott pont, mint referencia érték (DE éppen a kóstoló hozzáértését szeretnénk „mérni”, nem feltételezve referencia értéket)
- Egyszerű statisztikai elemzés, majd az alapján való rangsorolás
- Hálózatos (~ gráfos) megközelítés
  - Előnyök:
    - borkóstolókat nem individuálisan, hanem együttesen vizsgáljuk (rendszerszemlélet)
    - hatékonyan bizonyult rangsoroló algoritmusok: PageRank (Brin és Page, 1998), HITS (Kleinberg, 1999)
  - Hátrányok:
    - nem egyértelmű hogyan és milyen élsúlyokkal definiáljuk a gráfot
    - távoli célok tekintetében: talán nehezen „emészthető” a borászvilág számára





PageRank (Brin és Page, 1998)



Rekurzív formula:

$$pr(P) = \sum_{Q \in N_{in}(P)} \frac{pr(Q)}{d_{out}(Q)} \quad (1)$$

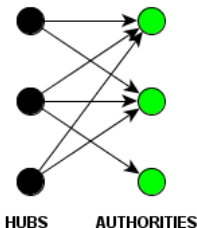
Ugró faktor (random szörföző),  $\lambda \in [0.1, 0.2]$

$$pr(P) = \frac{\lambda}{n} + (1 - \lambda) \sum_{Q \in N_{in}(P)} \frac{pr(Q)}{d_{out}(Q)} \quad (2)$$



A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

## Hyperlink-Induced Topic Search



$B(i)$ : azon pontok halmaza, melyekből van él  $i$ -be

$F(i)$ : azon pontok halmaza, melyekbe megy él  $i$ -ből.

*Authority pont*: pont, amely be-foka nem 0

*Hub pont*: pont, amely ki-foka nem 0



**Input**  $G$  directed graph

**Output** hub and authority scores of the nodes

- 1: Initialize all (node) weights to 1
- 2: **repeat**
- 3:   **for all** hub  $i \in H$  **do**
- 4:      $h_i = \sum_{j \in F(i)} a_j$
- 5:   **end for**
- 6:   **for all** authority  $i \in A$  **do**
- 7:      $a_i = \sum_{j \in B(i)} h_j$
- 8:   **end for**
- 9: **until** the weights converge
- 10: normalize





Legyen  $G = (X \cup Y, E)$  páros gráf  $X$  és  $Y$  színosztályokkal, ahol  $X = \{x_1, x_2, \dots, x_m\}$  és  $Y = \{y_1, y_2, \dots, y_n\}$

Legyen  $w(\overrightarrow{x_i y_j}) > 0$  (illetve  $w(\overleftarrow{x_i y_j}) > 0$ ) ha  $x_i \in X$  és  $y_j \in Y$  között van irányított él, különben legyen  $w(\overrightarrow{x_i y_j}) = w(\overleftarrow{x_i y_j}) = 0$ .  
(Feltehető, hogy  $\sum_{j \in Y} w(\overrightarrow{x_i y_j}) = 1$  és  $\sum_{i \in X} w(\overleftarrow{x_i y_j}) = 1$ .)

Tekintsünk egy véletlen sétát  $G$ -n úgy, hogy

$$w(\overrightarrow{x_i x_j}) = \sum_{k \in Y} w(\overrightarrow{x_i y_k}) w(\overleftarrow{x_j y_k}), \quad (3)$$

(Könnyen ellenőrizhető, hogy  $\sum_{j \in X} w(\overrightarrow{x_i x_j}) = 1$ .)



Az általánosított Co-HITS algoritmus (Deng et al., 2009)

Az alapötlet ugyanaz, mint a PageRank és a HITS  
 **SZÉCHENYI TERV**  
 esetén: iteratív módon értékeket rendelünk a gráf pontjaihoz.

Kezdetben legyen  $p_i^0$  az  $x_i$  és  $q_k^0$  az  $y_k$  értéke, majd

$$p_i = (1 - \lambda_x)p_i^0 + \lambda_x \sum_{k \in Y} w(\overleftarrow{x_i y_k})q_k, \quad (4)$$

$$q_k = (1 - \lambda_y)q_k^0 + \lambda_y \sum_{j \in X} w(\overrightarrow{x_j y_k})p_j. \quad (5)$$

(5)-öt (4)-helyettesítve kapjuk, hogy

$$p_i = (1 - \lambda_x)p_i^0 + \lambda_x(1 - \lambda_y) \sum_{k \in Y} w(\overleftarrow{x_i y_k})q_k^0 + \lambda_x \lambda_y \sum_{j \in X} w(\overrightarrow{x_j x_i})p_j. \quad (6)$$





Ha  $\lambda_x = \lambda_y = 1$ , akkor kapjuk, hogy

$$p_i = \sum_{j \in X} w(\overrightarrow{x_j x_i}) p_j \quad (7)$$

ami pontosan a HITS-egyenlet.

Ha pedig  $\lambda_y = 1$ , kapjuk, hogy

$$p_i = (1 - \lambda_x) p_i^0 + \lambda_x \sum_{j \in X} w(\overrightarrow{x_j x_i}) p_j, \quad (8)$$

ami nem más, mint a PageRank-egyenlet.



## A borkóstoló rangsorolási modell



Automatikus és objektív eljárást keresünk, minimális emberi beavatkozással, ami rangsorolja a kóstolókat a borokra adott pontjaik alapján

Rendelkezésre álló adatsorok:

- A 2009-es Szegedi Borfesztivál borversenyének adatai
- Villányi vörösborok versenyének adatsora



## A borkóstoló rangsorolási modell



## Feltevések:

- ① Első lépésként a borokat rangsoroljuk a rájuk leadott pontok alapján (azaz nincs referenciaérték)
- ② A borkóstolókat csak a borokra leadott pontok alapján rangsoroljuk (azaz csak egymással hasonlítjuk össze őket és így alakítjuk ki a rangsort)
- ③ Feltételezzük, hogy nincs „manipuláló” a kóstolók között (azaz többé-kevésbé azonos skálán pontoznak) *Példa: a szomszédai átlagát adó beugró kóstoló esete* (továbbá ld. Arrow-tétele)





## A borkóstoló rangsorolási modell



$G = (X \cup Y, E)$  páros gráf

$X$ : a borkóstolók halmaza

$Y$ : a borok halmaza

Minden  $x_i \in X$ -hez kezdetben ugyanaz a  $p^0$  kezdőértéket rendeljük (pl. legyen 1).

Legyen  $w'(\overrightarrow{x_i y_j})$   $x_i$  kóstoló által adott pont az  $y_j$  borra, normálva pedig  $w(\overrightarrow{x_i y_j}) = w'(\overrightarrow{x_i y_j}) / \sum_{j \in Y} w'(\overrightarrow{x_i y_j})$

$w(\overrightarrow{x_i y_j})$  definíciója természetesnek tűnik. Kérdés, hogyan definiáljuk a „bortól a kóstolóhoz vezető élt” pusztán a borra adott pontok ismeretében.



## A borkóstoló rangsorolási modell



Kiindulási alap, hogy a szakma azt elfogadja, hogy a jó borkóstolók konzisztens ítéleteket hoznak: azonos tesztanyagon az ő szórásuk kisebb, mint a laikusoké.

Az  $y_j$  bor kezdeti  $q_j^0$  értéke legyen a borra adott pontszámok átlaga (v.ö. 1. feltevés)



A borkóstoló rangsorolási modell



Tegyük fel, hogy az  $y_j$  bort  $1 < \ell \leq m$  kóstoló kóstolta és legyen  $D$  az összege a kóstolók által adott pontok átlagtól való eltérésének:

$$D = \sum_{i \in X} |q_j^0 - w'(\overline{x_i y_j})|, \quad (9)$$

Ezek után legyen a keresett élsúly:

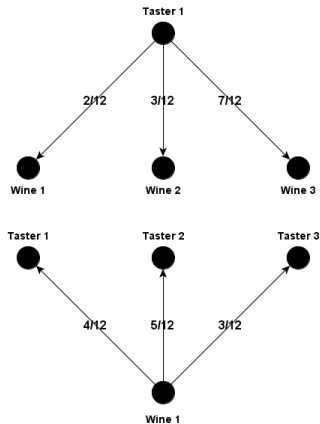
$$w(\overline{x_i y_j}) = \frac{|D - |q_j^0 - w'(\overline{x_i y_j})||}{(\ell - 1)D}. \quad (10)$$

(Itt is igaz, hogy  $\sum_{i \in X} w(\overline{x_i y_j}) = 1$ .)



### A borkóstoló rangsorolási modell

**Példa:** egy kóstoló által adott pontok rendre 20, 30, 70 (felül), míg egy bor rendre 20, 30, 70 pontot kapott 3 különböző kóstolótól (alul).



1. ábra.



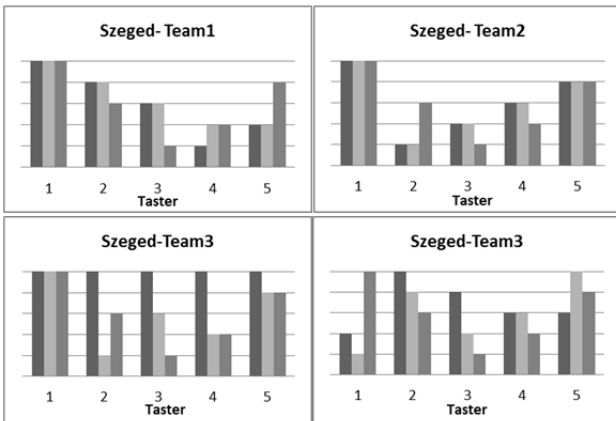


A HITS algoritmussal kapott eredményeket két egyszerű statisztikával hasonlítjuk össze:

- 1 Átlagtól való eltérések összege alapján rangsorolunk [SM1] (legkisebb eltérés adja az első helyezést)
- 2 Az átlagértékekkel való korreláció alapján rangsorolunk [SM2] (legjobb együttmozgás adja az első helyezést)



Szegedi borfesztivál, 2009



2. ábra. A sötét szürke, világos szürke, szürke oszlopok rendre a Co-HITS, SM1 és SM2 által adott helyezést mutatják (1-5-ig).



Villányi borok



	Co-HITS	Correlation	Sum of Diff
Romsics	1.00	0.75	0.86
Kertai	0.91	0.76	0.72
Kalocsai	0.93	0.62	0.71
Mészáros	0.90	0.48	0.58
Varga D.	0.92	0.70	0.79
Vörös G.	0.97	0.70	1.00
<i>Darrel (Master of Wine)</i>	0.97	0.61	0.91





## Néhány észrevétel:

- SM1 esetén az átlaghoz való közelség számít a rangsor alakulásánál, míg SM2 esetén az átlaggal való erősebb együttmozgás adja a jobb helyezést  $\implies$  Lehetőség a manipulációra
- A hálózati algoritmus nem pusztán az egyes kóistolók adatait veszi csak figyelembe, hanem egyszerre tekinti az összes adatot
- A két statisztikai módszer csak akkor működik, ha minden kóstoló ugyanazokat a borokat kóstolta, míg a hálózati algoritmus esetén ez nem feltétel  $\implies$  internetes adatbázis esetén is jól használható
- Az élsúlyozás módosítása más/általánosabb felhasználási lehetőségeket tesz lehetővé







## További tervek:

- Borkóstolók rangsorolására az algoritmus továbbfejlesztése, módosítása
- Null modellek, mesterséges adatok segítségével vizsgálni az előnyeit, alkalmazhatóságát
- Alkalmazás más értékelési rendszerek esetén:
  - Sportok: műkorcsolya, műugrás, szinkronúszás, síugrás
  - Egyéb versenyek: szépségverseny, dalverseny
  - Más kóstolók: főzőversenyek, sörivás, whiskey :-)





This work was partially supported by the European Union and the European Social Fund through project **FuturICT.hu** (grant no.: TÁMOP-4.2.2.C-11/1/KONV-2012-0013). The authors are gratefully thanks to to András Pluhár and Márk Jelasity for their useful advises and to Melinda Braun for typing the wine tasting data.

Köszönöm a figyelmet!

