

# Grafuri și clasamente

## 1 Descrierea problemei

În cazul unui singur turneu, cum e cel pe care l-ați simulat în Tema 1, realizarea clasamentului participanților nu pune probleme: locul întâi este ocupat de câștigătorul finalei, pe ultimul loc se clasează jucătorul/echipa care nu a câștigat niciun meci etc. Însă atunci când se iau în considerare rezultatele în mai multe competiții, lucrurile se complică: doi jucători pot avea același număr de victorii, însă dacă unul din ei a avut o victorie împotriva unui campion mondial, această victorie ar trebui să conteze mai mult.

Din acest motiv, în ultima vreme, atunci când se dorește realizarea clasamentelor (pentru sportivi, spre exemplu), se folosesc metode care analizează proprietățile rețelei [2, 3, 4], pentru a putea distinge între astfel de situații.

### 1.1 Modelarea victoriilor utilizând grafuri

Rețeaua sportivilor este un graf orientat în care fiecare vârf reprezintă un sportiv, iar între doi sportivi  $i$  și  $j$  există muchie de la  $i$  către  $j$  dacă cei doi au disputat un meci, iar  $j$  a câștigat. În cazul în care cei doi joacă mai multe meciuri, reprezentarea poate fi de multigraf (câte o muchie pentru fiecare meci) sau se pot agrega rezultatele tuturor meciurilor disputate de cei doi. De asemenea, graful poate fi ponderat, unde ponderea muchiei reprezintă numărul de victorii ale sportivului  $j$  în meciurile cu  $i$ .

Figura 1.1, reprodusă aici din lucrarea [5], prezintă în panoul A graful jucătorilor de tenis care au fost clasati numărul 1 în ATP. Nuanța de gri a muchiilor codifică ponderea acestora. Arborele din panoul B corespunde unui singur campionat, cum este cel pe care l-ați simulat în Tema 1, și este, practic, un subgraf al grafului mare. Panoul C reprezintă graful asociat acestui campionat.

### 1.2 Calculul clasamentului

În general, metodele bazate pe grafuri pentru calculul clasamentelor utilizează anumite proprietăți ale structurii grafului pentru a determina importanța unui jucător. Metoda pe care o veți implementa se aseamănă cu algoritmul PageRank [1], creat de Google pentru a clasa site-urile web.

Implementarea PageRank utilizează noțiuni pe care le veți învăța abia în Anul II (vectorii proprii ai grafului), însă principiul îl putem descrie pe scurt acum. PageRank se bazează pe conceptul de **centralitate**, care este o măsură a influenței unui nod dintr-o rețea. Un anumit nod va fi important dacă are multe legături către noduri cu scor mare (adică importante la rândul lor). Concret, legăturile pe care un nod le are cu noduri cu scor mare contribuie mai mult la scorul respectivului nod decât legăturile cu noduri cu scoruri mici.

Această idee este utilizată și în [5] pentru a clasa jucătorii de tenis, metodă pe care o veți implementa și voi. Calculul clasamentului pentru întreg graful presupune rezolvarea unui sistem de ecuații pentru care nu se cunoaște o soluție analitică, dar care poate fi calculată prin metode iterative (mai multe despre subiectul acesta în anul II).

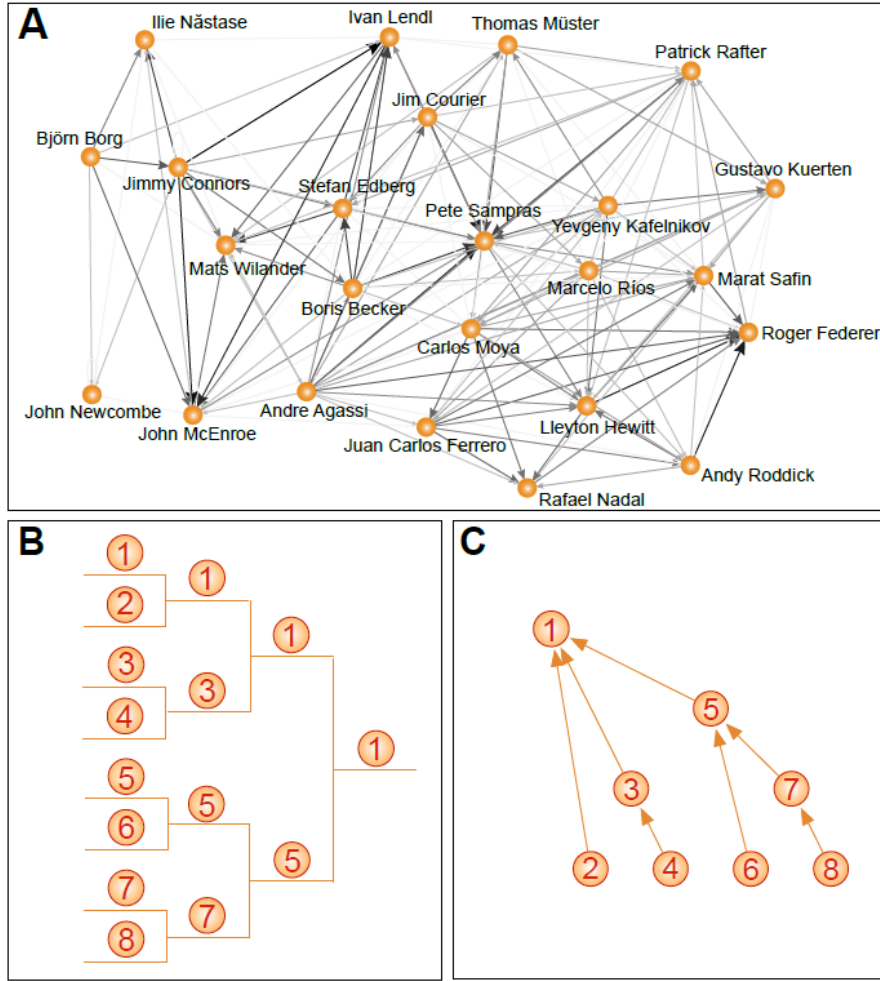


Figura 1: Rețeaua de jucători de tenis

Însă pentru un singur campionat, soluția se poate afla analitic, așa că o veți implementa pentru campionatul din Tema 1.

Notăm cu  $l$  numărul de meciuri necesare câștigării campionatului. Acesta corespunde înălțimii arborelui campionatului ( $l = 3$  pentru arborele din Figura 1.1, panoul B).

Notăm cu  $r$  numărul de victorii pe care le-a înregistrat un jucător în respectivul campionat. Scorul asociat cu  $r$  victorii, notat  $P_r$ , va fi

$$P_r = \frac{q(2-q)^r}{2^l + (2-q)^l(q-1)} \quad (1)$$

unde  $q \in [0, 1]$  este un parametru prin care se stabilește importanța numărului de victorii în scorul final: dacă  $q = 1$ , atunci toți jucătorii au același scor, iar cu cât  $q$  este mai mic, cu atât fiecare victorie în plus contribuie mai mult la scorul jucătorului (vedeți [5], Figura 3, pentru mai multe detalii).

În formula de mai sus este evident că scorul depinde de numărul de victorii câștigate și de numărul de victorii posibile și pare firesc ca aceste două mărimi să fie luate în considerare în calculul clasamentului. Ce nu apare evident, însă, este faptul că scorul ia în considerare structura rețelei, așa cum am specificat la începutul secțiunii. În realitate,

Ecuatia (1) este obținută dintr-o relație de recurență în care la începutul campionatului toți jucătorii au același scor, iar pe măsură ce câștigă un meci (se avansează în recurență), scorul lor crește și este redistribuit și către jucătorii cu care dispută meciuri în continuare. În felul acesta se asigură că importanța unui nod depinde de importanța nodurilor pe care le-a învins (observația despre victoria împotriva unui campion mondial de la început). Pentru mai multe detalii și relația de recurență, vedeți [5].

Așadar, scorul fiecărui jucător va fi dat de scorul asociat cu numărul de victorii,  $r$ , pe care le-a câștigat în timpul campionatului. Desigur, toți câștigătorii unei etape vor avea același scor (spre exemplu nodurile 1, 3, 5, 7 în prima etapă a campionatului din Figura 1.1, panoul B). Aceasta se întâmplă pentru că luăm în considerare exclusiv un anume campionat; dacă am lucra pe întreg graful jucătorilor, după fiecare meci jucătorii ar avea scoruri diferite.

Observați, de asemenea, că scorul depinde exponențial de înălțimea arborelui, ceea ce înseamnă că dacă un campionat are mai multe etape, el îi va aduce jucătorului un scor mai mare.

## 2 Cerințe

Fișierele din directorul `Input` al checker-ului conțin câte 32 de echipe care intră în turneu (similar cu cele rezultate în urma cerinței 2 din Tema 1), fiecare pe câte un rând.

Formatul este următorul: punctajul echipei (`float` formatat la 2 zecimale), urmat de spațiu, urmat de numele echipei.

Punctajul echipei reprezintă scorul de intrare în concurs (aka palmaresul până în acel moment).

Meciurile se desfășoară în aceeași maniera ca la Tema 1:

- Primele două echipe dispută un meci. Câștigă echipa cu scorul cel mai mare, iar dacă două echipe au același punctaj, atunci se folosește criteriul alfabetic (prima echipă în ordine alfabetică e cea câștigătoare). Se introduce echipa câștigătoare într-o coadă, iar cea învinsă în altă coadă.
- Se procedează la fel pentru restul echipelor, până se încheie prima etapă a campionatului.
- Se trece la următoarea etapă, unde câștigătorii dispută o nouă rundă de meciuri.
- Se procedează la fel până se obține un singur câștigător.
- După fiecare meci, **nu se mai crește scorul echipei câștigătoare, ca la tema 1.** Palmaresul rezultat în urma turneului va fi calculat la sfârșit, în funcție de numărul de victorii în competiție al fiecărei echipei, folosind relația (1).

1. Creați graful turneului (va semăna cu cel Figura 1.1, panoul C).

Vârfurile grafului vor fi cele 32 de echipe care au intrat în turneu. Spre exemplu (orientativ), fiecare vârf al grafului poate avea asociate următoarele câmpuri:

- un `int`, ce reprezintă poziția echipei în lista din fișierul de input.
- un `char*`, reprezentând numele echipei.
- un `float`, reprezentând scorul echipei la intrarea în concurs.

Scrieți matricea de adiacență a grafului orientat rezultat după turneu într-un fișier.

2. Calculați prestigiul fiecărei echipe, ca urmare a participării la turneu, utilizând formula din (1), cu  $q = 0.15$ . Scrieți într-un (alt) fișier prestigiul echipei, pe 4 zecimale, urmat de spațiu și numele echipei.

Echipele vor fi ordonate în funcție de locul lor în clasament, cu echipa câștigătoare

pe ultimul rând (în ordinea în care au pierdut meciurile).

**Observație:** Acest scor este independent de punctajul echipei la începutul campionatului, el depinde doar de performanțele sale în turneu. Însă el poate fi folosit în campionate viitoare, când va crește punctajul echipei.

## Bibliografie

- [1] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, Proceedings of the Seventh International World Wide Web Conference*, 30(1):107–117, 1998. <https://snap.stanford.edu/class/cs224w-readings/Brin98Anatomy.pdf>.
- [2] Henry E Daniels. Round-robin tournament scores. *Biometrika*, 56(2):295–299, 1969.
- [3] J. W. Moon and N. J. Pullman. On generalized tournament matrices. *SIAM Review*, 12(3):384–399, 1970.
- [4] Shun Motegi and Naoki Masuda. A network-based dynamical ranking system for competitive sports. *Scientific Reports*, 2(1):2045–2322, 2012.
- [5] Filippo Radicchi. Who is the best player ever? a complex network analysis of the history of professional tennis. *PloS one*, 6(2):e17249, 2011. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0017249>.