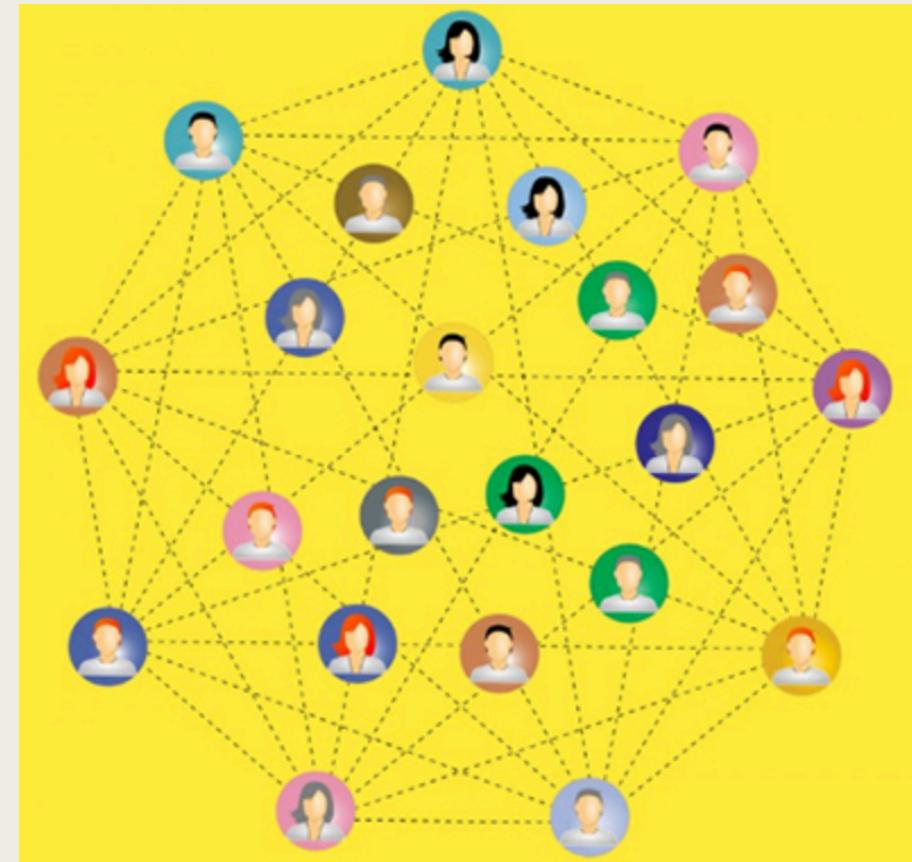


ALGORITMI I STRUKTURE PODATAKA PROJEKAT 2

2022/2023

EdgeRank algoritam

- Razvila ga je kompanija Facebook
- Predstavlja algoritam na osnovu kog se odlučuje koje objave će se prikazati korisniku
- Objavljen 2010. godine
- Graf korisnika i interakcija



Slika preuzeta sa: https://cdn.analyticsvidhya.com/wp-content/uploads/2020/04/featured_image-2.jpg

EdgeRank algoritam

Rank

$$= \text{Affinity} \times \text{Weight} \times \text{Time Decay}$$



Affinity score
between viewing
user and edge
creator.



Weight for this
edge type
(status,
comment, like,
tag, etc.)



Time Decay
factor based on
how long ago
the edge was
created.

Sklonost korisnika (Affinity) 1/4

- Sklonost korisnika predstavlja stepen bliskosti korisnika A sa autorom sadržaja (korisnik B)
- Sklonost se određuje na osnovu prethodnih interakcija
- Vrste interakcija su:
 - *klik (npr. otvaranje linka koji je korisnik objavio),*
 - *Reakcija na objavu*
 - *Komentarisanje objave*
 - *Deljenje objave*



* U projektu uzimamo u obzir samo reakcije, komentarisanje i deljenje objave

Sklonost korisnika (Affinity) 2/4

- Nemaju sve reakcije istu težinu (npr. klik na sadržaj ima mnogo manju težinu nego deljenje sadržaja)
- Na studentu je da odredi težine interakcija

Sklonost korisnika (Affinity) 3/4

- Na sklonost utiče i koliko je vremena prošlo od interakcije
- Ukoliko je korisnik A često reagovao na objave korisnika B pre godinu dana a u poslednjih mesec dana samo jednom, onda objave korisnika B treba da se prikažu korisniku A sa manjom verovatnoćom nego da su učestalost i intezitet reakcija ostali isti.

Sklonost korisnika (Affinity) 4/4

- Nije simetrična relacija
- korisnik A može da reaguje ređe/slabije na objave korisnika B nego obrnuto



Popularnost ili težina objave (Weight)

- Drugi faktor koji se uzima u obzir u EdgeRank algoritmu je popularnost objave
- Objava koja je izazvala puno pažnje, reakcija, komentarisanja i deljenja se smatra interesantnijom pa se povećava šansa da bude prikazana korisnicima
- Različite vrste objava imaju različit značaj
- Objava statusa je manje značajna od objave slike koja je manje značajna od objave videa.**
** U okviru projektnog zadatka ne uzimamo ovo u obzir

Vremenska komponenta raspada (Time Decay) 1/2

- Vreme objavljivanja utiče na verovatnoću prikazivanja objave
- Ukoliko je objava postavljena pre par sati, najveća je šansa da će se prikazati.
- Ukoliko je postavljena pre par dana, verovatnoća prikazivanja je manja.
- Ukoliko je stara mesec dana verovatnoća prikazivanja je minimalna.

Vremenska komponenta raspada (Time Decay) 2/2

The infographic is set against a blue background with a white rounded rectangle containing three main sections: a title area, a text area, and a graph area.

Title Area: On the left, a white rounded rectangle contains the symbol d_e . To its right is a yellow rounded rectangle containing the text: "The Recency Score is based on the time the object was posted". To the right of this text is a small icon of a grandfather clock.

Text Area: Below the title area is a white rounded rectangle containing a cartoon illustration of a boy's head on the left and a speech bubble on the right. The speech bubble contains the text: "As seen in the graph, Newer objects will have a higher recency score than objects that have been posted earlier".

Graph Area: To the right of the text area is a light blue rounded rectangle containing a graph. The vertical axis is labeled "Recency Score" and the horizontal axis is labeled "Time". A red line starts at the top-left corner and slopes downward to the bottom-right corner, representing a linear decay over time.

EdgeRank

- Na studentima je da definišu samostalno koliko svaki od ovih činilaca utiče na celokupan rezultat EdgeRank-a

Pretraga

- Obezbediti pretragu objava
- Podržati unos jedne ili više reči
- Na rangiranje objava utiče pojavljivanje (svake od) reči u samoj objavi kao i EdgeRank rezultat same objave
- Rezultat pretrage je niz objava
- Poželjno bi bilo označiti tražene reči u rezultatima pretrage
- Preporučuje se skladištenje reči u strukturi podataka Trie

Serijalizacija

- Ukoliko se EdgeRank graf i strukture podataka za podršku pretrage teksta učitavaju u trenutku pokretanja programa to može da potraje i nekoliko minuta
- Potrebno je sve potrebne strukture podataka formirati unapred i serijalizovati
- Strukture se pri pokretanju programa samo učitavaju
- Pogledajte modul **pickle**
- <https://docs.python.org/3/library/pickle.html>

Set podataka 1/4

- Objave (pa posledično i komentari, reakcije, deljenja) su podeljene u dve grupe.
- Veća grupa objava koja se koristi za formiranje grafa.
- Drugi set "svežijih" objava za testiranje
- Ovaj set bi trebalo učitati na odbrani i u tom trenutku ažurirati graf

Set podataka 2/4

- **statuses.csv** – sadrži podatke o objavama. Svaka objava sadrži identifikator objave (status_id), tekst objave (status_message), naziv veze (link_name), tip objave (status_type, može biti slika, link itd.), veza (status_link), datum i vreme objavljivanja (status_published), autor (author), broj reakcija (num_reactions), broj komentara (num_comments), broj deljenja (num_shares), broj sviđanja (num_likes), broj reakcija “voli” (num_loves), broj “wow” reakcija (num_wows), broj “haha” reakcija (num_hahas), broj tužnih reakcija (num_sads), broj besnih reakcija (num_angrys), broj posebnih reakcija (num_special).

Set podataka 3/4

- *comments.csv* – sadrži podatke o komentarima na objave. Svaki komentar sadrži identifikator komentara (comment_id), identifikator statusa kome pripada (status_id), identifikator roditeljskog komentara (parent_id, u slučaju da se komentariše komentar), poruka komentara (comment_message), autor (comment_author), datum i vreme objavljivanja komentara (comment_published), broj reakcija (num_reactions), broj sviđanja (num_likes), broj reakcija “voli” (num_loves), broj “wow” reakcija (num_wows), broj “haha” reakcija (num_hahas), broj tužnih reakcija (num_sads), broj besnih reakcija (num_angrys), broj posebnih reakcija (num_special).
- EdgeRank algoritam u našoj implementaciji ne treba da uzme u obzir reakcije na komentare

Set podataka 4/4

- ***friends.csv*** – sadrži podatke o prijateljstvima korisnika. Svaka linija fajla sadrži ime korisnika (person), broj prijatelja (number_of_friends), spisak prijatelja (friends).
- ***reactions.csv*** – sadrži podatke o reakcijama korisnika. Svaka reakcija sadrži identifikator statusa na koji se odnosi (status_id), vrstu reakcije (type_of_reaction, od dostupnih vrsta), korisnika koji je reagovao (reactor), datum i vreme reagovanja (reacted).
- ***shares.csv*** – sadrži podatke o deljenju objava korisnika. Svako deljenje sadrži identifikator statusa na koji se odnosi (status_id), korisnika koji je podelio (sharer), datum i vreme deljenja (shared).