

Analiza socijalnih mreža

Bipartitne mreže

Marko Mišić, Predrag Obradović

13M111ASM

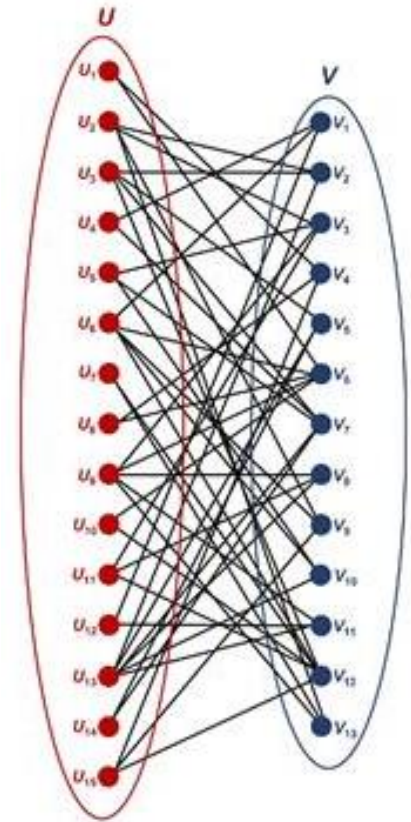
2023/2024.

Motivacija

- Postoji potreba u podacima iz stvarnog sveta za modeliranje odnosa između entiteta različitih tipova
 - Glumci i filmovi
 - Istraživači i publikacije
 - Korisnici i artikli za preporuku
 - Evolucioni odnosi u ekologiji
 - Interakcije proteina i proteina
 - Regulisanje gena
- Bipartitni grafovi (mreže) se često koriste za modeliranje takvih podataka
 - Entiteti su predstavljeni čvorovima
 - Njihove veze su predstavljene granama (linkovima)

Definicija bipartitnog grafa (1)

- Graf $G = (U, V, E)$ je bipartitan (bimodalni) ako:
 - Čvorovi mogu biti podeljeni u dva odvojena skupa U i V
 - Svaka grana iz skupa E povezuje dva čvora koja pripadaju različitim skupovima
 - Čvor iz U može biti povezan samo sa jednim ili više čvorova iz skupa V



Definicija bipartitnog grafa (2)

- Graf koji ne sadrži cikluse neparne dužine je po definiciji bipartitan
 - Može se obojiti korišćenjem dve boje
 - *2-colorable graph*
 - Bipartitivnost je mera koja kvantifikuje koliko je zadata mreža blizu toga da bude bipartitivna
- Veliki broj korisnih osobina i metrika
 - Stepennost čvora ograničena brojem čvorova suprotnog skupa
 - Suma stepenova jednog skupa je jednaka sumi stepenova drugog skupa
 - Svi putevi između dva čvora iz istog skupa su parne dužine

Definicija bipartitnog grafa (3)

- Mere sa uobičajenim značenjem
 - Centralnost po stepenu, bliskosti, svojstvenom vektoru, relaciona centralnost
 - Modularnost za otkrivanje komuna
- Koeficijent klasterizacije se ne može direktno koristiti
 - Problem sa direktnim projekcijama je pojava klika
- Ugrađenost predstavlja meru strukture u ekološkim bipartitnim mrežama
 - Hijerarhijska organizacija čvorova
 - Skup suseda nekog čvora (nižeg stepena) je podskup suseda drugog čvora (višeg stepena)

Primeri bipartitnih mreža (1)

- Ljudi koji posećuju socijalne događaje
 - *Deep South / Southern Women data* iz 1941. godine
- Mreže glumaca i filmova
 - Studije koje su sproveli *Watts-Strogatz* i *Newman*
- Članovi upravnih odbora i kompanije
- Analiza internet saobraćaja i detekcija anomalija
- Interakcije korisnika i grupa u socijalnim i kolaboracionim mrežama
 - Društvene mreže poput *Twitter-a*, *Reddit-a*
 - Forumi i diskusione grupe
 - Mreže autora i radovi, autora i tema

Primeri bipartitnih mreža (2)

- Primeri u biološkim naukama (*life sciences*)
- Ekološke mreže
 - Interakcije živog sveta u ekosistemima
 - Lanci ishrane
 - Food webs, simbiotske mreže, mreže domaćina i parazita
- Biomedicinske mreže
 - Mreže gena i bolesti (*diseasome*)
 - Mreže simptoma i bolesti ljudi (*Human Symptoms Disease Network*)
 - Mreže lekova i ciljnih proteina (*drugome*)

Primeri bipartitnih mreža (3)

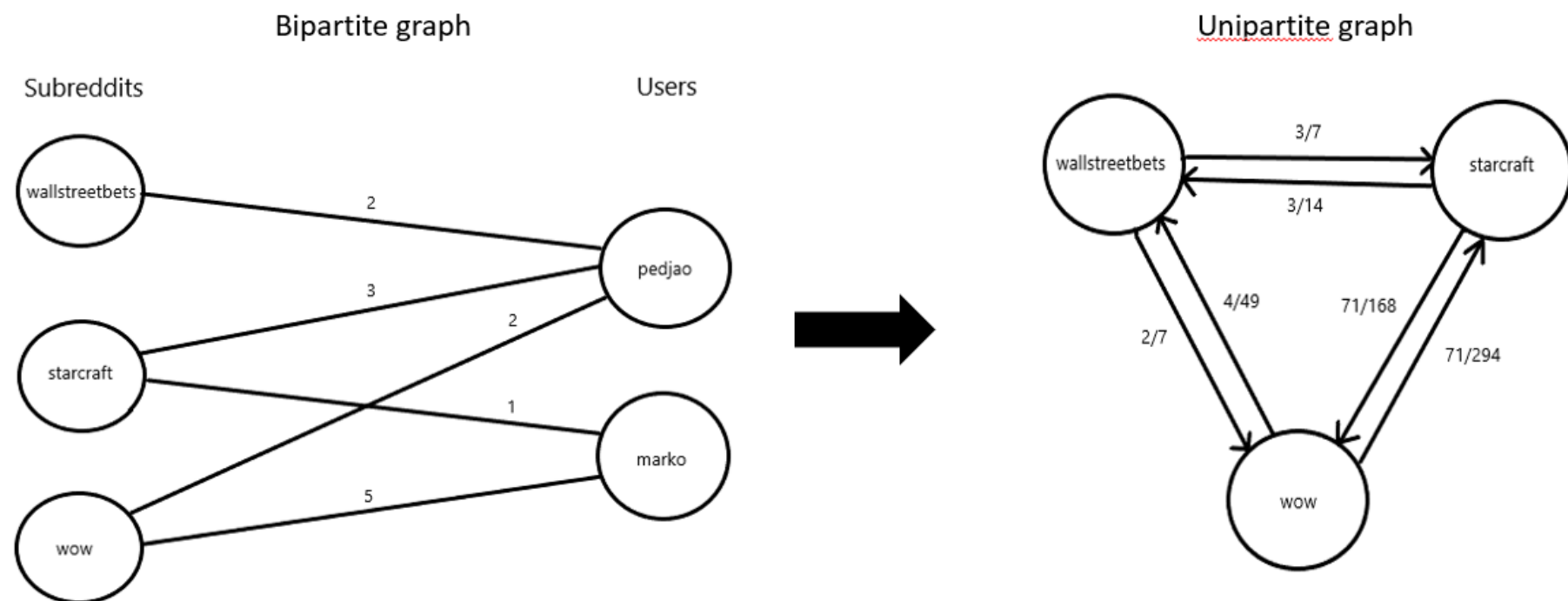
- Primeri u biološkim naukama (*life sciences*)
- Biomolekularne mreže
 - Veze između različitih klasa biomolekula u proteomici
 - Interakcija među proteinima i peptidima
 - Mreže regulacije gena i mreže koekspresije gena
 - Metaboličke mreže (međusobne interakcije metabolita)
- Epidemiološke mreže
 - Mreže seksualnih kontakata za analizu seksualno prenosivih bolesti i prevenciju
 - Analiza bolesti sa vektorom prenosiocem

Projekcija na unipartitivne mreže (1)

- Reprezentacija podataka bipartitnim grafom je nepogodna za mnoge primene
 - Mašinsko učenje
 - Problem veličine mreže
 - Milioni čvorova i grana
 - Problem interpretacije metrika
- Problem se rešava unipartitnom projekcijom mreže na jedan skup čvorova
 - $\mathcal{U}$ -projekcija ili $\mathcal{V}$ -projekcija
 - Potreban odgovarajući model agregacije

Projekcija na unipartitivne mreže (2)

- Korisnici i *subreddit*-i na mreži *Reddit*



Validacija projekcije

- Projekcija treba da zadrži važne veze između izabranog skupa čvorova
 - Jednostavne tehnike projekcije daju grafove sa velikim brojem grana koji nose malo informacija
 - Problem stvaranja klika
- Potrebna je ekstrakcija kičme mreže
 - *Backbone extraction* ili *network sparsification*
 - Izdvajanje najvažnijih relacija koje nose informacije
 - Problem za (težinske) bipartitne grafove

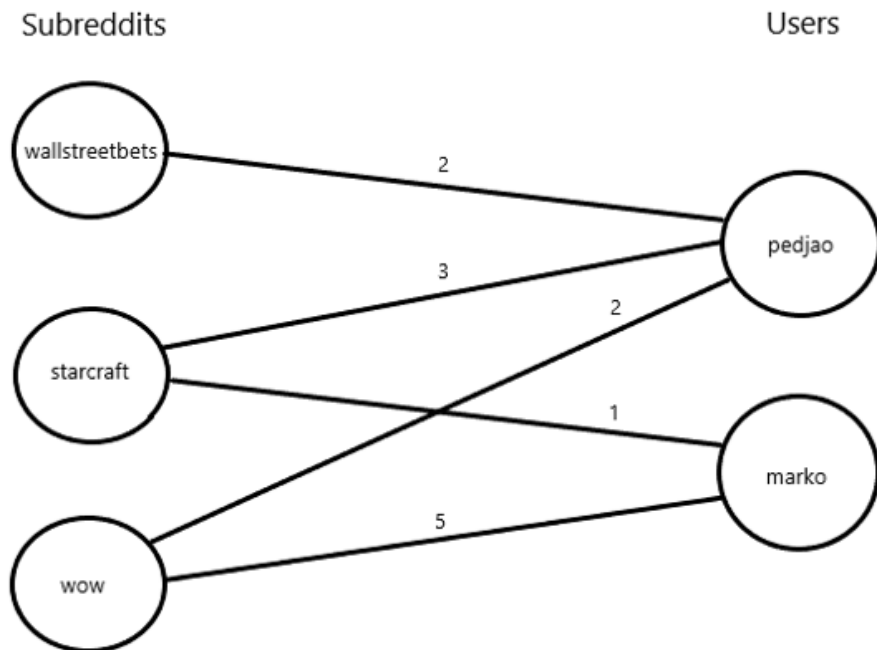
Izbor tehnike za agregaciju (1)

- Tri tipične tehnike agregacije za unipartitnu projekciju bipartitnog grafa:
 - Broj deljenih (preklapajućih) čvorova drugog skupa
 - Suma minimalnih težina grana koje spajaju čvorove prvog skupa preko čvora drugog skupa
 - Dvoprolazna agregacija (*two pass aggregation*)
- Rezultujući graf treba da bude značajno ređi
 - Potrebna dodatna obrada za eliminaciju grana koje nose malo informacija
 - Redundantne grane

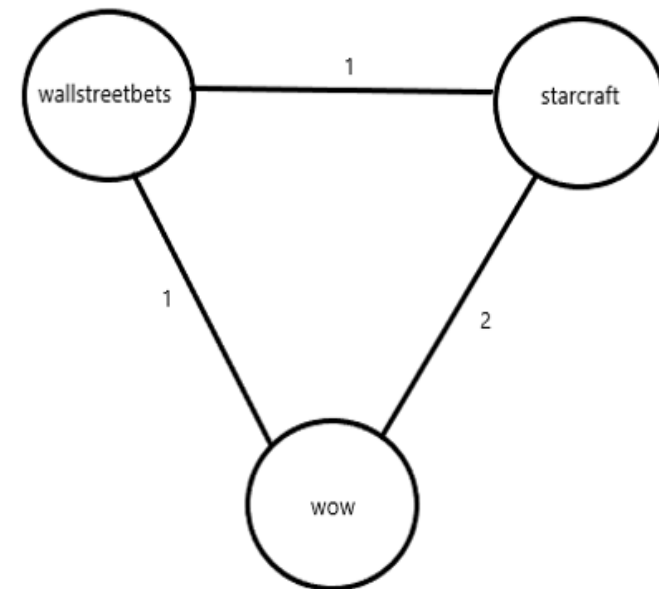
Izbor tehnike za agregaciju (2)

- Broj deljenih (preklapajućih) čvorova drugog skupa

Bipartite graph



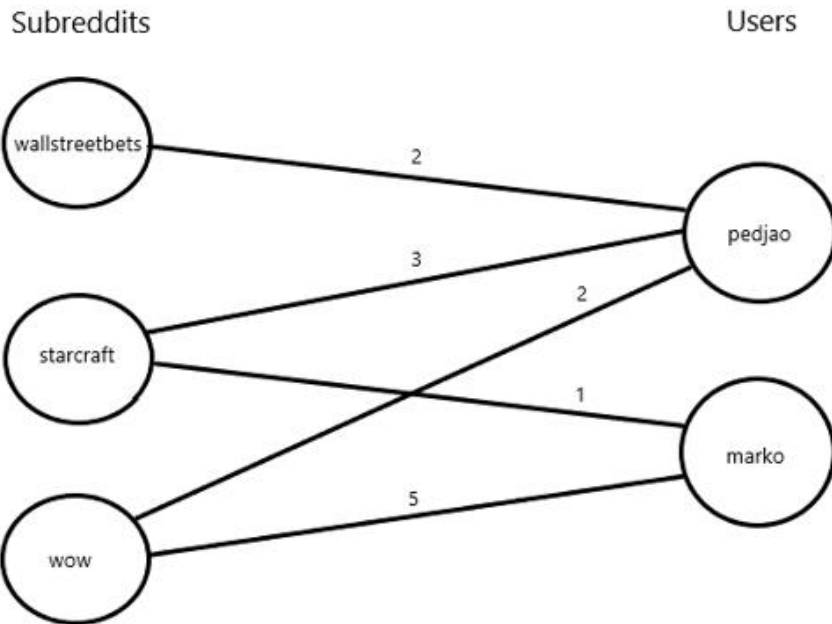
Unipartite graph



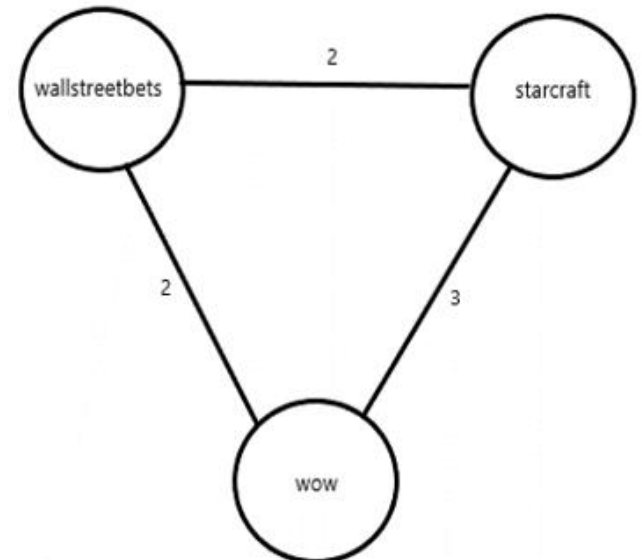
Izbor tehnike za agregaciju (3)

- Suma minimalnih težina grana koje spajaju čvorove prvog skupa preko čvora drugog skupa

Bipartite graph



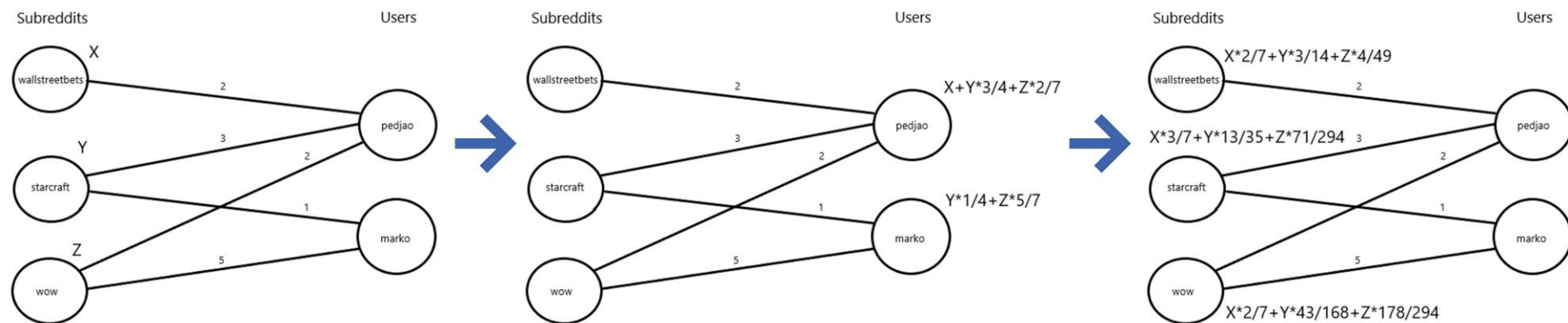
Unipartite graph



Izbor tehnike za agregaciju (4)

○ Dvoprolazna agregacija

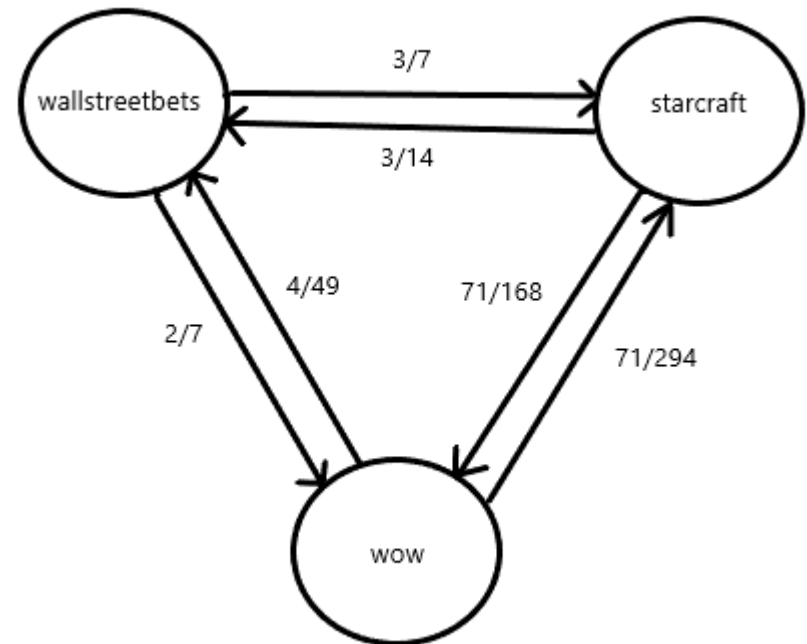
- Svakom čvoru prvog skupa se dodeli jedna jedinica resursa
- Resurs se raspodeli susedima iz drugog skupa proporcionalno težinama grana koje izlaze iz čvora
- Dobijeni resursi se na isti način vrate nazad



Izbor tehnike za agregaciju (5)

○ Dvoprolazna agregacija

- Količina resursa potekla iz jednog čvora i završila u drugom čvoru će biti projektovana težina grane u rezultujućem grafu
- Rezultujući graf je usmeren
- Male težine pojedinih grana



Konfiguracioni modeli (1)

- Konfiguracioni model predstavlja metod generisanja slučajne mreže
 - Na osnovu zadate sekvence stepena čvorova
 - Ne ograničava se na Poasonovu raspodelu kao kod *Erdos-Renyi* mreža
- Jednostavan algoritam
 - Alternativa Havel-Hakimi algoritam
- Referentni modeli za socijalne i kompleksne mreže
 - Prazni (*null*) modeli
 - Na osnovu njih se vrši filtriranje grana

Konfiguracioni modeli (2)

- Jednostavan algoritam:
 - Pridružiti redom svakom čvoru jedan od zadatih stepenova u obliku polu-linkova
 - Ukupan broj polu-linkova mora biti paran
 - Izabrati dva polu-linka i povezati ih, esencijalno formirajući granu
 - Ponavljati postupak dok svi čvorovi ne budu povezani
- Topologija dobijene mreže zavisi od redosleda izbora polu-linkova
- Problem petlji i paralelnih grana

Studija slučaja – Reddit (1)

- *Reddit* je socijalna mreža za agregaciju vesti, ocenu i diskusiju sadržaja na webu
- Postovi su organizovani po temama od strane korisnika u diskusione grupe (*subbredit-e*)
 - 50M+ dnevno aktivnih korisnika
 - 100K+ aktivnih diskusionih grupa
 - 50B+ mesečnih pregleda
- Veoma brz prenos informacija, odmah iza Twitter-a
- Jaka korelacija između događaja iz realnog sveta i sadržaja postovanog i diskutovanog na platformi
 - *Gamestop short squeeze* u januaru 2021. godine



Studija slučaja – Reddit (2)

- Konstruisane su dve bipartitivne mreže na osnovu komentara i postavljenih tema (submisija) korisnika na nekom *subreddit-u*
 - Izvršene su i upoređene tri projekcije
- Izvršena je ekstrakcija kičme metodom WABiCM
 - *Weighted Approximate Bipartite Configuration Model*
- Rezultujuće mreže su značajno manje od polaznih

Studija slučaja – Reddit (3)

○ Rezultati

	Comments graph (user count)	Comments graph (sum minimum)	Comments graph (two-pass)
number of edges	2 189 707	3 195 232	5 393 213
number of connected components	2 162	943	231
average node degree	45	58	85
max node degree	2 528	2 556	3 856
graph density	0.00046	0.00053	0.00067
clustering coefficient	0.292	0.261	0.351

Graph metrics for comments network

	Submissions graph (user count)	Submissions graph (sum minimum)	Submissions graph (two-pass)
number of edges	76 445	1 276 273	1 807 372
number of connected components	6 962	8 703	2 654
average node degree	7	32	29
max node degree	161	2 300	1 613
graph density	0.00028	0.00040	0.00022
clustering coefficient	0.440	0.410	0.458

Graph metrics for submissions network

Studija slučaja – Reddit (4)

- Redukcija grana u opsegu od 84% do 99%
 - Značajan stepen kompresije skupa grana
- Zadržavanje važnih grana iz originalnog modela
 - Potvrđeno dodatnom analizom na problemu mašinskog učenja

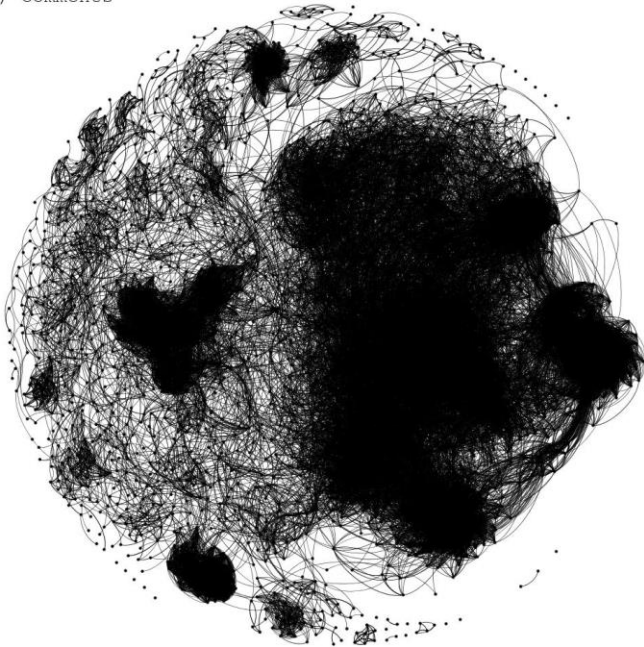
	before backbone extraction	after backbone extraction	Percent reduction
Overlapping neighbor count comments	110M	4.4M	96%
Sum minimum between weights comments	110M	6.4M	94%
Two-pass comments	110M	10M	91%
Overlapping neighbor count submissions	19.5M	150K	99%
Sum minimum between weights submissions	19.5M	2.5M	87%
Two-pass submissions	19.5M	3M	84%

Size of networks before and after backbone extraction

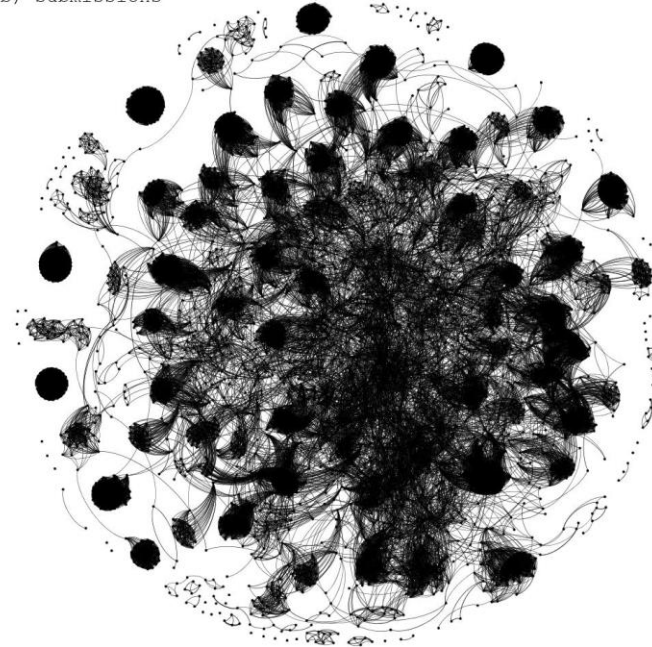
Studija slučaja – Reddit (5)

- Vizuelizacija rezultujućih grafova

a) Comments



b) Submissions



Literatura

- Hanneman, Robert A. and Mark Riddle, Introduction to social network methods, University of California, Riverside, 2005.
- N. Aleksić, E. Pajić, P. Obradović, W. Power, M. Mišić, Z. Obradović, Modelling subreddit interactions by activity overlap, Serbian International Conference on Applied Artificial Intelligence (SICAAI) 2022, best student paper award
- G. A. Pavlopoulos, P. Kontou, A. Pavlopoulou, C. Bouyioukos, E. Markou, P. G. Bagos, Bipartite graphs in systems biology and medicine: a survey of methods and applications, GigaScience, 7, 2018.