

Analiza omrežij *Facebook* in *Twitter*

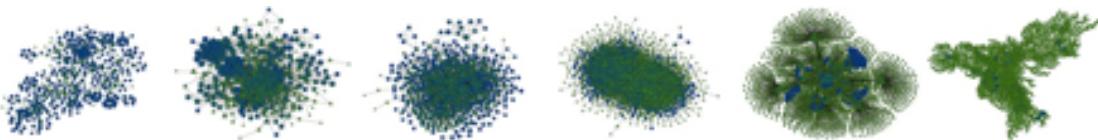
Vaje iz relacijske algebре (OPB)

Lovro Šubelj*

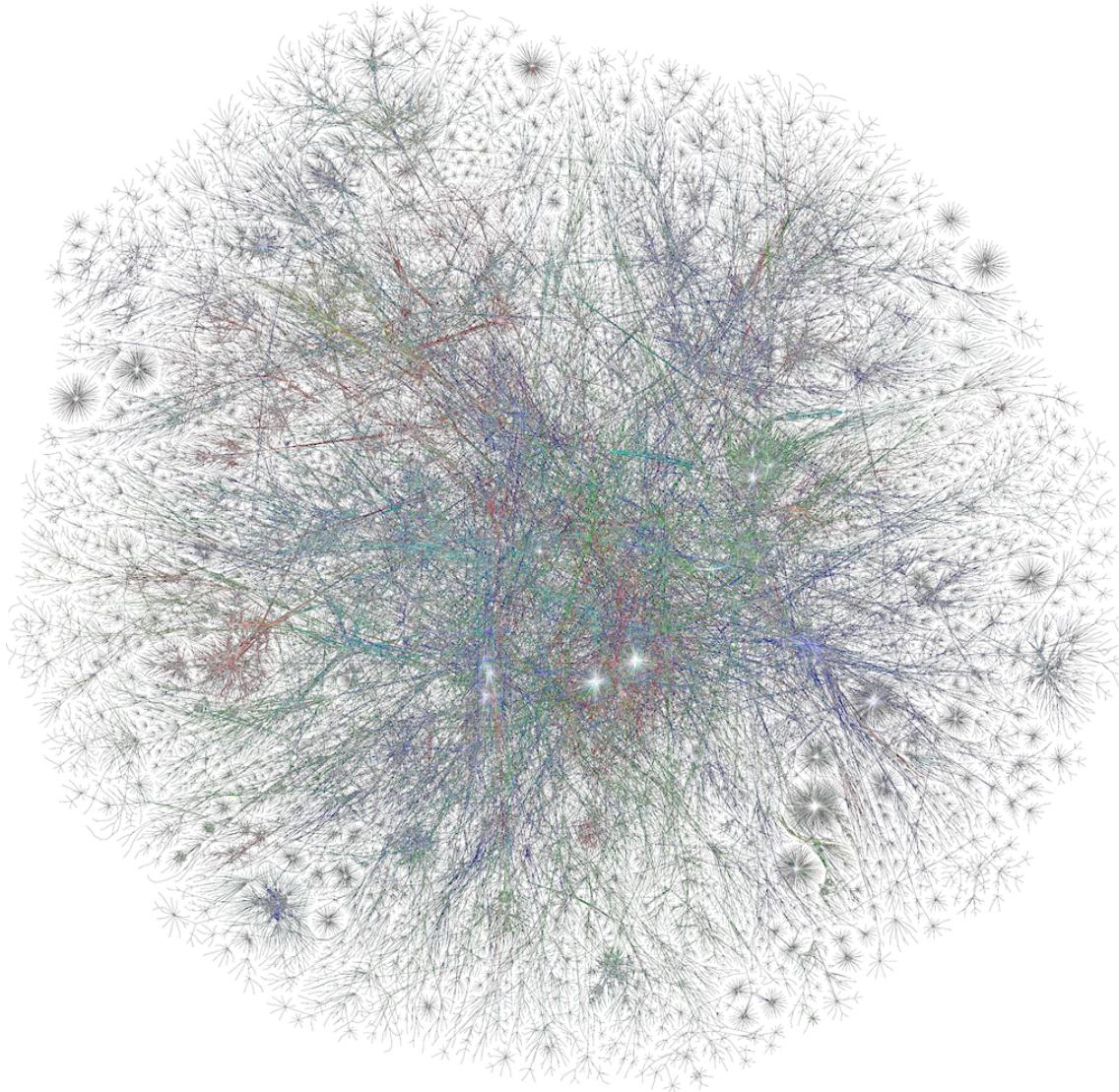
Kazalo

1 Omrežji <i>Facebook</i> in <i>Twitter</i>	3
2 Prijatelji in sledilci	4
2.1 Preimenovanje ρ in pritejanje \leftarrow	4
2.2 Projekcija π in selekcija σ	4
3 Splošne lastnosti omrežij	5
3.1 Unija \cup in presek \cap	5
3.2 Razlika \ in produkt \times	6
3.3 Agregacija τ	6
4 Brezlestvična omrežja	7
4.1 Agregacija z grupiranjem τ	8
4.2 Pogojni (in ekvi) stik \bowtie_θ	9
4.3 Naravni stik \bowtie	9
5 Omrežja malega sveta	10
5.1 Odprt stik \bowtie , \bowtie in \bowtie	11
5.2 Deljenje / in polstik \triangleright	12

*lovro.subelj@fri.uni-lj.si



Slika 1: Primeri različnih realnih omrežij – sodelovanje med znanstveniki, metabolično omrežje, prijateljstva v omrežju *Facebook*, spletni dnevniški, svetovni splet in cestno omrežje.



Slika 2: Internetno omrežje na nivoju avtonomnih sistemov

1 Omrežji Facebook in Twitter

Opis domene

Omrežji *Facebook* in *Twitter* sta predstavljeni s spodnjimi relacijskimi shemami (pripadajoče relacije so p , f in t). Pri tem relacija f hrani prijateljstva v neusmerjenem omrežju *Facebook*, dočim je vsako prijateljstvo shranjeno dvakrat (po potrditvi prijatelja). Na drugi strani pa relacija t hrani informacijo o sledilcih v usmerjenem omrežju *Twitter*, kjer je vsaka povezava shranjena le enkrat.

Relacijske sheme

$$\begin{aligned} \text{PERSON}(&\underline{ID}, \text{User}, \text{Gender}, \text{Birthday}, \text{Date}) \\ \text{FACEBOOK}(&\#ID, \#FID, \text{Date}) \\ \text{TWITTER}(&\#ID, \#FID, \text{Date}) \end{aligned}$$

Primeri relacij

Primeri relacij p , f in t so podani v Tabelah 1, 2 in 3.

Tabela 1: Primer relacije p

ID	$User$	$Gender$	$Birthday$	$Date$
1	ana	F	4.3.1993	1.1.2012
2	bojan	M	13.6.1992	14.6.1992
...

Tabela 2: Primer relacije f (tj. neusmerjeno omrežje *Facebook*)

ID	FID	$Date$
2	3	7.12.2000
3	2	8.12.2000
...

Tabela 3: Primer relacije t (tj. usmerjeno omrežje *Twitter*)

ID	FID	$Date$
2	3	5.5.2002
3	4	18.7.2010
...

2 Prijatelji in sledilci

Opis problema

Predpostavimo, da relacija p hrani vse osebe, ki so se kdajkoli pridružile omrežju *Facebook* ali *Twitter* (ter še nekatere druge osebe). Pri tem datum pridružitve v omrežji hranimo v atributu *Date*. (Podobno velja za atribut *Date* v relacijah f in t .)

Relacijske sheme

PERSON(*ID*, *User*, *Gender*, *Birthday*, *Date*)

FACEBOOK(#*ID*, #*FID*, *Date*)

TWITTER(#*ID*, #*FID*, *Date*)

2.1 Preimenovanje ρ in prirejanje \leftarrow

Naloga — Z uporabo relacijske algebre izpišite vse podatke o osebah v omrežjih *Facebook* in *Twitter*. Pri tem naj bo izpis naj bo naslednje oblike, novo ime relacije pa naj bo *oseba*.

<i>ID</i>	<i>Uporabnik</i>	<i>Spol</i>	<i>Rojstni datum</i>	<i>Datum</i>
1	<i>ana</i>	<i>F</i>	4.3.1993	1.1.2012
...

ali

2.2 Projekcija π in selekcija σ

Naloga — Z uporabo relacijske algebre poiščite uporabniška imena oseb, ki so se pridružile omrežju *Facebook* oziroma *Twitter* v preteklem letu. Izpis naj bo naslednje oblike.

<i>ID</i>	<i>User</i>
1	<i>ana</i>
...	...

3 Splošne lastnosti omrežij

Opis problema

Osebe v omrežjih predstavimo z vozlišči, prijateljstva in sledljivost pa s povezavami.

Relacijske sheme

PERSON(ID, User, Gender, Birthday, Date)

FACEBOOK(#ID, #FID, Date)

TWITTER(#ID, #FID, Date)

3.1 Unija \cup in presek \cap

Naloga — Z uporabo relacijske algebre poiščite vsa vozlišča v omrežju *Facebook* ali *Twitter*. Pri tem predpostavimo, da v danem trenutku v relaciji f ni nepotrjenih prošenj za prijateljstvo. Izpis pa naj bo naslednje oblike.

<i>ID</i>
1
2
...

Naloga — Z uporabo relacijske algebre poiščite vsa vozlišča, ki so hkrati v omrežju *Facebook* in *Twitter*. Pri tem so predpostavke enake kot zgoraj, izpis naj bo naslednje oblike.

<i>ID</i>
1
2
...

Relacijske sheme

$\text{PERSON}(\underline{ID}, \text{User}, \text{Gender}, \text{Birthday}, \text{Date})$

$\text{FACEBOOK}(\#ID, \#FID, \text{Date})$

$\text{TWITTER}(\#ID, \#FID, \text{Date})$

3.2 Razlika \ in produkt \times

Naloga — Z uporabo relacijske algebре poiščite najstarejšo osebo v relaciji p (ozioroma osebe, če obstaja več takih). Izpis naj bo naslednje oblike.

ID	$User$
-1	<i>oldest</i>
...	...

3.3 Agregacija τ

Naloga — Z uporabo relacijske algebре poiščite število vozlišč in povezav v omrežju Facebook. Pri tem so predpostavke enake kot zgoraj, izpis naj bo naslednje oblike.

n	m
$721,1 \cdot 10^6$	$68,7 \cdot 10^9$

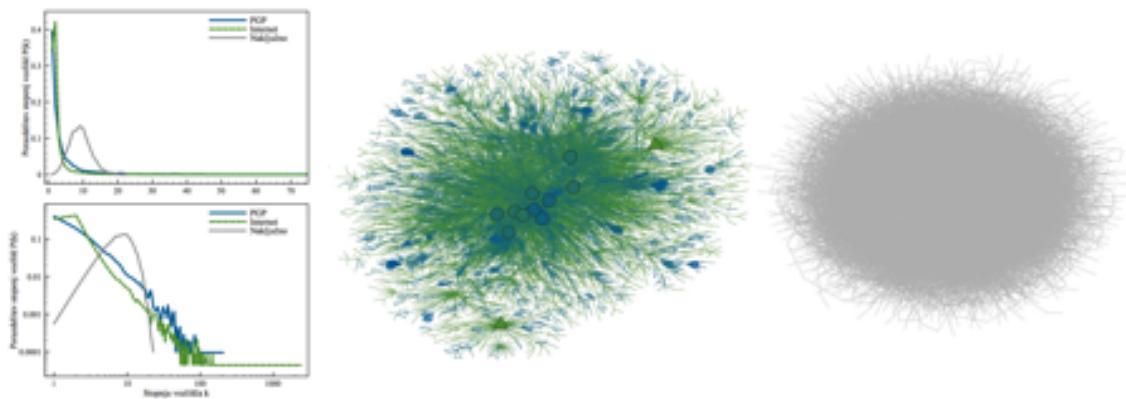
4 Brezlestvična omrežja

Opis problema

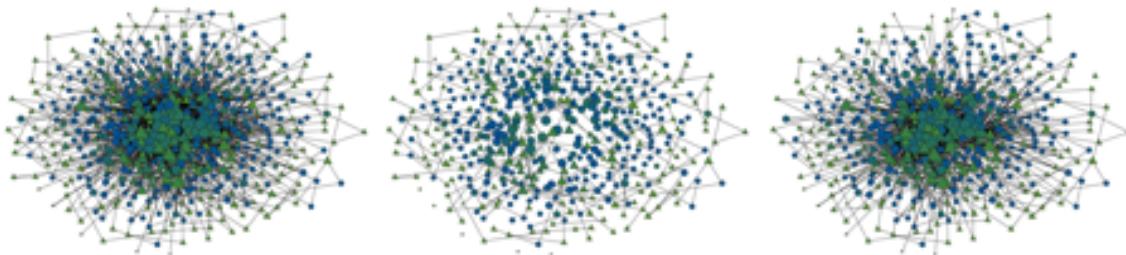
Številu prijateljev neke osebe v omrežju *Facebook* pravimo stopnja vozlišča. Na drugi strani pa v usmerjenem omrežju *Twitter* ločimo med vhodno in izhodno stopnjo, ki ponazarjata število sledilcev ter število drugih oseb, ki jim sledi preučevana oseba.

Porazdelitev stopenj vozlišč v velikih realnih omrežjih (npr. socialnih) značilno drugačna od tiste v primerljivih naključnih omrežjih. Dočim je slednja Poissonova, pa na drugi strani porazdelitev stopenj v realnih omrežjih navadno sledi potenčnemu zakonu (angl. *power-law*). Omrežja z omenjeno lastnostjo imenujemo brezlestvična (angl. *scale-free*).

Dodatno



Slika 3: Brezlestvično in naključno omrežje



Slika 4: Robustnost realnih omrežij

Relacijske sheme

$\text{PERSON}(\underline{ID}, \text{User}, \text{Gender}, \text{Birthday}, \text{Date})$
 $\text{FACEBOOK}(\#ID, \#FID, \text{Date})$
 $\text{TWITTER}(\#ID, \#FID, \text{Date})$

4.1 Agregacija z grupiranjem τ

Naloga — Z uporabo relacijske algebre poiščite stopnje vozlišč v omrežju *Facebook*. Pri tem kot prej predpostavimo, da v f ni nepotrjenih prošenj. Izpis naj bo naslednje oblike.

ID	k
1	2
47	87367
...	...

Naloga — Z uporabo relacijske algebre poiščite “porazdelitev” stopenj vozlišč v omrežju *Facebook*. Pri tem uporabite prejšnjo rešitev, izpis pa naj bo naslednje oblike.

k	$P(K = k)$
1	143892
2	4327897
...	...

Naloga — Z uporabo relacijske algebre poiščite izhodno “porazdelitev” stopenj vozlišč v omrežju *Twitter*. Pri tem si pomagajte z rešitvami prejšnjih nalog.

Naloga — Z uporabo relacijske algebre poiščite vhodno “porazdelitev” stopenj vozlišč v omrežju *Twitter*. Pri tem si pomagajte z rešitvami prejšnjih nalog.

Relacijske sheme

$\text{PERSON}(\underline{ID}, \text{User}, \text{Gender}, \text{Birthday}, \text{Date})$

$\text{FACEBOOK}(\#ID, \#FID, \text{Date})$

$\text{TWITTER}(\#ID, \#FID, \text{Date})$

4.2 Pogojni (in ekvi) stik \bowtie_θ

Naloga — Z uporabo relacijske algebре poiščite kumulativno “porazdelitev” stopenj vozlišč v omrežju Facebook (uporabite rešitev prejšnjih nalog). Izpis naj bo naslednje oblike.

k	$P(K \geq k)$
1	$721,1 \cdot 10^6$
2	$634,9 \cdot 10^6$
...	...

4.3 Naravni stik \bowtie

Naloga — Z uporabo relacijske algebре poiščite osebe v omrežju Facebook, ki ustrezajo vozliščem z največjo stopnjo (uporabite rešitev prejšnjih nalog). Izpis naj bo naslednje oblike.

ID	$User$	k
47	popular	87367
...

5 Omrežja malega sveta

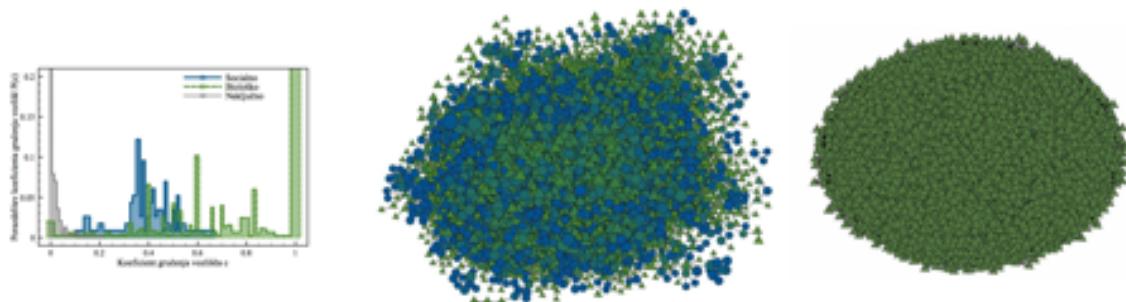
Opis problema

V številnih socialnih in drugih realnih omrežjih je opaziti pojav malega sveta (angl. *small-world*). Sledenje v prvi vrsti pomeni, da so vozlišča v omrežju povezana preko zelo kratkih poti. Na primer, povprečna razdalja med več kot 721 milijoni uporabniki omrežja *Facebook* v letu 2012 je znašala zgolj 4,74. Na drugi strani pa so omrežja malega sveta tudi (lokalno) zelo gosta, kar navadno merimo s koeficientom nakopičenosti (angl. *clustering coefficient*) vozlišča. Slednji je enak

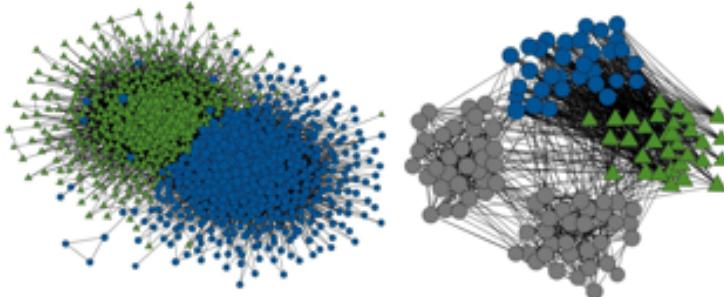
$$c = \frac{\Delta}{\binom{k}{2}},$$

kjer je k stopnja vozlišča, $\binom{k}{2} = k(k - 1)/2$ število vseh parov sosedov vozlišča ter Δ število povezanih parov. Povprečna vrednost koeficiente c je v omrežjih prijateljstev kot je *Facebook* kar enaka verjetnosti, da je prijatelj prijatelja prav tako prijatelj.

Dodatno



Slika 5: Omrežje malega sveta in naključno omrežje



Slika 6: Skupine vozlišč v realnih omrežjih

Relacijske sheme

$\text{PERSON}(\underline{ID}, \text{User}, \text{Gender}, \text{Birthday}, \text{Date})$

$\text{FACEBOOK}(\#ID, \#FID, \text{Date})$

$\text{TWITTER}(\#ID, \#FID, \text{Date})$

5.1 Odprti stiki \bowtie , \bowtie in \times

Naloga — Z uporabo relacijske algebре za vse osebe izračunajte koeficient nakopičenosti v omrežju *Facebook*. Pri tem podobno kot prej predpostavimo, da v danem trenutku v relaciji f ni nepotrjenih prošenj za prijateljstvo. V rešitvi lahko **izjemoma** enkrat uporabite aritmetično deljenje (v okviru priejanja), izpis pa naj bo naslednje oblike. (Namig: najprej za vsako vozlišče poiščite vse različne pare sosedov ter vse povezane pare.)

ID	$User$	c
47	popular	0.21
3	alone	NULL
...

ter

in tako

Naloga — Predpostavimo, da lahko omrežje *Facebook* vsebuje tudi vozlišča, ki ne pripadajo nobeni osebi v relaciji p . (Glede na dane relacijske sheme, slednje sicer ni možno.) Ustrezno popravite rešitev prejšnje naloge tako, da bodo v izpisu vsa vozlišča.

ID	$User$	c
NULL	nobody	0.64
3	alone	NULL
...

Relacijske sheme

PERSON(ID, User, Gender, Birthday, Date)

FACEBOOK(#ID, #FID, Date)

TWITTER(#ID, #FID, Date)

5.2 Deljenje / in polstik ▷

Naloga — Z uporabo relacijske algebре poiščite osebe, ki so v omrežju *Facebook* prijatelji prav z vsemi prijatelji osebe z identifikacijsko številko 1 (tj. ekvivalentna vozlišča). Izpis naj bo naslednje oblike.

<i>ID</i>	<i>User</i>	<i>Gender</i>	<i>Birthday</i>	<i>Date</i>
2	bojan	M	13.6.1992	14.6.1992
...

Naloga — Z uporabo relacijske algebре poiščite vse osebe v omrežjih *Facebook* oz. *Twitter*. Izpis naj bo naslednje oblike.

<i>ID</i>	<i>User</i>	<i>Gender</i>	<i>Birthday</i>	<i>Date</i>
2	bojan	M	13.6.1992	14.6.1992
...