

11. naloga – Izboljšani populacijski modeli

Jože Zobec

1 Uvod

V tem poglavju se bomo spet posvetili odbiranju parametrov s pomočjo metode najmanjših kvadratov in modeliranja populacijskih modelov s krajevno odvisnostjo, ali drugače povedano, grafov.

Algoritem, ki ga bomo uporabili za modeliranje, sta prva uporabila Watts in Strogatz in ni popolnoma realističen, vendar pa je dovolj dober, da ga lahko učinkovito uporabimo modeliranje manjših skupin. Začnemo s s periodičnim neorientiranim grafom z N elementi, ki je povezan z D sosedi (tj. $D/2$ na vsako stran). Tak graf lahko predstavimo z matriko povezav, M in ker je graf neorientiran (povezave tečejo v obe smeri) je matrika simetrična. Da lahko pravilno računamo premer grafa, bomo predpostavili, da je vsak povezan sam s seboj, tj. velja $\text{tr}(M) = N$, kjer smo vsako med i in j v matriki označili z $M_{ij} = M_{ji} = 1$.

Kot je povedal profesor na predavanjih ima tak graf predvsem dva neodvisna parametra – povprečno razdaljo poti in gručavost grafa. A ker je povprečno pot zahtevno računati, jo bomo nadomestili s premerom grafa, to je najdaljša pot v grafu. Premer grafa je tista potenca r matrike M , pri kateri je $M_{ij} > 0$, $\forall i, j$. Za povprečno razdaljo bi morali pri vsaki potenci pogledati koliko elementov matrike je postalo neničelnih – tisto so nove povezave, dolžine, ki je eneka potenci te matrike M .

Gručavost meri koliko sosedov nekega vozlišča ostane sosedov, če tisto vozlišče odmislimo, povprečeno po vseh vozliščih.

V drugem delu se bomo seznanili spet s prilagajanjem krivulj, tokrat so to krivulje, ki so samopodobne. S tujko jih lahko imenujemo tudi renormalizabilne krivulje. Renormalizabilnost je lastnost sistemov, ki so neodvisni od skale na kateri jih opazujemo. Tak primer krivulje je eksponentna funkcija.

2 Rezultati

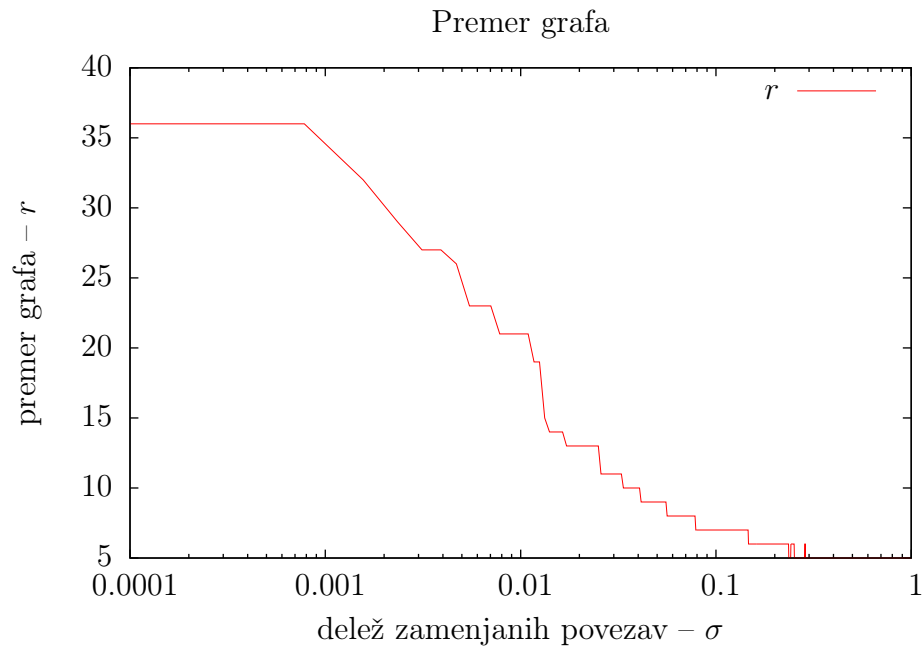
Deloma sem programiranje opravil v `Octave`, deloma v programskem jeziku `C`, da bi dosegel optimalno razmerje med hitrostjo delovanja in nazornostjo programa/skripte.

2.1 Model mali svet

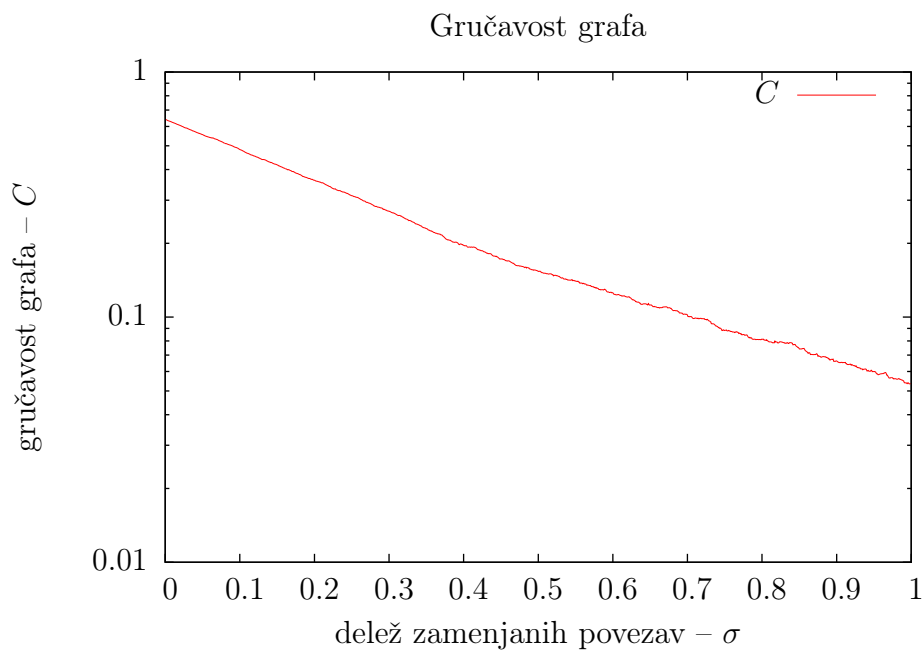
Pred kratkim se je v Kranju vršilo srečanje oktetov, ki je imelo ~ 320 poslušalcev. Mislimo si, da je nekdo sredi nastopa Okteta Gallus iz Ribnice imel hodomušno pripombo. Poglejmo, kako hitro bi ta novica prišla do njihovih ušes.

Pričnemo torej z matriko povezav $N = 320$ in $D = 8$, saj je oktet spremljal tudi njihov umetniški vodja, ki z njimi ni nastopal – to je devet ljudi. Sedaj moramo povezave

stohastično premikati, kar prikazujeta grafa 1 in 2. Število stohastičnih povezav sem računal z oceno, da vsakič izberemo drugo povezavo, ki še ni bila izbrana. Število vseh povezav je $T = ND/2$, torej je $\sigma = \kappa \cdot 2/ND$, kjer je κ število stohastično zamenjanih povezav.

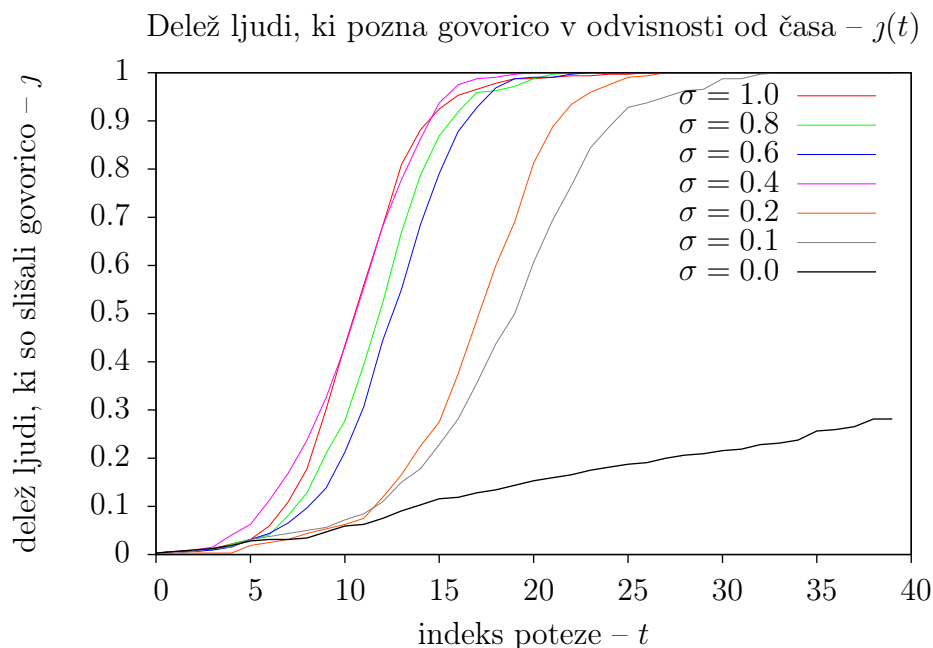


Slika 1: Premer grafa zelo hitro pade – minimum doseže že ko smo zamenjali šele eno desetino povezav.



Slika 2: Gručavost pada dosti počasneje.

V vsakem koraku je verjetnost, da se bosta soseda pogovarjala o tej pikolovski pripombi je $p = 10\%$.



Slika 3: V primeru, ko povezav ne premešamo, dobimo linerano širjenje informacije. Sicer pa že za zelo majhno število stohastično zamenjanih povezav dobimo ti. logistično funkcijo. Vidimo, da se hitrost širjenja govorice praktično ne spreminja, ko dosežemo $\sigma \sim 0.6$.

2.2 Sekularni model človeške populacije

Osnovni predlagani model je

$$N(t) \propto (T - T_0)^\alpha$$

katerega prilagojeni podatki so

$$N(t) \approx 2 \cdot 10^{11} (T - 2025)^{-1}.$$

Ta model torej leta 2025 napoveduje prej še nikoli tako izrazit „baby boom“ zaradi česar pod vprašaj postavlja lastno verodostojnost. Peter Kapitsa, ki je predlagal prejšnji model, je pokazal kako lahko ta model popravimo. Naša nova modelska krivulja je rešitev te diferencialne enačbe:

$$\dot{n} = \frac{dn}{dt} = K \sin^2(n/K) + 1/K,$$

kjer je $n = N/K$ in $t = (T - T_0)/\tau$. Enačba je separabilna, torej lahko odvod jemljemo kot pravi ulomek in na levi strani integriramo vse n , na desni pa vse t . Rešitev te enačbe je

$$n(t) = K \cdot \arctan \left\{ \frac{\tan \left[(t - t_1) \sqrt{1 + K^2/K} \right]}{\sqrt{1 + K^2}} \right\}.$$

Če ta izraz pretvorimo v enote N in T dobimo

$$N(T) = K^2 \arctan \left\{ \frac{\tan [(T - T_0)\sqrt{1 + K^2}/K^2\tau]}{\sqrt{1 + K^2}} \right\}$$

$$\xrightarrow{K \rightarrow \infty} K^2 \arctan \left\{ \frac{\tan [(T - T_0)/K\tau]}{K} \right\}$$

Funkcija ima pri polu tangensa populacijski prehod, ki ga ne opiše pravilno. Tam moramo zato upoštevati periodičnost in prišteti $K^2\pi$.

Vendar pa se je izkazalo, da tak model še vedno ni zvezen! Če hočemo zveznost, se mora ta skok zgoditi natanko enkrat! Potem, ga lahko določimo takrat, ko raste populacija, to je redefiniramo $T_0 \rightarrow T_0 + \tau K\pi/2$, torej se skok zgodi v okolici T_0 :

$$N(T) = K^2 \arctan \left\{ \frac{\tan [(T - T_0)/K\tau + \pi/2]}{K} \right\} + \begin{cases} 0; & T < T_0 \\ \pi K^2; & T \geq T_0 \end{cases} \quad (1)$$

Za prilagajanje sem uporabil metodo Levenberg-Marquardt, ki potrebuje še parcialne odvode po parametrih:

$$\frac{\partial N}{\partial K} = K^2 \left[K^2 + \tan^2 \left(\frac{T - T_0}{K\tau} + \frac{\pi}{2} \right) \right]^{-1} \left[- \frac{(T - T_0)/K\tau}{\cos^2 \left(\frac{T - T_0}{K\tau} + \pi/2 \right)} - \tan \left(\frac{T - T_0}{K\tau} + \pi/2 \right) \right] +$$

$$+ K \arctan \left\{ \frac{\tan [(T - T_0)/K\tau + \pi/2]}{K} \right\} + \begin{cases} 0; & T < T_0 \\ \pi K; & T \geq T_0 \end{cases}$$

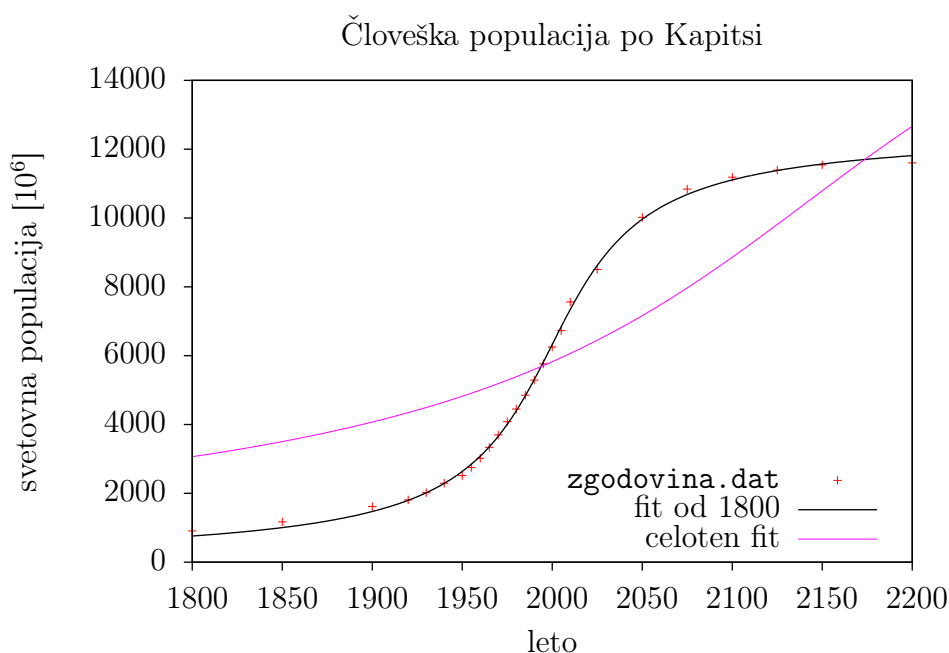
$$\frac{\partial N}{\partial \tau} = K^2 \left[K^2 + \tan^2 \left(\frac{T - T_0}{K\tau} + \frac{\pi}{2} \right) \right]^{-1} \left[- \frac{T - T_0}{\tau^2 \cos^2 \left(\frac{T - T_0}{K\tau} + \pi/2 \right)} \right]$$

$$\frac{\partial N}{\partial T_0} = K^2 \left[K^2 + \tan^2 \left(\frac{T - T_0}{K\tau} + \frac{\pi}{2} \right) \right]^{-1} \left[- \frac{1}{\tau \cos^2 \left(\frac{T - T_0}{K\tau} + \pi/2 \right)} \right]$$

Izkazalo se je, da je bolje odvode računati numerično, zato sem jih računal z diskretnimi diferencami. Uporabil sem paket `optim` iz `Octave Forge`, funkcijo `leasqr`, ki vrne rezultate z vsemi pritiklinami. Za more „dobrote“ prilagajanja ta funkcija uporablja R^2 , ki ga lahko interpretiramo kar kot verjetnost, da naš model ustreza naboru točk. Prilagajal sem na dve različni datoteki. V prvem primeru sem zavrgel vse kar ja bilo pred letom 1800. Drugič sem upošteval vse podatke. Rezultati so v tabeli 1. Graf je prikazan na sliki 4

Tabela 1: Rezultati za prilagaganje testne funkcije na datoteko `zgodovina.dat`. Vidimo, dobimo v obeh primerih sprejemljivo vrednost R^2 , vendar opazimo, da se okleščena datoteka veliko bolje prilega, kar se vidi tudi na tem, da so relativne napake parametrov dosti manjše.

	1800-		zgodovina.dat	
	P	δ_P	P	δ_P
K	63.249	$\pm 0.25\%$	81.518	$\pm 11.58\%$
τ	38.411	$\pm 1.80\%$	169.439	$\pm 11.55\%$
T_0	1999.604	$\pm 0.03\%$	2141.002	$\pm 2.48\%$
	$R^2 = 0.99918$		$R^2 = 0.82215$	



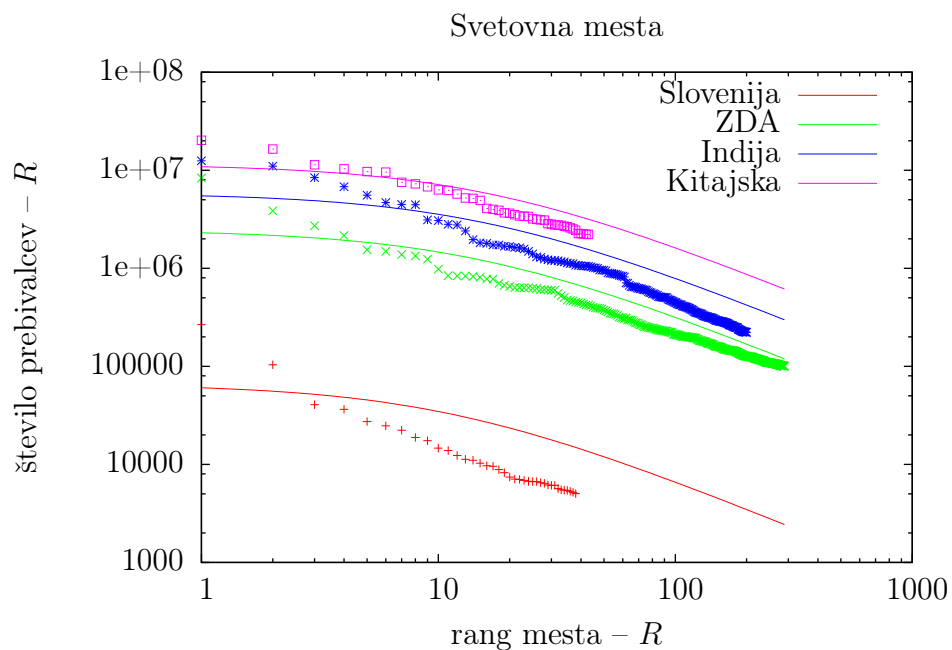
Slika 4: Vidimo, da je bolj upoštevati le podatke po letu 1800. Na ta račun lahko pravilno opišemo grbino, sicer bi s seboj nosili negotove podatke, ki tako ali tako nimajo tiste prave verodostojnosti.

2.3 Svetovna mesta

Za svetovna mesta naj bi veljala relacija

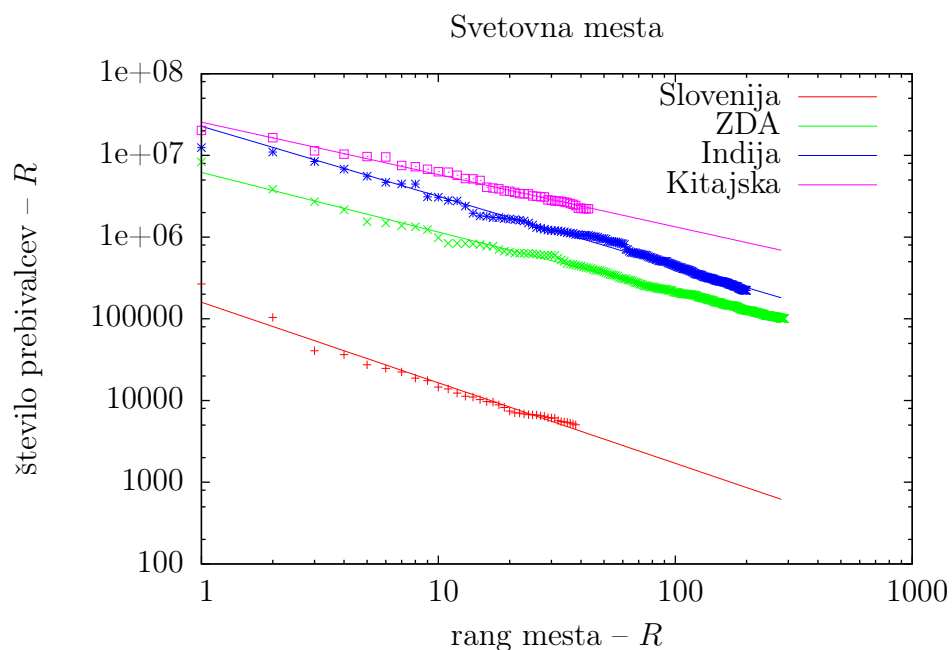
$$U(R) = U_0 \frac{\log U_0}{R + \log U_0},$$

kjer je R rang mesta (katero po velikosti je) in U populacije tega mesta. Veljavnost zakona sem preveril po podatkih za Slovenijo, ZDA, Kitajsko in Indijo. Presenetljivo ta zakon za ta mesta v veliki meri ne velja.



Slika 5: Na začetku je opis dober, vendar potem funkcija pada prepočasi, kot kaže trend iz podatkov.

Ker je slika 5 priča, da taka izbira ni najboljša (tudi tabela 2 tako pravi), bomo poskusili še s preprosto potenčno funkcijo.



Slika 6: Celotno ujemanje je veliko boljše, kot je bilo prej, vendar pa smo pri tem izgubili lepo zaobljeno obliko na začetku, ki pa je tako ali tako relevantna le v mestih Republike Indije.

Za prilagajanje sem v sliki 6 uporabil model

$$U(R) = A \cdot R^\alpha,$$

ki sem ga reševal kar z linearno regresijo. Parametri za posamezen graf so na tabeli 3.

Tabela 2: Parametri krivulj prilagojene podatkom s slike 5. Kot vidimo je $\overline{\chi^2}$ močno prevelik, kar pomeni da se testna krivulja zelo slabo prilega podatkom.

	Slovenija	ZDA	Indija	Kitajska
U_0	$6.6 \cdot 10^4$	$2.5 \cdot 10^6$	$5.8 \cdot 10^6$	$1.2 \cdot 10^7$
δ_{U_0}	$\pm 19.5\%$	$\pm 4.3\%$	$\pm 3.7\%$	$\pm 5.3\%$
$\overline{\chi^2}$	$1.44 \cdot 10^9$	$1.58 \cdot 10^{11}$	$7.03 \cdot 10^{11}$	$4.36 \cdot 10^{12}$

Tabela 3: Parametri prilagojenih funkcij s slike 6 in njihove relativne napake. Vidimo, da je $\chi^2 \sim 1$, kar je bistveno bolje, kot prej.

Slovenija		ZDA		Indija		Kitajska	
A	α	A	α	A	α	A	α
$1.59 \cdot 10^5$	-0.99	$6.18 \cdot 10^6$	-0.73	$2.27 \cdot 10^7$	-0.86	$2.55 \cdot 10^7$	-0.64
$\pm 4.5 \%$	$\pm 19.2 \%$	$\pm 1.9 \%$	$\pm 8.4 \%$	$\pm 1.9 \%$	$\pm 8.7 \%$	$\pm 3.0 \%$	$\pm 27.4 \%$
$\overline{\chi^2} = 0.71$		$\overline{\chi^2} = 0.55$		$\overline{\chi^2} = 1.60$		$\overline{\chi^2} = 0.19$	