

# Spekter grafov III: aplikacije

R. Škrekovski

26. maj 2014

Kompleksna omrežja lahko analiziramo s pomočjo informativnih mer. Mere so kvantitativni opis značilnosti mrež in nam pomagajo modelirati kompleksna omrežja, določiti njihove lastnosti, karakterizacijo in klasifikacijo.

Z merami središčnosti kvantitativno določimo, koliko je posamezno vozlišče pomembno. Poznamo naslednje mere središčnosti:

- središčnost stopnje
- središčnost glede na vmesnost
- središčnost lastnega vektorja
- Katz središčnost
- PageRank
- ...

# Središčnost lastnega vektorja

V omrežjih nas ponavadi ne zanima samo lokalni vpliv vozlišča, temveč tudi kako pomembno je to vozlišče v celem grafu. Oseba, ki pozna le nekaj zelo vplivnih oseb, je velikokrat mnogo pomembnejša od osebe, ki sicer pozna več oseb, a le-te nimajo nobenega vpliva v omrežju. Hočemo torej mero, ki bo upoštevala položaj točke glede na vsa ostala vozlišča in povezave med njimi. Ena izmed takih mer je središčnost lastnega vektorja.

Naj bo  $x$  normiran glavni dominantni vektor matrike sosednosti  $A$ ,  $x$ -u torej pripada največja lastna vrednost. Označimo z  $v_i \in V(G)$  vozlišča grafa. Tedaj vozlišču  $v_i$  pripada  $i$ -ti element vektorja  $x$ . To število je **središčnost lastnega vektorja** za vozlišče  $i$ .

Želimo, da so vrednosti središčnosti pozitivne za vsako vozlišče. Perron-Frobeniusov izrek nam zagotovi ravno to.

**Izrek 1 (Perron-Frobeniusov izrek za grafe)** *Naj bo  $G$  povezan graf in  $A$  matrika sosednosti. Potem*

- 1. je  $\lambda_1$  pozitivna,*
- 2. lastni vrednosti  $\lambda_1$  pripada lastni vektor  $x$ , za katerega so vse koordinate pozitivne, tj.  $x > 0$ ,*
- 3.  $\lambda_1 > |\lambda|$  za katerokoli drugo lastno vrednost  $\lambda \neq \lambda_1$ .*

Toda za velika omrežja je računanje lastnih vektorjev matrik zahteven postopek, v večini primerov pa celo neizvedljiv. Na srečo se iskanje dominantnega lastnega vektorja lahko zelo poenostavi.

Izberimo poljuben vektor  $x$ , ki ni pravokoten na glavni lastni vektor  $v_1$  matrike sosednosti  $A$ . Označimo z  $v_i$ ;  $2 \leq i \leq n$  še ostale lastne vektorje in z  $\lambda_i$ ;  $1 \leq i \leq n$ , pripadajoče lastne vrednosti. Tedaj lahko  $x$  zapišemo kot linearno kombinacijo lastnih vektorjev

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n,$$

kjer  $\alpha_1 \neq 0$ . Če na  $x$   $k$ -krat uporabimo preslikavo  $A$ , dobimo

$$A^k x = \lambda_1^k \alpha_1 v_1 + \lambda_2^k \alpha_2 v_2 + \cdots + \lambda_n^k \alpha_n v_n.$$

Levo in desno stran delimo z  $\lambda_1^k$ , ki je po Perron-Frobeniusovem izreku strogo pozitivna. Dobimo

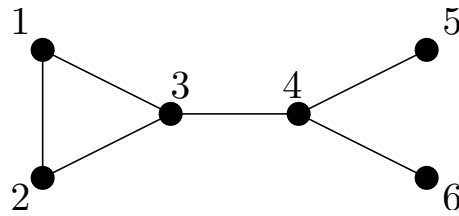
$$\frac{A^k x}{\lambda_1^k} = \alpha_1 v_1 + \frac{\lambda_2^k \alpha_2 v_2}{\lambda_1^k} + \cdots + \frac{\lambda_n^k \alpha_n v_n}{\lambda_1^k}.$$

Ker je  $\lambda_1 > \lambda_j$  pri  $2 \leq j \leq n$ , velja: ko gre  $k$  v neskončnost, se izraz  $\frac{A^k x}{\lambda_1^k}$  približuje  $\alpha_1 v_1$ . Označimo sedaj

$$d_k = \frac{A^k x}{\|A^k x\|}.$$

Po prejšnjem premisleku velja: ko  $k$  raste čez vse meje, se  $d_k$  približuje ravno  $v_1$ . Če torej izberemo poljubno začetno oceno za  $x$ , bo  $d_k$  skonvergirala proti dominantnemu lastnemu vektorju. Tako lahko dokaj hitro dobimo zelo dobro oceno za središčnost lastnega vektorja.

**Primer.** Poglejmo si primer izračuna centralnosti lastnega vektorja za spodnji graf



Njegova matrika sosednosti je

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Lastne vrednosti matrike so

$$\{-1.89122, -1., -0.704624, 0., 1.31743, 2.27841\}$$

med katerimi je največja

$$\lambda_1 \approx 2.27841.$$

S potenčno metodo poiščimo še dominantni lastni vektor, za začetni približek vzamimo  $d_0 = (1, 1, 1, 1, 1, 1)$ . Vmesni rezultati so predstavljeni v spodnji tabeli.

$k$	$d_k$	$ \lambda_1 d_k - A d_k $
1	(0.377964,0.377964,0.566947,0.566947,0.188982,0.188982)	0.415307
5	(0.456538,0.456538,0.563539,0.449404,0.178335,0.178335)	0.146430
10	(0.453986,0.453986,0.591403,0.408878,0.188221,0.188221)	0.057371
20	(0.456587,0.456587,0.58536,0.415745,0.183801,0.183801)	0.008907
30	(0.456925,0.456925,0.584395,0.416903,0.183186,0.183186)	0.001383
50	(0.456985,0.456985,0.584222,0.417111,0.183076,0.183076)	0.000033
51	(0.456987,0.456987,0.584214,0.417121,0.183071,0.183071)	0.000028

# Katz središčnost

Središčnost želimo uporabiti tudi na usmerjenih grafih. V primeru, ko ima vozlišče samo izhodne povezave, dobi središčnost vrednost 0, kar je smiselno. Toda vsako vozlišče, ki ima vhodno povezavo iz vozlišča z mero središčnosti 0, za tako povezavo ne dobi nobenega deleža. Da bomo take primere upoštevali, je Katz vpeljal naslednjo središčnost:

$$x_i = \alpha \sum_{j=1}^n A_{ji} x_j + \beta,$$

kjer sta  $\alpha$ ,  $\beta$  pozitivni konstanti. Tako vsaka točka grafa dobi majhno količino centralnosti  $\beta$ .

V matrični obliki to zapišemo takole

$$x = \alpha A^\top x + \beta \mathbf{1},$$

in od tukaj

$$x = \beta (I - \alpha A^\top)^{-1} \mathbf{1}.$$

Brez izgube za splošnost lahko vzamemo  $\beta = 1$ , ker nas ne zanima absolutna magnituda centralnosti. Tako dobimo

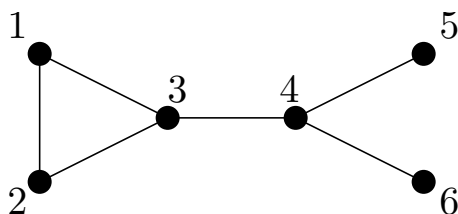
$$x = (I - \alpha A^\top)^{-1} \mathbf{1}.$$

Naslednji korak je določiti  $\alpha > 0$ . Nočemo  $\alpha = 0$ , ker v tem primeru ima vsaka točka enako centralnost  $\beta$ . Za določene  $\alpha > 0$  pa je lahko matrika  $I - \alpha A^\top$  brez inverza. Ker je  $\lambda_1$  največja lastna vrednost, velja  $\lambda_1 \geq \lambda_j$  za vsak  $j = 2, 3, \dots, n$  in zato tudi  $\frac{1}{\lambda_1} \leq \frac{1}{\lambda_j}$ . To pa pomeni, da za vsak  $j$  in izbrani  $\alpha$  velja  $\alpha < \frac{1}{\lambda_j}$  in ni nevarnosti, da bi bila matrika  $I - \alpha A^\top$  singularna.

EksPLICITNA rešitev pride iz enačbe, s katero lahko učinkoviteje in hitreje poiščemo  $x$  iterativno

$$x = \alpha A^\top x + \beta \mathbf{1}.$$

**Primer.** Na istem grafu izračunajmo Katz središčnost z iterativno metodo in  $\alpha = \frac{1}{2\lambda_1}$ .



Vozlišče $i$	$k = 1$	$k = 5$	$k = 10$	$k = 20$	$k = 50$
1	1.4389	1.90171	1.93296	1.93398	1.93398
2	1.4389	1.90171	1.93296	1.93398	1.93398
3	1.65835	2.28001	2.3207	2.32199	2.32199
4	1.65835	2.12658	2.1552	2.15614	2.15614
5	1.21945	1.45964	1.47276	1.47317	1.47317
6	1.21945	1.45964	1.47276	1.47317	1.47317

# PageRank

Google je v želji po čim boljšem vračanju rezultatov svojega iskalnika konstruiral mero središčnosti, ki se po enem od izumiteljev Larry-u Page-u imenuje PageRank. Mera je popolnoma neodvisna od vsebine spletnih strani, upošteva le njihove medsebojne povezave. Vektor PageRank središčnosti za vsako vozlišče, ki predstavlja neko spletno stran v svetovnem spletu, je rešitev enačbe

$$x_i = \alpha \sum_{j=1}^n A_{ji} \frac{x_j}{\deg^+(v_j)} + \beta,$$

kjer sta  $\alpha$ ,  $\beta$  pozitivni konstanti. V matrični obliki to zapišemo takole

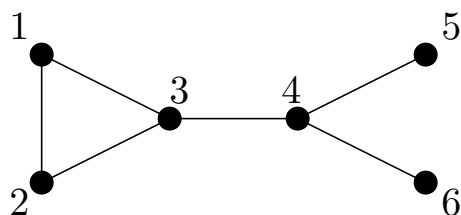
$$x = \alpha A^\top D^{-1} x + \beta \mathbf{1}, \quad (1)$$

kjer je  $D$  diagonalna matrika z  $D_{ii} = \max\{1, \deg^+(v_i)\}$  in od tukaj

$$x = \beta (I - \alpha A^\top D^{-1})^{-1} \mathbf{1} = \beta D (D - \alpha A^\top)^{-1} \mathbf{1}.$$

Veljati mora še  $\alpha < \frac{1}{\lambda_1}$ , kjer je  $\lambda_1$  največja lastna vrednost matrike  $A^\top D^{-1}$ , iz istega razloga kot pri Katz centralnosti – singularnosti. Enačba (1) nam podaja tudi način za iterativni izračun PageRank mere središčnosti.

**Primer.** Spet vzemimo naš testni graf in na njem izračunajmo PageRank središčnost. Uporabili bomo iteracijo po enačbi (1) parameter  $\alpha = \frac{1}{2\lambda_1}$ , kjer je  $\lambda_1$  glavna lastna vrednost, začetni približek pa bo kar vektor enic  $\mathbf{1} = (1, 1, 1, 1, 1, 1)$ . V spodnji tabeli so predstavljeni rezultati po 1., 5., 10. in 20.-ti iteraciji.



Vozlišče $i$	$k = 1$	$k = 5$	$k = 10$	$k = 20$
1	0.312967	0.161387	0.160961	0.160961
2	0.312967	0.161387	0.160961	0.160961
3	0.422693	0.181096	0.18056	0.18056
4	0.642144	0.207883	0.207045	0.207044
5	0.203242	0.145396	0.145237	0.145237
6	0.203242	0.145396	0.145237	0.145237

Največja lastna vrednost matrike  $A^\top D^{-1}$  je 1 oz. blizu 1, kar implicira  $\alpha_1 \in (0, 1)$ . V praksi Google uporablja  $\alpha = 0.85$ .



# Izrek o prepletanju

**Izrek 2** Lastne vrednosti  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  hermitske matrike reda  $n$   $H$  (tj.  $H = \overline{H}^T$ ) se prepletajo z lastnimi vrednostimi  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_{n-1}$  poljubne glavne podmatrike  $N$  reda  $n - 1$ , tj. velja

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \dots \geq \mu_{n-1} \geq \lambda_n.$$

**Posledica.** Naj bo  $G$  graf na  $n$  točkah in  $H$  induciran podgraf grafa  $G$  na  $n - 1$  točkah. Potem se jim lastne vrednosti prepletajo.

**Posledica.** Naj bo  $G$  graf na  $n$  točkah in  $H$  induciran podgraf grafa  $G$  na  $p (< n)$  točkah. Potem velja

$$\lambda_1(G) \geq \lambda_1(H) \quad \text{in} \quad \lambda_n(G) \leq \lambda_p(H).$$

**Trditev.** Za vsak graf  $G$  velja

$$\chi(G) \leq 1 + \lambda_1(G).$$

**Dokaz.** Če je  $\chi(G) = 1$ , potem je  $G$  prazen graf in zato  $\lambda_1 = 0$ . V tem primeru imamo enačaj. Zato predpostavimo, da je  $\chi(G) = k \geq 2$ . Iz kromatične teorije grafov je dobro znano, da  $G$  vsebuje  $k$ -kritičen podgraf  $H$ , tj.  $H$  je  $k$ -kromatičen in vsak njegov podgraf je  $(k - 1)$ -obarvljiv. Vsak  $k$ -kritičen graf pa ima minimalno stopnjo  $\geq k - 1$ . Torej iz

$$\lambda_1(G) \geq \lambda_1(H) \geq \delta(H) \geq k - 1,$$

takoj dobimo željeno zvezo. ■

# Laplaceove lastne vrednosti in število vpetih dreves

**Izrek 3** Naj bo  $G$  graf z Laplaceovo matriko  $L$ . Potem velja:

1. število vpetih dreves v grafu  $G$  je enako  $\det(L_i)$ , pri tem je  $L_i$  podmatrika matrike  $L$ , kjer izbrišemo  $i$ -to vrstico in  $i$ -ti stolpec.
2. število vpetih dreves je enako

$$\frac{1}{n} \prod_{i \geq 2} \mu_i(L).$$

**Izrek 4 (Cayley)** Število vpetih dreves polnega grafa  $K_n$  je  $\tau(K_n) = n^{n-2}$ .

**Zgled.**

# Laplaceove lastne vrednosti in Wienerjev indeks

Poznamo še en, čisto drugačen način izračuna Wienerjevega indeksa dreves, in sicer s pomočjo lastnih vrednosti Laplaceove matrike.

**Izrek 5** *Naj bo  $T$  drevo na  $n$  vozliščih. Če so  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n = 0$  Laplaceove lastne vrednosti za  $T$ , velja*

$$W(T) = n \sum_{i=1}^{n-1} \frac{1}{\mu_i}.$$

Zgornji izrek je nepričakovan rezultat, saj povezuje lastne vrednosti in Wienerjev indeks, ki je število, dobljeno na podlagi razdalj v grafu.

Spomnimo se: če za poljuben graf  $G$  lastne vrednosti Laplaceove matrike  $L$  razvrstimo  $\mu_1 \geq \dots \geq \mu_n$ , potem velja  $\mu_n = 0$ . Zgornji izrek takoj implicira, da je pri drevesih  $\mu_{n-1} > 0$ . Velja pa še več: za vse povezane grafe je  $\mu_{n-1} > 0$ .

**Zgled.**

# Energija grafa

**Energija** grafa  $G$ ,  $\mathcal{E}(G)$ , je vsota absolutnih vrednosti njegovih lastnih vrednosti:

$$\mathcal{E}(G) = \sum_{i=1}^n |\lambda_i|.$$

Energija grafa neke molekule ogljikovodika je v tesni povezavi z energijo, ki je potrebna za njegovo sintezo.

Naj bodo  $\lambda_1, \lambda_2, \dots, \lambda_n$  lastne vrednosti grafa  $G$ . Ker je  $M_1(G) = 0$  (glej posledico), od tod sledi:

$$\sum \{\lambda_i \mid \lambda_i > 0\} = - \sum \{\lambda_i \mid \lambda_i < 0\} = \frac{\mathcal{E}(G)}{2}.$$

Coulson je leta 1940 izpeljal naslednjo formulo za izračun energije grafa:

**Izrek 6** *Naj bo  $G$  graf. Energija grafa  $G$  je*

$$\mathcal{E}(G) = \frac{1}{\pi} \int_{-\infty}^{\infty} \left( n - \frac{ix\phi'(ix)}{\phi(ix)} \right) dx.$$

V zgornji formuli, ki jo imenujemo tudi Coulsonov integral, je  $\phi$  karakteristični polinom matrike sosednosti,  $\phi'$  pa njegov odvod.

Če je  $G$  acikličen graf (torej gozd oz. drevo, če je povezan), potem je karakteristični polinom oblike:

$$\phi(x) = x^n + \sum_{k \geq 1} (-1)^k m_k(G) x^{n-2k}, \quad (2)$$

kjer smo z  $m_k(G)$  označili število prirejanj velikosti  $k$  v grafu  $G$ . Če (2) vstavimo v Coulsonov intergal, dobimo naslednjo posledico:

**Posledica 7** *Naj bo graf  $G$  gozd. Potem velja:*

$$\mathcal{E}(G) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{x^2} \log \left( 1 + \sum_{k \geq 1} m_k(G) x^{2k} \right) dx .$$

Opazimo, da je energija gozda strogo naraščajoča funkcija parametrov  $m_k(G)$ . Ker ima med vsemi drevesi zvezda najmanj prirejanj, pot pa največ prirejanj (vseh velikosti), takoj dobimo naslednjo posledico:

**Posledica 8** *Naj bo  $S_n$  zvezda na  $n$  vozliščih (tj.  $K_{1,n-1}$ ) in  $P_n$  pot na  $n$  vozliščih. Nadalje naj bo  $T_n$  drevo z  $n$  vozlišči, ki ni niti zvezda, niti pot. Potem velja:*

$$\mathcal{E}(S_n) < \mathcal{E}(T_n) < \mathcal{E}(P_n) .$$

# Energija fulerenov in nanocevk

Grafi molekul fulerenov in nanocevk so 3-regularni in so brez trikotnikov in štirikotnikov. Velja naslednji izrek:

**Izrek 9** *Za energijo grafa molekule fulerena oz. nanocevke  $G$  z  $n$  ogljikovimi atomi velja ocena:*

$$\frac{3}{\sqrt{5}}n \leq \mathcal{E}(G) \leq \sqrt{3}n .$$

Za lažjo predstavo koeficienta v zgornji neenačbi zapišimo na dve decimalki natančno:

$$1,34 n \leq \mathcal{E}(G) \leq 1,73 n .$$

Zelo razveseljuje dejstvo, da je ocena na obeh straneh (zgornja in spodnja) linearna v številu  $n$  (število ogljikovih atomov). Energija fulerena leži v razmeroma ozkem intervalu in je bolj kot od same zgradbe molekule odvisna od števila ogljikovih atomov.