Polkolobarji in analiza omrežij

Vladimir Batagelj

IMFM Ljubljana in UP IAM Koper

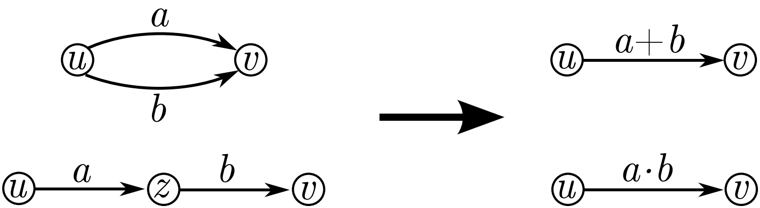
Srečanje raziskovalcev s področja analize omrežij
Ljubljana, 25. januar 2018

Kazalo

e-mail: vladimir.batagelj@fmf.uni-lj.si
wiki: polkolobarji.pdf
version: 26. januar 2018 ob 11 : 32

Polkolobarji in omrežja

Imamo uteženo omrežje $N = (\mathcal{V}, \mathcal{L}, w)$, $w : \mathcal{L} \to A$. Vrednost uteži bi radi posplošili na sprehode in množice (neurejene sezname) sprehodov.



Seštevanje mora biti komutativno in asociativno; množenje pa asociativno.

Polkolobar

$(A, +, \cdot, 0, 1)$ je polkolobar ntk. velja:

- $(A, +, 0)$ – Abelov monoid:
  $a + b = b + a, \quad a + (b + c) = (a + b) + c, \quad a + 0 = a;$
- $(A, \cdot, 1)$ – monoid:
  $a \cdot (b \cdot c) = (a \cdot b) \cdot c, \quad a \cdot 1 = 1 \cdot a = a$

in množenje $\cdot$ distribuira čez seštevanje $+$:

$$a \cdot (b + c) = a \cdot b + a \cdot c \qquad \text{in} \qquad (a + b) \cdot c = a \cdot c + b \cdot c$$

Polkolobar je zaprt, če v njem obstaja še enomestna operacija zaprtja (ovojnice) $a^\star$, za katero velja

$$a^\star = 1 + a \cdot a^\star = 1 + a^\star \cdot a$$

Polkolobar je idempotenten, če velja $a + a = a$.

Sprehodi po omrežju in polkolobarji

.

Utež w lahko razširimo na sprehode in končne množice sprehodov po grafu s predpisi:

- naj bo Z_v ničelni sprehod v vozlišču v, potem je: $w(Z_v) = 1$
- naj bo $S = v_0 v_1 v_2 \ldots v_k$ sprehod po grafu $\mathcal{G}$, potem je:

$$w(S) = \prod_{i=1}^{k} w((v_{i-1}, v_i))$$

- za prazno množico sprehodov $\emptyset$ velja $w(\emptyset) = 0$
- naj bo $\mathcal{S}$ končna množica sprehodov, potem je

$$w(\mathcal{S}) = \sum_{S \in \mathcal{S}} w(S)$$

Lastnosti vrednosti sprehodov

Naj bosta S_1 in S_2 sprehoda, tako da je konec sprehoda S_1 enak začetku sprehoda S_2. Tedaj lahko sprehoda staknemo v nov sprehod, ki ga označimo $S_1 \circ S_2$. Zanj velja

$$w(S_1 \circ S_2) = w(S_1) \cdot w(S_2)$$

Za končni množici sprehodov $\mathcal{S}_1$ in $\mathcal{S}_2$ pa velja

$$w(\mathcal{S}_1 \cup \mathcal{S}_2) + w(\mathcal{S}_1 \cap \mathcal{S}_2) = w(\mathcal{S}_1) + w(\mathcal{S}_2)$$

in v posebnem primeru, ko je $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$: $w(\mathcal{S}_1 \cup \mathcal{S}_2) = w(\mathcal{S}_1) + w(\mathcal{S}_2)$.

Polkolobar je poln ntk. vsota definirana tudi za neskončne množice in veljajo posplošena komutativnost za seštevanje, posplošena asociativnost in posplošena distributivnost. Poln kolobar je vselej zaprt za

$$a^\star = \sum_{i \in \mathbb{N}} a^i$$

Primeri polkolobarjev

Zgled: Kombinatorični polkolobar
$(\mathbb{N}, +, \cdot, 0, 1)$, $+$ seštevanje, $\cdot$ množenje; $w((u,v)) \in \mathbb{N}$ je število načinov, na katere lahko iz vozlišča u pridemo po povezavi v vozlišče v. $\square$

Zgled: Polkolobar regularnih izrazov
$(\mathcal{P}(\Sigma^\star), \cup, \cdot, \emptyset, \varepsilon)$, $\Sigma^\star$ – množica končnih nizov nad abecedo Σ, $\cdot$ stikanje nizov, $\varepsilon = \{\lambda\}$; $w((u,v)) \in \Sigma$ oznaka povezave. $w(S)$ – (beseda) zaporedje oznak povezav na sprehodu S. $\square$

Zgled: Polkolobar najkrajših poti
$(\mathbb{R}_0^+, \min, +, \infty, 0)$, $w((u,v)) \in \mathbb{R}_0^+$ dolžina povezave, $w(S)$ (prava) dolžina sprehoda S. $\square$

Zgled: Povezanostni polkolobar
$(\{0,1\}, \vee, \wedge, 0, 1)$, $w((u,v)) = 1 \Leftrightarrow (u,v) \in R$, $w(S) = 1$. $\square$

Zgled: Verjetnostni polkolobar
$([0,1], +, \cdot, 0, 1)$, $+$ seštevanje, $\cdot$ množenje; $w((u,v)) \in [0,1]$ je verjetnost prehoda iz vozlišča u po povezavi v vozlišče v. $\square$

Množice sprehodov

- $\mathcal{S}^{\star}_{uv}$ – množica vseh sprehodov iz vozlišča u v vozlišče v;
- $\mathcal{S}^{k}_{uv}$ – množica vseh sprehodov dolžine k iz vozlišča u v vozlišče v;
- $\mathcal{S}^{(k)}_{uv}$ – množica vseh sprehodov dolžine največ k iz vozlišča u v vozlišče v;
- $\mathcal{E}_{uv}$ – množica vseh enostavnih sprehodov iz vozlišča u v vozlišče v.

a. $\mathcal{S}^{k}_{uv} \subseteq \mathcal{S}^{(k)}_{uv} \subseteq \mathcal{S}^{\star}_{uv}$

b. $\mathcal{S}^{(k)}_{uv} = \bigcup_{i=0}^{k} \mathcal{S}^{i}_{uv} \qquad \mathcal{S}^{\star}_{uv} = \bigcup_{k=0}^{\infty} \mathcal{S}^{k}_{uv}$

c. $k \neq h \iff \mathcal{S}^{k}_{uv} \cap \mathcal{S}^{h}_{uv} = \emptyset$

d. $k \geq |\mathcal{V}| - 1 : \mathcal{E}_{uv} \subseteq \mathcal{S}^{(k)}_{uv}$

e. $w(\mathcal{S}^{(k)}_{uv}) = \sum_{i=0}^{k} w(\mathcal{S}^{i}_{uv})$

Enolična razcepnost

Pri razširitvi vrednosti na sprehode in množice sprehodov nismo uporabili distributivnosti. Kaj nam prinese?
Posplošimo najprej stikanje na množice sprehodov: $\mathcal{S} \circ \emptyset = \emptyset \circ \mathcal{S} = \emptyset$ in

$$\mathcal{S}_1 \circ \mathcal{S}_2 = \{S_1 \circ S_2 : S_1 \in \mathcal{S}_1, S_2 \in \mathcal{S}_2\}$$

Za množico sprehodov $\mathcal{S}$ bomo rekli, da je enolično razcepna za množici sprehodov $\mathcal{S}_1$ in $\mathcal{S}_2$, če je $\mathcal{S} = \mathcal{S}_1 \circ \mathcal{S}_2$ in ne obstajajo sprehodi $S_1, S_1' \in \mathcal{S}_1, S_1 \neq S_1'$ in $S_2, S_2' \in \mathcal{S}_2$, tako da bi veljalo $S_1 \circ S_2 = S_1' \circ S_2'$. Z drugimi besedami: vsak sprehod iz $\mathcal{S}$ je mogoče na en sam način zapisati kot stik sprehoda iz $\mathcal{S}_1$ in sprehoda iz $\mathcal{S}_2$.

IZREK

Naj bo množica $\mathcal{S}$ enolično razcepna za $\mathcal{S}_1$ in $\mathcal{S}_2$ in končna ali pa polkolobar poln. Potem velja

$$w(\mathcal{S}_1 \circ \mathcal{S}_2) = w(\mathcal{S}_1) \cdot w(\mathcal{S}_2)$$

Prehodna matrika

Uredimo množico vozlišč $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$, potem lahko preslikavo $w : \mathcal{L} \to A$ predstavimo z matriko, pravimo ji prehodna matrika $\mathbf{W} = [w_{ij}]$, katere elementi so določeni z predpisom

$$w_{ij} = \begin{cases} w((v_i, v_j)) & (v_i, v_j) \in \mathcal{L} \\ 0 & \text{sicer} \end{cases}$$

Množico vseh kvadratnih matrik reda n nad A označimo z $\mathcal{M}_n(A)$. Če v polkolobarju $(A, +, \cdot, 0, 1)$ zahtevamo, da velja še

$$\forall a \in A : a \cdot 0 = 0 \cdot a = 0$$

je tudi struktura $(\mathcal{M}_n(A), +, \cdot, \mathbf{0}, \mathbf{1})$ polkolobar, pri čemer sta operaciji nad matrikami definirani na običajni način:

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}] \qquad \text{in} \qquad \mathbf{A} \cdot \mathbf{B} = [\sum_{k=1}^{n} a_{ik} \cdot b_{kj}]$$

Potence prehodne matrike

Ničelna matrika $\mathbf{0}$ ima vse elemente enake 0, enotna matrika $\mathbf{1}$ pa le diagonalne enake 1. če je osnovni polkolobar poln, je poln tudi matrični polkolobar. Zaradi asociativnosti je enolično določena k-ta potenca poljubne kvadratne matrike nad A. Za k-to potenco prehodne matrike $\mathbf{W}$ velja:

IZREK

Element w_{ij}^k k-te potence $\mathbf{W}^k$ prehodne matrike $\mathbf{W}$ je enak vrednosti vseh sprehodov dolžine k iz vozlišča v_i v vozlišče v_j. $w_{ij}^k = w(\mathcal{S}_{ij}^k)$

Pozor, w_{ij}^k **ni** k-ta potenca w_{ij} !!!

Če je omrežje aciklično, velja $\exists k_0 < n : \forall k \geq k_0 : \mathbf{W}^k = \mathbf{0}$. Pri tem je k_0 dolžina najdaljšega sprehoda po grafu.

$$\mathbf{A}^{(k)} = \sum_{i=0}^{k} \mathbf{A}^i$$

Velja $w_{ij}^{(k)} = w(\mathcal{S}_{ij}^{(k)})$. V idempotentnih polkolobarjih je $\mathbf{A}^{(k)} = (\mathbf{1} + \mathbf{A})^k$.

Zaprtje (ovojnica) prehodne matrike

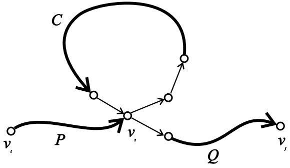
Kaj pa, če je množica sprehodov neskončna, kot je včasih $\mathcal{S}^{\star}_{ij}$? Tedaj stvari gladko tečejo, če je polkolobar poln.
Poglejmo si primer, ko v polkolobarju velja absorpcijski zakon:

$$\forall a, b, c \in A : a \cdot b + a \cdot c \cdot b = a \cdot b$$

Zaradi distributivnosti zadošča že zahteva:

$$\forall c \in A : 1 + c = 1$$

Neenostaven sprehod



Recimo, da v $\mathcal{S}_{ij}^{(k)}$ obstaja neenostavni sprehod S. Ker ni enostaven, se vsaj eno vozlišče, naj bo to v_t, pojavi večkrat. Del sprehoda med prvo in zadnjo pojavitvijo vozlišča v_t v S je obhod. Označimo ga s C. Tedaj lahko S zapišemo v obliki $S = P \circ C \circ Q$. Toda tudi $P \circ Q$ je sprehod. Poiščimo skupno vrednost obeh sprehodov:

$$w(\{P\circ Q, P\circ C\circ Q\}) = w(P\circ Q) + w(P\circ C\circ Q) =$$
$$= w(P)\cdot w(Q) + w(P)\cdot w(C)\cdot w(Q) =$$
$$= w(P)\cdot w(Q) = w(P\circ Q)$$

Torej je za vse dovolj velike k

$$w(\mathcal{S}_{ij}^{(k)}) = w(\mathcal{E}_{ij}) \qquad \text{in zato tudi} \qquad w(\mathcal{S}_{ij}^{\star}) = w(\mathcal{E}_{ij})$$

Enakost velja tudi v primeru, ko je $\mathcal{S}_{ij}^{\star} = \emptyset$.

Fletcherjev postopek

Za izračun matričnega zaprtja nad polnim polkolobarjem uporabljamo Fletcherjev postopek – posplošitev postopkov Kleene, Warshall, Floyd in Roy. Enomestno operacijo zaprtja * definiramo s predpisom: $a^\star = \sum_{i=0}^{\infty} a^i$.
Ne zahtevamo absorpcije ali idempotence.
Naj bo $\mathbf{C_0} = \mathbf{W}$, potem lahko Fletcherjev postopek zapišemo takole:

$$
\begin{array}{l}
\textbf{for } k := 1 \textbf{ to } n \textbf{ do begin} \\
\quad \textbf{for } i := 1 \textbf{ to } n \textbf{ do for } j := 1 \textbf{ to } n \textbf{ do} \\
\quad\quad \mathbf{C}_k[i,j] := \mathbf{C}_{k-1}[i,j] + \mathbf{C}_{k-1}[i,k] \cdot \mathbf{C}_{k-1}[k,k]^\star \cdot \mathbf{C}_{k-1}[k,j]; \\
\quad \mathbf{C}_k[k,k] := 1 + \mathbf{C}_k[k,k] \\
\textbf{end};
\end{array}
$$

$\mathbf{C}_k[i,j]$ je vrednost vseh sprehodov iz vozlišča v_i v vozlišče v_j, ki gredo skozi vozlišča z indeksom največ k. Torej je končna matrika $\mathbf{C}_n = \mathbf{W}^\star$.

Pri programiranju postopka seveda shajamo že z dvema matrikama; oziroma eno samo, če je seštevanje idempotentno. V primeru, ko velja tudi absorpcijski zakon, se postopek še poenostavi, saj je $a^\star = 1$.

Vmesnost in geodezični polkolobar

Vmesnost (betweenness)

$$b(t) = \frac{1}{(n-1)(n-2)} \sum_{u \neq t \neq v} \frac{g_{u,v}(t)}{g_{u,v}}$$

Če poznamo matriki $[d_{u,v}]$ in $[g_{u,v}]$ lahko določimo tudi $g_{u,v}(t)$ takole:

$$g_{u,v}(t) = \begin{cases} g_{u,t} \cdot g_{t,v} & d_{u,t} + d_{t,v} = d_{u,v} \\ 0 & \text{sicer} \end{cases}$$

Za hkratni izračun obeh matrik $[(d_{u,v}, g_{u,v})]$ lahko uporabimo Fletcherjev postopek nad geodezičnim polkolobarjem, ki ga uporabimo na sestavljeni matriki

$$(d,n)_{u,v} = \begin{cases} (1,1) & (u,v) \in \mathcal{L} \\ (\infty, 0) & (u,v) \notin \mathcal{L} \end{cases}$$

…Vmesnost in geodezični polkolobar

V množici $A = (\mathbb{R}_0^+ \cup \{\infty\}) \times (\mathbb{N} \cup \{\infty\})$ vpeljemo operaciji:
seštevanje:

$$(a,i) \oplus (b,j) = (\min(a,b), \begin{cases} i & a < b \\ i+j & a = b \\ j & a > b \end{cases})$$

in **množenje**: $(a,i) \odot (b,j) = (a+b, i \cdot j)$.

Dobljena struktura je poln in zaprt polkolobar

$$(a,i)^\star = \begin{cases} (0,\infty) & a = 0, i \neq 0 \\ (0,1) & \text{sicer} \end{cases}$$

Še učinkovitejši postopek za izračun vmesnosti je predlagal Brandes (2001).

Izračun vmesnosti v R-ju

```
mat.geodesics <- function(m)
{ n <- nrow(m)
  md <- m; md[m==0] <- Inf;  mc <- m; mc[m>0] <- 1
  for (k in 1:n) { for (u in 1:n){ for (v in 1:n){
    dst <- md[u,k] + md[k,v]
    if (md[u,v] >= dst) {
      cnt <- mc[u,k]*mc[k,v];
      if (md[u,v] == dst) {mc[u,v] <- mc[u,v] + cnt }
      else  {md[u,v] <- dst; mc[u,v] <- cnt }
    }
  }}}
  list(dis=md,cnt=mc)
}

vec.betweeness <- function(m)
{ mt <- mat.geodesics(m);  attach(mt)
  n <- nrow(m); bw <- rep(0,n)
  for (v in 1:n) {
    b <- 0
    for (u in 1:n) {for (w in 1:n) {
      if ((cnt[u,w] > 0) && (u != w) && (u != v) && (v != w) &&
            ((dis[u,v] + dis[v,w]) == dis[u,w]))
        {b <- b + cnt[u,v]*cnt[v,w] / cnt[u,w]}
    }}
    bw[v] <- b/((n-1)*(n-2))
  }
  bw
}
```

Še nekaj operacij z matrikami

Transponirana matrika $\mathbf{A}^T$ matrike $\mathbf{A}$ je določena z $a^T_{ij} = a_{ji}$. Velja

$$(\mathbf{A} \cdot \mathbf{B})^T = \mathbf{B}^T \cdot \mathbf{A}^T$$

Matrika $\mathbf{A}$ je simetrična ntk. $\mathbf{A}^T = \mathbf{A}$.
Za realno matriko $\mathbf{A}$ je prirejena dvojiška matrika $b(\mathbf{A}) = [b_{ij}]$ določena z

$$b_{ij} = \begin{cases} 0 & a_{ij} = 0 \\ 1 & \text{sicer} \end{cases}$$

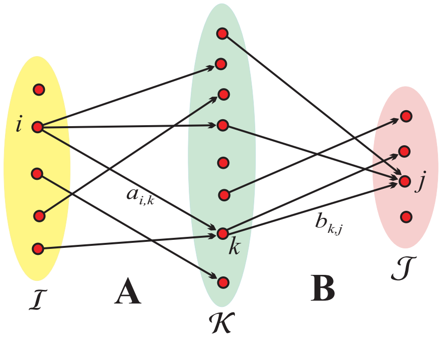
Množenje omrežij

Imejmo usklajeni (dvovrstni) omrežji $\mathcal{N}_A = (\mathcal{I}, \mathcal{K}, \mathcal{A}_A, w_A)$ in $\mathcal{N}_B = (\mathcal{K}, \mathcal{J}, \mathcal{A}_B, w_B)$ s pripadajočima matrikama $\mathbf{A}_{\mathcal{I}\times\mathcal{K}}$ in $\mathbf{B}_{\mathcal{K}\times\mathcal{J}}$. Produkt omrežij $\mathcal{N}_A$ in $\mathcal{N}_B$ je omrežje $\mathcal{N}_C = (\mathcal{I}, \mathcal{J}, \mathcal{A}_C, w_C)$, v katerem so $\mathcal{A}_C = \{(i,j) : i \in \mathcal{I}, j \in \mathcal{J}, c_{i,j} \neq 0\}$ in $w_C(i,j) = c_{i,j}$ za $(i,j) \in \mathcal{A}_C$. Produktna matrika $\mathbf{C} = [c_{i,j}]_{\mathcal{I}\times\mathcal{J}} = \mathbf{A} * \mathbf{B}$ je določena na običajni način

$$c_{i,j} = \sum_{k\in\mathcal{K}} a_{i,k} \cdot b_{k,j}$$

V primeru, ko velja $\mathcal{I} = \mathcal{K} = \mathcal{J}$, imamo opravka z navadnimi enovrstnimi omrežji s kvadratnimi matrikami.

. . . Množenje omrežij



$$c_{i,j} = \sum_{k \in N_A(i) \cap N_B^-(j)} a_{i,k} \cdot b_{k,j}$$

Če so vse uteži v omrežjih $\mathcal{N}_A$ in $\mathcal{N}_B$ enake 1, je vrednost $c_{i,j}$ enaka številu vseh načinov (poti), ki nas vodijo iz vozlišča $i \in \mathcal{I}$ čez množico $\mathcal{K}$ v vozlišče $j \in \mathcal{J}$. $c_{i,j} = |N_A(i) \cap N_B^-(j)|$.

...Množenje omrežij

Standardno matrično množenje ima zahtevnost $O(|\mathcal{I}| \cdot |\mathcal{K}| \cdot |\mathcal{J}|)$ – to je prezahtevno za velika omrežja. Večina velikih omrežij je redkih. Ta lahko zmnožimo veliko hitreje, če upoštevamo samo neničelne vrednosti:

> **for** k **in** $\mathcal{K}$ **do**
>   **for** (i, j) **in** $N_A^-(k) \times N_B(k)$ **do**
>   **if** $\exists c_{i,j}$ **then** $c_{i,j} := c_{i,j} + a_{i,k} \cdot b_{k,j}$
>   **else** new $c_{i,j} := a_{i,k} \cdot b_{k,j}$

`Networks/Multiply Networks`

V splošnem je množenje redkih omrežij "nevarna" operacija, ker lahko produkt "raznese" – sam ni redek.

. . . Množenje omrežij

Iz postopka hitrega množenja vidimo, da vsako vozlišče $k \in \mathcal{K}$ doda v produkt polno dvovrstno podomrežje $K_{N_A^-(k),N_B(k)}$ (ali, v posebnem primeru, polno enovrstno podomrežje $K_{N(k)}$).

Če obstaja tako vozlišče k, da sta obe stopnji $\deg_A(k) = |N_A^-(k)|$ in $\deg_B(k) = |N_B(k)|$ veliki (reda n), je že določitev tega podomrežja kvadratne zahtevnosti (čas in prostor) – rezultat "raznese".

Če pa v redkih omrežjih $\mathcal{N}_A$ in $\mathcal{N}_B$ obstaja v $\mathcal{K}$ le nekaj vozlišč velike stopnje in nobeno nima velike stopnje v obeh omrežjih, je tudi produktno omrežje $\mathcal{N}_C$ redko.

Izpeljane sorodstvene relacije

Iz datoteke GED `Pajek` ustvari 3 relacije (Orejevo omrežje):

F: _ **is a father of** _
M: _ **is a mother of** _
E: _ **is a spouse of** _

Dodamo še "diagonalni" relaciji:

L: _ **is a male** _ / *1-male, 0-female*
J: _ **is a female** _ / *1-female, 0-male*

$$\mathbf{F} \cap \mathbf{M} = \emptyset, \qquad \mathbf{L} \cup \mathbf{J} \subseteq \mathbf{I}, \qquad \mathbf{L} \cap \mathbf{J} = \emptyset$$

...Izpeljane sorodstvene relacije

Ostale sorodstvene zveze lahko izračunamo:

_ is a parent of _	P	$=$	$F \cup M$
_ is a child of _	C	$=$	P^T
_ is a son of _	S	$=$	$L * C$
_ is a daughter of _	D	$=$	$J * C$
_ is a husband of _	H	$=$	$L * E$
_ is a wife of _	W	$=$	$J * E$
_ is a sibling of _	G	$=$	$((F^T * F) \cap (M^T * M)) \setminus I$
_ is a brother of _	B	$=$	$L * G$
_ is a sister of _	Z	$=$	$J * G$
_ is an uncle of _	U	$=$	$B * P$
_ is an aunt of _	A	$=$	$Z * P$
_ is a semi-sibling of _	G_e	$=$	$(P^T * P) \setminus I$

in naprej

_ is a grand mother of _	M_2	$=$	$M * P$
_ is a niece of _	Ni	$=$	$D * G$

...

Bibliografska omrežja

Iz bibliografij (BibTEX) in bibliografskih storitev (Web of Science, Scopus, SICRIS, CiteSeer, Zentralblatt MATH, Google Scholar, DBLP Bibliography, US patent office, IMDb, in druge) lahko pridobimo nekaj dvovrstnih omrežij na izbrano temo:

dela $\times$ avtorji (**WA**),
dela $\times$ gesla (**WK**);

včasih še

dela $\times$ klasifikacija (**WC**)

ter enovrstno omrežje sklicevanj

dela $\times$ dela (**Ci**);

pri čemer dela vključujejo članke, poročila, knjige, patente itd.

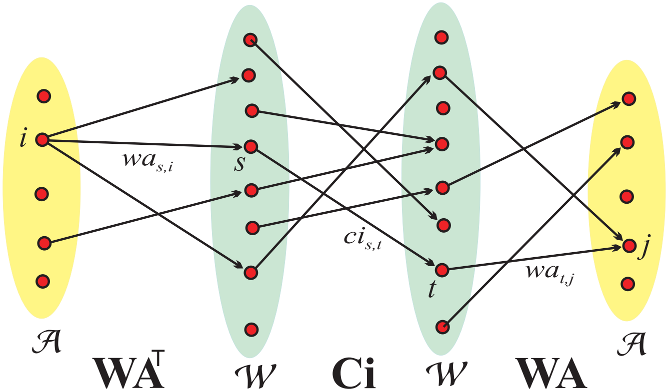
Poleg tega dobimo še razbitje del glede na časopis ali založnika, razbitje del glede na leto izida in vektor s številom strani.

Iz osnovnih bibliometričnih omrežij lahko dobimo z množenjem omrežij izpeljana omrežja. Npr. $\mathbf{Co} = \mathbf{WA}^T * \mathbf{WA}$ je omrežje soavtorstev

$$co_{ij} = \sum_{p \in W} wa_{pi} wa_{pj} = \sum_{p \in N^-(i) \cap N^-(j)} 1$$

co_{ij} = število del, ki sta jih avtorja i in j napisala skupaj
co_{ii} = število del, ki jih je napisal avtor $i = \mathrm{indeg}_{WA}(i)$
Velja: $co_{ij} = co_{ji}$.

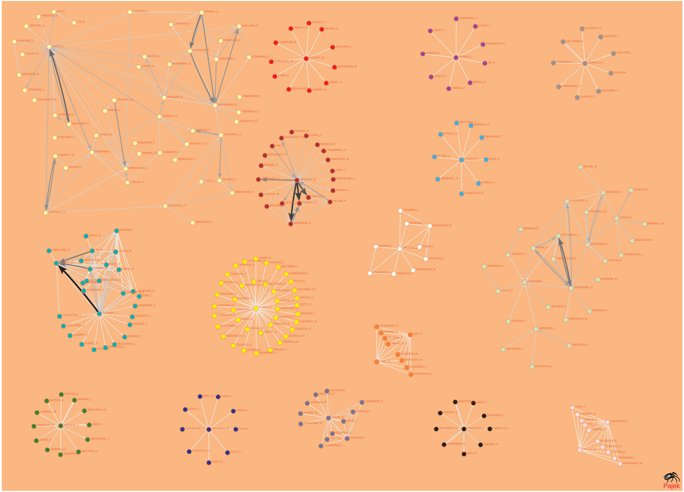
Omrežje sklicevanj med avtorji



Ca $=$ **AW** $*$ **Ci** $*$ **WA** je omrežje sklicevanj med avtorji. Utež $w(i,j)$ šteje sklice iz del, ki jih je napisal avtor i, na dela, ki jih je napisal avtor j.

Otoki v omrežju sklicevanj med avtorji iz SN5

Omrežja SN5 (2008): za "social network*" + najbolj citirana + okoli 100 pomembnih raziskovalcev; $|W| = 193376$, $|C| = 7950$, $|A| = 75930$, $|J| = 14651$, $|K| = 29267$



Časovne količine

Časovna količina je

$$a(t) = \begin{cases} a'(t) & t \in T_a \\ ⌘ & t \in \mathcal{T} \setminus T_a \end{cases}$$

kjer je T_a čas dejavnosti (prisotnosti) a-ja in je $a'(t)$ vrednost a-ja v trenutku $t \in T_a$ ter ⌘ označuje vrednost nedoločeno.

Predpostavimo, da vrednosti časovnih količin pripadajo množici A, ki je polkolobar $(A, +, \cdot, 0, 1)$ za operaciji $+ : A \times A \to A$ in $\cdot : A \times A \to A$. $A_⌘ = A \cup \{⌘\}$.

Z $A_⌘(\mathcal{T})$ označimo množico vseh časovnih količin nad $A_⌘$ v času $\mathcal{T}$. Za razširitev operacij na omrežja in njihove matrike najprej vpeljemo vsoto (vzporedni povezavi) $a + b$ kot

$$(a + b)(t) = a(t) + b(t) \quad \text{in} \quad T_{a+b} = T_a \cup T_b.$$

in produkt (zaporedni povezavi) $a \cdot b$ kot

$$(a \cdot b)(t) = a(t) \cdot b(t) \quad \text{in} \quad T_{a \cdot b} = T_a \cap T_b.$$

Vsota in produkt časovnih količin

```
a = [(1, 5, 2), (6, 8, 1), (11, 12, 3), (14, 16, 2),
     (17, 18, 5), (19, 20, 1)]
b = [(2, 3, 4), (4, 7, 3), (9, 10, 2), (13, 15, 5), (16, 21, 1)]
```

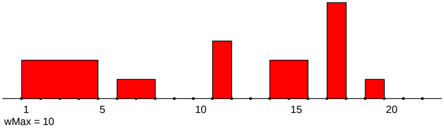
Za gornji časovni količini a in b sta njuna vsota $s = a + b$ in produkt $p = a \cdot b$ nad kombinatoričnim polkolobarjem naslednja

```
s = [(1, 2, 2), (2, 3, 6), (3, 4, 2), (4, 5, 5), (5, 6, 3),
     (6, 7, 4), (7, 8, 1), (9, 10, 2), (11, 12, 3),
     (13, 14, 5), (14, 15, 7), (15, 16, 2), (16, 17, 1),
     (17, 18, 6), (18, 19, 1), (19, 20, 2), (20, 21, 1)]
p = [(2, 3, 8), (4, 5, 6), (6, 7, 3), (14, 15, 10),
     (17, 18, 5), (19, 20, 1)]
```

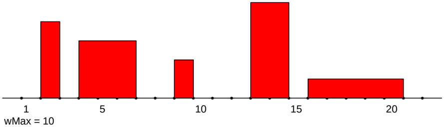
Operaciji lahko prikažemo tudi na sliki.

Vsota časovnih količin

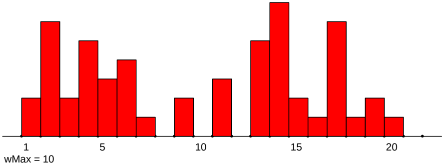
a :



b :

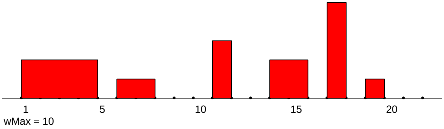


$a + b$:

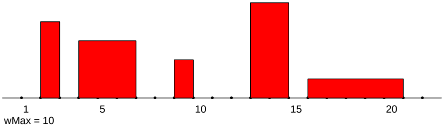


Zmnožek časovnih količin

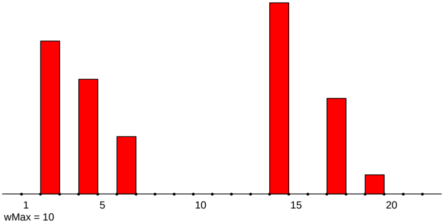
a :



b :



$a \cdot b$:



Časovna omrežja pripadnosti

Naj pripadnostna matrika $\mathbf{A} = [a_{ep}]$ opisuje dvovrstno omrežje z množico dogodkov E in množico udeležencev P:

$$a_{ep} = \begin{cases} 1 & p \text{ se je udeležil dogodka } e \\ 0 & \text{sicer} \end{cases}$$

Funkcija $d : E \to \mathcal{T}$ pripiše vsakemu dogodku e datum $d(e)$, ko se je ta zgodil. $\mathcal{T} = [\mathit{first}, \mathit{last}] \subset \mathbb{N}$. Na osnovi teh podatkov lahko omrežje predelamo v dve ustrezni časovni omrežji:

- **trenutno Ai** $= [ai_{ep}]$, kjer je

$$ai_{ep} = \begin{cases} [(d(e), d(e)+1, 1)] & a_{ep} = 1 \\ [\,] & \text{sicer} \end{cases}$$

- **nakopičeno Ac** $= [ac_{ep}]$, kjer je

$$ac_{ep} = \begin{cases} [(d(e), \mathit{last}+1, 1)] & a_{ep} = 1 \\ [\,] & \text{sicer} \end{cases}$$

Časovna omrežja sopojavljanj

Z množenjem časovnih omrežij pripadnosti nad kombinatoričnim polkolobarjem dobimo ustrezna trenutna oziroma nakopičena omrežja sopojavljanj

$$\mathbf{Ci} = \mathbf{Ai}^T \cdot \mathbf{Ai} \qquad \text{in} \qquad \mathbf{Cc} = \mathbf{Ac}^T \cdot \mathbf{Ac}$$

Tipičen primer tega je omrežje avtorstev **WA**, kjer je E množica del W, P je množica avtorjev A in je d leto izida. Dobimo časovna omrežja soavtorstev.

Trojica (s, f, v) v časovni količini ci_{pq} pove, da je v časovnem intervalu $[s, f)$ bilo v dogodkov, ki sta se jih p in q udeležila oba.

Trojica (s, f, v) v časovni količini cc_{pq} pove, da je v časovnem intervalu $[s, f)$ nabralo skupaj $v$dogodkov, ki sta se jih p in q udeležila oba.

Diagonalne vrednosti ci_{pp} in cc_{pp} vsebujejo časovne količine, ki štejejo dogodke, ki se jih je udeležil p.

BibTime

SN5 (2008)

	W	A	K	J
raw	193376	75930	29267	14651
DC=1	7950	12458		

S `Pajekom` izrežemo podomrežje **WAc** in pripadajoče razbitje **SN5yearC**. S programom `twoMode2netJSON` ju pretvorimo v časovni omrežji in shranimo v obliki netJSON.

Bibliografska omrežja so praviloma redka. Omrežje **WAcInst** ima 19488 povezav. Omrežje soavtorstev **CoInst** $=$ **WAcInst**^{T} $*$ **WAcInst** ima 64980 (neusmerjenih) povezav; pripadajoča matrika v knjižnici **TQ** bi imela $12458^2 = 155201764$ elementov. Z uporabo knjižnice **Graph** je izračun omrežja soavtorstev opravljen v sekundi in pol – precejšna pohitritev.

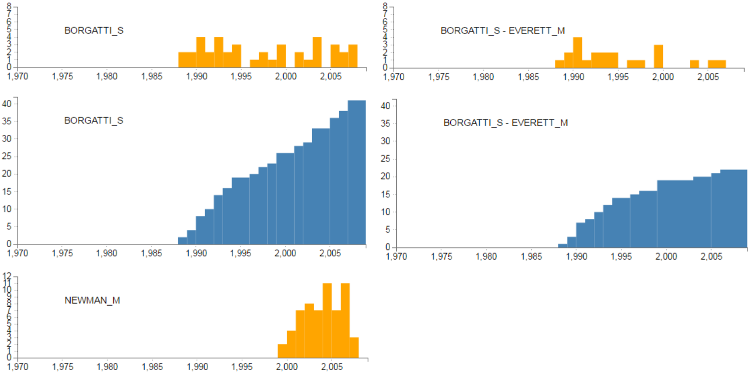
multiply.py

```
gdir = 'c:/users/batagelj/work/python/graph/graph'
wdir = 'c:/users/batagelj/work/python/graph/JSON/SN5'
cdir = 'c:/users/batagelj/work/python/graph/chart'
import sys, os, datetime, json
sys.path = [gdir]+sys.path; os.chdir(wdir)
import TQ
from GraphNew import Graph
# file = 'C:/Users/batagelj/work/Python/graph/JSON/WAtest.json'
file = 'C:/Users/batagelj/work/Python/graph/JSON/SN5/WAcInst.json'
# file = 'C:/Users/batagelj/work/Python/graph/JSON/SN5/WAcCum.json'
# file = 'C:/Users/batagelj/work/Python/graph/JSON/Gisela/papInst.json'
t1 = datetime.datetime.now()
print("started: ",t1.ctime(),"\n")
G = Graph.loadNetJSON(file)
t2 = datetime.datetime.now()
print("\nloaded: ",t2.ctime(),"\ntime used: ", t2-t1)
# T = G.transpose()
# Co = Graph.TQmultiply(T,G,True)
# CR = G.TQtwo2oneRows()
CC = G.TQtwo2oneCols()
t3 = datetime.datetime.now()
print("\ncomputed: ",t3.ctime(),"\ntime used: ", t3-t2)
ia = { v[3]['lab']: k for k,v in CC._nodes.items() }
# CC._links[(ia['BORGATTI_S'],ia['EVERETT_M'])][4]['tq']
# CC._links[(ia['IDI/B'],ia['HCL/B'])][4]['tq']
```

Časovna soavtorstva v SN5

```
======= RESTART: C:\Users\batagelj\work\Python\graph\graph\multiply.py =======
started:  Sun Nov 20 00:26:51 2016
loaded:  Sun Nov 20 00:26:51 2016
time used:  0:00:00.425024
computed:  Sun Nov 20 00:26:52 2016
time used:  0:00:01.165066
>>> BB = CC._links[(ia['BORGATTI_S'],ia['BORGATTI_S'])][4]['tq']
>>> BE = CC._links[(ia['BORGATTI_S'],ia['EVERETT_M'])][4]['tq']
>>> BB
[(1988, 1990, 2), (1990, 1991, 4), (1991, 1992, 2), (1992, 1993, 4),
 (1993, 1994, 2), (1994, 1995, 3), (1996, 1997, 1), (1997, 1998, 2),
 (1998, 1999, 1), (1999, 2000, 3), (2001, 2002, 2), (2002, 2003, 1),
 (2003, 2004, 4), (2005, 2006, 3), (2006, 2007, 2), (2007, 2008, 3)]
>>> BE
[(1988, 1989, 1), (1989, 1990, 2), (1990, 1991, 4), (1991, 1992, 1),
 (1992, 1995, 2), (1996, 1998, 1), (1999, 2000, 3), (2003, 2004, 1),
 (2005, 2007, 1)]
>>> TQmax = 8; Tmin = 1970; Tmax = 2009; w = 600; h = 120
>>> tit = 'BORGATTI_S'
>>> Graph.TQshow(BB,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='orange')
>>> tit = 'BORGATTI_S - EVERETT_M'
>>> Graph.TQshow(BE,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='orange')
>>> NN = CC._links[(ia['NEWMAN_M'],ia['NEWMAN_M'])][4]['tq']
>>> NN
[(1999, 2000, 2), (2000, 2001, 4), (2001, 2002, 7), (2002, 2003, 8),
 (2003, 2004, 7), (2004, 2005, 11), (2005, 2006, 7), (2006, 2007, 11),
 (2007, 2008, 3)]
>>> tit = 'NEWMAN_M'; TQmax = 12; h = 150
>>> Graph.TQshow(NN,cdir,TQmax,Tmin,Tmax,w,h,tit,fill='orange')
```

Prikazi


BORGATTI_S
BORGATTI_S - EVERETT_M
BORGATTI_S
BORGATTI_S - EVERETT_M
NEWMAN_M
1,970
1,975
1,980
1,985
1,990
1,995
2,000
2,005

Zaključek

1. z uporabo ustreznih polkolobarjev se lahko spoprimemo z različnimi problemi na omrežjih in jih postavimo na skupne temelje;
2. množenje omrežij omogoča pridobitev izpeljanih omrežij. Pri teh se pokaže, da moramo paziti tudi na normalizacijo vrednosti.

Viri I

- Batagelj, V: Semirings for social networks analysis. J Math Sociol 19 (1): 53-68 1994
- Batagelj, V.: Wos2pajek – networks from web of science (2007). http://vladowiki.fmf.uni-lj.si/doku.php?id=pajek:wos2pajek
- Batagelj, V.: Nets – Python package for network analysis. https://github.com/bavla/Nets
- Batagelj, V, Cerinšek, M: On bibliographic networks. Scientometrics 96 (2013) 3, 845-864.
- Batagelj, V., Doreian, P., Ferligoj, A., and Kejžar, N.: Understanding Large Temporal Networks and Spatial Networks. Wiley, 2014.
- Batagelj, V., Praprotnik, S.: An algebraic approach to temporal network analysis based on temporal quantities. Social Network Analysis and Mining, 6(2016)1, 1-22.
- Brandes, U.: A Faster Algorithm for Betweenness Centrality, Journal of Mathematical Sociology, 25 (2001), 163–177.

Viri II

Cerinšek, M, Batagelj, V: Semirings and Matrix Analysis of Networks. in Alhajj, R, Rokne, J (Eds.) Encyclopedia of Social Network Analysis and Mining. Springer, October 2017, 1-8.

De Nooy, W., Mrvar, A., Batagelj, V.: Exploratory Social Network Analysis with Pajek; Revised and Expanded Second Edition. Structural Analysis in the Social Sciences, Cambridge University Press, 2011.

Fletcher, J. G.: "A more general algorithm for computing closed semiring costs between vertices of a directed graph," *CACM* (1980), pp. 350-351.

Zaveršnik, M., Batagelj, V.: Islands. In: XXIV International Sunbelt Social Network Conference, Portorož, Slovenia, 2004.