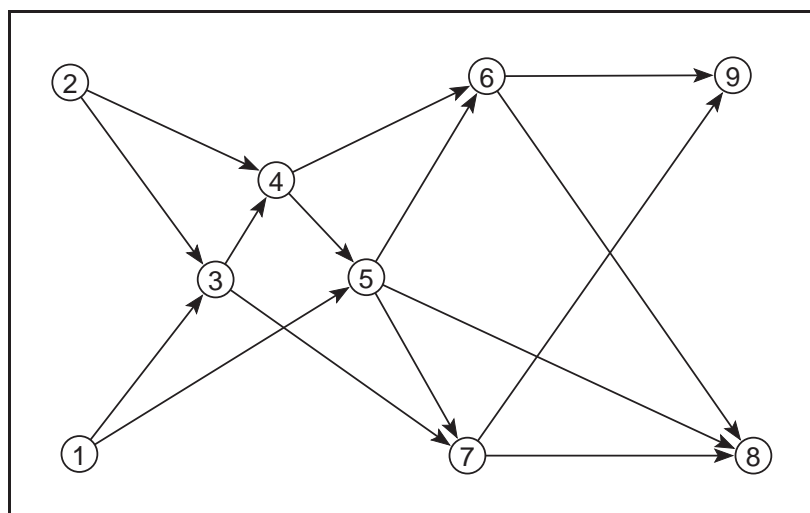


Analiza omrežij citiranj

Vhodni podatki: Aciklično omrežje, ki predstavlja citiranje med članki.

Rezultat: Relativne pomembnosti člankov in relativne pomembnosti citiranj. Primer omrežja citiranj:



V omrežju citiranj predstavljajo točke članke, usmerjena povezava od članka x do članka y pomeni, da y citira x (x je citiran v y).

Ugotoviti želimo, kateri članek je najpomembnejši in katera povezava oz. citiranje je bistveno za celotno omrežje.

Najpreprostejša mera pomembnosti je, da preštejemo kolikokrat je dani članek citiran. V naši obliki predstavitve to pomeni, da preštejemo koliko puščic vodi iz dane točke. Glede na ta kriterij je najpomembnejši članek 5 s tremi citirani.

Boljša mera za pomembnost točke v omrežju je zgrajena na osnovi števila poti, ki vodijo skozi dano točko.

Algoritem

- Omrežju dodamo enolični začetek in enolični konec.
- Opravimo dva prehoda skozi omrežje:

1. V prvem prehodu gremo od začetne do končne točke in si za vsako točko zapomnimo, koliko poti vodi vanjo iz začetne točke (to ugotovimo z zaporednim prištevanjem poti glede na točke iz katerih pridemo v dano točko).
2. V drugem prehodu gremo od končne točke nazaj proti začetni in si za vsako točko zapomnimo, koliko poti vodi iz te točke proti končni točki.

- Iz dobljenih podatkov lahko za vsako točko v grafu ugotovimo število vseh poti od začetne do končne točke, ki vodijo skozi to točko. Uporabimo osnovni izrek kombinatorike:

Če iz začetne do izbrane točke vodi M poti, od izbrane do končne pa N poti, potem je število vseh poti, od začetne do končne, ki gredo skozi izbrano točko, enako MN .

- Podobno ugotovimo koliko poti uporablja dano

povezavo. Naj vodi izbrana povezava od točke i do točke j .

Če iz začetne točke vodi v točko i M različnih poti, iz točke j pa vodi v končno točko N različnih poti, potem je število poti, ki uporablja povezavo (i, j) , enako MN .

- Za lažje primerjanje rezultatov iz različnih omrežij ponavadi omenjena števila delimo s številom vseh poti v grafu, tako da dobimo pomembnosti člankov/citiranj v intervalu $[0, 1]$.

Obstajata dve varianti algoritma:

- metoda *Paths Count (PC)* – to je ravnokar opisana metoda (štejemo poti od začetka do konca);
- metoda *Search Path Link Count (SPLC)* – ta metoda se od prejšnje loči po tem, da vsako točko v grafu obravnavamo kot začetek, zato iz enoličnega začetka dodamo še povezave do vseh točk grafa – število vseh poti v grafu se poveča.

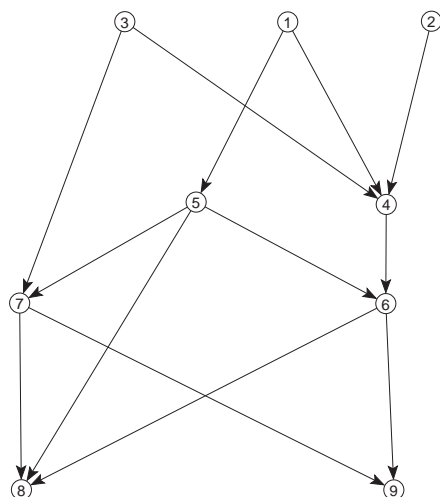
Absolutne in relativne pomembnosti člankov po metodah Paths Count (PC) in Search Path Link Count (SPLC)

članek	<i>PC</i>	<i>PC/30</i>	<i>SPLC</i>	<i>SPLC/58</i>
1	14	0.4667	14	0.2414
2	16	0.5333	16	0.2759
3	18	0.6000	27	0.4655
4	21	0.7000	35	0.6034
5	20	0.6667	35	0.6034
6	14	0.4667	26	0.4483
7	12	0.4000	22	0.3793
8	17	0.5667	32	0.5517
9	13	0.4333	25	0.4310

**Absolutne in relativne pomembnosti citiranj po metodah
Paths Count (PC) in Search Path Link Count (SPLC)**

od	do	<i>PC</i>	<i>PC/30</i>	<i>SPLC</i>	<i>SPLC/58</i>
1	3	9	0.3000	9	0.1551
2	3	9	0.3000	9	0.1551
1	5	5	0.1667	5	0.0862
2	4	7	0.2333	7	0.1207
3	7	4	0.1333	6	0.1034
3	4	14	0.4667	21	0.3621
4	5	15	0.5000	25	0.4310
4	6	6	0.2000	10	0.1724
5	7	8	0.2667	14	0.2414
5	6	8	0.2667	14	0.2414
5	8	4	0.1333	7	0.1207
6	8	7	0.2333	13	0.2241
6	9	7	0.2333	13	0.2241
7	8	6	0.2000	11	0.1897
7	9	6	0.2000	11	0.1897

Naloga 1



Naloga 2

