

# ECCENTRICITY, CLOSENESS AND BETWEENNESS CENTRALITY KRATEK OPIS

EVA DEŽELAK, NEJC LUKEŽIČ

## 1. UVOD

Pri najinem projektu bova s pomočjo treh različnih mer (ekscentričnosti, bližine in vmesne centralnosti) iskala in analizirala najpomembnejša vozlišča v grafih oziroma socialnih omrežjih. Vozlišče je pomembnejše, če ima višjo vrednost bližine in vmesne centralnosti ter nižjo vrednost ekscentričnosti. Natančnejši opisi pojmov so spodaj. Primerjala bova, kako pogosto je vozlišče, ki je pomembno z vidika ene mere, pomembno tudi v okviru ostalih dveh mer.

Najino glavno orodje za analizo grafov bo *Sage*, kjer bova generirala več grafov (približno 1000) v različnih velikostih (približno 10), grafe socialnih omrežij pa bova pridobila s spleta. Opazovala bova tudi, kako se vrednosti mer za najpomembnejša vozlišča spremenijo, če se omejimo na podgraf v določenem grafu. Grafi, ki jih bova pri projektu analizirala, bodo neusmerjeni.

## 2. RAZLAGA POJMOV

**2.1. Bližina (Closeness centrality/Closeness).** Bližina je v povezanem grafu mera centralnosti, ki jo izračunamo kot recipročno vsoto dolžin najkrajših poti med nekim vozliščem in vsemi drugimi vozlišči v grafu. Bližje kot je opazovano vozlišče ostalim vozliščem v grafu, bolj centralno je.

$$C(x) = \frac{1}{\sum_y d(y, x)},$$

kjer je  $d(y, x)$  razdalja med vozliščema  $x$  in  $y$ . Pogosto se namesto zgornje vrednosti izračuna povprečno dolžino najkrajše poti v grafu. Dobimo jo tako, da zgornjo formulo pomnožimo z  $N - 1$ , kjer je  $N$  število vseh vozlišč v grafu. Pri obsežnejših grafih se  $-1$  izpusti iz enačbe, zato se za bližino uporablja kar sledečo formulo:

$$C(x) = \frac{N}{\sum_y d(y, x)}.$$

Pri usmerjenih grafih je potrebno upoštevati tudi smer povezav. Določeno vozlišče ima lahko različno bližino za vhodne in izhodne povezave. V nepovezanih grafih namesto recipročne vsote dolžin najkrajših poti med vozlišči računamo vsoto recipročnih dolžin najkrajših poti med vozlišči. Pri tem upoštevamo, da  $\frac{1}{\infty} = 0$ .

$$H(x) = \sum_{y \neq x} \frac{1}{d(y, x)}.$$

**2.2. Ekscentričnost (Eccentricity).** Ekscentričnost nekega vozlišča  $v$  v povezanem grafu  $G$  označimo z  $\epsilon(v)$  in je definirana kot maksimalna dolžina med vozliščem  $v$  in katerikoli drugim vozliščem v grafu  $G$ . V nepovezanih grafih imajo vsa vozlišča neskončno vrednost ekscentričnosti. Maksimalno ekscentričnost v grafu imenujemo diameter (premer) grafa (najdaljša najkrajša pot med dvema vozliščema grafa), minimalno ekscentričnost pa polmer grafa.

**2.3. Vmesna centralnost (Betweenness centrality).** V teoriji grafov je "betweenness centrality" mera centralizacije grafa, ki temelji na najkrajših poteh v grafu. Za vsak par vozlišč v povezanem grafu, obstaja vsaj ena najkrajša pot med vozliščema tako, da je katerokoli število povezav, po katerih gre ta pot (za neutežene grafe) ali pa vsota uteži na povezavah (za utežene grafe) minimalna. "Betweenness centrality" za vsako vozlišče je število teh najkrajših poti, ki grejo skozi vozlišče. "Betweenness centrality" se uporablja v mnogih problemih v teoriji omrežij, tudi v problemih povezanih s socialnimi omrežji, biologijo in transportom. V telekomunijskem omrežju ima vozlišče z višjim "betweenness centrality" večjo kontrolo nad omrežjem, ker bo več informacij teklo čez to vozlišče. "Betweenness centrality" vozlišča  $v$  je podana z izrazom:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Kjer je  $\sigma_{st}$  skupno število najkrajših poti od vozlišča  $s$  do vozlišča  $t$  in  $\sigma_{st}(v)$  je število teh poti, ki grejo skozi  $v$ . Potrebno je upoštevati, da "betweenness centrality" nekega vozlišča skalira s številom parov vozlišč, na kar implicirajo tudi indeksi v vsoti. Zato se lahko računanje spremeni tako, da delimo s številom parov vozlišč, ki ne vsebujejo  $v$ , tako da je  $g \in [0, 1]$ . Deljenje se izvede z  $(N-1)(N-2)$  za usmerjene grafe in  $(N-1)(N-2)/2$  za neusmerjene grafe, kjer je  $N$  število vozlišč v povezani komponenti grafa. To skalira za najvišjo možno vrednost, kjer eno vozlišče seka vsaka najkrajša pot. Vendar pa ponavadi ni tako, zato je potrebna normalizacija

$$normal(g(v)) = \frac{g(v) - \min(g)}{\max(g) - \min(g)},$$

ki prinese rezultat:

$$\max(normal) = 1; \min(normal) = 0.$$

Skaliranje vedno poteka iz manjšega obsega v večji obsega zato se ne izgubi nič natančnosti.

V uteženih omrežjih se povezave med vozlišči ne štejejo več za binarne interakcije, ampak so utežene v razmerju z njihovo kapaciteto, kar doda novo dimenzijo heterogenosti znotraj omrežja. Moč vozlišča v uteženem omrežju je podana z vsoto uteži na sosednjih povezavah.

$$s_i = \sum_{j=1}^N a_{ij} \cdot w_{ij}$$

Kjer sta  $a_{ij}$  in  $w_{ij}$  matrika sosednosti in matrika uteži med vozliščema  $i$  in  $j$ . Moč danega vozlišča sledi "pravilu porazdelitve moči".