

Koeficient zhlukovania v sieťach

Beáta Stehlíková

2-EFM-155 Analýza sociálnych sietí

Fakulta matematiky, fyziky a informatiky, UK v Bratislave, 2019

Koeficient zhukovania v sieťach

- ▶ Základná myšlienka: Koeficient má vyjadrovať, nakoľko majú vrcholy tendenciu zhlukovať sa
- ▶ Zhlukovanie na príklade siete priateľstiev: ak sú A a B moji priatelia, sú priateľmi aj A a B?
- ▶ Celkovo ako sieť alebo samostatne pre každý vrchol

Globálny koeficient zhukovania

Definícia a príklad

Vypočíta sa ako podiel, kde

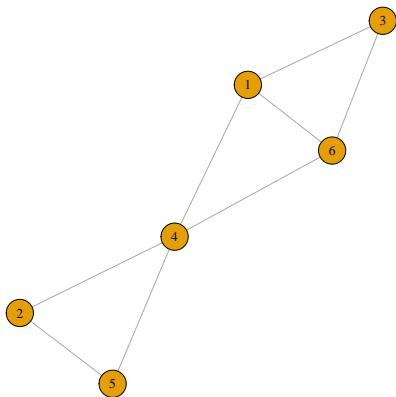
- ▶ v menovateli je dvojíc hrán so spoločným vrcholom, teda typu $i-j$, $i-k$
- ▶ v čitateli je počet takých dvojíc týchto hrán, pre ktoré existuje hrana $j-k$

Príklad Vypočítajme koeficient ručne na príklade siete

```
set.seed(12345)  
g = erdos.renyi.game(6, 0.6)
```

Definícia a príklad

```
plot(g, layout = layout_nicely)
```



Výpočet v R-ku

Funkcia transitivity s parametrom type = "global"

```
transitivity(g, type = "global")
```

```
## [1] 0.6
```

Lokálny koeficient zhukovania

Definícia a príklad

Pre každý vrchol i sa vypočíta sa ako podiel, kde

- ▶ v menovateli je dvojíc hrán so spoločným vrcholom i , teda typu $i-j$, $i-k$
- ▶ v čitateli je počet takých dvojíc týchto hrán, pre ktoré existuje hrana $j-k$

Pre celú sieť sa vypočíta ich priemer.

Príklad: Čo z nášho ručného výpočtu vyplýva pre lokálne koeficienty jednotlivých vrcholov?

Výpočet v R-ku

Funkcia `transitivity` s parametrom `type = "local"`, resp. `"localaverage"`.

```
round(transitivity(g, type = "local"), 4)
```

```
## [1] 0.6667 1.0000 1.0000 0.3333 1.0000 0.6667
```

```
transitivity(g, type = "localaverage")
```

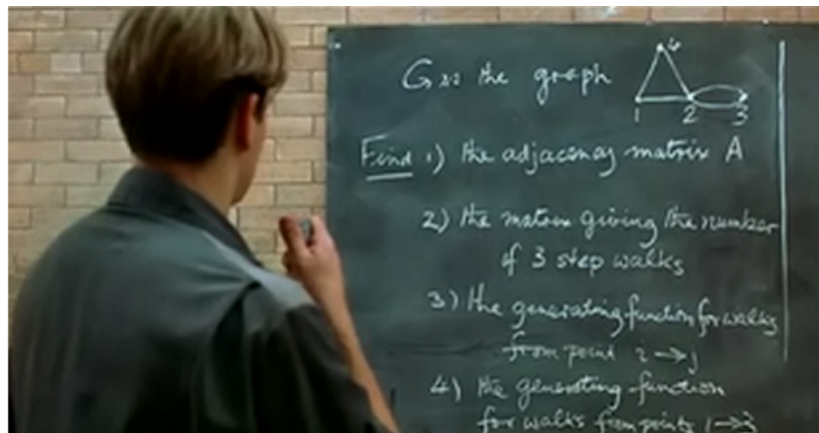
```
## [1] 0.7777778
```

Výpočet z matice susednosti

Ako by sme túto hodnotu vedeli získať z matice susednosti grafu

```
A <- as_adjacency_matrix(g) ?
```

Z filmu *Dobry Will Hunting*, tiež nás zaujíma druhá otázka:



Koeficient zhukovania vo vážených úplných
sieťach

Motivácia a referencia

- ▶ Minule sme pracovali so sieťami korelácií
- ▶ V tejto sieti existuje medzi každými dvoma vrcholmi hrana
- ▶ Ako merať mieru zhlukovania v takýchto sieťach?

McAssey, M. P., & Bijma, F. (2015). A clustering coefficient for complete weighted networks. Network Science, 3(2), 183-195.

https://www.researchgate.net/publication/272893386_A_clustering_coefficient_for_complete_weighted_networks

Prepokladá sa úplná sieť s váhami z intervalu $[0, 1]$.

V článku sa ďalej uvažujú aj orientované vážené siete.

Základná myšlienka

Vrchol má ako susedov, všetky ostatné vrcholy, ale budeme rozlišovať (voľne, bez presných hraníc)

- ▶ silných susedov (*strong*), ak je váha blízka jednotke
- ▶ slabých susedov (*weak*), ak je váha blízka nule
- ▶ stredných susedov (*moderate*) v ostatných prípadoch

Požiadavky na koeficient:

- C1. The clustering coefficient for node i should be large (close to one) if the set of strong neighbors of i are themselves strong neighbors of each other, and should become smaller as the proportion of its strong neighbors who are themselves weak neighbors increases.
- C2. As the weights of the links involving the remaining neighbors of node i increase, the clustering coefficient for node i should also increase proportionately.
- C3. The clustering coefficient for node i should be small if it has only weak neighbors, or at most one non-weak neighbor.

Príklady použitia

- ▶ *Amador, J., Cabral, S., Mastrandrea, R., & Ruzzenenti, F. (2018). Who's who in global value chains? A weighted network approach. Open Economies Review, 29(5), 1039-1059.*
- ▶ *Leonidov, A., & Serebryannikova, E. (2019). Dynamical topology of highly aggregated input-output networks. Physica A: Statistical Mechanics and its Applications, 518, 234-252.*
- ▶ *Clemente, G. P., Grassi, R., & Hitaj, A. (2019). Asset allocation: new evidence through network approaches. Annals of Operations Research, 1-20.*
- ▶ *Cerqueti, R., Clemente, G. P., & Grassi, R. (2018). A network-based measure of the socio-economic roots of the migration flows. Social Indicators Research, 1-18.*

Odkazy na uvedené publikácie:

- ▶ <https://link.springer.com/article/10.1007/s11079-018-9499-7>
- ▶ <https://www.sciencedirect.com/science/article/pii/S0378437118315097>
- ▶ <https://link.springer.com/article/10.1007/s10479-019-03136-y>
- ▶ <https://link.springer.com/article/10.1007/s11205-018-1883-6>

Algoritmus

- ▶ Pre každé $t \in (0, 1)$ sa definuje (nevážená) sieť, ktorá obsahuje hrany s váhou aspoň t
- ▶ Pre každý z vrcholov sa vypočíta lokálny koeficient zhlukovania, označíme ho $C_i(t)$, kde index i predstavuje vrchol (ak nie je definovaný kvôli jedinej hrane niektorého vrcholu, položíme ho rovný nule)
- ▶ Výsledný koeficient zhlukovania potom vypočítame ako

$$C_i = \int_0^1 C_i(t) dt$$

(prakticky len konečná suma)

- ▶ Pre celú sieť sa zoberie priemer koeficientov jednotlivých vrcholov

Vysvetlite maticový výpočet z článku:

2. Let $\gamma_i(t)$ denote the number of triangles formed by consecutive edges with node i at one vertex and any two neighbors of node i as the other two vertices, and let $\Gamma_i(t)$ denote the number of triangles that would be formed with node i at one vertex if every pair of neighbors of node i were also neighbors of each other, i.e.,

$$\gamma_i(t) = \sum_{j \neq i} \sum_{k \neq i, j} a_{ij}^t a_{jk}^t a_{ik}^t = [A_t^3]_{ii} \quad \text{and} \quad \Gamma_i(t) = \sum_{j \neq i} \sum_{k \neq i, j} a_{ij}^t a_{ik}^t = [A_t O A_t]_{ii}. \quad (4)$$

Here $O = \mathbf{1} \cdot \mathbf{1}' - I$, that is, a matrix consisting of zeros on the diagonal and ones in all other positions, I is the $N \times N$ identity matrix, $\mathbf{1}$ is a vector of length N consisting of ones in every position, and $\mathbf{1}'$ is its transpose. The clustering coefficient $C_i(t)$ for node i corresponding to \mathcal{N}_t is then defined as the ratio of these two quantities (which is the established clustering coefficient for a node in a BUN), i.e.,

$$C_i(t) = \frac{\gamma_i(t)}{\Gamma_i(t)} = \frac{[A_t^3]_{ii}}{[A_t O A_t]_{ii}}, \quad (5)$$

provided $[A_t O A_t]_{ii} \neq 0$. Otherwise set $C_i(t) = 0$.

Príklad 1: Priebeh pomocných funkcií

- ▶ Pre sieť z obrázku 1 v článku znázorníme priebeh funkcií $C_i(t)$ pre jednotlivé vrcholy a vypočítame pre ne koeficienty zhlukovania.

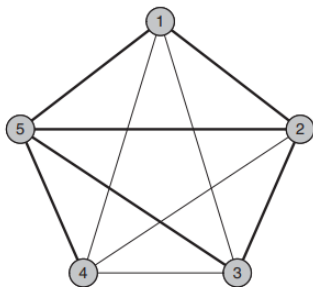


Fig. 1. A complete weighted undirected network consisting of five nodes, with line thickness corresponding to edge weights (thick edge = 0.9, thin edge = 0.1).

Príklad 1: Priebeh pomocných funkcií

- ▶ Zrekonštruujeme obrázok 4 z článku ($C_i(t)$ pre vygenerované váhy - kvôli náhodnosti môže výjsť iný výsledok)

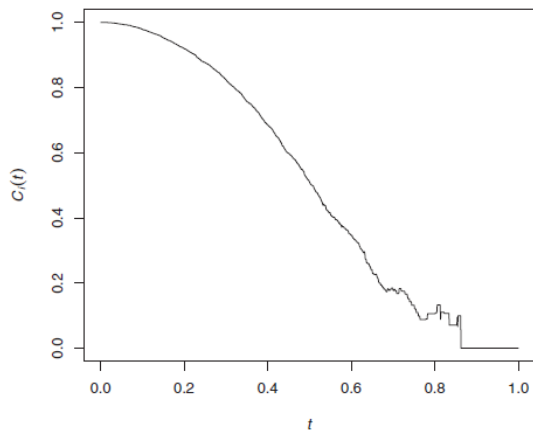


Fig. 3. Evolution of $C_i(t)$ as t ranges from 0 to 1 for one node in a WUN consisting of 100 nodes with edge weights drawn from a uniform distribution on $(0, 1)$.

Príklad 2: Výpočet koeficientov

Všeobecné zadanie: Implementovať túto metódu, funkcia dostane ako vstup maticu váh a vráti vektor hodnôt C_i .

Pokračujme v príklade:

- ▶ Vypočítajme koeficienty zhlukovania pre vrcholy siete na obrázku 1.

Príklad 3: Zmena zhlukovania v čase

Zoberme dáta o indexoch a výnosoch z minulého cvičenia:

```
library(datasets)
data(EuStockMarkets)
vynosy <- diff(log(EuStockMarkets))
```

- ▶ Rozdelíme dáta po rokoch a z každého roka vypočítame korelačnú maticu.
- ▶ Na základe korelačnej matice vypočítame pre každý index a pre každý rok koeficient zhlukovania.
- ▶ Ako sa koeficienty menia v čase?