

# Hľadanie komunit v sieťach - časť I.

Beáta Stehlíková

2-EFM-155 Analýza sociálnych sietí

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

## Hľadanie komunití v sieťach

V tejto časti:

## **Klika**

- ▶ “každý pozná každého”
- ▶ podmnožina vrcholov siete, každé dva vrcholy v tejto množine sú spojené hranou

## **K-jadro**

- ▶ “každý pozná aspoň K ľudí”
- ▶ podmnožina vrcholov siete, každý vrchol je spojený hranou s aspoň K vrcholmi tejto množiny

Nabudúce:

## **Komunita**

- ▶ nemá presnú definíciu
- ▶ vrcholy patriace do komunity by mali mať veľa hrán medzi sebou a málo hrán s vrcholmi mimo komunity

Klika (clique)

## Klika

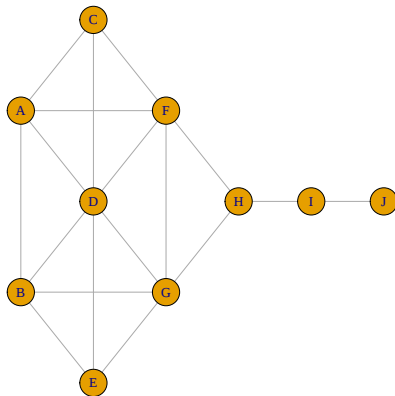
- ▶ Podgraf, každé dva jeho vrcholy sú spojené hranou

Poznámky:

- ▶ Každé dva vrcholy spojené hranou sú klikou veľkosti 2 → prakticky nás budú zaujímať kliky veľkosti aspoň 3
- ▶ Podgrafy kliky veľkosti  $N$  sú tiež kliky → definuje sa **maximálna klika** (*maximal clique*), ktorá nie je vlastným podgrafom inej kliky
- ▶ v R-ku funkcie
  - ▶ `cliques` (výstup je typu `list`)
  - ▶ `maximal.cliques`
  - ▶ `largest.cliques` - kliky s najväčším počtom vrcholov

# Príklad 1

```
library(igraphdata)  
data(kite)  
plot(kite)
```



## Príklad 1: kliky

```
kliky.vsetky <- cliques(kite)
length(kliky.vsetky)
```

```
## [1] 41
```

```
kliky.vsetky
```

```
## [[1]]
## + 1/10 vertex, named, from 6b7ddad:
## [1] D
##
## [[2]]
## + 1/10 vertex, named, from 6b7ddad:
## [1] H
##
## [[3]]
## + 1/10 vertex, named, from 6b7ddad:
```

## Príklad 1: kliky

```
kliky.vsetky.2 <- cliques(kite, min = 3)
length(kliky.vsetky.2)
```

```
## [1] 13
```

```
kliky.vsetky.2
```

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

```
## + 3/10 vertices, named, from 6b7ddad:
```

```
## [1] A D F
```

```
##
```

```
## [[2]]
```

```
## + 3/10 vertices, named, from 6b7ddad:
```

```
## [1] F G H
```

```
##
```

```
## [[3]]
```

```
## + 3/10 vertices, named, from 6b7ddad:
```

## Príklad 1: maximálne kliky

```
kliky <- maximal.cliques(kite, min = 3)
length(kliky)
```

```
## [1] 5
```

*Cvičenie:* Ktoré kliky sú to? Zistite z grafu a overte vypísaním výstupu z R-ka

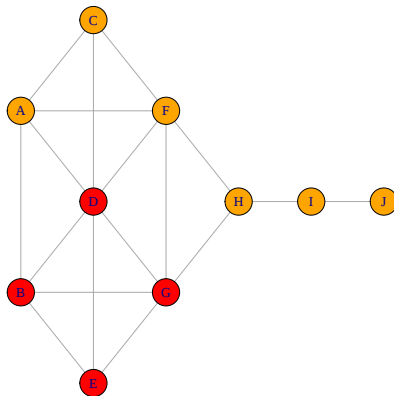
## Príklad 1: najväčšie kliky

```
kliky <- largest.cliques(kite)
kliky
```

```
## [[1]]
## + 4/10 vertices, named, from 6b7ddad:
## [1] D G B E
##
## [[2]]
## + 4/10 vertices, named, from 6b7ddad:
## [1] D A C F
```

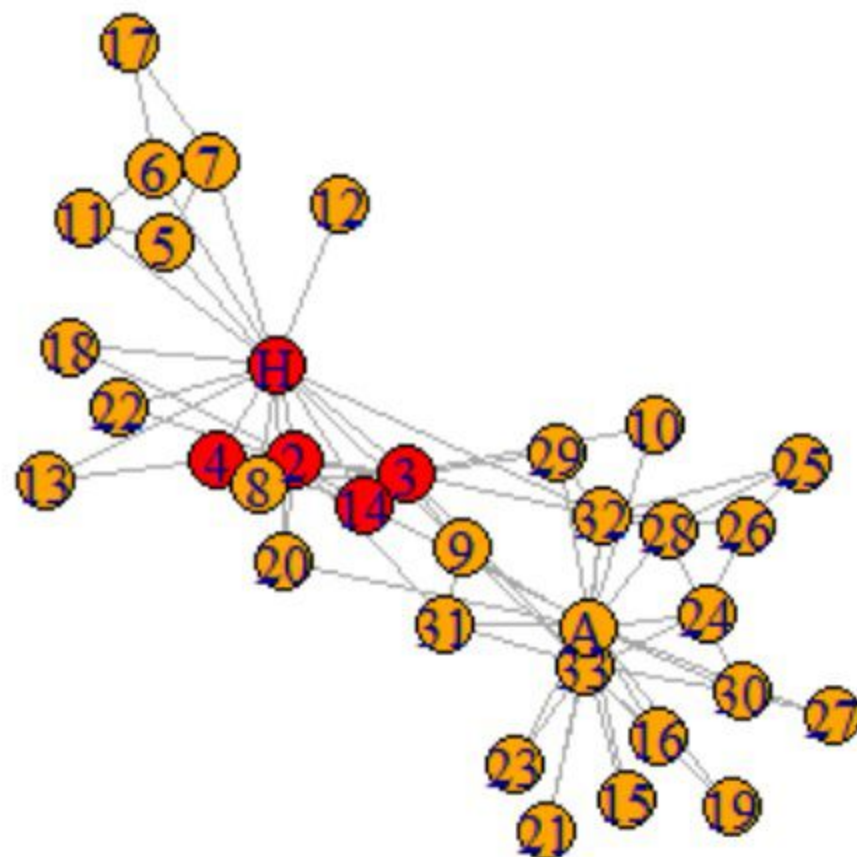
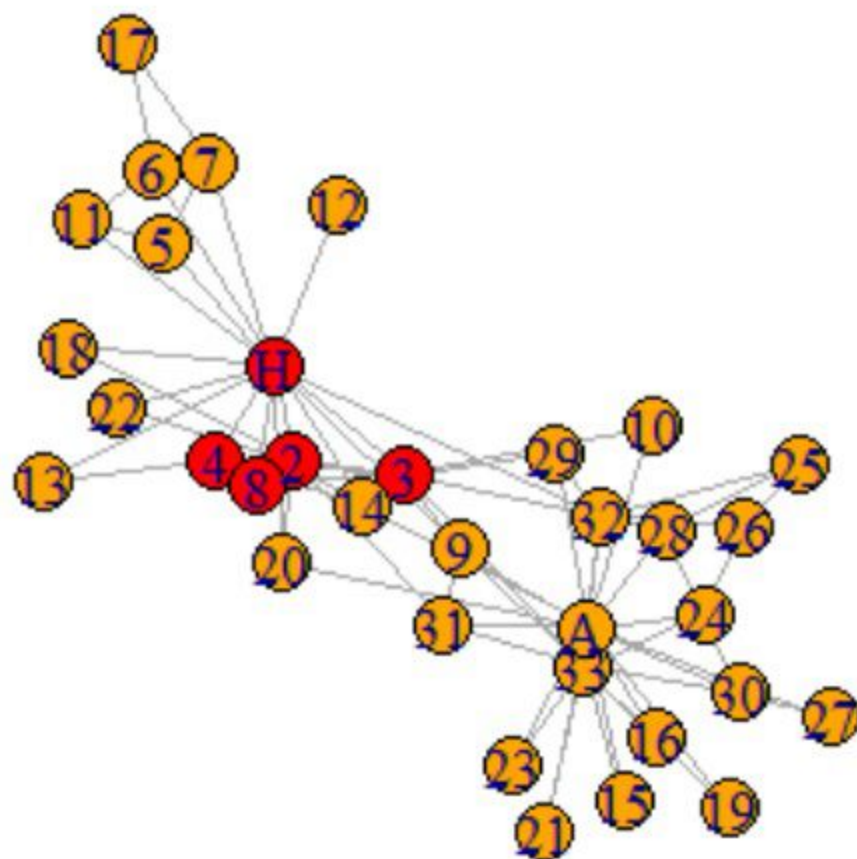
## Príklad 1: grafické znázornenie vrcholov

```
plot(kite,  
     vertex.color = ifelse(V(kite) %in% kliky[[1]],  
                           "red", "orange"))
```



## Príklad 2

Zopakujte na príklade Zacharyho karate klubu. Aká je veľkosť najväčšej kliky? Znázornite ju graficky.



## Príklad 3: Davis' Southern Women Club

Zdroj dát: <https://toreopsahl.com/datasets/#southernwomen>

Popis: *This dataset was collected by Davis and colleague in the 1930s. It contains the observed attendance at 14 social events by 18 Southern women*

Dáta, ktoré použijeme:

- ▶ [http://opsahl.co.uk/tnet/datasets/Davis\\_southern\\_club\\_women-binary.txt](http://opsahl.co.uk/tnet/datasets/Davis_southern_club_women-binary.txt)
- ▶ zoznam hrán v sieti (napríklad 1-2, 1-3 atď. )
- ▶ hrana znamená, že ženy sa obidve zúčastnili na nejakej akcii

```
women <- read.table("http://opsahl.co.uk/tnet/datasets/Davis_southern_club_women-binary.txt")  
head(women)
```

```
##    V1 V2 V3  
## 1   1  2  1  
## 2   1  3  1  
## 3   1  4  1
```

## Príklad 3: Davis' Southern Women Club

Použijeme funkciu `graph_from_data_frame`, ktorá je ďalším spôsobom, ako z dát spraviť v R-ku sieť.

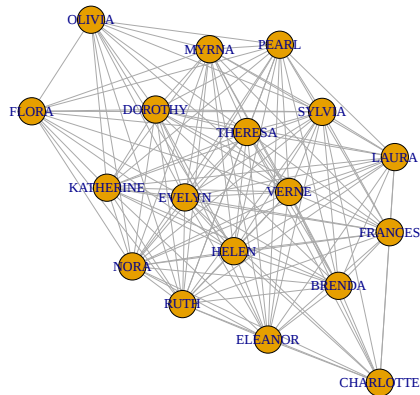
```
w <- graph_from_data_frame(women, directed = FALSE)
```

Každá hrana je tu dvakrát (pozrite si súbor so zoznamom hrán), napríklad 1-2 aj 2-1. Preto:

```
w <- simplify(w)
```

- ▶ Na stránke sú dostupné aj mená žien. Priradte ich k vrcholom siete a zobrazte získanú sieť.
- ▶ Nájdite najväčšiu kliku. Akú má interpretáciu pre túto konkrétnu sieť?

## Príklad 3: Davis' Southern Women Club



K-jadro (k-core)

## K-jadro

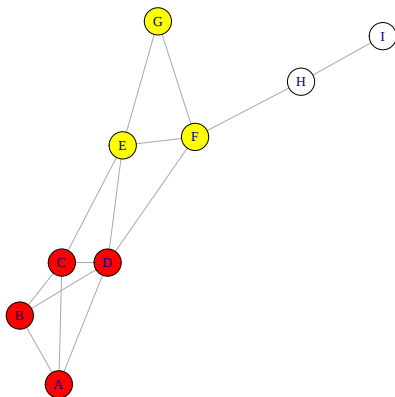
Maximálny podgraf, ktorého každý vrchol je spojený s aspoň  $K$  vrcholmi tohto podgrafu.

V R-ku:

- ▶ `coreness` - hodnota je  $k$ , ak vrchol je v  $k$ -jadre, ale nie je v  $(k+1)$ -jadre

# Príklad 1

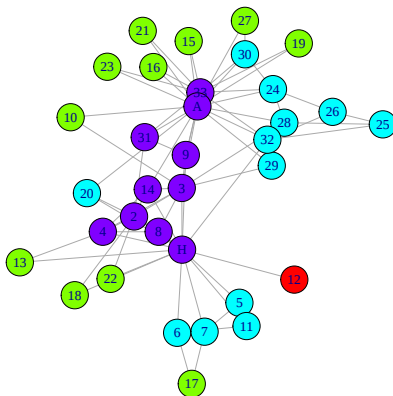
- Biele vrcholy majú hodnotu coreness 1, žlté 2 a červené 3
- 3-jadro: A, B, C, D; 2-jadro: A, B, C, D, E, F, G; 1-jadro: všetky



# Príklad

Vypočítajte hodnoty coreness pre sieť Zacharyho karate klubu.  
Farebne vyznačte vrcholy s rôznymi hodnotami.

*Poznámka:* Pri automatickom nastavovaní farieb sa môže zísť funkcia rainbow



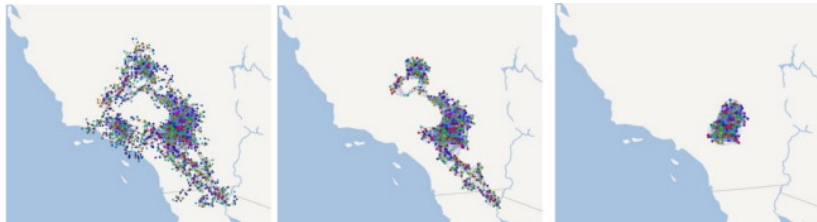
# Aplikácia 1

*He, X., Zhao, H., Cai, W., Li, G. G., & Pei, F. D. (2015). Analyzing the structure of earthquake network by k-core decomposition. Physica A: Statistical Mechanics and its Applications, 421, 34-43.*

https:

[//www.sciencedirect.com/science/article/pii/S037843711400973X](https://www.sciencedirect.com/science/article/pii/S037843711400973X)

1-, 6-, 12-jadro siete zemetrasení v Kalifornii



## Aplikácia 2

*Lordan, O., & Sallan, J. M. (2019). Core and critical cities of global region airport networks. Physica A: Statistical Mechanics and its Applications, 513, 724-733.*

https:

[//www.sciencedirect.com/science/article/pii/S0378437118310665](https://www.sciencedirect.com/science/article/pii/S0378437118310665)

Jadro s najvyšším koeficientom (vľavo), periféria (v strede, iba v 1-jadre), mosty (vpravo, ostatné vrcholy)



# Grafické znázornenie

Upravte funkciu na blogu <https://jcasasr.wordpress.com/2015/02/03/plotting-the-coreness-of-a-network-with-r-and-igraph/>, potom môžete spraviť napr. `plot(g, layout=CorenessLayout(g))` (+ ostatné parametre)

