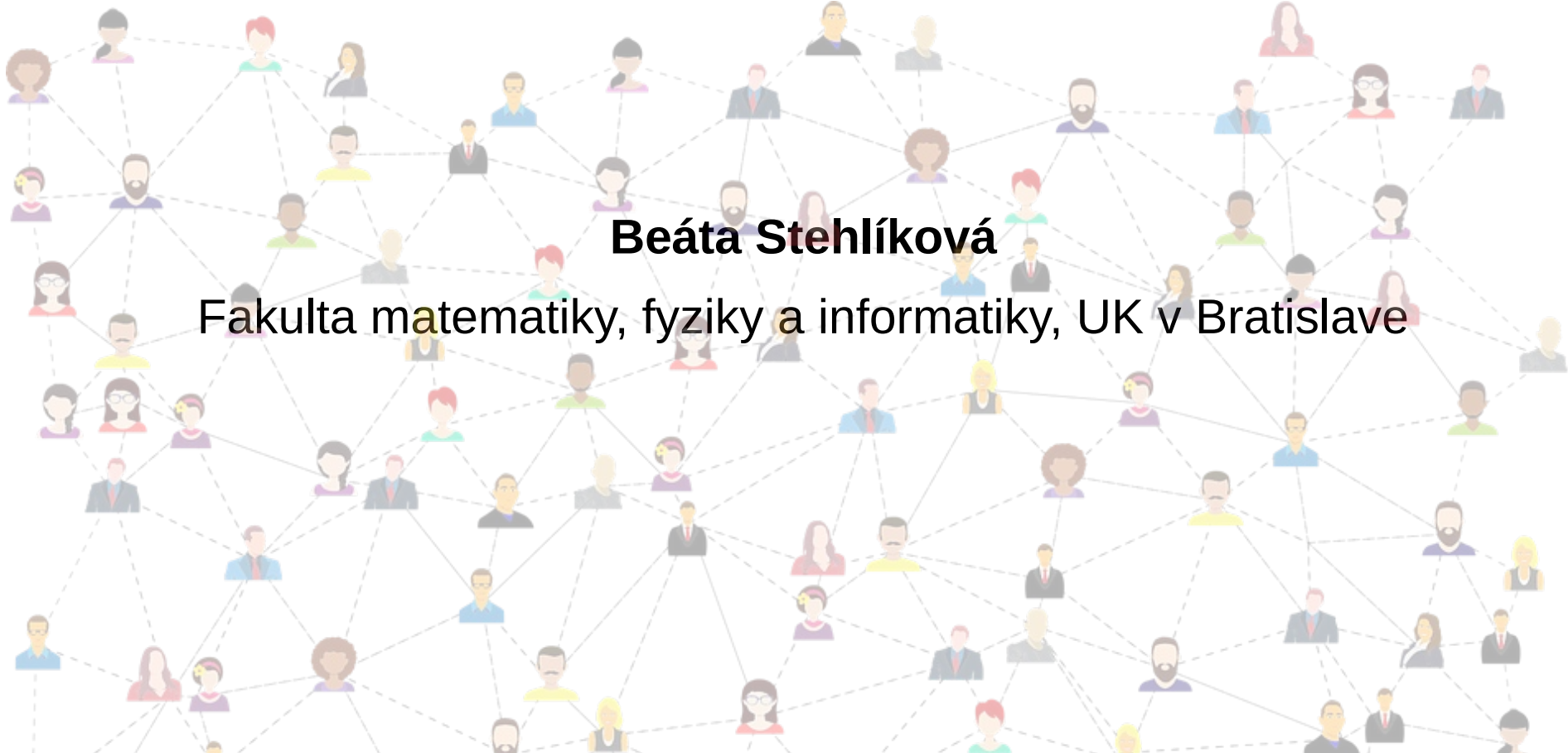


# Analýza sociálnych sietí – ukážky



**Beáta Stehlíková**

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

# Čo sú sociálne siete



- A a B sú priatelia na Facebooku
- A sleduje B na Instagrame

## **(Sociálna) sieť**

- vrcholy siete
- niektoré sú pospájané hranami

# Príklad: Články na Wikipédii

Článok:

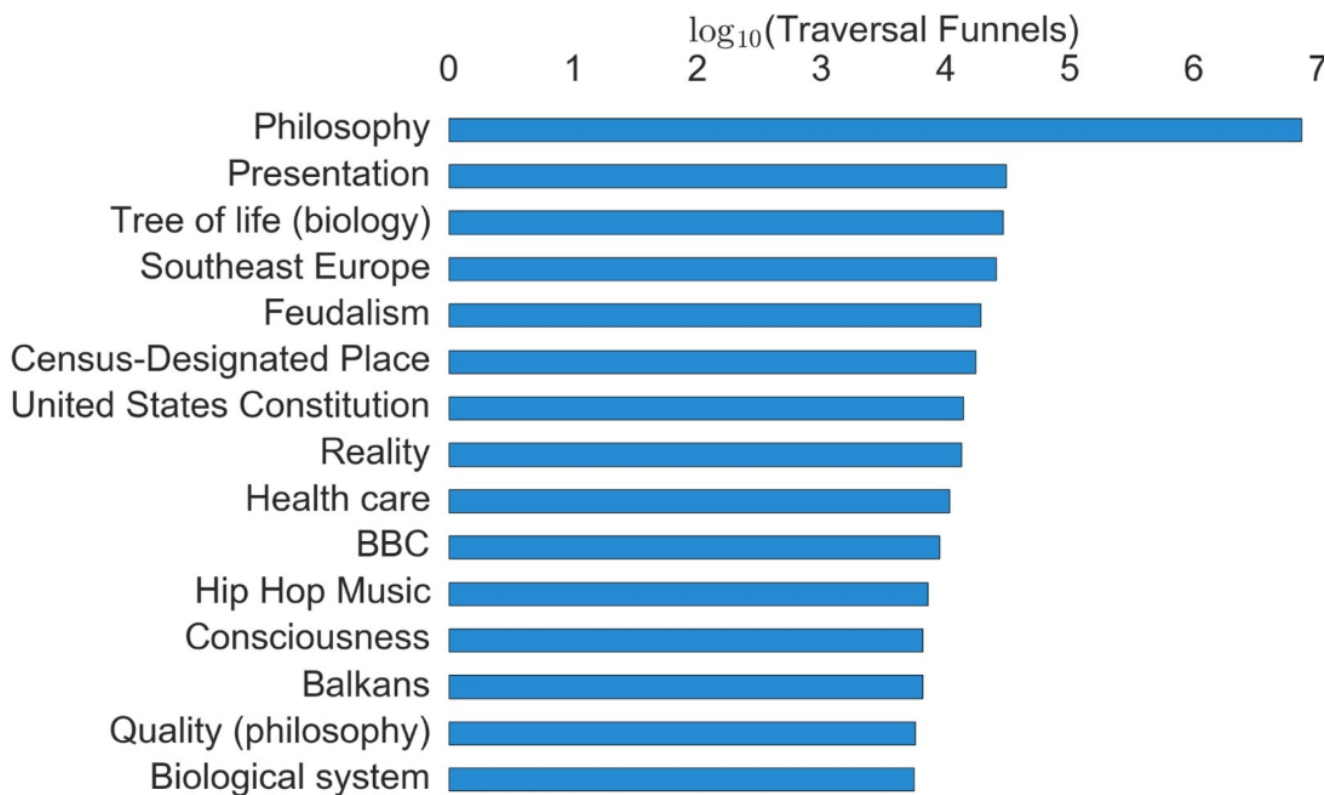
Ibrahim, M., Danforth, C. M., & Dodds, P. S. (2017). **Connecting every bit of knowledge: The structure of Wikipedia's First Link Network.** Journal of Computational Science, 19, 21-30.

Z abstraktu:

By following the first link in each article, we algorithmically construct a directed network of all 4.7 million articles: Wikipedia's First Link Network.



# Príklad: Články na Wikipédii



# Študentský projekt: Hlasovanie poslancov

- Vrcholy siete: poslanci
- Hrana: ak v niektorom hlasovaní hlasovali rovnako
- Váha hrany: počet rovnakých hlasovaní



# Študentský projekt: Hlasovanie poslancov

- Súčasť práce: naprogramovanie získania týchto informácií z webu [www.nrsr.sk](http://www.nrsr.sk)

Schôdza

[Schôdza č. 55](#)

Dátum a čas

1. 2. 2022 13:30

Názov hlasovania

Hlasovanie o programe 55. schôdze NR SR ako o celku so schválenými zmenami a

Výsledok hlasovania

Návrh prešiel

Prítomní

126

[?] Zdržalo sa hlasovania

12

Hlasujúcich

126

[N] Nehlasovalo

0

[Z] Za hlasovalo

86

[0] Neprítomní

24

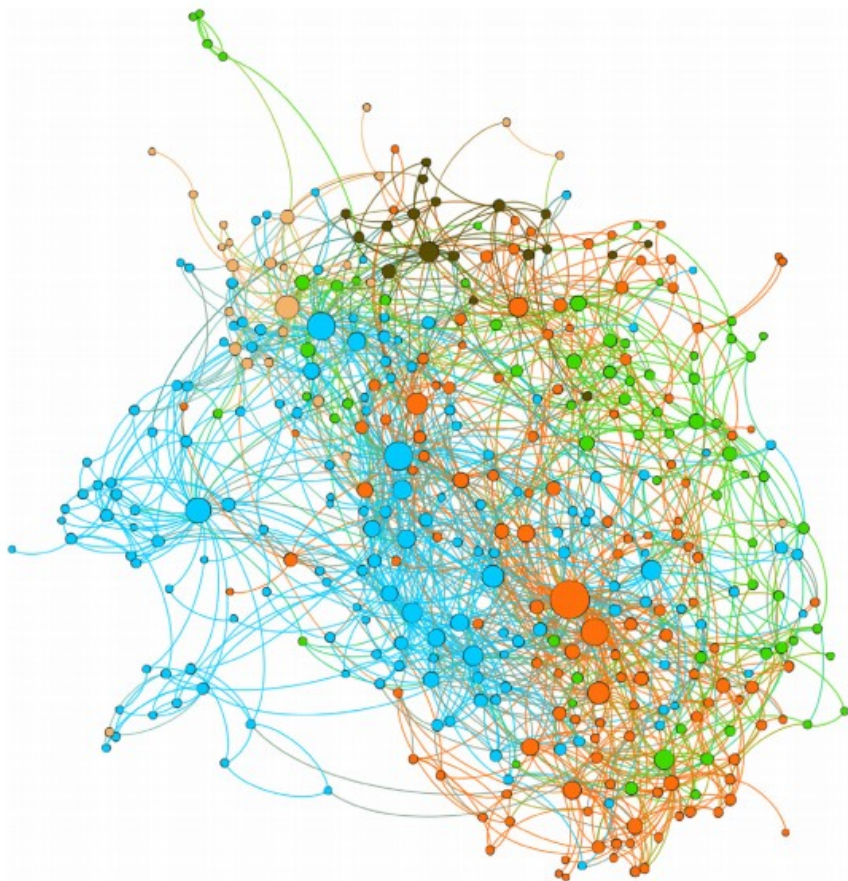
Klub OĽANO

[0] [Andrejuvová, Anna](#)

[Z] [Brisudová, Martina](#)

[Z] [Čekovský, K](#)

# Príklad: Zločinecké organizácie



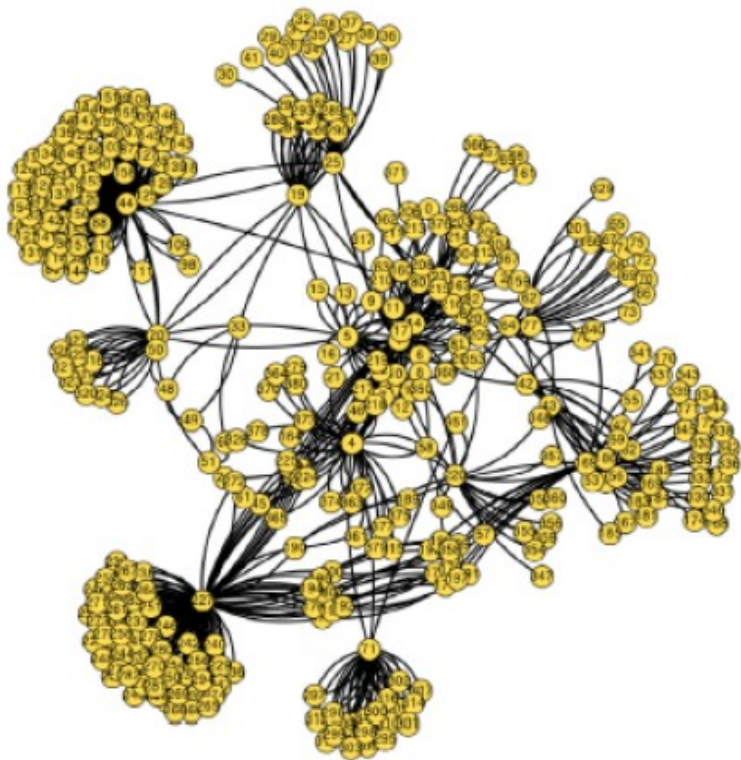
*Článok:*

DellaPosta, D. (2017). **Network closure and integration in the mid-20th century American mafia.** Social Networks, 51, 148-157

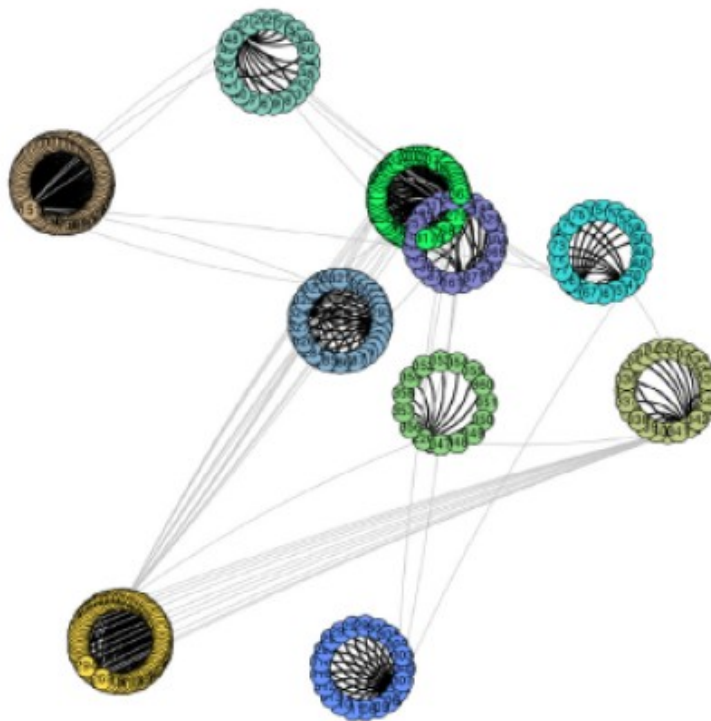
*Popis obrázku:*

The Five Families of New York City.  
Note: Nodes are colored according to family membership and sized according to degree.

# Príklad: Komunikácia, zhukovanie



(a) Phone call network of 148 nodes and 210 edges.



(b) Clusters detected after 46 edges deleted.

Článok:

Ferrara, E., De Meo, P., Catanese, S., & Fiumara, G. (2014). **Detecting criminal organizations in mobile phone networks.** Expert Systems with Applications, 41(13), 5733-5750.

# Zacharyho karate klub

*Článok:*

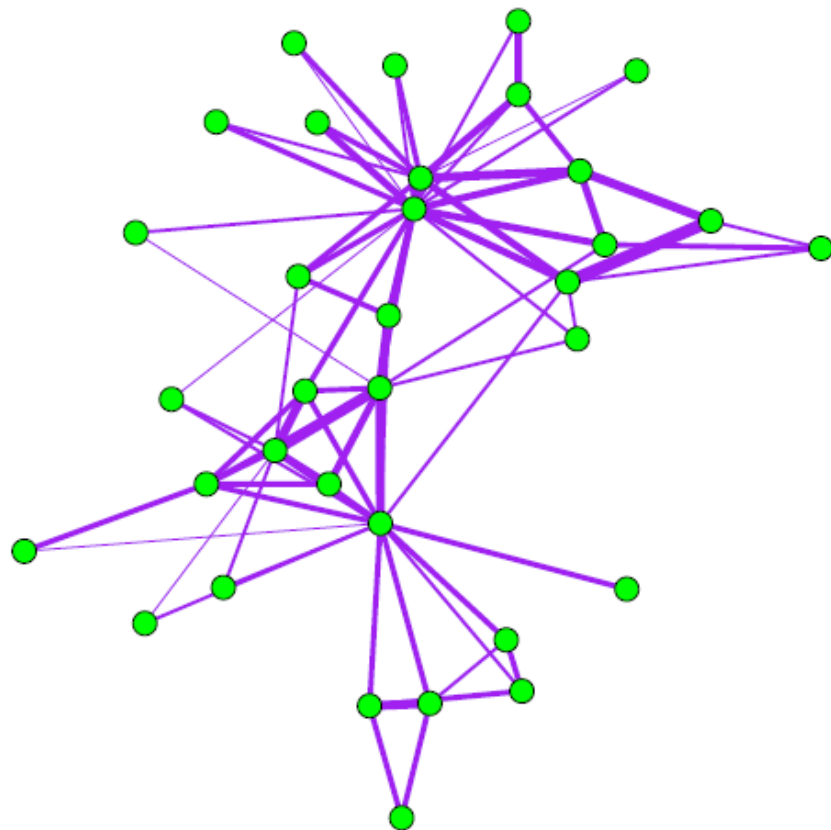
Zachary, W. W. (1977). **An Information Flow Model for Conflict and Fission in Small Groups.** Journal of Anthropological Research, 33, 452-473.

*Zobrazená sieť:*

Členovia klubu, hrany znázorňujú kontakty medzi nimi, hrúbka hrany je intenzita kontaktov

*Čo sa stalo:*

Klub sa rozpadol na dva. Vedeli by sme predpovedať, ako?



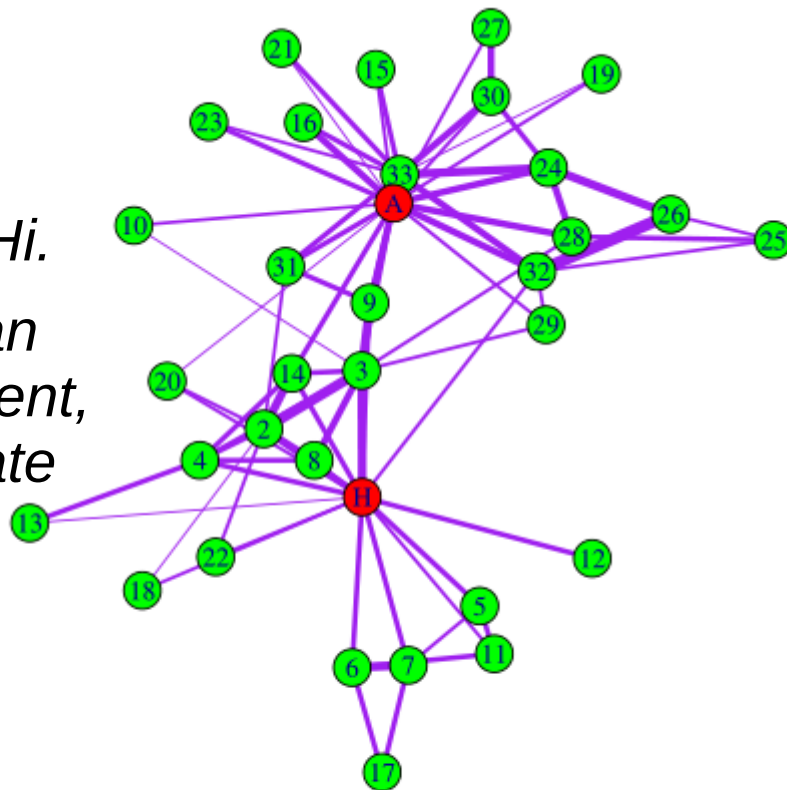
# Putovná cena

Kto ako prvý spomenie na konferencii vo svojej prezentácii Zacharyho karate klub?



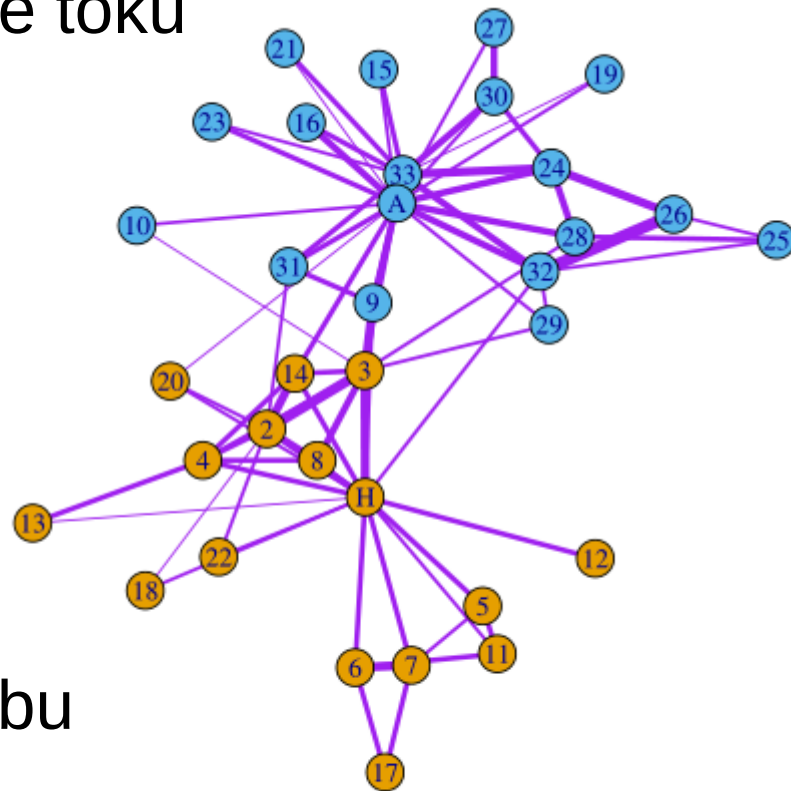
# Čo sa dialo v karate klube

- Čo písal Zachary o klube:
  - *The club employed a part-time karate instructor, who will be referred to as Mr. Hi.*
  - *At the beginning of the study there was an incipient conflict between the club president, John A., and Mr. Hi over the price of karate lessons.*
  - *As time passed the entire club became divided over this issue*



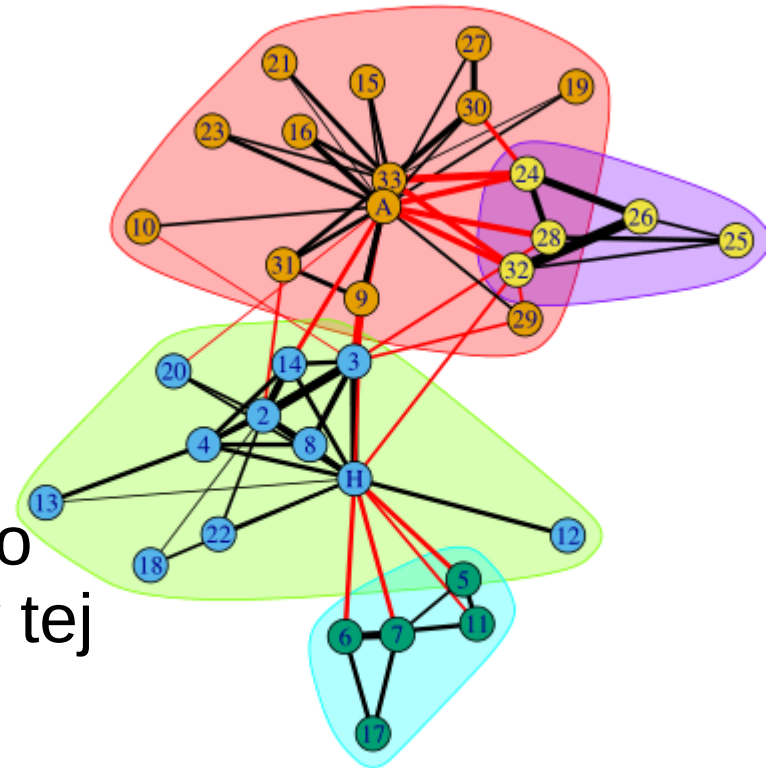
# Karate klub: predpoveď a realita

- Zacharyho predpoveď na základe toku informácií:
- V porovnaní s realitou zaradil nesprávne len jedného člena klubu



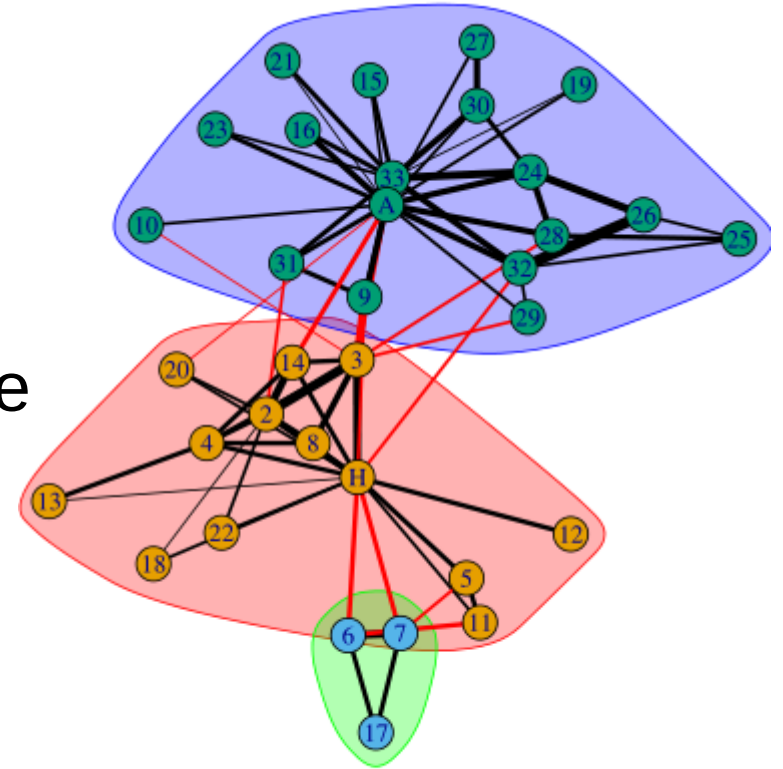
# Zhlukovanie / hľadanie komúní: algoritmy

- Existuje veľa rôznych algoritmov
- Napríklad „**walktrap**“ algoritmus
  - *walk* – prechádzka
  - *trap* – pasca
  - pri krátkej náhodnej prechádzke po hranách máme tendenciu zostať v tej istej komunite



# Zhlukovanie / hľadanie komúní: algoritmy

- Iný príklad: „**label propagation**“ algoritmus
  - *label* – štítok, značka
  - *propagation* – šírenie, prenášanie
  - vrcholy dostanú značky a postupne sa budú meniť podľa toho, aké značky má väčšina ich susedov



# Zhlukovanie / hľadanie komúní: algoritmy

- Veľa rôznych algoritmov **a stále vznikajú nové**

- Len na ukážku:

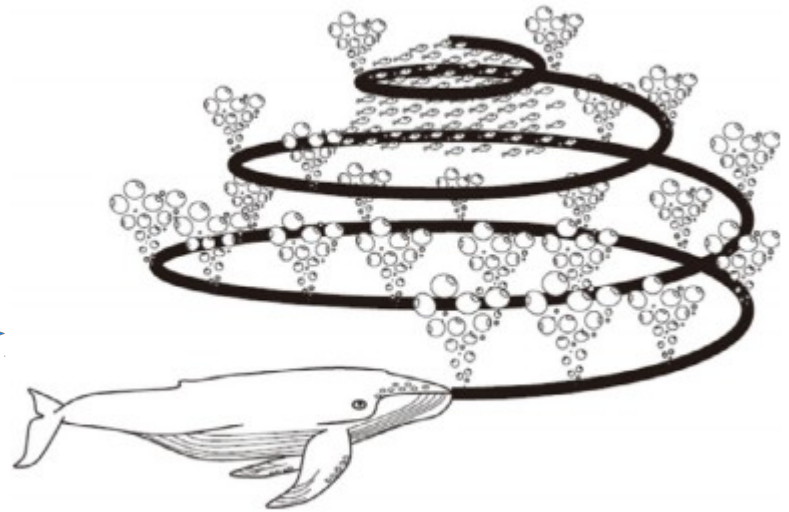
Zhang, Y., Liu, Y., Li, J., Zhu, J., Yang, C., Yang, W., & Wen, C. (2020).

**WOCDA: A whale optimization based community detection algorithm.**

Physica A: Statistical Mechanics and its Applications 539, 122937.



Y. Zhang, Y. Liu, J. Li et al. / Physica A 539 (2020) 122937

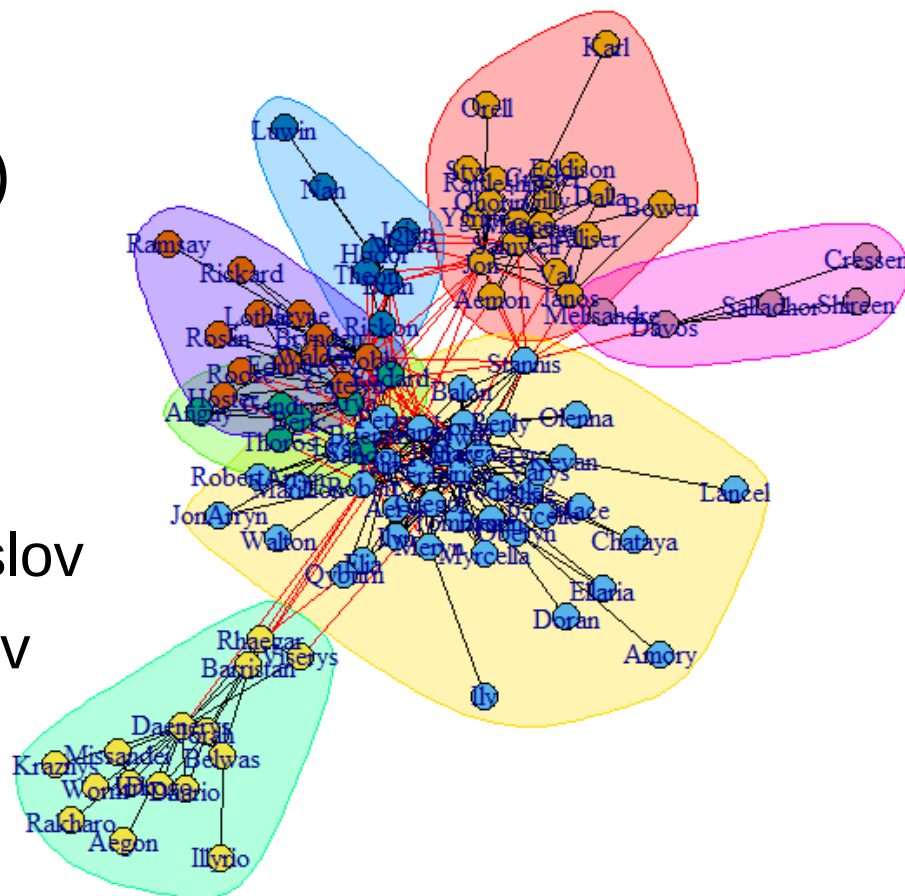


**Fig. 1.** Bubble-net feeding behavior of humpback whales [19].

# Študentský projekt: Hra o tróny



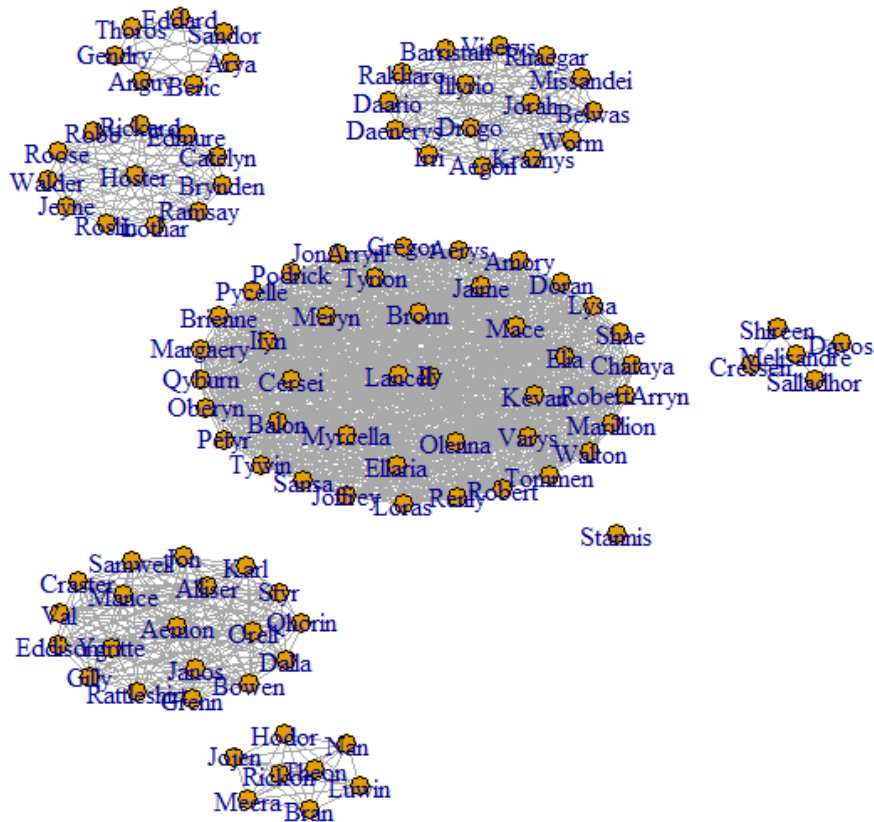
- Knižná séria (George R. R. Martin) a televízny seriál HBO
- Analýza knižnej predlohy 3. série
  - 108 postáv
  - hrana - výskyt mien v rozmedzí 15 slov
  - váha hrany - počet takýchto výskytov
- Rôzne algoritmy zhlukovania



# Študentský projekt: Hra o tróny

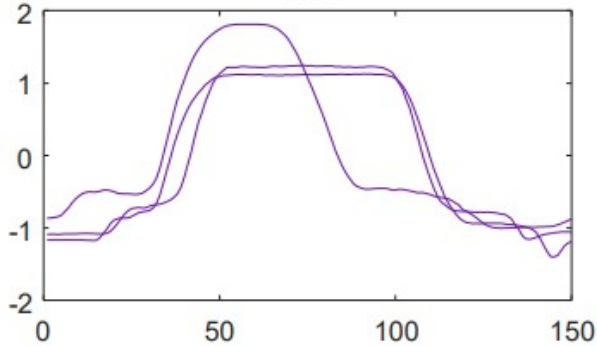


- Algoritmy dávajú podobné, ale nie úplne rovnaké výsledky
- Použijeme 5 algoritmov a vytvoríme novú sieť:
  - Vrcholy – ako predtým, postavy
  - Hrana – ak boli zaradení do spoločného zhluku aspoň 4 krát

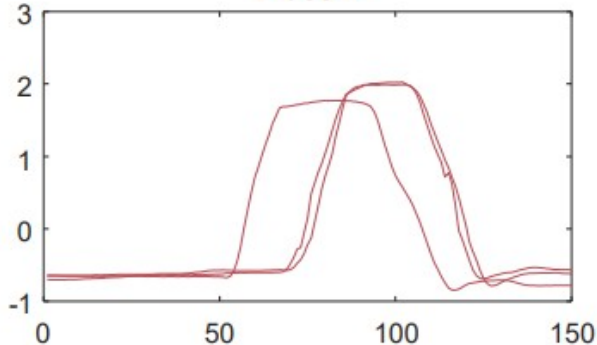


# Študentský projekt & diplomovka: Zhlukovanie časových radov

Class 1



Class 2

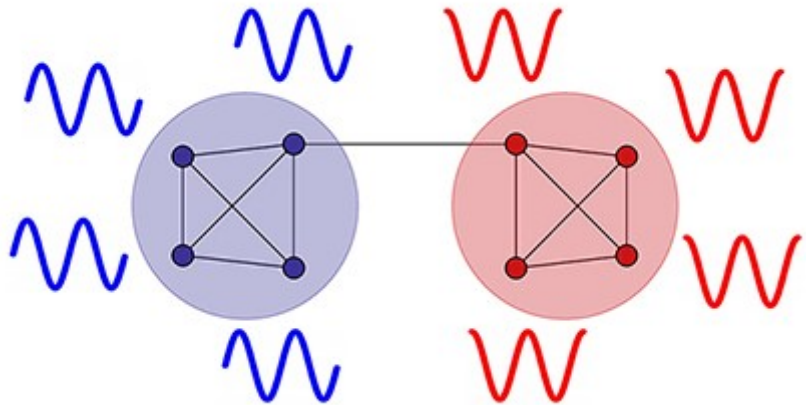


Left) GunPoint recording of 2003, right) GunPoint recording of 2018.  
Top) Ann Ratanamahatana, bottom) Eamonn Keogh.  
The female and male actors are the same individuals recorded fifteen years apart.

- Ukážka aplikácie: *Gun or Point?*
- Zdroj dát: *UCR Time Series Classification Archive*

# Študentský projekt & diplomovka: Zhlukovanie časových radov

- Vrchol: časový rad
- Hrana: ? (treba sformulovať kritérium)
- Váha hrany: ? (nejaká miera „vzdialenosti časových radov“)



Ferreira, L. N., & Zhao, L. (2016). **Time series clustering via community detection in networks**. Information Sciences, 326, 227-242.

# Článok len nedávno prijatý na publikovanie

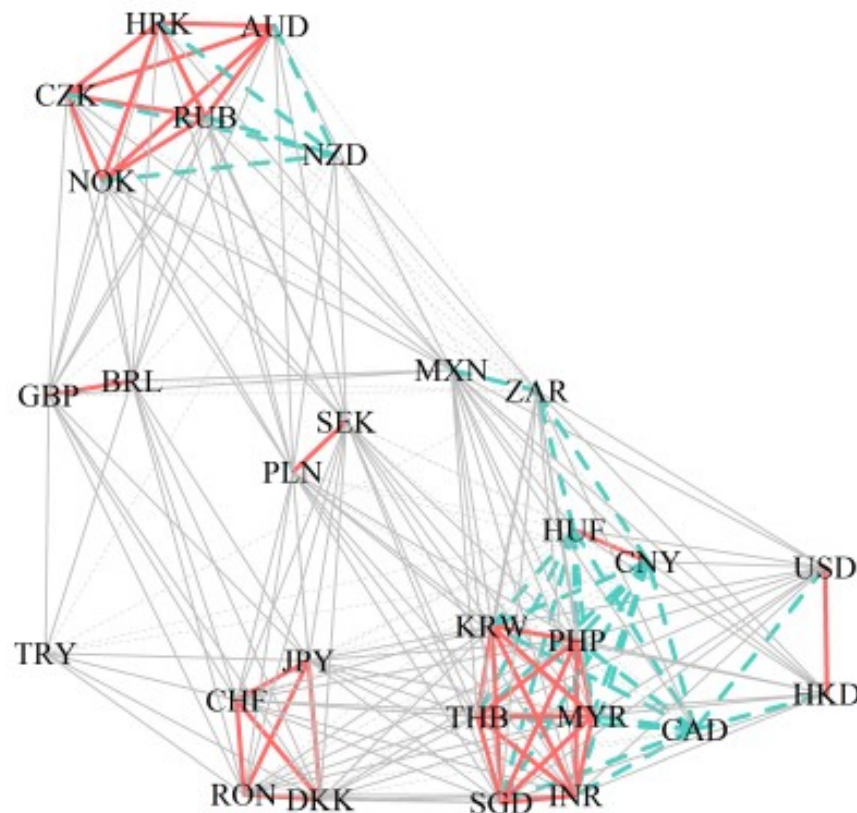
## Time series clustering based on time-varying Hurst exponent

Alex Babiš<sup>1</sup>

Beáta Stehlíková<sup>2</sup>

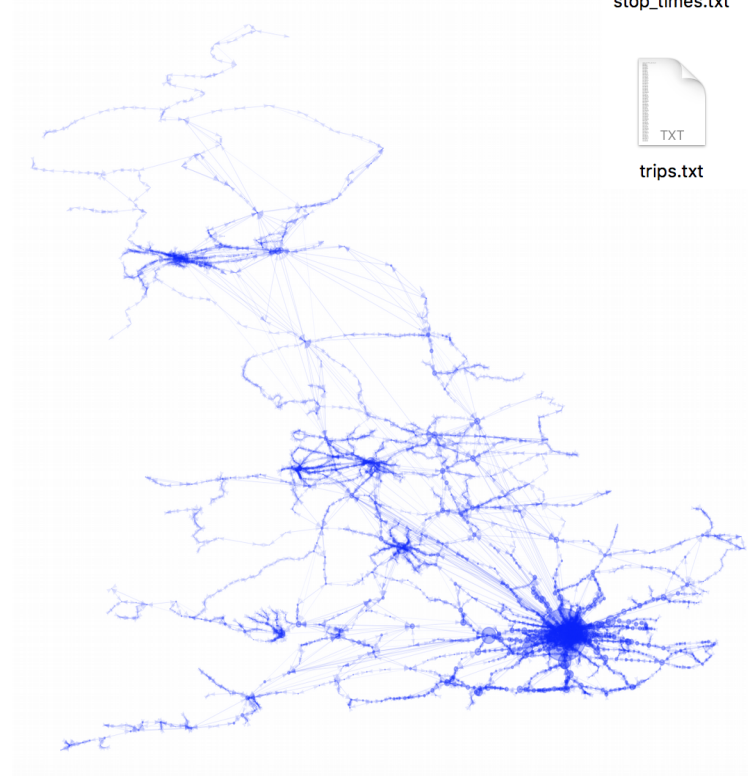
### Abstract

We consider the problem of clustering time series which are assumed to possess the long term memory. We propose an approach based on combining the results obtained by applying different methods for estimating time-varying Hurst exponent and apply it to Euro exchange rates. Firstly, we fit AR-GARCH models to every time series to reduce bias of rescaled range analysis method. We only consider model with residuals, in which no autocorrelation and ARCH effect is present; among them we choose the model with the lowest value of the Bayesian information criterion. Afterwards, we estimate the Hurst exponent from the residuals by means of the rolling window approach using four different estimation methods. Vectors of Hurst exponents are clustered for each of the four cases and the clusters are compared in order to obtain the final clustering.



# Študentský projekt: Železničná sieť

- Ktoré stanice v sieti sú významné / dôležité / centrálné / ... ?
- A ako to vlastne merať?



agency.txt



calendar.txt



routes.txt



stop\_times.txt



stops.txt

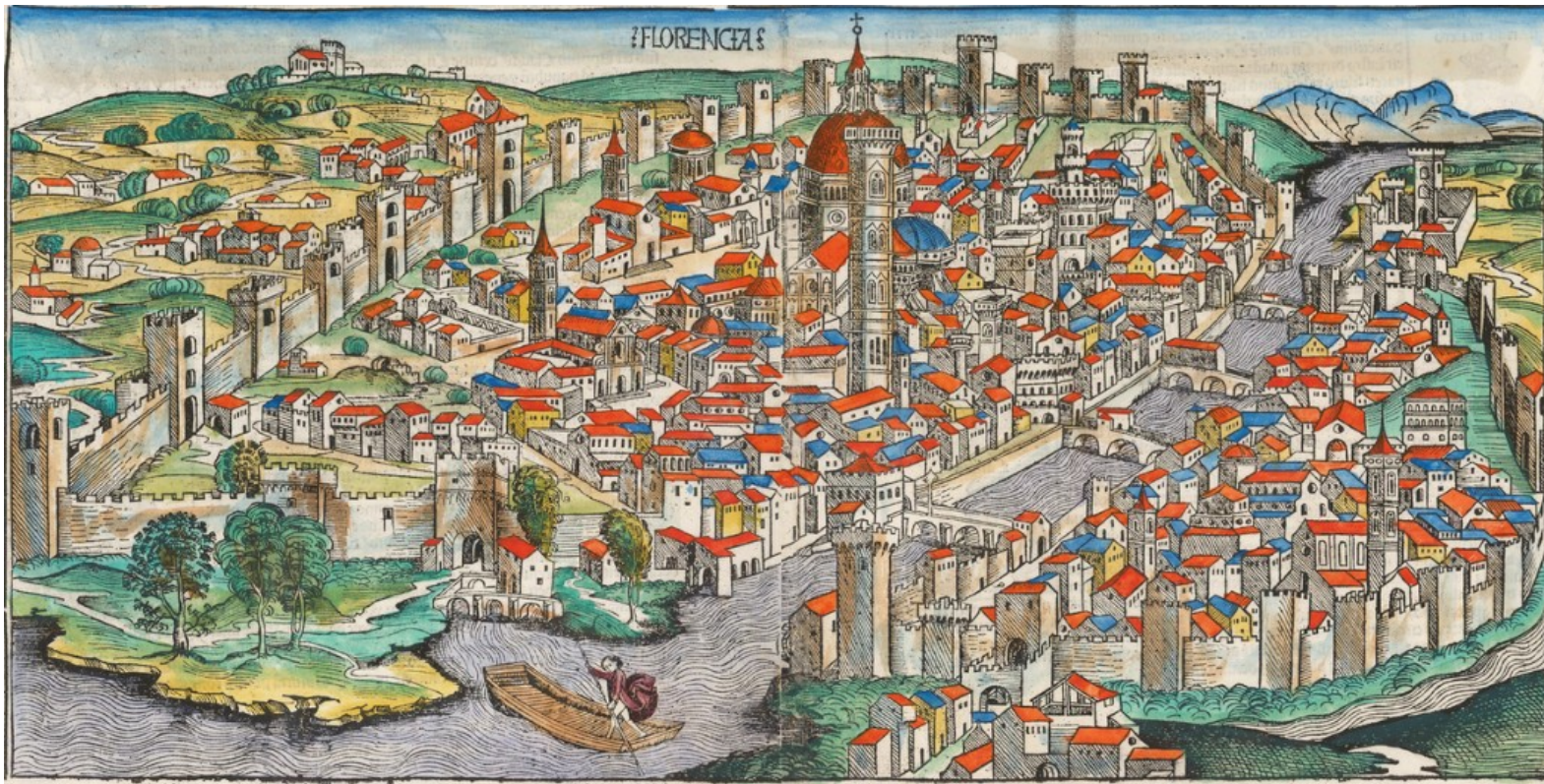


transfers.txt



trips.txt

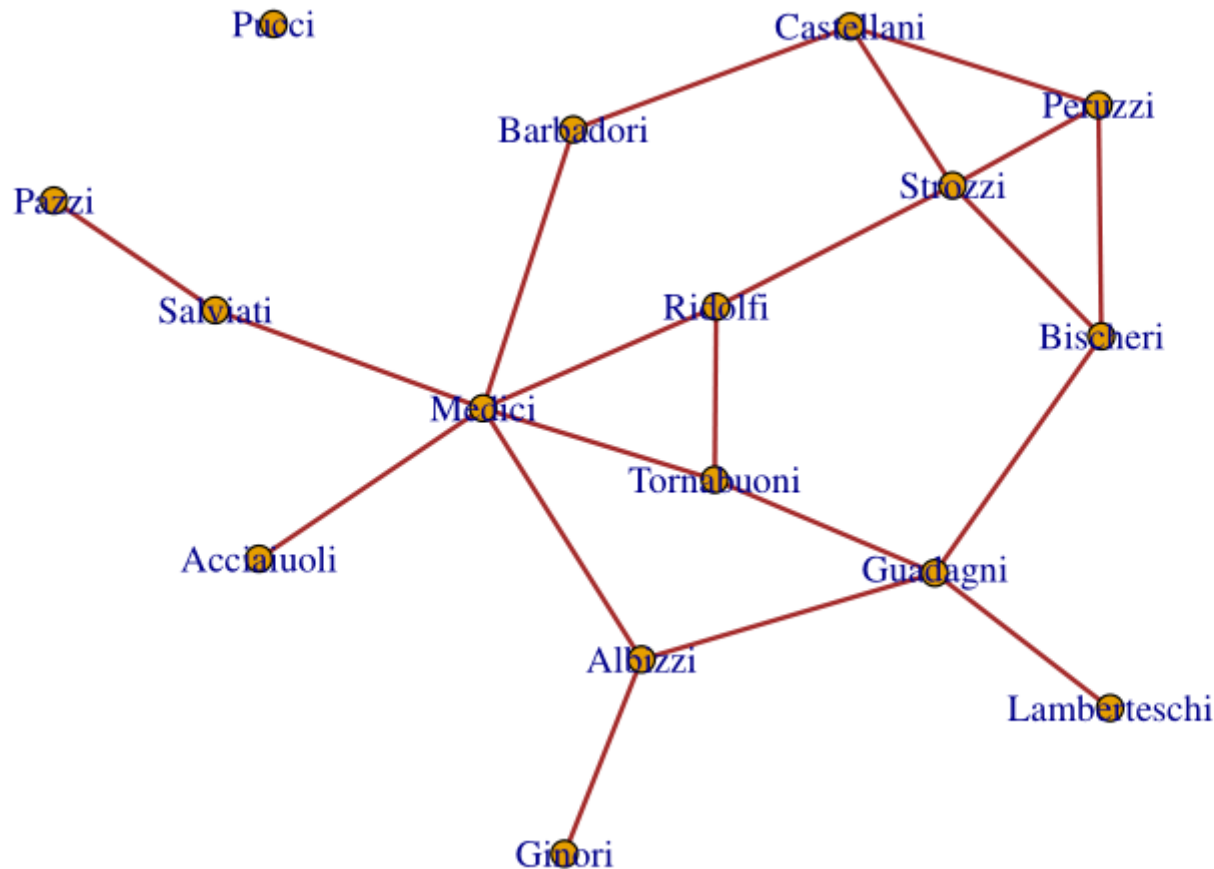
# Renesančná Florencia



Padgett, John F. (1994). **Marriage and Elite Structure in Renaissance Florence, 1282-1500**. Paper delivered to the Social Science History Association.

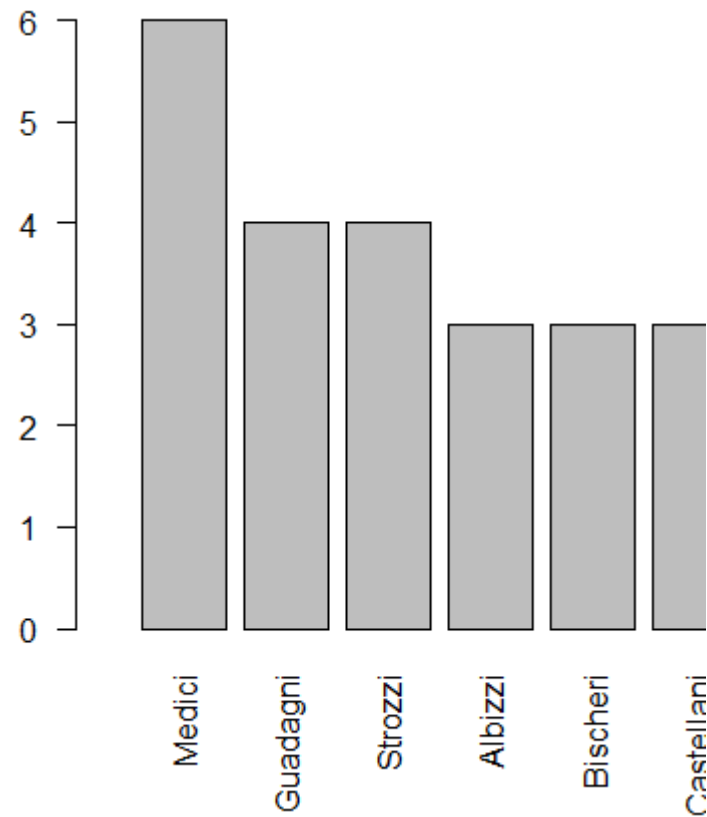
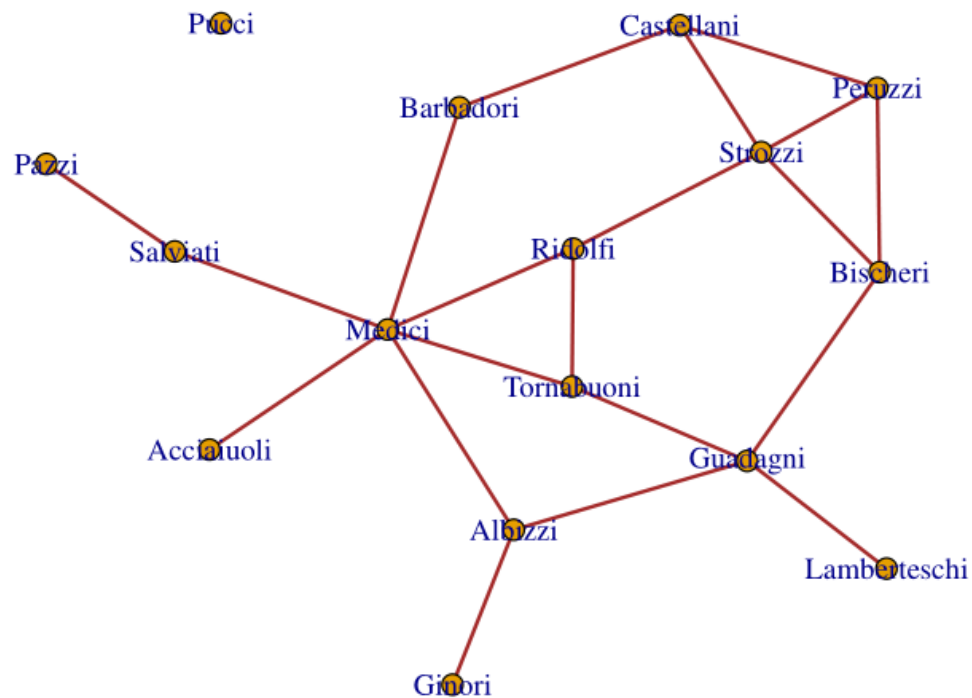
# Renesančná Florencia

- Manželstvá medzi členmi významných rodín
- Ktorá rodina má v tejto sieti „centrálne postavenie“?
- Čo to vlastne znamená „**centrálne postavenie**“? Ako ho merať?



# Miery centrality: centralita stupňa

**S koľkými vrcholmi daný vrchol susedí?**



# Miery centrality: centralita blízkosti

Nepriamo úmerná  
súčtu vzdialeností  
k ostatným  
vrcholom

$vzdialenosť = 4$



Salviati

$vzdialenosť = 3$

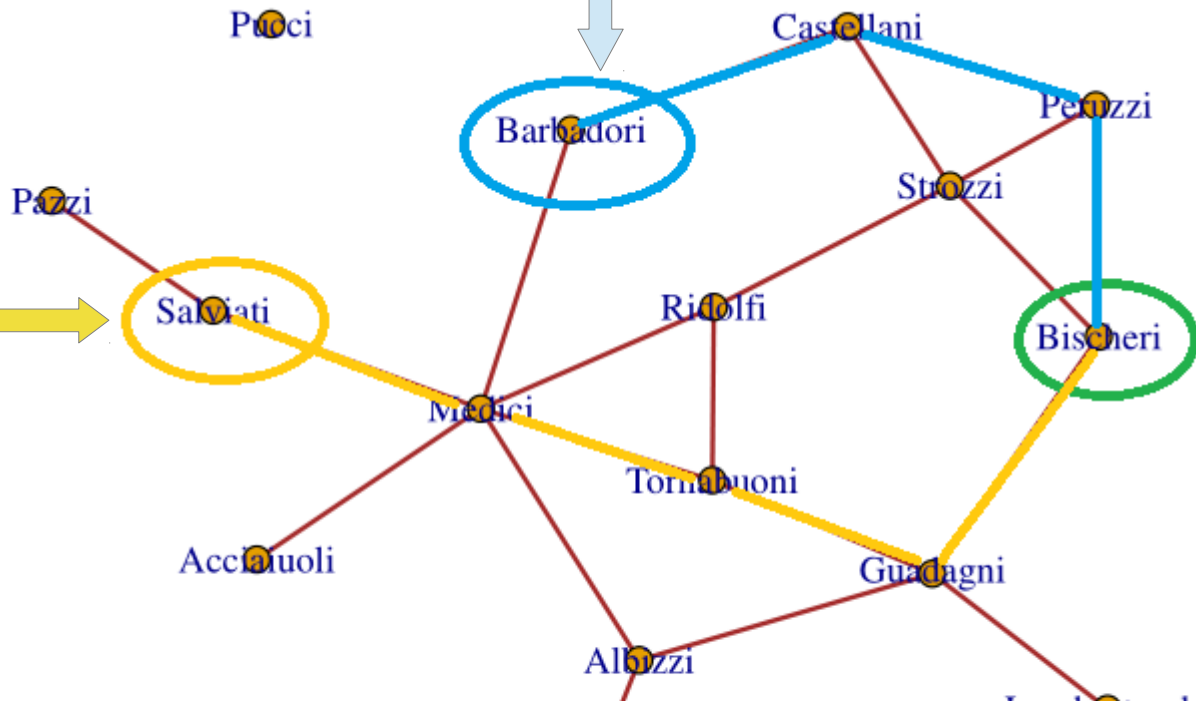


Barbadori

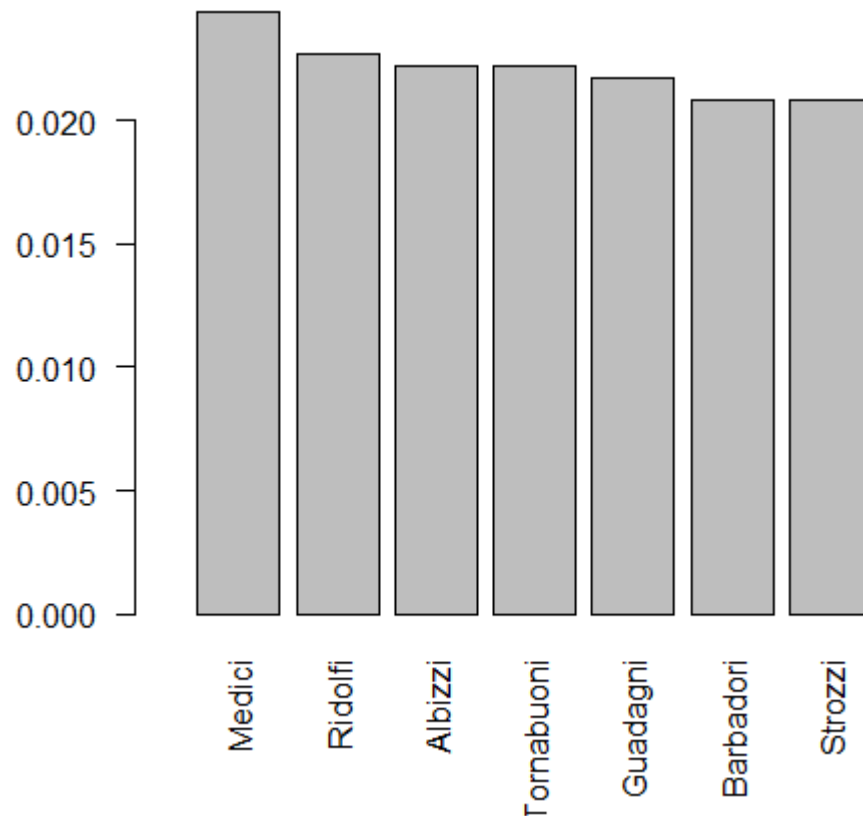
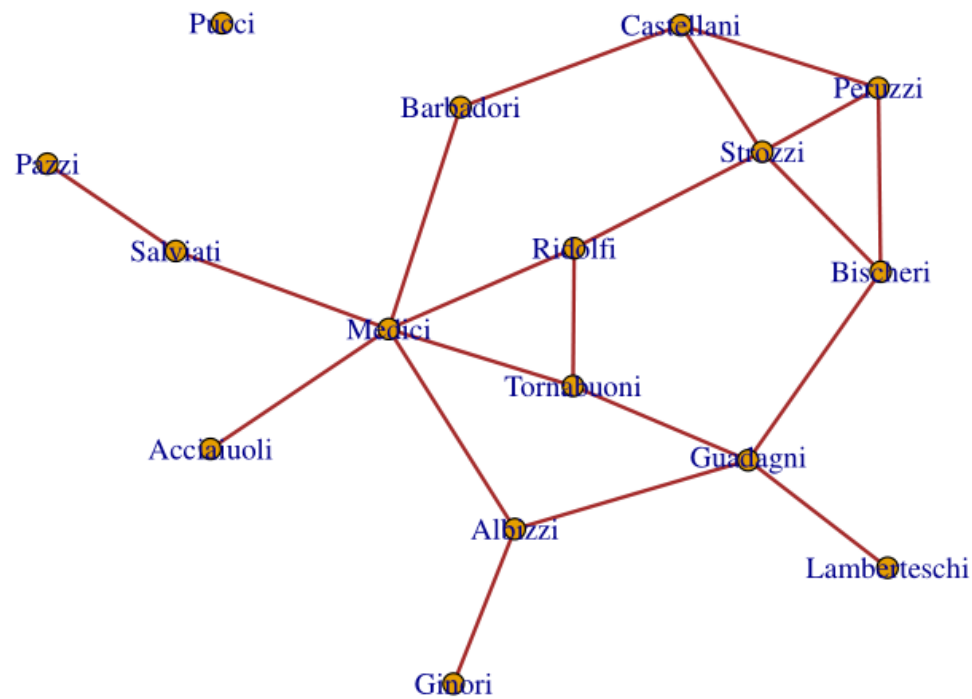
Ako sú  
ďaleko od  
ostatných  
rodín?



Bischeri

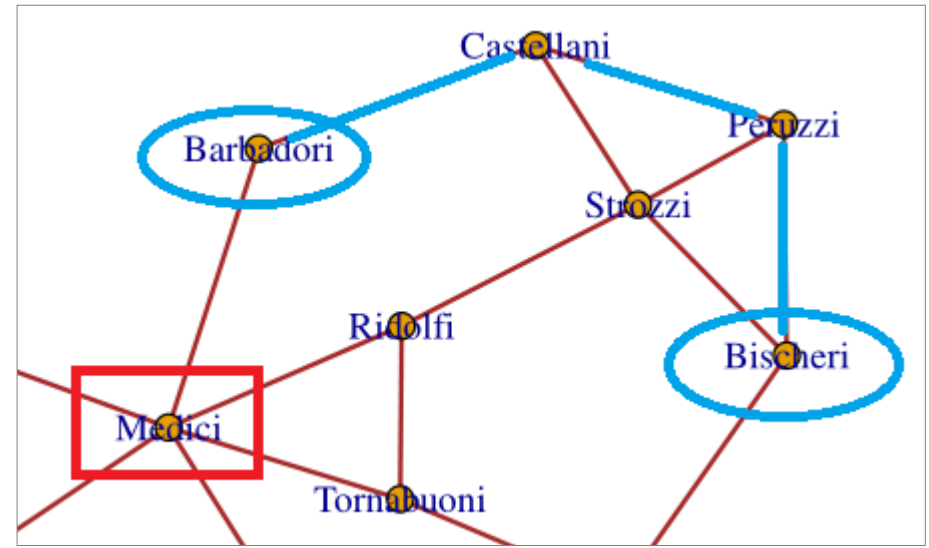
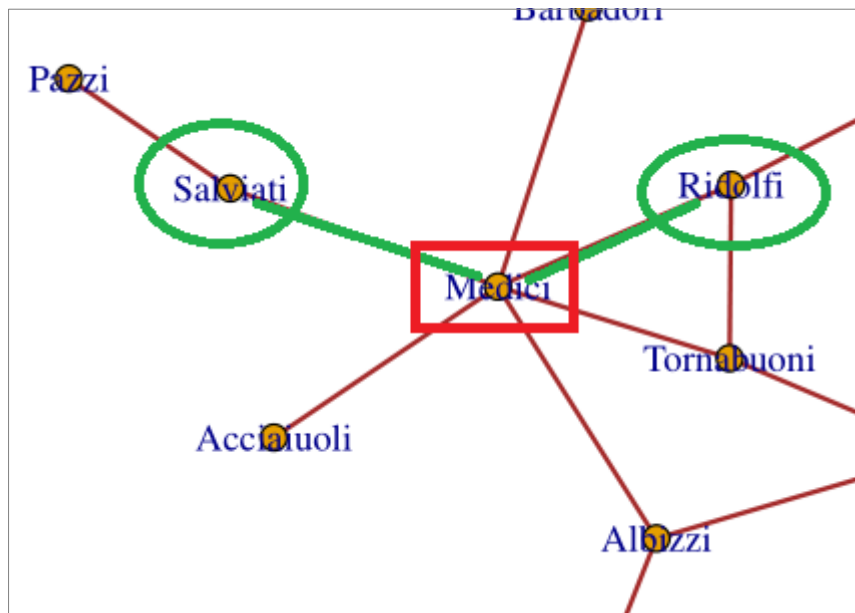


# Miery centrality: centralita blízkości

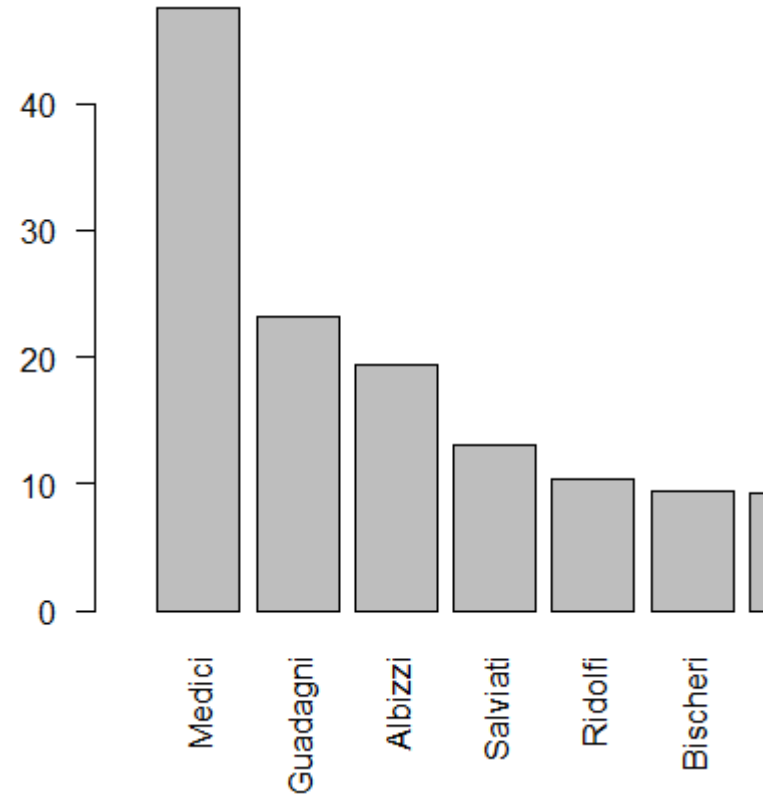
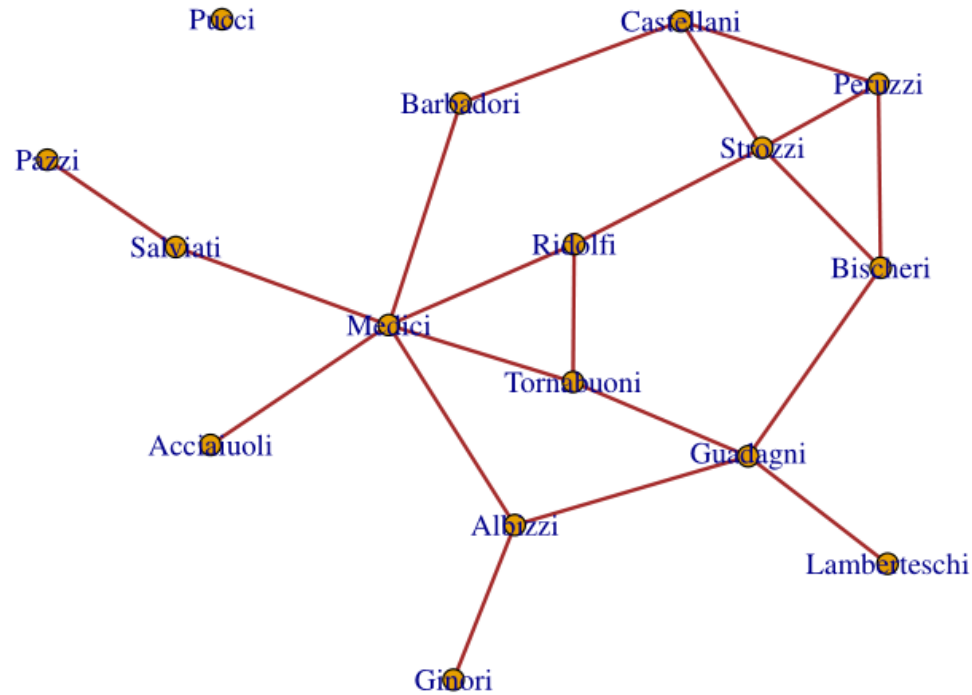


# Miery centrality: centralita medzipolohy

Ako často sa vrchol vyskytuje v najkratších cestách medzi inými vrcholmi?



# Miery centrality: centralita medzipolohy

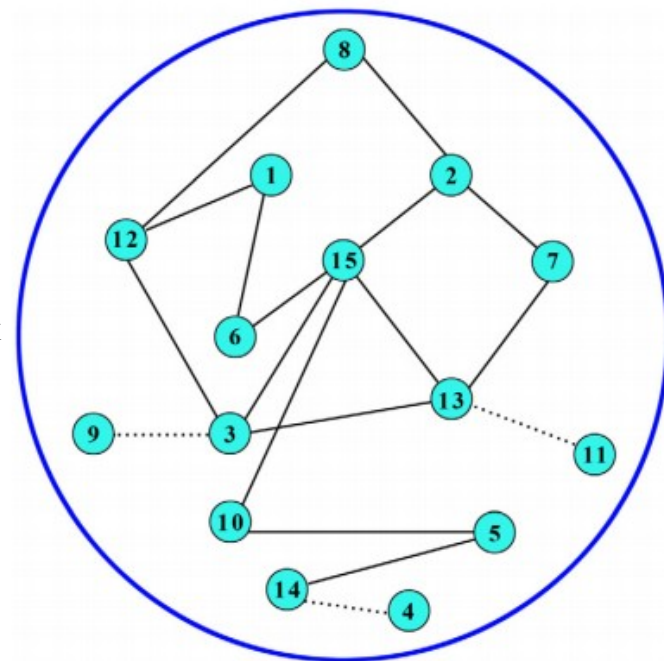


# Miery centrality: FAQ

- **Sú aj iné miery centrality ako tieto tri?**

Áno a stále vznikajú nové. (Túto konkrétnu budeme aj programovať.)

Yang, F. et al. (2018). **Ranking the spreading influence of nodes in complex networks: An extended weighted degree centrality based on a remaining minimum degree decomposition.** Physics Letters A, 382(34), 2361-2371.



- ***Ktorú centralitu teda použiť?***

Závisí od okolností...

(b)  $md = 1$ ,  $CurrentIter = 1$ , delete node 4, 9 and 11. The RMD index of these nodes is 1.

# Miery centrality: FAQ

- Ktorá centralita zodpovedá intuícii o centrálnej postave filmu, knihy, ... (napr. *Hra o tróny* z projektu)?

Diplomovka:

*Centralita vrcholov v sociálnych sieťach a jej subjektívne vnímanie*

- analyzoval sa jeden z filmov *Star Wars*



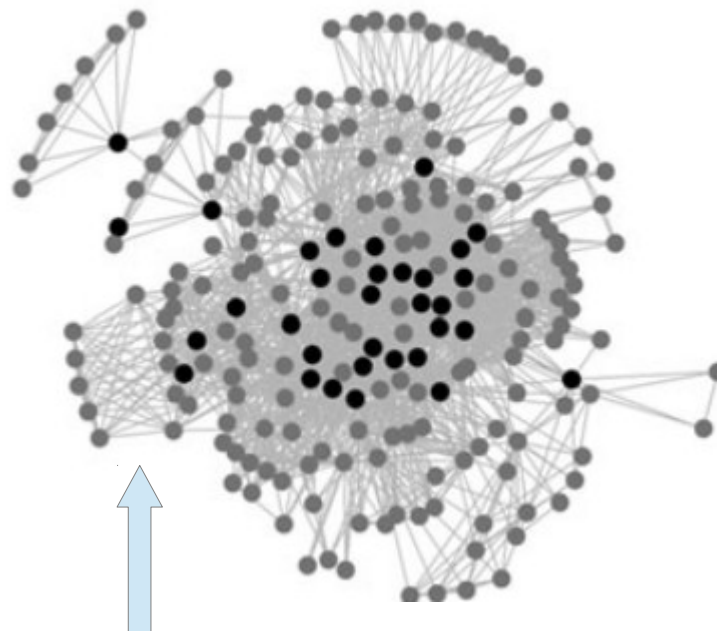
# Miery centrality: FAQ

- **Je niekedy „dôležitý vrchol v sieti“ definovaný jasnejšie?**

Áno, napríklad mafia má hierarchickú štruktúru a môžeme sa snažiť identifikovať *bossov*.

- **A nedali by sa rôzne centrality navzájom skombinovať?**

Dali... aj v prípade tej mafie.



Calderoni, F. (2014). **Identifying mafia bosses from meeting attendance**. In Networks and network analysis for defence and security (pp. 27-48). Springer, Cham.

# Študentský projekt: Mafiáni a Bedári



chasing

adopted

lost



mother

raised

grew up together

in love



parents

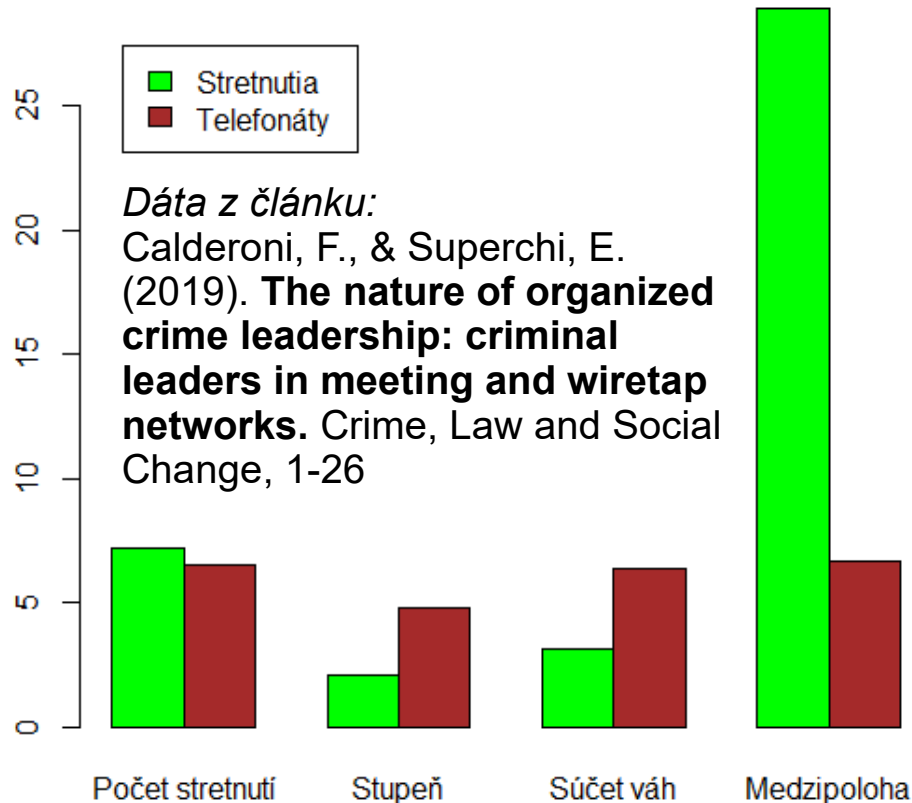
friends

friends

- Známe postupy z iných aplikácií:
  - porovnávanie dôležitosti kritérií (subjektívne, ale dá sa skontrolovať konzistenciu)
  - porovnávanie s ideálnym riešením
- Kritériá – centrality, ideálne riešenie – centrálny vrchol



# Študentský projekt, následne spoločný výskum: Bossovia mafie



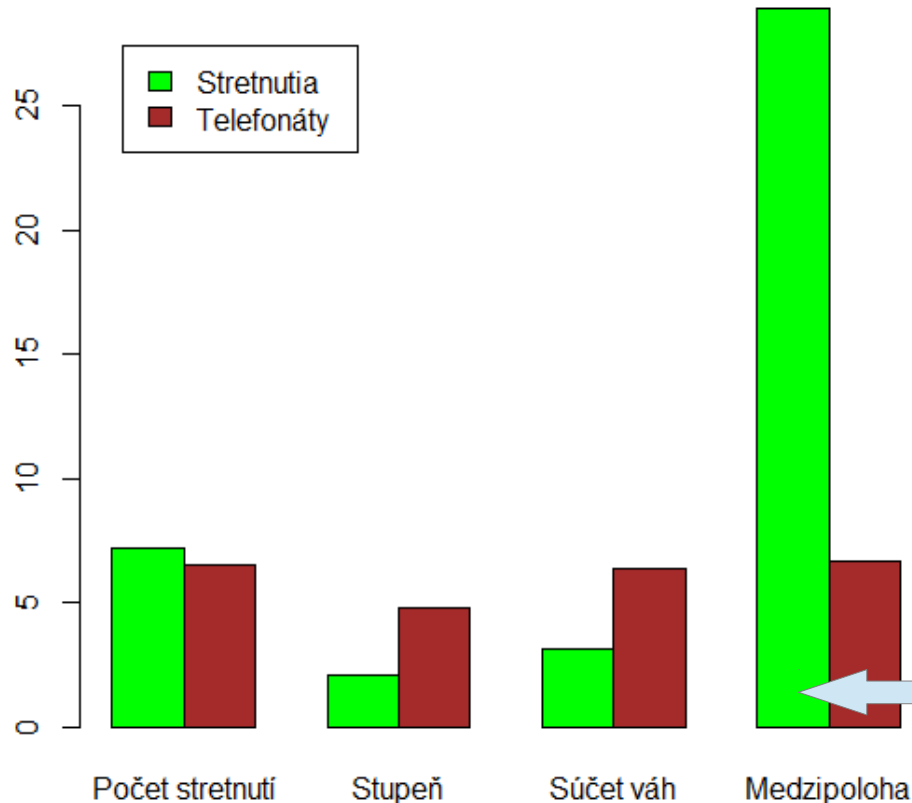
- **Calderoni (2014)**

- Koľkokrát je vyššia centralita jednotlivých typov u bossov ako u ostatných
- Veľmi dôležitá je centralita medzipolohy
- Výber charakteristík, konštrukcia skóre, z toho pravdepodobnosť, že ide o bossa

- **Náš prístup (spisujeme článok)**

- Na základe Calderoniho výsledkov porovnanie vplyvu centralít a iných charakteristík (už nie subjektívne)
- Aký je „ideálny boss“ a nakoľko sa podozriví k tomu blížia

# Bossovia mafie – aktuálne vedená bakalárka, jedna z aplikácií



- **Dáta:**

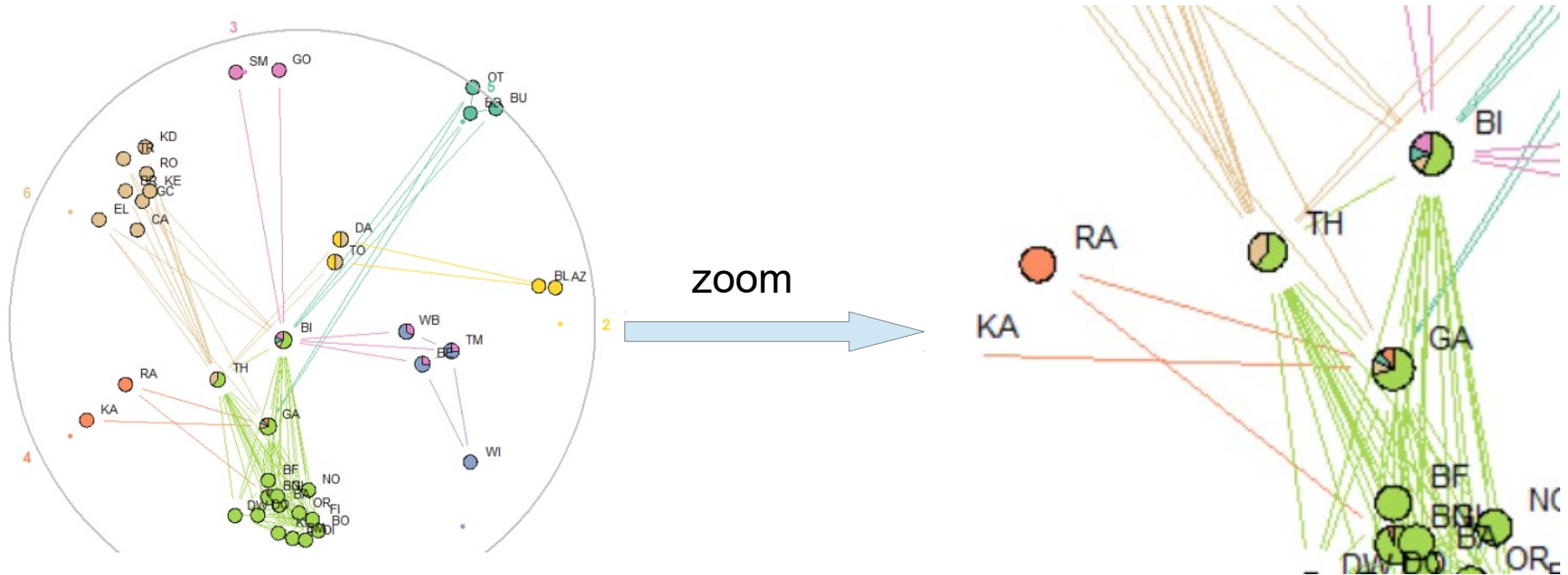
- Nielen, či sa daní ľudia niekedy stretli
- Ale aj účasť na jednotlivých stretnutiach

- Základná myšlienka centrality **BiRank**:

- Dôležití ľudia sa zúčastňovali na dôležitých stretnutiach
- Dôležité stretnutia sú tie, na ktorých boli dôležití ľudia
- Môžeme zahrnúť aj známu informáciu (napríklad že u ľudí je dôležitá centralita medzipolohy)

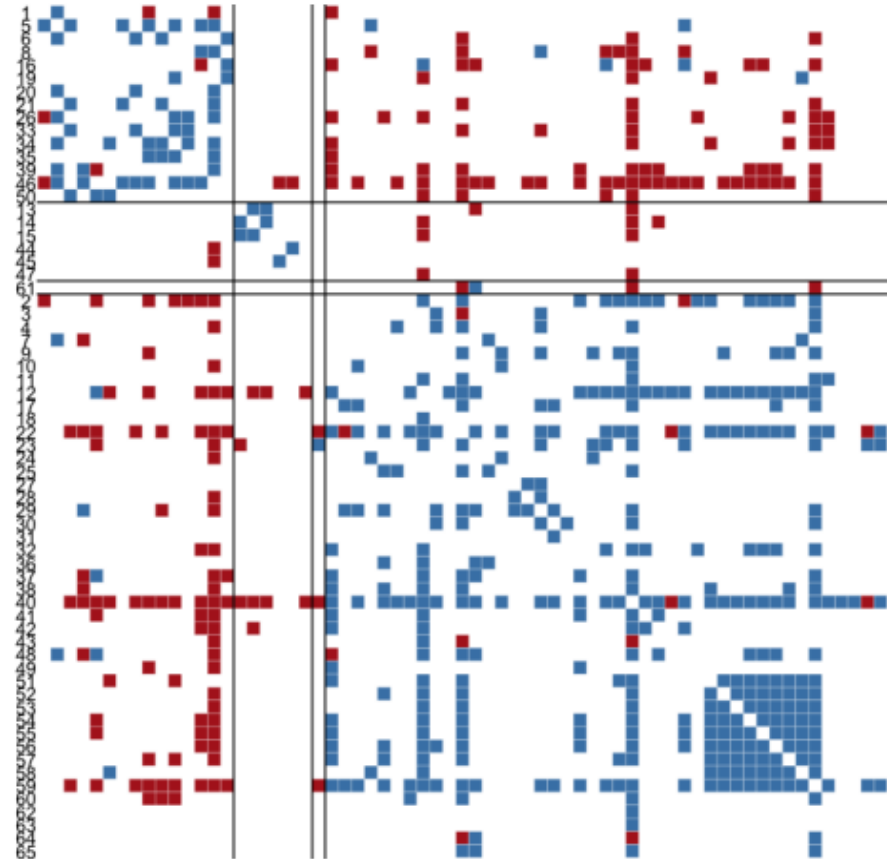
# Študentský projekt – postavy z Hobita

- **Špeciálny typ zhukovania:**
  - Postavy môžu patriť aj do viacerých zhukov



# Študentský projekt – postavy z Harryho Pottera

- **Vrcholy:**
  - Jednotlivé postavy
- **Hrany:**
  - Rozlíšenie podľa toho, či ide o pozitívny alebo negatívny vzťah
  - Rozdiel oproti napríklad karate klubu (merali sme intenzitu kontaktov, resp. existenciu kontaktu) alebo od mafiánov (sledovali sme, či sa stretli alebo nie)



# Projekty v tomto roku...

Sociálne siete sú zaujímavá téma,  
nápadov je veľa :-)

