

# Siete vytvorené z korelácií časových radov

Beáta Stehlíková

2-EFM-155 Analýza sociálnych sietí

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

Siete vytvorené z korelácií

# Siete vytvorené z korelácií

- ▶ Majme stacionárne časové rady  $\rightarrow$  môžeme počítať korelácie medzi nimi
- ▶ Sieť:
  - ▶ vrcholy budú časové rady (resp. štáty, firmy, . . . , ktorým zodpovedajú)
  - ▶ hrana bude existovať medzi ľubovoľnými dvoma vrcholmi
  - ▶ váha bude zodpovedať “vzdialenosti” medzi časovými radmi a bude odvodená od korelácie medzi nimi

# Príklad 1

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

```
##           DAX           SMI           CAC           FTSE
## DAX  1.0000000  0.7031219  0.7344304  0.6394674
## SMI  0.7031219  1.0000000  0.6160454  0.5847791
## CAC  0.7344304  0.6160454  1.0000000  0.6485679
## FTSE 0.6394674  0.5847791  0.6485679  1.0000000
```

## Výpočet vzdialenosti časových radov

- ▶ Majme *standardizované* časové rady  $x$  a  $y$  (teda s nulovým priemerom a jednotkovou disperziou)
- ▶ Nech  $d_e$  je ich euklidovská vzdialenosť, potom:

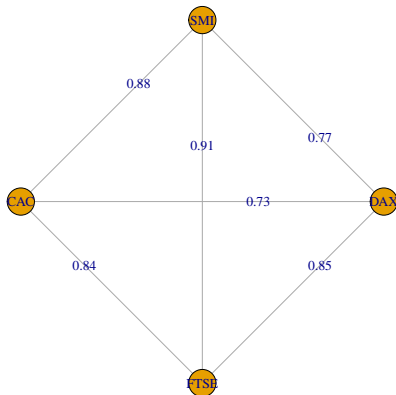
$$d_e^2 = \sum (x_i - y_i)^2 = \sum x_i^2 + \sum y_i^2 - 2 \sum x_i y_i = n[2 - 2\rho(x, y)]$$

- ▶ Pre ľubovoľné časové rady  $x$  a  $y$  definujeme ich vzdialenosť ako

$$d(x, y) = \sqrt{2(1 - \rho(x, y))}$$

## Príklad 1 - pokračovanie

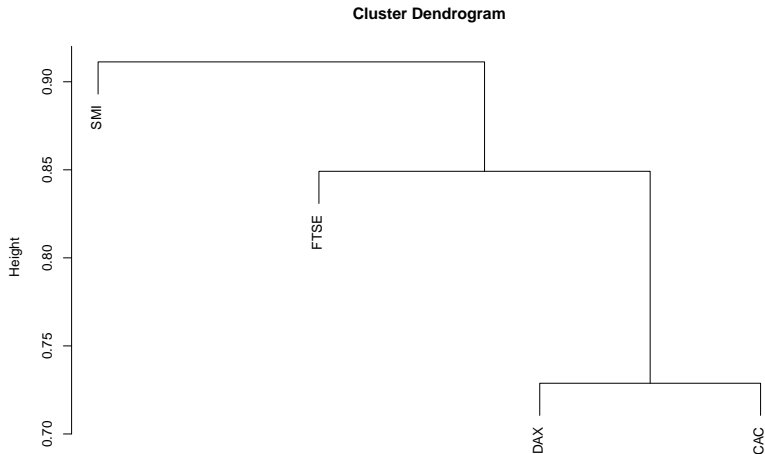
```
d <- ... # matica vzdialenosti (t.j. vahy) -> DOPLNTE  
g <- graph.adjacency(d, weighted=TRUE, mode="lower")  
plot(g, ) # -> DOPLNTE PARAMETRE
```



## Hierarchické zhlukovanie

Máme definované vzdialenosti medzi vrcholmi → môžeme spraviť hierarchické zhlukovanie:

```
clust <- hclust(as.dist(d))  
plot(clust)
```

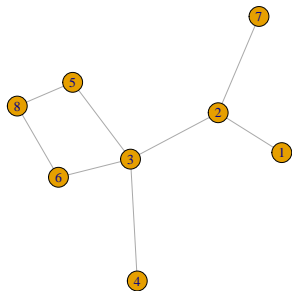
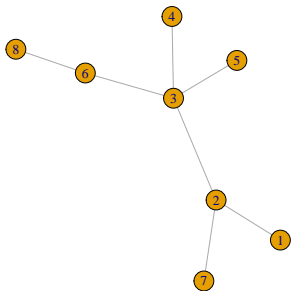




## Minimálna kostra grafu

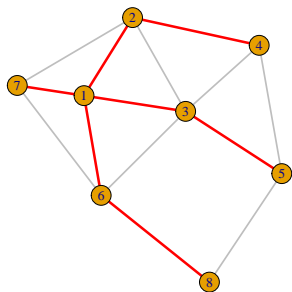
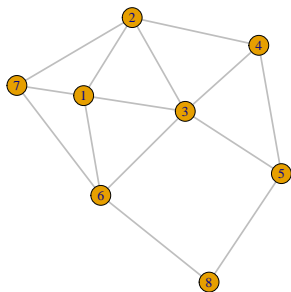
# Strom

- ▶ **Strom** - súvislý graf, ktorý neobsahuje cyklus
- ▶ Graf vľavo je strom, graf vpravo nie je strom



# Kostra

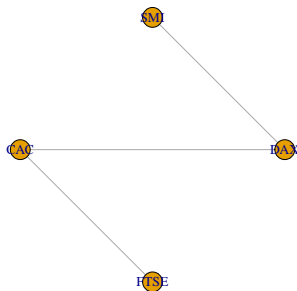
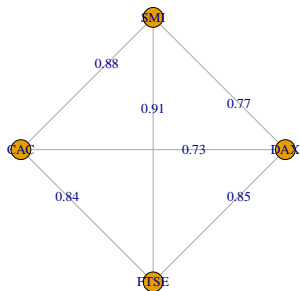
- ▶ **Kostra** súvislého grafu je podgraf obsahujúci všetky vrcholy, ktorý je stromom
- ▶ Z toho anglický názov *spanning tree*



# Minimálna kostra

- ▶ Ak majú hrany váhy, tak kostra s najmenším súčtom váh hrán sa nazýva **minimálna kostra** (*minimum spanning tree*)
- ▶ V R-ku funkcia `minimum.spanning.tree`

```
g.mst <- minimum.spanning.tree(g)
```



# Minimálna kostra - algoritmus

- ▶ Funkcia `minimum.spanning.tree` používa *Primov algoritmus*: Vojtech Jarník (1930), Robert C. Prim (1957), Edsger W. Dijkstra (1959)
- ▶ Algoritmus:
  - ▶ Vstupom je súvislý graf
  - ▶ Na začiatku dáme do kostry ľubovoľný vrchol.
  - ▶ V každom kroku vyberieme hranu s najmenšou váhou, ktorá spája vrchol z kostry s vrcholom mimo kostry. Hranu a vrchol pridáme do kostry.
  - ▶ Pokračujeme, kým v kostre nie sú všetky vrcholy grafu.
  - ▶ Ak kostra obsahuje všetky vrcholy, algoritmus končí a jeho výstupom je minimálna kostra.
- ▶ Dá sa dokázať správnosť algoritmu, teda že výstupom je vždy minimálna kostra.

## Aplikácia na sieť z korelácií

- ▶ Zostrojíme sieť z korelácií (váhy = vzdialenosti)
- ▶ Nájdeme minimálnu kostru (“najvýznamnejšie” korelácie)
- ▶ Na ňu použijeme niektorý z algoritmov na hľadanie komunit

```
set.seed(123)
E(mst)$weight <- 1-(E(mst)$weight^2)/2 # naspat korelacie
com <- walktrap.community(g.mst)
plot(com, g.mst)
```

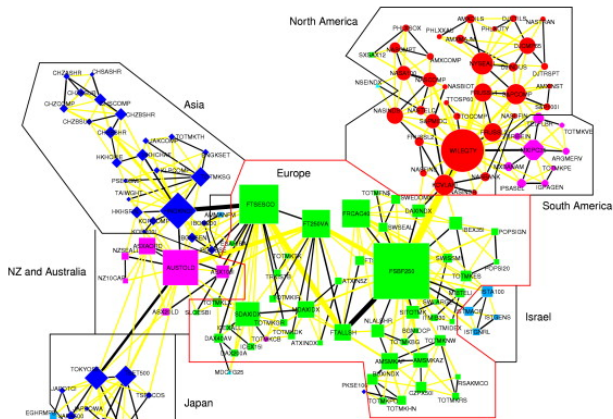
## Ukážky publikací

# Indexy I.

*Eryiğit, M., & Eryiğit, R. (2009). Network structure of cross-correlations among the world market indices. Physica A: Statistical Mechanics and its Applications, 388(17), 3551-3562.*

https:

//www.sciencedirect.com/science/article/pii/S0378437109003318

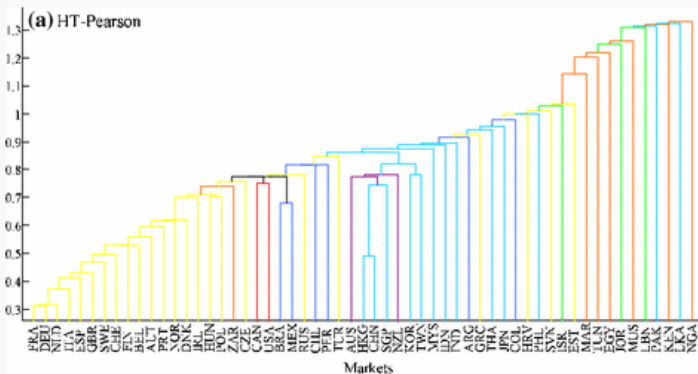




## Indexy II.

Wang, G. J., Xie, C., & Stanley, H. E. (2018). Correlation structure and evolution of world stock markets: Evidence from Pearson and partial correlation-based networks. *Computational Economics*, 51(3), 607-635.

<https://link.springer.com/article/10.1007/s10614-016-9627-7>

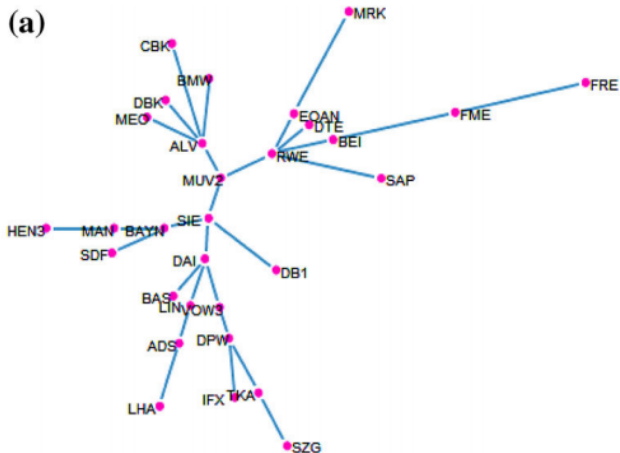




# Výnosy akcií

Birch, J., Pantelous, A. A., & Soramäki, K. (2016). Analysis of correlation based networks representing DAX 30 stock price returns. *Computational Economics*, 47(4), 501-525.

<https://link.springer.com/article/10.1007/s10614-015-9481-z>

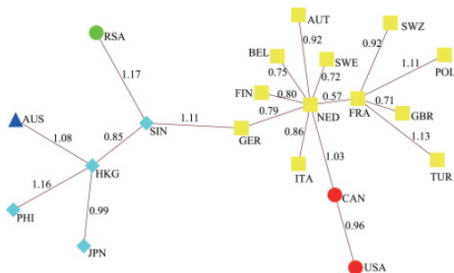


# Trh s nehnutelnostami

Wang, G. J., & Xie, C. (2015). Correlation structure and dynamics of international real estate securities markets: A network perspective. *Physica A: Statistical Mechanics and its Applications*, 424, 176-193.

https:

//www.sciencedirect.com/science/article/pii/S0378437115000278



[Download full-size image](#)

Fig. 3. Minimum spanning tree (MST) with weighted edges of the set of 20 country indices in international real estate securities markets during the period 2006–2012. The weight on each edge represents the distance ( $d_{ij}$ ) between the corresponding two vertices

Námet na projekt

## Parciálne korelácie, regularizácia

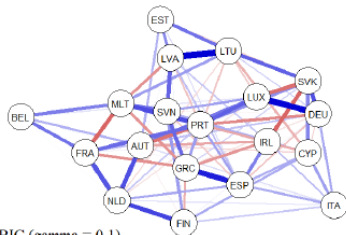
- ▶ Dve premenné môžu byť korelované kvôli tomu, že sú obidve ovplyvňované nejakou inou, nie preto, že je medzi nimi priamy vzťah - toto odstraňuje **parciálna korelácia**, meria vzťah očistený od vplyvu iných premenných → môžeme pracovať so sieťami z parciálnych korelácií
- ▶ Pomocou MST sme vyberali dôležité korelácie, iným prístupom je **regularizácia** - penalizovanie nenulových korelácií v sieti, výsledkom sú niektoré korelácie nulové

### V R-ku

- ▶ balík `glasso`
- ▶ balík `bootnet`, funkcia `estimateNetwork`
- ▶ treba si naštudovať, čo sa tam vlastne deje

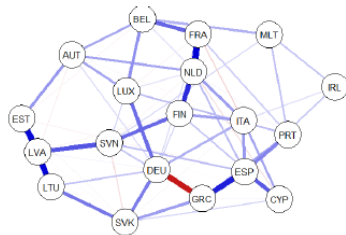
# Parciálne korelácie, regularizácia

Ukážka regularizácie (Radka Litvajová, 2018, diplomovka)



EBIC ( $\gamma = 0.1$ )

Počet hrán = 84



EBIC ( $\gamma = 0.221$ )

Počet hrán = 66