

# Network Analysis Methods 2

Network Science

Šíření vlivu v sociálních sítích

2023/24

# Problém

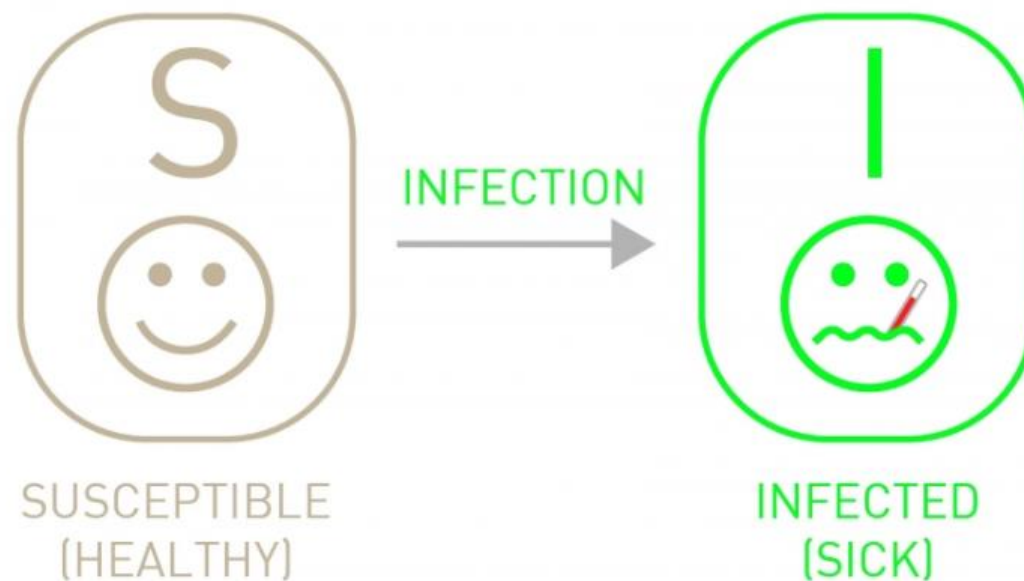
- Jedním ze studovaných fenoménů sociálních sítí je způsob, kterým se šíří informace. Klíčovou úlohou je zde nalezení takových vrcholů sítě (lidí), které mají vysokou schopnost šířit informace tak, aby se dostaly k velkému množství jiných vrcholů sítě.
- Vrcholy s velkým vlivem jsou významnými hráči například při šíření reklamy na nové produkty. Na druhé straně ale mohou šířit i negativní informace (např. tzv. fake news). Není jednoduché takové vrcholy v obrovských sociálních sítích nalézt.

# Procesy šíření

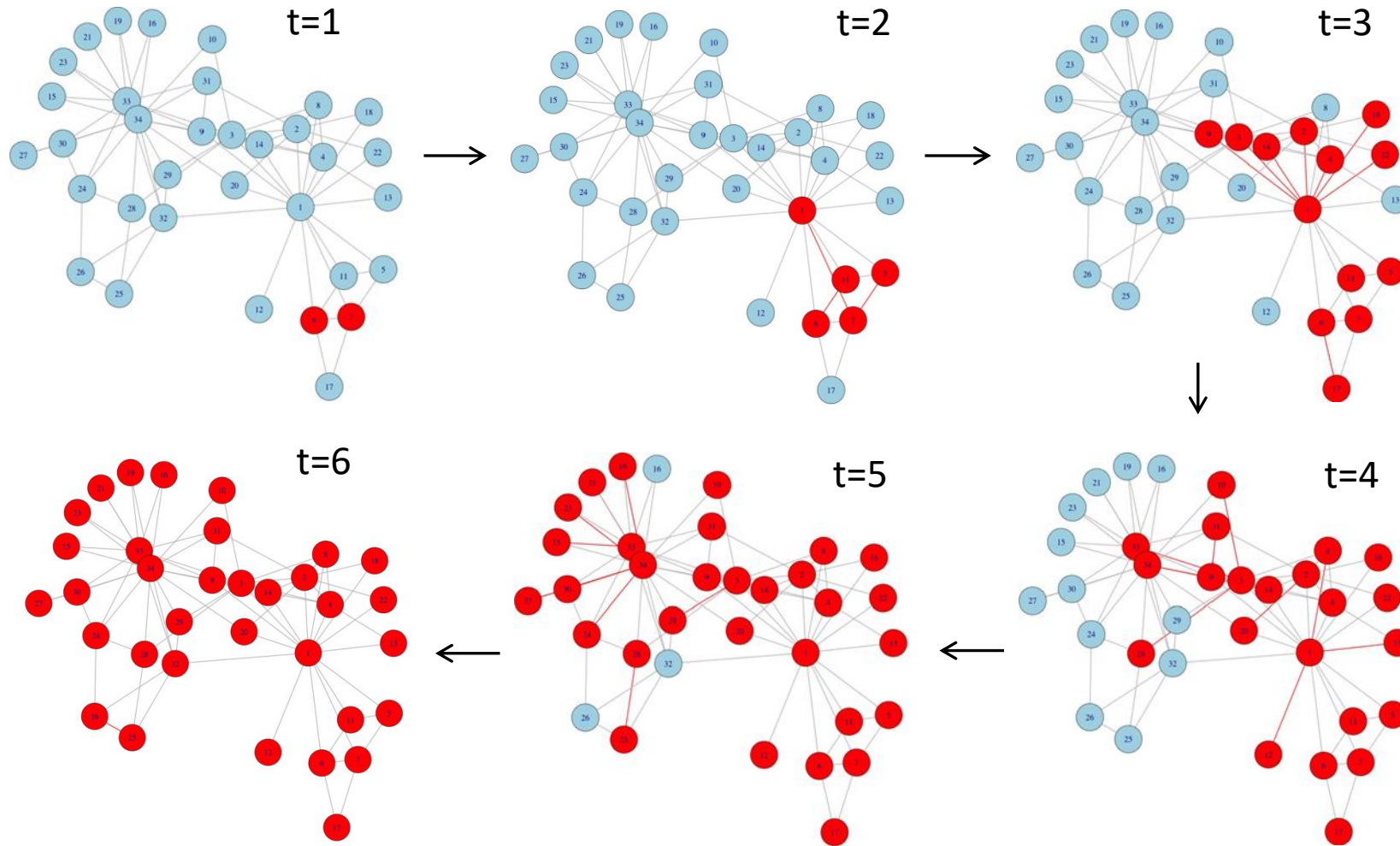
- Díky vazbám (hranám) reprezentujícím interakce mezi vrcholy je možné pozorovat šíření informací od jednoho k druhému (a k dalším a dalším vrcholům).
- Je to jako *...jedna paní povídala...*
- Můžeme se inspirovat přírodními jevy, např. epidemiemi...
- ...a pracovat s tzv. *modely*, kterými se snažíme (byť ne zcela přesně) popsat *JAK* a *PROČ* dochází k nějakému jevu.

# Inspirace přírodou: Susceptible-Infected (SI) model

- *Vnímavý* (S, Susceptible):  
Zdraví jedinci, kteří se ještě nenakazili.
- *Infekční* (I, Infectious):  
Nakažlivé osoby, které mohou infikovat ostatní.
- $\beta$  *míra infekce* (infection rate):  
Pravděpodobnost nákazy v čase  $t$ .



# SI model: $\beta=0,5$ (Karate klub)



# Šíření vlivu

- Co znamená, že je něco *virální*? Jak se nějaká informace (zpráva, obrázek, audio, video, produkt, služba, značka apod.) stane virální?
- Nejde jen o to, *CO se šíří*, ale také *KDO to šíří*...
- ...a v podstatě to jde v prostředí sociálních sítí snadno, a to díky *sdílení*, které může mít různou formu.
- Předpokladem virálního marketingu je, že když se zpočátku zaměříme na několik málo *vlivných členů* sítě - například jim dáme zdarma vzorky produktu - můžeme spustit kaskádu vlivu, díky níž přátelé doporučí produkt dalším přátelům, a nakonec ho vyzkouší mnoho lidí.

# Klíčová otázka

*Zvolme v síti nějaký vrchol a předpokládejme, že mu dáme k dispozici zajímavou informaci. Dokážeme alespoň odhadnout, kam (ke komu) se tato informace rozšíří?*



# Nezávislý kaskádový (IC) model šíření

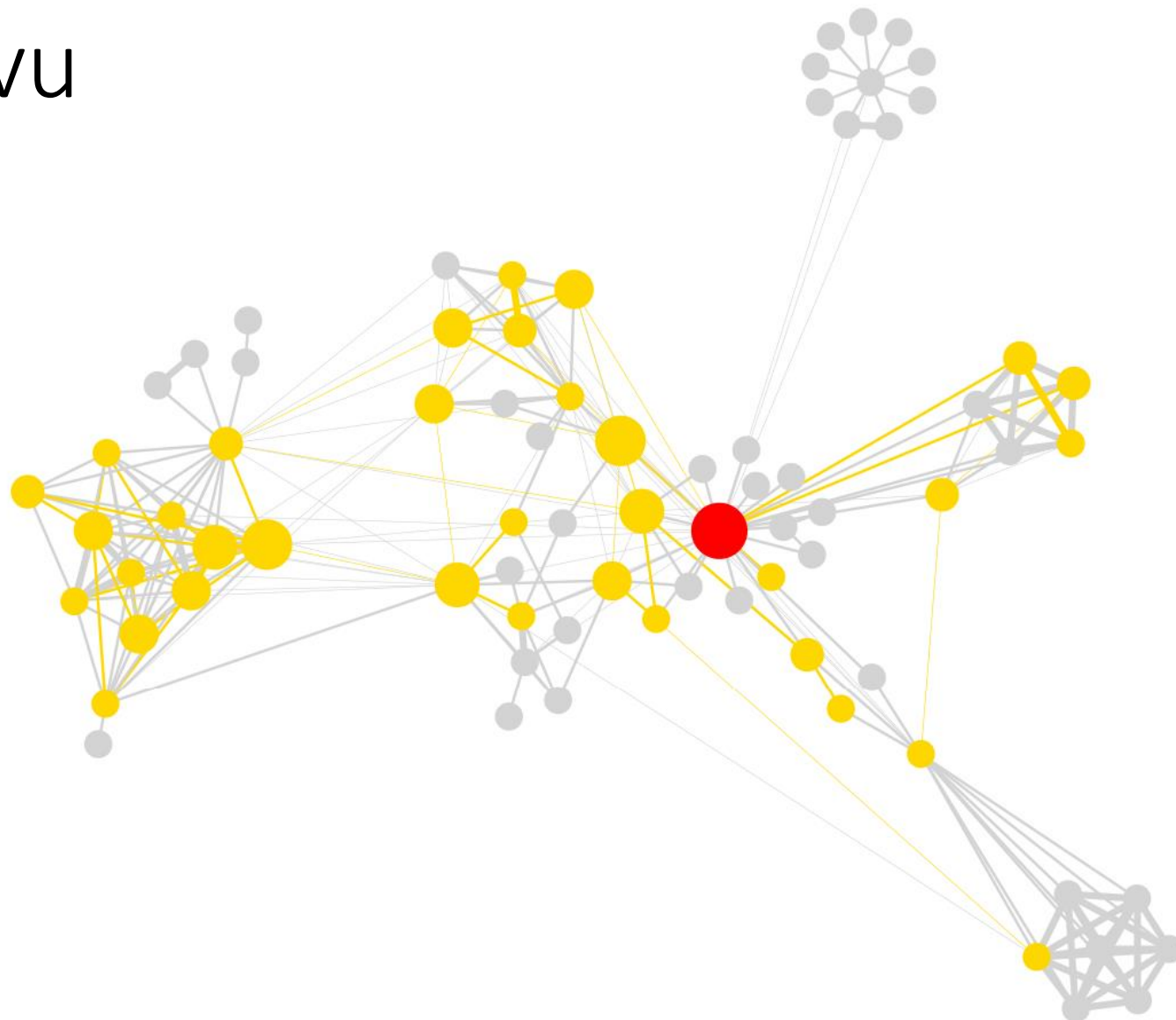
Začneme s počáteční neprázdnou množinou aktivních vrcholů a zvolenou pravděpodobností  $p$ . Proces šíření probíhá v diskrétních krocích podle následujícího pravidla:

1. Když se uzel poprvé stane aktivním v kroku  $t$ , má jedinou šanci aktivovat každého ze svých aktuálně neaktivních sousedů a u každého z nich uspěje s pravděpodobností  $p$ .
2. Ty vrcholy, u kterých uspěje, se stanou aktivními v kroku  $t+1$ ; ale ať už aktivní vrchol uspěje nebo ne, nemůže v následujících krocích učinit žádný další pokus o aktivaci svých neaktivních sousedů.
3. Proces pokračuje opět od bodu 1 dokud není možná žádná další aktivace (neexistují žádné aktivní vrcholy).



# Simulace šíření vlivu

- Simulace je aplikace algoritmu založeného na IC modelu.
- Simulaci je nutné provádět opakovaně a poté „zprůměrovat“ její výsledky.
- *Opakovaně* znamená alespoň 100x a více, a to v závislosti na velikosti sítě.

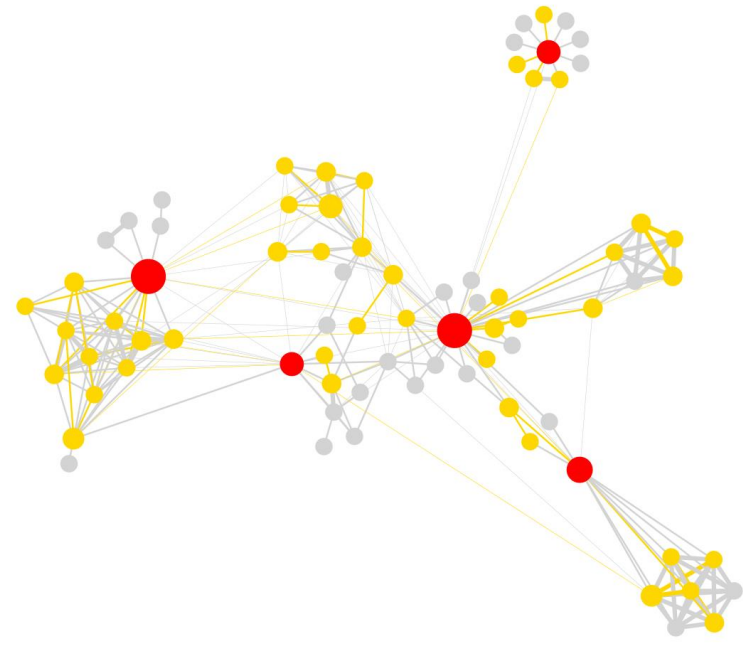
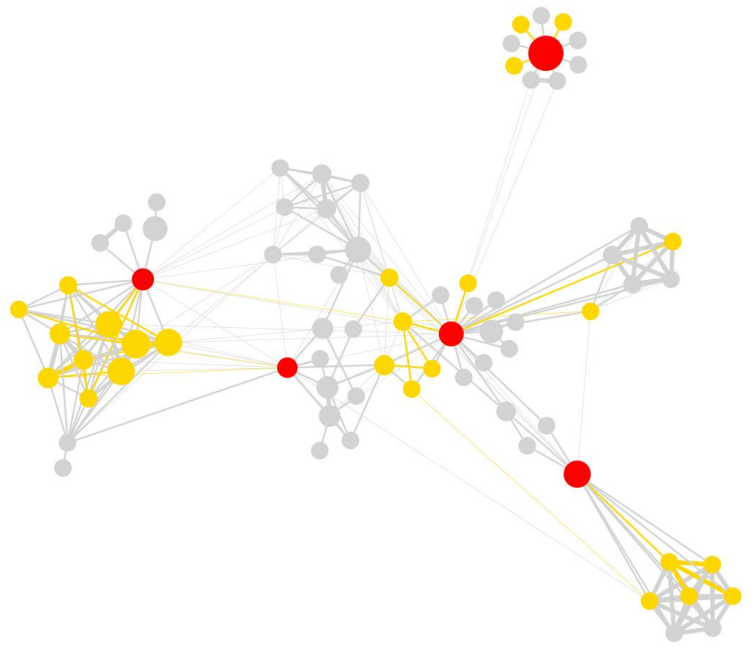
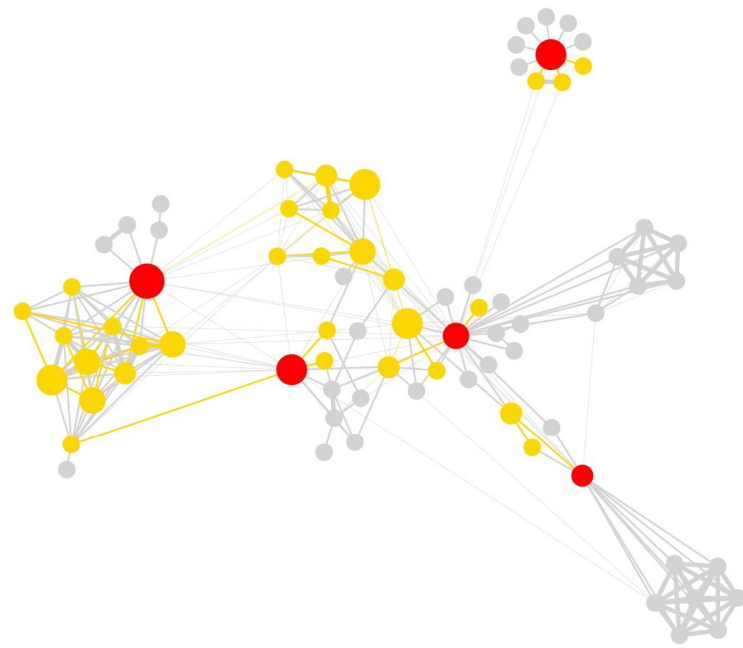
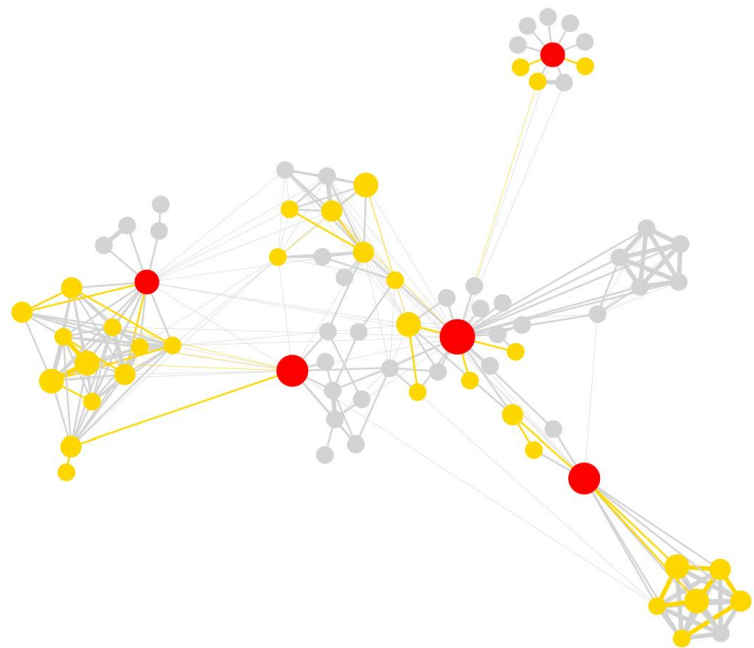


# Obecně o počítačové simulaci

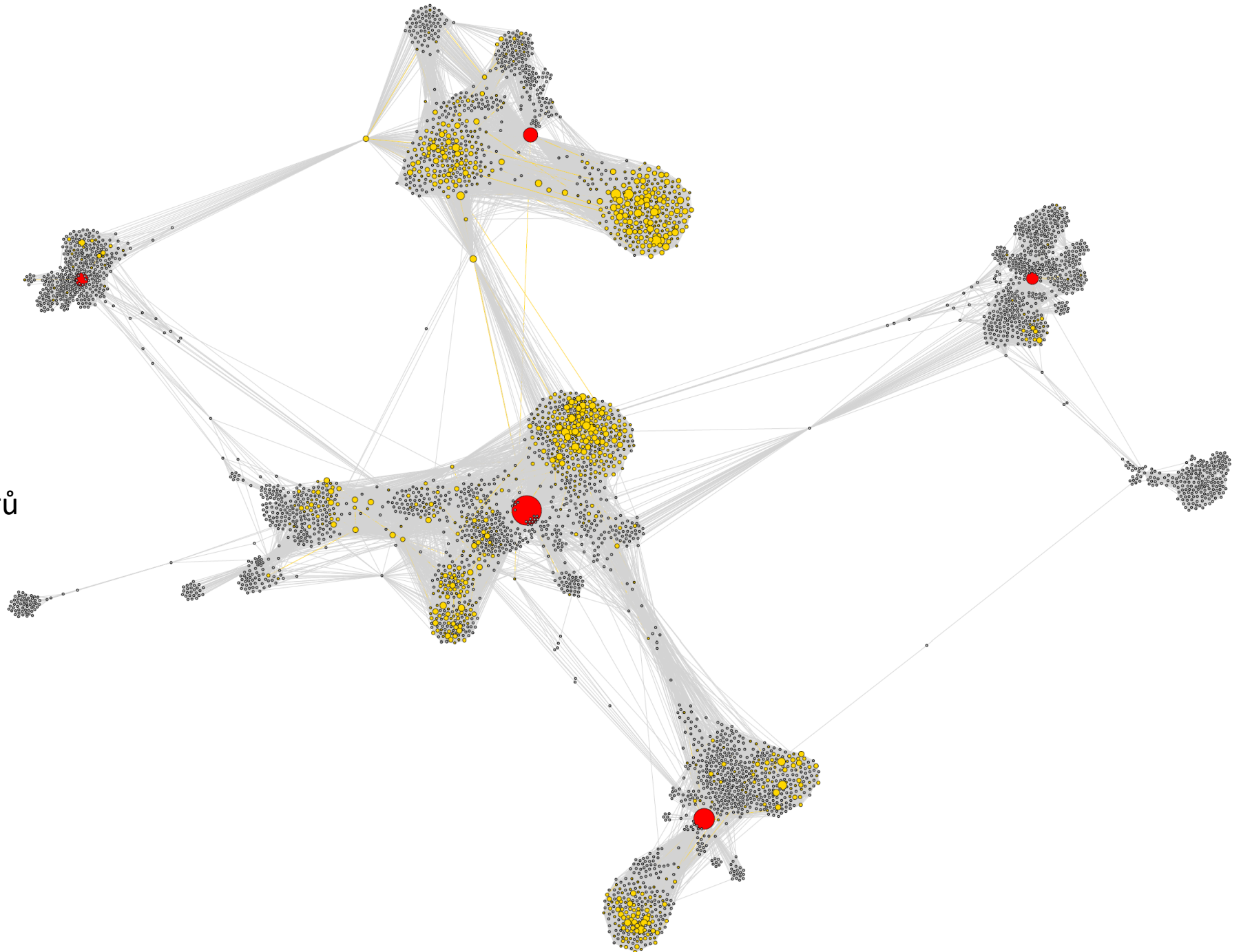
- Neformálně řečeno, počítačová simulace je experiment s počítačovým modelem. Musíme mít tedy model, který zjednodušuje realitu, přičemž zachovává nějakou vlastnost(i) této reality. Simulace je spojena s několika kroky:
  - Návrh modelu
    - Redukce (zjednodušení) reality
    - Konstrukce modelu
  - Realizace simulace
    - Implementace počítačového programu
    - Nastavení parametrů, spuštění popř. vizualizace
    - Pozorování
  - Analýza a statistické vyhodnocení získaných dat

# Maximalizace vlivu

- Lidé např. z marketingu, správy veřejného zdraví a dalších oblastí mají často omezený rozpočet na kampaně, takže se mohou zaměřit pouze na malý počet osob. Proto v případě potřeby šíření informací (o nových produktech nebo třeba o nové nákaze) řeší strategii, *JAK* a *NA KOHO* v sítích zacílit tak, aby se informace účinně šířila.
- Problém maximalizace vlivu spočívá ve *výběru malé skupiny  $K$  vrcholů* v síti zajišťujících *maximální očekávaný rozsah šíření*.
- Nalezení *optimální* skupiny je NP-úplný problém (pro velké sítě jej nelze v reálném čase algoritmicky zvládnout). Proto se používají heuristiky založené zejména na výběru málo vzájemně propojených vrcholů majících vysoký stupeň.
- Není tedy zaručeno, že najdeme *optimální skupinu*. Cílem proto je najít *dostatečně dobrou skupinu* vrcholů s vlivem blížícím se optimu.



- Síť sestavená kolem několika lidí na FB
- 4 039 vrcholů
- 88 234 hran
- $K = 5$  automaticky detekovaných influemcerů
- $p = 0,02$
- Průměrný stupeň 43,7

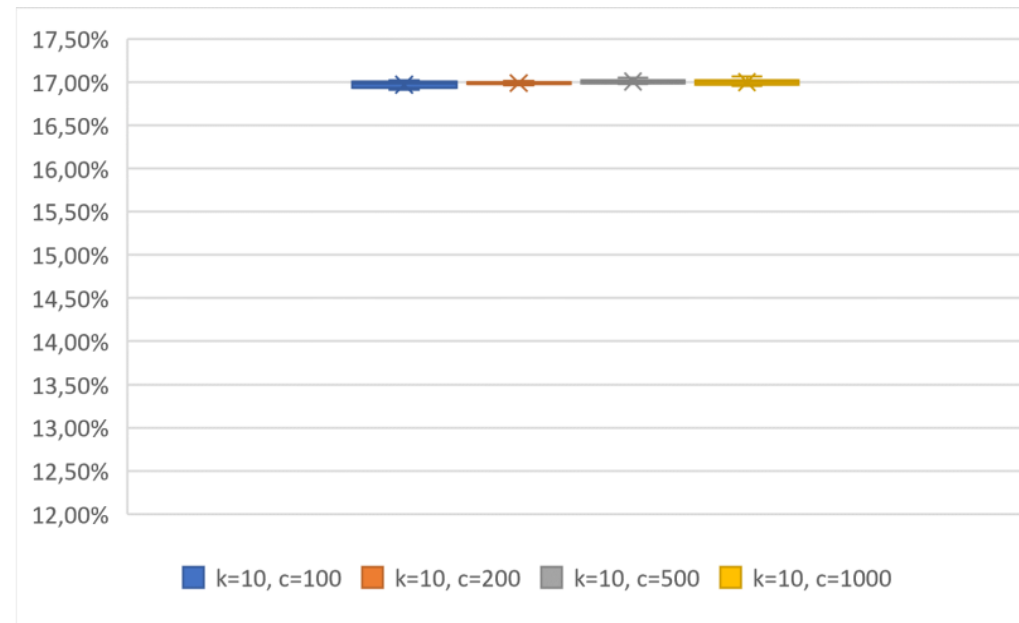
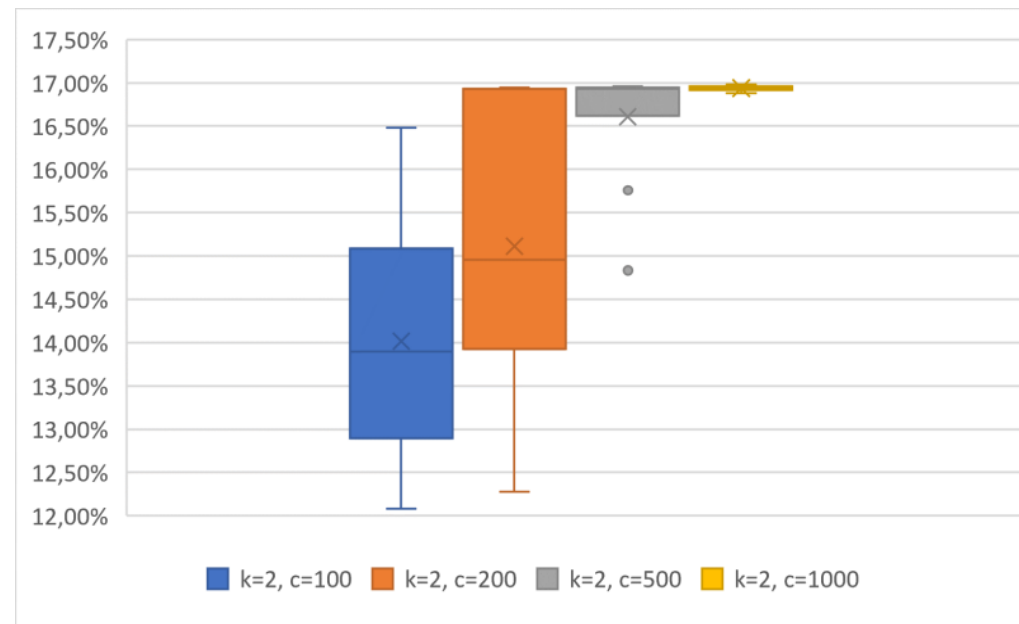
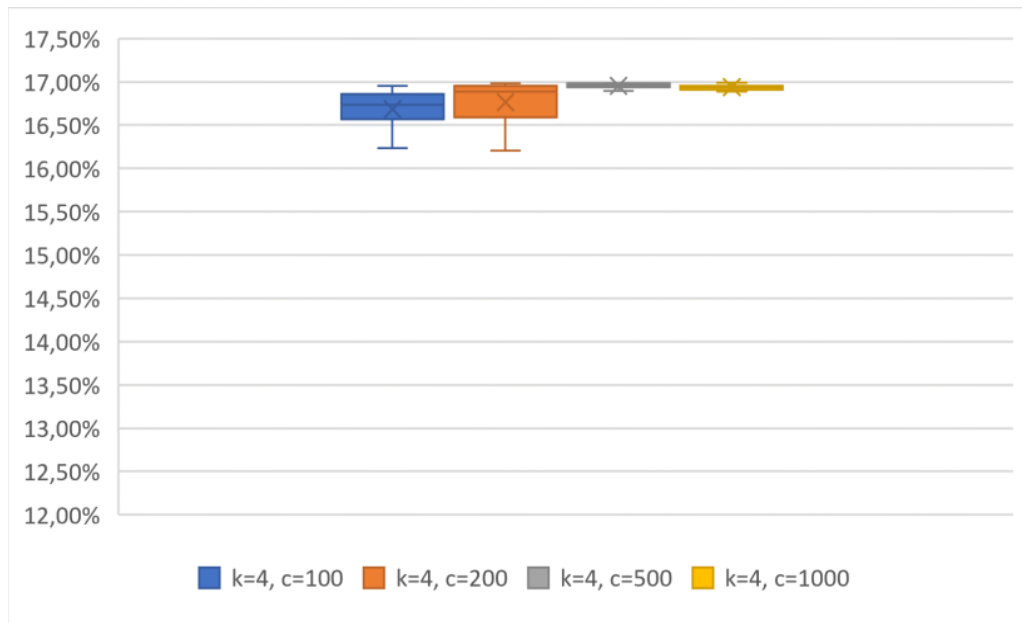


# A co velké sítě?

- Problém je, jak nalézt vrcholy, které by *MOHLY* být influencery. Musíme nějak zkusit nalézt dobré kandidáty (zejména s vyšším stupněm).
  - Můžeme to udělat např. tak, že náhodně zkoušíme hrany. Jde o to, že s vysokou pravděpodobností má alespoň jeden z vrcholů každé hrany vysoký nebo alespoň nadprůměrný stupeň.
- Z těchto kandidátů pak musíme vybrat ty, kteří mezi sebou nejsou silně propojeni (nemají hranu nebo alespoň mají málo společných susedů).
- Můžeme také pro jednotlivé kandidáty spustit simulaci a nakonec vybrat ty kandidáty, kteří dopadnou nejlépe.
- Obvykle platí, že jen *několik málo vybraných influencerů* (2-4) pokryje svým vlivem *potřebnou část* zkoumané části sítě (ve smyslu optimálního poměru *cena-výkon*).

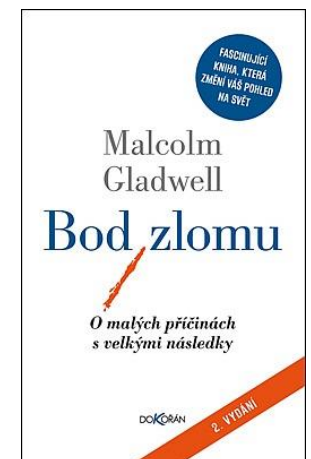
# Velká síť (statistika)

- FB: studenti Pensylvánské univerzity
- 41 536 vrcholů, 1 362 220 hran, stupeň 65,6
- $p = 0,01$ , 10x500 simulací



# Shrnutí

- Detekce lidí se silným vlivem (influencerů) je nesmírně zajímavá úloha s praktickým využitím.
- Používali jsme pouze strukturální vlastnosti sítě.
- Obvykle nám to ale nestačí, protože je potřeba brát v úvahu také další vlastnosti, jako jsou *společné zájmy* sousedících vrcholů nebo *délka a pravidelnost* jejich vzájemné interakce.
- To je ovšem zá rámecem dnešní přednášky, tak snad někdy příště...
- **Malcolm Gladwell: *Bod zlomu*. Dokořán, 2007**
  - **The Tipping Point: How Little Things Can Make a Big Difference, 2000)**





# Úkoly na cvičení

- Na malé síti provedte simulaci šíření vlivu. Pro každou jednotlivou simulaci vizualizuje výsledek podobně, jako je to v prezentaci (použijte např. Gephi a GDF soubor).
- Provedte simulaci šíření vlivu na třech velkých sítích z předchozích prezentací.
- Obě simulace statisticky vyhodnoťte.

# Odkazy

- Kempe, D., Kleinberg, J., Tardos, É. (2003, August). *Maximizing the spread of influence through a social network*. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 137-146).