

Úvod do analýzy sociálních sítí

Historie a přehled typických úloh

2024-25

Organizace výuky

- doc. Mgr. Miloš Kudělka, Ph.D.
 - EA 439, +420 597 325 877
 - homel.vsb.cz/~kud007
 - milos.kudelka@vsb.cz
- Přednáška
 - Zaměřeno na znalosti
 - Ukázky práce se sítěmi
- Cvičení
 - Diskuze
 - Prezentace připravených úkolů
 - Interpretace toho, co je možné o sítích zjistit
 - Práce s nástroji pro analýzu sítí

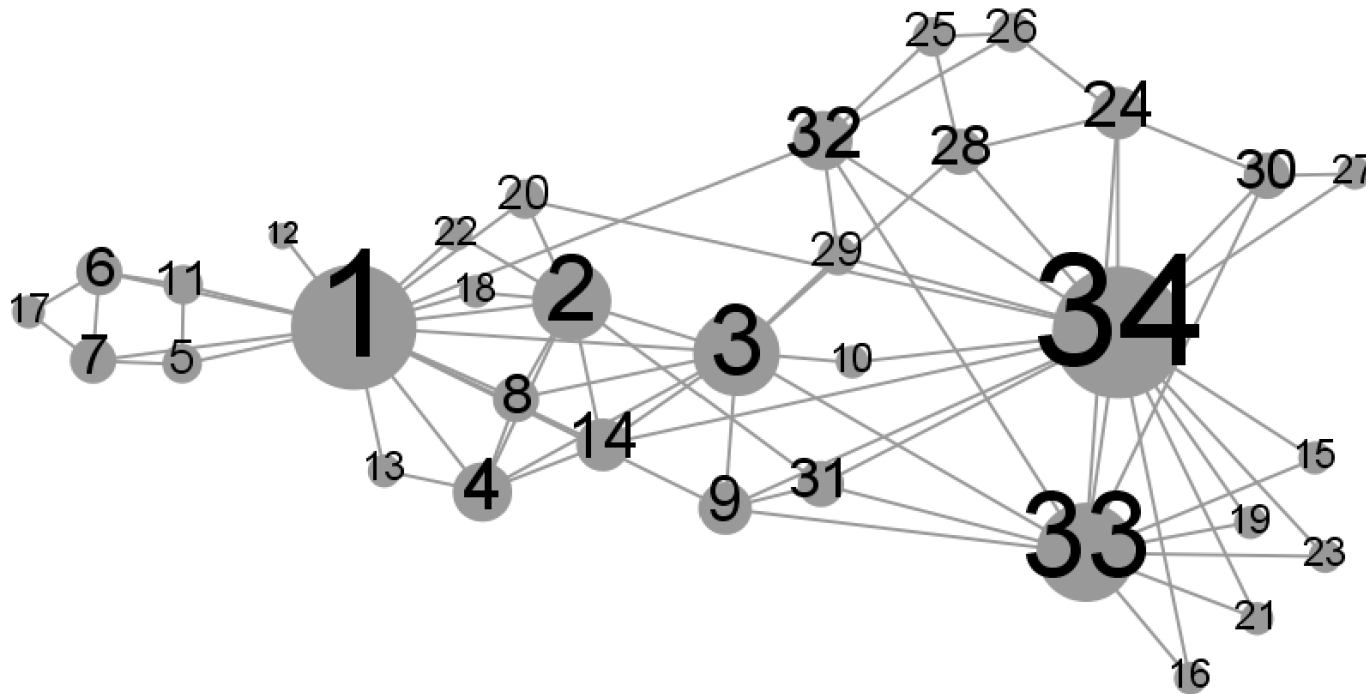
Podmínky pro získání zápočtu

- Cvičení jsou povinná, povoleny jsou nejvýše dvě absence. Další odůvodněné absence jsou řešeny individuálně s garantem předmětu.
- Aktivita na cvičení:
 - 1 bod za účast s tím, že se očekává alespoň základní znalost toho, co bylo předmětem přednášky.
 - Až 2 body za splnění převážné části zadání úkolu na cvičení.
 - Až 2 body za dokončení úkolu ze cvičení (lze získat také za dokončení úkolu doma)
 - Je bodováno 8 cvičení, celkově je tedy možné za semestr získat až 40 bodů, minimem za aktivitu je 20 bodů.
- Semestrální práce:
 - Analytický úkol (využití znalostí a schopností), zdokumentovaná analýza zvolené sítě s interpretací výstupů.
 - Celkově je možno získat až 60 bodů, minimem je 30 bodů.

Síť (graf) je struktura obsahující

vrcholy (uzly)

hrany (vazby, spojení)

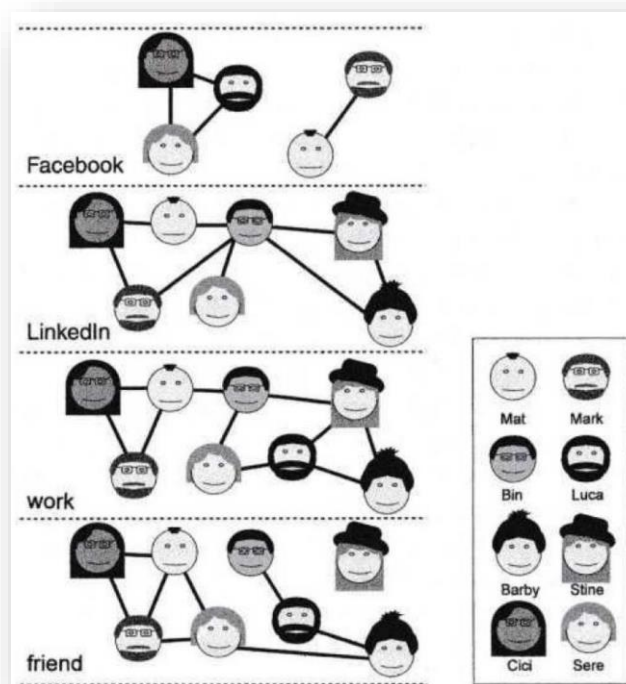
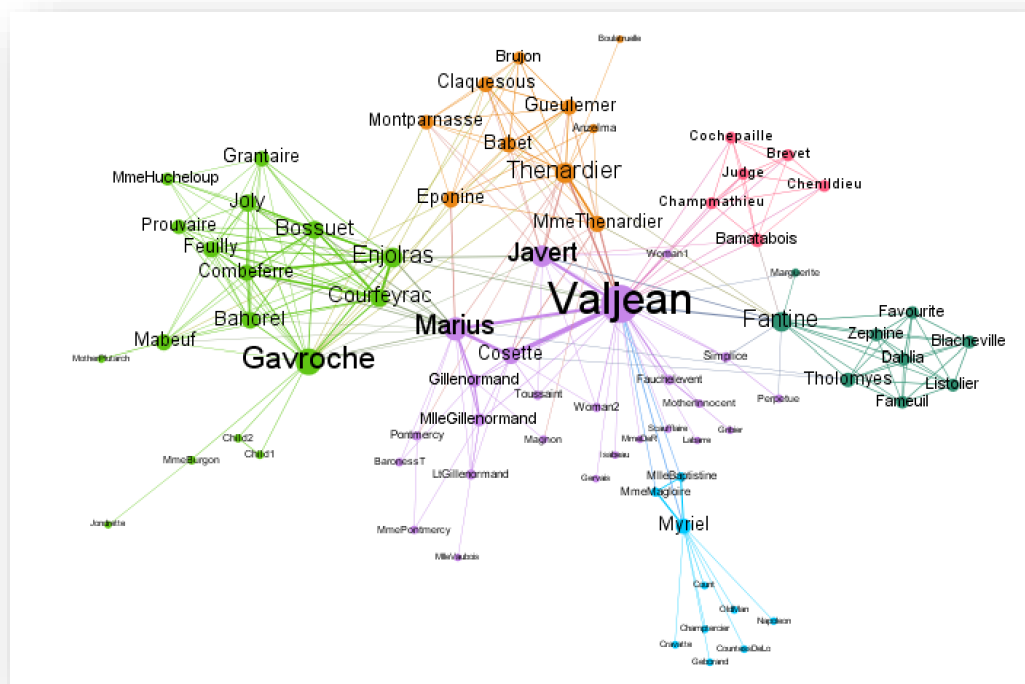


Není to tak jednoduché, protože...

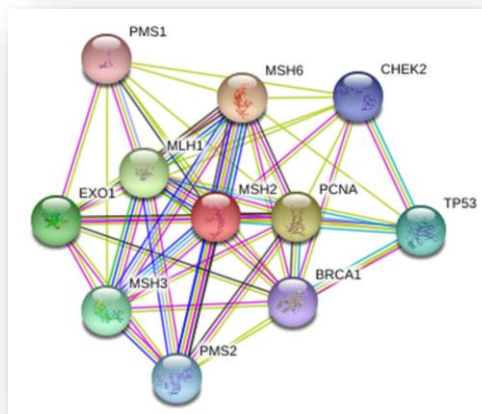
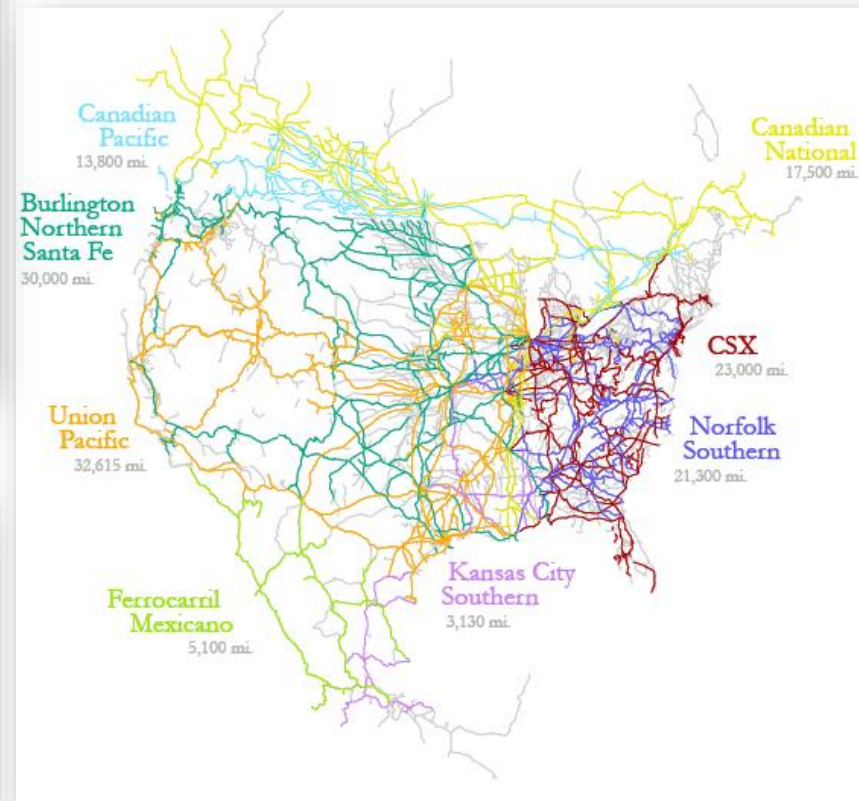
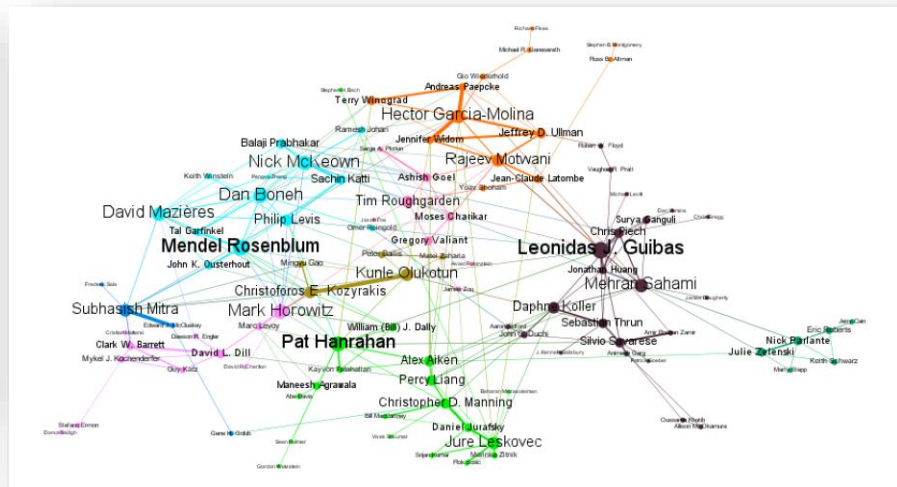
Sítě mohou být orientované nebo neorientované

Sítě mohou být vážené nebo nevážené

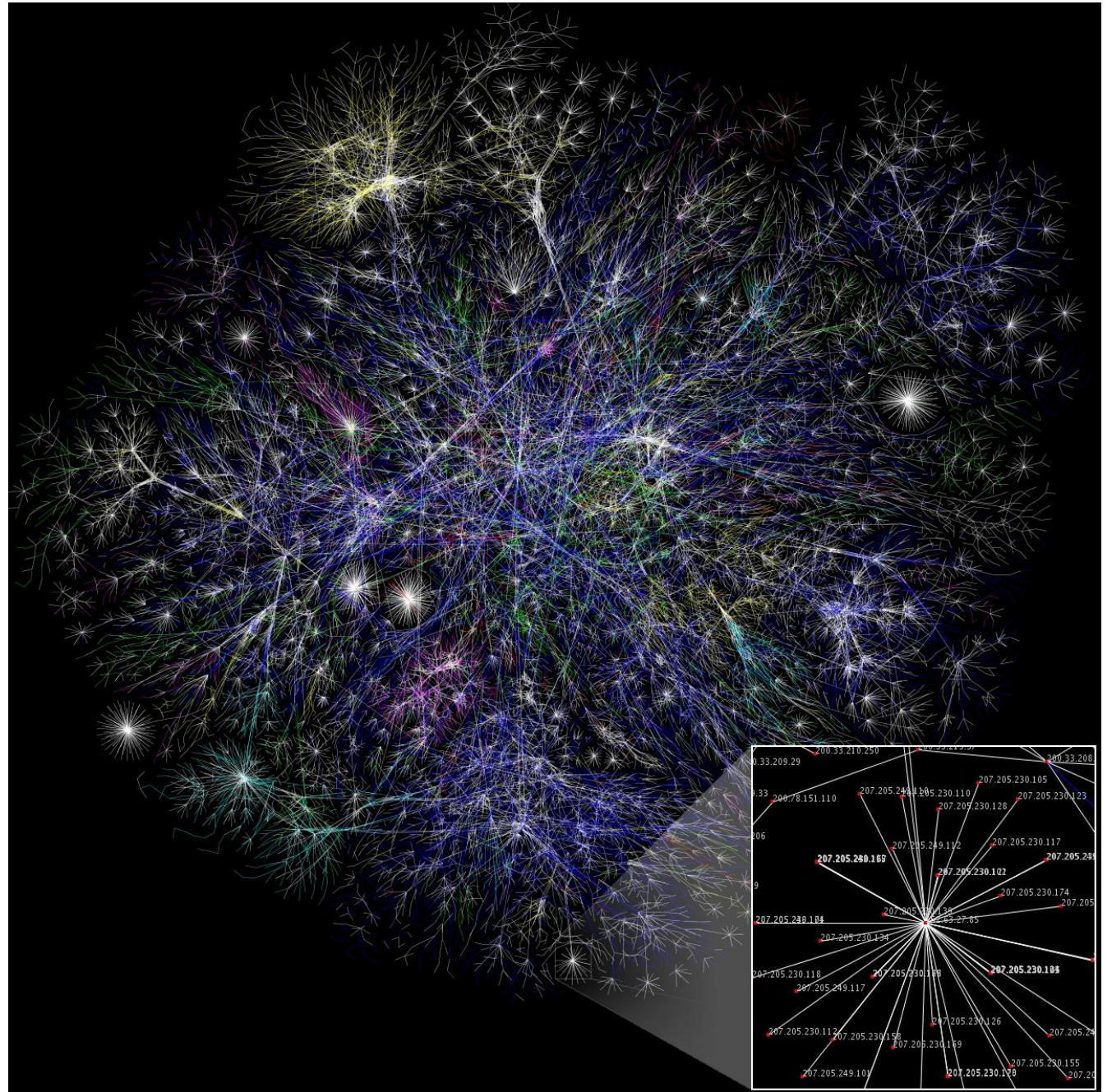
Sítě mohou mít jednu nebo více vrstev



Sítě jsou všude, kam se podíváme



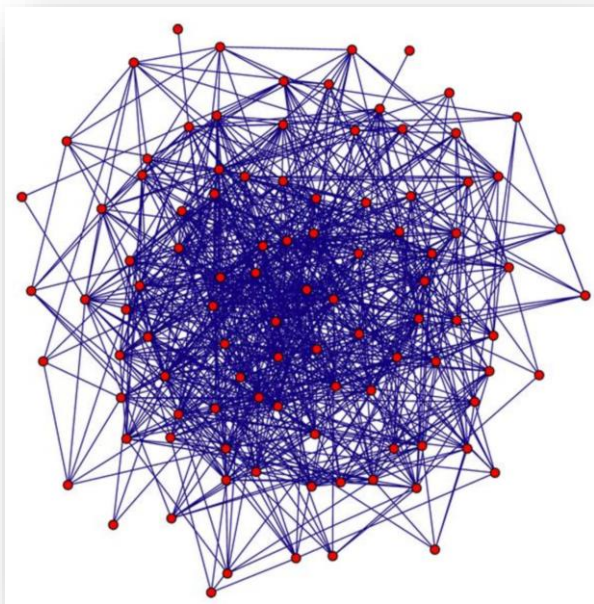
Internet (Wikipedia)



Pál Erdős (1913 - 1996)

1959 – první model sítě (náhodný graf)

Zahájil obrovský rozmach zkoumání sítí





Stanley Milgram (1933-1984)

1967 – šest stupňů odloučení (**malý svět**)

Experiment s doručením dopisu
neznámé osobě (v Bostonu)

Pro doručení bylo potřeba přibližně 6
kroků (šest osob předávajících si dopis)

Vzdálenosti mezi lidmi jsou malé

Albert-László Barabási (1967)

1999 – model sítě založený na
preferenčním připojování pozorovaném
na webu

Bez-škálové (scale-free) sítě
s malým počtem center

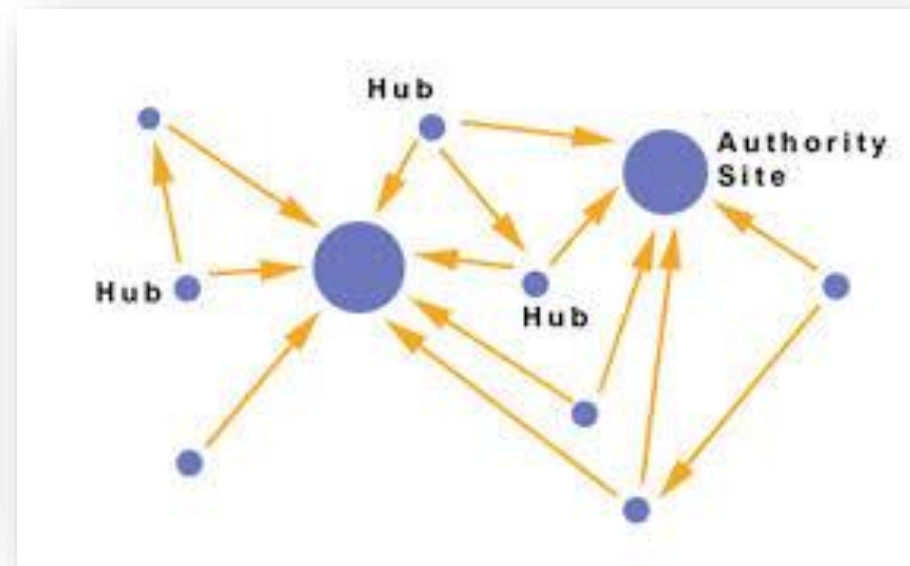
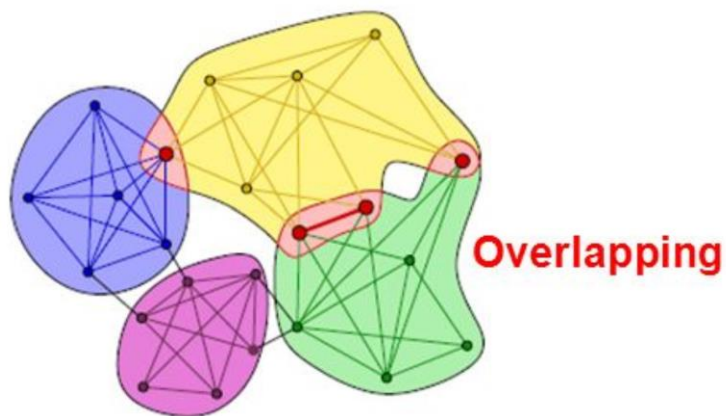
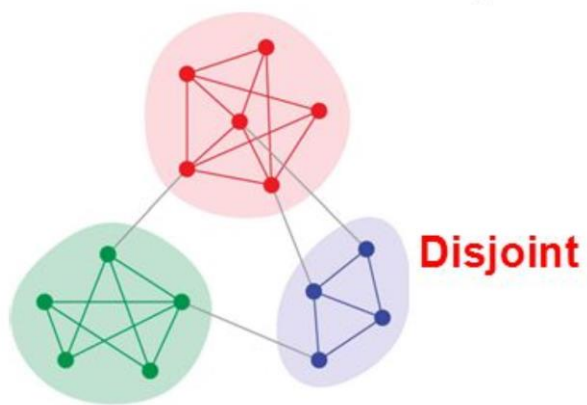
Nejdůležitější krok
v intenzivním rozvoji
analýzy sítí



Komunity

Důležité věci se odehrávají ve skupinách
Důležité vrcholy jsou v překryvech těchto skupin (propojují komunity)

Důležité vrcholy jsou centra (hub x authority)

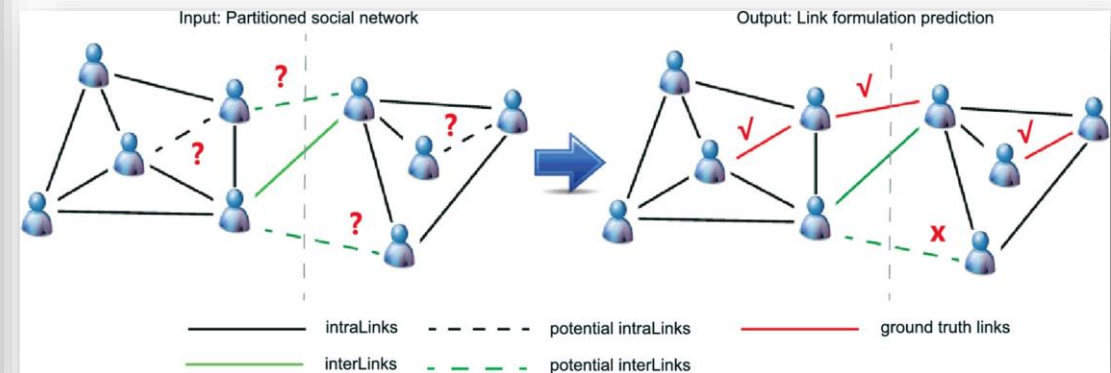
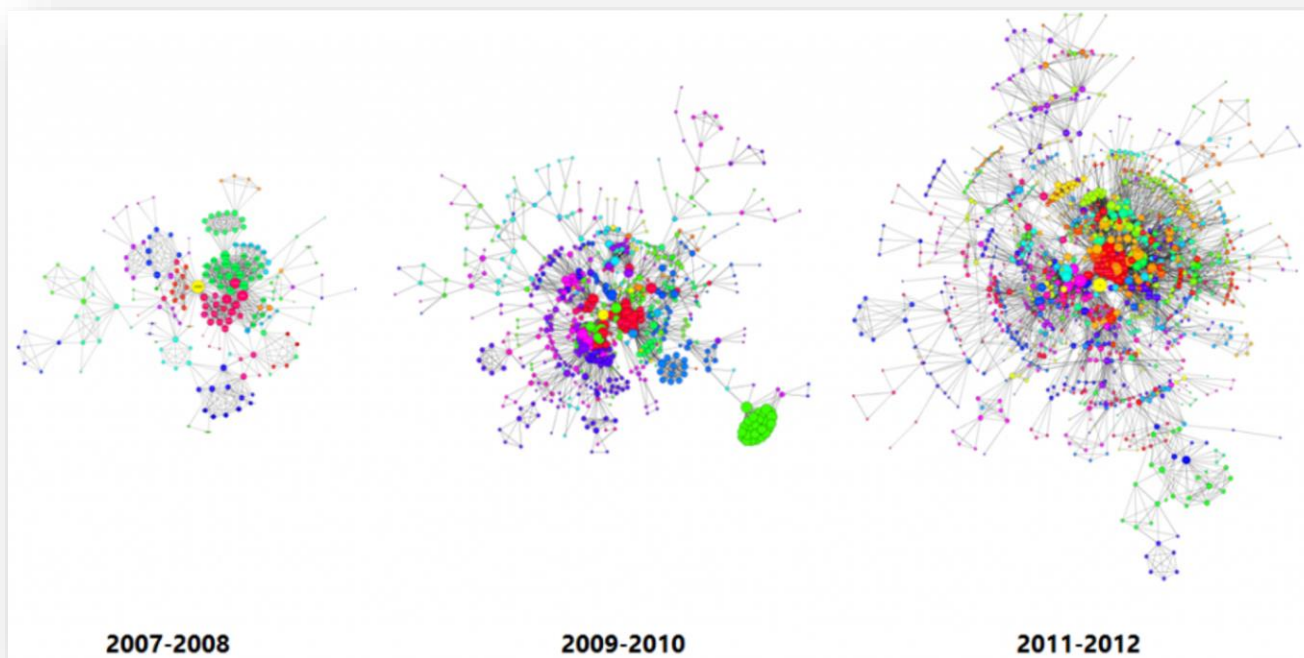


Evolute, dynamika, predikce

Vrcholy i hrany mohou přibývat i ubývat

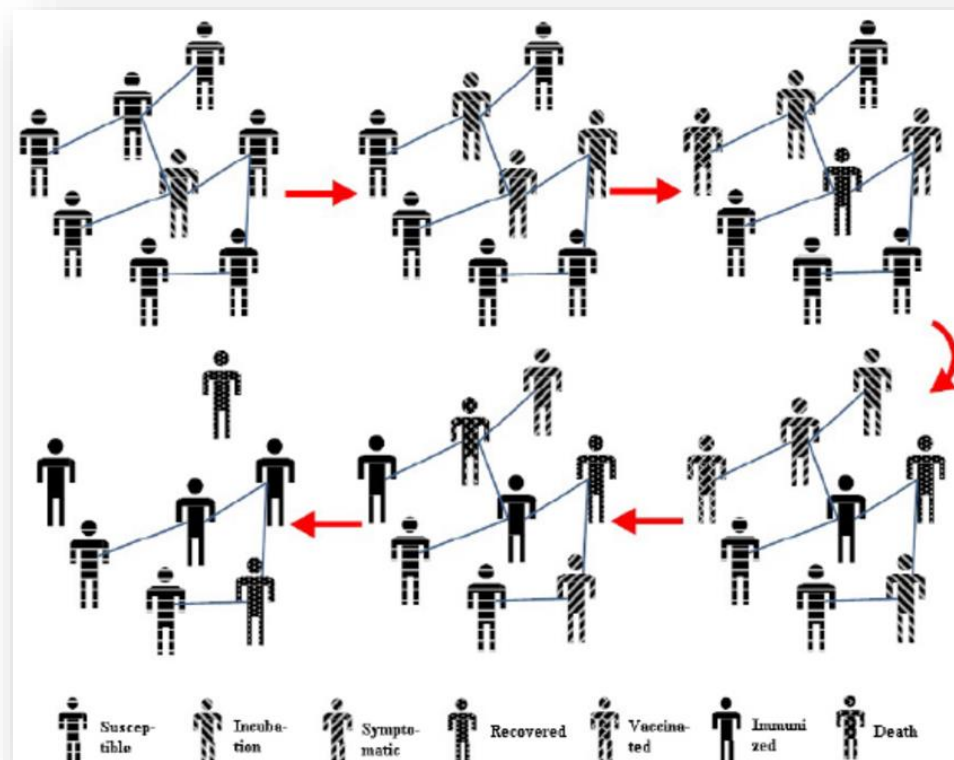
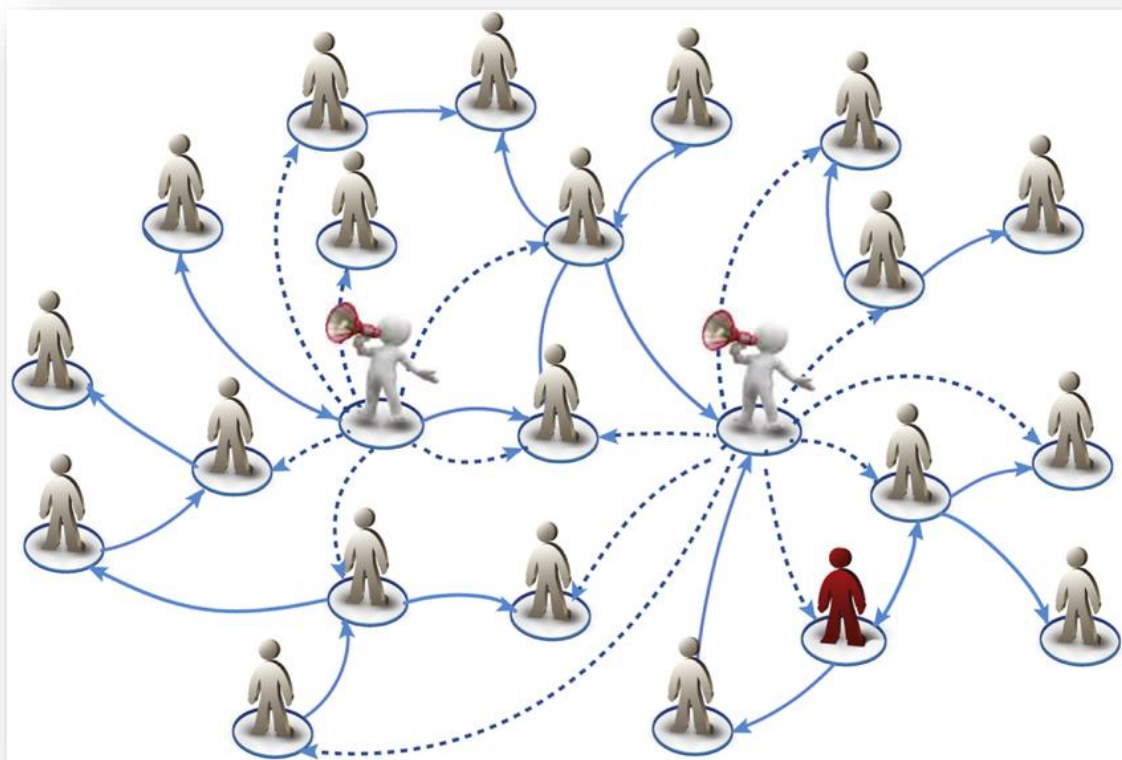
V různém čase a v různých místech sítě se to děje s různou intenzitou

Jak bude vypadat síť za nějaký čas?



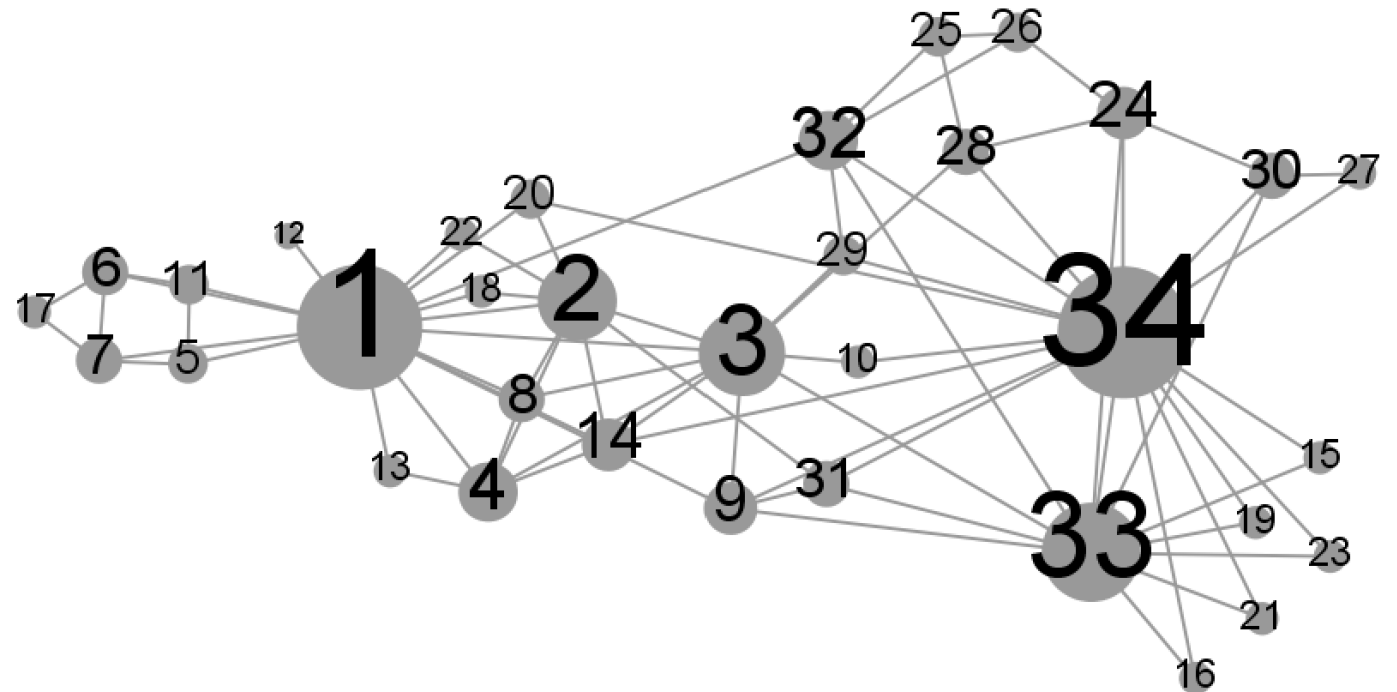
Šíření informací, viralita

Šíření informací všeho druhu a stejně tak šíření nákaz mají stejnou podstatu a zákonitosti

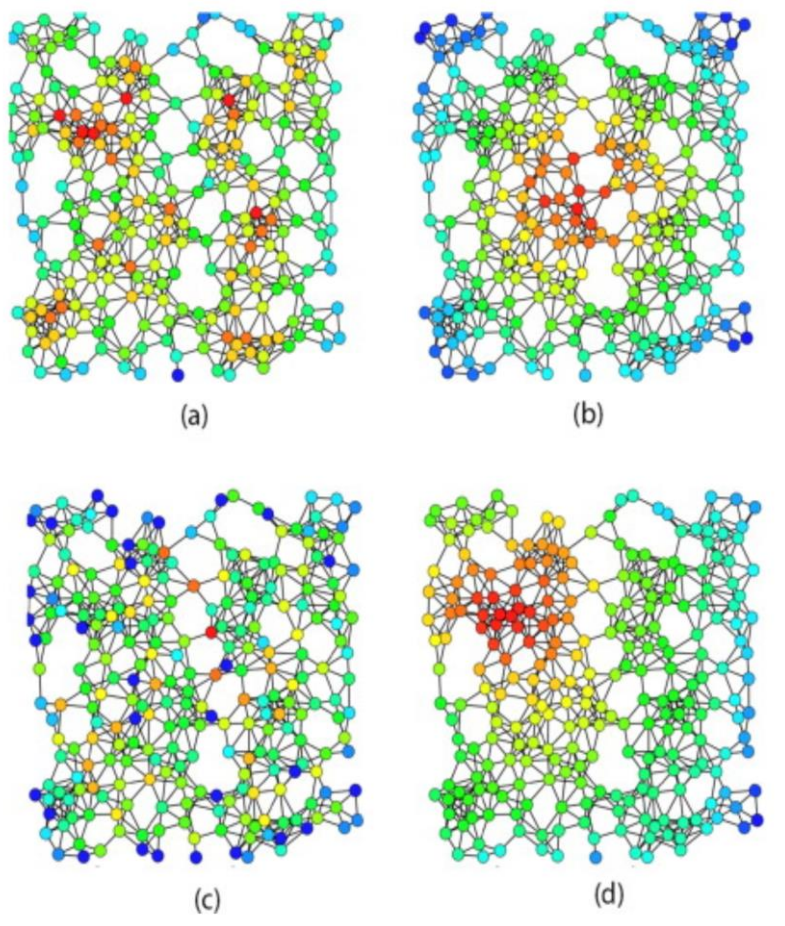


A budeme měřit...

stupeň vrcholu
nejkratší cesta mezi
vrcholy
průměr sítě
shlukovací koeficient



Není důležitost jako důležitost



(a) Stupeň

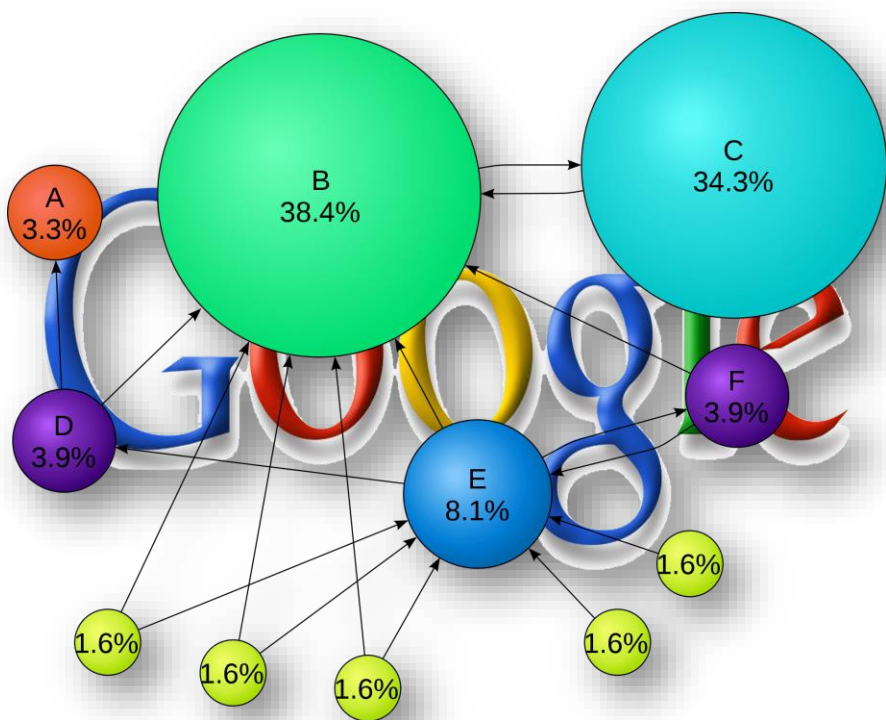
(b) Closeness

(c) Betweenness

(d) Vlastní vektor

PageRank: pochopili, jak sítě fungují

Larry Page and Sergey Brin (1999)



$$\text{PageRank of site} = \sum \frac{\text{PageRank of inbound link}}{\text{Number of links on that page}}$$

Vizualizujme to...

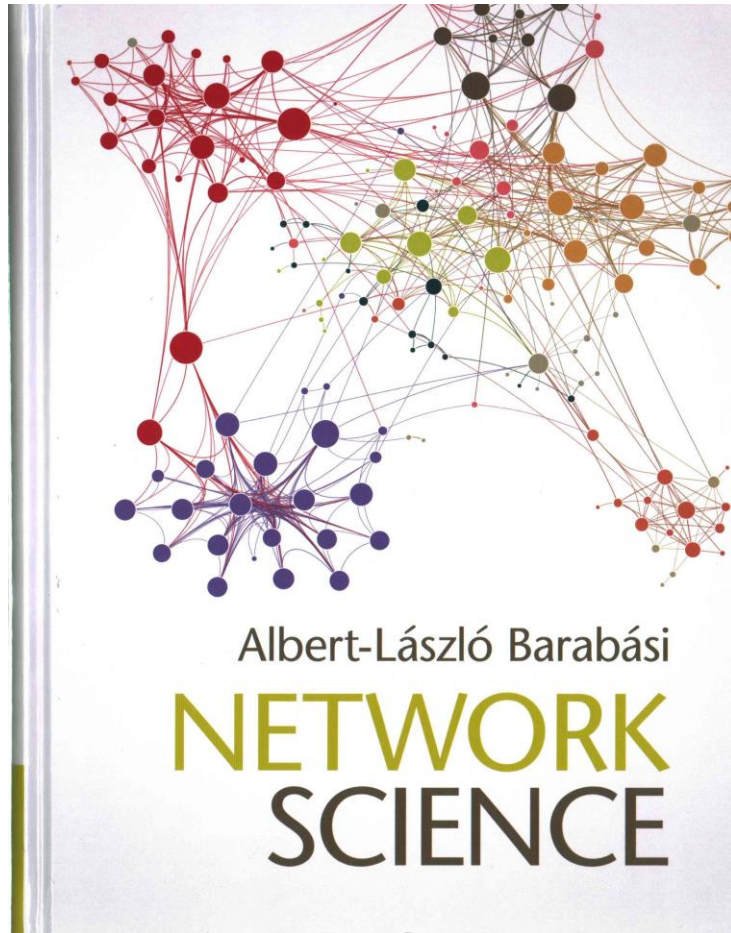
Použijme vhodný nástroj, např.

GEPHI

The Open Graph Viz Platform

<https://gephi.org/>

Network Science



<http://networksciencebook.com/>

Disciplína zaměřená na analýzu sítí kombinující

Sociologii, biologii apod.
dodávají data a interpretují výsledky

Fyziku
zkoumá vlastnosti, procesy a dynamiku

Matematiku
poskytuje teorii grafů a další nástroje, zejména pravděpodobnost a statistiku

Informatiku a data mining
zabývá se předzpracováním dat, algoritmy a výpočty, efektivním ukládáním
rozsáhlých dat a jejich analýzou

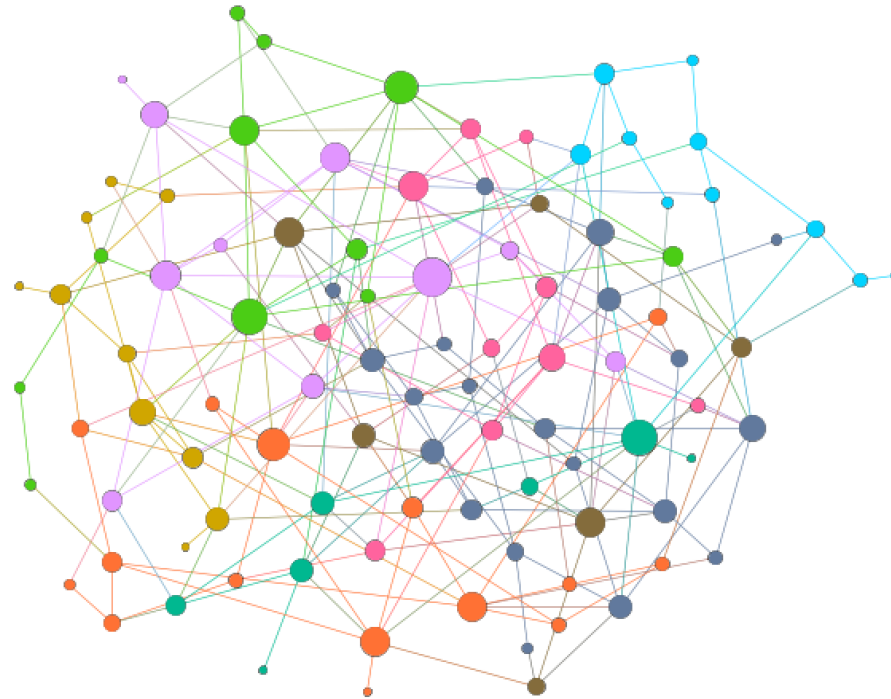
Potřebujeme zkoumat vlastnosti sítí...

....abychom zjistili odpovědi na otázku „proč jsou věci tak, jak jsou“

náhodná síť
malý svět a bezškálovost
komunitní struktura

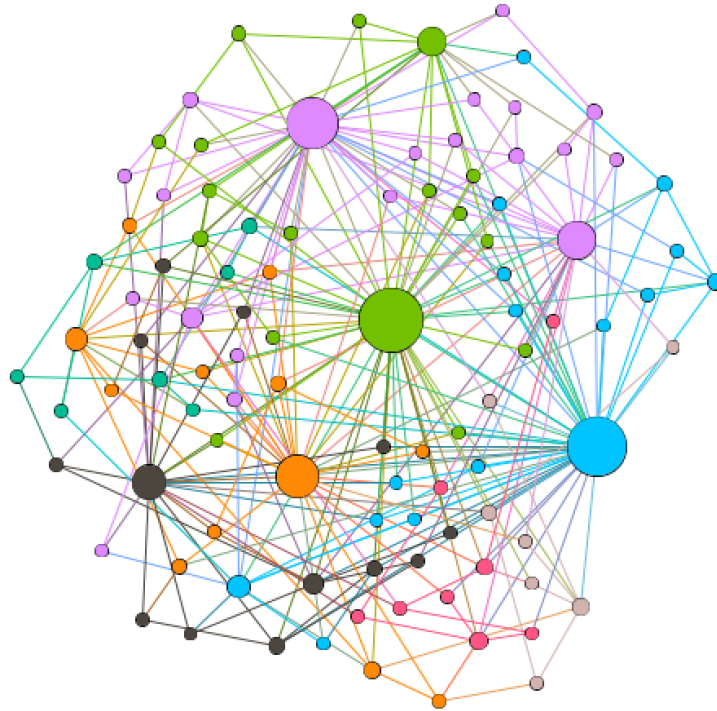
Erdős–Rényi model

Hrany mezi vrcholy vznikají se stejnou (předem známou)
pravděpodobností



Barabási–Albert model

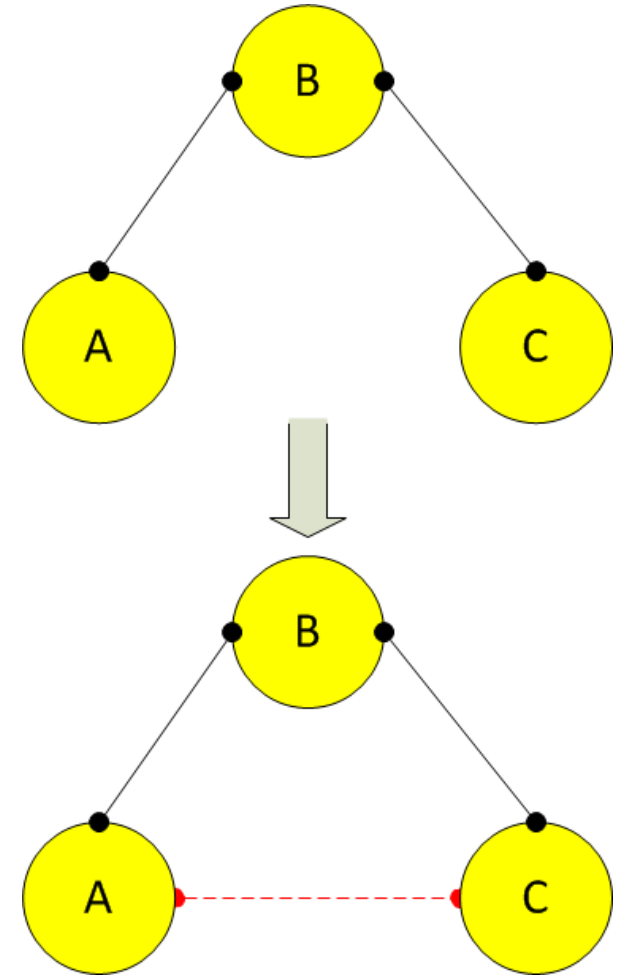
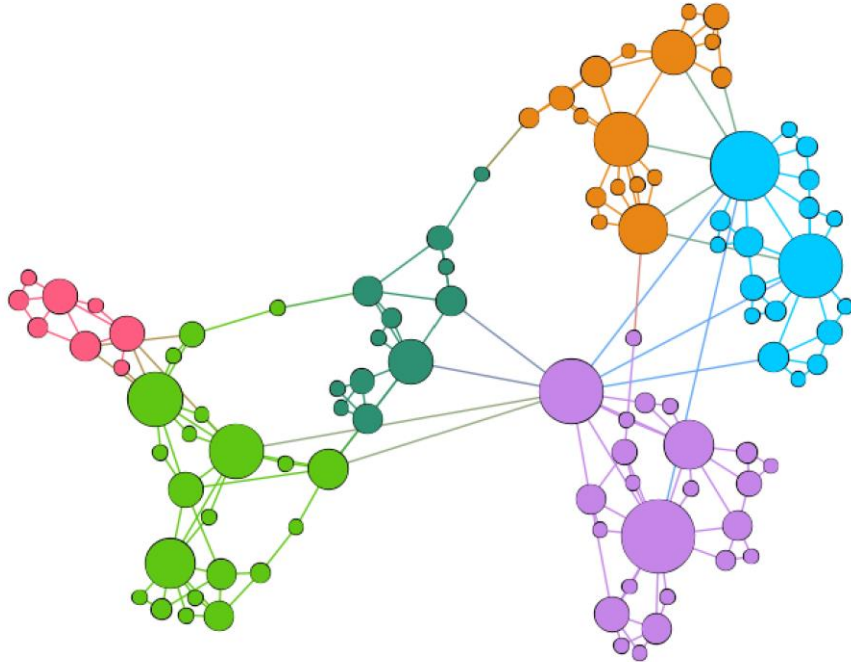
Nový vrchol se v síti spojuje přednostně s vrcholy s vyšším stupněm
s pravděpodobností odpovídající stupni vrcholu



Triadic-closure komunitní model

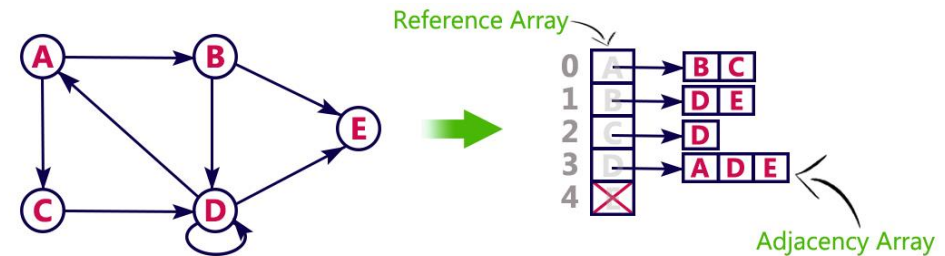
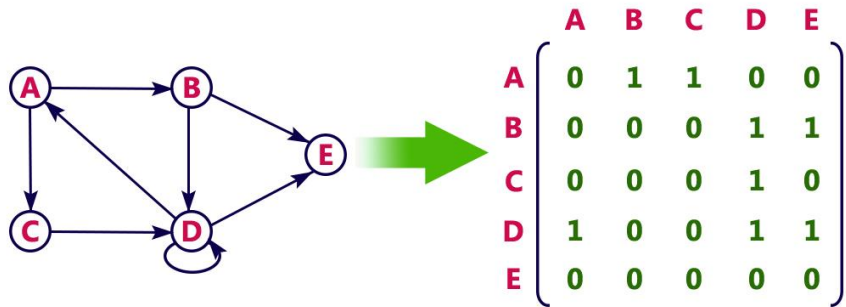
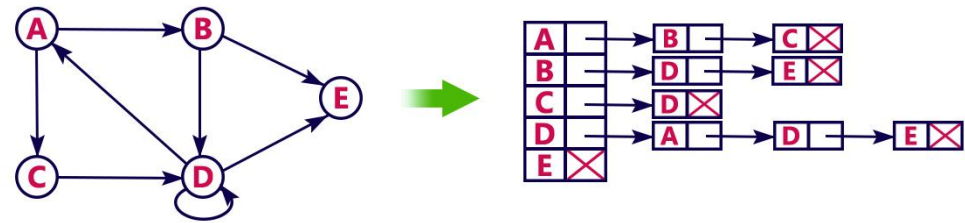
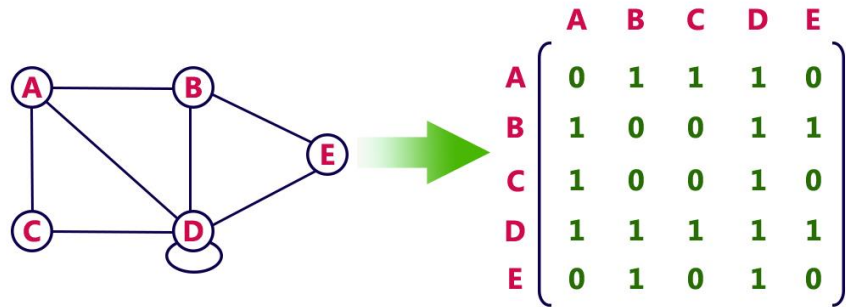
Triadický uzávěr předpokládá, že pokud existují hrany mezi vrcholy B a A a zároveň mezi vrcholy B a C, pak s vysokou pravděpodobností existuje i hrana mezi vrcholy A a C.

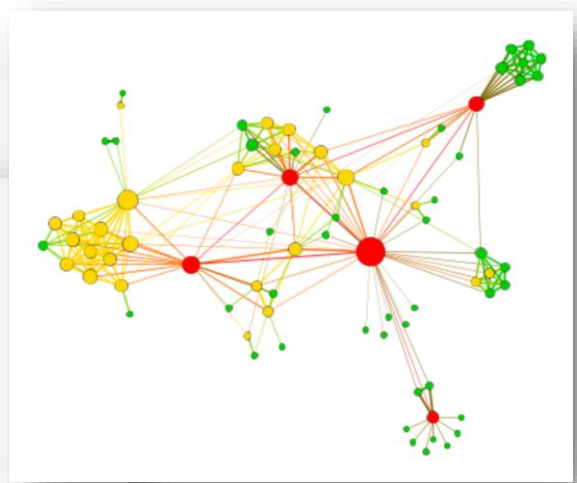
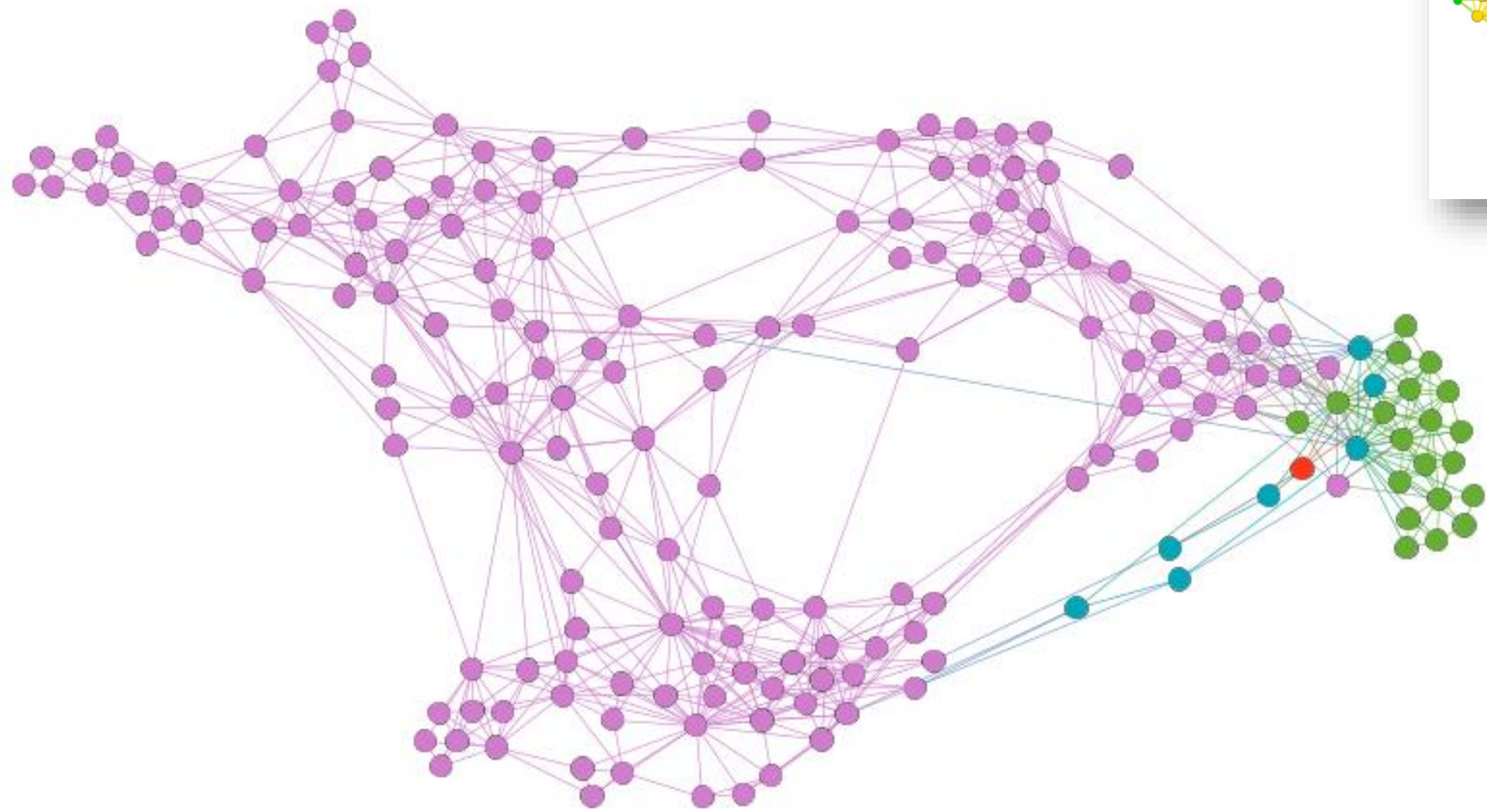
Nový vrchol se v síti spojí s náhodně vybraným vrcholem. Poté se s předem známou pravděpodobností buď připojí k nějakému jeho sousedovi nebo k dalšímu náhodně vybranému vrcholu.



Co potřebujeme pro algoritmy...

...reprezentovat síť v počítači nějakou datovou strukturou





Kontrolní otázky – pojmy

- Síť (graf), vrchol (uzel), vazba (hrana)
- Sled, tah, cesta, průměr
- Centralita, stupeň, closeness, betweenness, Pagerank
- Hustota, shlukovací (clustering) koeficient, komunita
- Bezškálovost (scale-freeness), malý svět (small world), triadický uzávěr
- Core-periphery struktura
- Matice sousedností, seznam (seznamů, polí) sousedností

Úkoly

- Nainstalujte si na své počítače systém Gephi.
- Vyzkoušejte si do Gephi načíst alespoň jednu síť a vizualizovat ji.
- Zjistěte prostřednictvím Gephi některou z probíraných vlastností (např. průměrný stupeň).

Ke čtení

- Barabási, A. L. (2014). *Network Science*. Cambridge University Press (online <http://networksciencebook.com>).
 - Kapitoly 1 a 2.
 - Kapitoly 3, 4, 5, 9 (jen lehoučce).
- Pojmy lze najít i na *Wikipedii*.