

# Network Analysis Methods 2

Network Science

Velké a dynamické sítě

2023/24

# Co je velká (large-scale) síť?

- Zjednodušeně a velmi neformálně řečeno, za velkou síť je možno považovat takovou, kterou je nemožné (nebo obtížné) vizualizovat tak, aby byly na první pohled vidět strukturální detaily.
- Pokud lze síť nějak vizualizovat, při analýze velkých sítí se můžeme vizuálně zaměřit na vlastnosti, které nevyžadují detailní pohled. Jedná se o sítě s tisíci až desetitisíci vrcholů, nicméně překážkou zde může být vysoká hustota sítě (mnohonásobně více hran než vrcholů).
- Pokud není možné síť vizualizovat, musíme se spolehnout na výsledky aplikace výpočetních metod (nebo provést tzv. sampling). Jedná se o sítě se statisíci, miliony až miliardami vrcholů a násobně více hranami.

# O jakých sítích se tu mluví?

- Sociální sítě
- Sítě spoluprací
- Komunikační sítě
- Biologické sítě

# Co můžeme ve velkých sítích hledat?

- Souvislé komponenty
- Kliky nebo struktury podobné klikám (near-clique)
- Hvězdy nebo struktury podobné hvězdám (near-star)
- Velká (globální) centra (hubs)
- Hustě propojené a od okolí oddělené struktury (komunity)

# Největší souvislá komponenta...

- Je taková komponenta, která je “úplná” (neexistuje vrchol mimo komponentu, ke kterému by z komponenty vedla cesta) a...
- ...pokud má síť více komponent, je z nich největší.
- V prostředí reálných sítí je převažující část vrcholů (téměř všechny) právě v největší souvislé komponentě.
- Ve většině případů se při analýze sítí stačí zaměřit na vlastnosti největší souvislé komponenty.

# Struktury blízké klikám

- Klika je struktura s relativně malým množstvím vrcholů (sociologové uvádějí max. 12), v nichž je hranou spojena každá dvojice vrcholů.
- Maximální klikou se rozumí taková klika, která není součástí jiné kliky.
- Jednotlivé vrcholy kliky však mohou být propojeny s dalšími vrcholy mimo kliku (klika pak není izolovaná).
- Často bývají všechny vrcholy kliky propojené s jedním nebo více centry (hubs).

# Struktury blízké hvězdám

- Hvězda je struktura s jedním centrem, které je propojeno s více vrcholy ale tyto vrcholy nejsou propojeny mezi sebou.
- Vrcholy izolované hvězdy (kromě centra) nejsou propojeny s dalšími vrcholy.
- V méně striktních případech může být hvězda jen zčásti (převážná část vrcholů hvězdy je propojena pouze s centrem).
- Za hvězdu se často považuje i situace, kdy centrum je propojeno s mnoha malými klikami (např. o velikosti 2-5).

# Sociální sítě

- Jde o sítě, ve kterých se propojují vrcholy (obvykle lidé) na základě společných vlastností (zájmy apod.).
- Sítě obvykle obsahují velká centra (hubs), ale neobsahují příliš často ani „čisté“ kliky ani „čisté“ hvězdy.
- Velmi rychle se mění v čase a velmi efektivně se v nich šíří informace na velké vzdálenosti.
- Vyskytují se v nich velké komunity (stovky i o mnoho více vrcholů).
- Sítě mohou být vážené, ale je obtížné stanovit jedinou váhu.



# Sítě spoluprací

- Vznikají jako výsledkem přímé interakce (spolupráce) skupin vrcholů.
- Příkladem mohou být spoluautorské sítě, kdy malé skupiny autorů publikují společně.
- Sítě jsou vážené, váha reprezentuje počet spoluprací (např. článků, účasti na společných projektech – filmy apod.).
- Typickým rysem, který je odlišuje od ostatních sítí, je vysoký počet klik a spíše menších (lokálních) center...
- ... a poměrně malé komunity.

# Komunikační sítě

- Jedná se o sítě, které jsou výsledkem interakce dvojic vrcholů (telefonování, posílání mailů, apod.).
- Mohou být podobně jako sítě spolupráce vážené počtem interakcí.
- Velmi záleží na tom, jak jsou sítě vytvořeny.
  - Pokud se zaměříme na interakce dvojic, pak komunikační sítě mají mnoho velkým hvězdám podobných struktur a podobají se spíše sociálním sítím.
  - Pokud bereme v úvahu skupinové komunikace (např. více adresátu v mailech nebo skupinové hovory), pak se komunikační sítě podobají spíše sítím spoluprací.

# Biologické sítě

- Sítě vznikající jako výsledek interakce v rámci nějakého biologického systému (např. mezi neurony, molekulami, enzymy, mezi léky, nemocemi a jejich projevy, genetickými poruchami apod.)
- Bývají (nebo by měly být) vážené, ale je obtížné váhu přesně změřit.
- V mnoha ohledech bývají svou strukturou podobné sociálním sítím (centra, komunity apod.).

# Vývoj sítě v čase

- Dynamic, evolving, temporal networks?
- Jde o sítě, jejichž stav se v čase mění a o procesy, které se v čase realizují.
- V nejjednodušším případě se předpokládají tzv. rostoucí sítě (growing networks), kdy v čase přibývají vrcholy i hrany.
- Jak „evolving“ tak „temporal“ sítě se dají chápat jako dynamické. Liší s v úlohách, které se řeší a které mají jako jeden z parametrů **CAS**.

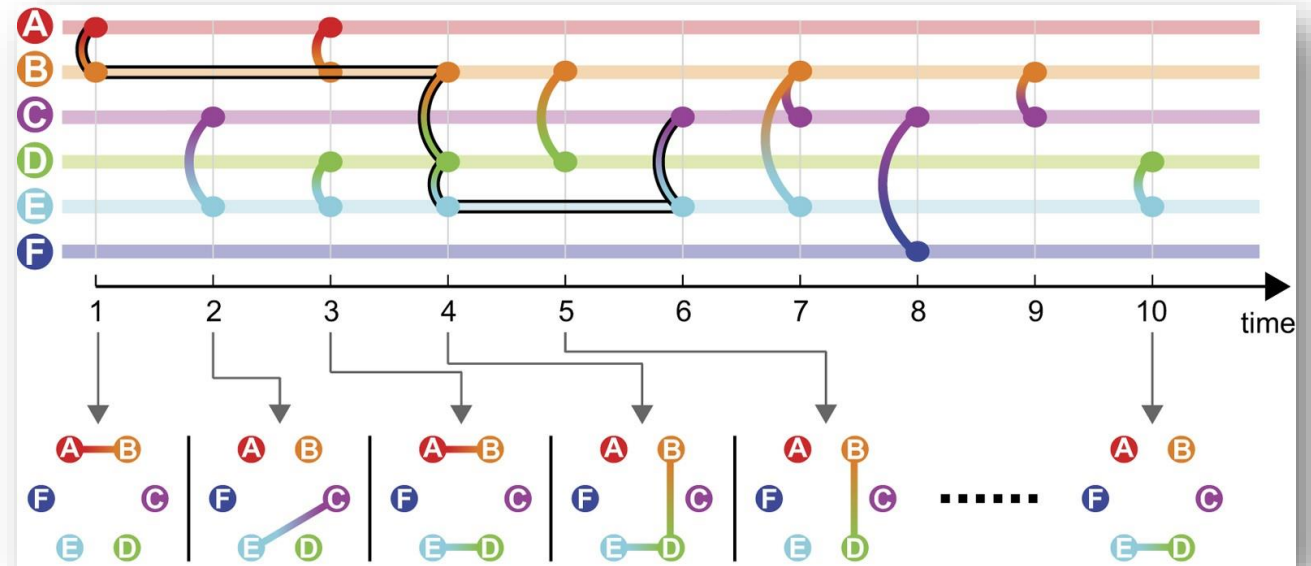
# Dynamické sítě

- Základní charakteristikou je stochastický přístup k analýze.
- Pracuje se pravděpodobností existence (nebo mírou aktivity) vrcholu resp. hrany v daném časovém okamžiku.
- Jevy se často v dynamických sítích modelují pomocí simulace.
- Často se také předpokládá, že vrcholy sítě mohou být různého typu (heterogenní síť).

# Evolving (vyvíjející se) sítě

- Sítě, ve kterých v čase přibývají vrcholy a hrany.
  - V některých případech mohou jak vrcholy, tak hrany ubývat („úmrtí“, „stárnutí“).
  - Obvykle se pracuje s neváženými sítěmi.
- Existují poměrně jednoduché modely popisující principy růstu a umožňující simulace (např. Barabasi-Albert model).
  - Modely jsou zaměřené např. na růst webu (zaměřené na velká centra).
- K přečtení: <http://networksciencebook.com/chapter/6>.

# Temporální síť 1/2



- Také se označují jako „time-varying networks“.
- Jsou charakteristické tím, že hrany existují jen v jistém čase a mohou mít měnící se váhu. To prakticky znamená, že hrany mohou existovat „přerušovaně“, tedy v určitých časových intervalech v síti jsou a v jiných ne.
- K přečtení: <https://arxiv.org/pdf/1108.1780.pdf>

# Temporální sítě 2/2

- S tím, jak se temporální síť v čase mění, souvisí i další vlastnosti.
  - Hustota lokálního propojení.
  - Stupně či jiné centrality vrcholů.
  - apod.
- Obvykle se tyto sítě analyzují tak, že se vybere určitý časový úsek a v něm se temporální síť transformuje na obyčejnou.



# Příklad – spoluautorské sítě

- Autoři článků v určité doméně (např. v informatice) publikují v různých týmech či komunitách a s různou mírou pravidelnosti.
- Mohou publikovat v několika týmech současně.
- Týmy mění svou strukturu, mohou zanikat a vznikat nové, např. podle toho, jakým tématům se věnují.
- Vazby (nebo simplexy) mezi některými autory se posilují opakovaným publikováním, mohou však také slábnout a zanikat.

# Síť pro experimenty (DBLP)

- Spoluautorská síť (informatika)
- *Simplexem* je klika (jedna publikace, na níž má hranu každý s každým spoluautorem)
- <https://www.cs.cornell.edu/~arb/data/coauth-DBLP/>

# Úkoly

- Napište program v C++/C#/Java), který ze zdrojových dat vytvoří definovaný počet časových snímků (sítí měnících se v čase).
- Zjistěte, jak se v čase (v jednotlivých snímcích) mění průměrný stupeň, průměrný vážený stupeň a průměrný shlukovací koeficient sítě.
- Najděte simplex s nejvyšší průměrnou váhou hrany.

# E-mailová síť

- E-mailovou síť můžeme chápat jako síť, ve které zpracováváme mailovou komunikaci.
- Simplexem je vlastník účtu, ostatní vrcholy simplexu jsou ti, kterým vlastník mail odeslal, nebo od kterých mail obdržel.
- Několik emailových sítí (reprezentujících několik různých účtů) lze sloučit do jedné, tzv. komunikační, sítě.

Zehnalova, S., Horak, Z., Kudelka, M. (2015). *Email conversation network analysis: Work groups and teams in organizations*. In 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 1262-1268).

# Firemní komunikace

- Záznam komunikace zaměstnanců jedné firmy mezi sebou, se zákazníky a uživateli za ~5 let.
- Všechna data jsou anonymizována – jména, doména, slova z předmětu mailu (tzv. *stopslova* jsou odstraněna).
- Obsah mailu a přílohy se nezpracovávaly.

```
</Message><Message Sender="name00008@domain0004.cz" Sent="2010-03-15T14:40:28.000Z"  
MessageId="f75bf701-ecd5-48ac-a8ed-b865a0c60230" InReplyTo="0a376a9a-80b9-48c0-9ad3-  
fb2c5db61186">
```

```
<Recipient Type="To">name00188@domain0029.cz</Recipient>
```

```
<Recipient Type="To">name00188@domain0029.cz</Recipient>
```

```
<Recipient Type="To">name00189@domain0029.cz</Recipient>
```

```
<Recipient Type="To">name00189@domain0029.cz</Recipient>
```

```
<Recipient Type="Cc">name00140@domain0029.cz</Recipient>
```

```
<Recipient Type="Cc">name00056@domain0002.cz</Recipient>
```

# Zaměstnanci firmy

- Jeden SW vývojář se třemi mailovými adresami, další dva SW vývojáři, SW vývojář na specifických platformách, CEO, tým leader + architekt.

```
<Accounts>
  <Account smtpAddress="name00001@domain0001.com" />
  <Account smtpAddress="name00002@domain0002.cz" />
  <Account smtpAddress="name00003@domain0003.cz" />
  <Account smtpAddress="name00004@domain0004.cz" />
  <Account smtpAddress="name00005@domain0005.cz" />
  <Account smtpAddress="name00006@domain0004.cz" />
  <Account smtpAddress="name00007@domain0004.cz" />
  <Account smtpAddress="name00008@domain0004.cz" />
  <Account smtpAddress="name00009@domain0004.cz" />
  <Account smtpAddress="name00010@domain0004.cz" />
</Accounts>
```

# Struktura jednoho záznamu (e-mailu)

- Sender: Adresa odesílatele
- Sent: Datum a čas odeslání
- messageId: unique ID (může chybět)
- InReplyTo (nepovinné): messageId předcházejícího mailu v konverzaci (může chybět)
- Subject (optional): Předmět – skupina slov
- Recipient: Adresa příjemce typu "To" nebo "Cc"

```
</Message><Message Sender="name00008@domain0004.cz" Sent="2010-01-13T17:00:48.000Z"  
MessageId="9fce8b1b-2db3-447a-b422-aba74d8e4264" InReplyTo="1f29f3b6-f77e-42a5-b272-  
f37adcf6b50d" Subject="term00053 term00054 term00055 ">  
  <Recipient Type="To">name00045@domain0028.cz</Recipient>  
  <Recipient Type="Cc">name00056@domain0002.cz</Recipient>
```

# Úkoly

- Zkonstruujte temporální síť (z definovaných snímků).
- V síti, která byla konstruována a vyčištěna z několika agregovaných anonymizovaných účtů se pokuste najít významné uživatele (Ega).
  - Slovo „významný“ si definujte např. vysokým stupněm vrcholu.
- V každém snímku sítě spočítejte vlastnosti.