

## Analýza sítě (min. 17, max. 34 bodů)

1. Cílem je použít vybrané metody k analýze reálné sítě s alespoň deseti tisíci vrcholy. Sít musí být
  - a) buď sít přímo stažená z úložiště (např. <https://networkrepository.com/> )
  - b) nebo sít zkonstruovaná z vektorových dat (stažených z např. <https://www.kaggle.com/> , <https://archive.ics.uci.edu/> apod. )
2. Úloha se bude skládat z dílčích kroků:
  - a) Analýza základních strukturálních vlastností sítě vyjádřená buď tabulkou v případě globálních vlastností, grafem distribuce hodnot dané lokální vlastnosti a vizualizací.
    - i. Počet vrcholů, hran, hustota.
    - ii. Průměrný stupeň, maximální stupeň, distribuce stupňů.
    - iii. Centrality.
    - iv. Shlukovací koeficient CC – průměr. Dále graf pro tzv. shlukovací efekt. Shlukovací efekt se určí jako průměrný CC pro vrcholy daného stupně (distribuci průměrného CC (osa Y) vůči stupni (osa X)).
    - v. Souvislost - počet souvislých komponent a distribuce jejich velikostí.
    - vi. Vizualizace (včetně zvýraznění center/hubů).
  - b) Analýza komunitní struktury (alespoň dva různé algoritmy).
    - i. Počet komunit, průměrná, minimální a maximální velikost. Hodnota modularity.
    - ii. Distribuce velikostí komunit.
    - iii. Vizualizace podle komunitní struktury.
3. Použité nástroje - R, Python, Gephi nebo/a jejich kombinace.
4. Výstup bude ve formě textového dokumentu (**pouze PDF!**). Bude obsahovat popis datového souboru (kde a jak byl získán, co obsahuje atd.), výsledky analýzy a interpretaci výsledků (co znamenají).
5. Deadline je 16.12.2024, předpokládaný termín prezentací výsledků analýzy po písemném testu. Pdf soubor zašlete na mou emailovou adresu. V předmětu mailu bude „MAS1 analýza váš\_login“, soubor bude pojmenován stejně.