



Detekcia komunití v sieťach

(Community detection in networks)

Zadanie z EA
2020/2021

Modelovanie systémov grafom

- Niekedy systém môže mať tvar siete, kedy grafová reprezentácia je prirodzeným nástrojom
 - Vrcholy grafu sú hlavné objekty/zložky systému
 - Interakcie medzi zložkami sú reprezentované pomocou hrán grafu
- Cieľom je odhaľovať
 - Topologické atribúty grafu (napr. distribúcia stupňa vrcholov, priemer grafu,...)
 - *Skryté štruktúry* obsiahnuté v grafe

Príklady sietí

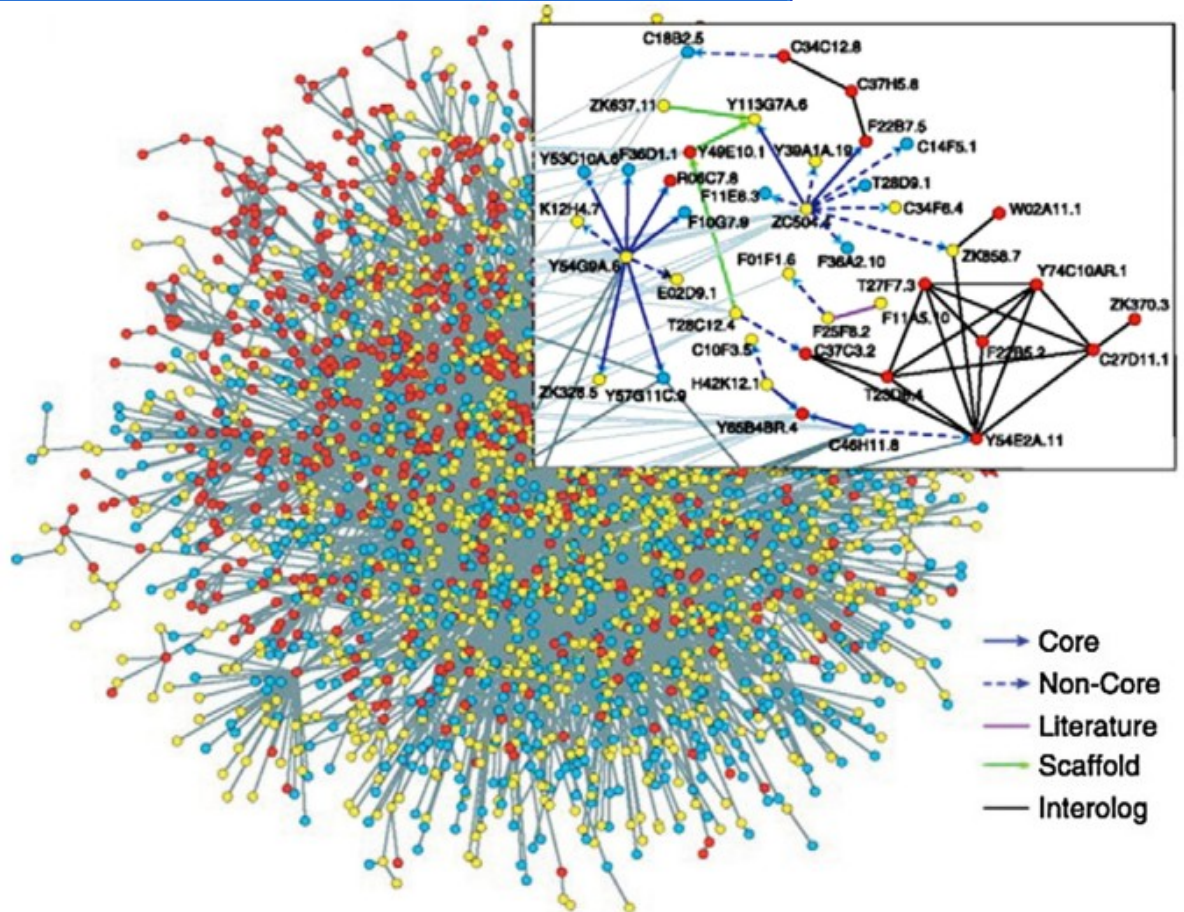
- Sociálne siete
 - Citačné siete
 - Komunikačné siete (e-mail, facebook)
- Biologické siete
 - Chemické interakcie medzi proteínmi v bunkách
- Technologické siete
 - Transportné siete

Komunity v sieťach

- Komunita je podmnožina vrcholov grafu
 - Niektoré oblasti používajú pojmy ako modul, komplex
- Elementy komunity majú
 - Silný vzťah a veľa spoločných charakteristík k iným elementom danej komunity
 - Minimálny vzťah a málo spoločných charakteristík k elementom iných komunít

Príklad biologickej siete

- C.elegans protein interaction



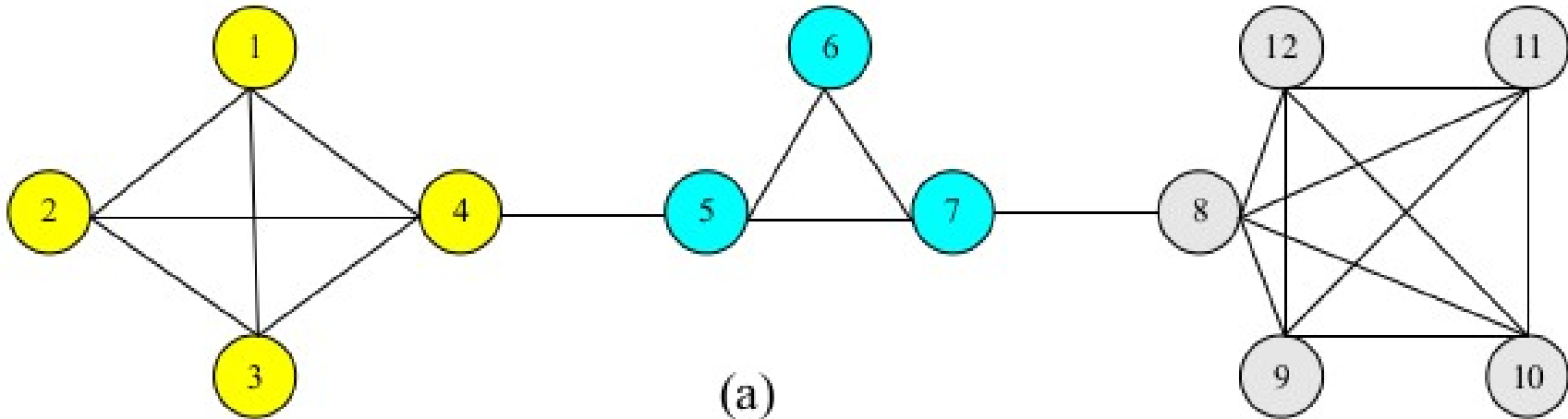
- © R. Albert, Scale-free networks
in cell biology, J. Cell Sci. 118 (21) (2005)

Definícia problému (1)

- Použijeme zjednodušenú podobu problému
- Systém reprezentovaný grafom $G(V,E)$, kde
 - $V = \{v_1, \dots, v_n\}$ je množina vrcholov
 - $E = \{e_1, \dots, e_m\}$ je množina hrán, E je podmnožina $V \times V$
- Hrany e_i grafu
 - Sú neorientované
 - Nie sú vážené (resp. každá hrana má jednotkovú váhu)

Definícia problému (2)

- Úlohou je pomocou evolučného algoritmu rozdeliť vrcholy grafu do disjunktných podmnožín so
 - silným prepojením vo vnútri podmnožín
 - slabým prepojením medzi podmnožinami



Testovacia úloha - zadanie

- Použitý graf 115 uzlov, 616 hrán (približne) a neznámy počet komunít
- Použitý formát GML (Graph Modelling Language)
 - `node [id X label Y value Z]`
 - `edge [source id target id]`
- Vstupné dáta boli anonymizované (Y="?", Z=?)

Testovacia úloha - výstup

- Požadovaný výstup riešenia = rovnaký formát GML ako vstup
 - Namiesto hodnoty Z (aktuálne ?) bude id komunity
 - Komunity identifikujte číselne: 0, 1, 2, ...
- Odovzdávate
 - Zdrojový text programu
 - Popis riešenia – štruktúra algoritmu, reprezentácia, vhodnosť, realizácia použitých blokov algoritmu, nastavenie parametrov – v detailnosti umožňujúcej replikáciu (formát pdf)
 - Najlepšie riešenie
 - Popis podielu členov skupiny na riešení – presne kto sa podieľal na čom

Hľadanie komunití

- Hľadanie komunití je typ zhukovacieho problému
 - NP-hard kombinatorický optimalizačný problém
- Používajúci
 - Optimalizáciu modularity
 - Betweenness score
 - Community score
 - ...