**4. Tesztelés**

Ebben a fejezetben az algoritmusok eredményeit hasonlítjuk össze NMI pontosság és futásidő szerint. Az algoritmusok teljesítményeit valós világbeli és mesterségek hálózatokon értékeljük ki. A teszteléshez legfeljebb néhány száz csúcsból álló gráfokat használunk az algoritmusok komplexitása miatt.

A tesztek egy Intel(R) Core(TM) i5-1035G1 CPU @ 1.00GHz 1.19 GHz processzorral, 8,00 GB memóriával ellátott számítógépen, Windows 10 operációs rendszeren lettek elvégezve.

**4.1. NMI pontosság**

A klaszterezések kiértékeléséhez az egyik jól ismert teljesítmény mérő az *NMI (normalized mutual information)*. Az NMI az optimális közösségek és az algoritmus által talált közösségek hasonlóságát minősíti. Adott G hálózat optimális közösségeinek halmaza legyen C(A), és az adott algoritmusból kapott közösségeinek halmaza legyen C(B).

ahol n a hálózat pontjainak száma, CA az alapigazság közösségek száma, CB a talált közösségek száma, nij azon i valós alapigazság lévő csúcsok száma, amelyek megtalálhatók a j talált közösségben is , ni az i alapigazság közösségben lévő csúcsok száma, nj a j talált közösségben lévő csúcsok száma.

Az NMI értéke 0 és 1 közötti valós szám, minél jobban hasonlít az adott algoritmus által visszaadott közösségi hozzárendelés az alapigazságra, annál nagyobb. Abban az esetben, ha a talált közösségek azonosak az alapigazsággal, az NMI egy, ha teljesen különböznek, akkor nulla.

A pontosság mérésére elkészítettem Python nyelven a saját implementációmat nmi.py néven. Az nmi(A,B) függvény két paramétert vár, az egyik az alapigazságot, a másik a talált közösségeket tartalmazó DAT fájl. Az input fájlokban a közösségszerkezet él és a hozzárendelt közösség párok listájaként írják le. A program beolvassa a fájlokat, kiszámítja a talált közösségszerkezet NMI értékét, és ezt az értéket adja vissza.

**4.2. Alapigazság gráfok**

Léteznek közösségkereső algoritmusok tesztelésére használt valós életbeli standard teszt gráfok, amelyeknek ismerjük a közösségszerkezetét. A teszt hálózatok megtalálhatók Mark Newman oldalán [] GML formátumban. A gráfokat a NetworkX csomag függvényeivel alakítottam át.

Zachary karate klub [karate] hálózata egy amerikai egyetem karate klubjának barátságait jelöli. Tudjuk, hogy a 34 tagból álló klub egy konfliktus következtében két külön klubra bomlott, az algoritmusok által talált közösségszerkezetet ehhez tudjuk hasonlítani.

Dolphin social network

A Les Miserables hálózat csúcsai a Nyomorultak című regény szereplői. Ha két szereplőt összeköt él, az azt jelenti, hogy a szereplők megjelennek ugyanabban a fejezetben.

College Football gráf [fball] LP 0.6046 labdarúgó mérkőzéseket reprezentál, ahol a csapatok a gráf csúcsainak felelnek meg, és két csapat között akkor helyezkedik el él, ha rendszeresen mérkőztek meg egymással a szezonban. Az alapigazság közösségszerkezet a csapatokat konferenciákra osztja. Általában több meccs kerül megrendezésre konferencián belül, mint konferenciák között.

Books about U.S. politics [books] hálózat pontjai a az amazon.com-ról vásárolt amerikai politikáról szóló könyveket reprezentálják, az élek a gyakran együtt vásárolt könyveket kötik össze. Newman szétosztotta a könyveket liberális és konzervatív könyvek osztályára. A könyvek egy részéről nem tudta egyértelműen eldönteni, melyik ideológiához tartoznak. LP 0.5272

**4.3. Gráfgyár bemutatása (röviden)**

Benchmark graphs for testing community detection algorithms

A mesterséges hálózatainkat a gráf generátorral [] állítjuk elő.

A program három fájlt készít. A network.dat a generált hálózat éllistáját tartalmazza soronként, a csúcsokat egész számokkal jelöli 1-től kezdve, növekvő sorrendben. Minden él kétszer szerepel. A community.dat-ban a hálózat csúcsainak listája szerepel, a csúcsokhoz hozzárendelve az őket tartalmazó közösséget. A közösségeket is egész számokkal jelöli, 1-től indítva a számozást. A statistics.dat

paraméterek