

Véletlen séta alapú eljárás átfedő közösségek felderítésére

Borzási Henrietta

II. évf. programtervező informatikus MSc

Témavezető: Dr. Vinkó Tamás

SZTE TTIK Számítógépes Optimalizálás Tanszék

Napjainkban mindenkit és mindent körülvesznek a hálózatok. Gondolhatunk például az internetre, vállalatok elhelyezésére vagy akár betegségek terjedésére, különféle hálózatok mindenhol megjelennek. Azt is megfigyelhetjük, hogy a valós hálózatokban sokszor csoportosulások, úgynevezett közösségek alakulnak ki. Például egy vállalat dolgozói egymással szorosabb kapcsolatban vannak, mint más vállalat dolgozóival, ezáltal a vállalatok diszjunkt közösségekként is felfoghatók. Viszont egy másik szemszögből tekintve a problémát egy személy többnyire nem egy, hanem több közösségbe is beletartozik, hiszen egyaránt tagja a családtagjai, a munkahelye, vagy a hobbija által reprezentált közösségeknek. Ily módon egy szociális hálóra tekinthetünk úgy, mint átfedő közösségek halmazára. Ezek alapján jogosan vetődik fel a kérdés: hogyan lehetne detektálni a közösségeket? Általánosságban elmondható, hogy az egy közösségbe eső csúcsok között szorosabb kapcsolat van, azaz egy alkalmas metrika szerint közel vannak egymáshoz, míg a közösségek egymástól távolabb helyezkednek el. Z. Jiang et al. 2016-os cikkükben¹ javasoltak egy diszjunkt közösségeket detektáló algoritmust, amelynek alapötlete az, hogy a TSP algoritmus a hálózat csúcsait olyan sorrendben látogatja meg, hogy előbb az egy közösségbe esőket járja be, utána lép a következő közösségre, és így tovább. Az algoritmus kulcsa olyan távolságfogalom meghatározása, ami jól reprezentálja két csúcs kapcsolatát, és egy alkalmas küszöb megadása, ami szerint a TSP által kapott utat szétvághatjuk, így azonosítva a csoportosulásokat. Ezt a gondolatmenetet visszük tovább. Diszjunkt közösségek detektálásához nincs szükség a TSP algoritmusra; kizárólag a távolságmátrix alapján végezhetünk vágásokat egy alkalmas küszöb mentén, melynek eredményeképp egyértelműen beazonosíthatók a közösségek. Átlapoló közösségek esetén nem ilyen egyszerű a kérdés, hiszen az átlapoló csúcsok megszüntetik a közösségek közötti nagy távolságokat. Ebből kifolyólag egyértelmű, hogy csupán a csúcsok közötti távolságok vizsgálata nem elég átfedő csoportosulások meghatározásához. A dolgozatban megvizsgáljuk, mi történik utóbbi esetben, és milyen módosítások vihetnek közelebb az átlapoló közösségek megtalálásához. Eredményeinket a kiindulási cikkben felvetett algoritmussal, illetve a rendkívül hatékony CFinder átlapoló közösségkereső algoritmussal hasonlítjuk össze futásidő illetve pontosság szempontjából.

¹ Zhongzhou Jiang, Jing Liu, Shuai Wang, Traveling salesman problems with PageRank Distance on complex networks reveal community structure, *Physica A: Statistical Mechanics and its Applications*, Volume 463, 2016, Pages 293-302, ISSN 0378-4371