

3. félévi beszámoló

Sulyok Bendegúz (bsulyok94@gmail.com)

Statisztikus fizika, biológiai fizika és kvantumrendszerek fizikája PhD program

Témavezető: Pollner Péter (pollner@angel.elte.hu)

A dolgozat címe: Beágyazások GRS optimalizációja.

Bevezetés

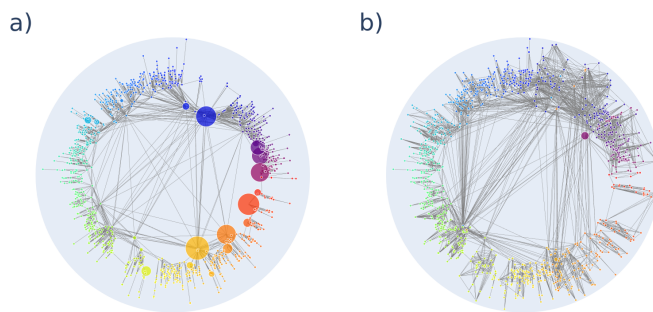
A félév során folytattam a hiperbolikus térbeli beágyazási módszerek fejlesztését. A korábban megkezdett, greedy routing mértékek direkt optimalizálásán alapuló eljárás mellett most egy másik módszerrel dolgozom mely a csoportkeresési és beágyazási problémák dualitását hivatott feltárni: topologikus csoportok ismeretében ágyazok be gráfokat.

Kutatás

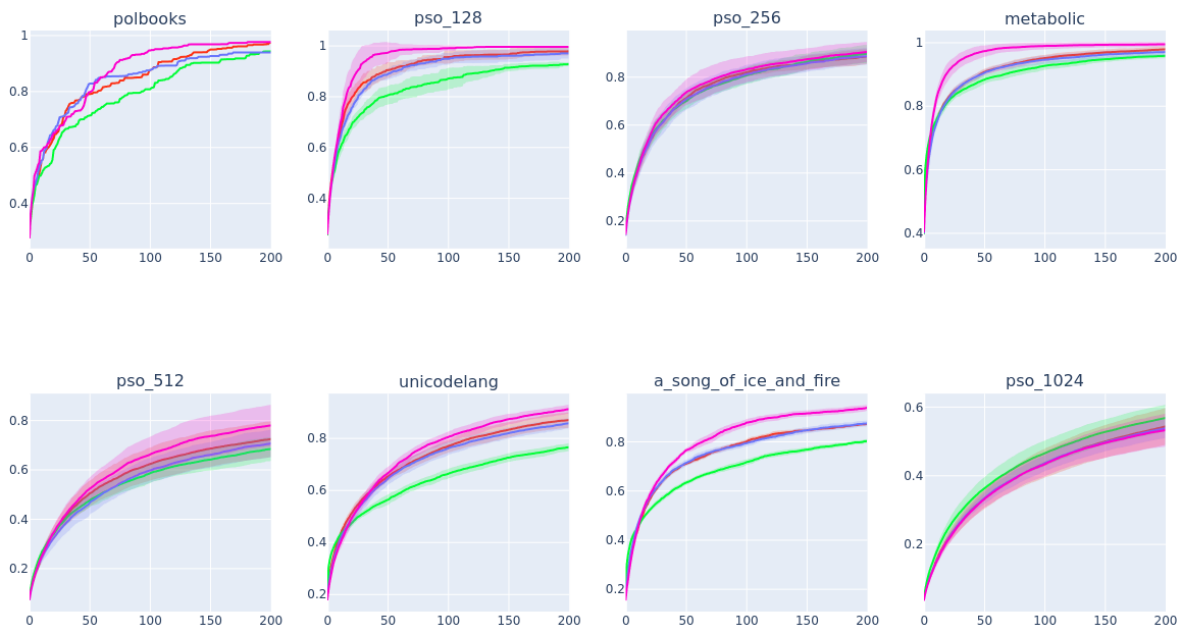
Greedy útvonalak beágyazása

A greedy útvonalak sikerességi arányára optimalizálok simulated *Simulated Annealing*-gel. Greedy útvonalról egy olyan gráf esetében beszélhetünk, melynek gráfpontjait valamilyen metrikus térbeli koordinátákkal láttuk el, ekkor egy kiindulási és egy cél pont közötti greedy útvonal az az egyedi útvonal, mely során egy walker (pl. üzenet, csomag) mindig az aktuális pontból a célhoz legközelebbi szomszédos pontba lép. Amennyiben a walker egy hurokba szorul (visszalép ugyanabba a pontba, ahonnan érkezett), az útvonalkeresés sikertelen.

A szimuláció röviden összefoglalva a következő: minden iterációban kiválasztok egy gráfpontot, a környezetében kiválasztok egy véletlenszerű térbeli pontot és megnézem, hogy az új pozíció javítja-e a teljes gráf bejárhatóságát greedy útvonalak mentén. Amennyiben nettó javulás tapasztalható, a módosítást elfogadom. A módosítást véges nagy valószínűséggel



1. ábra. Kezdeti (a) és szimuláció utáni (b) beágyazása ugyanannak a gráfnak (N=1024).



2. ábra. A szimuláció lefolyása nyolc választott gráfra.

akkor is elfogadom, ha az rontja a gráf bejárhatóságát, de ez a valószínűség csökken, ahogy a virtuális *hőmérséklet* paraméter csökken a szimuláció során.

A 1 ábra egy konkrét példát mutat a a szimuláció előtti és utáni beágyazására. Az eredeti beágyazás a Mercator algoritmussal ([1]) készült, ami egy gyors és ebben a kontextusban megfelelő beágyazási módszer, a kutatás során ezt veszem alapvonalnak. A gráfpontok mérete indikálja, hogy az adott pontban mennyi greedy útvonal akad el, a színezés mindösszesen esztétikai célú. Jól látható, hogy míg a kezdeti beágyazásban számos nagyobb pont található, ezek a szimuláció során eltűnnek.

A greedy útvonalak sikerességi arányát kiszámolni intenzív folyamat: $O(N^2 \log N) + O(N^2 < k >)$, ahol N a gráf pontjainak száma, $< k >$ pedig az átlagos élszám. Mivel közvetlenül erre a mennyiségre optimalizálok, minden iteráció során ki kell számolni ezt. Ennek érdekében számos optimalizációs lépést végrehajtottam, de még így is jelentős a futási idő, ezért viszonylag kicsi gráfokra alkalmazható csak a módszer.

A 2. ábra mutatja az optimalizáció lefolyását nyolc választott gráfra. A különböző színek négy különböző algoritmus teljesítményét ábrázolja (a különbség a mozgásra kiválasztott gráfpont mintavételezési eljárása).

Ez a kutatás már a végső stádiumban jár, rövidesen benyújtjuk a kéziratot.

Csoportok beágyazása

Hierarchikus csoportkereső segítségével megkeresem a gráfban fellelhető csoportokat, majd a hierarchia különböző szintjein végig haladva fokozatosan ágyazom be a csoportokat, ilyen módon csökkentve az eljárás komplexitását. Mivel a csoportkeresés egy igen sokat publikált probléma, számos algoritmus kínálkozik, ami a beágyazás alapját szolgáltathatja, egy konkrét példa erre a Leiden algoritmus [2].

Publikációk

Az előző szemeszter során részt vettem a [ai4covid CXR hackathon](#) versenyen, az itt elért eredményekből írt kézirat jelenleg elbírálás alatt áll, várhatóan *Automated prediction of COVID-19 severity upon admission by chest X-ray images and clinical metadata aiming at accuracy and explainability* címen fog megjelenni.

Tanulmányi tevékenység

A félév során a következő tárgy(ak)at végeztem el:

- *Klaszterezés hálózatokkal* (FIZ/3/064E),
- *Környezeti áramlások fizikája* (FIZ/3/017E).

Oktatási tevékenység

A félév során mérésvezetőként vettem részt a Modern fizika laboratórium (fizlab3f19la / ff1c4s13) lebonyolításában, ahol idén a *Spektrofotometria* és a *Diffúzió* mérésekért voltam felelős.

Az *Adatmodellek és adatbázisok* (dsmodelsf20vm) tárgyból két gyakorlatot tartottam *noSQL* témákból.

További szakmai tevékenység

Folytattam korábbi munkámat a *Semmelweis Egyetemenél* (SE), ahol vastagbél szövetmintákon daganatos zónákat azonosítunk mesterséges intelligencia segítségével.

Az európai CounteR projekt keretében tovább dolgoztam szociális hálók analitikai módszereinek fejlesztésén. A félév során elsősorban adatgyűjtést és gráf neurális hálók vizsgálatát és implementációját végeztem.

A *Broad Institute* rendezésében részt vettem a [Cancer Immunotherapy Data Science Grand Challenge](#)-en, ahol génmódosítások (CRISPR-Cas9 knockout) kimenetelét kellett megjósolni. A verseny rövid lefolyású és intenzív volt, sok hasznos tapasztalatot szereztem.

Hivatkozások

- [1] Guillermo García-Pérez, Antoine Allard, M Ángeles Serrano, and Marián Boguñá. Mercator: uncovering faithful hyperbolic embeddings of complex networks. *New Journal of Physics*, 21(12):123033, dec 2019.
- [2] V. A. Traag, L. Waltman, and N. J. van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233, Mar 2019.