

Egyéni kutatómunka 1.

Beszámoló

Mikulás Zsófia

0.1. Bevezetés

A kutatómunkám során pythonos programozás segítségével vizsgáltam agygráfokban a különböző területek közötti maximális mértékű összeköttetéseket.

A gráfokat a <https://braingraph.org/download-pit-group-connectomes/> honlapról töltöttem le. Elsőként a '86 nodes set, 1064 brains, 1 000 000 streamlines, 10x repeated averaged (18 MB)'-ba tartozókat vizsgáltam, majd legutóbb irányított gráfokra ('Directed graphs, 423 brains, 20 000 streamlines (99 MB)') is futtattam az algoritmust.

0.2. Az agygráfok

Az agygráfokat az úgynevezett diffúziós tenzoros képalkotási módszerrel (DTI), illetve traktográfia segítségével hozták létre.

A DTI olyan speciális MRI-módszer, amellyel rekonstruálható a vízmolekulák háromdimenziós mozgása. A DTI során extra, meghatározott irányú mágneses gradienseket kapcsolnak be, amelyek irányában a vízmolekula mozgása mérhető. Elegendő gradiens esetén (minimum hat, de minél több gradiens van, annál pontosabban írható le a mozgás) a háromdimenziós mozgás rekonstruálható. Az agy fehérállományában a vízmolekulák a rostokkal párhuzamosan szabadon tudnak mozogni, de azokra merőlegesen csak limitáltan (a sejtmembránok barriert képeznek). Az adott voxelben (3D-s pixel) ezért a vízmolekulák fő diffúziós iránya feltételezeten párhuzamos a voxel által lefedett agyállományban haladó rosttal. A voxelekben rekonstruált diffúziós mozgások alapján ún. traktográfiát, vagy rostkövetést végezhetünk. Az agy egy bizonyos területéről elindított traktográfia során a diffúziós irányoknak megfelelően a voxeleket a rostok lefutásának megfelelően összekapcsoljuk, és így az MR-képen kirajzolható például a corticospinalis pálya lefutása.

Az általam először vizsgált agygráfok megalkotásához az agy szürkeállományában 83 csúcsot jelöltek ki, az ezek között futó rostszálakat térképezték fel a fent említett módszerrel, ezek lettek a gráf élei, amelyekhez a rostszálak mennyiségét rendelték hozzá kapacitásként. Az adathalmaz létrehozásakor nem vették figyelembe a rostok futási irányát, viszont a mérések bizonytalansága miatt azokat tízszer elvégezték, majd a legkisebb illetve legnagyobb értékeken kívül a mért rostszámok átlagát rendelték az egyes élekhez kapacitás gyanánt. A másik adathalmazban lévő gráfokat, amelyekkel dolgoztam, egyszeri méréssel hozták létre, azonban egyrészt nagyobb felbontású, 1015 csúcsos ráccsal dolgoztak, illetve figyelembe vették a rostok futásának irányát is.

0.3. Az adathalmaz feldolgozása, algoritmus futtatása

A létrehozott agygráfok külön fájlban vannak eltárolva. Az egyes csúcsok sorszámokkal vannak jelölve, a hozzájuk tartozó adatok közül az agyfélteki (jobb vagy bal), illetve agyterületi elhelyezkedést használtam fel. Emellett el vannak tárolva a nem nulla kapacitású élek tulajdonságai, amik közül a kezdő és végpontjára, illetve a d11 azonosítóval eltárolt átlagos rostszámára volt szükség.

Tehát a fájlokat egyesével be kellett olvasni, kinyerni a megfelelő adatokat, vagyis az egyes csúcsok területi elhelyezkedését, a közöttük futó élek kapacitásértékeit, majd a 83x83-as kapacitásmátrixot létrehozni. Mivel a később használt algoritmus irányított gráfokon fut, viszont az első adathalmaz irányítatlan gráfra vonatkozik, ezeknél az éleket mindkét irányítással hozzá kellett adni.

A maximális folyamat a különböző jobb- és baloldali magok (Caudate, Putamen, Pallidum, Accumbens-area) és nyolc lebenyterület között (jobb és bal: temporal, occipital, parietal, frontal) kellett nézni. Az utóbbiakhoz több csúcs is

tartozott, amelyeknek az azonosítása során segítségemre volt a http://uratum.com/diversity/Figure_S1-Krona.html oldal. Az algoritmus két megadott csúcs között számolja a folyamértékeket, így az oda futó maximális folyamok kiszámításához hozzávettem egy-egy plusz csúcsot és az egyes lebenyekhez tartozó csúcsokat megfelelően nagy (100000-es) kapacitással hozzákötöttem, és aztán ezekre a csúcsokra futtattam az algoritmust. Tehát a kapacitásmátrixot kibővítettem ezekhez a csúcsokhoz tartozó nyolc-nyolc sorral és oszloppal és a megfelelően magas kapacitású élekkel.

A nyolc-nyolc vizsgálni kívánt csúcs között futó maximális folyamok kiszámításához Edmonds-Karp algoritmust használtam, amihez GitHub-ról leszedtem egy python-kódot (<https://github.com/anxiaonong/Maxflow-Algorithms/blob/master/Edmonds-Karp%20Algorithm.py>), amin lényegében nem változtattam. Ezután létrehoztam egy 8x8-as mátrixot, amibe eltároltam a magok és a lebenyek közötti maximális folyamok értékeit. Az irányítatlanság miatt mindegy, hogy az algoritmust a magokból a lebenyek irányába, vagy fordítva futtatjuk, azonban a későbbi okok miatt a lebenyekből a magokba menő maximális folyamokat számoltam. Végül a kapott adatokat kiírtam egy txt-fájlba.

0.4. Eredmények

A kód első értelmes eredményei olyan mátrixok voltak, amelyeknek minden sora ugyan a nyolc értéket tartalmazta, vagyis az egy adott mag és bármely lebenyterület közötti maximális folyamértékek megegyeztek. Ennek egyik oka az volt, hogy amikor az egy lebenyhez tartozó csúcsokat összecsíptem, az abszolút csúcsba menő nagy kapacitású éleket szintén irányítatlanként, azaz mindkét irányítással hozzáadtam. Ez azért eredményezte ugyanazokat az értékeket egy oszlopon belül, mert ezzel, ha az adott magból az adott lebeny valamelyik csúcsába eljut valamennyi

információ, az abszolút csúcson keresztül eljut az a mennyiség, tehát a maximális folyam legalább ugyanannyi, vagyis egy adott magból az egy lebenyhez tartozó minden csúcsba ugyanannyi lesz a maximális folyam, ami viszont az eredeti gráfban nem teljesül. Ezt módosítva, tehát a lebenyek összecspésének számító csúcsba irányított nagy kapacitású élek kivételével, a kapott maximális folyamokat tároló mátrixok sorai azonban még mindig nagyon hasonló értékeket tartalmaztak.

A gráfok élhálózatát vizsgálva látható (vagy biológiában jártasság esetén tudható), hogy a talamuszon (az agygráfok adataiban 35-ös, illetve 76-os sorszámú csúcsok) szinte minden keresztülhalad, vagyis kis túlzással minden út, ami a magok és a lebenyek között fut, áthalad a jobb- vagy a baloldali talamuszon, ráadásul a magok és a talamuszhoz tartozó csúcsok között futó utak nagyrészt függetlenek. Így ami valamelyik mag és egy adott lebenyhez tartozó csúcs között fut, az, vagy legalábbis egy jelentős hányada, eljut a többi lebenybe is.

102513	R-Caud.	R-Put.	R-Pall.	R-Ac.	L-Caud.	L-Put.	L-Pall.	L-Ac.
r-tem.	19627.0	17858.0	4871.5	4.25	11152.125	11152.125	5063.25	160.625
r-occ.	16816.0	16816.0	4871.5	4.25	11152.125	11152.125	5063.25	160.625
r-par.	24425.5	17858.0	4871.5	4.25	11152.125	11152.125	5063.25	160.625
r-fr.	24425.5	17858.0	4871.5	4.25	11152.125	11152.125	5063.25	160.625
l-tem.	11152.125	11152.125	4871.5	4.25	17030.25	17030.25	5063.25	160.625
l-occ.	11152.125	11152.125	4871.5	4.25	19459.875	19459.875	5063.25	160.625
l-par.	11152.125	11152.125	4871.5	4.25	32747.5	21159.375	5063.25	160.625
l-fr.	11152.125	11152.125	4871.5	4.25	32747.5	21183.625	5063.25	160.625

Tehát nem meglepő, hogy a mátrixok oszlopaiban ennyi azonos érték szerepel. Viszont nem erre számítottunk, ezekből az adatokból nem lehet olyan következtetéseket levonni, amilyeneket szerettünk volna, így az algoritmust lényegesen nagyobb felbontású, irányított gráfokra is lefuttattuk. Ehhez át kellett írni a kódokat az

újabb adathalmaz kezelésére. Mivel ezekben a fájlokban kevésbé konzisztensen szerepeltek a szükséges adatok (például nem volt egységes a csúcsok elhelyezkedését jelző azonosító), az adatok beolvasására vonatkozó kódon is jelentősen változtatni kellett. Emellett figyelni kellett arra, hogy bizonyos élek irányítatlanként voltak feltüntetve, ezeket mindkét irányítással, ugyanazzal a kapacitással hozzá kellett adni a gráfhoz, majd ebből létrehozni a megfelelő kapacitásmátrixot. Ugyanazok között az agyi területek közötti maximális folyamra voltunk kíváncsiak, és hasonlóan a korábbiakhoz, itt is összecsíptem az egy lebenyhez tartozó csúcsokat egy-egy plusz csúcs hozzáadásával, megfelelően nagy kapacitású élek segítségével.

Itt már számított, hogy a magokból a lebenyekbe, vagy fordítva futó maximális folyamot számoljuk-e, tehát meg kellett vizsgálni, hogy az összecsípett lebenycsúcsokból ennek megfelelően kifelé, vagy befelé kell-e irányítani az éleket. Mindkét irányban lefuttatva azt kaptam, hogy a magokból a lebenyekbe az érték szinte mindig 0, míg a lebenyekből a magokba értelmes mennyiség kerültek a 8x8-as mátrixba (ez a biológiai háttér miatt várható is volt).

	R-Caud.	R-Put.	R-Pall.	R-Ac.	L-Caud.	L-Put.	L-Pall.	L-Ac.
r-tem.	60	190	893	309	484	112	181	720
r-occ.	60	190	893	309	484	112	181	720
r-par.	60	190	893	309	484	112	181	720
r-fr.	60	190	893	309	484	112	181	720
l-tem.	60	190	893	309	484	112	181	720
l-occ.	60	190	893	309	484	112	181	720
l-par.	60	190	893	309	484	112	181	720
l-fr.	60	190	893	309	484	112	181	720

A kapott eredménymátrixokban azonban itt is azonosak lettek a sorok, és eddig ezekkel az értékekkel sem dolgoztunk tovább, részben időhiány miatt is.