

4. félévi beszámoló

Kovács Bianka (bianka.kovacs@ttk.elte.hu)

Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája PhD program

Témavezető: Palla Gergely

Bevezetés

Munkánk során a komplex hálózatok topológiája és a hiperbolikus tér geometriája közötti kapcsolat [1] különböző vonatkozásait vizsgáljuk: egyrészt hiperbolikus beágyazást (vagyis a hálózat pontjainak a hiperbolikus térben történő, a hálózati topológiát jól tükröző elrendezését) megvalósító algoritmusokat, másrészt pedig hiperbolikus geometriára építő hálózatgeneráló modelleket.

Az előző három félévben elért kutatási eredmények összegzése

Elsőként kiegészítettük a dimenzióredukciós módszerek közé sorolható, noncentered minimum curvilinear embedding (ncMCE) [2] nevű hiperbolikus beágyazó algoritmust egy, a beágyazásnak az E-PSO modellre [3] vonatkozó likelihoodjának maximalizálását végző, a szögkoordinátákat optimalizáló eljárással. A likelihood kiszámításához szükséges modellparaméterek meghatározására kidolgoztam egy szintén likelihood maximalizáción alapuló eljárást a szokásosan alkalmazott komplikált procedúrák helyett. Az E-PSO modellben az ún. internal link-ek számát szabályozó L paraméter értelmezési tartományát kiterjesztettem a negatív számokra is, ezáltal képessé téve a modellt arra, hogy az eredetileg nem összekötött pontok idővel bekövetkező összekötődése mellett korábban megjelent élek megszűnését is szimulálja. Az így kapott beágyazó módszer nem csak a likelihood, hanem a greedy routing score [2] szempontjából is jobbnak bizonyult az eredeti ncMCE-nél és sok esetben a tisztán likelihood optimalizáción alapuló HyperMap-nál [3] is, valamint csoportstruktúrával rendelkező hálózatok esetén olyan elrendezéseket generál, amelyek az ncMCE-hez képest jobban kihangsúlyozzák a csoportokat.

Készítettünk továbbá egy részletes elemzést a popularity-similarity optimisation (PSO) [4], az E-PSO [3,5] és az S^1/\mathbb{H}^2 [6] modellekkel generált, a pontok szög szerinti elrendezésében homogén hálózatok csoportstruktúrájára vonatkozóan. Ebben a detektálható csoportstruktúrák modularitását, szög szerinti szétválását, különböző csoportkereső algoritmusok eredményeinek összeegyeztethetőségét, a csoportméreteloszlást, illetve mindezeknek a különböző modellparaméterektől való függését vizsgáltuk. Eredményeink szerint az erős csoportszerkezet kialakulásának nem szükséges feltétele, hogy egy hálózat pontjai szög szerint különböző csoportokba tömörüljenek a hiperbolikus síkon.

Végezetül kidolgoztam a hiperbolikus síkon kisvilág tulajdonsággal, skálafüggetlen fokszámeloszlással, magas átlagos klaszterezettségi együtthatóval és erős csoportstruktúrával rendelkező hálózatokat generálni képes PSO modell tetszőleges $2 < d$ dimenziós hiperbolikus térre történő kiterjesztését, mely elősegítheti a magasabb dimenziós beágyazó módszerek [2] fejlesztését. A beágyazások szempontjából alapvető fontosságú eredmény, hogy a radiális koordinátákat meghatározó popularity fading paraméter értéke a korábbi feltételezésekkel [2] ellentétben nem csak a fokszámeloszlás hatványkitevőjétől, hanem a dimenziószámtól is függ.

Az aktuális félévben elvégzett kutatások ismertetése

A félév során elsősorban a d -dimenziós PSO modellt bemutató kéziratban dolgoztam. A modell analitikus elemzésének némi kiegészítésével bekapcsolódott a munkába Balogh Gáspár Sámuel doktorandusz is, a kézirat megszövegezésén így már hárman dolgozunk. A korábbi tervekkel ellentétben az E-PSO modellt, vagyis a PSO modell azon általánosított változatát, amelyben a pontok megjelenéskori élszáma a pontok sorszámától függ, végül nem vizsgáljuk ebben a kéziratban, mivel a dimenzióknak ebben az általánosított modellben betöltött szerepének

igazolásához még további analitikai levezetésekre van szükség, melyeket inkább már egy következő publikációhoz szeretnénk felhasználni.

Új témaként irányított hálózatok hiperbolikus beágyazásával kezdtem el foglalkozni. Az irányított hálózatokat euklideszi térben elhelyező ún. High-Order Proximity preserved Embedding (HOPE) [7] algoritmus a hálózat pontjaira felírt egyfajta aszimmetrikus távolságmátrixnak a szinguláris érték felbontásával (SVD) kapott mátrixok alapján a hálózat minden pontjához hozzárendel két helyvektort, amelyek közül az egyik a kérdéses pontnak mint a hozzá kapcsolódó élek kiindulópontjának (source) a pozícióját, a másik pedig a kérdéses pontnak mint a hozzá kapcsolódó élek végpontjának (target) az elhelyezkedését adja meg. Szintén távolságmátrixon végrehajtott SVD-n alapszik két, irányítatlan hálózatokra kidolgozott hiperbolikus beágyazó módszer, az Isomap (ISO) és a noncentered Isomap (ncISO) [2]. Mindkettő a hálózat pontjai közötti legrövidebb úthosszakat tartalmazó szimmetrikus mátrixból indul ki, utóbbi ezt, előbbi pedig ennek a centrált formáját bontja fel $\mathbf{D} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{U}^T$ valamint $\mathbf{D} = \mathbf{X} \cdot \mathbf{X}^T$ alakban, amely képletek összeegyeztetése alapján a szögkoordinátákat megadó mátrixot $\mathbf{X} = \mathbf{U} \cdot \sqrt{\mathbf{\Sigma}}$ módon számítják ki. Ehelyett a HOPE-nál aszimmetrikus mátrixokra alkalmazott $\mathbf{D} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T$ valamint $\mathbf{D} = \mathbf{X}_{\text{source}} \cdot \mathbf{X}_{\text{target}}^T$ alakú felbontásokat használva irányított hálókra felírt legrövidebb úthossz mátrixokból is kiszámíthatóak a hiperbolikus síkon vett szögkoordináták az $\mathbf{X}_{\text{source}} = \mathbf{U} \cdot \sqrt{\mathbf{\Sigma}}$ és az $\mathbf{X}_{\text{target}} = \mathbf{V} \cdot \sqrt{\mathbf{\Sigma}}$ képletekkel.

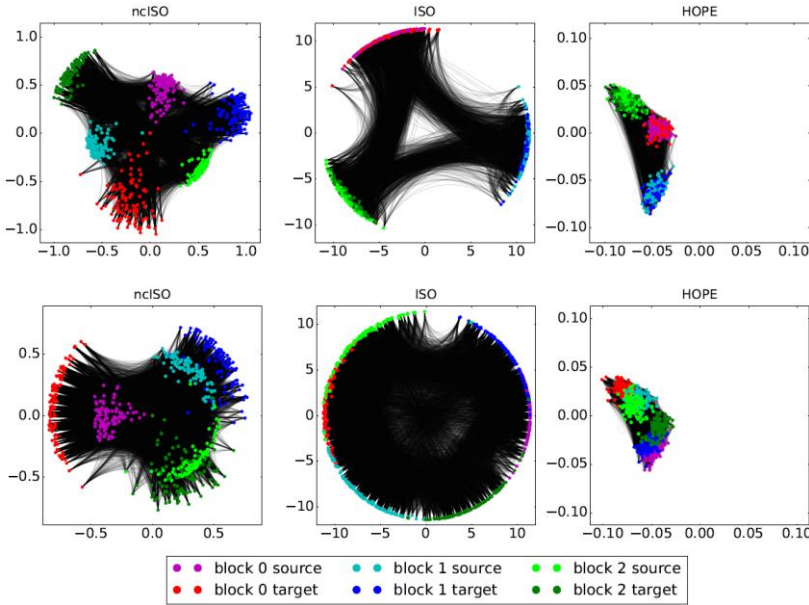
A radiális koordináták meghatározása az eredeti ISO és ncISO módszernél is független a szögkoordinátáktól. Mindkettő esetén a PSO modellnek megfelelően feltételezték, hogy minél több élt gyűjtött össze egy pont a hálózat növekedése során, annál korábban jelent meg a hálózatban, így annál közelebb van a beágyazáshoz használt hiperbolikus korong origójához. Ez alapján az i pontnak a fokszámok csökkenő sorrendjében vett sorszámát ID_i -vel, a fokszámeloszlás hatványkitevőjéből becsült popularity fading paramétert β -val, a hálózat pontjainak számát N -el és a hiperbolikus sík görbületének abszolút értékének négyzetgyökét ζ -val jelölve az (1) képlettel írható fel az i pont r_i radiális koordinátája [2]. Irányított hálózatok esetén csakúgy, mint szög szerinti pozícióból, radiális koordinátából is kettőt rendelünk minden ponthoz: egyet source-ként, egyet pedig target-ként. Előbbi értéke a kimenő élek száma szerinti pontsorrenden és a kifokok eloszlásának hatványkitevőjén, utóbbié pedig a bemenő élek száma szerinti pontsorrenden és a befokok eloszlásának hatványkitevőjén alapul.

$$r_i = \beta \cdot \frac{2}{\zeta} \cdot \ln(ID_i) + (1 - \beta) \cdot \frac{2}{\zeta} \cdot \ln(N) \quad (1)$$

Az 1. ábrán látható, hogy a szokásostól eltérően az irányított ncISO a sűrűn összekötött ponthalmazokat egymással átellenesen, tehát szög szerint egymástól távol helyezi el, mintha a PSO helyett egy olyan modellből származna a beágyazott hálózat, amelyben a kötési valószínűség a hiperbolikus távolságnak *növekvő* függvénye. Ennek megfelelően az irányított ncISO esetén egy fordított radiális sorrendet alkalmaztam, amelyben a nagy fokszámú pontokhoz nagy radiális koordinátát rendeltem, mintha a hálózat növekedése során az origóhoz egyre közelebb jelentek volna meg a pontok. Emellett az (1) képletben a pontoknak a popularity fading-et szimuláló kifelé történő mozgásából adódó $+(1 - \beta) \cdot (2/\zeta) \cdot \ln(N)$ tagot az ncISO módszer esetén elhagytam azt feltételezve, hogy a radiálisan fordított irányú hálónövekedés esetén a pontok befelé mozognak és az utolsóként érkező (azaz az N -edik) pont radiális koordinátája $(2/\zeta) \cdot \ln(N)$ helyett éppen 0.

Ugyan a [2] cikkben még csak 2- és 3-dimenziós beágyazásokat készítettek az ISO és az ncISO módszerekkel és feltételezték, hogy a β popularity fading paraméter csak a fokszámeloszlás γ hatványkitevőjétől függ és a dimenziók d számától nem, én a jellemzően magas dimenziós (pl. 100) beágyazások elkészítésére használt HOPE algoritmussal való jobb összevethetőség érdekében úgy implementáltam az ISO és az ncISO irányított hálókra alkalmazható változatát, hogy tetszőleges dimenziós hiperbolikus beágyazásokat el lehessen készíteni velük, valamint a

d -dimenziós PSO modellre vonatkozó eredményeink ismeretében $\beta^{\text{in/out}} = \frac{1}{(d-1) \cdot (\gamma^{\text{in/out}} - 1)}$ alakban számítottam ki a popularity fading paramétereket.



1. ábra: Sztochasztikus blokk modellel (SBM) generált assortatív (fent) és diszassortatív (lent) csoportstruktúrájú egy-egy irányított hálózat különböző beágyazásai. A pontok színezése mutatja, hogy a hálózat egy adott pontjának a source vagy a target pozíciója van-e az adott helyen, valamint hogy az SBM-be betáplált 3 blokk közül melyiknek a tagja az adott pont.

A beágyazó algoritmusok teljesítményének összevetéséhez a greedy routing score-nak az irányítatlan hálózatok 2- vagy 3-dimenziós hiperbolikus beágyazásának kiértékelésére kidolgozott implementációját [2] módosítottam úgy, hogy irányított hálózatok tetszőleges dimenziós hiperbolikus és euklideszi beágyazásánál is használható legyen. A hiperbolikus távolságok mérését a d -dimenziós PSO modellben már használt (2) képlettel oldottam meg. Az ncISO esetén módosítottam a greedy routing léptetési szabályát: itt azt vizsgáltam, hogy mennyire sikeres az átjutás a hálózat bármely két pontja között, ha egy adott pontból mindig a célponttól hiperbolikusan *legtávolabbi* szomszédra lépünk – ez az ncISO-ból kapott beágyazások GR-score-ját jelentősen javította. Egy 803 pontból és 24729 élből álló e-mail hálózat [8] esetén például a HOPE által elért GR-score csupán 0.28 lett, ami elmarad mind az ncISO 0.58-os, mind pedig az ISO 0.79-os eredményétől.

$$x(\underline{u}, \underline{v}) = \frac{1}{\zeta} \cdot \operatorname{arccosh} \left(\cosh(\zeta r_u) \cdot \cosh(\zeta r_v) - \sinh(\zeta r_u) \cdot \sinh(\zeta r_v) \cdot \frac{\underline{u} \cdot \underline{v}}{r_u r_v} \right) \quad (2)$$

Elkezdtem ezen kívül hálózatoknak a különböző beágyazások alapján történő rekonstruálhatóságát is vizsgálni. Ehhez az összes ismert élt felhasználva beágyaztam hálózatokat mindhárom módszerrel, majd a HOPE esetén a helyvektorok skaláris szorzatának a csökkenő sorrendje szerint, az ISO és az ncISO esetén pedig a pontok közti hiperbolikus távolságnak a növekvő illetve csökkenő sorrendje szerint sorba rendeztem a lehetséges pontpárok egy véletlenszerűen kiválasztott halmazát, amelyben 1000 összekötött pontpár szerepelt. A példaként felhozott e-mail hálózat egy-egy beágyazásából így kapott pontpár sorrendben az első 1000 helyen szereplő pontpárból a HOPE esetén 980, az ISO esetén 416, az ncISO esetén pedig 887 db volt valóban összekötve, míg az éleket csupán a vizsgált mintában érvényes élsűrűség alapján véletlenszerűen megjósolva a helyes találatok száma 38 volt.

Tanulmányi tevékenység az aktuális félévben

A félév során a Biológiai Doktori Iskola által meghirdetett alábbi tárgyat végeztem el:

- Számítógépes modellezés a biológiában (BIO/10/32G)

Oktatási tevékenység az aktuális félévben

A Biológiai Fizika Tanszék által meghirdetett *Modern fizika laboratóriumi gyakorlatok* című tárgy *Spektrofotometria* mérését vezettem 5 alkalommal.

Publikációk

A félév során az ncMCE módszerrel előállított hiperbolikus beágyazásoknak az E-PSO modellre vonatkozó likelihood maximalizációján alapuló optimalizálásáról szóló kézirat *Optimisation of the coalescent hyperbolic embedding of complex networks* címmel **megjelent** a *Scientific Reports* folyóiratban.

Emellett a hiperbolikus geometrián alapuló, egyenletes szög szerinti eloszlást alkalmazó E-PSO és S^1/\mathbb{H}^2 modellekkel generált hálózatok csoportszerkezetét bemutató munkánkról szóló kéziratot *The inherent community structure of hyperbolic networks* címmel beküldtük a *Scientific Reports* folyóiratba, jelenleg az első körös bírálatok nyomán elvégzett módosításaink elfogadását várjuk.

A PSO hiperbolikus síkbeli hálózatgeneráló modell tetszőleges dimenziós térre történő kiterjesztéséről szóló kézirat *Growing hyperbolic networks beyond two dimensions: the generalised popularity-similarity optimisation model* címmel várhatóan hamarosan felkerül az arXiv weboldalra.

Konferenciák az aktuális félévben

A félév során beküldött absztraktom alapján lehetőséget kaptam, hogy júliusban szóbeli prezentáció formájában bemutassam az E-PSO és az S^1/\mathbb{H}^2 hálózatok csoportszerkezetével foglalkozó munkánkat a *Networks 2021* konferenciához tartozó *Satellite meeting on Communities in Networks* című online tudományos rendezvényen.

Konferenciák a teljes képzés alatt

- 2019. december, Lisszabon: *Complex Networks 2019 – The 8th International Conference on Complex Networks and their Applications* – itt poszter prezentációt tartottam
- 2020. április, online: *OSA Biophotonics Congress: Biomedical Optics* – itt csak hallgatóként voltam jelen
- 2020. szeptember, online: *NetSci 2020* – itt poszter prezentációt tartottam
- 2020. december, online: *Complex Networks 2020 – The 9th International Conference on Complex Networks and their Applications* – itt szóbeli prezentációt tartottam
- 2021. július, online: *Communities in Networks* – itt szóbeli prezentációt fogok tartani

Hivatkozások

- [1] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, Marián Boguñá, *Hyperbolic geometry of complex networks*, Physical Review E, 2010, 82, 036106
- [2] Alessandro Muscoloni, Josephine Maria Thomas, Sara Ciucci, Ginestra Bianconi, Carlo Vittorio Cannistraci, *Machine learning meets complex networks via coalescent embedding in the hyperbolic space*, Nature Communications, 2017, 8, 1615
- [3] Fragkiskos Papadopoulos, Constantinos Psomas, Dmitri Krioukov, *Network Mapping by Replaying Hyperbolic Growth*, IEEE/ACM Transactions on Networking, 2015, 23, 198-211
- [4] Fragkiskos Papadopoulos, Maksim Kitsak, M. Ángeles Serrano, Marián Boguñá, Dmitri Krioukov, *Popularity versus similarity in growing networks*, Nature, 2012, 489, 537-540
- [5] B. Kovács, G. Palla, *Optimisation of the coalescent hyperbolic embedding of complex networks*, Scientific Reports, 2021, 11, 8350
- [6] M. Ángeles Serrano, Dmitri Krioukov, Marián Boguñá, *Self-Similarity of Complex Networks and Hidden Metric Spaces*, Phys. Rev. Lett., 2008, 100, 078701
- [7] M. Ou, P. Cui, J. Pei, Z. Zhang, W. Zhu, *Asymmetric transitivity preserving graph embedding*, Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016, 1105-1114
- [8] <http://snap.stanford.edu/data/email-Eu-core.html>