

Bevezetés

A hálózatelmélet egyik, jelenleg intenzíven kutatott témaköre a komplex hálózatok topológiájának kapcsolata a hiperbolikus tér geometriájával. Mivel a legtöbb hálózat rendelkezik egyfajta hierarchikus struktúrával (pl. dominancia viszonyok, vagy akár a különböző attribútumokkal rendelkező pontokat magukba foglaló kategóriák bennfoglalási hierarchiája) és az ezt leíró faszzerű struktúra egy negatív görbületű (hiperbolikus) tér diszkrétizált változatának tekinthető, így adja magát az az elképzelés, miszerint a hiperbolikus beágyazással (azaz a hálózat pontjainak a hiperbolikus térben történő elhelyezésével) a hálózat egyfajta természetes reprezentációja állítható elő, amelyben a pontok térbeli elrendeződése a hálózat mögött rejlő hierarchiának megfelelő [1]. A hiperbolikus beágyazás a hálózat pontjainak egy olyan elrendezését eredményezi, amely a hálózati topológiát jól tükrözi, ugyanis a hiperbolikus térben az egymással összekötött, illetve a hasonló szomszédokkal rendelkező pontok egymáshoz közel helyezkednek el, valamint a hálózat csoportjai jól kirajzolódnak. Következésképp a hálózat hiperbolikus térbeli elrendeződésének, a pontok közti hiperbolikus távolságoknak a figyelembevétele elősegítheti a hiányzó élek predikcióját, vagy akár a hálózat csoportjainak azonosítását is. Hálózatok hiperbolikus beágyazására többféle módszer is létezik már [2,3,4], de ezek fejlesztése illetve újabb módszerek kidolgozása folyamatosan zajlik.

A hiperbolikus geometria relevanciájára utal az a megfigyelés is, hogy a hiperbolikus síkot reprezentáló Poincaré-korongon véletlenszerűen leszórt pontok között a hiperbolikus távolságtól függő (a hiperbolikusan közeli pontokra nagy) valószínűséggel létrehozva éleket olyan hálózati topológiák állíthatóak elő, amelyek a valós hálózatoknál általánosan megfigyelt három tulajdonsággal is rendelkeznek, úgymint a kis-világ tulajdonság, a skálafüggetlen fokszámoszlás és a magas átlagos klaszterezettség [1]. Mostanra már több olyan modell is létezik, amely a hiperbolikus síkban generál hálózatot, ilyen többek közt a *popularity-similarity optimization (PSO)* modell [5] és az S^1 modell [6]. Kutatásunk során a hiperbolikus beágyazási módszerek fejlesztése mellett a hiperbolikus geometriára építő hálózatmodellekben rejlő további lehetőségek feltárása is célunk.

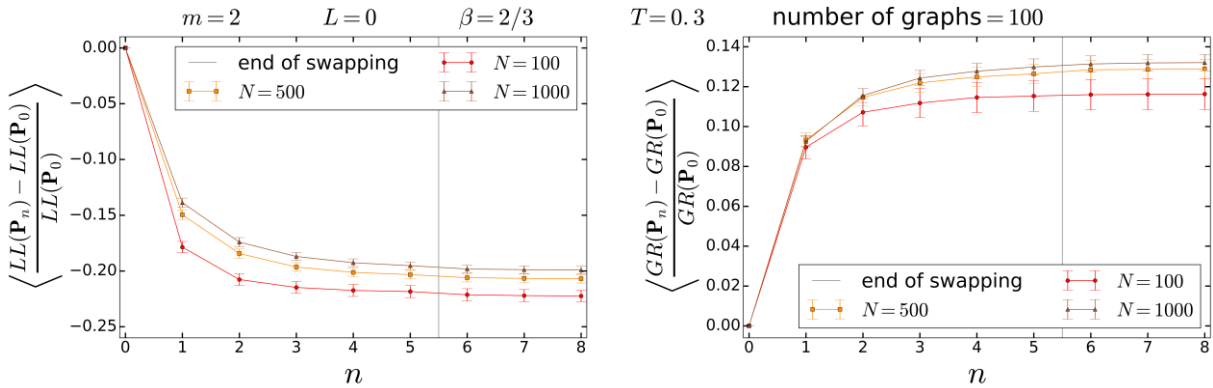
Az aktuális félévben elvégzett kutatások ismertetése

Munkánk során kidolgoztunk egy likelihood [2] maximalizáción alapuló eljárást, amellyel a Poincaré-korongra elhelyezni kívánt hálózat pontjainak az ún. *noncentered minimum curvilinear embedding (ncMCE)* beágyazó módszerből [3,4] kapott szögkoordinátáit optimalizáltuk úgy, hogy a beágyazott hálózat végső hiperbolikus elrendezése a lehető legnagyobb mértékben feleljen meg az E-PSO modell [2] által generált azonos szomszédsági mátrixú hálózaténak.

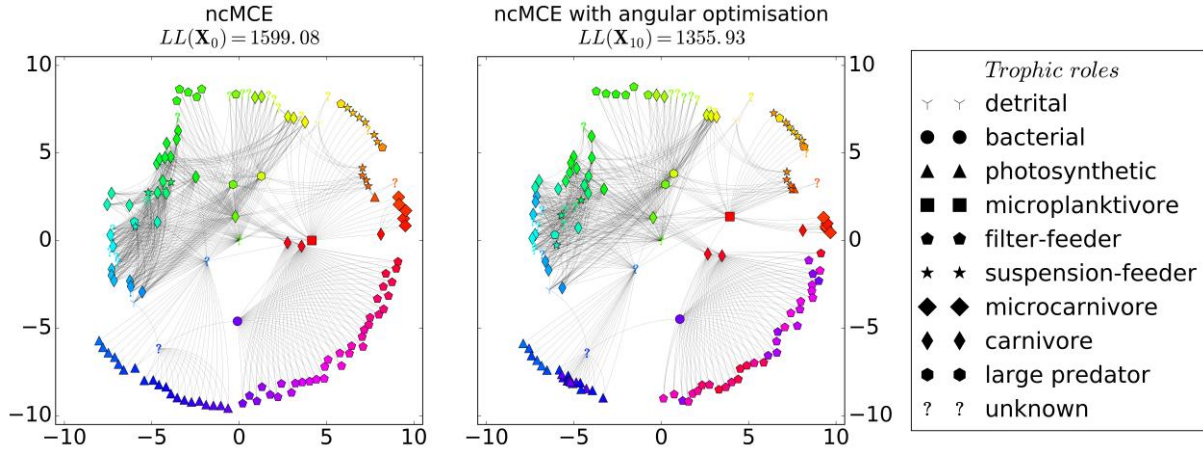
Az ideális beágyazást elsőként a *HyperMap* algoritmusban [2] próbálták az E-PSO-ra vonatkozó likelihood maximalizálásával megtalálni. A módszer alapja az a feltételezés, hogy mivel az E-PSO reális hálózatokat generál, így tekinthetünk a beágyazandó valós hálózatunkra úgy, mintha azt az E-PSO modell generálta volna – ekkor pedig a beágyazás során

értelemszerűen azt az elrendezést kell megkeresni, amelynek a valószínűsége (likelihoodja) az E-PSO modellben a legnagyobb. Mivel azonban a likelihood maximalizálása a szögkoordinátákra nézve analitikusan nem megoldható, a HyperMap viszonylag lassú: egy N pontos háló beágyazásához $O(N^3)$ számítási lépésre van szüksége. Ezzel szemben a gyors, dimenzióredukción alapuló ncMCE módszerből kapott szögkoordinátákból indítva a likelihood maximalizálását elegendő a teljes $[0, 2\pi)$ szögtartomány helyett a kiindulási szögkoordináták körüli leszűkített régiót feltérképezni a hálózat minden pontjára. Feltételezve, hogy az ncMCE-ből kapott szögkoordináták viszonylag jók, de lehetővé téve a pontok szög szerinti sorrendjében jelentkező esetleges hibák javítását a szöghorrekciós algoritmusunk a hálózat pontjainak kezdetben az aktuális másodsomszédok, majd – mikor ez már nem okoz jelentős likelihoodnövekedést – az aktuális közvetlen szomszédok közötti elmozdulását engedi meg. A radiális koordináták szerinti növekvő sorrendben $n \ll N$ -szer végighaladva a hálózat pontjain és minden lépésben a beágyazás likelihoodját az adott pont $q \ll N$ tesztpozíciójában mintavételezve a likelihood maximalizáláshoz szükséges futásidő csupán $O(N^2)$ -es.

Méréseink alapján már $n \approx 10$ korrekciós kör és ezek mindegyikében minden pontra $q \approx 6$ tesztpozíció vizsgálatának elvégzése elegendő ahhoz, hogy akár 20%-ot csökkenjen a beágyazás negatív log-likelihoodja (*logarithmic loss*) az ncMCE eredményéhez képest. Figyelemre méltó, hogy az 1. ábra tanulsága szerint az E-PSO-ra vonatkozó likelihood növekedésével egyidejűleg a beágyazott hálózat keletkezéséről semmit nem feltételező, a *greedy routing*-nak [7] az adott hiperbolikus pozíciók esetén a hálózat egyes pontpárjai közötti átlagos hatékonyságát jellemző ún. *GR-score* [4] is javul (nö) a szögek optimalizálása során. A 2. ábra egy valós hálózat [8] hiperbolikus beágyazását mutatja a szögek korrekciója előtt (vagyis az ncMCE módszer esetén) és után, eszerint a likelihood maximalizálása a hálózat csoportjainak szög szerinti jobb szétválását is elősegíti.



1. ábra: A beágyazás E-PSO modellre vonatkozó negatív log-likelihoodjának (bal) és GR-score-jának (jobb) relatív javulása a végrehajtott korrekciós körök n számának függvényében különböző méretű, az E-PSO modellel generált hálózatokra. A függőleges vonal előtti körökben a szög szerinti másodsomszédok, míg utána a szög szerinti közvetlen szomszédok pozíciója közötti 6 pontban mintavételeztem a likelihoodot a hálózat összes pontjára. A generáláshoz használt E-PSO modell paraméterei a következők: a hálózat pontjainak száma N , az átlagos fokszám $2 \cdot (m + L)$, a fokszámeloszlás hatványfüggvényesen lecsengő végén a hatványkitevő abszolút értéke $\gamma = 1 + 1/\beta$, a klaszterezettséget meghatározó hőmérséklet T .



2. ábra: Egy tápláléklánc [8] ncMCE módszerből kapott (bal) és a szögek optimalizálása utáni (jobb) elrendezése a Poincaré-korong natív reprezentációjában. Az E-PSO modellre vonatkozó likelihood maximalizálásával már 10 szögkorrekciós kör alatt határozottabbá vált a hálózatban található csoportok szög szerinti szeparációja.

Az E-PSO modellre vonatkozó likelihoodnak a beágyazási minőség mérőszámaként való alkalmazását eddig nagyban nehezítette, hogy kiszámításához előzetesen ki kell választani az E-PSO modell megfelelő paramétereit. Ez főleg a β és a T paraméterek esetén volt problémás: előbbit a fokszámeloszlásra – kis hálózatoknál igen nehézkesen – illeszthető hatványfüggvény kitevője, utóbbit pedig egy viszonylag komplikált eljárás során, több különböző T érték mellett kapott beágyazáson a hiperbolikus távolság függvényében megmért összekötési valószínűségekre illesztett egyenes meredekségének az elméleti értéktől való eltérését minimalizálva becsülték meg [2]. Az általunk bevezetett módszer ennél lényegesen egyszerűbb: az ncMCE módszerből kapott, a likelihood maximalizálás paramétereiről független szögkoordináták mellett történő, a koordináták helyett a paraméterekre vonatkozó likelihood maximalizálással határozzuk meg β és T értékét. Ezen kívül felismertük, hogy az m paraméter szokásos becsült értéke (a hálózatban előforduló minimális foksám) jellemzően eltér a likelihood függvénynek az m szerinti maximumhelyétől, emiatt az m paraméter esetén is áttértünk a likelihoodalapú becslésre.

A hiperbolikus geometriára építő hálózatgeneráló modellek jelenleg mind a 2-dimenziós hiperbolikus térben működnek. A félév során elkezdtem kidolgozni az E-PSO modell tetszőleges d -dimenziós hiperbolikus térbe történő kiterjesztését. A kettőnél magasabb dimenziós terekben az eredő szögtávolság meghatározásának nehézsége miatt már nem használható a hiperbolikus koszinusztétel a két pont közötti hiperbolikus távolság meghatározására. Azonban az ún. *Poincaré ball* modellben [9] érvényes (1) távolságképlet tetszőleges dimenziós térben könnyen alkalmazható – itt $x(\underline{u}, \underline{v})$ a hiperbolikus teret reprezentáló gömb középpontjába eső origóval rendelkező Descartes-féle koordináta-rendszerben egy \underline{u} és egy \underline{v} helyvektorú pont közötti hiperbolikus távolság, $\|\cdot\|$ pedig az euklideszi normát jelöli. Így az E-PSO modellben a hiperbolikus sík *natív* reprezentációjában megadott radiális koordinátákról áttérve a Poincaré ball-beli radiális koordinátákra tetszőleges d -dimenziós hiperbolikus térben leszórt hálózati pontok esetén megoldhatóvá válik a hiperbolikus távolságok mérése és az ezektől függő összekötési valószínűségek alkalmazása. A következő félévben tervezem elvégezni a d -dimenziós modell részletes analitikus és szimulációs vizsgálatát.

$$x(\underline{u}, \underline{v}) = \operatorname{arccosh} \left(1 + 2 \cdot \frac{\|\underline{u} - \underline{v}\|^2}{(1 - \|\underline{u}\|^2) \cdot (1 - \|\underline{v}\|^2)} \right) \quad (1)$$

Publikációk

Jelenleg szerkesztés alatt áll az ncMCE módszerrel előállított hiperbolikus beágyazásoknak az E-PSO modellre vonatkozó likelihood maximalizációján alapuló optimalizálásáról szóló kézirat.

Tanulmányi tevékenység az aktuális félévben

A félév során a PhD programomban meghirdetett alábbi két tárgyat végeztem el:

- Fejlődésbiológiai mechanizmusok kvantitatív modelljei (FIZ/3/056E)
- Adatbányászat és gépi tanulás (FIZ/3/084)

Konferenciák az aktuális félévben

Decemberben részt vettem egy konferencián (*The 8th International Conference on Complex Networks and their Applications*) Lisszabonban, ahol poszter prezentáció formájában bemutattam az ncMCE módszerekből kapott hiperbolikus beágyazásoknak a likelihood maximalizációjára irányuló szögkorrekciójával elért eredményeimet.

Oktatási tevékenység az aktuális félévben

A Biológiai Fizika Tanszék által meghirdetett *Modern fizika laboratóriumi gyakorlatok* című tárgy háromórás *Spektrofotometria* mérését vezettem 11 alkalommal. A Czirók András által írt, 2013 óta használatban lévő mérésleírást a félév során frissítettem, a hallgatói visszajelzések alapján átdolgoztam.

Hivatkozások

- [1] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, Marián Boguñá, *Hyperbolic geometry of complex networks*, *Physical Review E*, 2010, **82**, 036106
- [2] Fragkiskos Papadopoulos, Constantinos Psomas, Dmitri Krioukov, *Network Mapping by Replaying Hyperbolic Growth*, *IEEE/ACM Transactions on Networking*, 2015, **23**, 198-211
- [3] Carlo Vittorio Cannistraci, Gregorio Alanis-Lobato, Timothy Ravasi, *Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding*, *Bioinformatics*, 2013, **29**, 199-209
- [4] 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
- [5] Fragkiskos Papadopoulos, Maksim Kitsak, M. Angeles Serrano, Marián Boguñá, Dmitri Krioukov, *Popularity versus similarity in growing networks*, *Nature*, 2012, **489**, 537-540
- [6] M. Angeles Serrano, Dmitri Krioukov, Marián Boguñá, *Self-Similarity of Complex Networks and Hidden Metric Spaces*, *Phys. Rev. Lett.*, 2008, **100**, 078701
- [7] Marián Boguñá, Dmitri Krioukov, K. C. Claffy, *Navigability of complex networks*, *Nature Phys.*, 2009, **5**, 74–80
- [8] Dunne J. A., Williams R. J., Martinez N. D., Wood R. A., Erwin D. H., *Compilation and Network Analyses of Cambrian Food Webs*, *PLoS Biol*, 2008, **6**(4), e102
- [9] M. R. Bridson and A. Haefliger, *Metric Spaces of Non-Positive Curvature*, Springer-Verlag, Berlin, 1999