

Bevezetés

Munkánk során a komplex hálózatok topológiája és a hiperbolikus tér geometriája közötti, a hálózatok egyfajta hierarchikus struktúráján alapuló kapcsolat [1] különböző vonatkozásait vizsgáljuk. Célunk egyrészt hiperbolikus beágyazást (vagyis a hálózat pontjainak a hiperbolikus térben történő, a hálózati topológiát jól tükröző elrendezését) megvalósító algoritmusok, másrészt pedig hiperbolikus geometriára építő hálózatgeneráló modellek fejlesztése.

Az előző félév végére kidolgoztunk egy optimalizálási eljárást, amelyben a hiperbolikus síkon kis-világ tulajdonsággal, skálafüggetlen fokszámeloszlással és magas átlagos klaszterezettséggel rendelkező hálózatokat generáló E-PSO modellre [2] vonatkozó likelihoodot [2] maximalizálva nem csak a likelihood, hanem a greedy routing score [3] szempontjából is sikerült a gyors, dimenzió redukción alapuló ncMCE módszerrel [3,4] kapott beágyazások minőségét feljavítani. Eközben a likelihood m , β és T paramétereinek becslésére szokásosan alkalmazott komplikált procedúrákról áttértünk a szükséges paraméterek együttes, gyors, szintén likelihood maximalizáción alapuló optimalizálására.

Emellett az 1. félévben elkezdtem foglalkozni a különböző dimenziójú hiperbolikus terekben történő hálózatgenerálással. Felismertem, hogy ugyan a tér natív reprezentációjában [1] rendkívül nehézkes meghatározni két pont egymástól vett távolságát a 2-nél magasabb dimenziós hiperbolikus terek esetén, azonban a Poincaré ball modellre [5] áttérve ez a számítás egyszerűen elvégezhető, tehát a Poincaré ball modellt alkalmazva lehetőség nyílik arra, hogy az eredetileg a hiperbolikus *sík* natív reprezentációját használó E-PSO hálózatgeneráló modell algoritmusát tetszőleges dimenziós térre kiterjesszük. A dimenzióknak a generált hálózatok jellemzőire gyakorolt hatásának megismerése pedig elősegítheti a háromdimenziós hiperbolikus térbeli beágyazó módszerek [3] fejlesztését, amelyek sok esetben határozottabb szétválást eredményeznek a hálózat csoportjai között, mint ami a hálózat pontjainak hiperbolikus síkbeli elrendezése esetén tapasztalható [6].

Az aktuális félévben elvégzett kutatások ismertetése

Ebben a félévben elkészítettük a kéziratot, ami bemutatja az E-PSO modellre vonatkozó likelihoodot maximalizáló algoritmusunkat, amellyel az ncMCE módszerrel előállított hiperbolikus beágyazásokat optimalizáltuk. Az eddigi eredményeinket még kiegészítettük azzal az észrevétellel, hogy irányított hálózatok beágyazása esetén a greedy routing score szempontjából kedvezőbb lehet, ha a hálózat pontjainak a hiperbolikus sík natív reprezentációját képező korongon vett radiális pozícióit az élek irányítottságának figyelembevétele mellett határozzuk meg.

Az ncMCE módszer a pontok radiális sorrendjét a HyperMap [2] beágyazó algoritmusban bevezetett (az E-PSO modellre vonatkozó likelihood maximalizálásán alapuló) módon, a pontok fokszám szerinti sorrendjének megfelelően állítja fel. Irányított hálózatok esetén azonban háromféle fokszámról is beszélhetünk: az egyes pontokhoz kapcsolódó összes él számát megadó teljes fokszám mellett vizsgálhatjuk az egyes pontokba befutó és az egyes

pontokból kiinduló élek számát is. Az 1. táblázat adatai alapján az általunk vizsgált tápláléklánc [7] esetén az ncMCE módszerből származó beágyazás, valamint az ncMCE eredményének likelihood alapú optimalizálásával kapott hiperbolikus elrendezés is lényegesen jobb lett a GR-score szempontjából akkor, amikor a radiális koordináták növekvő sorrendjét a pontok teljes fokszámanak csökkenő sorrendje helyett az egyes pontokba befelé irányuló élek számának a csökkenő sorrendjével azonosítottuk.

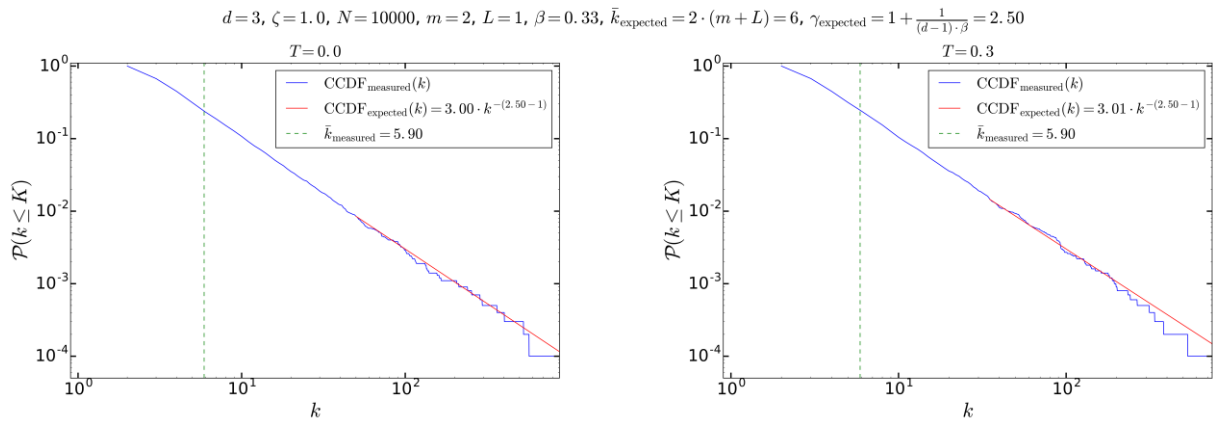
	teljes fokszaám	befok	kifok
ncMCE	0.674	0.792	0.672
ncMCE optimalizált szögkoordinátákkal	0.736	0.890	0.720

1. táblázat: Az ncMCE módszerből és annak szög szerint optimalizált változatából kapott beágyazások greedy routing score-ja a vizsgált tápláléklánc [7] esetén a radiális elrendezés különböző típusú fokszaámok alapján történő meghatározása mellett.

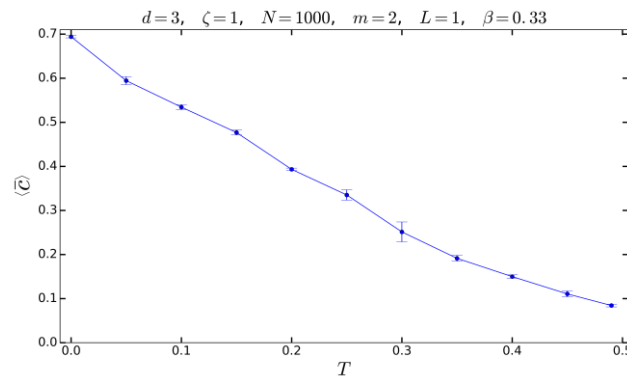
A félév során elkészítettem az E-PSO hálózatgeneráló modell d -dimenziós kiterjesztését megvalósító programot, amelyben minden lépésben megjelenik egy új pont a hiperbolikus teret reprezentáló d -dimenziós gömbben és a pontok közötti x hiperbolikus távolságoktól függő $p(x) = 1/[1 + e^{(x-R)/T}]$ valószínűségek szerint hozzákötött adott számú korábban megjelölt ponthoz. Az egyes időlépésekben kialakított élek számának várható értékét a síkbeli E-PSO modellben alkalmazott levezetéssel [2] analóg módon számítottam ki.

A PSO modellnél [8] leírtak mintájára levezettem a nagy fokszaámok tartományán várható fokszaámeloszlás képletét a determinisztikus ($T = 0$, $p(x)$ lépcsőfüggvény) és a sztochasztikus ($0 < T < 1/(d - 1)$) élkialakítási szabály esetén is. Az 1. ábrán látható egy-egy példa arra, hogy a kapott képletből számolt görbe jól közelíti a generált hálózatokban megjelenő fokszaámok komplementer kumulatív eloszlását. A fokszaámeloszlás a kétdimenziós esethez hasonlóan tetszőleges d dimenzió esetén skálafüggetlennek adódott, azaz nem túl kis k fokszaámokra $\mathcal{P}(K = k) \sim k^{-\gamma}$ alakban írható fel, amelynek megfelelően a fokszaámok komplementer kumulatív eloszlására a $CCDF(k) \equiv \mathcal{P}(k \leq K) \sim k^{-(\gamma-1)}$ képlet érvényes. Fontos kiemelni, hogy az itt szereplő γ nem csak a pontok radiális sűrűségét meghatározó $\beta \in (0,1]$ *popularity fading parameter* értékétől függ, hanem a d dimenziótól is, amit az eddig publikált háromdimenziós beágyazási algoritmusok [3] nem vettek figyelembe. A fokszaámeloszlás kitevőjére kapott $\gamma = 1 + \frac{1}{(d-1)\beta}$ kifejezés értelmében a d dimenzió növekedésével az elérhető legkisebb γ kitevő csökken ($\gamma_{\min} = 1 + 1/(d - 1)$), a valós hálózatokra leginkább jellemző $2 < \gamma < 3$ tartomány pedig az $1/(2d - 2) < \beta < 1/(d - 1)$ beállítással érhető el.

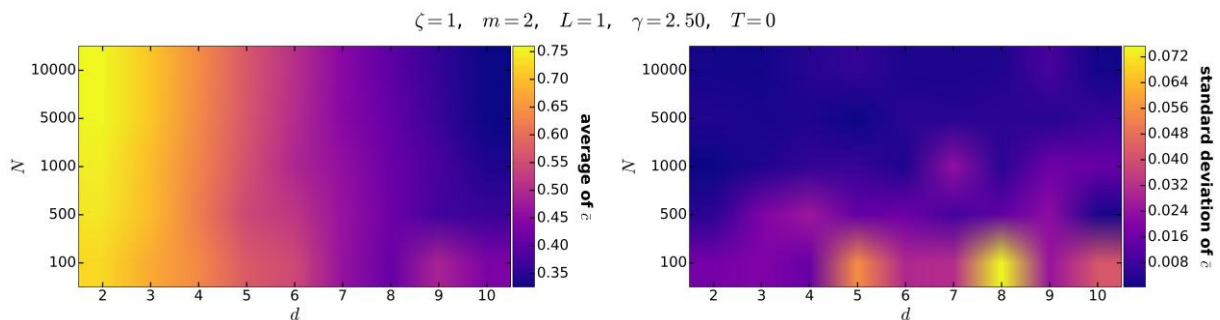
Szimulációs eredmények alapján a \bar{c} átlagos klaszterezettség tetszőleges d dimenzió esetén ugyanúgy viselkedik a T hőmérséklet függvényében, mint a kétdimenziós esetben: ahogy a 2. ábra példája bemutatja, $T = 0$ -ra maximális, és fokozatosan 0 felé közelít, ahogy a hőmérséklet $1/(d - 1)$ -hez tart. A 3. ábra tanulsága szerint az adott dimenziós térben elérhető legnagyobb átlagos klaszterezettség nem változik jelentősen a hálózat méretének növekedésével, azonban az azonos méretű hálózatok közül a magasabb dimenziós térben generáltak kevésbé klaszterezettek.



1. ábra: Példa a foksámok mért és várt komplementer kumulatív eloszlása közötti hasonlóságra az újonnan bevezetett d -dimenziós E-PSO modell esetén, determinisztikus (bal) és sztochasztikus (jobb) élkialakítási szabály alkalmazása mellett. Az ábrák egy-egy hálózat adatai alapján készültek, a hálózatok generálásához használt paraméterértékek a grafikonok felett vannak megadva. A paraméterek jelentése a következő: d a hiperbolikus tér dimenziója, $\zeta \in \mathbb{R}^+$ a hiperbolikus tér K görbületét jellemzi ($K = -\zeta^2$), N a hálózat pontjainak száma, $m \in \mathbb{R}^+$ és $L \in \mathbb{R}$ az átlagfoksámot határozza meg a $\bar{k} = 2 \cdot (m + L)$ képlet szerint, $\beta \in (0,1]$ a hálózat pontjainak radiális irányú sűrűségét befolyásoló popularity fading parameter és $T \in \left[0, \frac{1}{d-1}\right)$ a hálózat átlagos klaszterezettségét szabályozó hőmérséklet.



2. ábra: A T hőmérséklet növekedésével az E-PSO modell $2 < d$ dimenzió esetén is egyre kisebb \bar{c} átlagos klaszterezettségű hálózatokat generál. Minden ábrázolt adatpont 3 hálózatra vett átlagot mutat, a grafikonon fel vannak tüntetve az ezekhez tartozó 95%-os konfidencia intervallumok is.



3. ábra: Az E-PSO modellel elérhető legnagyobb ($T = 0$ hőmérséklet beállítása mellett kapott) átlagos klaszterezettségnek a d dimenziótól és a hálózat pontjainak N számától való függése. Minden $N - d$ beállítás mellett 3 hálózat \bar{c} átlagos klaszterezettségét mértem meg, a bal oldalon ezen értékeknek az átlaga, a jobb oldalon pedig a szórása látható.

Publikációk

Jelenleg az ncMCE módszerrel előállított hiperbolikus beágyazásoknak az E-PSO modellre vonatkozó likelihood maximalizációján alapuló optimalizálásáról szóló kézirat a *Scientific Reports* folyóiratba való beküldés alatt áll.

Tanulmányi tevékenység az aktuális félévben

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

- Sejtszignalizációs hálózatok kvantitatív analízise (FIZ/3/055E)
- Mélytanulás és gépi tanulás a tudományokban (FIZ/3/089)

Február 7. és 9. között meghallgattam az ELTE-n megrendezett, biofizika témájú Mafihe Téli Iskola előadásait.

Konferenciák az aktuális félévben

Áprilisban részt vettem egy online konferencián (*2020 OSA Biophotonics Congress: Biomedical Optics*), ahol elsősorban a mély- és gépi tanulással kapcsolatos szekciók előadásait hallgattam meg.

Terveztem előadást tartani a munkámról a *Statiztikus Fizikai Nap* keretei között, az esemény azonban a koronavírus-járvány miatt elmaradt.

Oktatási tevékenység az aktuális félévben

A Biológiai Fizika Tanszék által meghirdetett *Modern fizika laboratóriumi gyakorlatok* című tárgy *Spektrofotometria* mérését vezettem 4 alkalommal.

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] 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
- [4] 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
- [5] M. R. Bridson and A. Haefliger, *Metric Spaces of Non-Positive Curvature*, Springer-Verlag, Berlin, 1999
- [6] Alessandro Muscoloni, Carlo Vittorio Cannistraci, *Angular separability of data clusters or network communities in geometrical space and its relevance to hyperbolic embedding*, arXiv:1907.00025, 2019
- [7] 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
- [8] Fragkiskos Papadopoulos, Maksim Kitsak, M. Angeles Serrano, Marián Boguñá, Dmitri Krioukov, *Popularity versus similarity in growing networks*, *Nature*, 2012, **489**, 537-540