

4. Félévi beszámoló

Kovács Zoltán (kovzoltan@student.elte.hu)

Statisztikus fizika, biológiai fizika és kvantumrendszerek fizikája PhD program

Témavezetők: Palla Gergely és Zafeiris Anna

A dolgozat címe: Komplex hálózatok és hierarchiák evolúciója társas rendszerekben

Bevezetés

A kutatásom során három fő területtel foglalkoztam. Az első terület a véleménydinamika, amelyen belül én az ágens alapú szimulációkkal foglalkoztam. Ez egy aktívan kutatott terület, ami a hálózatelmélethez is szorosan kapcsolódik, mivel az általános megközelítés szerint az ágenseket egy szociális hálózatban helyezik el, ahol az egyének a csúcsok és a köztük lévő kapcsolatok az élek. Az általános kiindulási feltétel az, hogy az ágensek rendelkeznek saját véleményekkel, amelyek a csoporttagok közti kommunikáció hatására megváltozhatnak. A hagyományos modellekben a véleményeket egy skalár írja le, míg az újabb megközelítésekben az ágensek belső szerkezettel bírnak, ami általában egy másik hálózat (hiedelemrendszer). Mi az utóbbi irányt fejlesztettük tovább egy olyan modellel, amelyben egy pszichológiailag realisztikusabb hiedelemrendszert feltételeztünk, ahol a kognitív disszonanciát is figyelembe vettük.

A kutatásom második fő területe egy régészeti adatokon alapuló adatanalitikai feladat volt. Itt az MTA-ELTE Lendület Innováció kutatócsoportban dolgoztam. A régészek által felvett, különböző lelőhelyekről származó kerámia adatok alapján készítettem hálózatokat, és segítettem az elemzésükben. A lelőhelyek közti kapcsolatok erősségét a legegyszerűbben az ott talált leletanyagok közti hasonlósággal tudjuk jellemezni. A hasonló stílusban, technológiával készült kerámia leletek utalhatnak kereskedelemre, közös kultúrára vagy leszármazásra. A különböző tulajdonságok mentén felvett hálózatok segítik a kapcsolatok mélyebb megértését.

A harmadik nagyobb terület, amin elkezdtem dolgozni, egy genetikai hasonlóságon alapuló családfa-rekonstrukciós algoritmus volt. Ez a kutatás is a régészekkel közös munka eredménye, ahol a cél egy feltárt temetőből származó egyének rokonsági kapcsolatainak meghatározása.

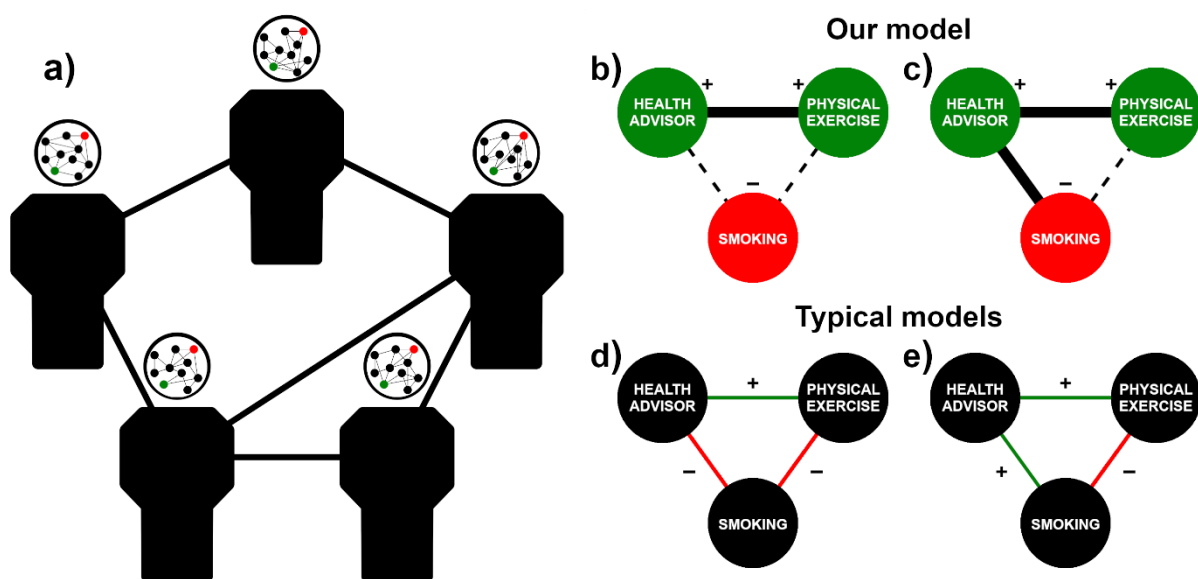
Az előző három félévben elért kutatási eredmények összefoglalása

A kutatásom során implementáltam egy ágens alapú véleménydinamikai modellt. A modellünkben minden ágens rendelkezik egy saját, belső hiedelemrendszerrel. Ezt a hiedelemrendszert egy hálózat írja le, ahol a csúcsok adott fogalmaknak felelnek meg, míg a köztük lévő súlyozott élek a fogalmak közti asszociáció erősségét adják meg. Minden fogalomhoz tartozik egy -1 és 1 közötti vélemény érték. Az egymással asszociált, azonos előjelű fogalmak növelik az ágens koherenciaszintjét, míg az ellentétes előjelű fogalmak közötti élek kognitív disszonanciát okoznak, és csökkentik a koherenciát. Ez eltér az irodalomban megtalálható többi modelltől, ahol a fogalmak közötti éleknek van pozitív vagy negatív előjele. Itt a koherenciaszintet úgy határozzák meg, hogy a fogalmak triádjait vesszük, és összeszorozzák a köztük lévő élek súlyait. Ha pozitív az eredmény, a triád növeli a koherenciát, ha negatív, akkor csökkenti. A hiedelemrendszerek ismert tulajdonsága, hogy az ágensek a maximális koherenciára törekednek. Mi ezen kívül egy külön paraméterrel figyelembe vettük azt is, hogy

az ágensek mennyivel tartják fontosabbnak a koherencia maximalizálásához képest a kognitív disszonancia minimalizálását.

Az ágensek egy szociális hálózatban helyezkednek el, ahol a köztük lévő élek az egymással való kommunikáció valószínűségét adják meg. Annak érdekében, hogy az ágensek ne csak a közvetlen ismerőseikkel kommunikáljanak, bevezettünk egy új paramétert, ami azt adja meg, hogy az ágensek mennyire szeretnek az ismerőseik ismerőseivel kommunikálni. A kommunikáció hatására az ágensek hiedelemrendszerében megváltozik az asszociáció a különböző fogalmak között. Ez potenciálisan inkoherens állapotot eredményezhet, amit az ágensek úgy tudnak orvosolni, hogy stabilizálják a saját hiedelemrendszerüket, az adott fogalmakhoz tartozó véleményeik változtatásával. A kommunikáció után a szociális hálóban is megváltozik az élsúly az ágensek között, annak függvényében, hogy mennyire értettek egyet a tárgyalt fogalmakban. A szimulációkat addig futtattuk, amíg stabil struktúrákat nem kaptunk a szociális hálóban. Ez a kezdeti paraméterektől függően lehetett konszenzus (egy nagy, azonos véleményű csoport), polarizáció (két nagy, ellentétes véleményű csoport) vagy fragmentáció (sok kisebb, egymástól elkülönülő csoport) is. A modellünkben ezek az állapotok akkor is létrejöttek, amikor a kezdeti időpontban az ágensek klónok voltak, azonos hiedelemrendszerekkel.

Az eredményeinket témavezetőimmal egy közös cikkben gyűjtöttük össze, és a Nature Scientific Reports-ba benyújtottuk, amely most a bíráló második fordulójánál tart.

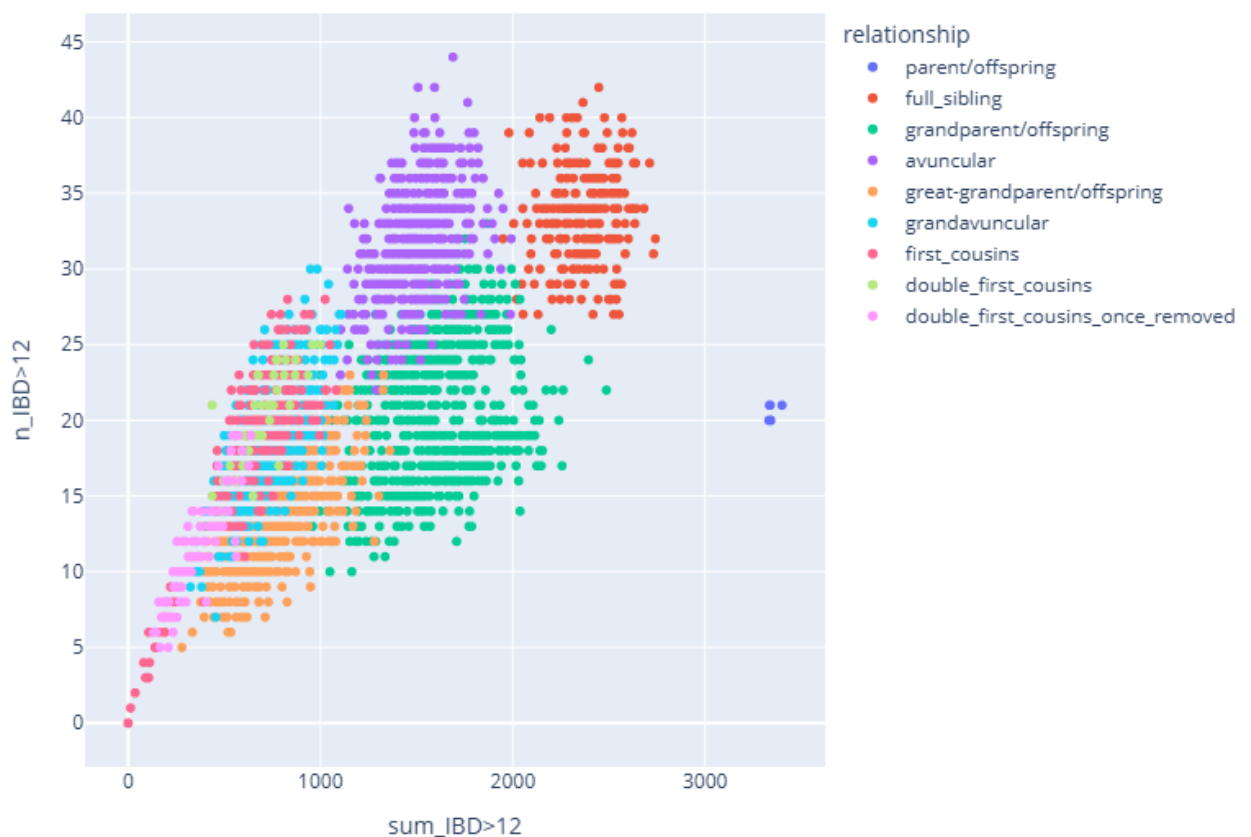


1. Ábra – Az a) ábra a hiedelemrendszerrel rendelkező szociális hálót mutatja be. A b) ábra egy magas, míg a c) ábra egy alacsony koherenciaszinttel rendelkező állapotot mutat be a hiedelemrendszerben a mi modellünkben. Az irodalomban korábban vizsgált modellek stabil és instabil állapotát a d) és e) ábrák mutatják.

A régészekkel közösen folytatott kutatás fő fókuszja az első 3 félévemben az általuk felvett adatbázis értelmezése és kezdeti elemzése volt. Az adatbázis különböző, Budapest környéki lelőhelyekről származó kerámia leletek adatait tartalmazta. A leleteket 3 különböző szempont alapján elemezték a régészek: stílus, készítési technológia és funkció. Ezek közül a stílust és a technológiát kezdtem el feldolgozni. Az adatoknál az UMAP beágyazót használtam a klaszterek

és hálózat kialakítására. Az így kapott eredményeket a régészekkel egy streamlit alapú dashboard-ban megosztottam, amit folyamatosan fejlesztettem a visszajelzéseik alapján.

A családfa rekonstrukció terén a legfőbb eredményem a ped-sim program használatának elsajátítása volt. Ez lehetővé tette, hogy tetszőleges, általam definiált családfák genetikai adatait legyártsam, és összehasonlítsam őket, megkapva az ún. identity by descent (IBD) értékeiket. Az IBD értékek megadják, hogy két ember genomja között mekkora az átfedő szakaszok hossza. Minél több és hosszabb az ilyen átfedés, annál közelebbi rokonságot lehet feltételezni. Nagy mennyiségű IBD adat elemzése azt az eredményt adta, hogy a közeli kapcsolatok megkülönböztethetők csak IBD értékek alapján, de az egyre távolabbi kapcsolatok összemosódnak, azok megfigyeltéhez összetettebb algoritmusokat kell kidolgozni.



2. *Ábra – A ped-sim programmal kapott IBD adatok, különböző rokonsági típusokra. Az első és másodrendű kapcsolatok jól megkülönböztethetők (parent/offspring, full_sibling, grandparent/offspring, avuncular). A távolabbi kapcsolatok viszont nem válnak szét, sok az átfedés, ami megnehezíti a pontos kapcsolatok kinyerését.*

Az aktuális félévben elvégzett kutatások ismertetése

A mostani félévemben folytattam a korábban megkezdett irányokat. A hiedelemrendszer alapú véleménydinamikai modellünk bírálatát megkaptuk, és az ott kapott visszajelzések alapján átirtuk a kéziratot, valamint a szimulációt is részben módosítottuk. Az új eredmények a korábban látott megfigyeléseinket nem változtatták meg.

A régészekkel közös projekten belül folytattam a kerámia adatok elemzését. A megbeszéléseink alapján kibővítettük az elemzett lelőhelyeket az egész Kárpát-medencére, nem csak a Budapest-környékiekre. Mivel ezeknél a lelőhelyeknél többről még nem volt felvéve a stílus és funkció adat, ezért arra a megállapodásra jutottunk, hogy a technológia adatokat magukban is elemezzük, és egy külön cikket írunk belőle a Journal of Archaeological Method and Theory-ba.

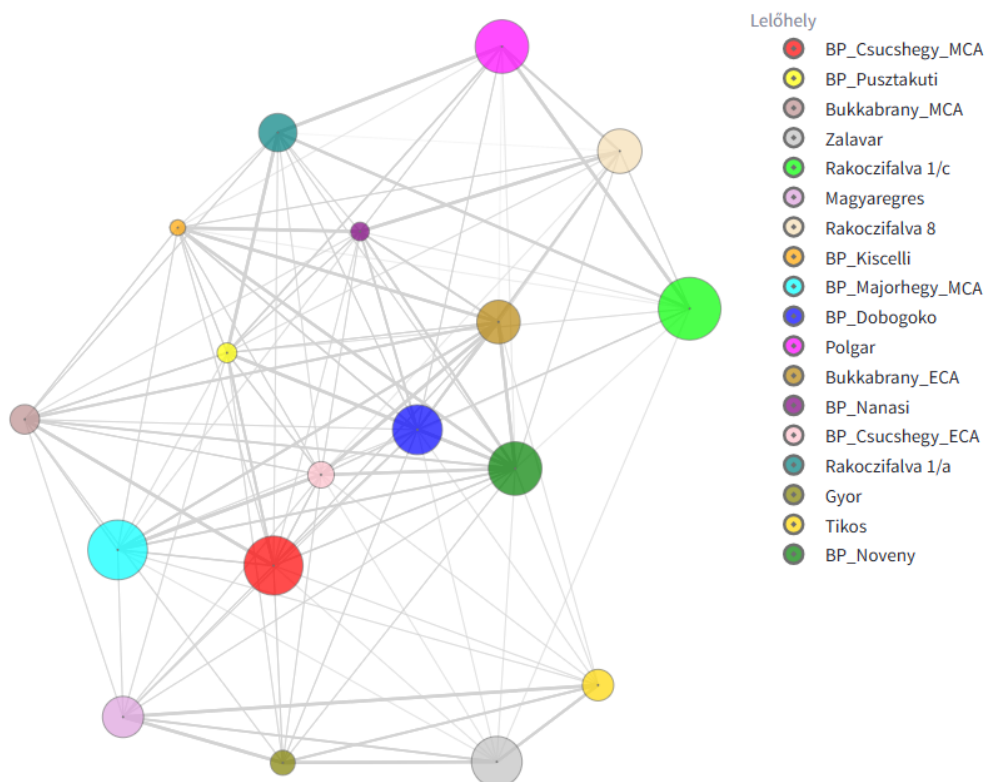
A technológiához tartozó adatbázis tárgyalapon van felvéve, ahol minden tárgyat 4 különböző kategorikus változó ír le. Az alapján, hogy az egyes kategóriákban milyen értékek szerepelnek egy tárgynál, meg lehet határozni, hogy milyen „mintához” tartozik. Egy adott minta egyértelműen jellemez egy tárgyat, ami előfordul az adatbázisunkban. A lelőhelyeket tudjuk úgy jellemezni, hogy mintánként hány lelet fordul elő bennük. Itt az a fő probléma, hogy ilyen reprezentáció mellett az új adatbázisunk oszlopai egymástól teljesen függetlenek, míg a valóságban ez nem igaz. Lesznek olyan minták, amik több kategóriában is megegyeznek és hasonlóbbak lesznek egymáshoz, mint azok, amik minden kategóriában különböznek (Pl., ahol a készítési módszer megegyezik, de a felületi megmunkálás különbözik, egy olyan esethez képest, ahol egyik sem egyezik meg). Az elmúlt félévben kidolgoztam egy olyan módszert, ami segít feloldani ezt a problémát, és a bonyolult leletanyagok feldolgozása informatikailag lehetővé válik. A módszer lényege, hogy a különböző minták között páronként meghatározom a hasonlóságot, és ennek segítségével eltranszformálok az egyes lelőhelyekhez tartozó adatokat. Az összehasonlítást a lelőhelyek között már ezeken a transzformált adatokon végzem.

Ezt a módszert sikerült alkalmazni a technológiai adatokra, és az UMAP beágyazót használva a régészek szerint is jól értelmezhető eredményeket kaptam. A régészekkel továbbra is folyamatos megbeszéléseket tartunk, a közös cikk írása folyamatban van.

A genetikai adatokon vett elemzések azt mutatták, hogy nem lehet egyszerűen meghatározni a pontos kapcsolatokat az IBD értékek alapján. Ennek orvoslására témavezetőm, Zafeiris Anna kidolgozott egy genetikai algoritmuson alapuló módszert, ami a kapcsolatokat nem páronként, hanem hármasokban, ún. triádokban elemzi. Egy ilyen triád 3 olyan személyből áll, akik páronként össze vannak hasonlítva egymással. Egy adott triád lehet valid vagy invalid. Ezt az alapján határoztuk meg, hogy egy valós családfában előfordulhatnak-e. Ezeknek a pontos meghatározására több családfát is szimuláltam a ped-sim programmal, és meghatároztam a valid triádok halmazát. Mivel a legtöbb pár több triádban is szerepel, ezért a párok közti kapcsolatok változtatásával a többi triád is változik a valid és az invalid konfigurációk között. A genetikai algoritmikusnak azt a célfüggvényt megadva, hogy a valid triádok számát maximalizálja vissza tudunk kapni egy olyan kapcsolati hálót, ami egy jó családfa rekonstrukciónak felel meg.

A jelenleg meglévő algoritmust még bővíteni fogjuk, például figyelembe véve a mitokondriális DNS és Y-kromoszomális DNS hatásait, és az általunk szimulált családfákon fogjuk tesztelni.

Amennyiben ezeken jó eredményt ad, úgy a valós, régészekről kapott adatokon is alkalmazni fogjuk a módszert.



3. Ábra - A technológia adatok módosított hasonlósággal vett UMAP hálózata. Az alul elhelyezkedő 4 lelőhely (Győr, Tikos, Magyaregres, Zalavár) a Dunántúlnak felelnek meg, és egymással szoros csoportot alkotnak. A Budapesti és Alföldi lelőhelyek kevésbé válnak szét.

Publikációk

- Kovács, Z., Zafeiris, A., & Palla, G. (2024). Opinion polarisation in social networks driven by cognitive dissonance avoidance. arXiv preprint arXiv:2409.16343. – benyújtva a Nature Scientific Reports-ba, első körös review utáni javítás alatt
- Journal of Archaeological Method and Theory – kerámia technológia hálózatelemzés cikk. – Írás alatt

Tanulmányi tevékenység az aktuális félévben

A félév során egy kurzuson vettem részt:

- FIZ/3/093 – Haladó gépi tanulás labor

Konferenciák a képzés alatt

A képzés során eddig nem vettem részt konferencián. A következő évben az alábbi konferenciákon tekinteném hasznosnak a részvételt:

- NetSci
- Conference on Complex Systems
- CompleNet - International Conference on Complex Networks

Oktatási tevékenység

A félév során méréseket felügyeltem a Modern Fizika laboratóriumban, összesen 6 alkalommal. Ezen kívül a laborból két mérés jegyzőkönyveit is én értékeltem.