



M Ű E G Y E T E M 1 7 8 2

TDK DOLGOZAT

Homofilia hatásának modellezése a szociális hálózatokban

Tamás Gábor

Fizikus MSc I. évfolyam
Természettudományi Kar

Témavezető:

Dr. Török János

egyetemi docens
Elméleti Fizika Tanszék
Természettudományi Kar

Budapesti Műszaki és Gazdaságtudományi Egyetem
2018

Tartalomjegyzék

1. Bevezetés	2
1.1. Szociális hálózati modellek	2
1.2. Szociális tulajdonságok	5
1.3. Internetes kommunikációs csatornák linkszelekcíós mechanizmusa	6
1.4. Dunbar elmélete	8
1.5. Közösségek felismerése	9
2. A vizsgált modell	11
2.1. Definíció	11
2.2. Mért mennyiségek	13
3. Fázisátalakulás	15
3.1. Relaxációs idő	16
3.2. A jelenség háttere	18
4. Összehasonlítás adatokkal	22
4.1. iWiW	22
4.2. Modell	26
5. Összefoglalás, kitekintés	29
Irodalomjegyzék	31

1. fejezet

Bevezetés

A szociológusok mindig is nagy érdeklődéssel vizsgálták a társas kapcsolatok kialakulásának folyamatát, melynek számos korábbi megfigyelés szerint egyik fő szervező ereje a homofília. Ezen alapelv szerint hasonló emberek között nagyobb eséllyel jön létre kapcsolat, mint különböző emberek között. Erre utal az egyik ismert közmondás is: „Madarat tolláról, embert barátjáról”. A homofília jelenségének hátterét vizsgálta Miller McPherson 2001-ben megjelent cikkében [1].

Mára az információ- és kommunikációtechnológia (ICT) fejlődésének következtében lehetővé vált az egyének közötti kapcsolatok kvantitatív, adatalapú tanulmányozása. Így az emberek szociális hálózata - amely tartalmazza az egyének tulajdonságait és kapcsolatrendszerét - részleteiben is vizsgálható.

A társadalmat leíró modellek - például az 1.2. fejezetben tárgyalt Axelrod-modell - sok esetben a fizikai jelenségek tanulmányozása során tapasztalt fázisátalakulásokhoz hasonló viselkedést mutatnak, éppen ezért a fizikában alkalmazott modellek is alkalmazhatóak lehetnek szociológiai rendszerek vizsgálatára.

1.1. Szociális hálózati modellek

TDK munkám során szociális hálózatokkal foglalkoztam, melyek tárgyalásához elengedhetetlen a szociális kapcsolat fogalmának definiálása. Szociális kapcsolatnak akkor nevezzük két ember ismeretségét, ha egy véletlen találkozás során képesek öt percen keresztül egymás magánéletét érintő beszélgetést folytatni, illetve hajlandóak egy kisebb

szívességet megtenni a másikért [2].

A Kumpula-modell [3] az elsők között tudott viszonylag egyszerű módon reprodukálni szociális hálózatokon végzett kvantitatív méréseket. Célja egy olyan súlyozott hálózat előállítása, melynek tulajdonságai megfelelnek az emberi szociális hálózatokban tapasztalhatóaknak. A modell a hálózat kialakulását három mechanizmus együttes hatásaként írja le: GA - globális kapcsolódás (global attachment), LA - lokális kapcsolódás (local attachment) és ND - nódusok törlése (node deletion).

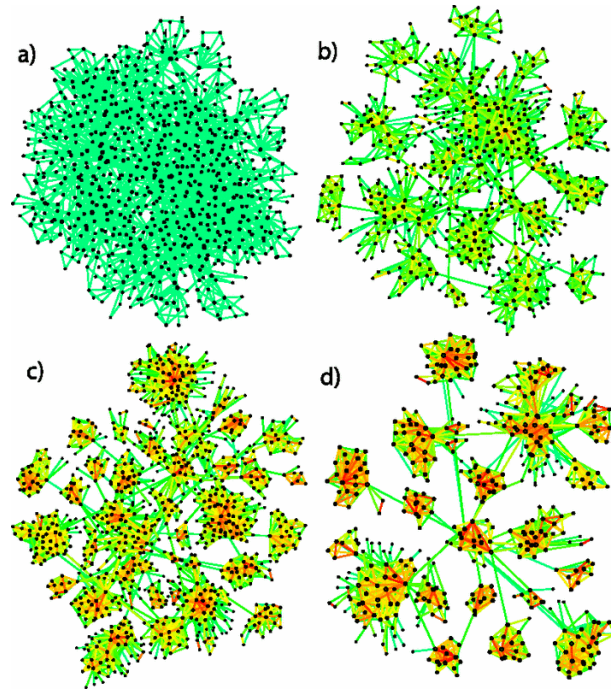
A GA és az LA az emberi kapcsolatok kialakulásának és fejlődésének két alapvető, szociológusok által megfigyelt mechanizmusát modellezzik [4]. A GA a véletlen találkozást szimulálja, amely során olyan emberrel kerülünk szociális kapcsolatba, aki nem része ismerőseink köreinek. Ez egy ritka jelenségnek tekinthető. Ezzel szemben gyakori az LA esemény, amely a szociális hálózatokban megfigyelt háromszögek erős jelenlétét próbálja leírni. Ennek során vagy egy létező baráti háromszög erősödik meg a szociális interakció során, vagy közepes valószínűséggel egy még nyitott háromszög záródik be, azaz ekkor az adott egyén kapcsolatba kerül a barátja barátjával.

A stacionárius állapot megvalósulásáért a harmadik lépés, azaz egy véletlen egyén összes kapcsolatának törlése (ND) felel. Utóbbi a valóság durva leegyszerűsítése, hiszen nem jellemző, hogy az emberek egy hálózatot - és így az abban megtalálható ismerőseiket - pillanatszerűen hagynának el.

A modell egyik fontos paramétere δ , ami azt méri, hogy a szociális interakció milyen erős hatással van a kapcsolatokra. A modell viselkedését kiértékelve jól megfigyelhető fázisátalakulás megy végbe δ függvényében, amit az 1.1. ábra szemléltet. Kis δ értékek esetén a hálózat homogén, nagyon hasonlít egy Erdős-Rényi véletlen gráfhoz. δ növelésével a hálózat jól láthatóan egyre inkább csoportokra oszlik.

Az 1.1. (d) ábra megfelel a Granovetter által tett megállapításnak [5], amely szerint a hálózatokban az erősen csatolt közösségeket gyenge linkek (kapcsolatok) tartják össze, azok nélkül a hálózat szétesik. Ezen állítást úgy igazolhatjuk kvantitatív módon, ha a hálózatban megtalálható éleket (kapcsolatokat) kétféleképpen, eltérő sorrendben távolítjuk el.

Elsőként a nagy súlyal rendelkező élektől haladunk az egyre kisebb súlyúak felé, míg a második esetben fordított sorrendben töröljük a kapcsolatokat. Ezen módszert linkel-



1.1. ábra. A Kumpula-modellben megfigyelhető átmenet a homogén (a) és a széttagolt (d) hálózat között a (a) $\delta = 0$, (b) $\delta = 0,1$, (c) $\delta = 0,5$, (d) $\delta = 1$ paraméterértékek mellett. A zöld szín a gyenge, míg a pirosba hajló az erősebb kapcsolatokat szemlélteti. Az új kapcsolatok egységnyi súllyal jönnek létre [3].

távolítási perkolációnak (link removal percolation) nevezzük [6]. A Kumpula-modellt a tagolt fázisban vizsgálva az eredmények - a várakozásoknak megfelelően - azt mutatják, hogy a második esetben (amikor a gyenge éleket távolítjuk el először) a hálózat jóval hamarabb egymással kapcsolatban nem álló részgráfokra esik szét. Ez igazolja a gyenge élek nagy jelentőségét, és így megfelel a Granovetter-képnek. A Kumpula-modell továbbá kvalitatíve sikerrel adja vissza az empirikusan megfigyelt mennyiségeket, többek között az asszortativitás és a klaszterezettség fokszámfüggését is.

A modellel kapcsolatban két fontos kritikát fogalmazhatunk meg. Egyrészt az ND lépés a valóságnak egyáltalán nem felel meg, annak finomítására lenne szükség. Másrészt a modell a társadalmi kapcsolatoknak csupán egy rétegét vizsgálja, holott a társadalomban az emberek tipikusan több szociális réteggel rendelkeznek, amelyek csak együttesen tudják feltárni az egyén valódi kapcsolati hálóját.

Témavezetőm kutatócsoportja a Kumpula-modellt a következőképpen fejlesztette tovább [7]: az ND lépés helyett - amely során gyakorlatilag egyes emberek teljesen

eltűnnek a hálózatból - két új típusú, életszerűbb mechanizmust vezettek be, melyek közelebb állnak a valós szociológiai folyamatokhoz. Ezek közül az egyik az LD - linkek törlése (link deletion), amelynek hatására csak egyes kapcsolatok szakadnak meg idővel. Ennél még realiztikusabb az, ha a linkek nem pillanatszerűen szűnnek meg, hanem az idő elteltével fokozatosan halványulnak el. Ez az AG - linkek gyengítése (link aging) mechanizmus. Mindkét módszert megvizsgálva megállapítást nyert [7], hogy az ND-nél jobb tulajdonságokkal rendelkeznek, segítségükkel a valóság jobban leírható. Véleményem szerint az emberi kapcsolatokra sokkal inkább a fokozatos elhalványulás jellemző a véletlenszerű megszűnéssel szemben, ezért TDK munkám során az AG lépést építettem be a vizsgált modellbe.

1.2. Szociális tulajdonságok

A Kumpula-modell egy további hiányossága, hogy a kapcsolatok létrejötte során nem veszi figyelembe az egyének jellemzőit, azokat megkülönböztethetetlennek tételezi fel. Ez azonban természetesen a valóságban nem tehető meg, hiszen egy szociális kapcsolat létrejötte során nagy szerepe van annak, hogy az érintett két egyén mennyire hasonlít egymáshoz, szociális tulajdonságaik mennyire hasonlóak.

Az emberek szociális tulajdonságai alatt a továbbiakban demográfiai és kulturális tulajdonságaik összességét értem, így többek között nemüket, életkorukat, iskolázottsági szintjüket, lakóhelyüket, sportolási preferenciáikat, politikai irányultságukat és vallási hovatartozásukat. A szociális tulajdonságok (features) számát jelölje F , míg az egyes tulajdonságok lehetséges értékeinek számát (traits) q . A lehetséges értékek száma függ az adott tulajdonság jellegétől, így például az adott egyén neme két különböző értéket vehet fel (férfi vagy nő), míg iskolázottsági szintje már csak egy többfokozatú skála használatával határozható meg (példa egy ilyen skálára: alapfokú végzettség, középfokú végzettség, alapszakos diploma, mesterszakos diploma, doktori fokozat).

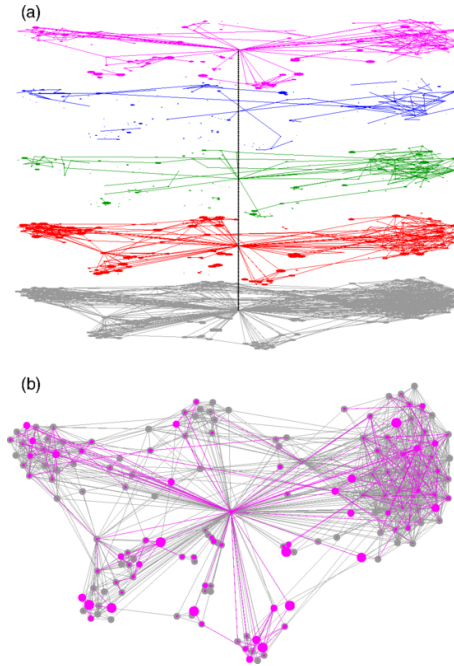
A hálózatot alkotó egyének tulajdonságait figyelembe vevő első, nagy hatású modell az Axelrod-modell [8]. Az Axelrod-modell minden egyént egy F hosszúságú vektorral jellemez, ezen vektor egyes elemei q különböző értéket vehetnek fel. Az algoritmus egy lépésében egy véletlenszerűen kiválasztott i nódus (egyén), valamint annak egy j

szomszédja között a szociális tulajdonságaik hasonlóságának mértékével arányos valószínűséggel jön létre kölcsönhatás, melynek során az i egyén lemásolja j valamely tulajdonságának értékét.

Az Axelrod-modell $F=1$ és $q=2$ paraméterek mellett megfeleltethető a statisztikus fizikából ismert Ising-modellnek $T=0$ hőmérséklet esetében. Bár arra számítunk, hogy ferromágneses alapállapot - és így egy globális domén - jön létre egyforma nódusokkal, azonban egy konkrét esetet vizsgálva előfordulhat azon stabil konfiguráció is, amelyben két ellentétes spinnel rendelkező domént látunk egy egyenes doménfallal elválasztva. Axelrod megmutatta, hogy nagyobb F esetén is hasonló jelenség tapasztalható. Azt várhatnánk, hogy elegendően hosszú idő letelte után egy homogén klasztert kapunk, azonban a szimuláció eredményeként domének, kulturális szigetek alakulnak ki, melyeket jól definiált kulturális határvonalak választanak el egymástól. Mivel a határvonalak két oldalán megtalálható nódusok minden tulajdonsága eltérő lesz, ezért a szimuláció elegendően nagy számú lépés után már nem fog változást hozni a nódusok tulajdonságaiban, hiszen ha két szomszédos egyén egymástól teljesen különbözik, akkor zérus valószínűséggel hatnak kölcsön a modell defíciója alapján. Emiatt a doménfalak tetőszőleges alakúak lehetnek. Az Axelrod-modellt sokan tanulmányozták, viselkedése sok esetben fázisátalakulást mutat, amelyet divergáló időskála kísér [9].

1.3. Internetes kommunikációs csatornák linkszelekciós mechanizmusa

Egy adott egyén egocentrikus hálózata - amely definíció szerint a kiválasztott egyént, valamint ismerősinek kapcsolatrendszerét tartalmazza - a legtöbb esetben nem jelenik meg teljes valójában egyik internetes kommunikációs csatorna esetében sem, így reprodukálni is csak az egyes csatornák együtteseként tudjuk. Ezt az alapvetést az 1.2. ábra szemlélteti.



1.2. ábra. A különböző kommunikációs rétegek sematikus ábrája (a), valamint egy réteg és az eredeti egocentrikus hálózat közös szemléltetése (b) [10].

Meg lehet mutatni, hogy amennyiben egy realiztikus szociális hálózatból csak bizonyos szabályoknak megfelelő éleket tartunk meg, akkor kvalitatíven reprodukálhatók a nagy adatokon mért eredmények [10]. Ezt a következőképpen mutatták meg.

Mivel a valódi egocentrikus hálózat ismeretlen, ezért helyettesítő hálózatokat hozunk létre adott észszerű tulajdonságokkal (például adott fokszámeloszlással), majd ezekre alkalmazzuk az ún. linkszelekciós mechanizmust (sampling), ami egy adott kommunikációs csatorna használatát szimulálja. Minden nódushoz hozzárendelünk egy a_i affinitást, amely azt hivatott reprezentálni, hogy az adott felhasználó mennyire kedveli az adott kommunikációs csatornát. A linkszelekció során az a_i affinitás értékét (1.1) csökkenő eloszlásból véletlenszerűen választjuk ki minden i -re.

$$P(a_i) = \frac{1}{A} e^{-\frac{a_i}{A}} \quad (1.1)$$

Az emberek, amikor kommunikálni akarnak egy ismerősükkel, általában több kommunikációs csatorna közül választhatnak. Azonban nem feltétlen a kedvencüket (a legnagyobb affinitásút) választják, hiszen lehet, hogy a másik azt nem is használja (pl. szülő hiába ír emailt gyermekének, ha az csak a cseten lóg). A legkézenfekvőbb meg-

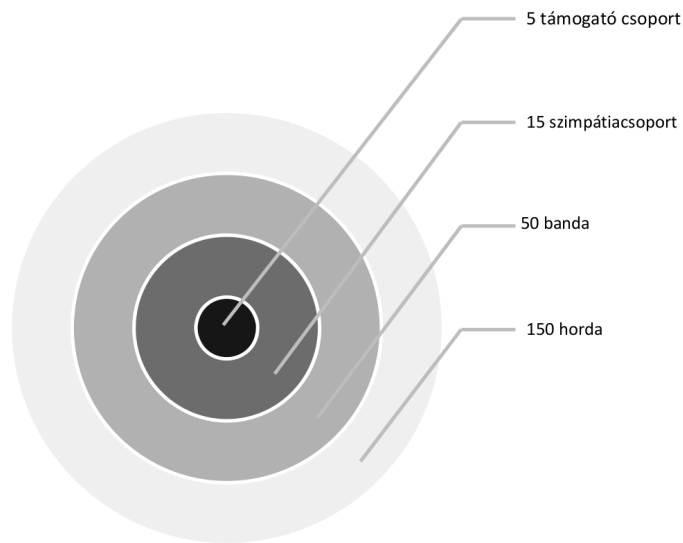
oldás olyan csatornát választani, ami mindkettejük számára a legkevésbé ellenszenves. Ezen elv szerint a két affinitás minimumát kell vennünk a kapcsolat megtartását leíró valószínűség kiszámítása során. Tehát az i és j nódusok között a link (1.2) valószínűséggel lesz látható a kommunikációs csatornán.

$$p_{ij} = \min\{a_i, a_j\} \quad (1.2)$$

Ésszerű tulajdonságú helyettesítő hálózatokból kiindulva tehát kvalitatíve reprodukálni tudjuk a nagy adatokon (big data) mért mennyiségeket.

1.4. Dunbar elmélete

Dunbar ún. társas agy (social brain) hipotézise [11] szerint az ember a neocortex (új agykéreg) méretéből adódóan közelítőleg 150 emberrel (ez az ún. Dunbar-szám) képes fenntartani szociális kapcsolatot. Ezt az eredményt a főemlősökön megfigyelt adatok emberre történő extrapolálásával kapta. Egy adott egyén ismerőseit a kapcsolat erősségét alapul véve hálózati rétegekbe sorolja, melyek közül a belső rétegeknél - Dunbar elnevezése szerint a támogató és szimpátia csoportban - gyakoribb a személyes, intenzívebb érzelmi kapcsolat. Ezzel szemben a külső rétegeknél - Dunbar elnevezése szerint a bandában és hordában - ritkák a személyes találkozások, és az intimitás is kevésbé jellemző [12]. A hálózati rétegeket az 1.3. ábra szemlélteti.



1.3. ábra. Az egyén hálózatának rétegei és az azokban megtalálható egyének száma Dunbar hipotézise alapján [12].

1.5. Közösségek felismerése

A szociális hálózatok szerkezetének feltárásában, valamint a hálózatban lejátszódó folyamatok megértése során fontos szerep jut a közösségeknek, ahogy azt az 1.1. fejezetben tárgyalt Kumpula-modell esetében is láthattuk. A közösség fogalmára széles körben elfogadott meghatározás az, hogy a közösség egy lokális, sűrűn összekapcsolt részgráf a hálózatban. Ez azt jelenti, hogy egy közösség minden tagja elérhető ugyanazon közösség többi tagja által. Emellett egy közösségben az egyének nagyobb eséllyel kapcsolódnak egymáshoz, mint a hálózat más - közösségen kívüli - nódusaihoz [13].

A hálózatokban megtalálható közösségek felismerésére számtalan eljárás született az elmúlt évtizedekben. TDK munkám során ezek közül az egyik legnépszerűbbet, az Infomap algoritmust használtam, mely egyúttal az egyik leggyorsabb is [14]. Az algoritmus a hálózaton történő véletlen bolyongáson alapul. Minden nódust egy címkével látunk el, és ezen címkék segítségével próbáljuk meg hatékonyan kódolni a bolyongást a hálózaton a Huffman-kódoláshoz hasonlóan. Azonban kétszintes struktúrát feltételezünk, a feltételezett közösségeket külön címkézzük. Így a közösségek közötti átmenetet a hozzájuk tartozó címkével írjuk le.

Az Infomap algoritmus célja, hogy az ún. map equation (1.3) alapján minél jobban tömöríteni tudja a bolyongás leírását.

$$L = q_{\curvearrowright} H(Q) + \sum_{i=1}^{n_c} p_{\circlearrowleft}^i H(P^i), \quad (1.3)$$

ahol $H(Q)$ a gyakorisággal súlyozott átlagos hossza a közösségek közötti átmenetek során a címkéknek, míg $H(P^i)$ a gyakorisággal súlyozott átlagos hossza az i közösséghez tartozó címkéknek.

Amennyiben nem a valóságnak megfelelően jelöltünk ki közösségeket, azok címkéjét túlságosan gyakran fogjuk használni a közösségek közötti vándorlás során, és így akár az eredeti kódolásnál rosszabb eredményt is kaphatunk. Abban az esetben azonban, ha optimális a választás, az egyenlet által leírt kifejezés minimális lesz. Az algoritmus kiindulási állapotában minden egyes nódust egy külön közösségként jelölünk meg.

2. fejezet

A vizsgált modell

2.1. Definíció

A TDK munkám során vizsgált, témavezetőm kutatócsoportja által kidolgozott [15] modell az 1. fejezetben bemutatott Kumpula- és Axelrod-modellek kombinációjaként jött létre. Az Axelrod-modellhez hasonlóan minden i nódushoz hozzárendel egy F elem-ből álló vektort, amely az egyén szociális tulajdonságait reprezentálja. Minden egyes tulajdonság q különböző értéket vehet fel. Azonban az Axelrod-moddellel ellentétben a tulajdonságok a szimuláció során rögzítettek, azokat az egyének közötti kölcsönhatás nem változtatja meg. Emellett még megengedett, hogy F értéke nódusról nódusra változzon.

Így tehát az i nódust leíró tulajdonságvektor (feature vector) összesen $F_i \in [F_{\min}, F]$ egész számot tartalmaz (2.1).

$$\mathbf{v}_i = (\varphi_i^1, \varphi_i^2, \dots, \varphi_i^{F_i}), \quad (2.1)$$

ahol $\varphi_i^f \in \{1, \dots, q\}$ minden $f \in \{1, \dots, F_i\}$ -re.

Az F_i érték nagysága az egyén szociális komplexitását jellemzi: minél hosszabb a \mathbf{v}_i vektor, annál több tulajdonsággal lesz leírható az egyén a hálózatban. A q érték a sokszínűséget hivatott megőrizni: minél nagyobb, annál többfajta tulajdonságvektor állhat elő a lehetséges értékek kombinációjaként.

A modell a szociális hálózat létrejöttét három effektus együttes hatásaként írja le.

- GA - globális kapcsolódás

A hálózatból véletlenszerűen kiválasztunk egy i nódust és annak egy f tulajdonságát. Ezt követően egy olyan j csúcsot jelölünk ki, amelyre teljesül a $\varphi_j^f = \varphi_i^f$ összefüggés. Amennyiben i -nek van olyan már meglévő ismerőse, akivel az f tulajdonsághoz tartozó értékük megegyezik, akkor i és j között p_{glob} valószínűséggel jön létre kapcsolat. Abban az esetben, ha nincs ilyen ismerőse i -nek, az i és j közötti ismeretséget $p = 1$ valószínűséggel hozzuk létre. Az újonnan létrehozott link súlya w_0 .

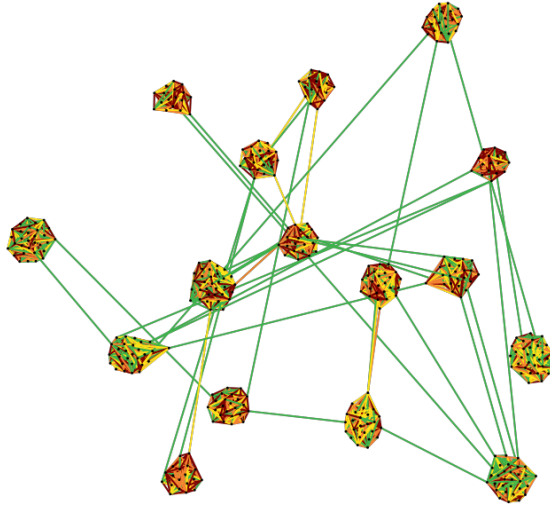
- LA - lokális kapcsolódás

Véletlenszerűen kiválasztunk egy i egyént, valamint annak egy f tulajdonságát. Ezt követően a linkek w_{ij} súlyával arányos valószínűséggel kijelöljük i egy olyan j szomszédját, akivel osztoznak ezen tulajdonságon. Ezután j szomszédai közül is kiválasztunk az élek súlyával arányos valószínűséggel egy megegyező tulajdonságú k nódust. Így tehát teljesülnie kell a $\varphi_k^f = \varphi_j^f = \varphi_i^f$ összefüggésnek. A háromszögben érintett - azaz a három nódust összekötő, már létező - linkek súlyát megnöveljük w_0 -lal, majd amennyiben k és i között még nincs kapcsolat, akkor p_{tri} valószínűséggel és w_0 súllyal létrehozuk azt.

- AG - linkek gyengítése (link aging)

Minden link w_{ij} súlyát megszorozzuk egy egynél kisebb c_{ag} számmal. Amennyiben $w_{ij} < w_{tr}$, azaz a link súlya egy előre megadott küszöbérték alá esik, az adott kapcsolat eltávolításra kerül a hálózatból.

A modellünket fel lehet fogni úgy, mint a Kumpula-modell többrétegű változatát, ahol egy rétegben csak az emberek egy részével léphetünk kapcsolatba, még hozzá azokkal, akik velünk megegyező tulajdonsággal rendelkeznek. A teljes szociális hálózat ezen rétegek összesége adja. Azt szeretnénk elérni, hogy egy-egy rétegen továbbra is teljesüljön a Granovetter-hipotézis, ezért F és q kivételével olyan modellparamétereket választottunk (params llist), amik ezt visszaadják. A 2.1. ábrán egy mintahálózatot látunk $q = 1$, azaz egy réteg esetén.



2.1. ábra. Az általunk vizsgált modell alapján kapott hálózat $N = 500$ nódus esetén, $F = 4$ és $q = 1$ paraméterértékek mellett. A zöld szín a gyenge, míg a pirosba hajló az erősebb kapcsolatokat szemlélteti.

Megfigyelhetjük, hogy a létrejött hálózatban - várakozásainknak megfelelően - a kialakult közösségeket gyenge (zöld színű) élek tartják össze.

2.2. Mért mennyiségek

A kapott eredmények kiértékelése során két fontos mennyiséget mértünk az F és q paraméterek függvényében.

- $\langle o \rangle$ átlagos hasonlóság (average overlap) a hálózatban

$$\langle o \rangle = \frac{1}{N} \sum_{i=1}^N \langle o_i \rangle, \quad (2.2)$$

ahol N a hálózatban megtalálható egyének száma és $\langle o_i \rangle$ az átlagos hasonlóság az i nódusra vonatkozóan (2.3).

$$\langle o_i \rangle = \frac{1}{n_i} \sum_{j=1}^{n_i} o_{ij}, \quad (2.3)$$

ahol n_i az i nódus ismerőseinek száma, o_{ij} pedig egy i és egy j nódus \mathbf{v}_i és \mathbf{v}_j tulajdonságvektorainak hasonlósága (2.4).

$$o_{ij} = \frac{1}{\tilde{F}} \sum_{f=1}^{\tilde{F}} \delta_{v_i^f, v_j^f}, \quad (2.4)$$

ahol $\tilde{F} = \min\{F_i, F_j\}$, illetve $\delta_{i,j}$ a Kronecker-delta.

- $\langle k \rangle$ átlagos fokszám (average degree)

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i, \quad (2.5)$$

ahol N a hálózatban megtalálható egyének száma, k_i pedig egy adott i nódus fokszáma.

A valóságban F és q minden ember esetében más és más lehet. Az egyszerűség kedvéért feltételeztük, hogy q minden nódus esetében rögzített, F pedig csak megadott határokon belül változhat. A szimulációk futtatása során használt paramétereket a 2.1. táblázatban foglaltam össze.

Paraméter	Érték
F	[1, 8]
q	{1, 2, 3, 4, 5, 7, 10, 20, 100, 300}
N	10000
p_{glob}	0,001
p_{tri}	0,2
w_{tr}	0,01
c_{ag}	0,9

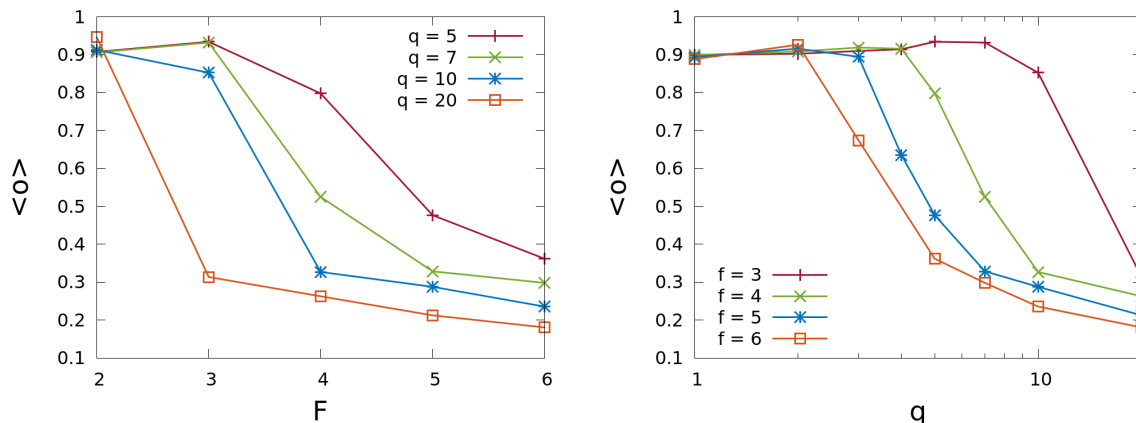
2.1. táblázat. A szimulációs paraméterek

3. fejezet

Fázisátalakulás

A vizsgált modell implementálását követően realizáltam, hogy a program futtatása a vártnál jelentősen több időt vesz igénybe. Így a szimulációkhoz szükséges nagy számítási igény miatt azokat nem én magam végeztem el, hanem témavezetőm kérésére a kutatócsoport egy tagja, Yohsuke Murase, akinek köszönettel tartozom segítségéért.

Az $\langle o \rangle$ átlagos hasonlóság F -től és q -tól való függését a 3.1. ábrán mutatom be.



(a) Rögzített q érték mellett

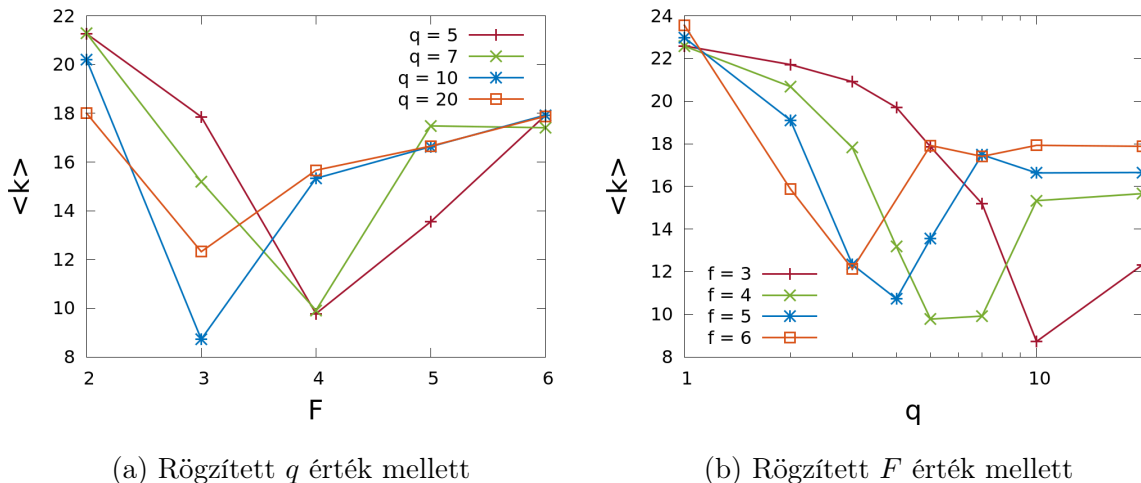
(b) Rögzített F érték mellett

3.1. ábra. Az $\langle o \rangle$ átlagos hasonlóság függése a paraméterektől

A 3.1. ábrákon látható görbék szemlélve megállapíthatjuk, hogy alacsony F , illetve q érték esetén az átlagos hasonlóság $\langle o \rangle = 1$ -hez közeli értéket vesz fel, azonban F , illetve q növelésével hirtelen jelentősen lecsökken. Rögzített F mellett megfigyelhetjük (3.1a ábra), hogy a hirtelen letörés nagyobb q -k esetén hamarabb megtörténik. Ugyanezen jelenség tapasztalható rögzített q mellett is (3.1b ábra), F növelése során.

Az $\langle o \rangle$ átlagos hasonlóság által mutatott viselkedés alapján úgy tűnhet, hogy alacsony F , illetve q értékek mellett az emberek csak a velük szinte teljesen megegyező tulajdonságokkal rendelkezőkkel kötnek ismeretséget. Ám F és q növelésével egy ponton valami hirtelen megváltozik. Az általunk vizsgált kérdés az volt, hogy mi bújhat meg ezen jelenség hátterében.

A $\langle k \rangle$ paraméter F -től és q -tól való függése a 3.2. ábrán látható.



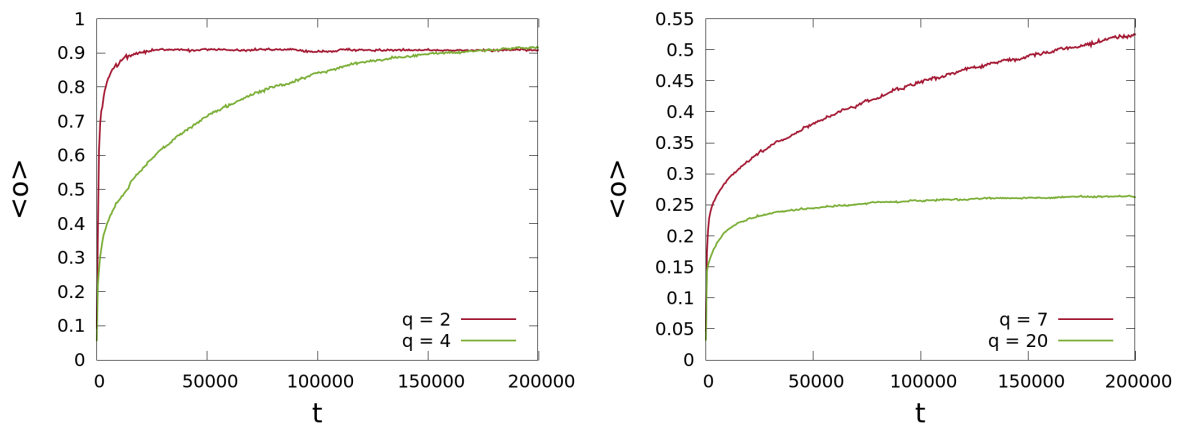
3.2. ábra. Az $\langle k \rangle$ átlagos fokszám függése a paraméterektől

Összevetve a 3.2. ábrákat a 3.1. ábrákkal megállapíthatjuk, hogy $\langle k \rangle$ -nak éppen ott van minimuma, ahol $\langle o \rangle$ értéke hirtelen lecsökken. Ezt látva feltételezhető, hogy a két jelenség nem független egymástól: a kapott eredmények fázisátalakulást sejtetnek. Arra következtethetünk, hogy az $\langle o \rangle$ átlagos hasonlóság rendparaméterként, míg az F és q kontrollparaméterként viselkednek. A grafikonok ugyanis azt mutatják, hogy alacsony F és q értékek mellett $\langle o \rangle$ alapján „rendezett”, míg nagy F és q esetén - az átalakulási pont elérését követően - „rendezetlen” fázisban vagyunk.

3.1. Relaxációs idő

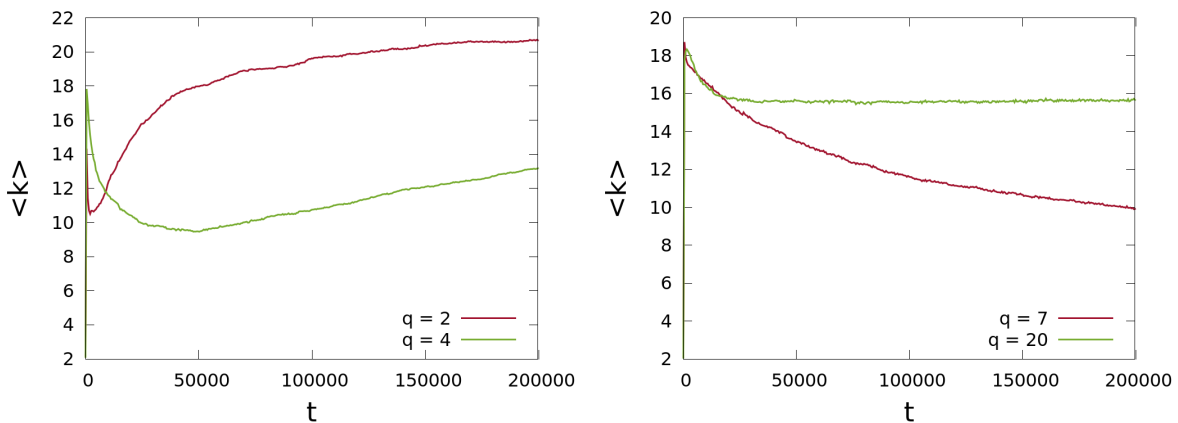
A továbbiakban megpróbáltunk utánajárni annak, hogy találunk-e más, fázisátalakulásra utaló jelet is az adatsorokban. Abban bízunk, hogy az átlagos hasonlóság és az átlagos fokszám által mutatott viselkedés együtt jár a rendszer relaxációs idejének megváltozásával. A relaxációs idő vizsgálatához az $\langle o \rangle$ és $\langle k \rangle$ mennyiségek időfüggését

leíró adatokra volt szükség, melyeket adott F és változó q értékek mellett a 3.3. ábrákon mutatok be.



(a) $\langle o \rangle$ a fázisátalakulást megelőzően

(b) $\langle o \rangle$ a fázisátalakulást követően



(c) $\langle k \rangle$ a fázisátalakulást megelőzően

(d) $\langle k \rangle$ a fázisátalakulást követően

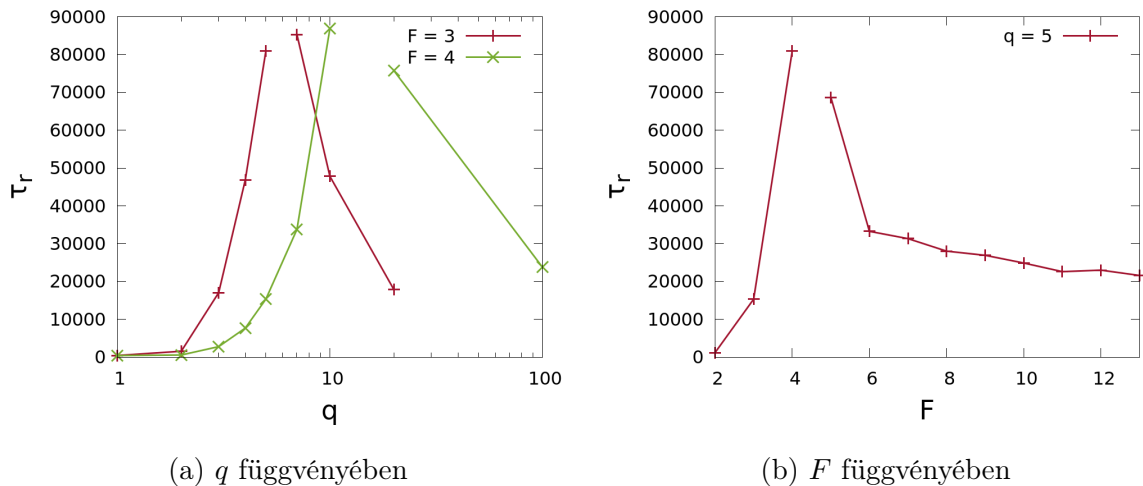
3.3. ábra. Az $\langle o \rangle$ átlagos hasonlóság és a $\langle k \rangle$ átlagos fokszám a t szimulációs idő függvényében, $F = 4$ számú tulajdonság mellett

A 3.3 ábrák alapján megállapítható, hogy $\langle o \rangle$ különböző q értékek esetén eltérő sebességgel relaxál. Hasonló jelenséget figyelhetünk meg a $\langle k \rangle$ átlagos fokszám időfüggését leíró görbéken is, bár azok esetében komplexebb folyamat állhat a háttérben. Először az átlagos fokszám értéke hirtelen megnő, majd ezt követően kezd el csak relaxálni. Feltételezésünk szerint elsőként egy rövidtávú folyamat során létrejönnek a kapcsolatok, majd ezt követően a nódusok egocentrikus hálózatának a stacionárius állapothoz történő relaxációja alakítja ki a hálózat szerkezetét.

A relaxációs idő meghatározását az $\langle o \rangle$ -t görbékre való függvényillesztéssel 3.1) végeztem el, mivel azok szabályosabb viselkedést mutattak a $\langle k \rangle$ -t görbékhez képest.

$$g(t) = a - b \cdot e^{-\frac{t}{\tau_r}} \quad (3.1)$$

A függvényillesztés során kapott τ_r értékeket a 3.4. ábrákon mutatom be.



3.4. ábra. A τ_r relaxációs idő

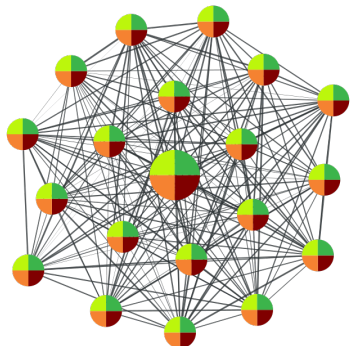
Fontos megjegyezni, hogy F és q csak egész értékeket vehetnek fel. Ez azt jelenti, hogy a kritikus pontot nem tudjuk tetszőlegesen megközelíteni, mint például a ferromágneses fázisátalakulás során.

Mivel a rendszerünk mérete véges ($N=10000$ nódust vizsgáltunk), ezért nem várjuk azt, hogy a mért relaxációs idő ténylegesen divergáljon. Ennek ellenére rendkívül nagy csúcsokat láthatunk a 3.4. ábrákon, a relaxációs idő F és q függvényében is divergálás-hoz hasonló viselkedést mutat. Ez megerősíti az általunk feltételezett fázisátalakulás tényét.

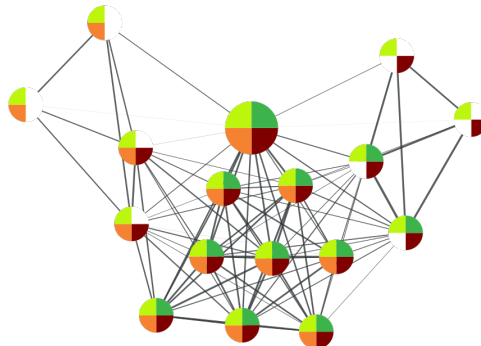
3.2. A jelenség háttere

Az eddigiek alapján tehát arra jutottunk, hogy $\langle o \rangle$ rendparaméter értéke szerint két különböző fázisban (rendezett és rendezetlen) lehet a rendszerünk. Ahhoz, hogy

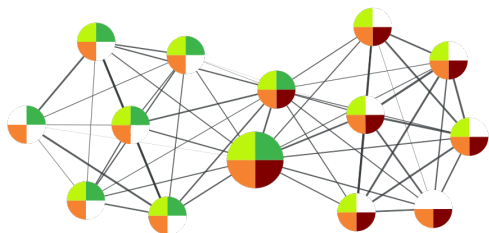
ezen két fázis kialakulásának hátterét felfedhessem, a 3.5. ábrákon négy egocentrikus hálózatot mutatok be $F=4$ esetében, különböző q értékek mellett.



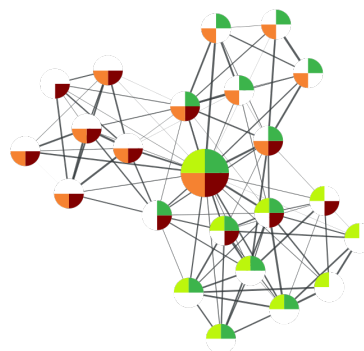
(a) $q = 2$



(b) $q = 4$



(c) $q = 7$



(d) $q = 20$

3.5. ábra. Példa a létrejött egocentrikus hálózatokra $F = 4$ esetén, különböző q értékek mellett. A legnagyobb kör jelöli az egót. Az egyes körcikkek az adott nódus lehetséges tulajdonságainak felelnek meg, a fehértől különböző színek az egoéval megegyező tulajdonságokat mutatják.

A 3.5. ábrákon azt látjuk, hogy a kritikus pont alatt, attól távol (3.5a. ábra) a kiválasztott nódus (nagyobb mérettel megkülönböztetve) minden ismerősével teljesen megegyező tulajdonságokkal rendelkezik. Az egocentrikus hálózat ekkor egy nagy klikk (teljes gráf). A kritikus pont felé közeledve (3.5b. ábra) már megjelennek olyan ismerősök is, akik nem minden tulajdonságban egyeznek meg a kiválasztott egyénnel. A kritikus pont felett egy-egy tulajdonság vagy tulajdonsághalmaz köré csoportosulva

jól látható közösségek alakulnak ki (3.5c. ábra), melyek száma q növelésével egyre nő (3.5d. ábra).

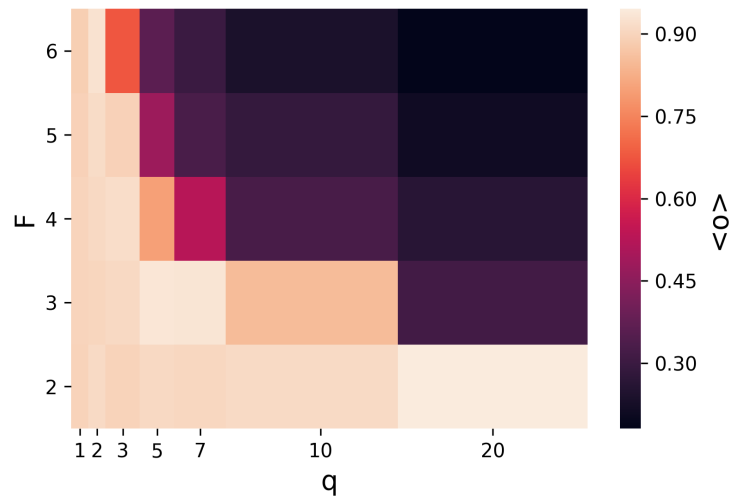
A rendezett fázisban egy adott nódus ismerősei nagyon hasonlítanak hozzá, szinte ugyanolyanok szociális tulajdonságaikat tekintve, míg a rendezetlen fázisban sokkal jobban különböznek az ismerősök \mathbf{v} tulajdonságvektorai a kiválasztott nódusétól. Véleményünk szerint ennek oka az, hogy az átalakulási pontot elérve az adott egyén már nem tud az igényeinek megfelelő, azonos tulajdonságú ismerősre szert tenni, így új kapcsolatok után kutatva olyanokkal is ismeretséget fog kötni, akikkel csupán egy-két közös tulajdonságon osztoznak. Ehhez szorosan kapcsolódva magyarázható az átlagos foksám viselkedése: $\langle k \rangle$ értéke a kritikus pont felé haladva - a rendparaméterek növelése során - feltételezésünk szerint azért lesz alacsonyabb, mert az egyén úgy reagál a hozzá hasonlók számának csökkenésére, hogy inkább nem hoz létre több kapcsolatot. Az átalakulási pont után azonban kilép a rendezett fázisból, és kisebb hasonlóságú egyénekkel is keresni fogja az ismeretséget, így az átlagos foksám újból növekedésnek indul.

Mivel az egyes tulajdonságok értékét véletlenszerűen választjuk ki q lehetséges érték közül, és összesen F tulajdonsággal írunk le egy egyént, ezért q^F különböző \mathbf{v}_i tulajdonságvektort hozhatunk létre. Így tehát egy adott tulajdonságkombináció megvalósulásának valószínűsége $p_i = \frac{1}{q^F}$. Ez a valószínűség közelítőleg megadja az ezen kombinációval rendelkező egyének N' , és az összes egyén N számának hányadosát (3.2).

$$N' = \frac{N}{q^F} \quad (3.2)$$

A fázisátalakulás akkor történik meg, amikor $\frac{N}{q^F} \sim \langle k \rangle$, hiszen ez az a pont, ahol az emberek képtelenek megfelelő számú, velük azonos tulajdonságú barátot találni. Az összefüggés alapján termodinamikai limeszben ($N \rightarrow \infty$ esetén) nem lesz fázisátalakulás tapasztalható a hálózatban, hiszen mindig rendelkezésre fog állni elegendő számú, teljesen megegyező tulajdonságvektorú nódus.

Az $\langle o \rangle$ átlagos hasonlóság F és q kontrollparamétereiktől való függését a 3.6. ábrán látható szintériképen mutatom be. A sötét rész határa jó közelítéssel megfelel a (3.2) egyenletnek.



3.6. ábra. Az $\langle o \rangle$ átlagos hasonlóság értéke az F és q paraméterek függvényében

A társadalomban az emberek tipikusan nagy számú tulajdonsággal - így nagy F értékkel - rendelkeznek, aminek eredményeként a társadalom sokszínű lesz, átfedő közösségek alakulnak ki. Azonban vannak olyan folyamatok, melyek csupán néhány tulajdonság hangsúlyossá válása miatt jönnek létre, és így szegregációt eredményeznek. Ilyen folyamatra példa az internetes szociális hálózatokban az ún. véleménybuborék (echo chamber) kialakulása is, melynek következményeként az adott felhasználó csak a saját maga elvárásainak, világnézetének megfelelő tartalmakkal találkozhat az online közösségi oldalakon.

Felmerül a kérdés, hogy miként alakulhatnak ki az interneten véleménybuborékok, ha az emberek F értéke nagy? Kézenfekvő lehet a felelősséget a számítógépes algoritmusokra hárítani, amelyek igyekeznek minél jobban kiszolgálni a felhasználói igényeket. Azonban vizsgálataink nyomán azt mondhatjuk, hogy a véleménybuborékok kialakulásának egyik fő oka az interneten elérhető egyének nagy számában keresendő. Ugyanis a nagy N érték azt eredményezi, hogy az adott egyén a virtuális térben képes lesz a számára legfontosabb szociális tulajdonságokban vele megegyező ismerősöket találni (3.2 alapján ilyen felhasználó nagy számban lesz jelen a hálózatban), így a szegregációt eredményező, rendezett fázisba tud kerülni.

4. fejezet

Összehasonlítás adatokkal

A vizsgált modell használhatóságának megállapításához elengedhetetlen, hogy az eredményeket valamilyen módon össze tudjuk hasonlítani tényleges hálózatok adataival. Erre a legkézenfekvőbb megoldás az internetes közösségi portálok (pl. Facebook, Twitter) adatainak elemzése. Témavezetőm kutatócsoportjának rendelkezésére állt az iWiW internetes közösségi oldal adatbázisa, amely tartalmazza a felhasználók anonimizált adatait, valamint az egyének kapcsolatrendszerét. A továbbiakban ezen adatokat vettem össze a szimulációk eredményeivel.

4.1. iWiW

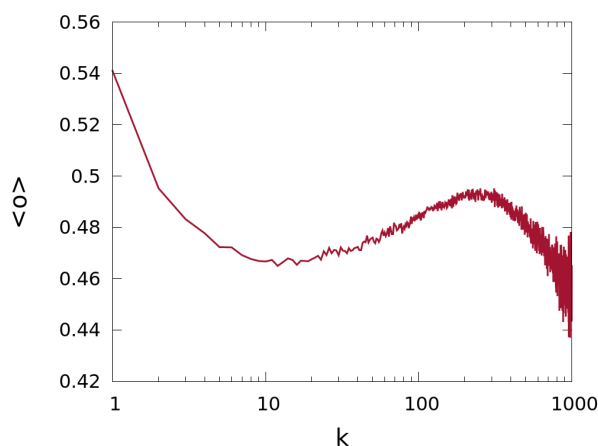
Az iWiW egykor Magyarország legnépszerűbb közösségi oldala volt, több, mint 4,5 millió felhasználóval büszkélkedhetett. Így akkoriban az interneteléréssel rendelkező lakosság [16] közel kétharmadát tagjai között tudhatta, ami kiemelkedően jó aránynak tekinthető. A felhasználók által megadott tulajdonságokat alapul véve lehetőségem nyílt az $\langle o \rangle$ átlagos hasonlóság vizsgálatára. A felhasznált attribútumok listáját, valamint azt, hogy azokat a felhasználók hány százaléka adta meg az oldalon, a 4.1. táblázatban foglaltam össze.

A 2. fejezetben definiált $\langle o \rangle$ átlagos hasonlóság kiszámításának módját úgy módosítottam az iWiW-es adatok vizsgálata során, hogy csak azon tulajdonságokat vettem figyelembe o_{ij} kiszámításakor, melyeket mind az i , mind a j egyén megadott magára vonatkozóan.

Tulajdonság	Rendelkezésre áll (%)
Nem	100
Lakóhely	88,6
Életkor	61,1
Iskolázottsági szint	52,5

4.1. táblázat. A felhasználók vizsgált tulajdonságai és az azokat az oldalon megadók százalékos aránya

Elsőként megvizsgáltam, hogy az $\langle o \rangle$ átlagos hasonlóság hogyan függ a felhasználók k fokszámától. A kapott grafikont a 4.1. ábrán mutatom be. Előzetesen azt vártam, hogy a fokszám függvényében monoton csökkenő függvényt fogok kapni.



4.1. ábra. Az $\langle o \rangle$ átlagos hasonlóság értéke a k fokszám függvényében - iWiW

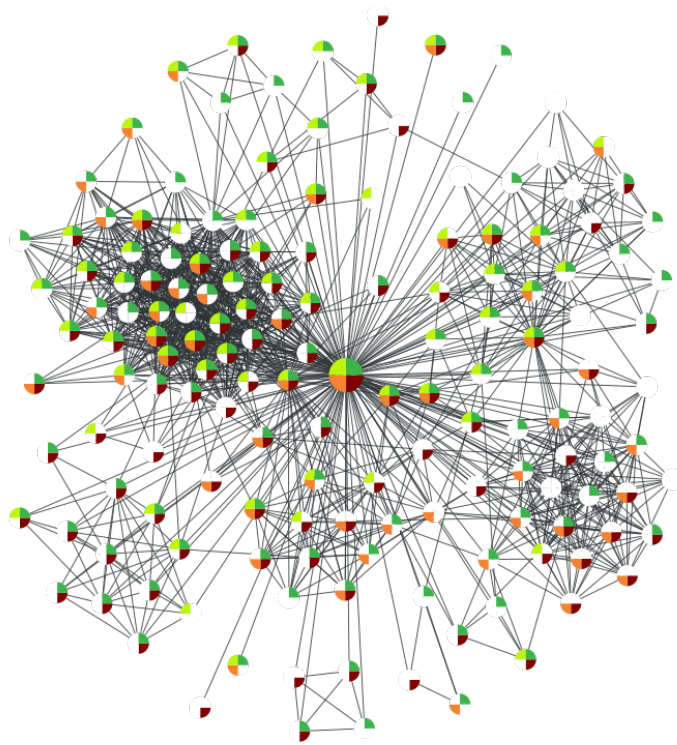
A 4.1. ábrán azt látjuk, hogy a kapott függvény - bár alacsony, illetve magas k értékekre a várt módon viselkedik - várakozásainkkal ellentétben nem monoton. Abban az esetben, ha egy embernek kevés kapcsolata van a hálózaton, akkor nagyobb a valószínűsége, hogy azokat a hozzá hasonlók közül választotta ki. Ez látható is a grafikonon, kis k -ra gyorsan csökken $\langle o \rangle(k)$ értéke. Nagy k mellett feltételezhető, hogy a virtuális ismerősök már nem felelnek meg valódi szociális kapcsolatnak, tehát a szociális hasonlóság - azaz $\langle o \rangle(k)$ - csökkenő függvény lesz. Azonban $k=200$ fokszám körül meglepő viselkedés tapasztalható, valamilyen nem triviális oknál fogva a $\langle o \rangle$ átlagos hasonlóság emelkedni kezd. Ez egy olyan jellegzetes vonás, amit a modell esetében is lehetséges és

érdeemes ellenőrizni.

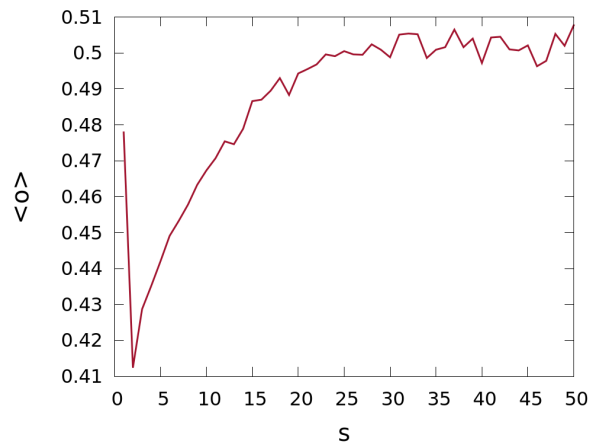
Az átlagos hasonlóság görbéjén $k=200$ fokszám környékén megtalálható csúcs feltételezhetően Dunbar elméletének [11] (ld. 1.4. fejezet) segítségével magyarázható. Fontos megjegyezni, hogy az iWiW-en mért fokszám nem az egyén ismerőseinek számát jelenti [10], hanem csak a valós kapcsolatokból az iWiW-en megjelenők számát. A kiválasztott egohoz legerősebben kötődő ismerősök az első két Dunbar-rétegben vannak (ez közelítőleg 15 ismerőst jelent). Meglátásunk szerint ezek a linkek nem egy kiválasztott közösségbe mutatnak, hanem több különböző közösség az ego számára legfontosabb nódusaihoz. Így érthető, hogy kis k értékek mellett miért csökken gyorsan az átlagos hasonlóság. Ezt követően az adott felhasználó kapcsolatba lép a valóságban hozzá kevésbé kötődő ismerőseivel is a közösségi oldalon. Abban az esetben, ha a legnagyobb közösség egyben a leghasonlóbb is, akkor a külső rétegeknek megfelelő kapcsolatok hálózathoz adását követően nagyobb lesz az $\langle o \rangle$ átlagos hasonlóság. Ezen állítás az intuícióval erősen ellentétes, hiszen azt várhatnánk, hogy egyre több embert hozzáadva a hálózathoz az átlagos hasonlóság csökkenni fog az emberek sokféleségéből adódóan.

A modell által létrehozott hálózattal való összehasonlítás, valamint a közösségek vizsgálatának céljából a 4.2. ábrán egy, az iWiW közösségi oldalon kiépült egocentrikus hálózatot ábrázoltam.

A 4.2. ábrán megfigyelhetjük, hogy a közösségek várakozásainknak megfelelően tipikusan egy-két tulajdonságot alapul véve szerveződnek. Emellett látható az is, hogy ebben a hálózatban a legnagyobb közösség valóban a leghasonlóbb is egyben. Azt, hogy ez az iWiW esetében globálisan is teljesül-e, az $\langle o \rangle$ átlagos hasonlóságot az s közösségméret függvényében vizsgálva tudhatjuk meg. A kapott görbét a 4.3. ábrán mutatom be. A közösségek felderítése során az Infomap algoritmust (1.5. fejezet) használtam.



4.2. ábra. Példa egy egocentrikus hálózatra - iWiW

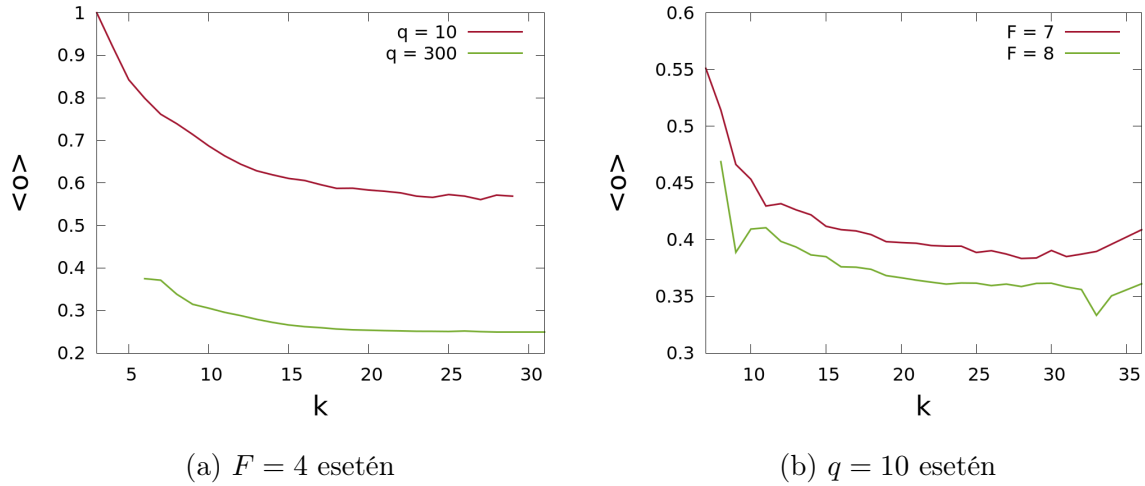


4.3. ábra. Az $\langle o \rangle$ átlagos hasonlóság értéke az s közösségméret függvényében - iWiW

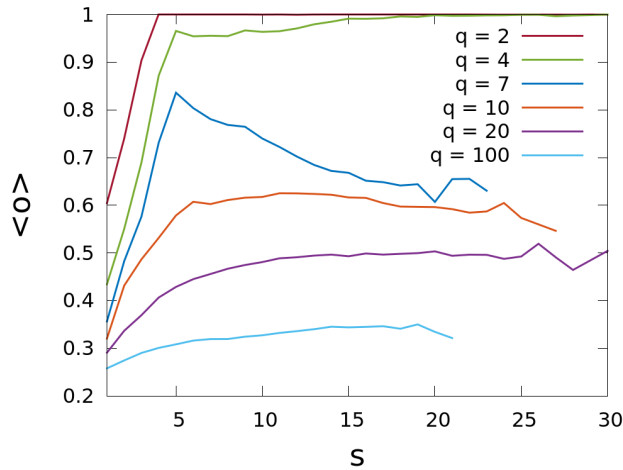
A 4.3. ábrán jól látható, hogy s növelésével egy adott s értékig $\langle o \rangle$ monoton nő. Ez a megfigyelés igazolja azon felvetésünket, amely szerint a nagyobb közösségek egoval való átlagos hasonlósága nagyobb lesz.

4.2. Modell

A szimulációk futtatásának eredményeként, különböző F és q értékek mellett a 4.4. és a 4.5. ábrákon látható grafikonokat kaptuk.



4.4. ábra. Az $\langle o \rangle$ átlagos hasonlóság a k foksám függvényében - modell



4.5. ábra. Az $\langle o \rangle$ átlagos hasonlóság értéke az s közösségméret függvényében, $F = 4$ és változó q értékek mellett - modell

Megállapíthatjuk, hogy a 4.5. ábra alapján van olyan paramétertartomány, amely esetében az átlagos hasonlóságot a közösségméret függvényében leíró görbe az iWiW esetében kapotthoz (4.3. ábra) nagyon hasonló ($F=4$ és $q \geq 10$ mellett). A 4.4. ábrák alapján az átlagos hasonlóság fokszámfüggése viszont már nem felel meg az iWiW-en

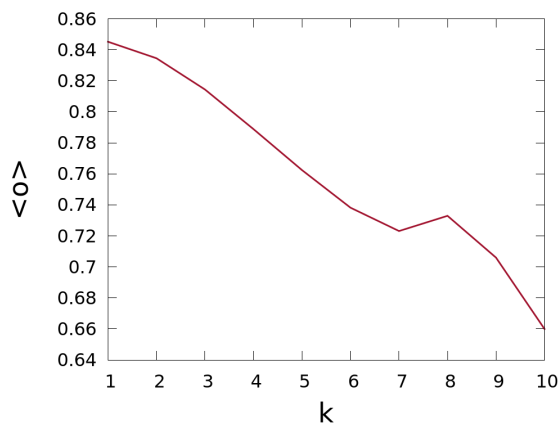
látottaknak (4.1. ábra), nem jelenik meg csúcs nagyobb k értékek esetén. Valójában viszont nem is várjuk azt, hogy megjelenjen a grafikonon egy csúcs, ugyanis a vizsgált hálózat minden egyén teljes kapcsolatrendszerét tartalmazza, így az 1.3. fejezetben tárgyalt effektus itt nem jelenik meg.

Ahhoz, hogy egy internetes közösségi portál kapcsolatrendszeréhez hasonló hálózatot tudjunk elemezni, szükség van egy ún. linkszelekciós mechanizmusra (sampling), amely által a teljes hálózatból előállíthatunk egy réteget [10]. Az átlagos hasonlóság fokszámfüggésének reprodukálásához a szimulációk eredményeként kapott hálózatokból véletlenszerűen, az élek w_{ij} súlyával arányos valószínűséggel, exponenciálisan csökkenő eloszlás szerint szelektáltam ki éleket. Az így kapott hálózat már megfeleltethető a társadalom egy rétegének.

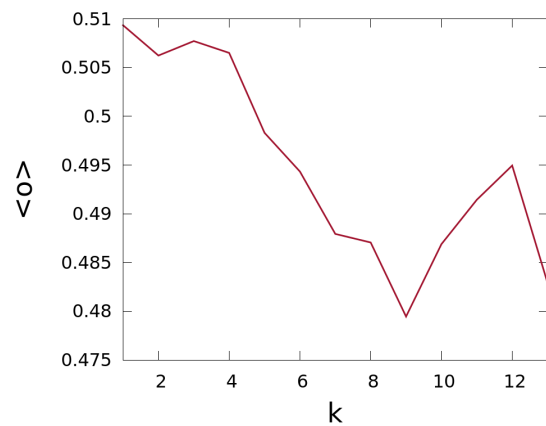
Az 1.3. fejezetben leírtakat követve a linkszelekciót a $P(a_i)$ eloszlás szerint, $A=0,3$ értéket választva végeztük el. Egy i és egy j nódus közötti link kiválasztását megadó p_{ij} valószínűség kiszámítása során figyelembe vettük a link w_{ij} súlyát is. Mivel a linkek súlyai nagyságrendileg 10^{-2} és 10 közé estek, ezért azokat átskáláztuk 0.25 és 2 közé a (4.1) képlet szerint.

$$p_{ij} = \min\{a_i, a_j\} 2^{\log_{10}(w_{ij})} \quad (4.1)$$

A hálózat „ritkítését”, azaz a linkszelekciót követően az újonnan kapott hálózatok esetében is megvizsgáltam az átlagos hasonlóság fokszámtól és közösségmérettől való függését, melyeket a 4.6. és a 4.7. ábrákon mutatok be.

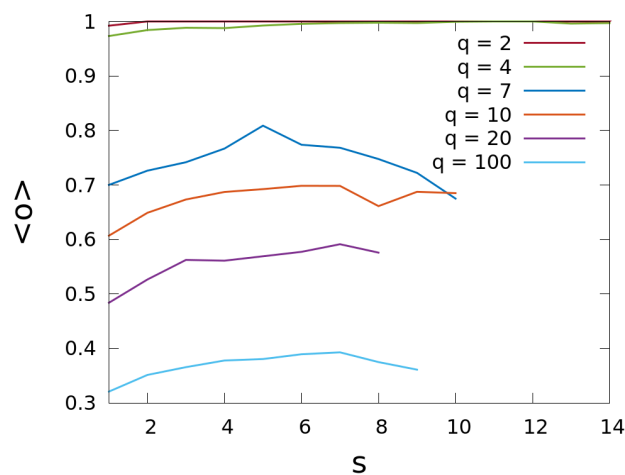


(a) $F = 4$ és $q = 7$ esetén



(b) $F = 6$ és $q = 10$ esetén

4.6. ábra. Az $\langle o \rangle$ átlagos hasonlóság a k foksám függvényében - ritkített modell



4.7. ábra. Az $\langle o \rangle$ átlagos hasonlóság értéke az s közösségméret függvényében, $F = 4$ és változó q értékek mellett - ritkített modell

A 4.6. ábrákon két olyan paraméterkombinációhoz tartozó görbe látható, melyek esetében úgy tűnik, hogy a várt viselkedés tapasztalható. Azt gondoljuk, hogy ezek a csúcsok már nem véletlenül jelentek meg a görbéken, és közülük lehet az iWiW-en megfigyelt jelenségekhez (4.1. ábra). A probléma esetünkben az lehet, hogy már a kiindulási hálózat fokszáma is alacsony, így a „ritkítást” követően túlságosan ritka hálózatot kapunk. A jövőben ezt a kérdéskört részletekbe menően tervezzük vizsgálni.

A 4.7. ábrán látható, hogy az átlagos foksám közösségmérettől való függése nem módosult jelentősen a linkszelekció során.

5. fejezet

Összefoglalás, kitekintés

TDK dolgozatomban bevezettünk egy szociális hálózatokat leíró modellt, ami a korábbi modellekkel szemben figyelembe veszi az emberi kapcsolatok létrejötte során a homofiliát és a kapcsolatok öregedését is.

Megfigyeltem, hogy a rendszer két különböző fázist mutat az F és q paraméterek függvényében. Adott rendszerméret esetén kis F és q értékeknél a rendezett fázisban vagyunk, ekkor egy kiválasztott egyén ismerősei vele teljesen azonos tulajdonságokkal rendelkeznek. Azonban a rendszer egy rendezetlen fázissal is rendelkezik, amelyben egy kiválasztott nódustól ismerősei jelentősen különböznek, a közösségek pedig az emberek egyes tulajdonságcsoportjai alapján szerveződnek.

A rendszer tanulmányozásához bevezettem egy rendparamétert, az $\langle o \rangle$ átlagos hasonlóságot, melynek értéke a rendezett fázisban egyhez közeli, a rendezetlen fázisban pedig alacsony. A másodrendű fázisátalakulásokhoz hasonlóan azt találtam, hogy a fázisátmenetet divergáló relaxációs idő kíséri. Egyszerű indoklás segítségével sikerült azonosítanunk az átalakulási pontot, amelyet akkor érünk el, amikor az egyén már nem tud kellő számú, hozzá teljesen hasonló ismerőst találni. Ebből kifolyólag termodinamikai limeszben nincs fázisátalakulás. Így tehát abban az esetben, ha megfelelően nagy számú ember kerül szociális kapcsolatba egymással (például a Facebook közösségi oldalon), kialakulhatnak teljesen homogén közösségek (ld. echo chamber).

Végül a vizsgált modell alapján kapott eredményeket összevetettem az iWiW internetes közösségi oldal adataival. A modell rendparaméterét, az átlagos hasonlóságot a fokszám függvényében ábrázolva a várt monoton csökkenő függvény helyett egy olyan

görbét kaptam, amelyen egy csúcsot láthatunk a Dunbar-számhoz közel. Ezt azzal magyarázhatjuk, hogy a kiválasztott egohoz legerősebben kötődő ismerősök az első két Dunbar-rétegben vannak (ez közelítőleg 15 ismerőst jelent). Meglátásunk szerint ezek a linkek nem egy kiválasztott közösségbe mutatnak, hanem több különböző közösség az ego számára legfontosabb nódusaihoz. Így érthető, hogy kis fokszámértékek mellett miért csökken gyorsan az átlagos hasonlóság. Ezt követően az adott felhasználó kapcsolatba lép a valóságban hozzá kevésbé kötődő ismerőseivel is a közösségi oldalon. Abban az esetben, ha a legnagyobb közösség egyben a leghasonlóbb is, akkor a külső rétegeknek megfelelő kapcsolatok hálózathoz adását követően nagyobb lesz az átlagos hasonlóság. Ebből az következik, hogy a nagyobb közösségek jobban hasonlítanak az egohoz. Ezt a rendelkezésekre álló adatokon mérésével is igazoltam.

A modell eredményeként kapott hálózatból [10] alapján, linkszelekció segítségével szimuláltam egy internetes közösségi oldalon létrejövő kapcsolatrendszert. A kapott eredmények - bár nem bizonyító erejűek - azt mutatják, hogy az adatokhoz hasonlóan a modell alapján kapott görbék esetében is megjelenhet egy csúcs az átlagos hasonlóságot a fokszám függvényében leíró görbén, míg a közösségméret függvényében monoton növekedő átlagos hasonlóság-görbét a paraméterkombinációk nagy része mellett visszakapjuk. A linkszelekció által adott eredmények pontosabb validálása az egyik rám váró jövőbeli feladat.

Irodalomjegyzék

- [1] M. McPherson, L. Smith-Lovin, and J. M. Cook, „Birds of a feather: Homophily in social networks,” *Annual review of sociology*, vol. 27, no. 1, pp. 415–444, 2001.
- [2] R. Dunbar, *How many friends does one person need?: Dunbar’s number and other evolutionary quirks*. Faber & Faber, 2010.
- [3] J. M. Kumpula, J.-P. Onnela, J. Saramäki, K. Kaski, and J. Kertész, „Emergence of communities in weighted networks,” *Physical Review Letters*, vol. 99, no. 22, p. 228701, 2007.
- [4] G. Kossinets and D. J. Watts, „Empirical analysis of an evolving social network,” *science*, vol. 311, no. 5757, pp. 88–90, 2006.
- [5] M. S. Granovetter, „The strength of weak ties,” in *Social networks*, pp. 347–367, Elsevier, 1977.
- [6] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, „Structure and tie strengths in mobile communication networks,” *Proceedings of the national academy of sciences*, vol. 104, no. 18, pp. 7332–7336, 2007.
- [7] Y. Murase, H.-H. Jo, J. Török, J. Kertész, and K. Kaski, „Modeling the role of relationship fading and breakup in social network formation,” *PloS one*, vol. 10, no. 7, p. e0133005, 2015.
- [8] R. Axelrod, „The dissemination of culture: A model with local convergence and global polarization,” *Journal of conflict resolution*, vol. 41, no. 2, pp. 203–226, 1997.

- [9] F. Vázquez and S. Redner, „Non-monotonicity and divergent time scale in axelrod model dynamics,” *EPL (Europhysics Letters)*, vol. 78, no. 1, p. 18002, 2007.
- [10] J. Török, Y. Murase, H.-H. Jo, J. Kertész, and K. Kaski, „What big data tells: sampling the social network by communication channels,” *Physical Review E*, vol. 94, no. 5, p. 052319, 2016.
- [11] R. I. Dunbar, „The social brain hypothesis,” *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, vol. 6, no. 5, pp. 178–190, 1998.
- [12] E. Berán, Z. Unoka, P. Soltész, A. Rácz, P. Kardos, and C. Pléh, „Az ego-központú szociális háló és a kötődési stílus összefüggései,” *Magyar Pszichológiai Szemle*, vol. 72, no. 1, pp. 47–68, 2017.
- [13] A.-L. Barabási, *Network Science*. Cambridge University Press, 2016.
- [14] D. Hric, R. K. Darst, and S. Fortunato, „Community detection in networks: Structural communities versus ground truth,” *Physical Review E*, vol. 90, no. 6, p. 062805, 2014.
- [15] Yohsuke Murase, Hang-Hyun Jo, Török János, Kertész János és Kimmo Kaski munkája alapján, 2018. augusztus.
- [16] Eurostat, „Households with broadband access.” URL: <http://ec.europa.eu/eurostat/tgm/table.do?tab=table&plugin=1&language=en&pcode=tin00073>. Online; 2018. október 25.