

# AZ EMBERI VISELKEDÉS DINAMIKÁJÁNAK SZÁMÍTÓGÉPES VIZSGÁLATA

## COMPUTATIONAL STUDIES OF HUMAN DYNAMICS

Az MTA Doktora cím elnyeréséhez készített

tézisfüzet

**Dr. Karsai Márton**



Rényi Alfréd Matematikai Kutatóintézet

Hálózat és Adattudományi Tanszék

Central European University

2022



## Bevezetés, a kutatások előzménye

A dinamikus emberi viselkedési minták számítógépes vizsgálata viszonylag új tudományterület, ami nagyban támaszkodik más tudományágakban elért eredményekre. Erre a területre én a fizika irányából érkeztem, így munkáim során mindig is a "miért?" kérdésre kerestem a választ, és nem elsősorban új módszerek vagy technológiák kifejlesztésére összpontosítottam. Munkáimban közös, hogy habár a tudományos jelenségek, amelyeknek a megértését célként tűztem ki, számos tudományterülethez kapcsolódnak, ezekhez elsősorban a fizikából és számítástudományból kölcsönzött elméletek és módszerek segítségével közelíték. Ez a hozzáállás egy nagyon változatos és igazán interdiszciplináris tudományos portfóliót eredményezett, ami első olvasatban talán nem illeszthető be egyértelműen a tudományágak hagyományos kategóriáiba. Ezt a multidiszciplinaritást azonban előnyként értékelem, mivel teret biztosít látszólag távoli tudományágak ismereteinek kombinálására, számos fontos és érdekes probléma megértésének érdekében.

Tudományos munkáim természetesen a komplex rendszerek fizikájához kapcsolódnak leginkább. Ennek a területnek a fő célja annak megértése, hogyan alakulnak ki makroszkopikusan megfigyelhető kollektív jelenségek nagyszámú részecske, ágens vagy egyén kölcsönhatásának eredményeként, úgy hogy azok egyéni tulajdonságai nem elégségesek a jelenségek megmagyarázására. Ilyen jelenségeket tradicionálisan a fázisátmenetek és kritikus jelenségek statisztikus fizikájának keretében szokás tanulmányozni [64]. Azonban ilyen kollektív jelenségek nemcsak a fizikai rendszerekben fordulnak elő, hanem számos más területen, például biológiai, társadalmi és gazdasági rendszerekben is. Egy világvárvány globális terjedése [13], egy széles körben elfogadott vélemény kialakulása [15], az állatok kollektív vándorlása [75], a daganatok növekedése [59] vagy az idegsejtek összehangolt aktivitása [18] mind jó példa kollektív jelenségekre, amelyek megértéséhez a komplex rendszerek fizikai megközelítése jól használható [63].

Az egyéni és kollektív viselkedési minták pontos megfigyelése a társadalmi rendszerekben a közelmúltig szinte lehetetlen volt. Kollektív jelenségek kvantitatív vizsgálatához nagy mennyiségű és pontos adatra van szükség, aminek az összegyűjtése, különösen társadalmi méretekben, lehetetlennek tűnt. Sok szempontból könnyebb volt megfigyelni apró baktériumok vagy távoli galaxisok sokaságát mint a társadalomban körülöttünk élő embereket [73]. Ezeket az alapvető korlátokat döntötte le a digitális adatforradalom [48, 17]. A digitalizált társadalmak mára elérték azt a szintet, hogy 1,2

évente több társadalmi-gazdasági adatot generálnak, mint az emberiség a teljes korábbi történelme során összesen. Ennek következményeként milliók egyéni és kollektív viselkedési mintáit tudjuk "in vivo" követni, anélkül hogy közvetlenül befolyásolnánk az emberek döntéseit a megfigyeléseinkkel [73].

Az ilyen fajta adatgyűjtést széles körben alkalmazott új technológiák megjelenése tette lehetővé. A mobiltelefonunk vagy a különböző online szolgáltatások használata során rengeteg adat gyűlik össze, aminek feldolgozása és elemzése új számítási módszerek kifejlesztéséhez vezetett [14]. Nagy teljesítményű számítási erőforrásokat igénybevevő "social computing" rendszerek [54] vagy a statisztikai alapokra épülő fejlett gépi tanulási módszerek [46, 24] megjelenése mind hozzájárult az emberi viselkedés adatvezérelt kutatásának sikeréhez. Ugyanakkor az így elért eredmények rávilágítottak a komplexitásról alkotott tudásunk korlátaira is. A kollektív emberi viselkedési minták természetszerűleg más törvényszerűségek mentén alakulnak ki, mint a fizikában megfigyelt hasonló jelenségek. Így azok az alapvető fogalmi, elméleti és modellezési szempontok, amelyek korábban a komplex rendszerek megértésének alapját képezték, sokszor alkalmatlannak bizonyultak az újonnan megfigyelt jelenségek megértéséhez.

Ebben az összefüggésben a hálózattudomány megjelenése [8] fontos mérföldkő volt a komplex rendszerek fogalmi keretének kialakulásában. A hálózattudomány alapjai a statisztikus fizikában [2] és a gráfelméletben gyökereznek [35]. Ez a megközelítés olyan struktúrákkal foglalkozik, ami valós entitásokat gráf csomópontokként és a közöttük kialakuló kölcsönhatásokat pedig gráf kapcsolatokként (élekként) definiálja. Ez a leírás a komplex rendszerek többségénél hasznosnak bizonyult, mivel egy jól reprezentálható és viszonylag könnyen vizsgálható matematikai eszközt biztosított, első látásra végtelenig bonyolult rendszerek struktúrájáról [77, 9, 2, 52]. A hálózati megközelítés nagyban hozzájárult számos valós komplex rendszer jobb megértéséhez a fizika, a biológia, a kémia, és a társadalomtudományok területén. Ez a közös matematikai nyelv ahhoz a felfedezéshez vezetett, hogy megannyi látszólag független valós rendszert leíró hálózat közös tulajdonságokat mutat. Mi több, ezek a közös mintázatok annyira egybevághóak, hogy látszólag teljesen idegen rendszerekben hasonló háttér-folyamatok jelenlétére utalnak. A hálózati csomópontok kapcsolatszámainak heterogén eloszlása, a csomópontokat összekötő rövid gráftávolságok, vagy a nagy sűrűségű lokális klaszterek számtalan komplex hálózatot jellemeznek biológiától a fizikai rendszereken át a társadalmi hálózatokig. Ezek a megfigyelések vezettek a komplex hálózatok mérési és modellezési módszertanának egyre pontosabb kidolgozásához [51, 34, 21, 81, 42]. Példaként említhető a széles körben ismert Watts-Strogatz model [77],

ami a hálózatokra jellemző rövid gráftávolságok és magas klaszterezedés duális jelenségét magyarázza meg, vagy a Barabási-Albert preferenciális kapcsolódási model [9] ami a hálózati csomópontok heterogén kapcsolati eloszlására ad magyarázatot. Ahogy a modern hálózat tudomány fejlődött, egyre több hálózati tulajdonságot sikerült leírni [2, 7]. Ilyen volt a hálózatok dinamikájának leírására [34, 36] [T3], többretegűségének felismerésére [38], a kialakuló klaszterek (community) detektálására [22], a hálózatok térbeli beágyazottságának jellemzésére [11] és a magasabb rendű struktúráinak azonosítására [79, 31, 12] alkalmas matematikai leírás. Mindezek az erőfeszítések a komplex hálózatok tudományának látványos fejlődéséhez vezettek, betekintést engedve a körülöttünk található komplex rendszerek architektúrájába. Ez nagyban hozzájárult például globális világjárványok, nemzetközi pénzügyi vagy közlekedési rendszerek, az agy vagy maga a társadalom hálózati struktúrájának megértéséhez.

A digitális adatforradalommal párhuzamosan a komplex rendszerek, a gépi tanulás és a komplex hálózatok tudományának közös fejlődése vezetett ahhoz, hogy korábban lehetetlennek tűnő megfigyeléseket tudjunk tenni egyének és csoportok társadalmi viselkedéséről. Ez a fejlődés hívott életre egy új tudományterületet, amit számítógépes társadalomtudományként (*computational social science*, CSS) ismer a tudományos közösség [44, 45, 33, 19]. Az ilyen típusú kutatások fő célja a társadalmi rendszerek kvantitatív leírása, modellezése és megértése számítás- és adatalapú módszerekkel. A CSS módszerei, azon túl, hogy minden esetben egy számítási problémaként is definiálhatóak, a hagyományos társadalomtudományok, a kognitív és viselkedéstudományok, a pszichológia, a statisztika, a fizika, az informatika és a hálózat tudomány módszertanára épülnek. Azonban a módszerek nem minden esetben bizonyultak adaptálhatónak, így az emberi viselkedés számítógépes vizsgálatához teljesen új fogalmakra is szükség volt. Ez további olyan speciális területek kialakulásához vezetett, mint a számítógépes közgazdaságtan [65] (*computational economics*), az emberi dinamika [7, 36, 61] (*human dynamics*), a társadalmi szimulációk [26] (*social simulations*) vagy a számítógépes társadalmi nyelvészet [53] (*computational sociolinguistics*). Ezen irányok megjelenése nem csupán a társadalmi jelenségek pontosabb megértését rejti magában, hanem egy lehetséges paradigmaváltást is előre vetít, amely a kvantitatív érvelés nagyobb térhódítását ígéri a társadalomtudományokon belül.

Számos terjedési folyamat, mint az információ vagy járványok terjedése, vagy új termékek és mémek adaptációja, hatékonyan tanulmányozható a hálózat tudomány és a kollektív jelenségek fizikájának módszereivel [10, 55]. Az irodalomban alapvetően kétfajta mechanizmust feltételeznek a terjedési folyamatok modellezésére. A főként bi-

ológiai terjedési folyamatok mögött egy úgynevezett *egyszerű terjedési folyamatot* lehet feltételezni, ahol egy egészséges egyén megfertőződési valószínűsége más fertőzött egyénnel való bináris kölcsönhatásoktól függ [10, 55, 37]. Ez általában sztochasztikus folyamatként értelmezhető, ahol a legegyszerűbb modellben minden időegység alatt egy beteg egyén egy bizonyos rátával fertőzhet meg egy egészséges egyént akivel kölcsönhat, ugyakkor egy másik rátával figyelembe vehető a gyógyulás (vagy immunizálódás) is. Az eredmény egy terjedési folyamat, ahol a rendszer a fenti két ráta arányának függvényében egy fázisátalakuláson megy keresztül egy kiháló és egy kifejlődő terjedési fázis között. Ez a fajta megközelítés számos módon továbbfejleszthető, és így nagyon jól alkalmazható eszközt biztosít valós járványok prediktív modellezésére, ahogy ez az elmúlt évtizedben számos pandémia, például a H1N1, Zika, Ebola, vagy COVID-19 világjárványok esetén bebizonyosodott [6, 66, 40, 49, 49, 74].

A terjedés szempontjából fontos, az egyének közötti kölcsönhatás modellezésére több fajta közelítés alkalmazható. A legegyszerűbb esetben homogén módon összekevert populációt feltételezhetünk, ahol egy időlépésen belül minden egyén mindenki mással ugyanakkora valószínűséggel találkozik [10]. Egyel pontosabb feltételezés, ha az egyénekhez egy attribútumot rendelünk (például a korukat) és így tudjuk modellezni korcsoportok közötti inhomogén kölcsönhatási mintákat [50, 57, 80, 41]. Azonban ez a megközelítés a korcsoportokon belül még mindig homogén keveredést feltételez. A kölcsönhatások legpontosabb modellezésére a társadalmi hálózat statikus vagy temporális [T3, T4] [55, 47] reprezentációja használható ami pontosan megmondja kik között terjedhet a fertőzés a hálózaton belül. Ugyanakkor ez a megközelítés egy társadalmi méretű populációra alkalmazva mind adatgyűjtési, mind adatvédelmi szempontból nehézkes, mitöbb még modellezett hálózatok esetén is igen nagy számítási kapacitást igényel.

Ez a fajta pontos hálózati leírás azonban nélkülözhetetlen az úgynevezett *komplex terjedési folyamatok* modellezésénél [16], ami főleg társadalmi terjedési folyamatok leírására irányul. Ilyen folyamatok például különböző (ál-)információk, mémek, viselkedési minták vagy új termékek vásárlásának elterjedése. Az ilyen folyamatoknál az egyénre irányuló, az ismerősöktől érkező személyes befolyás fontos szerepet játszik, és ezt úgynevezett küszöbérték mechanizmusokkal szokás modellezni [32, 76]. Ezekben a modellekben minden egyén egy  $\phi$  küszöbértékkel jellemezhető, ami azt határozza meg, hogy az ismerősei közül hányan (vagy mekkora arányban) kell, hogy egy adott viselkedési formát (információt, memet, stb) átvegyenek, mielőtt maga az egyén is adoptálja azt [76]. Az egyszerű terjedési folyamatokkal szemben ez lehet egy

determinisztikus folyamat, mivel ha egy egyén  $\phi$  küszöbértékét meghaladja a szomszédoktól érkező összes befolyás erőssége, az egyén feltétlenül állapotot vált és hasonlóan kezd el viselkedni [28]. Ez a mechanizmus lavinákhoz hasonló hirtelen kialakuló makroszkópikus jelenségeket tud előidézni, amilyenek valós esetekben is megfigyelhetők például népszerű termékek vagy mémek terjedése esetén [T7,T8] [70, 71, 70], vagy akár széleskörű társadalmi jelenségek (például demonstrációk, forradalmak) kialakulása során [30].

Bármely, a fent említett egyszerű vagy komplex terjedési folyamat esetét az egyik központi kérdés, hogy a közvetítő hálózat szerkezete hogyan befolyásolja a terjedési folyamat kritikus viselkedését [10]. Hagyományosan ezt a kérdést matematikai módszerekkel, például fok- vagy állapot-kompartimentalizált mesteregyenletekkel [27] vagy számítógépes szimulációs módszerekkel szokás modellezni [1]. Azonban az elmúlt néhány évben egy újfajta modellezési technika jelent meg, aminek fejlesztésében magam is szerepet vállaltam [T3] [25]. Ezek az úgynevezett adatvezérelt modellek számos dinamikus folyamatot valós hálózatokon szimulálnak [60], így definiálva egy, a valósághoz lehető legközelebb álló modellrendszert. Ezen túl gépi tanulási algoritmusok [23] jelöltek ki egy ígéretes irányt a dinamikus folyamatok prediktív modellezésében [58, 78]. Mindezen munkák eredményeként számos különböző hatásra sikerült rámutatni, ami a hálózati kölcsönhatások szerkezeti és időbeli heterogenitását, rétegződését, valamint szerkezeti és időbeli korrelációit azonosítja mint a dinamikus folyamatok kimenetelére ható kritikusan tulajdonságok.

A tézisem az emberi dinamika számítógépes vizsgálatára (computational human dynamics) [62] koncentrálni ami egyrészt egyéni viselkedésminták kutatását foglalja magába, így tárva fel az emberek tetteinek, autonóm döntéseinek, véleményük kialakításának, vagy térbeli és absztrakt környezetben történő mobilitásának dinamikai mintázatait. Azonban minden egyén társadalmi-gazdasági környezetbe van beágyazva, amelyet nem lehet figyelmen kívül hagyni a viselkedésük tanulmányozása során. Első közelítésben ezt az egyén közvetlen társadalmi kapcsolatainak vizsgálatával lehet figyelembe venni, amik úgynevezett egocentrikus hálózatokba rendeződnek. Következő lépésként a társadalmi hálózatot formáló csoportok vizsgálhatjuk, ahol az egyén magához hasonló másokkal áll kapcsolatban. A társadalom szintjén az egész hálózatnak (vagy annak egy megfigyelhető rész-hálózatának) a formálódása és az azon kialakuló kollektív folyamatok megértése a kutatás tárgya. Ily módon ez a terület egyrészt a számítógépes társadalomtudomány egyik részterületeként azonosítható, másrészt erősen összefonódik a hálózattudomány, dinamikus rendszerek fizikája, az adattudomány, és

a számítástudomány számos területével.

Az emberi dinamika mint tudományterület az egyéni viselkedés közvetlen megfigyelésére koncentrál [7, 29]. Azonban, ebben a dolgozatban a terület egy tágabb értelmezését javaslom, ami nemcsak a dinamikus emberi viselkedés közvetlen tanulmányozását tekinti célnak, hanem a rajta feljövő dinamikus folyamatokra gyakorolt hatásait is vizsgálja. Így módon olyan problémákat is a kutatásaim céljának tekintek, amelyek természetesen más területhez kapcsolódnak, de megértésük döntően az emberi dinamika egyes aspektusaitól függ.

## Célkitűzések

A kutatásaim célkitűzése nagy vonalakban az emberi viselkedés dinamikai rendszereinek a megfigyelése és megértése volt, a hálózattudománynak, a komplex rendszerek fizikájának és a számítás- és adattudomány eszköztárainak a segítségével. Az első céloom az emberi viselkedésdinamika megannyi aspektusát jellemző, azon időbeli heterogenitás megértéséhez kapcsolódott, amire a szakirodalom a "bursty" ("villanásos") kifejezést alkalmazza. Ez a jelenség egyéni és csoportos viselkedési mintákat jellemez ahol az események sorozatában az egymást követő események nem hasonló időközönként jelennek meg, hanem időben heterogén módon oszlanak el [36]. Jellemzően ilyen rendszerekben rövid idő alatt sok esemény történik, és ezt követően a rendszer hosszú ideig inaktív marad. Első fő célkitűzésem [T1] ebben az irányban egy entitás által generált "bursty" események közötti időbeli korrelációk jellemzése volt és annak megértése, hogy az ilyen eseménysorozatok miért rendeződnek csoportokba. Ezt követően [T2] a bursty események kollektív mintázatait vizsgáltam időben változó hálózatokban, hogy rámutassak ezek csoportos vagy egyéni jellegére. Kollégáimmal az emberi viselkedés "bursty" dinamikájáról a közelmúltban könyvet is írtunk [36], aminek első szerzője vagyok.

Az általam vizsgált második fő kutatási irány olyan hálózati struktúrákra irányult, ahol a hálózat csomópontjait összekötő kölcsönhatások időben változnak. Ez a megközelítés sokban hozzájárult ahhoz, hogy a statikus hálózati reprezentációnál pontosabb leírást adjunk számos időben fejlődő komplex rendszerről. Az ilyen temporális hálózatok sok kihívás elé állították a hálózatkutatókat, mivel az addig kidolgozott hálózatokat leíró eszköztárat át kellett ültetni időben változó struktúrákra, ahol a kölcsönhatások sorrendje és az így megjelenő események közötti kauzalitás nagy szerepet játszik. Az első kutatási témám ebben az irányban olyan mikrokanonikus véletlen hálózati mo-



dellek definíciója volt [T3] [39, 25], amelyek kontrollált módon tudják érvényteleníteni különböző strukturális és temporális korrelációk hatásait időben változó valós hálózatokban. Az így nyert véletlen hálózatokon szimulált információterjedési folyamatokat használtam annak a kimutatására, hogy mely korrelációk fontosak az információt közvetítő temporális utak kialakulásában. Ez a munka nem csak az újfajta véletlen referencia modellek definíciója miatt volt fontos, hanem mert ez volt az egyik első példája a terjedési folyamatok adatalapú modellezésére, ahol a terjedést valós hálózatokon szimuláltunk azért, hogy a struktúra teljes komplexitását és annak hatásait figyelembe tudjuk venni. Temporális hálózatok generatív modellezésével is foglalkoztam, főleg aktivitásvezérelt hálózati modelleket használva [56, 67, 43]. A kutatásaim során arra voltam kíváncsi, hogy miként tudnak egyéni memória folyamatok a valós hálózatokban megfigyelt globális strukturális és időbeli heterogenitásokhoz vezetni [T4] [68, 69]. Ezen túlmenően arra is kerestem a választ, hogy ezek a memóriefolyamatok milyen hatással vannak a temporális hálózaton végbemenő információterjedésre [T4]. Végül az utóbbi néhány évben egy új matematikai reprezentáción dolgoztam [T5] [3, 5, 4], ami az időbeli hálózatokat statikus irányított aciklikus hálózatokként értelmezi. Ez a fajta leírás mutatott rá arra, hogy véletlen temporális hálózatok egy széles halmaza jellemezhető egy olyan fázisátalakulással, ami az irányított perkolációs univerzalitási osztályba tartozik. Hasonló jelenségeket valós temporális hálózatokban is meg lehet figyelni.

Kutatásaim jelentős része különböző terjedési folyamatok modellezésére koncentrált, azok közül is leginkább társadalmi terjedési folyamatok jobb megértését tűzve ki célul. Ezek során az akkor elérhető egyik legnagyobb online társadalmi hálózat, a Skype felhasználóit leíró teljes adatbázison dolgozhattam, ami több mint 300 millió felhasználó kommunikációs hálózatát és termékadaptációs dinamikáját írta le 8 éven keresztül. A munkám fő célja az volt, hogy a felhasználók hálózatán terjedő termékvásárlás globális mintázatát mint egy társadalmi terjedési folyamatot írjam le, és modellezzem. Ez a munka több tanulmányon [T6,T7,T8] keresztül mutatta be, hogy milyen, korábban nem várt dinamikai jellemzői vannak az ilyen társadalmi hálózaton terjedő folyamatoknak [T6,T7], valamint ezek viselkedését hogyan lehet komplex terjedési folyamatokként modellezni [T7] a Watts küszöbérték-modellből kiindulva [76]. Közelítő mesteregyenletek [28] és stabilitási elemzési módszerek, valamint nagyméretű számítógépes szimulációk eredményét felhasználva a kritikus jelenségek egy gazdag csoportját figyeltem meg. Ezek a jelenségek a hálózat, és az egyéni fertőzési határértékektől függő, a háttérben meghúzódó perkolációs folyamatokkal magyarázhatóak [T8], és akár nem-monoton [71] vagy többszörös fázisátmenetekhez is vezethetnek [72, 70].

## Új tudományos eredmények

**1. Tézis:** Számos valós rendszerekben megfigyelhető inhomogén időbeli folyamat rövid de nagyon intenzív intervallumokból épül fel, amelyeket hosszú, alacsony aktivitású időszakok választanak el egymástól. Ilyen dinamikai minták jellemzik az egyéni kommunikációs folyamatokat, de neuronok aktivitásában vagy a földrengések sorozatában is megfigyelhetők. Ilyen dinamikai jelenségek esetén az egymást követő események között eltelt  $\tau$  időintervallumok  $P(\tau) \sim \tau^{-\alpha}$  lassú lecsengésű hatványfüggvény-eloszlást követnek, ahol az események között feltételezett időbeli korrelációkat az  $A(t_d)$  autokorrelációs függvénnyel szokás jellemezni. Valós adatokból elemzéséből kiindulva, analitikusan megmutattam, hogy az autokorrelációs függvény nem alkalmas időbeli heterogén pont-folyamatokban megjelenő korrelációk egyértelmű kimutatására, mivel akár független eseményekből álló sorozatok esetén is lehet, hogy  $A(t_d) \sim t_d^{-\gamma}$  szerint cseng le (itt  $t_d$  a kezdetiként kijelölt esemény óta eltelt idő).

Megoldásként [T1] kidolgoztam egy új mérési módszert ami akár bursty esemény-sorozatokban is egyértelműen kimutatja az időbeli korrelációk jelenlétét. A módszer lényege, hogy egy pont-folyamatban az egymást követő eseményeket, melyek egy adott  $\Delta t$  maximális időablakon belül követik egymást, vonatokba csoportosíthatjuk. Ezen vonatok  $E$  hosszának eloszlása definíció szerint  $P(E) \sim a^{E-1}$  exponenciális alakot vesz fel (ahol  $a < 1$ ), ha az egymást követő események függetlenek, bármely események közötti időintervallum-eloszlás esetén. Azonban, ha az egymást követő események nem függetlenek egymástól, ez az eloszlás ettől eltérő alakot ölt. Egyéni által generált mobiltelefon-kommunikációs (hívás és sms) folyamatokat, egyéni neuronok által kibocsátott aktivitási jelek sorozatát, valamint egy adott helyen történt földrengések időpontjait vizsgálva megmutattam, hogy ezekben az igen különböző természetes és viselkedési rendszerek esetén, melyek mind a bursty aktivitást mutatnak, a bursty vonatok mérete  $P(E) \sim E^{-\beta}$  hatványfüggvény-eloszlásokkal jellemezhetőek.

A megfigyelt jelenségek magyarázatára bevezettem egy memóriavezérelt kétállapotú esemény-generáló modellt. Ebben az automatában a fenti eredmények alapján egy  $p(n) = (n/(n+1))^\nu$  alakú memória függvényt feltételeztem (ahol  $\beta = \nu + 1$ ), ami annak a valószínűségét határozta meg, hogy egy  $n$  hosszú bursty vonat esetén megjelenik-e egyik  $n + 1$  esemény, vagy a vonat véget ér. Ennek a memória függvénynek a segítségével sikerült bevezetnem nem-markovi korrelációkat a modellezett pont-folyamatba, amely így jól reprodukálta az empirikus eredményeket.

**2. Tézis:** Nagyméretű mobil kommunikációs adatok elemzésével megfigyeltem, hogy a bursty vonatok, melyek mind az egyéni, mind a csoportos kommunikációs dinamikát jellemzik, nem egyedülálló egyének vagy nagyobb csoportok által generált mintákból tevődnek össze, hanem főleg páros kommunikáció eredményei [T2]. 6.2 millió egyén egocentrikus hálózatának dinamikáját vizsgálva kimutattam, hogy ha az egocentrikus hálózatban generált csoportos bursty vonatokot szétbontom egyéni kapcsolatokra, a vonatokat jellemző statisztikai eloszlás nem változik jelentősen. Ebből az alapmegfigyelésből kiindulva több módszert dolgoztam ki az állítás empirikus igazolására, melyek mind azt mutatták, hogy bursty viselkedés inkább emberek kapcsolati dinamikáját jellemzi, mint az egyéni vagy csoportos aktivitási dinamikákat. Megmutattam, hogy ez a viselkedés nagyban független a megfigyelt vonatok  $E$  hosszától.

Kimutattam, hogy a társadalmi kapcsolatokon kialakult bursty vonatok kommunikációs egyensúlya nagyban függ a használt kommunikációs csatornától. Megfigyeléseim azt mutatták, hogy mobiltelefon hívásokból álló egyéni vonatok sokkal kevésbé kiegyensúlyozottak, mint az az adott kapcsolat globális kommunikációs egyensúlyából adódna. Ugyanakkor, az ellenkező állítás igaz rövid üzenetekből álló vonatok esetén, amik sokkal kiegyensúlyozottabbak mint azt az adott kapcsolat teljes megfigyelési periódusból számított egyensúlyából várnánk.

Végül, a megfigyelt egyensúlyi viselkedési minták megmagyarázására modellt dolgoztam ki, amiben megerősítési mechanizmusokat tételeztem fel. Egész pontosan, a vártnál kevésbé kiegyensúlyozott dinamika modellezésének érdekében feltételeztem, hogy ha egy bursty vonat első  $n$  eseményét csak az egyik  $\sigma_1$  kölcsönható fél indukálta, akkor annak a valószínűsége, hogy az  $n + 1$ -ik eseményt is ő fogja kezdeményezni az  $p_\sigma(n|\sigma_1) = n/(n + 1)$ , ami nő a vonat  $n$  aktuális méretével. Ugyanakkor, ha az üzenetek kiegyensúlyozott dinamikáját modellezzük, ez a valószínűség  $p_\sigma(n|\sigma_{n-1}) = n/(n + 1)$ , tehát akkor nagyobb, ha nem az aktuálisan aktív résztvevő generálta az utolsó eseményt. Ezen modellek segítségével sikerült reprodukálni a különböző kommunikációs csatornákra jellemző egyensúlyi dinamikákat, és így megerősítési mechanizmusok által generált nem-markovi kölcsönhatásokkal magyarázni meg a társadalmi kapcsolatokon kialakuló bursty vonatok egyensúlyát.

**3. Tézis:** Adataalapú szimulációs módszerekkel arra kerestem a választ, hogy valós temporális hálózatokon miért terjednek nagyságrendekkel lassabban szimulált információs folyamatok, mint azt a hálózatok kisvilág-struktúrája sugallná [T3]. A modellek vizsgálatához egy nagyméretű mobiltelefon kommunikációs hálózatot használtam ami

4.5 millió felhasználó 325 millió kommunikációs eseményét rögzítette. Bevezettem egy újfajta referenciamodellezési paradigmát, ami időtől függő hálózati kölcsönhatásokat kever össze úgy, hogy különböző, a hálózati dinamikát és struktúrát jellemző peremfeltételeket rögzítve tart. Az így keletkező mikrokanonikus csoportokból választott hálózatokon szimulált determinisztikus Susceptible-Infected (SI) ("kitett-fertőzött") folyamatokból arra következtettem, hogy az információterjedési folyamat sebességét leginkább a hálózatra jellemző statikus klaszterstruktúra, valamint a kölcsönhatások időben heterogén (bursty) eloszlása lassítja. Ugyanakkor, a szomszédos hálózati kölcsönhatások kauzális kapcsolódásai gyorsítólag hatnak a folyamatra. Továbbá rámutattam arra, hogy a folyamat sebességét az emberi viselkedést globálisan vezérlő napi ritmusok kevésbé befolyásolják.

A bursty hatást analitikailag is megvizsgáltam egy hatványfüggvény alakú  $P(\tau) \sim \tau^{-\alpha}$  események közötti idő eloszlást feltételezve, és megmutattam, hogy ha az eloszlás eléggé heterogén ( $\alpha > 2 + \sqrt{2}$ ), akkor a hálózaton terjedő folyamat lassúbb, mint amikor a kölcsönhatások dinamikáját egy Poisson folyamattal modellezzük.

**4. Tézis:** Valós temporális hálózatok elemzésével, valamint numerikus és analitikus modellezési módszerekkel rámutattam egy új statisztikai törvényszerűsége, ami az emberek közötti kapcsolatok ismétlődését vezérli, és így nagyban meghatározza az egocentrikus hálózatok fejlődését is [T4]. Megmutattam, hogy ez a törvényszerűség egyszerű memóriefolyamatokkal megmagyarázható, és kritikus hatása van a temporális hálózatokon fejlődő információterjedési folyamatokra.

A kutatás során egy  $\sim 6.2$  millió anonim felhasználó  $\sim 634$  millió híváseményét tartalmazó mobiltelefon-hívási hálózatot elemeztem, amit egy temporális hálózatként reprezentáltam. Az egyes egyének egocentrikus hálózatának időfejlődését elemezve megállapítottam, hogy a megfigyelés kezdetétől számítva annak a valószínűsége, hogy egy egyén egy új, még nem megfigyelt kapcsolatot létesít, csökken a már megfigyelt kapcsolatok számával. Másképpen fogalmazva, minél több ismerőse van egy egyénnek (akiket már megfigyeltünk), annál kisebb a valószínűsége, hogy új ismeretséget fog kialakítani. Skálázási analízissel megmutattam, hogy ez egy általános viselkedési minta ami minden emberre igaz, függetlenül attól hogy összesen hány ismerőse van egy egyénnek. Mitöbb, ez a csökkenő valószínűségi függvény meglepően jól illeszkedik egy egyszerű univerzális függvény alakra.

Ezt a törvényszerűséget feltételezve definiáltam egy aktivitás vezérelt temporális hálózati modellt [56] [T4]. A modellben az adatokból szintetizált valószínűségi függ-

vényt egy memóriefolyamatként értelmeztem, ami azt határozza meg, hogy egy adott pillanatban aktív csomópont mekkora valószínűséggel fog kapcsolódni egy véletlenül kiválasztott másik csomóponthoz, vagy ennek komplementereként, mekkora valószínűséggel fog megismételni egy kölcsönhatást egy már korábban kapcsolódott szomszédal. Ezzel a modellel rámutattam, hogy ez a fajta memóriefolyamat nagy hatással van a kialakult globális hálózati struktúrára mivel (a) lelassítja az egocentrikus hálózatok méretének növekedését, és így a hálózat kapcsolódó komponenseinek növekedését is; (b) visszafogja a globális fokheterogenitást, és így a valós megfigyeléshez sokkal jobban illeszkedő, lassan lecsengő fokeloszlást indukál; (c) a valós hálózatokra jellemző heterogén élsúlyokat indukál, amit a korábbi aktivitás vezérelt hálózati modellek nem tudtak reprodukálni. Összevetve, az így definiált temporális hálózati modell sok szempontból reprodukálni tudta a valós megfigyeléseket. Végül, a Delay-Kendall információterjedési modellt [20] szimulálva valós és modellezett temporális hálózatokon, rámutattam arra, hogy ilyen típusú folyamatok esetén az erős hálózati kapcsolatok inkább akadályozzák az információk terjedését, mivel a sokszor ismételt kölcsönhatások az információ lokális érdektelenségét indukálják így a globális terjedés korai elhalásához vezethetnek.

**5. Tézis:** Temporális hálózatok egy új matematikai reprezentációját definiáltam, ami az időben változó kölcsönhatások hálózatát egy statikus irányított aciklikus gráfként értelmezi [T5]. Ez a leírás előrelépést jelent a temporális hálózatokat jellemző magasabb rendű kauzális mintázatok mint például a hálózatokat felépítő időbeli utak megértésében, mivel ezek szuperpozícióját reprezentálja mint egy statikus gráf. Ezt a reprezentációt felhasználva, kollegáimmal megmutattam [T5][5, 4, 3], hogy időben változó hálózatok irányított perkolációhoz hasonló jelenségként értelmezhetőek, és egy másodrendű strukturális fázisátalakuláson mennek keresztül. Mitöbb, véletlen temporális hálózati modellek egy széles csoportja pontosan leképezhető ebbe az univerzális osztályba.

Egy  $G = (V, E, T)$  temporális hálózatot a  $v \in V$  csomópontok közötti időben változó  $E \subset V \times V \times [0, T]$  kapcsolódási események sorozataként írhatunk le egy adott  $T$  hosszú periódus alatt. Ha feltételezzük, hogy a csomópontok nem kapcsolódhatnak önmagukhoz és egy csomópont egyszerre csak egy másik csomóponttal hathat kölcsön, akkor bevezethetjük a kapcsolódó események fogalmát. Az  $e = (i, j, t)$  és  $e' = (i', j', t')$  eseményeket akkor hívjuk kapcsolódottnak, ha nem egyszerre történnek ( $t \neq t'$ ) és legalább egy csomópont közös bennük (például  $j = i'$ ). Így a  $G$  temporális

hálózat egy  $D = (E, E_D, w)$  magasabb rendű reprezentációját definiálhatjuk, ahol az  $E$  kapcsolódási események a csomópontok és a közöttük lévő  $E_D$  éleket a kapcsolódó kölcsönhatások definiálják. Mitöbb, minden  $e_D \in E_D$  élhez egy  $w = |t' - t|$  súlyt rendelhetünk amit a kapcsolódó események között eltelt idő definiál.

Az így definiált  $D$  hálózat, úgynevezett eseménygráf, a  $G$  temporális hálózat egy magasabbrendű egzakt reprezentációját adja, ami az összes  $G$ -ben lévő kapcsolódó eseményekből felépülő temporális utak szuperpozíciójaként is felfogható. Ezek a temporális utak határozzák meg bármilyen információ terjedésének struktúráját és dinamikáját a temporális hálózaton, így szabva korlátot bármiféle kollektív jelenség kialakulásának. Dinamikus folyamatok egy széles csoportját olyan jelenségként definiálhatjuk, ahol az információ (pl egy vírus fertőzőképessége, utas maximális várakozási ideje egy repülőtéren, vagy egy információ újdonsága) egy idő után elévülhet egy adott csomóponton. Az ilyen típusú folyamatok fejlődését temporális gráfokon egzaktul le lehet írni olyan eseménygráfokkal, amelyekben csak azokat az éleket tartjuk meg, melyek súlya kisebb mint egy  $\delta t$  határérték. Megmutattam, hogy ezt a  $\delta t$  határértéket kontrollparaméterként definiálva a véletlen temporális hálózatokat leíró eseménygráfok széles csoportja strukturális fázisátalakuláson megy keresztül ami az irányított perkoláció univerzalitási osztályba tartozik. Ezt az állításomat az eseménygráfokra definiált perkolációs paraméterek, mint a perkolációs rendparaméter, szuszceptibilitás, térbeli és időbeli korrelációs hossz vizsgálatával igazoltam. Számításaim alapján ezen paraméterek a kritikus perkolációs pont körül olyan exponensekkel skálázódnak melyek értéke pontosan megegyezik az irányított perkolációnak a kritikus dimenzió felett talált átlagtér exponenseivel. Ezt a módszert nagyméretű valós hálózatokra is alkalmaztam, amelyek akár több százmillió eseményt is tartalmaztak. Ezekben a rendszerekben hasonló perkolációs fázisátalakulásokat figyeltem meg amik több karakterisztikus időskálán jelentek meg.

**6. Tézis:** A Skype társadalmi hálózatát és felhasználók termékválasztási dinamikáját vizsgálva valós társadalmi terjedési folyamatokat vizsgáltam az akkor elérhető legnagyobb online kommunikációs hálózaton [T6]. 34 ország összes felhasználójának vizsgálatával arra kerestem a választ, hogy a felhasználók termékválasztását hogyan befolyásolja a hálózati partnereik (barátaik) korábbi termékválasztási eseményei, és hogy ez a folyamat komplex vagy egyszerű terjedési modellekkel írható-e pontosabban.

Megmutattam, hogy a felhasználókat saját és szomszédaik viselkedése alapján csoportosítva, akik a barátaik közül elsőként, látszólag spontán módon választottak egy

terméket állandó rátával jelentek meg a rendszerben. Ellenben, azok akik az ismerőseiket követve "társadalmi nyomás" hatására vásároltak, növekvő rátával csatlakoztak, így kölcsönözve nem-lineáris dinamikát a rendszernek. Megmutattam, hogy az egyén szempontjából ez a külső nyomás lineárisan növekszik a befolyásoló ismerősök számával. Ugyanakkor azt találtam, hogy ez a hatás a termékelhagyás esetén nem releváns, mivel ebben az esetben a spontán és befolyásolt termékelhagyási események állandó rátával fejlődnek az időben.

Ezeket a megfigyeléseket egy lineáris határértékmodellben egyesítettem, ahol hálózatba összekötött entitások egy terjedő állapot adaptálása vagy elhagyása érdekében befolyásolták egymást. A modell analitikai közelítését heterogén átlagtér egyenletekkel végeztem [27], melyek numerikus megoldását nagyméretű számítógépes szimulációkkal validáltam. A modellszámítások pontos középtávú előrejelzéseket adtak különböző termékek választásának dinamikájáról számos országban. Ezen predikciók segítségével rámutattam, hogy a termékválasztási dinamika nagyban függ egy adott ország gazdasági fejlettségi szintjétől, lassúbb eladási dinamikát prognosztizálva kevésbé fejlett országokban.

**7. Tézis:** A konvencionális terjedési modellek alapján [76] a társadalmi befolyás által gerjesztett terjedési folyamatok lavinaszerű fejlődését várnánk, azonban a megfigyelések alapján számos társadalmi terjedési folyamat sokkal lassúbb dinamikát követ. Ennek az ellentmondásnak a feloldására vállalkoztam a teljes Skype online társadalmi hálózatának és az azon terjedő termékválasztási szokások elemzésével és modellezésével [T7].

Kutatásaim során definiáltam egy módszert, aminek segítségével felső becslést tudtam adni arra a határértékre, ami a szomszédok felől érkező befolyás miatt egy egyént az adott viselkedési minta elfogadására készítet. Viselkedési mintaként egy online termék megvásárlását vizsgáltam a teljes, több mint 300 millió embert összekötő Skype hálózaton. Ily módon a befolyásolási küszöbértéknek az eloszlását az irodalomban először figyeltem meg sikeresen, és a korábbi elméleti feltevésekkel ellentétben egy heterogén eloszlást találtam amit egy log-normális eloszlással tudtam illeszteni. A termékvásárlás dinamikáját egyéni szinten vizsgálva arra jutottam, hogy a korábbi megfigyelésekkel megegyezően a spontán adaptációs ráta konstans volt, azonban a Watts-modellel ellentétben nem azok a felhasználók indukálták a globális terjedési folyamatot, akiknek csak egy befolyásoló szomszédra volt szükségük, hanem azok akiknek több szomszéd együttes hatásának következményeként vásárolták meg a terméket. Mitöbb,

a hálózatban kapcsolódó legtöbb felhasználó nem vette meg a terméket annak ellenére, hogy erre lehetőségük lett volna a termék 8 éves története alatt.

Definiáltam egy komplex terjedési modellt [T7, T8] amiben egy hálózaton terjedő küszöbérték-folyamatot feltételeztem, ahol a csomópontok, azon túl, hogy üresek vagy betöltöttek lehetnek, a folyamat elejétől kezdve blokkolt állapotot is felvehettek, vagyis sosem "fertőződhetnek" meg a terjedő állapottal. Ez a blokkolt populáció ritkító hatásként lépett fel a folyamat szempontjából és képes volt nagyban lelassítani a terjedési folyamatot. Ezt a jelenséget analitikusan és numerikusai is szemléltettem közelítő mester-egyenletek [27] és stabilitásanalízis segítségével. Ennek eredményeként az találtam, hogy blokkolt csomópontok arányának növelésével a terjedési folyamat egy átmeneti fázison keresztül halad át a globális és lokális terjedést indukáló fázisok között. Ennek határai a spontán, gyenge és erős befolyást igénylő felhasználók által indukált alhálózatok kritikus perkolációs pontjaihoz köthető. Ezek után a modellt a valós rendszerből kimért paramterek segítségével szimulálva megmutattam, hogy a vizsgált valós terjedési folyamat ebbe az átmeneti fázisba esik, ami megmagyarázza az empirikus folyamat viszonylag lassú terjedését.

## **Társadalmi hasznosítás**

Habár eredményeim nagy része alapkutatásból származik, számos, a társadalom számára hasznos felhasználási lehetőséget kínálnak. Első sorban, a heterogén dinamikus viselkedési minták jobb megértése hatékonyabb telekommunikációs protokollokhoz vezethetnek, valamint a felhasználók igényeit jobban kielégítő kommunikációs szolgáltatások tervezéséhez használhatóak fel. Az terjedési folyamatok sebességét befolyásoló temporális hálózati hatások pontosabb leírása fontos szerepet játszik a különböző járványok terjedési dinamikájának megértésében és modellezésében. Ezen túl olyan kommunikációs stratégiák kidolgozását segíthetik elő, ami vészhelyzetek esetén hatékonyabb információ közléshez vezethetnek, vagy ellentétben ezzel, a hamis információk széleskörű elterjedését lassíthatják a társadalomban. A társadalmi terjedési folyamatok általam fejlesztett modelljei sokféle innováció, mém vagy viselkedési minta társadalmi szintű elterjedését tudja előrejelezni, így elősegítve újabb marketing vagy kontrol stratégiák kidolgozására. Az elmúlt időben számos munkám irányult a COVID-19 pandémia magyarországi alakulásának és annak egyéni viselkedési és társadalmi következményeinek megfigyelésére és modellezésére. Ezen eredményeim közvetlenül szolgáltattak támaszpontot az országos járványügyi védekezési stratégiák kialakításában.



## A tézispontokhoz kapcsolódó saját publikációk

- [T1] M. Karsai, K. Kaski, A.-L. Barabási and J. Kertész, Universal features of correlated bursty behaviour. *Scientific Reports* 2, 397 (2012)
- [T2] M. Karsai, K. Kaski and J. Kertész, Correlated dynamics in egocentric networks. *PLoS ONE* 7(7), e40612 (2012)
- [T3] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási and J. Saramäki, Small But Slow World: How Network Topology and Burstiness Slow Down Spreading. *Phys. Rev. E* 83, 025102(R) (2011)
- [T4] M. Karsai, N. Perra and A. Vespignani, Time varying networks and the weakness of strong ties. *Scientific Reports* 4, 4001 (2014)
- [T5] M. Kivelä, J. Cambe, J. Saramäki, M. Karsai, Mapping temporal-network percolation to weighted, static event graphs. *Scientific Reports* 8, 12357 (2018)
- [T6] M. Karsai, G. Iniguez, K. Kaski, J. Kertész, Complex contagion process in spreading of online innovation. *J. R. Soc. Interface* 11, 101 (2014)
- [T7] M. Karsai, G. Iniguez, R. Kikas, K. Kaski, J. Kertész, Local cascades induced global contagion: How heterogeneous thresholds, exogenous effects, and unconcerned behaviour govern online adoption spreading. *Scientific Reports* 6, 27178 (2016)
- [T8] Z. Ruan, G. Iniguez, M. Karsai, J. Kertész, Kinetics of Social Contagion. *Phys. Rev. Lett.* 115, 218702 (2015)



# Irodalomjegyzék

- [1] Marco Ajelli, Bruno Gonçalves, Duygu Balcan, Vittoria Colizza, Hao Hu, José J Ramasco, Stefano Merler, and Alessandro Vespignani. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC infectious diseases*, 10(1):1–13, 2010.
- [2] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [3] Arash Badie-Modiri, Márton Karsai, and Mikko Kivelä. Efficient limited-time reachability estimation in temporal networks. *Physical Review E*, 101(5):052303, 2020.
- [4] Arash Badie-Modiri, Abbas K Rizi, Márton Karsai, and Mikko Kivelä. Directed percolation in random temporal network models with heterogeneities. *arXiv preprint arXiv:2110.07698*, 2021.
- [5] Arash Badie-Modiri, Abbas K Rizi, Márton Karsai, and Mikko Kivelä. Directed percolation in temporal networks. *arXiv preprint arXiv:2107.01510*, 2021.
- [6] Duygu Balcan, Hao Hu, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al. Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility. *BMC medicine*, 7(1):1–12, 2009.
- [7] Albert-László Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.
- [8] Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.
- [9] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [10] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008.

- [11] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.
- [12] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, 874:1–92, 2020.
- [13] Andrea L Bertozzi, Elisa Franco, George Mohler, Martin B Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of covid-19. *Proceedings of the National Academy of Sciences*, 117(29):16732–16738, 2020.
- [14] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):10, 2015.
- [15] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009.
- [16] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.
- [17] Min Chen, Shiwen Mao, and Yunhao Liu. Big data: A survey. *Mobile Networks and Applications*, 19(2):171–209, 2014.
- [18] Dante R Chialvo. Emergent complex neural dynamics. *Nature physics*, 6(10):744–750, 2010.
- [19] Rosaria Conte, Nigel Gilbert, Giulia Bonelli, Claudio Cioffi-Revilla, Guillaume Defuant, Janos Kertesz, Vittorio Loreto, Suzy Moat, J-P Nadal, Anxo Sanchez, et al. Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1):325–346, 2012.
- [20] Daryl J Daley and David G Kendall. Epidemics and rumours. *Nature*, 204(4963):1118, 1964.
- [21] Sergey N Dorogovtsev and Jose FF Mendes. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002.
- [22] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.
- [23] Ian Foster, Rayid Ghani, Ron S Jarmin, Frauke Kreuter, and Julia Lane. *Big data and social science: A practical guide to methods and tools*. CRC Press, 2016.
- [24] Katerina Fragkiadaki, Sergey Levine, Panna Felsen, and Jitendra Malik. Recurrent network models for human dynamics. In *Proceedings of the IEEE international conference on computer vision*, pages 4346–4354, 2015.

- [25] Laetitia Gauvin, Mathieu Génois, Márton Karsai, Mikko Kivelä, Taro Takaguchi, Eugenio Valdano, and Christian L Vestergaard. Randomized reference models for temporal networks. *arXiv preprint arXiv:1806.04032*, 2018.
- [26] Nigel Gilbert and Klaus Troitzsch. *Simulation for the social scientist*. McGraw-Hill Education (UK), 2005.
- [27] James P Gleeson. High-accuracy approximation of binary-state dynamics on networks. *Physical Review Letters*, 107(6):068701, 2011.
- [28] James P Gleeson. Binary-state dynamics on complex networks: Pair approximation and beyond. *Physical Review X*, 3(2):021004, 2013.
- [29] Marta C Gonzalez, Cesar A Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- [30] Sandra González-Bailón, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. The dynamics of protest recruitment through an online network. *Scientific reports*, 1(1):1–7, 2011.
- [31] Palash Goyal and Emilio Ferrara. Graph embedding techniques. *Applications, and Performance: A Survey. arXiv preprint*, 2017.
- [32] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.
- [33] Jake M Hofman, Duncan J Watts, Susan Athey, Filiz Garip, Thomas L Griffiths, Jon Kleinberg, Helen Margetts, Sendhil Mullainathan, Matthew J Salganik, Simine Vazire, et al. Integrating explanation and prediction in computational social science. *Nature*, 595(7866):181–188, 2021.
- [34] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.
- [35] Gerardo Iñiguez, Federico Battiston, and Márton Karsai. Bridging the gap between graphs and networks. *Communications Physics*, 3(1):1–5, 2020.
- [36] Márton Karsai, Hang-Hyun Jo, and Kimmo Kaski. *Bursty human dynamics*. Springer, 2018.
- [37] Matt J Keeling and Ken TD Eames. Networks and epidemic models. *Journal of the royal society interface*, 2(4):295–307, 2005.
- [38] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

- [39] Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, Jari Saramäki, and Márton Karsai. Multiscale analysis of spreading in a large communication network. *Journal of Statistical Mechanics: Theory and Experiment*, page P03005, 2012.
- [40] Pei-Ying Kobres, Jean-Paul Chretien, Michael A Johansson, Jeffrey J Morgan, Pai-Yei Whung, Harshini Mukundan, Sara Y Del Valle, Brett M Forshey, Talia M Quandelacy, Matthew Biggerstaff, et al. A systematic review and evaluation of zika virus forecasting and prediction research during a public health emergency of international concern. *PLoS neglected tropical diseases*, 13(10):e0007451, 2019.
- [41] Júlia Koltai, Orsolya Vásárhelyi, Gergely Röst, and Márton Karsai. Reconstructing social mixing patterns via weighted contact matrices from online and representative surveys. *Scientific Reports*, 12(1):1–12, 2022.
- [42] Gautier Krings, Márton Karsai, Sebastian Bernhardsson, Vincent D Blondel, and Jari Saramäki. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1):4, 2012.
- [43] Guillaume Laurent, Jari Saramäki, and Márton Karsai. From calls to communities: a model for time-varying social networks. *The European Physical Journal B*, 88(11):301, 2015.
- [44] David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915):721, 2009.
- [45] David MJ Lazer, Alex Pentland, Duncan J Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts, et al. Computational social science: Obstacles and opportunities. *Science*, 369(6507):1060–1062, 2020.
- [46] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [47] Naoki Masuda and Petter Holme. *Temporal network epidemiology*. Springer, 2017.
- [48] Viktor Mayer-Schönberger and Kenneth Cukier. Big data: A revolution that transforms how we work, live, and think, 2012.
- [49] Stefano Merler, Marco Ajelli, Laura Fumanelli, Marcelo FC Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis L Chao, Ira M Longini Jr, M Elizabeth Halloran, and Alessandro Vespignani. Spatiotemporal spread of the 2014 outbreak of ebola virus disease in liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *The Lancet Infectious Diseases*, 15(2):204–211, 2015.

- [50] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3):e74, 2008.
- [51] Mark Newman. *Networks: an introduction*. Oxford university press, 2010.
- [52] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [53] Dong Nguyen, A Seza Doğruöz, Carolyn P Rosé, and Franciska De Jong. Computational sociolinguistics: A survey. *Computational linguistics*, 42(3):537–593, 2016.
- [54] Manoj Parameswaran and Andrew B Whinston. Social computing: An overview. *Communications of the association for Information Systems*, 19(1):37, 2007.
- [55] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.
- [56] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Scientific reports*, 2:469, 2012.
- [57] Kiesha Prem, Alex R Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Computational Biology*, 13(9):e1005697, 2017.
- [58] P Rajendra and V Brahmajirao. Modeling of dynamical systems through deep learning. *Biophysical Reviews*, 12(6):1311–1320, 2020.
- [59] Tiina Roose, S Jonathan Chapman, and Philip K Maini. Mathematical models of avascular tumor growth. *SIAM review*, 49(2):179–208, 2007.
- [60] Csilla Rudas, Olivér Surányi, Taha Yasseri, and János Török. Understanding and coping with extremism in an online collaborative environment: A data-driven modeling. *PloS one*, 12(3):e0173561, 2017.
- [61] Shih-Lung Shaw, Ming-Hsiang Tsou, and Xinyue Ye. Human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, 30(9):1687–1693, 2016.
- [62] A. Skalka. *Understanding Human Dynamics: Report of the Defense Science Board Task Force*. DIANE Publishing Company, 2009.
- [63] Neil J Smelser. *Theory of collective behavior*. Quid Pro Books, 2011.

- [64] H. Eugene Stanley. *Introduction to Phase Transitions and Critical Phenomena (International Series of Monographs on Physics)*. Oxford University Press, USA, July 1987.
- [65] Leigh Tesfatsion and Kenneth L Judd. *Handbook of computational economics: agent-based computational economics*. Elsevier, 2006.
- [66] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, José J Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm. *BMC medicine*, 10(1):1–31, 2012.
- [67] Mario V Tomasello, Nicola Perra, Claudio J Tessone, Márton Karsai, and Frank Schweitzer. The role of endogenous and exogenous mechanisms in the formation of r&d networks. *Scientific reports*, 4:5679, 2014.
- [68] Enrico Ubaldi, Nicola Perra, Márton Karsai, Alessandro Vezzani, Raffaella Burioni, and Alessandro Vespignani. Asymptotic theory of time-varying social networks with heterogeneous activity and tie allocation. *Scientific reports*, 6:35724, 2016.
- [69] Enrico Ubaldi, Alessandro Vezzani, Márton Karsai, Nicola Perra, and Raffaella Burioni. Burstiness and tie activation strategies in time-varying social networks. *Scientific reports*, 7:46225, 2017.
- [70] Samuel Unicomb, Gerardo Iñiguez, James P Gleeson, and Márton Karsai. Dynamics of cascades on burstiness-controlled temporal networks. *Nature communications*, 12(1):1–10, 2021.
- [71] Samuel Unicomb, Gerardo Iñiguez, and Márton Karsai. Threshold driven contagion on weighted networks. *Scientific reports*, 8(1):3094, 2018.
- [72] Samuel Unicomb, Gerardo Iñiguez, János Kertész, and Márton Karsai. Reentrant phase transitions in threshold driven contagion on multiplex networks. *Physical Review E*, 100(4):040301, 2019.
- [73] Alessandro Vespignani. Predicting the behavior of techno-social systems. *Science*, 325(5939):425–428, 2009.
- [74] Alessandro Vespignani, Huaiyu Tian, Christopher Dye, James O Lloyd-Smith, Rosalind M Eggo, Munik Shrestha, Samuel V Scarpino, Bernardo Gutierrez, Moritz UG Kraemer, Joseph Wu, et al. Modelling covid-19. *Nature Reviews Physics*, 2(6):279–281, 2020.
- [75] Tamás Vicsek and Anna Zafeiris. Collective motion. *Physics reports*, 517(3-4):71–140, 2012.



- [76] Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.
- [77] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.
- [78] Yuexin Wu, Yiming Yang, Hiroshi Nishiura, and Masaya Saitoh. Deep learning for epidemiological predictions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1085–1088, 2018.
- [79] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE transactions on pattern analysis and machine intelligence*, 29(1):40–51, 2007.
- [80] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, et al. Changes in contact patterns shape the dynamics of the covid-19 outbreak in china. *Science*, 368(6498):1481–1486, 2020.
- [81] Kun Zhao, Márton Karsai, and Ginestra Bianconi. Entropy of dynamical social networks. *PloS one*, 6(12):e28116, 2011.