

A hálózatelemzés alkalmazásai a meteorológiában

SZAKDOLGOZAT

FÖLDTUDOMÁNYI ALAPSZAK

METEOROLÓGUS SZAKIRÁNY



Készítette:

Kovács Ádám

Témavezető:

Pieczka Ildikó

Eötvös Loránd Tudományegyetem

Földrajz- és Földtudományi Intézet

Meteorológiai Tanszék

Budapest, 2016

Tartalomjegyzék

1. Bevezetés.....	3
2. A hálózatelmélet története	4
3. A hálózatelmélet matematikai alapjai.....	6
4. Hálózatelmélet és meteorológia	9
5. Hogyan lesz hálózatunk meteorológiai adatokból?	10
6. Klímahálózatok és a mintavételezés térbeli eloszlásának problémája	13
7. Esettanulmányok.....	19
I.: Domináns transzport útvonalak egy légköri blocking helyzetben	19
II.: Klímahálózatok fejlődésének jellemzése.....	26
8. Jövőbeli kilátások.....	31
Köszönetnyilvánítás.....	33
Irodalomjegyzék	34

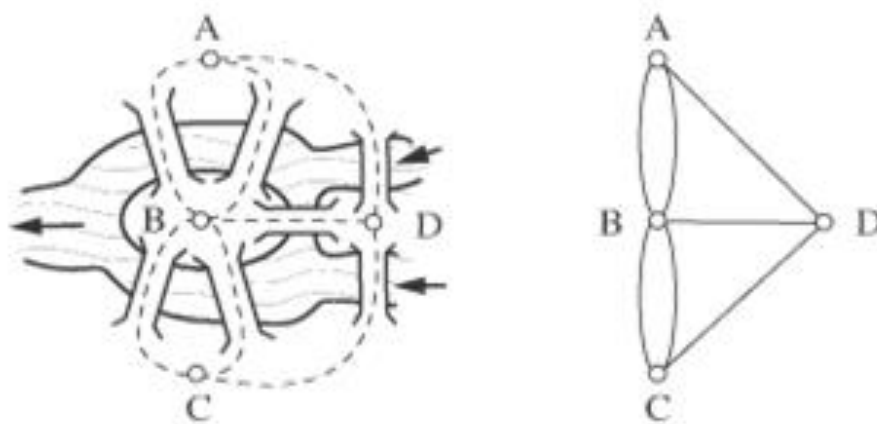
1. Bevezetés

A hálózatemélet a meteorológia egyik legfrissebb ága, az ide kapcsolódó kutatások az elmúlt tíz évben kezdődtek meg. Az ötlet onnan származik, hogy az elmúlt évtizedekben rengeteg komplex rendszerekkel foglalkozó tudományág sikerrel alkalmazta a hálózatokat jelenségeik leírására, például a biológia a sejtek együttműködésére, a fehérjeszintézis modellezésére, a szociológia az emberi kapcsolatok szerveződésére, vagy a számítástechnika az internet topológiájának vizsgálatára. A meteorológiával és klimatológiával kapcsolatos hálózateméleti kutatások célja is hasonló: a rendelkezésre álló hatalmas adatmennyiségből olyan fontos kapcsolatokat szeretnénk kinyerni, amelyek rávilágítanak a légkört kormányzó fizikai, dinamikai mechanizmusokra.

A következő fejezetben áttekintjük a hálózatemélet eddigi történetének legfontosabb állomásait és magyar vonatkozásait. A 3. fejezetben megismerkedünk a hálózatemélet matematikai alapjaival (gráfelmélet), legfontosabb fogalmaival, valamint a hálózatok szemléltetésének lehetőségeivel. Ezt követően részletesebben lesz szó a meteorológia és a hálózatemélet kapcsolatáról és az eddigi főbb kutatási irányokról (4. fejezet). Az 5. fejezet arról szól, hogy milyen lehetőségeink vannak hálózatok konstruálására a meteorológiai mérési adatokból, ennek kapcsán megismerkedünk az eddigi kutatásokban felhasznált legfontosabb statisztikai, információelméleti fogalmakkal és módszerekkel. Ezután a hálózatok térbeli függésének problematikáját járjuk körbe, azaz megvizsgáljuk a mérések térbeli eloszlásának a hálózatok topológiájára gyakorolt hatását (6. fejezet). A hálózateméleti módszerek konkrét alkalmazásának példáit a 7. fejezetben esettanulmányok segítségével tekintjük át, végül pedig a jövőbeli lehetséges kutatási irányok, illetve a hálózateméletben rejlő további lehetőségek kerülnek összefoglalásra (8. fejezet).

2. A hálózatelmélet története

Az első, hálózatokhoz köthető matematikai probléma az azóta elhíresült königsbergi hidakhoz, annak megoldása pedig az 1700-as évek egyik, ha nem a legnagyobb matematikusához, Leonhard Eulerhez köthető. Königsberg a tizennyolcadik századi Poroszország fejlődő városa volt, a Pregel folyó mentén helyezkedett el. A város vezetői hét hidat építettek a folyón keresztül, amelyek a város különböző pontjait kötötték össze egymással, illetve a folyóba ékelődő Kneiphof szigettel (1. ábra).



1. ábra: A königsbergi hidak és a probléma gráfja (Makó és Téglási (2012)).

A híres feladat a következő volt: át lehet-e sétálni a hét hídon úgy, hogy mindegyiken csak egyszer haladunk át. Euler 1736-ban egzakt matematikai bizonyítást adott arra, hogy ezen elrendezésben ilyen útvonal nem létezik. Számunkra azonban nem a bizonyítás a lényeges, hanem Euler egy apró ötlete, amely új utat nyitott a matematikában: létrehozta a gráfelméletet azzal, hogy a négy földterület mindegyikét egy-egy ponttal (1. ábra, A-tól D-ig), a hidakat pedig éllekké helyettesítette, így egy négy pontból és hét élből álló gráfot kapott.

A gráfelmélet Euler után olyan nagy matematikusoknak köszönhetően lendült fel, mint például Cauchy, Hamilton, Cayley, Kirchoff, vagy Pólya. Szinte nekik köszönhetünk mindent, amit manapság a gráf-, illetve hálózatelméletről tudunk. Az 1950-es évekig a matematika ezen ágának célja egyszerű volt: megismerni és rendszerezni a különböző szerkezetű gráfok tulajdonságait. Szintén a híres problémák közé tartozott, hogy megtalálják egy labirintusból a kiutat (1873-ból ismert az első megoldása), vagy a sakktáblán egy olyan lépéssorozat, azaz útvonal meglelése, amelyen a huszár mind a hatvannégy mezőt egyszer érinti, majd visszaér a kiindulási pozíciójába (Barabási (2013)).

A huszadik század második felében két magyar matematikus, Erdős Pál és Rényi Alfréd forradalmasította a gráfokról való gondolkodásmódunkat azzal, hogy a gráfok tulajdonságai helyett a gráfok, illetve a valódi hálózatok kialakulásával és fejlődésével kezdtek el foglalkozni. Ők ketten fektették le az úgynevezett *véletlen hálózatok* elméletének matematikai alapjait. A véletlen hálózatokban (gráfokban) a kapcsolatokat (éleket) véletlenszerűen, például kockadobás eredményével alakították ki, majd az így kapott hálózatok topológiáját vizsgálva rengeteg érdekes tételt és bizonyítást alkottak, például hogy a véletlen hálózatok fokszámainak eloszlása Poisson-eloszlást követ. Ez azt jelenti, hogy a legtöbb pont átlagos számú kapcsolattal rendelkezik, és kevés az extrém magas, vagy extrém alacsony kapcsolattal rendelkező pont. Erdős és Rényi felfogásának a hálózatokról akkora hatása volt, hogy a témában kutatók a mai napig nehezen tudnak elszakadni az ő alapvető elképzeléseiktől (*Barabási (2013)*).

A valóságban előforduló hálózatok azonban nem véletlenszerűen fejlődnek, gondoljunk csak a szociális hálózatunk alakulására, egy barátunk ismerősével sokkal nagyobb valószínűséggel kerülünk kapcsolatba, mint egy busman törzsfőnök unokájával. Ez a felismerés vezetett az úgynevezett *skálafüggetlen hálózatok* felfedezéséhez (1980-as évek vége), amelyek fokszámeloszlása hatványfüggvényeket követ (sok pont rendelkezik kis számú kapcsolattal és kevés pontnak van extrém sok linkje). A skálafüggetlen hálózatok magyarázatot adtak a természetben előforduló hálózatokban megfigyelhető csomópontok kialakulására (ezek a rengeteg kapcsolattal rendelkező pontok), valamint lehetőséget adtak nagyon sok tudományág számára a hálózatok modellezésére. Ez a forradalmi újítás az ezredforduló környékére lehetővé tette az internet topológiájának vizsgálatát, a gazdasági és szociális kapcsolatok újabb szemléletű analízisét, részecskefizikában meglepő analógiára derült fény a hálózatok szerkezete és a Bose-Einstein-kondenzátum tulajdonságai között, a biológia hálózatok segítségével sikeresen modellezte a sejtek fehérjeszintézisét, és ezt a sort a végtelenségig folytathatnánk (*Barabási (2013)*). Az elmúlt tíz évben a meteorológiai és klimatológiai kutatások is megkezdődtek a hálózatelmélet eszköztárának alkalmazásával. Erről a későbbi fejezetekben részletesebben lesz szó, elsőként azonban ismerkedjünk meg a hálózatelmélet matematikai alapjaival, a gráfelmélettel.

3. A hálózatelemélet matematikai alapjai

A hálózatok nagyon lecsupaszítva pontokból, valamint az azokat összekötő vonalakból (kapcsolatokból) állnak. Gondoljunk csak az elektromos hálózatokra, ahol kapcsolási rajzokon az áramforrásokat és fogyasztókat (pontok) vezetékek (vonalak) kötik össze. Ezen geometriai megfontolás vezet ahhoz az elgondoláshoz, hogy a hálózatokat legegyszerűbben gráfokkal reprezentálhatjuk.

Definíció: Egy $G = (V, E)$ párt *egyszerű gráfnak* nevezünk, ha V egy véges halmaz, és $E \subset \binom{V}{2}^1$ (Hajnal (1997)). V elemeit a G gráf *csúcsainak* vagy *pontjainak* nevezzük. E elemei a G gráf *élei*.

Az $e = \{u, v\} \in E(G)$ relációt úgy értelmezhetjük, hogy u és v az e él végpontjai, azaz e él összeköti u és v pontokat. Ha $e = \{u, u\}$, akkor e -t hurokélnek nevezzük. Az $E(G)$ zárójeles jelölést akkor használjuk, ha hangsúlyozni szeretnénk, hogy a G gráfról van szó.

A gráfok általánosan megfogalmazott definíciója lehetővé teszi, hogy a csúcsoknak és az éleknek bármilyen szemléletes jelentést adhassunk, így rengeteg tudományág számára eszközt biztosít. Például közlekedéstudományban a csúcsok városokat, az élek pedig ezen városokat összekötő útvonalakat jelölhetnek, szociális kutatások esetén a gráf csúcsai személyeket, az élek ismeretségeket reprezentálhatnak, míg meteorológiában a csúcsok mérőállomásokat jelenthetnek, az élek pedig az egymás adatsoraival erősen korreláló pontokat köthetik össze.

Egy adott gráf definiálásához tehát nincs más dolgunk, mint megadni a $V(G)$, valamint $E(G)$ halmazokat, ez azonban nagy gráfok esetén körülményes, nehezen áttekinthető. Szintén egy lehetőség a gráfok geometriai ábrázolása: minden csúcsnak megfeleltetünk a síkon egy kis karikát, valamint minden élnek egy egyszerű görbét, amely a megfelelő két csúcsot reprezentáló karikát köti össze, és másik karikán nem halad át (2. ábra). Ezen geometriai interpretáció ugyan szemléletes, ám nagy gráfok esetén az átláthatósága szintén nehézkes. A gráfok ábrázolásának másik lehetséges módja a mátrixok használata.

¹ $\binom{V}{2}$ a V halmaz kételemű részhalmazait jelöli.

Definíció: Legyen I_G egy $|V(G)|$ sorral és $|E(G)|$ oszloppal rendelkező mátrix. A sorok a pontoknak, az oszlopok az éleknek felelnek meg. Egy v pont sorának és egy e él oszlopának kereszteződésében álló $I_G(v, e)$ elem:

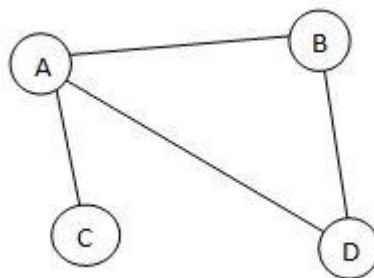
$$I_G(v, e) = \begin{cases} 1, & \text{ha } v \in e \\ 0, & \text{egyébként} \end{cases}$$

Ezt a mátrixot a gráf *pont-él illeszkedési mátrixának* nevezzük. Ekkor I_G oszlopaiban két darab 1-es kivételével 0-k állnak (Hajnal (1997)).

Definíció: Az A_G mátrix négyzetes mátrix, sorainak és oszlopainak száma $|V(G)|$. A sorok és az oszlopok is a pontokkal vannak azonosítva. Egy (u, v) pontpárnak megfelelő helyen

$$A_G(u, v) = \begin{cases} 1, & \text{ha } \{u, v\} \in E(G) \\ 0 & \text{egyébként} \end{cases}$$

áll. Ezt a mátrixot a G egyszerű gráf *szomszédsági mátrixának* nevezzük. Ekkor A_G szimmetrikus mátrix, elemei nullák és egyesek, főátlójában 0-k szerepelnek (Hajnal (1997)).



2. ábra: G gráf szemléltetése. $V(G) = \{A, B, C, D\}$, $E(G) = \{AB, AC, AD, BD\}$ (Makó és Téglási (2012)).

A 2. ábrán látható G gráf szomszédsági mátrixa:
$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

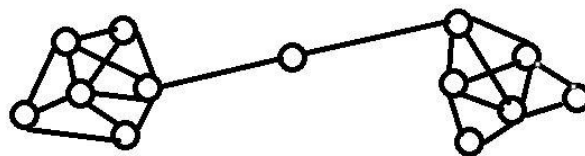
A gráf fogalmának megismerése után tekintsük át a hálózatokat, valamint azok pontjait jellemző legfontosabb mérőszámokat.

Definíció: Legyen $v \in V(G)$. A v csúcs $d_G(v)$ fokszáma vagy foka a G gráfban $|\{e \in E(G) : v \in e, \text{ és } e \text{ nem hurokél}\}| + 2|\{e \in E(G) : v \in e, \text{ és } e \text{ hurokél}\}|$ (Hajnal (1997)).

A fokszám tehát egyszerű gráfok esetén a csúcsra illeszkedő élek számát adja meg, a hurokélek pedig 2-vel növelik egy pont fokát (ahogy az a geometriai megfontolásból is látszik). Ha az egész gráfot szeretnénk jellemezni csúcsainak fokszáma alapján, azokból hisztogramot készítve megkaphatjuk a *fokszámeloszlást*, vagy a fokszámokat nagyság szerint sorba rendezve megkaphatjuk az úgynevezett *fokszámsorozatot*.

Definíció: A $g(v) = \sum_{s \neq v \neq t} A/B$ számot a v csúcs *közöttisége*nek nevezzük, ahol B az s és t csúcsok közötti legrövidebb utak száma, A pedig azok száma, amelyek ezek közül áthaladnak v -n. A definícióból egyértelműen következik, hogy $g(v)$ értéke 0 és 1 közé esik (Freeman (1977)).

Egy pont közöttisége azt mutatja meg, hogy ennek a csúcsnak mennyire van központi szerepe a gráfban. Minél nagyobb egy pont közöttisége, annál nagyobb hatása van a hálózaton belüli információáramlásban, azzal a feltételezéssel élve, hogy az információ a legrövidebb úton terjed tovább. Fontos látnunk, hogy egy csúcs fokszáma és közöttisége között nincs összefüggés, attól, hogy egy pontnak nagy a fokszáma, még nem feltétlenül nagy a közöttisége, és fordítva. Például a 3. ábrán bemutatott két gócot egymással összekapcsoló pont fokszáma 2, ami kicsi a gráf többi csúcsáéhoz képest, közöttisége viszont nagy, hiszen a két csoport közötti út csak ezen a csúcson keresztül lehetséges.



3. ábra: Fokszám és közöttiség.

Definíció: Legyen $x, y \in V(G)$. x és y *távolsága* ($d_G(x,y)$) az x és y csúcsokat összekötő utak közül a legrövidebb hossza. Speciálisan $d(x,x) = 0$, valamint ha nincs út x és y csúcsok között, akkor $d(x,y) = \infty$ (Hajnal (1997)).

A távolság definíciója elég szemléletes, azt mutatja meg, hogy egy gráf két csúcsa között hány ponton át vezet a legrövidebb út. Fontos megjegyeznünk, hogy egy hálózat két pontjának „térbeli” közelsége nem jelenti azt, hogy a gráfelméleti értelemben vett távolságuk is kicsi lenne: ennek legegyszerűbb ellenpéldája, ha a két térben legközelebbi pont nincsen összekötve.

4. Hálózatelmélet és meteorológia

A hálózatokat az egymással dinamikus kölcsönhatásban álló részekből felépülő komplex rendszerek egy praktikus reprezentációjának tekinthetjük. A huszadik század második felétől rengeteg tudományág sikeresen alkalmazta a vizsgálati körökben előforduló jelenségek leírására, modellezésére a komplex hálózatokat, például a szociológia, a számítástechnika, a mérnöki- vagy a természettudományok is. Az elmúlt évtizedben a meteorológiával és klimatológiával foglalkozó kutatók is felfedezték a hálózatelméletben rejlő lehetőségeket a légköri és óceáni mozgásrendszerek vizsgálatára. Az alapvető elgondolás az, hogy az atmoszférát (óceánokat) leíró állapotváltozók, paraméterek (például hőmérséklet, légnyomás, csapadékmennyiség) karakterisztikáját összhangba hozzuk a komplex hálózatok topológiájával. Ahogy azt majd későbbi fejezetekben látjuk, statisztikai és információelméleti eljárások segítségével kapcsolatokat lehet definiálni a rendszer pontjai között az adott földrajzi helyek idősorainak felhasználásával. Ezen kapcsolatok lesznek a hálózatunk (gráfunk) élei, míg az említett pontok a hálózat csúcsai. Az eredményül kapott hálózatot gráf- és hálózatelméleti vizsgálatnak alávetve pontosabb képet kaphatunk az éghajlati rendszerben zajló folyamatokról.

A hálózatelmélet egy viszonylag fiatal tudományág, a meteorológiában mindössze tízéves „múlta” tekint vissza, így egyelőre nincsenek meg a kiforrott, jól működő eljárások, az eddigi kutatások eredményei azonban bizakodásra adnak okot. A hálózatelmélet új területeket nyitott a Föld klímarendszerének vizsgálatában, segítségével légköri távkapcsolatokat detektáltak, vizsgálták többek között az ENSO (El Niño Southern Oscillation) és az indiai monszun jelenségkörét, valamint bolygónk éghajlati rendszerének stabilitását is. Az eddigi kutatások három nagy témakör köré csoportosultak: készítettek folyadékdinamikai modelleket hálózatok segítségével, vizsgáltak nagyobb skálájú légköri mozgásrendszereket (ENSO, monszun), valamint a légköri/óceáni távkapcsolatokat (*Dong Zhou et al. (2015)*). Ez utóbbi két kategória erősen összefügg, például több kutatás is megjelent arról, hogy az ENSO és a monszun jelenség hogyan befolyásolja egymást. A folyadékdinamikai kutatások további három nagyobb csoportra oszthatók: az elsőben geometriai objektumokat (vonalak, felszínek) kerestek, amelyek a különböző tulajdonságú folyadékrészeket kapcsolják össze, egy másik csoportba sorolhatók azok, amelyek a rendszer dinamikáját leíró mérőszámokat (például Ljapunov-exponenst) számoltak. Ezekon kívül készültek kutatások a véges méretű folyadékrészek mozgásának hálózatelméleti leírásáról is (*Ser-Giacomi et al. (2015)*).

5. Hogyan lesz hálózatunk meteorológiai adatokból?

Ahogy azt már a korábbi fejezetekben láttuk, egy hálózat definiálásához kettő dolgot kell megadnunk: a hálózatot reprezentáló gráf csúcsait, valamint éleit. Folytonos közegekben, mint például a légkör, vagy az óceánok, azonban nem magától értetődő, hogy hogyan határozzuk meg a rendszert legpontosabban leíró diszkrét gráf-pontokat, és szintén nehézkes a folyamatokat jellemző kapcsolatokat (a gráf éleit) definiálnunk. Ezen probléma megoldására az eddigi kutatásokban leginkább az úgynevezett „kölcsonhatás-hálózatokat” használták, amelyek a következőképpen épülnek fel: a légköri (vagy óceáni) mérési helyek, vagy rácsponti adatbázisok pontjai felelnek meg a gráf csúcsainak, a linkeket két pont között pedig valamilyen hasonlóságot kifejező statisztikai mutató (például korrelációs együttható) segítségével definiáljuk, amelyet a két vizsgált mérőállomás idősorából számítunk ki. A következőkben áttekintjük a legfontosabb statisztikai és információelméleti mérőszámokat, amelyeket meteorológiai hálózatok konstruálásához használhatunk (Deza és Ihsaish (2015)).

Az eddigi kutatásokban legtöbbet alkalmazott hasonlósági mérőszám a keresztkorreláció volt. A korrelációs együttható a két adatsor közötti lineáris kapcsolatot méri, értéke -1 és 1 közé esik, az alábbi képlettel definiálható:

$$E[(X - m_x)(Y - m_y)] = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_x) * (y_i - m_y) \quad (1)$$

ahol az x_i , valamint y_i mintához tartozó X és Y valószínűségi változók m_x és m_y várható értékkel, E a várható értéket jelöli.

Az idősorok karakterisztikáját jellemző mutató az úgynevezett *Shannon-entrópia* (H), amely a p_i valószínűségi sűrűségfüggvényből számított érték:

$$H = - \sum_{i \in A} p_i * \log p_i , \quad (2)$$

ahol A teljes eseményrendszer. A Shannon-féle entrópia egy 1 szabadsági fokkal rendelkező mérőszám, amely egy adott ponton áthaladó információmennyiséget írja le (az információ nevű elvont fogalmat fizikai, számszerűsíthető mennyiséggé alakítja). Bár önmagában nem határozható meg vele egy hálózat, hiszen nem két idősor, hanem egy függvénye, ám segítségével többet tudhatunk meg az információáramlás jellegéről.

Az úgynevezett *közös információ* (Mutual Information, MI) két pont idősorát jellemző p_i és p_j sűrűségfüggvényekből, valamint a p_{ij} közös sűrűségfüggvényből származtatott mennyiség:

$$M_{ij} = \sum_{m,n} p_{ij}(m,n) * \log \frac{p_{ij}(m,n)}{p_i(m)*p_j(n)} \quad (3)$$

M_{ij} szimmetrikus ($M_{ij} = M_{ji}$), és az $i(t)$ és $j(t)$ idősorok statisztikai értelemben vett egymástól való függését jellemzi. Ha a két idősor független egymástól, akkor $M_{ij} = 0$, hiszen ebben az esetben a közös sűrűségfüggvény az individuális sűrűségfüggvények szorzatával egyenlő.

Az *irányítottsági index* (Directionality Index, DI) az alábbi képlettel definiált mennyiség:

$$DI_{XY}(\tau) = \frac{I_{XY}(\tau) - I_{YX}(\tau)}{I_{XY}(\tau) + I_{YX}(\tau)}, \quad (4)$$

ahol $I_{XY}(\tau)$ és $I_{YX}(\tau)$ a *feltételes közös információ*:

$$I_{XY}(\tau) = I(X;Y|X_\tau) = H(X|X_\tau) + H(Y|X_\tau) - H(X,Y|X_\tau) \quad (5)$$

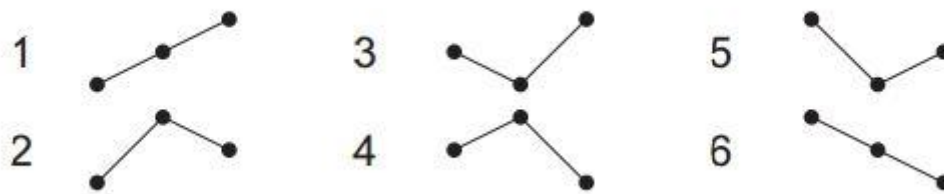
$$I_{YX}(\tau) = I(Y;X|Y_\tau) = H(Y|Y_\tau) + H(X|Y_\tau) - H(Y,X|Y_\tau), \quad (6)$$

$X_\tau = X(t-\tau)$, $Y_\tau = Y(t-\tau)$, valamint $H(X/Y)$ a feltételes entrópia.

Az irányítottsági index az egész hálózatban történő információáramlást számszerűsíti. Az (5) egyenletből következik, hogy $DI_{XY} = -DI_{YX}$. A DI értéke -1 és 1 között változhat: $DI_{XY} = 1$ akkor és csak akkor, ha $I_{XY} \neq 0$ és $I_{YX} = 0$, azaz az információ-áramlás iránya X pontból Y felé mutat és nincsen visszacsatolás Y-től X felé. $DI_{XY} = -1$ pedig csak úgy lehetséges, ha $I_{XY} = 0$ és $I_{YX} \neq 0$, azaz az információ $Y \rightarrow X$ irányban halad, és nincsen fordított irányú visszacsatolás. A $\tau > 0$ paramétert az adatsor időlépcsőinek megfelelően kell beállítani.

A fent említett információelméleti mennyiségek (Shannon-entrópia, közös információ, irányítottsági index) kiszámításához szükség van az adott valószínűségi változó sűrűségfüggvényére. Ez többféle módon is meghatározható, a leggyakrabban alkalmazott módszer, hogy a mintából hisztogramot, vagy relatív gyakorisági függvényt készítünk, az erre illesztett függvény pedig megfelelően nagy elemszám esetén jól közelíti az adott eloszlásból származó minta sűrűségfüggvényét. Egy másik lehetséges eljárás a sorrendi mintázatok (*Ordinal Patterns*, OP) használata. Ennek lényege, hogy az adatsorunk értékeit egymáshoz viszonyítva ábrázoljuk, például 3 értéket vizsgálva $3! = 6$ különböző mintázatot kaphatunk (4. ábra). Ha egy érték ($x_i(2)$) nagyobb, mint az előző ($x_i(1)$), de kisebb, mint az azt követő ($x_i(3)$), akkor a mintázatunk '123' lesz, míg fordított esetben ($x_i(1) > x_i(2) > x_i(3)$) '321'. Általános

esetben, amennyiben nem engedünk meg két egymás utáni egyenlő értéket, a mintázatok számának növekedését a minta elemszámának (n) faktoriálisa határozza meg. Ezen egyszerűsítéssel az idősort terhelő zaj miatt nem követünk el nagy hibát. Az OP használatával az adatsorok értékeinek sorrendje közötti korrelációt is vizsgálhatjuk, amit a hisztogram készítésénél nem vettünk számításba, hiszen ott az adatok sorrendje nem számított, csak az értékek nagysága. Ezzel szemben az OP-módszer hátránya, hogy semmilyen információt nem nyújt az értékek nagyságáról, abszolút magnitúdójáról.



4. ábra: Sorrendi mintázatok (Deza és Ihshaish (2015))

Az OP idősorok elkészítése után hisztogramot számíthatunk belőlük. A hisztogram „dobozainak” számát a mintázatok hossza szabja meg: n elemű minta esetén $n!$ különböző OP-t kaphatunk, és szintén $n!$ részre kell osztanunk a hisztogramot is. Ezzel a technikával elkerülhetőek a hisztogramkészítésnél gyakorta előforduló intervallummeghatározási problémák.

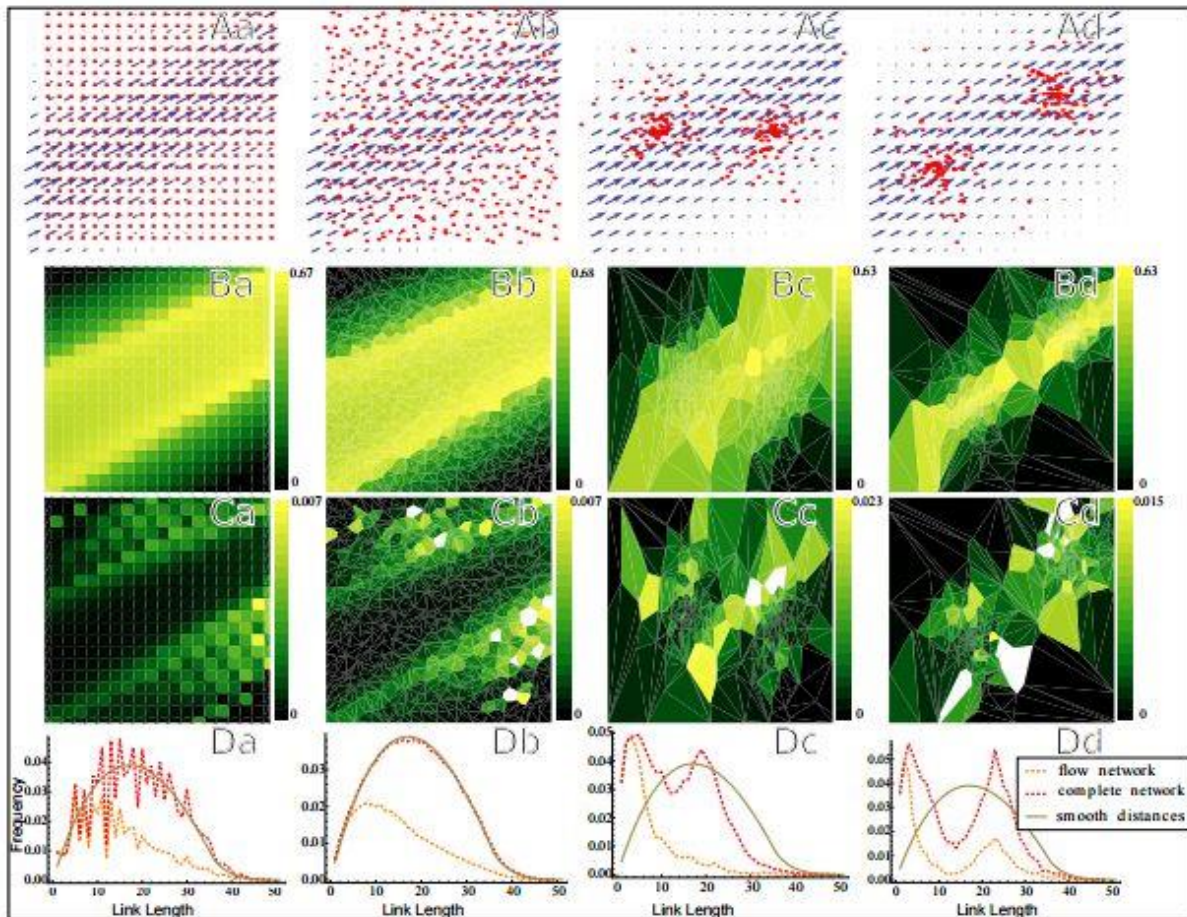
Fontos még megjegyezni, hogy a szokásos hisztogramokból, valamint az OP-s hisztogramokból számított mérőszámok nagyban eltérhetnek egymástól, hiszen ahogy az korábban is említésre került, az OP nem tartalmaz információt az adatok abszolút nagyságáról, csak az értékek egymásutániságáról.

A választott statisztikai, illetve információelméleti mérőszámok kiszámítása után nincs más dolgunk, mint meghatározni egy, a kapott értékekhez igazított küszöbszámot. Ha a két pontot jellemző érték (például MI) nagyobb, mint ez a küszöbszám, akkor ez a két csúcsa a gráfnak kapcsolatban áll egymással, össze van kötve. Szintén van lehetőség a kapcsolatok erősségének súlyozására a mérőszámok függvényében. Ez pontosíthatja a klimatológiai hálózatok gráfelméleti modellezését, azonban nagyobb számításigénnyel jár.

6. Klímahálózatok és a mintavételezés térbeli eloszlásának problémája

Az előző fejezetben az éghajlati hálózatok kapcsolatainak témakörét jártuk végig, most pedig a hálózatok másik alkotóelemének, a csúcsok térbeli eloszlásának kérdését vizsgáljuk meg. Ahogy azt korábban láttuk, folytonos közegek (mint például a légkör) hálózati modellezésénél a közeg térbeli leírását néhány diszkrét pontra korlátozzuk, ezek között valamilyen hasonlóságot kifejező mérőszámot (például korrelációt, MI-t) kiszámítva pedig a kapcsolatokat is meghatározzuk. Az így kapott hálózat struktúráját gyakran magyarázza a pontok földrajzi elhelyezkedése. Az egymáshoz térben közel lévő csúcsok között általában nagyobb korreláció adódik, orografikus akadályok, például nagyobb hegyek azonban korlátozhatják a hálózaton belüli információáramlást. Állandó szelek (pl. passzát), vagy óceáni áramlások pedig térben nagyobb kiterjedésű kapcsolatokat hozhatnak létre (*Molkenthin et al.* (2014)). Mindezek alapján joggal lehet feltételezni, hogy a pontok (mérőállomások) térbeli eloszlása hatással van az éghajlati hálózatok szerkezetére és viselkedésére.

Elsőként tekintsük át a térbeli hatások szerepét az úgynevezett áramlási hálózatokban. Áramlási hálózatokat az advekción-diffúziós egyenlet analitikus megoldásával lehet előállítani, a korábban bemutatott korrelációs módszer helyett (*Molkenthin et al.* (2013)). A hálózaton végzett vizsgálatot az 5. ábra mutatja.



5. ábra: A diagonális áramlás négy különböző rácshálózaton: Aa) szabályos rács, Ab) perturbált rács, Ac) két csomópont az áramlás szélén, Ad) két csomópont az áramlás közepén. B-sor: foks szám, C-sor: közöttiség, D-sor: kapcsolatok hosszának eloszlása (Molkenthin et al. (2014))

A diagonális áramlás (az 5. ábra képeinek bal alsó sarkából a jobb felső irányába tart) hálózatát négy különböző rácson hasonlították össze: egy szabályos 20x20-as, egy perturbált 20x20-as, valamint kettő, egyenként két csomópontot tartalmazó hálón.

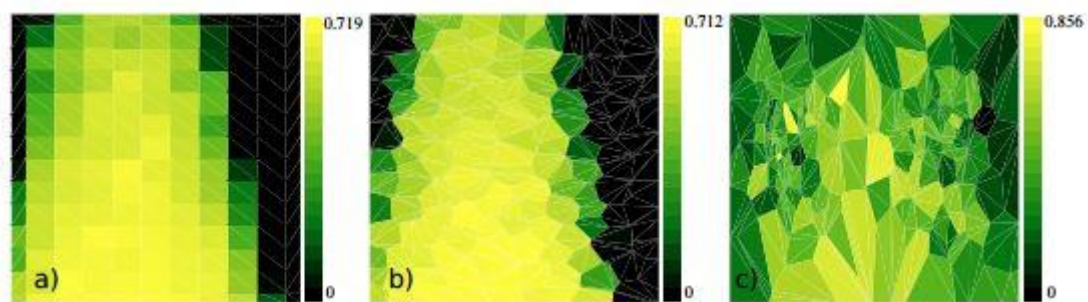
A foks számok (5. ábra, B-sor) mind a négy esetben ott a legmagasabbak, ahol az áramlási sebesség a legnagyobb. Megállapítható, hogy a foks szám nem érzékeny a mintavételezés térbeli eloszlására, bár a csomópontok esetén torzulhat a kép a kevésbé reprezentált területeken.

A szabályos és a perturbált rács esetén is hasonló szerkezetet mutat a csúcsok közöttisége (5. ábra, C-sor): a legmagasabb közöttiség értékkel rendelkező pontok a nagy és kis áramlási sebességű területeket összekapcsoló átmeneti zónában találhatóak. Ez a struktúra egyik csomópontokkal rendelkező hálózat esetén sem figyelhető meg, itt a legmagasabb értékek az

áramlás központjában vannak. Ez azzal állhat összefüggésben, hogy a c) és d) esetén is van olyan része az átmeneti zónának, ahol csak csekély mértékű a mintavételezés.

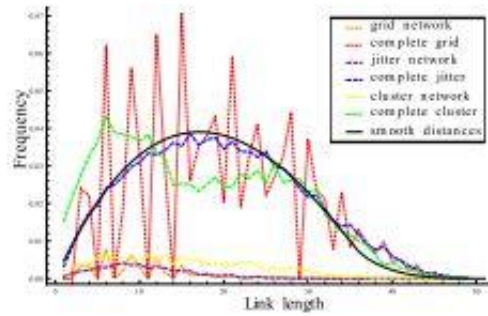
A kapcsolatok hosszának eloszlását a D-sor mutatja (5. ábra). Szabályos rács esetén az eloszlás „tüskés”, a perturbált rács esetén azonban jóval simább, a csomópontos rácsok esetén pedig két maximum figyelhető meg. A „tüskés” eloszlást a szabályos rács esetén bizonyos távolságok, például a rácscsillapok többszöröseinek a túlréprezentáltsága okozza, ezt értelemszerűen kiküszöböli a perturbáció, így ott simább eloszlás számítható. A c), valamint d) esetben lévő két maximum a csomópontokon belüli távolságoknak és a két csomópontot összekötő kapcsolatok hosszának köszönhetőek.

Ezen négy rács típusra szintén végeztek összehasonlító elemzést a START-modell (Stream Transported Auto Regressive Temperature) felhasználásával (Molkenthin et al. (2014)). A modell három forrásból származó véletlenszerű információ útját írja le, melyeket három egymástól független áramlás szállít. Bár a START is az advekción-diffúziós egyenlet segítségével állítja elő a hálózatot, ez a modell nagy mértékben különbözik az előzőtől a források megléte miatt. A hálózat csúcsainak fokszáma ennek megfelelően nem csak az áramlási sebességtől függ, mint az előző esetben, hanem a forrástól való távolság növekedésével csökken (6. ábra). A szabályos (a-) és a perturbált (b-) rácson a foksámok térbeli eloszlása hasonló, ez a mintázat azonban a csomópontokkal rendelkező rácson (c-) nem figyelhető meg. Azt is láthatjuk, hogy a forrástól legtávolabbi helyeken az áramláshoz adott véletlenszerű fluktuációk miatt a foksámok nullához tartanak.



6. ábra: A START-modellből készített hálózat fokszámai a) szabályos, b) perturbált, c) két csomóponttal rendelkező rácson (Molkenthin et al. (2014))

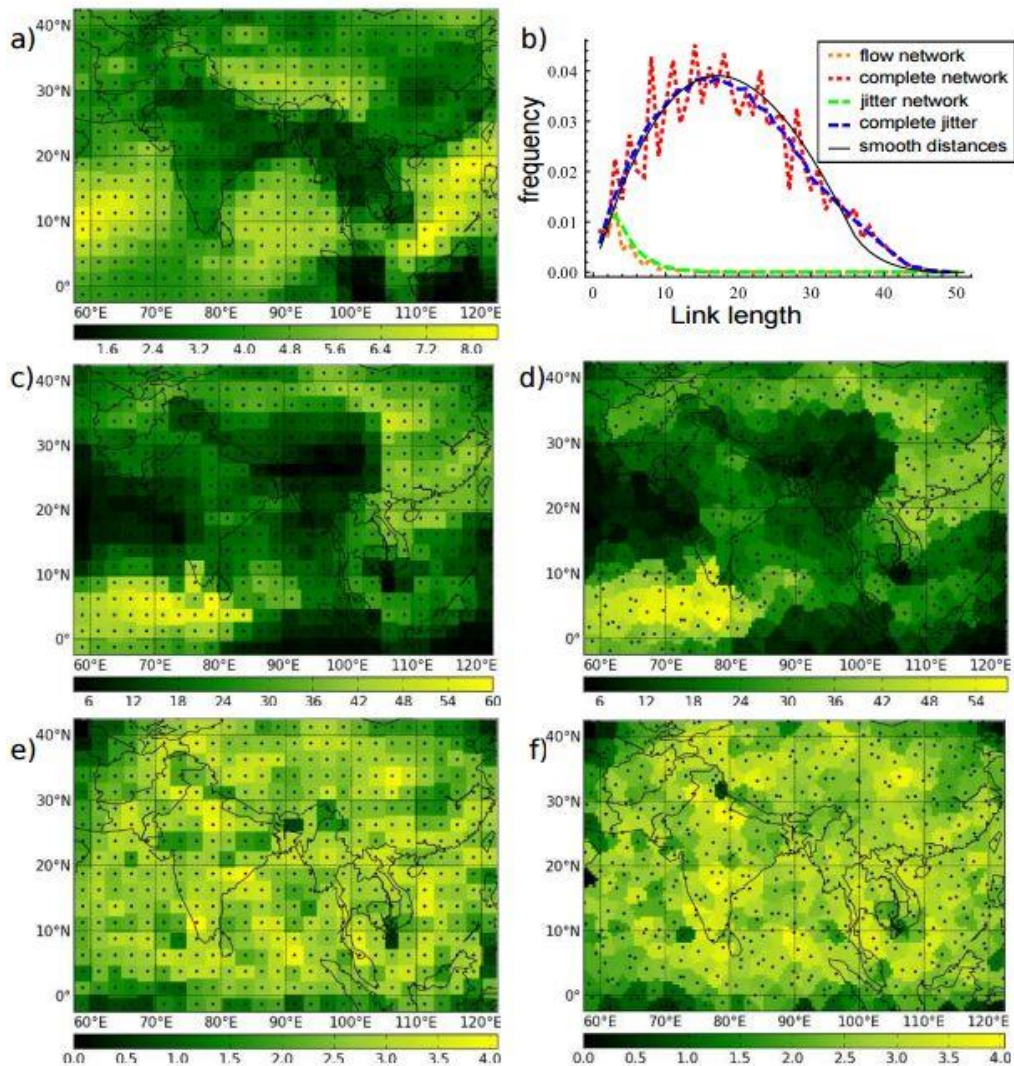
A kapcsolatok hosszának eloszlása a START-modell esetén is hasonló az áramlási hálózatnál megfigyeltekhez (7. ábra).



7. ábra: A kapcsolatok hosszának eloszlása a három rács típusban a START-modell alkalmazásával
(Molkenthin et al. (2014))

A csúcsok közöttiségének eloszlása a foksámokhoz hasonló mintázatot követ, ez szintén azzal magyarázható, hogy az áramláshoz adott véletlen fluktuációk miatt az áramlás belsejében elhelyezkedő pontok egymással dinamikus kapcsolatban állnak, míg a többi terület gyakorlatilag nincs összeköttetésben semmivel.

A két modell analízisét követően valós meteorológiai adatokból képzett hálózaton is elvégezték a mintavételezés térbeli eloszlásának különbözőségéből adódó eltérések vizsgálatát (Molkenthin et al. (2014)). A mérési adatok az indiai monszun térség 2011. évi napi hőmérséklet-anomáliái (NCEP/NCAR) voltak, $2,5 \times 2,5$ fokos rácsfelbontással, ami körülbelül $a = 280$ km-es rácsállandónak felel meg. A hálózatokat a korábbi fejezetben leírtak szerinti korrelációs módszerrel alkották meg az anomália idősorokból, 5%-os küszöbértékkel, azaz az 5%-nyi legmagasabb korrelációs együtthatójú pontok között létesítettek kapcsolatot. A térbeli hatások vizsgálatának érdekében kettő rácson határozták meg a meteorológiai hálózatot: egy szabályoson, melynek pontjai a $2,5 \times 2,5$ fokos cellák közepén helyezkedtek el, valamint egy perturbált rácson, amelynek pontjai úgy adódtak, hogy a szabályos rács csúcsait egy véletlenszerű, $\pm a/2$ közötti értékkel eltolták a hosszúsági és szélességi körök mentén is. Az új idősorokat minden pontra a cellákat határoló csúcsok adatainak súlyozott átlagával adták meg, ahol a súlyozás az új pont és a rács pont euklideszi értelemben vett távolsága. Az így kapott két hálózaton végzett elemzés eredményét a 8. ábra mutatja.



8. ábra: NCEP/NCAR adatok korrelációs hálózata. a) átlagos szélesség, b) kapcsolatok hosszának eloszlása, c) foksám szabályos rácson, d) foksám perturbált rácson, e) közöttiség szabályos rácson, f) közöttiség perturbált rácson (*Molkenthin et al. (2014)*)

A foksámok a térképen ábrázolt régió délnyugati részén a legnagyobbak, ami egybeesik az Indiai-óceán felszíni áramlataival. Megállapíthatjuk, hogy az óceáni áramlatok nagyobb mértékben befolyásolják a hálózatok foksámát ebben az esetben, mint a légköri áramlatok. A korábbi modelleredményekkel szemben a nagy szélességek és foksámok közötti összefüggés itt nem látható, a legnagyobb áramlási sebességekkel rendelkező tibeti területekhez tartoznak a legkisebb foksámok (8. ábra, a), c) és d)). Ez vélhetően orografikus okokkal magyarázható, hiszen ez a térség meglehetősen elzárt a hálózat többi pontjától magas tengerszint feletti magassága miatt, így még ha sok kapcsolat is lenne ezen a területen belül, a foksám kicsi lenne, mert a többi ponttal kisebb valószínűséggel van összeköttetésben. Egy másik nagy foksámú terület a kínai partvidék (északi szélesség 35°–40°), ez a nyugatias

szelekkel esik egybe. A szabályos és perturbált rács hálózatának (c) és d) ábrák) összehasonlításával azt a megállapítást tehetjük, hogy a mintavételezés térbeli különbségei nem okoznak nagy eltéréseket a hálózat csúcsainak fokszámában. Ezzel szemben a közöttségek szempontjából nagyobb eltérés mutatkozik a két hálózat esetén (8. ábra e) és f)). A legtöbb pontnak közepes közöttsége van, az extrém értékeket azonban más területeken találhatjuk meg a szabályos és a perturbált rácson (például az é. sz. 25° , k. h. 90° -on elhelyezkedő pont közöttsége a szabályos rács esetén az egyik legkisebb, míg a másik esetben teljesen átlagos). Az egyelőre tisztázatlan, hogy a közöttségek terén mutatkozó különbségnek milyen fizikai okai lehetnek, de ez arra is utalhat, hogy a közöttség túl érzékeny mérőszám a rendszer dinamikájának leírásához ebben a vizsgálatban.

Összefoglalásként tehát azt mondhatjuk, hogy a fokszám kevésbé, a közöttség azonban érzékenyebben reagál a mintavételezés térbeli különbségeire. Kellően sima (differenciálható) kapcsolathossz-eloszlásfüggvény meglétét azzal biztosíthatjuk, ha a hálózat csúcsainak távolsága között a vizsgált skálán minden távolságot megtalálunk, azaz a szabályos rács-hálózat nem feltétlenül a legalkalmasabb választás. A hálózatban lévő esetleges csomópontok hatása azonban még nem teljesen tisztázott, különböző elrendezésű clusterok más-más módon befolyásolják hálózatunk topológiáját (5. ábra, c) és d) oszlopok), ez a terület mindeképpen további kutatásokat igényel. Szintén hasznos lehet a jövőben a csomópontok nélküli, de különbözőképpen perturbált rácsok hálózatának összehasonlítása. Utolsó és talán legfontosabb konklúzióként pedig megállapíthatjuk, hogy amíg a mintavételezés térbeli eloszlása homogénnek tekinthető, addig a konkrét térbeli eloszlások közti különbségek nem okoznak nagy változásokat a hálózatok topológiájában egyik vizsgált esetben sem. Látványos torzulásokat kizárólag a szignifikánsan inhomogén mintavételezésnél tapasztalhatunk (5. és 6. ábrák).

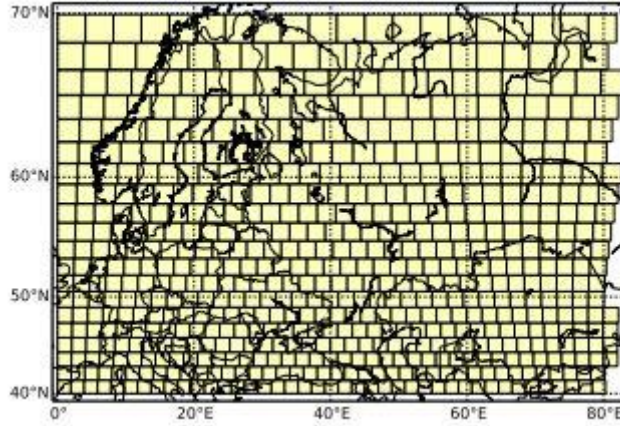
7. Esettanulmányok

I.: Domináns transzport útvonalak egy légköri blocking helyzetben

Az első esettanulmány (*Ser-Giacomi et al. (2015)*) a Kelet-Európát és Nyugat-Oroszországot érintő, 2010 nyarának blocking helyzetét elemzi hálózatelméleti módszerekkel. A tanulmány egy lagrange-i típusú áramlási hálózat segítségével olyan *legnagyobb valószínűségű útvonalakat* (most probable paths, MPP) számít ki, amelyeket a folyadékrezecskék követnek a légköri mozgások során, ezek a trajektóriák pedig visszaadják az ómega-blocking helyzet geometriáját. Arra az esetre, ha a legnagyobb valószínűségű útvonal nem lenne elég reprezentatív, definiálták az úgynevezett *nagy valószínűségű útvonalakat* (highly probable paths, HPP) is, ez pedig lehetőséget biztosít a légköri transzportfolyamatok pontosabb leírására.

Individuális légrészecskék útját alapvetően a kezdeti feltételekből számított trajektóriák adják meg. Ez azonban csak egy determinisztikus rendszerben, precízen meghatározott kezdeti értékekkel működik, gyakorlati alkalmazások során, főleg földtudományok terén a mozgáshoz adott sztochasztikus taggal modellezik annak véletlenszerűségét. A pontatlanul megadott kezdeti feltételek nagyon eltérő trajektóriákhoz vezethetnek (pl. spagetti-diagramok), ezért szükség van a mozgásokat legjobban leíró trajektóriák kiválasztására. Ezek az optimális útvonalak (MPP-k és HPP-k) pedig megadják a vizsgált mozgásrendszer alapvető geometriáját.

Az optimális útvonalakat a 2010. július 20–30. közötti időszakban Kelet-Európa és Nyugat-Oroszország felett uralkodó ómega-blocking helyzetre számították ki. A tér diszkretizációja érdekében a térséget felosztották 626 db egyenlő területű boxra, minden egyes doboz a hálózat egy csúcsának felelt meg (9. ábra).



9. ábra: A vizsgált terület és annak diszkretizációja (Ser-Giacomi et al. (2015))

A hálózat pontjai közötti kapcsolatok az egyes térrészek között kicserélődő légrézecskek trajektóriái segítségével határozhatók meg. A vizsgált $[t_0, t_M]$ időtartamot τ hosszúságú időlépcsőkre bontva ($t_l = t_0 + l \cdot \tau$, $l = 1, 2, \dots, M$), majd minden egyes időlépcsőre a mozgásegyenletet integrálva nyomon követhetjük az ideális folyadékrezsceskek útvonalát. A rendszer dinamikáját az $\mathbf{A}^{(l)}$, ($l = 1 \dots M$) szomszédsági mátrixokkal írhatjuk le, melyben az $\mathbf{A}_{IJ}^{(l)}$ elem értékét azon részecskek száma adja meg, amelyek a hálózat I pontjából J pontjába kerültek (t_{l-1}, t_l) időlépcső alatt.

Az így kapott trajektóriákat a hálózat pontjaira diszkretizált útvonalakként értelmezzük, egy M hosszúságú-, μ útvonal I és J csúcsok között annak az $(M + 1)$ db pontnak a sorrendjét jelenti, amiket a vizsgált részecske t_0 időpontban I -ből indulva, t_M időpillanatban J -be érve az útja során érintett. Amennyiben a transzportra Markov-folyamatként tekintünk (azaz a rendszer jövőbeli állapota csak a jelentől függ, a múlttól nem, még akkor sem, ha ismert a múltbeli állapota), akkor minden ilyen μ útvonal valószínűségét az alábbi képlet határozza meg:

$$(p_{IJ}^M)_\mu = \mathbf{T}_{Ik_1}^{(1)} * \left[\prod_{l=2}^{M-1} \mathbf{T}_{k_{l-1}k_1}^{(l)} \right] * \mathbf{T}_{k_{M-1}J}^{(M)}, \quad (7)$$

$$\text{ahol } \mathbf{T}_{k_{l-1}k_1}^{(l)} = \frac{A_{k_{l-1}k_1}^{(l)}}{\sum_j A_{k_{l-1}j}^{(l)}} \quad (8)$$

Az I és J csúcsok közötti lehetséges M hosszúságú útvonalak közül (8) alapján számított legnagyobb valószínűségű lesz az MPP. Ezt az útvonalat a gráfelméletben gyakran alkalmazott Dijkstra-algoritmussal, számítógép segítségével megkaphatjuk.

M magas értéke (vagyis hosszú útvonal) esetén az MPP nem elég reprezentatív, általában nem rendelkezik szignifikáns valószínűséggel, azonban ilyen esetekben is kiválasztható a rendszer dinamikáját legjobban leíró trajektóriák egy halmaza, ezek a nagy valószínűségű útvonalak (HPP). Egy ilyen $K_{IJ}^M(r, \varepsilon)$ halmazt két paraméter, a rangja ($0 \leq r \leq M-1$) és küszöbértéke ($0 \leq \varepsilon \leq 1$) jellemez, és tartalmazza azon r -hosszúságú útvonalakat, amelyek valószínűsége az MPP valószínűségének ε -szorosánál nagyobb. Az adott útvonal kezdő és végpontja között r db pontot fixnek választva a korábban bemutatott módszerrel kereshetők olyan MPP trajektóriák, amelyek áthaladnak a kiválasztott fix pontokon. Amennyiben ezen útvonalak valószínűsége meghaladja az ε küszöbérték által meghatározott értéket, akkor ez az új útvonal HPP lesz. A HPP-k kiszámítása magas rang esetén meglehetősen számításigényes feladat (r függvényében exponenciálisan nő), azonban már r értékét alacsonynak (1, vagy 2) megválasztva is érdekes eredményeket kaphatunk.

A HPP-k relevanciájának vizsgálata esetén fontos kérdés lehet, hogy ezen trajektóriák mekkora távolságra futnak az MPP-től. Ezt egy átlagos távolságot kifejező függvénnyel adhatjuk meg legegyszerűbben:

$$\mathbf{d}(\mu_1, \mu_2) = \frac{1}{M-1} \sum_{i=1}^{M-1} d(k_i l_i), \quad (9)$$

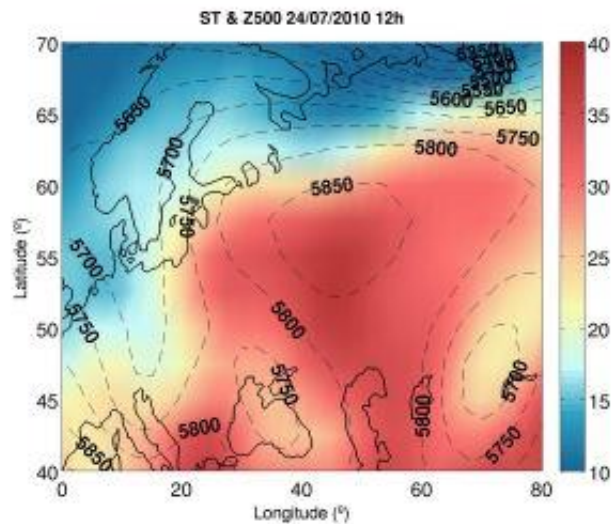
ahol $\mu_1 = \{I, k_1, \dots, J\}$ és $\mu_2 = \{I, l_1, \dots, J\}$ két útvonal I és J pontok között, $d(k_i, l_i)$ pedig a hálózat két csúcsának távolságát megadó metrika, melynek legcélszerűbb megválasztása esetünkben a két pont földrajzi távolsága. Az adott (I, J) pontok között futó MPP és összes HPP átlagos távolságát pedig a

$$\mathbf{D}_{IJ}^M = \frac{1}{N_{IJ}^M} \sum_{\mu} \mathbf{d}(\mu, \eta_{IJ}^M), \quad (10)$$

mérőszám adja meg, N_{IJ}^M a HPP-k összes száma (I, J) pontok között, η_{IJ}^M az MPP trajektóriát, μ pedig a HPP-eket jelöli. \mathbf{D}_{IJ}^M nagy értéke azt jelenti, hogy az I és J pontok közötti utak valószínűsége térben „szét van kenve”, azaz szükség van a HPP-k vizsgálatára is, míg alacsony \mathbf{D}_{IJ}^M esetén a HPP-k alig térnek el a legvalószínűbb útvonaltól, így az MPP önmagában is elég reprezentatív, még akkor is, ha kicsi valószínűség érték tartozik hozzá.

Az előbb bemutatott optimális útvonalak segítségével megvizsgálták a 2010. nyarán Kelet-Európát sújtó erős hőhullámot. Az extrém magas hőmérsékletek több, mint 50 000 ember halálát és hatalmas gazdasági károkat okoztak. A hőhullámot egy június végétől augusztus elejéig tartó légköri blocking helyzet okozta, ami nagyon stabil, anticiklonális körülményeket

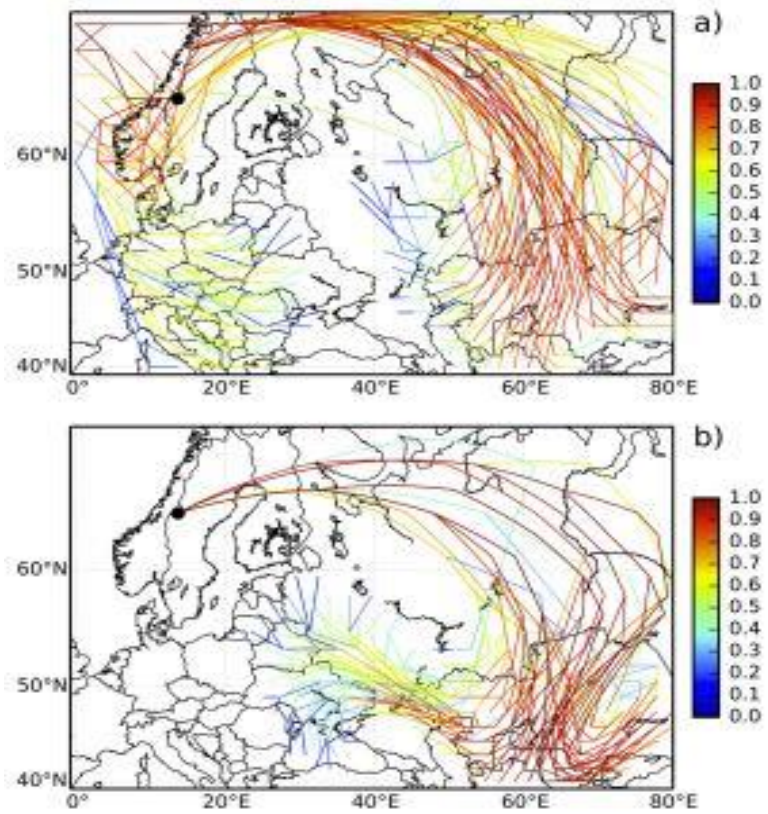
eredményezett egész Kelet-Európában. Morfológiáját tekintve egy ómega-blocking helyzet volt. Ez arról kapta nevét, hogy a nyomási mező alacsony-magas-alacsony szerkezetét leíró geopotenciál vonalak a görög Ω betű alakját követik (10. ábra). Az ómega-blocking az átlagosnál szárazabb és melegebb időjárást hoz az általa érintett területeken, elő- és hátoldalán pedig hűvösebbet, nedvesebbet.



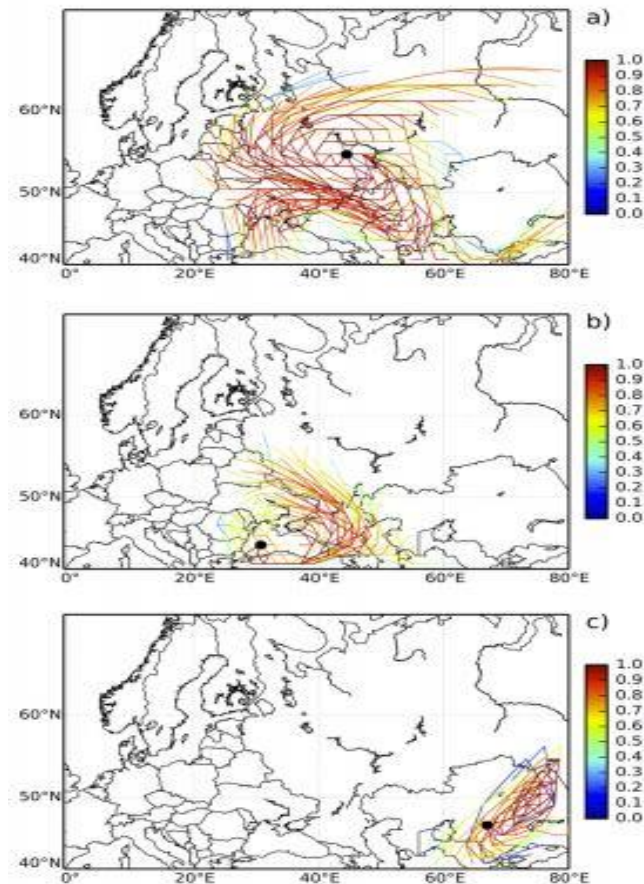
10. ábra: 500 hPa geopotenciál (szaggatott vonalak, méterben) és hőmérséklet (színek, °C-ban) a vizsgált területen, 2010.07.24. 12:00 UTC (*Ser-Giacomi et al. (2015)*)

A hálózat mind a 626 db csúcsából 800 db ideális folyadékreszecske útját követték végig a FLEXPART terjedési modell segítségével 5000 méteres magasságban $\tau = 12$ h időlépcsővel. A vertikális mozgásokat elhanyagolták, csak a nagyskalájú horizontális transzportot vizsgálták. Az eredményeket a 11.–13. ábrák mutatják.

A legnagyobb valószínűségű értékek (11. ábra, vörös szín) követik az anticiklonális (óramutató járásával megegyező) áramlás irányát a magas nyomású terület körül, anélkül, hogy benyúlnának a központba. Néhány kisebb valószínűségű MPP (11. ábra, sárgás és kékes színek) viszont ciklonális mozgásiránnyal dél felé mutat. Az is megállapítható, hogy az MPP-k időben elég változékonyak (11. ábra a) és b) összehasonlítása): -ugyanazt a kezdőpontot vizsgálva néhány nap elteltével teljesen eltűnt a déli ciklonális áramlás.



11. ábra: $M = 9$ hosszúságú MPP-k, $\tau = 12$ h időlépcsővel (4,5 nap) az áramlási hálózat egy skandináviai pontjából, a) július 25., b) július 20. kezdődátummal. A színek a valószínűségeket adják meg normált logaritmikus skálán. A valószínűségek 10^{-3} – 10^{-15} közé esnek. (Ser-Giacomi et al. (2015))

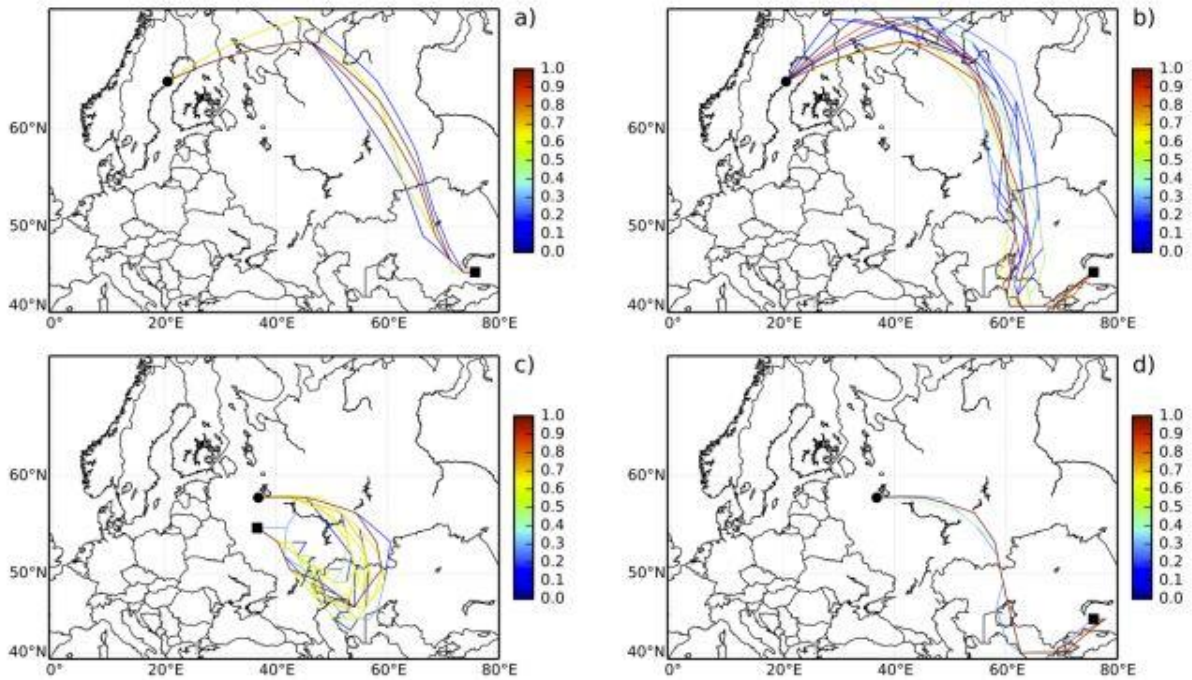


12. ábra: $M = 9$ hosszúságú MPP-k, $\tau = 12$ h időlépcsővel (4,5 nap) július 20. kezdő dátummal a) a térség egy magasnyomású, b) és c) a térség egy alacsony nyomású pontjából indítva. A színezés ugyanaz, mint a 11. ábra esetén. (Ser-Giacomi et al. (2015))

A 12. a) ábra szintén tisztán mutatja az anticiklonális áramlást. A b) és c) ábrákon az ómega-block két oldalán elhelyezkedő ciklonális áramlás figyelhető meg. Mindkét esetben a trajektóriák nagyon kompaktnak tekinthetők, a vizsgált 4,5 nap alatt a légrészecskék alig változtatták meg helyüket, ami jól mutatja a blocking helyzet stabilitását (12. b) és c) ábrák).

Fontos megemlíteni, hogy a 11. és 12. ábrákon nem spagetti-diagramokat látunk, amik több lehetséges trajektóriát rajzolnának ki különböző kezdeti feltételekkel, ez ebben az esetben a hálózat minden pontjából 800 db kiinduló utat jelentene. Itt azonban csak minden lehetséges pontpárra 1 db trajektóriát látunk kirajzolva (ez az MPP, a trajektóriák közül a legnagyobb valószínűségű).

A HPP-k és MPP-k viszonyát különböző területeken és különböző időpontokban a 13. ábra mutatja be.



13. ábra: MPP-k és HPP-k $r = 1$, $\varepsilon = 0,1$ paraméterekkel, különböző kezdő- (kör) és végpontokhoz (négyzet). A színskála normált logaritmus, az MPP valószínűségétől annak ε -szorosáig csökken. a) $M = 8$, július 25-től, b) $M = 12$, július 25-től, c) $M = 11$, július 20-tól, d) $M = 11$, július 20-tól

Már $r = 1$ esetén is sokat segítenek a HPP-k az áramlás geometriájának leírásában. Mind a 4 esetben a HPP-k szorosan követik az MPP trajektóriát (13. a) – d) ábrák), az ezek távolságát jellemző D_{ij}^M mérőszám értéke az összes esetben jóval a rácsállandó (kb. 166 km) alatt marad, ami azt mutatja, hogy az optimális útvonalak egymás közelében futnak.

A bemutatott optimális útvonalak (MPP-k és HPP-k) tehát segítséget nyújtanak légköri transzport folyamatok analizésében és vizualizációjában. Az esettanulmányban ezen trajektóriák futása visszaadta a vizsgált blocking helyzet alapvető geometriáját, egy központi anticiklonális, valamint az ezt közrefogó két ciklonális áramlást. Segítségükkel betekintést kaptunk a vizsgált mozgásrendszer stabilitásába. Megállapíthatjuk, hogy a HPP-k az összes vizsgált esetben közel futottak az MPP-khez, és ezek együttese jól leírta az áramlással haladó légréseccskék útját.

II.: Klímahálózatok fejlődésének jellemzése

A második esettanulmányban (*Tupikina et al. (2014)*) a Föld klímahálózatának időbeli fejlődésének számszerűsített jellemzésének lehetőségeiről lesz szó. Mivel más komplex rendszerekhez hasonlóan bolygónk éghajlata is változik időben, ezért az ezt reprezentáló hálózat tulajdonságai is időfüggőek. A komplex hálózatok topológiájában történő változásokat nehéz szemmel megfogni, így szükség van a különbségek számszerűsítésére. Erre már korábban is voltak kísérletek, például az 1950-ben Richard Hamming, amerikai matematikus által bevezett *Hamming-távolság*, ami azon élek száma, amelyek az egyik gráfban megtalálhatók, a másikban pedig nem. Ez a mérőszám azonban önmagában keveset árul el a hálózatok fejlődéséről, hiszen két különböző gráf egy referencia-gráfhoz viszonyított *Hamming-távolsága* lehet ugyanannyi, miközben teljesen más szerkezettel rendelkeznek. Ezen probléma áthidalására vezették be az úgynevezett CCEF-t (Common Component Evolution Function), melynek kiszámítása a két hálózat közös élein alapszik.

Tekintsünk egy n csúcsból álló hálózatot, amely éleinek struktúráját egy $n \times n$ -es szomszédsági mátrix (\mathbf{A}) jellemez, és kövessük ezen hálózat időbeli alakulását (N_1, \dots, N_T). A csúcsok helyzetét vegyük állandónak, csak a pontok összekötöttsége változzon. N_i és N_j hálózatok *közös hálózatának* ($CC(N_i, N_j)$) azt a gráfot nevezzük, amely ugyanazon csúcsokból áll, és azokat az éleket tartalmazza, amelyek mindkét hálózatban megtalálhatók. Ez a közös hálózat bármely $k+1$ hálózatra indukcióval előállítható: $CC(N_i, \dots, N_{i+k+1}) = CC(CC(N_i, \dots, N_{i+k}))$.

A CCF-függvény (Common Component Function) megszámlálja k db hálózat közös kapcsolatait: $CCF(N_i, \dots, N_{i+1}) = \|CC(CC(N_i, \dots, N_{i+k-1}), N_{i+k})\|$, ahol $\|N_i\|$ az N_i gráf éleinek számát jelöli. Megállapodás szerint $CCF(N_i) = CCF(N_i, N_i)$.

A CCF-ekből átlagot számolva kaphatjuk a CCEF* (non-normalized Common Component Evolution Function) értékét:

$$CCEF^*(\tau) = \frac{1}{T-\tau} \sum_{i=1}^{T-\tau} CCF(N_i, N_{i+\tau}), \quad (11)$$

ahol τ a két hálózat közötti időlépcső, $\tau \in [0, T-1]$. CCEF* maximális értékét a kezdeti hálózat adja meg:

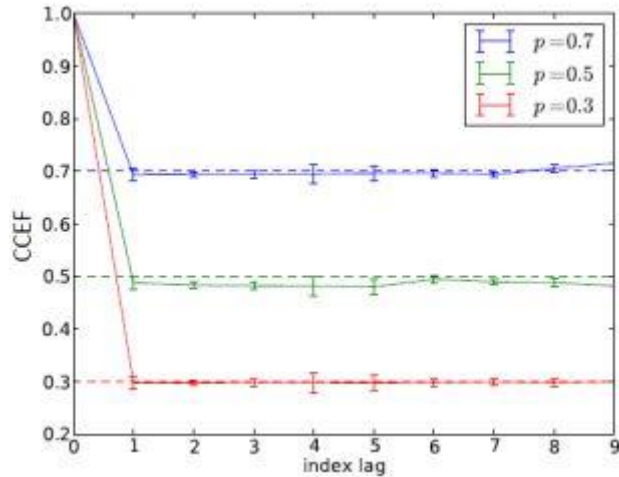
$$CCEF^*(0) = \frac{1}{T} \sum_{i=1}^T CCF(N_i), \quad (12)$$

melynek segítségével megkapjuk a normált CCEF értéket:

$$CCEF(\tau) = \frac{CCEF^*(\tau)}{CCEF^*(0)}, \quad (13)$$

amelyet a következőkben a klímahálózatok jellemzésére használunk. A CCEF bizonytalanságát pedig az összes CCEF érték szórásával adhatjuk meg.

A CCEF-t elsőként az Erdős-Rényi-féle véletlen hálózatokon tesztelték, melyekben minden pontpár közötti kapcsolat valószínűsége p . A hálózatok 100 csúcsból álltak, 0,3; 0,5 és 0,7 összekötöttségi valószínűségekkel. Mindhárom esetre készítettek egy időben változó gráfsorozatot, majd kiszámították rá a CCEF-t. Az eredmények a 14. ábrán láthatók.



14. ábra: CCEF különböző összekötöttségi valószínűségekkel (p) véletlen hálózatokon (Tupikina et al. (2014))

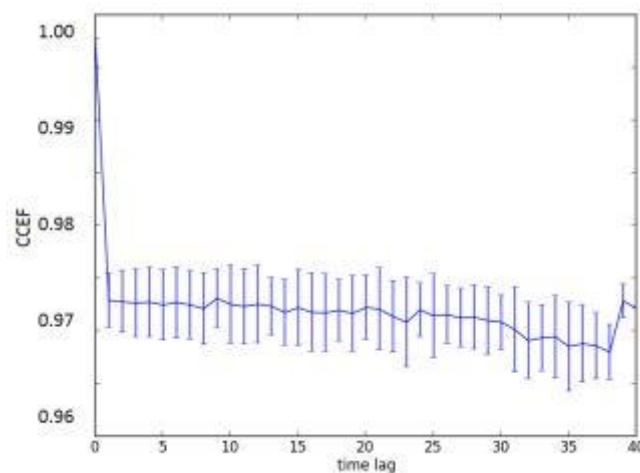
Mindhárom esetben a CCEF értéke 1-től a $CCEF(\tau) \approx p$ értékéig csökken. $CCEF(0)$ értéke a definícióból triviálisan következően 1, $\tau \neq 0$ -ra pedig kiszámíthatjuk a CCEF várható értékét: n db pont esetén a kapcsolatok összes lehetséges száma $n(n-1)/2$, várható értéke pedig ennek a p -szerese. Annak a valószínűsége, hogy minden él megjelenik mindkét gráfban $p^2 n(n-1)/2$, amelyből a normálás után megkaphatjuk a közös kapcsolatok várható értékét:

$$f(p) = \frac{p^2 n(n-1)}{pn(n-1)} = p. \quad (14)$$

Emiatt a CCEF értéke p körüli lesz.

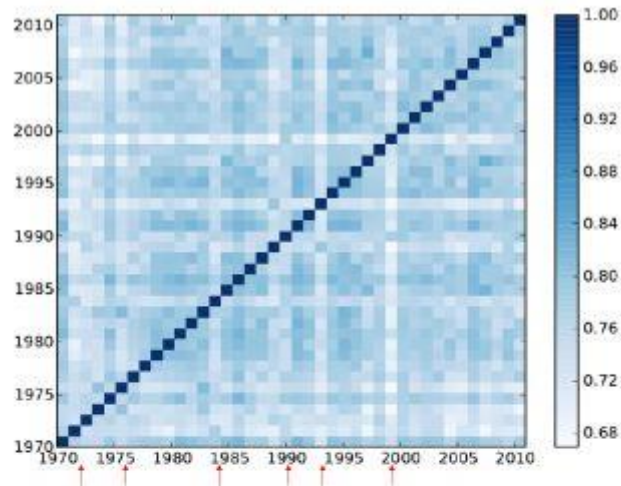
A CCEF számítógépes áramlási modellekből készített hálózatokra is jól működött, a lineárisan, valamint a periodikusan változtatott áramlás időbeli alakulását is precízen követte (Tupikina et al. (2014)).

A CCEF-t végül az ázsiai monszun jelenségének vizsgálatára is alkalmazták. A hálózatokat az 1970-2011 évek hőmérsékleti anomáliáinak reanalízise alapján (NOAA, 2014), $2,5 \times 2,5$ fokos rácsfelbontással készítették el az ötödik fejezetben bemutatott korrelációs technika felhasználásával, 5%-os küszöbértékkel. Az elemzés eredményét a 15. ábra mutatja be.

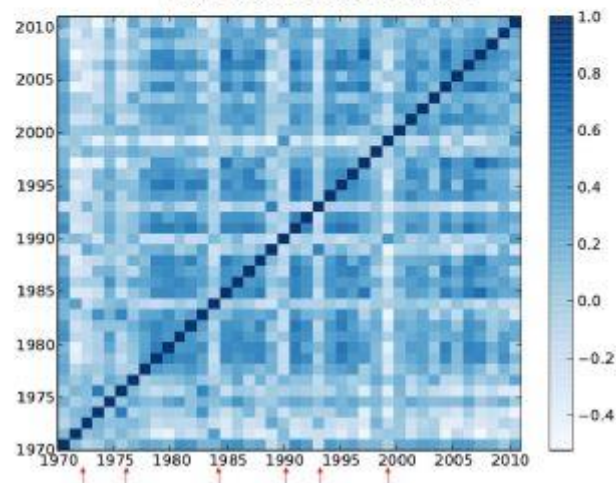


15. ábra: CCEF az 1970-2011 évek monszun adataira, éves időlépcsővel (Tupikina et al. (2014))

A 15. ábráról egyértelműen látszik, hogy az éves időskálán az ázsiai monszun hőmérsékleti klímahálózatának kicsi a tér- és időbeli változékonysága. A különbségek elemzésének érdekében kiszámolták a $CCF(N_i, N_j)$ értékeket minden (i, j) párra, i -t fixnek választva, $i, j \in [1970, 2011]$. A 16. (a) ábrán láthatók a szignifikánsan alacsonyabb értékekkel rendelkező CCF sorok és oszlopok. Az $i \in (1971-1973, 1975, 1984, 1989, 1993, 1999)$ évekre kapták a legkisebb CCF értékeket. Ezeket összehasonlítva a legerősebb ENSO jelenségekkel (El Niño 3.4 index: 1972, 1982, 1988, 1992, 1997) megállapíthatjuk, hogy az ENSO jelentősen befolyásolhatja az ázsiai monszunt (mivel az ENSO az északi félteke telén játszódik le, ezért van egy 1 éves csúszás a két évszám-sorozat között). El Niño években tehát kevesebb közös link van a monszun hálózatban, és a CCF-ek korrelációja is alacsonyabb (16. (b) ábra).



(a) Common link recurrence



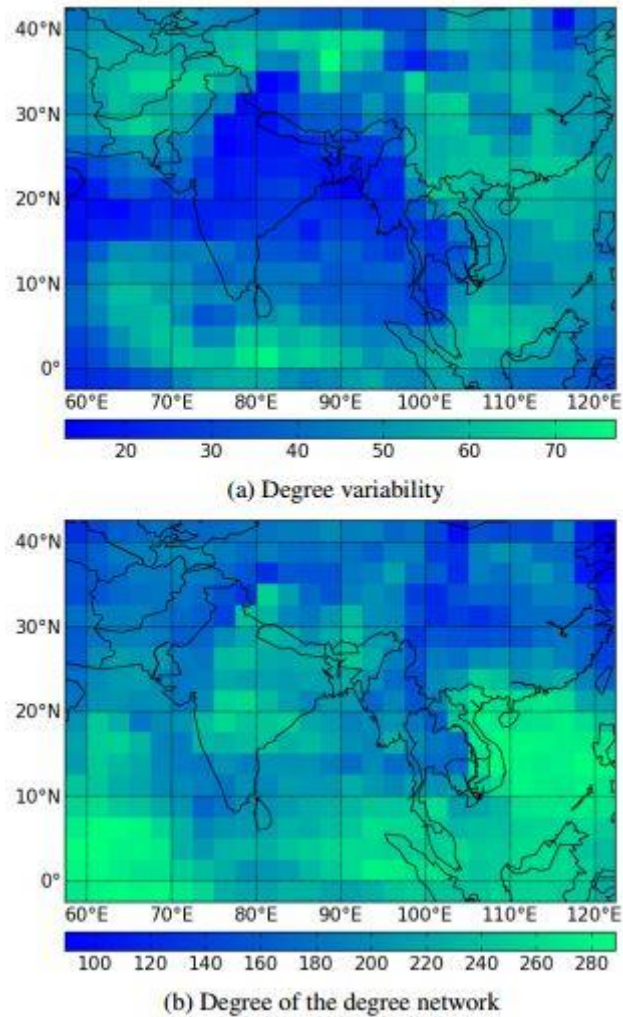
(b) Correlation of the common link evolution

16. ábra: (a) CCF-ek az 1970-2011 időszak monszun hőmérséklet adataiból. (b) A CCF-ek korrelációja

(*Tukipina et al. (2014)*)

Az ázsiai monszun klímahálózat topológiájának viszonylagos változatlanságának ellenére tehát az ENSO-monszun távkapcsolat mégis kimutatható.

Az itt tapasztalt változásokat hálózatelméleti szempontból a legmagasabb fokszámú pontok fokszámának megváltozása okozhatta (17. ábra).



17. ábra: (a) A monszon klímahálózatának fokszám-változékonysága az 1970–2011 időszakban. (b) A fokszám-idősorokból korrelációval készített hálózat fokszámai 1970–2011 időszakra (Tupikina *et al.* (2014))

A fokszámok változékonysága ott magas (alacsony), ahol a fokszámok magasak (alacsonyak) (Tupikina *et al.* (2014)). A 17. (a) és (b) ábrák összehasonlításával az is megállapítható, hogy az alacsony változékonyságú területeken (India) magas a (b) hálózat fokszáma, azaz ez a régió hálózati szempontból állandónak tekinthető. A nagy változékonyságú területek (Indiai-óceán, Dél-kínai-tenger) esetén is magas fokszámokat találunk a (b) hálózatban, ez arra utalhat, hogy itt a változások nagyok, de szinkronizáltak történnek. Ez a szinkronizáció a térség északi részén (Afganisztán-Pakisztán-Himalája vonalban) nem figyelhető meg, ennek oka a nyugati szelek befolyásoló hatása lehet.

A fejezetben bemutatott módszereket (CCF, CCEF) tehát sikerrel alkalmazták véletlenszerű-, modell- és valós klímahálózatok időbeli változásának számszerűsítésére, segítségével pedig az ENSO-monszun távkapcsolatra is rávilágítottak.

8. Jövőbeli kilátások

A dolgozatban áttekintettük a hálózatelmélet rövid történetét, matematikai alapjait, valamint meteorológiai vonatkozásait. Megismerkedtünk a meteorológiai és klimatológiai hálózatok készítésének lehetőségeivel és ezen hálózatok térbeli függésével. Két esettanulmány segítségével pedig megvizsgáltuk egy-egy hálózatelméleti módszer (az optimális útvonalak, illetve a CCEF) meteorológiai alkalmazását, beillesztését a folyadékdinamikai kutatások, illetve a klímahálózatok időbeli fejlődésének és a távkapcsolatok vizsgálatának témakörébe.

Ahogy az már korábban említésre került, a hálózatelmélet egy nagyon fiatal tudományág, meteorológiai és klimatológiai alkalmazása pedig mindössze az elmúlt évtizedben kezdődött meg. Az eddigi eredmények nagyon biztatók, azonban még sok a homályos folt, amelyek további kutatásokat igényelnek.

Nagyon fontos lenne egy adott adatsorból az 5. fejezetben bemutatott különböző statisztikai és információelméleti módszerekkel előállított meteorológiai hálózatok topológiájának összehasonlító elemzése. Ez abban nyújthat segítséget a jövőben, hogy az adott kutatás vizsgálati tárgyához melyik eljárás passzol, azaz mely mérőszámokkal mutathatók ki a legszignifikánsabb eredmények. A 6. fejezetben felmerülő probléma, vagyis a különböző módon elhelyezkedő csomópontok hálózatok topológiájára gyakorolt hatásának kutatásának is megvan a létjogosultsága, hiszen a meteorológiai mérőhálózatok eloszlása a Földön közel sem tekinthető homogénnek. Szintén érdekes lehet a hálózatok harmadik dimenzióba történő kiterjesztése, ezáltal pedig a légköri vertikális mozgások hálózatelméleti analízise, –erre néhány kutató már ígéreteket is tett (*Ser-Giacomi et al. (2015)*)–, valamint a hálózatelmélet időjárás- és éghajlatelőrejelzési gyakorlatba történő beillesztése.

További ötletekhez csak meg kell találnunk néhány már ismert hálózatelméleti probléma meteorológiai analógiáját, egy ilyen lehetőség például a számítógépes hálózatokat érő támadások, valamint a légköri szennyezőanyag terjedés párhuzama: hackerek általában a központi szerepkörű (nagy fokszámú, vagy magas közöttség értékkel rendelkező) computereket támadják meg, hogy ezzel hátráltassák a hálózatban történő információáramlást. Ennek mintájára a meteorológiai hálózatban az ilyen kulcsfontosságú pontok közelében elhelyezkedő szennyezőanyag kibocsájtókat szigorúbb előírásoknak kitéve, védelmüknek nagyobb figyelmet szentelve több területet megkímélhetünk az esetleges környezeti károktól.

A lehetőségeink száma tehát gyakorlatilag végtelen, hatalmas adatmennyiség áll rendelkezésre, a sok meteorológiai paraméterből (például hőmérséklet, légnyomás, csapadékmennyiség, szélesség, stb.) különböző módszerekkel készíthetünk hálózatokat. A kutatásoknak csak a magas számítás- és tárhelyigény szabhat határt, egyéb matematikai, illetve számítástechnikai problémába ritkán ütközhetünk, a hálózatok konstruálását és elemzését elvégző programok, például *ClimNet*, *Par@Graph* (Deza és Ihshaish (2015)) rendelkezésre állnak.

Ami Magyarországot és a Kárpát-medencét illeti, a hálózatelméleti vizsgálatok nehézségét a térség csekély kiterjedése okozhatja (az eddigi kutatások szinte mindegyike nagyobb területet ölelt fel), azonban a jövőben érdekes eredményeket kaphatunk mérsékelt övezeti légköri mozgásrendszerek Magyarország időjárását alakító hatásairól, vagy hálózatelméleti módszerekkel olyan távkapcsolatokat detektálhatunk, melyek hazánk éghajlatát befolyásolják.

Köszönetnyilvánítás

Rengeteg köszönet illeti a témavezetőmet, Pieczka Ildikót, akinek odaadó munkája, ötletei, türelme és biztatása nélkül nem készült volna el a dolgozatom. A kisebb részletekben nyújtott segítséget szeretném megköszönni Farkas Ritának és Berényi Alexandrának. Szintén hálával tartozom a családomnak és barátaimnak, akik a szakdolgozat készítése alatt végig támogattak.

Irodalomjegyzék

Barabási, A.-L., 2013: Behálózva. Helikon, Budapest. 320p.

Deza, I., Ihsaish, H., 2015: The construction of complex networks from linear and nonlinear measures – Climate Networks. *Procedia Computer Science*, 51, 404-412p.

Dong Zhou, Gozolchiani, A., Ashkenazy, Y., Havlin, S., 2015: Teleconnection paths via climate network direct link detection. *Physical review letters*, 115, 268501

Freeman, L., 1977: A set of measures of centrality based on betweenness. *Sociometry*, 40, 35-41p.

Hajnal, P., 1997: Gráfelmélet. Polygon, Szeged. 293p.

Makó, Z., Téglási, I., 2012: Indoklás és bizonyítás. Eszterházy Károly főiskola, Eger. 193p.

Molkenthin, N., Rehfeld, K., Marwan, N., Kurths, J., 2013: Networks from flows – from dynamics to topology. *Science Reports*, 4, 4119

Molkenthin, N., Rehfeld, K., Stolbova, V., Tupikina, L., Kurths, J., 2014: On the influence of spatial sampling on climate networks. *Nonlinear Processes Geophysics*, 21 (3), 651-657p.

Ser-Giacomi, E., Vasile, R., Recuerda, I., Hernández-García, H., López, C., 2015: Dominant transport pathways in an atmospheric blocking event. *Chaos*, 25, 087413

Tupikina, L., Rehfeld, K., Molkenthin, N., Stolbova, V., Marwan, N., Kurths, J., 2014: Characterizing the evolution of climate networks. *Nonlinear Processes Geophysics*, 21, 705-711p.