Holme-Kim és Konfigurációs modellek összehasonlítása

Bedécs Flóra Blanka  
Bognár Bence  
Kovács Adél

**Bevezetés**

A hálózati modellek tanulmányozása egyre nagyobb szerepet kap az összetett rendszerek, például társadalmi kapcsolatok és biológiai hálózatok megértésében. Két széles körben tanulmányozott hálózati modell a Holme-Kim és a Konfigurációs modell. Ez a jelentés átfogó összehasonlítást nyújt ezekről a modellekről, beleértve elvi hátterüket, módszertanukat és alkalmazásaikat. Részletes technikai elemzést is tartalmaz Python-alapú szimulációk segítségével, vizualizációkkal és metrikákkal kiemelve a sajátosságaikat.

A Holme-Kim modell a Barabási-Albert modell kiterjesztése, amelyet a preferenciális csatlakozási mechanizmus alapján definiálnak. Matematikailag a csomópont i-hez való csatlakozás valószínűsége:

P(i) = k\_i / Σ(k\_j),

ahol k\_i az i-edik csomópont fokszáma. A Holme-Kim modell továbbá egy triadikus záródási mechanizmust vezet be, amely valószínűségi alapon hozzáad egy további élt az új csomópont és egy meglévő, közös szomszéd csomópont között. Ez a mechanizmus növeli a hálózat klaszterezési együtthatóját, miközben megőrzi a skálamentes tulajdonságokat.

A Konfigurációs modell egy olyan véletlen gráfgenerálási módszer, amely egy előre meghatározott fokszám-vektort vesz bemenetként. Olyan hálózatot hoz létre, amelyben a fokszámeloszlás megfelel a bemenetnek, miközben az élkapcsolatok véletlenszerűek maradnak. Ez a modell az egyik alapvető eszköz a nullhipotézis teszteléshez a hálózattudományban, mivel lehetővé teszi a fokszámeloszlás hatásának vizsgálatát különböző hálózati tulajdonságokra.

**Módszertan**

**Eszközök és könyvtárak**

* **Python**: Szimulációkhoz és vizualizációkhoz használt.
* **NetworkX**: Egy erőteljes könyvtár, amely komplex hálózatok létrehozására, elemzésére és vizualizációjára szolgál.
* **Matplotlib**: Részletes hálózati tulajdonságokat ábrázoló grafikonok generálására.
* **Scipy**: Statisztikai elemzések támogatására, mint például a hatványtörvény-eloszlás illesztése.

# Kulcsmértékek

**1. Hálózati Növekedés és Dinamika**

* **Holme-Kim Modell:** A Holme-Kim modellben az új csomópontok nem véletlenszerűen kapcsolódnak a hálózathoz, hanem preferenciális kötődés alapján. Ez a **preferenciális csatlakozási szabály** a következő matematikai formulával írható le:

P(u)∝deg(u)P(u) \propto \text{deg}(u)P(u)∝deg(u)

ahol P(u)P(u)P(u) annak a csomópontnak a valószínűsége, hogy az új csomópont uuu-hoz csatlakozik, és deg(u)\text{deg}(u)deg(u) a csomópont fokszáma (kapcsolatok száma). Tehát az új csomópontok inkább azokhoz csatlakoznak, amelyeknek nagyobb fokszámuk van. Ezen kívül, az új csomópontok nemcsak egyetlen csomóponthoz csatlakoznak, hanem annak **szomszédaihoz** is, ami növeli a hálózaton belüli klaszterezést.

* **Configurációs Modell:** A Configurációs modell egy statikus modell, amelynél a csomópontok fokszáma előre meghatározott. A modell célja, hogy az élek véletlenszerű módon jöjjenek létre úgy, hogy a fokszám-eloszlás megfeleljen az előre meghatározott eloszlásnak. A csomópontok közötti élek létrejötte a következő szabályok szerint történik:

P(edge between i and j)∝kikj∑i,jkikjP(\text{edge between } i \text{ and } j) \propto \frac{k\_i k\_j}{\sum\_{i,j} k\_i k\_j}P(edge between i and j)∝∑i,j​ki​kj​ki​kj​​

ahol kik\_iki​ és kjk\_jkj​ az iii és jjj csomópontok fokszámai. Ez a **preferenciális kötődés** egy statikus változata, ahol az élek létrejötte a fokszámok szorzataként arányos.

**2. Fokszám-eloszlás**

* **Holme-Kim Modell:** Mivel az új csomópontok inkább a magas fokszámú csomópontokhoz csatlakoznak, a fokszám-eloszlás általában **power-law** típusú lesz. Ez azt jelenti, hogy a csomópontok többsége alacsony fokszámú (kevés kapcsolattal rendelkezik), míg egy kisebb részük kiemelkedően magas fokszámú (sok kapcsolattal rendelkezik). A power-law eloszlás matematikailag így írható le:

P(k)∼k−γP(k) \sim k^{-\gamma}P(k)∼k−γ

ahol P(k)P(k)P(k) a fokszám eloszlása, kkk a csomópont fokszáma, és γ\gammaγ az eloszlás kitevője, amely tipikusan 2 és 3 között van.

* **Configurációs Modell:** A fokszám-eloszlás előre meghatározott, és nem dinamikusan alakul, mint a Holme-Kim modellben. Az élek véletlenszerű hozzárendelése miatt a hálózat fokszám-eloszlása is **power-law eloszlású** lehet, ha az előzőleg megadott fokszámok így alakítják a hálózatot. Az eloszlás a következőképpen alakulhat:

P(k)∼k−γP(k) \sim k^{-\gamma}P(k)∼k−γ

itt a fokszámok eloszlása szintén power-law típusú, de a differencia az, hogy a fokszámok előre meghatározottak, és a modellezés nem dinamikusan, hanem a statikus kapcsolatok függvényében történik.

**3. Hálózati Szabályosság és Összetettség**

* **Holme-Kim Modell:** Az élek eloszlása szabálytalan és dinamikus, hiszen a csomópontok fokszámától függően egyre több új kapcsolat jön létre. A **lokális összekapcsolódottság** és a **klaszterezési hajlam** magas, mivel a kiegészítő kapcsolatok révén a csomópontok szorosabb kapcsolatokra tehetnek szert a szomszédjaikkal, így növelve a **klaszterezési koeficienst** (CCC):

C=3×haˊromszo¨gek szaˊmahaˊromszo¨gre eˊpu¨lo˝ lehetseˊges eˊlek szaˊmaC = \frac{3 \times \text{háromszögek száma}}{\text{háromszögre épülő lehetséges élek száma}}C=haˊromszo¨gre eˊpu¨lo˝ lehetseˊges eˊlek szaˊma3×haˊromszo¨gek szaˊma​

A Holme-Kim modellben ez a koefficiens magasabb, mivel az új csomópontok több szomszédos csomóponthoz is csatlakozhatnak.

* **Configurációs Modell:** A **Configurációs modellben** az élek véletlenszerűen jönnek létre, így a klaszterezés mértéke alacsonyabb lehet, mint a Holme-Kim modellben. A klaszterezési koefficiens alacsonyabb, mivel az élek nem a szomszédságok között keletkeznek, hanem az előre meghatározott fokszámok szerint.

**4. Skálázhatóság és Hálózati Méret**

* **Holme-Kim Modell:** Mivel a Holme-Kim modell **dinamikusan növekvő** hálózatot generál, az új csomópontok folyamatosan csatlakoznak a hálózathoz, így könnyen skálázható. A hálózatok méretének növekedésével az eloszlás folyamatosan alkalmazkodik, és a fokszámok dinamikusan alakulnak.

A növekvő hálózatban az új csomópontok csatlakozási valószínűsége a fokszám alapján a következőképpen változik:

P(csatlakozaˊs uˊj csomoˊponthoz)∝deg(u)∑ideg(i)P(\text{csatlakozás új csomóponthoz}) \propto \frac{\text{deg}(u)}{\sum\_{i} \text{deg}(i)}P(csatlakozaˊs uˊj csomoˊponthoz)∝∑i​deg(i)deg(u)​

* **Configurációs Modell:** A **Configurációs modell** előre meghatározott fokszámokkal dolgozik, így a hálózatok növekedését nem dinamikusan, hanem statikusan modellezi. A hálózatok skálázhatósága itt inkább a csomópontok fokszám-eloszlásának megfelelően történik, de az élek eloszlása nem változik a hálózat bővítésével.

**5. Hálózati Kialakulás és Evolúció**

* **Holme-Kim Modell:** A Holme-Kim modellben az új csomópontok csatlakozása a fokszámoktól és a szomszédok kapcsolataitól függően dinamikusan alakul. Ez azt jelenti, hogy az élek és a hálózati struktúra folyamatosan **evolválódik** az idő előrehaladtával. Az új csomópontok **preferenciális kötődése** és **szomszédos kapcsolatai** folyamatosan módosítják a hálózati szerkezetet.
* **Configurációs Modell:** A Configurációs modell nem dinamikus, tehát egyszeri hálózatgenerálásról van szó, amelyben az élek és a csomópontok kapcsolatai már az elején meghatározottak. A hálózat nem változik idővel, csak egyszerűen egy adott fokszám-eloszlás szerint generálódik.

**6. Klaszterezési Koeficiens**

* **Holme-Kim Modell:** A klaszterezési koeficiens CCC magasabb lesz, mivel az új csomópontok nemcsak egyetlen csomóponthoz, hanem annak szomszédaihoz is csatlakoznak, ami fokozza a **lokális csoportosulásokat**. Matematikailag a klaszterezési koefficiens magasabb lesz, mivel az élek közvetlen környezeti kapcsolatokhoz vezetnek.
* **Configurációs Modell:** A klaszterezési koeficiens alacsonyabb lehet, mivel az élek véletlenszerűen jönnek létre, és nem biztos, hogy közvetlenül a csomópontok szomszédságaihoz kapcsolódnak.

# Hálozat generálasa

**Holme-Kim modell megvalósítása**

holme\_kim\_graph = nx.powerlaw\_cluster\_graph(

n=2000, m=5, p=0.8

)

A powerlaw\_cluster\_graph függvény a NetworkX-ben Holme-Kim hálózatot generál, kombinálva a preferenciális csatlakozást és a triadikus záródást.

**Konfigurációs modell megvalósítása**

degree\_sequence = [holme\_kim\_graph.degree(n) for n in holme\_kim\_graph.nodes]

configuration\_graph = nx.configuration\_model(degree\_sequence)

configuration\_graph = nx.Graph(configuration\_graph) # Multi-edges és hurkok eltávolítása

A Konfigurációs modell a Holme-Kim hálózat fokszám-vektorát használja bemenetként, hogy összehasonlítható véletlen hálózatot hozzon létre.

**Fokszám-eloszlás Kikényszerítése**

A folyamat a következő lépéseket tartalmazza:

1. **Hurkok és párhuzamos élek eltávolítása:** A gráf generálása során minden olyan élt eltávolítunk, amely hurkot vagy párhuzamos kapcsolatot képezne a már meglévő élekkel. Ez biztosítja, hogy a hálózat tiszta és érvényes struktúrával rendelkezzen, elkerülve az olyan problémákat, amelyek a vizualizálás és az elemzés során zűrzavart okozhatnának.
2. **Csúcsok fokszámának kiegyenlítése:** Azokhoz a csúcsokhoz, amelyek fokszáma az eredeti fokszám alatt van, új éleket adunk hozzá. Ezzel biztosítjuk, hogy a csúcsok fokszám-eloszlása megfeleljen az elvárt modellezési szabályoknak.
3. **Fokszám-deficit kiválasztása:** A csúcsok közül azokat választjuk ki, amelyeknek a legnagyobb a fokszám-deficitjük, azaz azok, amelyeknél a fokszám a legnagyobb mértékben elmarad az eredetitől.
4. **Élek hozzáadása:** A csúcsokhoz új éleket keresünk, és hozzáadjuk őket a gráfhoz, így a csúcsok fokszáma fokozatosan eléri az előírt értékeket. Ez iteratív módon történik, amíg minden csúcs fokszáma el nem éri az eredeti célfokszámot.

# Megjelenítés

A green and blue circles

Description automatically generated

## Spring vs Kamada-Kawai

**Spring**

* **Erő-alapú elrendezés:** A Spring algoritmusban a csomópontokat olyan erők tartják össze, amelyek a gráf éleihez hasonlóan **rugóként viselkednek**. A csomópontok között **vonzó erők** hatnak, míg az élek között **taszító erők** lépnek fel.
* **Iteratív optimalizálás:** Az algoritmus iteratív módon dolgozik, és fokozatosan javítja a csomópontok elhelyezkedését az erők egyensúlyozásával. Az elrendezés folyamata fokozatosan finomodik, míg a gráf struktúrája egyre jobban tükrözi a kívánt elrendezést.
* **Alacsonyabb számítási költség:** A Spring algoritmus általában alacsonyabb számítási költséggel rendelkezik, mivel egyszerűbb erő-alapú számításokat alkalmaz a csomópontok elhelyezésére.
* **Nagy és bonyolult grafikonok:** Különösen jól alkalmazható **nagy és bonyolult gráfok** esetén, mivel az algoritmus képes kezelni a gráf komplexitását és nagy számú csomópontot.
* **Zsufoltság és átfedések:** Mivel iteratív módszert használ, előfordulhatnak zsúfolt területek, ahol a csomópontok túl közel kerülnek egymáshoz, különösen kisebb csomópontok között.

**Kamada-Kawai**

* **Ideális távolságok:** A Kamada-Kawai algoritmus az egyes csomópontok közötti **ideális távolságokat** próbálja optimalizálni a gráf topológiájának megfelelően. Az élek **a rövidebb utak hosszát** figyelembe véve vannak optimalizálva, így a csomópontok közötti távolságok a valós gráf struktúrához igazodnak.
* **Energia minimalizálás:** Az algoritmus az **energiafüggvényt** minimalizálja, amely a csomópontok közötti távolságok eltéréseit méri. Az energia minimalizálása segít a gráf pontosabb és esztétikusabb elrendezésében, de magasabb számítási költséggel jár.
* **Magasabb számítási költség:** A Kamada-Kawai algoritmus hajlamos magasabb számítási költségre, mivel részletesebb energiaoptimalizálást végez, és pontosabb elrendezést eredményez.
* **Kisebb gráfok esetén ideális:** A Kamada-Kawai elrendezés a **kisebb gráfok** esetében ideális, ahol a pontos topológiai távolságok vizualizációja kritikus, például **társadalmi hálózatokban** vagy **biológiai hálózatokban**.
* **Valós kapcsolati távolságok:** Az algoritmus a csomópontok közötti valódi kapcsolati távolságokat próbálja tükrözni, így **esztétikusabb és informatívabb vizualizációt** adhat, mivel az élek valódi kapcsolatokat tükröznek.

# Megfigyelések

## Klusterzési Koefficiens Eloszlása

A graph of a graph with a red and blue bar

Description automatically generated A graph of a graph

Description automatically generated with medium confidence

A grafikonon jól látható, hogy a **Holme-Kim modell** jelentősen magasabb klaszterezési koefficienst mutat, míg a **Configurációs modell** alacsonyabb értéket ad, tükrözve a két modell közötti különbséget a lokális közösségek kialakulásában.

**Holme-Kim Modell:**

* Az eloszlás széles tartományban oszlik el, és magasabb klaszterezési együtthatók is előfordulnak.
* A **Holme-Kim modell** esetén több csomópont rendelkezik magas lokális klaszterezési együtthatóval, ami azt jelenti, hogy a szomszédok közötti kapcsolatok sűrűbbek.
* A hálózat lokális közösségeket (klikkeket) hoz létre, ahol a csomópontok szorosabban kapcsolódnak egymáshoz.

**Configurációs Modell:**

* A klaszterezési együtthatók nagy része 0.0 körüli értékekkel rendelkezik, ami azt jelenti, hogy a csomópontok többségénél nincs jelentős klasztereződés.
* A **Configurációs modell** esetében a gráf több csomópontja rendelkezik alacsony lokális klaszterezési együtthatóval, és a szomszédok közötti kapcsolatok nem olyan sűrűek, mint a Holme-Kim modellben.
* Csak néhány csomópont rendelkezik nem nulla klaszterezési együtthatóval, és ezek viszonylag alacsonyak (max. 0.2 körül).

## Skálamentesség

A graph of a power-law fit comparison

Description automatically generated

A grafikonon látható a **Holme-Kim adatainak** és a hozzájuk illesztett **power-law görbe**. Az illesztés meredeksége körülbelül -1.55, amely az eloszlás jellegzetes szóródását mutatja.  
A következő dián a **skálamentesség** (scale-freeness) jellemzőjét és annak hatványfüggvényes eloszlását vizsgáljuk. A skálamentesség azt jelenti, hogy a hálózaton belüli csomópontok fokszám-eloszlása (degree distribution) hatványfüggvényt követ, azaz a csomópontok többsége alacsony fokszámmal rendelkezik, míg egy kisebb részük kiemelkedően magas fokszámú.

## Klasterezési Együttható vs Fokszám

A graph of a graph with red and blue dots

Description automatically generated

A **klaszterezési együttható** és a **fokszám** közötti kapcsolatot vizsgáljuk. Ez segít megérteni, hogyan függ a csomópontok közötti kapcsolatok sűrűsége a csomópont fokszámától, és milyen különbségek figyelhetők meg a két modell között.

A grafikonon a **Holme-Kim adatainak** eloszlása egy szoros összefüggést mutat, ahol a magas fokszámú csomópontok klaszterezési együtthatója gyorsan csökken.

A **Configurációs modell** esetén, ahogy a fokszám növekszik, a csomópontok klaszterezési együtthatója gyorsan **0.0 körüli értékhez közelít**.

## Aszortivitás

A graph with a red and blue rectangle

Description automatically generated

A **Holme-Kim modell** negatív aszortivitást mutat, ami azt jelenti, hogy a magas fokszámú csúcsok inkább **alacsony fokszámú csúcsokkal** kapcsolódnak.

A **Configurációs modell** szintén negatív aszortivitást mutat, de az érték alacsonyabb, ami azt jelzi, hogy a modell **kevésbé erősen** hajlik a negatív aszortivitásra.

## Átlagos Szomszéd Fokszám (KNN) vs Fokszám

A graph showing a number of different colored lines

Description automatically generated with medium confidence

Az **átlagos szomszéd fokszám** (Average Neighbor Degree, KNN) és a **fokszám** közötti kapcsolatot vizsgáljuk. Ez segít megérteni, hogyan befolyásolja egy csomópont fokszáma a szomszédai fokszámait, és ezáltal a hálózat **asszortatív** vagy **disszortatív** viselkedését.

A grafikonon látható, hogy a **nagy fokszámú csomópontok** esetében a **KNN érték** alacsonyabb, míg a **kis fokszámú csomópontok** szomszédai jellemzően nagyobb fokszámúak.

A **Configurációs modell** esetében a csomópontok szomszédainak fokszáma **kevésbé mutat erős korrelációt** a csomópont fokszámával.

# Életbeli Alkalmazásai

## 1. Holme-Kim Modell

**a) Szociális Hálózatok**

A Holme-Kim modell a szociális hálózatokban alkalmazott preferenciális kötődés szabályokkal és a csomópontok közötti kiegészítő kapcsolatokkal jól modellezi a valódi interakciókat.

* A **Facebook** vagy más szociális média platformok példáján látható, hogy az új felhasználók valószínűleg azokhoz a már meglévő felhasználókhoz csatlakoznak, akiknek nagyobb számú barátjuk van, így gyorsan nő a hálózatuk. Ezen kívül az emberek hajlamosak olyan ismerősökkel kapcsolódni, akiknek közvetlen kapcsolataik vannak az ő baráti körükkel, ami tovább növeli a hálózat összetettségét.

**b) Infektivitás terjedése**

A Holme-Kim modellt alkalmazzák járványok és fertőző betegségek terjedésének modellezésére, mivel a fertőzés nemcsak véletlenszerű kapcsolatok révén terjedhet, hanem preferenciális kapcsolatokon keresztül is, ahol a nagyobb közösségekkel rendelkező személyek gyorsabban terjeszthetik a betegséget.

* A **COVID-19** terjedésének modellezésekor figyelembe kell venni, hogy a fertőzött személyek közvetlen kapcsolatai (pl. családtagok, barátok) nagyobb valószínűséggel fertőzhetik meg másokat, miközben a közvetett kapcsolatokat is figyelembe kell venni. A Holme-Kim modell segíthet ezen kapcsolatok dinamizmusának feltérképezésében.

**c) Hálózati Képzés és Képességfejlesztés**

Az oktatási hálózatok, amelyek különböző intézményeket és egyéneket kapcsolnak össze, szintén alkalmazhatják a Holme-Kim modellt. Az egyes tanulók (csomópontok) preferenciálisan választanak mentorokat és tanárokat (képzett egyéneket), akik szélesebb kapcsolati hálóval rendelkeznek.

## 2. Configurációs Modell

**a) Szociális Hálózatok (Fokszám-eloszlás modellezés)**

A **Configurációs modell** gyakran alkalmazható szociális hálózatokban, ahol a csomópontok fokszáma (azaz, hogy hány kapcsolatuk van) előre meghatározott eloszlás alapján alakul. A hálózatok fokszám-eloszlása jellemzően **hatalmas fokú varianciát** mutat, ahol a legtöbb csomópontnak csak néhány kapcsolata van, míg néhány csomópontnak sok kapcsolata van.

* A **Twitter** vagy **Instagram** példáján látható, hogy bár a felhasználók többségének viszonylag kevés követője van, néhány influencer vagy híresség hatalmas követői bázissal rendelkezik. Ezen hatások modellezésére a Configurációs modell hasznos, mivel segít megérteni, hogyan alakul ki a hálózatban az egyenlőtlen eloszlás a csomópontok fokszámában.

**b) Internet és Weboldalak Hálózatai**

Az internet és különböző weboldalak közötti kapcsolatok is jól modellezhetők a konfigurációs modell segítségével, különösen, ha előre meghatározott fokszám-eloszlással dolgozunk. Az internetes forgalom és a weboldalak közötti linkek gyakran követik a **power-law eloszlást**, amely a Configurációs modell egyik jellemzője.

* A különböző weboldalak közötti linkek eloszlása gyakran nem egyenletes, néhány weboldal rendkívül sok linket kap (pl. nagy népszerűségű weboldalak), míg a legtöbb kisebb weboldal kevesebb kapcsolattal rendelkezik. A konfigurációs modell segíthet a webes struktúrák, linkek és weboldalak közötti kapcsolatok modellezésében.

**c) Biológiai és Ökológiai Hálózatok**

A biológiai és ökológiai hálózatokban a csomópontok (pl. különböző fajok, állatok) fokszámát előre meghatározott eloszlás szerint lehet modellezni, és ezen eloszlások alapján lehet tanulmányozni az ökoszisztémák működését és a fajok közötti interakciókat.

* A **fajok közötti kapcsolatok**, ahol egyes fajok sok más fajjal, míg mások csak kevés fajjal interakcióba lépnek, jól modellezhetők a Configurációs modell segítségével. Az ilyen típusú hálózatok segíthetnek a biológiai sokféleség és a fajok közötti kölcsönhatások jobb megértésében.

**d) Telekommunikációs és Távközlési Hálózatok**

A telekommunikációs rendszerek, mint például mobiltelefon-hálózatok, vezeték nélküli hálózatok és más távközlési infrastruktúrák is alkalmazhatják a Configurációs modellt. A különböző állomások vagy cellák kapcsolatait a fokszám-eloszlás alapján modellezhetjük.

* A mobilhálózatokban a bázisállomások közötti kapcsolatok fokozatosan alakulnak ki, és a kapcsolatokat előre meghatározott szabályok alapján hozhatják létre a különböző csomópontok (pl. bázisállomások), figyelembe véve azok terhelését és kapacitását.