

# **TDK dolgozat**

Leitold Dániel, Pigler Csaba

# **TÚLÉLÉS-ELEMZÉS HÁLÓZATOKBAN – ÚJ ESZKÖZ A KOMPLEX RENDSZEREK DINAMIKÁJÁNAK FELTÁRÁSÁRA**

Szerzők: Leitold Dániel<sup>1</sup>, Pigler Csaba<sup>2</sup>

<sup>1-2</sup>mérnökinformatikus, MSc

Témavezetők: Dr. Fogarassyné Dr. Vathy Ágnes, adjunktus

Dr. Abonyi János, egyetemi tanár

Pannon Egyetem

Műszaki Informatikai Kar

Rendszer- és Számítástudományi Tanszék

## Tartalomjegyzék

|                                                                  |    |
|------------------------------------------------------------------|----|
| Kivonat .....                                                    | 2  |
| 1 Bevezetés .....                                                | 3  |
| 2 Elméleti háttér .....                                          | 5  |
| 2.1 A túlélés analízis elméleti háttere .....                    | 5  |
| 2.1.1 A Kaplan-Meier túlélési modell .....                       | 6  |
| 2.1.2 A hazard ráta .....                                        | 8  |
| 2.1.3 Log-Rank teszt .....                                       | 9  |
| 2.1.4 Cox regresszió .....                                       | 10 |
| 2.2 Hálózatelemzés .....                                         | 11 |
| 2.2.1 Gráfelméleti alapok .....                                  | 11 |
| 2.2.2 Kapcsolati hálók és elemzésük .....                        | 13 |
| 3 Kidolgozott megközelítésmód és alkalmazásának eredményei ..... | 16 |
| 3.1 Célkitűzés .....                                             | 16 |
| 3.2 Hálózat felépítése .....                                     | 16 |
| 3.3 Szomszédok hatásának vizsgálata .....                        | 17 |
| 3.4 Cégek túlélés analízise .....                                | 18 |
| 3.5 Bedőlési folyamatok vizsgálata .....                         | 28 |
| 4 Összefoglalás .....                                            | 31 |
| Jelölésjegyzék .....                                             | 32 |
| Irodalomjegyzék .....                                            | 33 |

## Kivonat

A számítástudomány egyik fontos hajtóereje a komplex rendszerek dinamikájának tanulmányozására alkalmas eszközök iránti társadalmi és gazdasági igény. A túlélés-elemzést először életbiztosítási statisztikusok alkalmazták az 1950-es években. E folyamatosan fejlődő tudományterület olyan eszköztárat eredményezett, amely segítségével biológiai, műszaki, társadalmi és gazdasági rendszerek jól definiált eseménysorainak bekövetkezési valószínűségének meghatározására, és ezen keresztül a rendszer élettartamának becslésére alkalmas. A megközelítésmód fontos hátránya, hogy egy statisztikailag homogén populáció egyedeit vizsgálja, nem veszi figyelembe azt a fontos tényt, hogy a társadalom, a műszaki rendszerek, és a gazdasági élet szereplői is komplex hálózatokat alkotnak, és e hálózatok struktúrája jelentős hatással van az egyedek viselkedésére.

E felismerés alapján a komplex rendszerek vizsgálatára egy új, a túlélés-elemzés és hálózatelemzés eszköztárának integrálásán alapuló megközelítésmódot javasolunk. A koncepció alkalmazása nem triviális feladat, ugyanis a hálózatelemzés szemünk előtt zajló robbanásszerű fejlődése ellenére az összetett kapcsolatrendszerekből a különböző ok-okozati összefüggések ma még nehezen tárhatók fel, a komplex függőségekkel rendelkező hálózatok dinamikájának identifikációjára alkalmas eszközök még nem állnak rendelkezésre.

Kutatásunk alapgondolata, hogy a statikus hálózatanalízis (fokszám elemzés, centrális elemek és független komponensek keresése) révén kapott eredmények alapján feltételes valószínűségek számításán keresztül túlélési modellek készíthetők, és 'számolási folyamat elmélet' alkalmazásával a csomópontokon értelmezett több egyidejű, sorrendi esemény vizsgálatát lehetővé tevő dinamikus modell identifikálható.

A javasolt megközelítésmóddal gazdasági szervezetek csődjének korai előrejelzésére alkalmas rendszert fejlesztettünk. Dolgozatunkban a javasolt ketrrendszer elméleti alapjainak ismertetésén túl részletesen bemutatjuk az összes magyar gazdasági szereplőt átfogó Dun & Bradstreet adathalmazon elvégezett hálózati kapcsolatokat figyelembe vevő csődvalószínűség előrebecslési rendszer megalkotásához szükséges vizsgálatokat, és azokat az eredményeket, melyeket kizárólag az általunk kidolgozott új eseményanalízist és a hálózatelemzést integráló megközelítéssel érthettünk el.

## 1 Bevezetés

Felgyorsult világunkban a komplex rendszerek dinamikájának feltárása fontos és alapvető feladatnak számít mind gazdasági, mind műszaki, mind társadalmi szempontból. A rendszerek változásának előrejelzésében hatékony segítséget nyújt számunkra a túlélés analízis matematikai eszköztára, amely a statisztika egy speciális részterületének tekinthető [1]. A túlélés analízis elsődleges célja a túlélési görbe meghatározása, mellyel műszaki rendszerek élettartamára, biológiai fertőzések, betegségek lefolyására, organizmusok életképességére, gazdasági események bekövetkezésére vonatkozóan adhatunk predikciós leírásokat [2]. Bár a túlélés analízist, mint előrejelző módszert életbiztosítási statisztikusok alkalmazták először az 1950-es években, a folyamatos fejlődésnek köszönhetően mára már olyan eszköztárral rendelkezik, melynek segítségével komplex rendszerek jól definiált eseménysorainak bekövetkezési valószínűségét tudjuk meghatározni tetszőleges tudományterületen [1]. A túlélés analízis alapvetően egymástól jól elkülöníthető objektumokra vonatkozóan fogalmaz meg előrejelzéseket, mely objektumok lehetnek például műszaki berendezések, személyek, gazdasági társulások. Ezen megközelítés fő hátránya, hogy nem veszi figyelembe az objektumok közötti kialakuló komplex kapcsolatokat, s ilyen módon nem számol az objektumok egymásra gyakorolt hatásával. Márpedig a napjainkban is zajló tudományos forradalom egyik legfontosabb üzenete az, hogy a valós világ objektumai egymással kapcsolatban állnak, egy individuális objektum működése nem izolálható el a környezetétől, azzal szoros kapcsolatban van. Ezen szoros kapcsolat által az önállóan is működőképes egységek egymásra hatással vannak, egymás működését akár jelentős mértékben is befolyásolni képesek.

A hálózattudomány a matematikából ismert gráfelméleten alapul, mivel az objektumok és a köztük lévő kapcsolatrendszer gráfokkal modellezhető [3]. A hálózatok elemzése betekintést nyújt az objektumok kapcsolatrendszerébe, általa feltárhatók az elemek közti csoportosulások, azonosíthatók a hálózatok meghatározott szerepben lévő elemei, elemezhetők a hálózatokban zajló folyamatok [4]. A hálózatok elemzése alapulhat statikus gráfjellemző tulajdonságokon [5], illetve tekinthetünk rájuk úgy is, mint dinamikus változó modellekre [5]. A statikus elemzések alapját a statikus gráfokra jellemző tulajdonságok adják, mint például a fokszám, átlagos gráfátmérő, csúcsok centralitása. Amennyiben figyelembe vesszük a hálózatok dinamikus változását is, akkor vizsgálhatjuk a gráfok strukturális változásának jellegét, valamint a hálózatokban végbemenő folyamatokat [6]. Mind a statikus, mind a dinamikus megközelítésmód implicit módon magában foglal olyan eszközöket, amelyek a hálózatokra, illetve azok csomópontjaira vonatkozóan adnak „túlélési” információt. A hálózatok léte, felbomlása ugyanis nagyban függ azok struktúrájától. Amennyiben ugyanis a hálózatban megjelennek központi szerepet betöltő centrális csúcsok, akkor ezen csúcsok támadása, kivétele a hálózattól a hálózat egészének működésképtelenségét, szétesését okozza. A csúcsok centralitása, mint statikus gráfjellemző, tehát a hálózat stabilitásának egy alapköve. Másrészt viszont, a hálózatban zajló folyamatok a hálózatok csomópontjaira vetítve adhatnak prediktív leírásokat. A hálózatban terjedő fertőzések modellezése által jóslásokba bocsátkozhatunk arra vonatkozóan, hogy egy információ milyen csúcsokhoz, mikorra jut el [7]. Ezen fertőzési modellek nem csak az információ terjedésének modellezésére alkalmasak, hanem akár biológiai fertőzések terjedésének leírására, vagy hitelbedőlési láncok modellezésére is [7]. Vagyis a fertőzési modellek ezáltal szintén alkalmasak egyfajta túlélési vizsgálat végrehajtására, mellyel az egyedek, mint csomópontok túlélése modellezhető.

Dolgozatunkban egy új megközelítési módot javasolunk, melyben a hálózat analízis és túlélés elemzés eszközeinek integrálását valósítottuk meg. Az új szemlélet előnye, hogy hálózatos környezetbe ültetve a túlélés elemzés eszköztárát az egyedek közti komplex kapcsolatok is figyelembe vehetők. Szakirodalmi kutatásaink alapján azt tapasztaltuk, hogy ezen két terület integrálása még nem elterjedt, a szakirodalomban csupán néhány példát találunk az együttes alkalmazásukra. Így például Kossinets és szerzőtársai által írt cikkben [8] egy egyetem hallgatóinak, dolgozóinak e-mail forgalmából képzett hálózat vizsgálatánál jelenik meg a hazard ráta fogalma, ami a túlélés analízis egyik alapvető leíró mennyisége. Butler és társai művében [9] az elektronikus hálózatokat érintve jelenik meg a túlélés analízis, Christodoulou és szerzőtársai pedig vízhálózatokon végeztek survival elemzéseket [10]. Bár ezekben a munkákban együttesen jelenik meg a két terület, mindamellett az jellemzi őket, hogy döntően az egyik terület eszköztárát alkalmazzák, s a másik terület csupán mint kiegészítő módszer jelenik meg. Kossinets és munkatársai munkájukban döntően hálózat analízist végeztek, a túlélés analízis csak marginális szerephez jut. A másik két publikációban a túlélés vizsgálat kapja a nagyobb szerepet. Kutatásaink során mi nem csupán segédeszközként alkalmaztuk a hálózattudományt, vagy a túlélés elemzés által nyújtott eszközkészletet, hanem mindkét eszközt egyenrangúként kezelve, az előnyöket ötvözve egy új megközelítésmódot hoztunk létre, amellyel a túlélés elemzést komplex kapcsolatrendszerekbe helyeztük.

Elméleti kutatásainkat mindamellett gyakorlati környezetbe ültetve validáltuk. Ezen tevékenység folyamán gazdasági társulások felszámolás alá kerülésének előrejelzésére alkalmas rendszer alapjait fejlesztettük ki. Kutatásainkat valós adatok felhasználásával végeztük, melynek során az összes magyar gazdasági szereplőt átfogó Dun & Bradstreet országos cégadatbázist használtuk. Ezen adatbázis nem csupán a gazdasági szereplők között felépülő kapcsolatrendszerre vonatkozóan tartalmaz sokrétű információt, hanem a gazdasági társulásokra vonatkozóan tartalmazza azt is, hogy mikor kerültek felszámolás alá. Ezen adatok felhasználása révén az új módszertan segítségével hatékonyan tudjuk előrejelezni a működő cégek csődbemenetelének valószínűségét amely eszköz hasznos lehet például a cégek számára hitelt nyújtó pénzintézetek számára.

A továbbiakban dolgozatunk második fejezetében ismertetjük a túlélés elemzés és hálózat analízis elméleti hátterét. A harmadik fejezetben a gazdasági szereplők által képzett hálózatokon elvégzett vizsgálatainkat, továbbá az ezekből kapott eredményeket mutatjuk be. Az eredmények bemutatásával párhuzamosan ismertetjük az eredményekből származó egyértelmű állításokat, következtetéseket.

## 2 Elméleti háttér

### 2.1 A túlélés analízis elméleti háttére

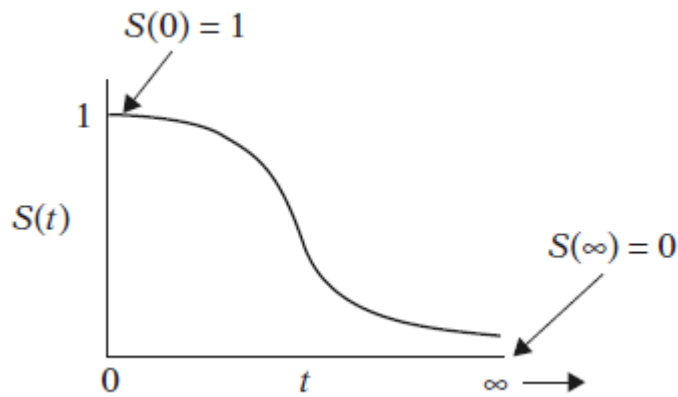
A túlélés analízis [11] statisztikai eszközök gyűjteménye, melyek egy megfigyelési időszak alatt bekövetkező események elemzésére használhatóak. Megfigyelési időszak alatt olyan előre definiált időintervallumot értünk, mely alatt egyedekhez köthető események bekövetkezését vizsgáljuk. Eseményként lényegében bármilyen megfigyelés eredményét tekinthetjük, maga a vizsgálandó probléma írja le azt, lehet például betegek vizsgálata alatt bekövetkező halál, vagy műszaki berendezések esetén egy esetlegesen fellépő meghibásodás. Utóbbi miatt mérnöki területeken a túlélés analízist gyakran megbízhatósági analízisnek is nevezik.

A túlélés analízis népszerű megközelítése különböző, főleg egészségügyi adatok vizsgálatának. Az első túlélés analízisre alkalmas modellt Boag fejlesztette ki 1949-ben, és Cure rate modell néven vált ismertté, majd ezt 1952-ben Berkson és Gage fejlesztették tovább [12]. Munkájuk végül a standard cure rate néven híresült el. Később több túlélési modell is született, melyek közül a két legismertebb a Kaplan-Meier (1958) [13] és a Cox (1972) [14] modellek. A megfigyelési időszak alatt a megfigyelt egyedekhez egy túlélési időt rendelhetünk, mely annak túlélési periódusát adja meg. A legtöbb túlélés elemzés figyelembe vesz egy kulcsfontosságú problémát, mégpedig a cenzorálás problémáját. Ez akkor lép fel, ha a megfigyelési időszak alatt egyes egyedekről hiányos információkkal rendelkezünk, és nincs ismeretünk a vizsgált esemény bekövetkezéséről. Tipikus példája ennek, mikor egyének elhalálozását tekintjük végpont eseménynek, de az egyének kikerülnek a vizsgálat látóköréből (pl. elköltöznék), s ezáltal egy adott időpont után nem rendelkezünk információval a végpont bekövetkezésére vonatkozóan. Ilyen esetekben jobbra-cenzorálásról beszélünk, és a legtöbb túlélési analízist alkalmazó feladatnál is ilyen jellegű adatokkal dolgozunk. Ellenkező esetben, ha arról van információnk, hogy egy egyén élettartama kevesebb, mint egy bizonyos ismert időtartam, akkor balra-cenzorálásról beszélhetünk. A cenzorálással kapcsolatban vezethetjük be a  $\delta=(0,1)$  változót, mely az esemény bekövetkezését vagy a cenzorálást hivatott jelölni.  $\delta=1$  esetén az esemény bekövetkezése,  $\delta=0$  esetén pedig a cenzorálás történik meg. Fontos megjegyezni, hogy azok az egyedek, melyeknél a megfigyelt esemény és cenzorálás sem következik be a megfigyelési időszak alatt, cenzorálással fognak rendelkezni a megfigyelés lezárása után.

A megfigyelés során az egyedekhez köthető esemény bekövetkezési idejének jelölésére a  $T \geq 0$  valószínűségi változót használjuk. Ekkor definiálhatunk egy túlélési függvényt, melyet az 1. egyenlet szemléltet.

$$S(t) = P(T > t) \quad (1)$$

Ez a függvény annak a valószínűségét adja meg, hogy a vizsgált esemény  $t$  idő után következik be, vagyis annak az esélyét, hogy az esemény bekövetkezésének időpontját jelző  $T$  valószínűségi változó értéke nagyobb, mint a megadott  $t$  idő. A túlélési függvény meghatározása alapvető eszköze a túlélés analízisnek. Az elemzések egy általános eszköze a különböző túlélési függvények reprezentálása túlélési görbékkel, melyek a túlélés valószínűségét ábrázolják az eltelt idő függvényében. Egy ilyen túlélési görbe látható az 1. ábraán.

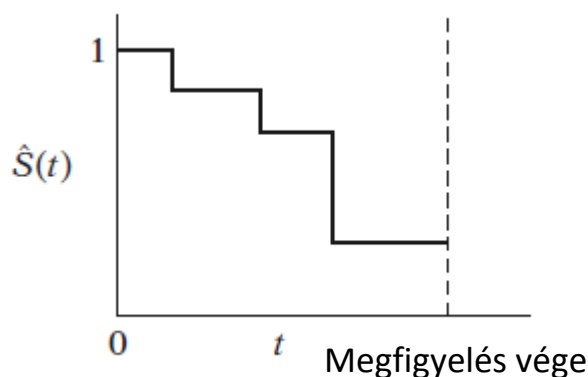


1. ábra Példa  $S(t)$  túlélési görbére

A grafikon abszcisszáján az eltelt idő ( $t$ ), ordináta tengelyén pedig a túlélés valószínűsége ( $S(t)$ ) jelenik meg. A túlélési görbékre a következő tulajdonságok jellemzőek:

- a függvények értéke  $t$  idő növekedtével csökken,
- $t=0$  időpontban  $S(t)=1$ , mivel kezdetben minden megfigyelt egyén „életben van”,
- $\lim_{t \rightarrow \infty} S(t) = 0$  mivel ha a megfigyelési periódust egyre hosszabbnak választjuk, akkor egyre kevesebb lesz az esélye annak, hogy a megfigyelt egyének túléljék a vizsgált időszakot.

A túlélési függvények jellemzője, hogy egy meghatározott sűrűségfüggvénnyel megadhatók. A gyakorlatban azonban sokkal inkább az előzetes megfigyelésekből származó adatok alapján meghatározott empirikus túlélési függvények az elterjedtek. Ezen függvénygörbék jellemzője a lépcsőzetes jelleg, mint ahogy azt a 2. ábra is mutatja.



2. ábra Empirikus túlélési görbe

### 2.1.1 A Kaplan-Meier túlélési modell

A vizsgált csoportok túlélési görbéinek megállapítását kétféle módon lehet megközelíteni. Az egyik, általános módszer egy úgynevezett life-time táblázat használatán alapszik. A táblázat bejegyzései a teljes követési időt egységes időszakaszokra osztottan megadják a túlélési időkre vonatkozó információkat, a  $\delta$

## 2. Elméleti háttér

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására

cenzorálási változót, és szükség esetén egyéb, értelmezéshez szükséges információkat ( $X_i$ ). Egy ilyen táblázat felépítését mutatja az 1. táblázat.

1. táblázat Life-time táblázat felépítése

| Megfigyelt<br>egyén | $t$   | $\delta$   | $X_1$    | ... | $X_p$    |
|---------------------|-------|------------|----------|-----|----------|
| 1                   | $t_1$ | $\delta_1$ | $X_{11}$ | ... | $X_{1p}$ |
| .                   | .     | .          | .        | .   | .        |
| .                   | .     | .          | .        | .   | .        |
| .                   | .     | .          | .        | .   | .        |
| $n$                 | $t_n$ | $\delta_n$ | $X_{n1}$ | ... | $X_{np}$ |

A túlélési grafikonok meghatározásának másik lehetséges eszköze a Kaplan-Meier módszer, mely Edward L. Kaplan és Paul Meier nevéhez fűződik [13]. A módszer abban tér el a life-time táblázatos megközelítéstől, hogy egy alternatív adatkezelési módszert biztosít a vizsgálatok során kapott túlélési eredményekhez, és nem követeli meg az egységes időszakaszokra osztott vizsgálatot sem. A módszer alkalmazásához a következő értékek meghatározása szükséges:

- $t_j$  – megfigyelési időpontok, növekvő sorrendben
- $n_j$  – vizsgált csoport mérete a  $j$ -edik időpontban
- $m_j$  – „halálozások” száma  $t_j$  és  $t_{j+1}$  időpontok között
- $q_j$  – cenzorált egyének száma  $t_j$  és  $t_{j+1}$  időpontok között
- $R(t_j)$  – veszélyeztetett egyének halmaza, melyek  $t_j$  időpontig túléltek.

Ezen információk meghatározása után a Kaplan-Meier módszer segítségével megállapíthatóak a túlélési valószínűségek és a grafikonok meghatározásához szükséges adatok. A  $t_j$  időpillanathoz tartozó túlélési valószínűség kiszámítási módját a 2. egyenlet szemlélteti.

$$\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)}) * \hat{P}(t > t_{(j)} | T \geq t_{(j)}) \quad (2)$$

A  $t_j$  időpontbeli túlélési valószínűség tehát megkapható a  $t_{j-1}$  időpontbeli túlélési valószínűség és annak a valószínűségnek a szorzatából, mely azt adja meg, hogy mennyi a túlélési valószínűsége  $t_j$  időpillanat után a vizsgált egyénnek, feltéve ha  $t_j$  időpontig túlélte.  $\hat{S}(t_{(j)})$  a 3. egyenleten látható formula szerint adható meg.

$$\hat{S}(t_{(j)}) = \prod_{i=1}^j \hat{P}(t > t_{(i)} | T \geq t_{(i)}) \quad (3)$$

A 2. táblázatban a túlélési valószínűségek meghatározására alkalmas Kaplan-Meier módszer alkalmazására láthatunk példát.

2. táblázat Példa Kaplan-Meier módszer alkalmazásával meghatározott túlélési valószínűségekre

| $t_j$ | $n_j$ | $m_j$ | $q_j$ | $\hat{S}(t_j)$               |
|-------|-------|-------|-------|------------------------------|
| 0     | 21    | 0     | 0     | 1                            |
| 6     | 21    | 3     | 1     | $1 * \frac{18}{21} = 0.8571$ |

## 2. Elméleti háttér

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására

|    |    |   |   |                                            |
|----|----|---|---|--------------------------------------------|
| 7  | 17 | 1 | 1 | $0.8571 * \frac{16}{17} = \mathbf{0.8067}$ |
| 10 | 15 | 1 | 2 | $0.8067 * \frac{14}{15} = \mathbf{0.7529}$ |
| 13 | 12 | 1 | 0 | $0.7529 * \frac{11}{12} = \mathbf{0.6902}$ |
| 16 | 11 | 1 | 3 | $0.6902 * \frac{10}{11} = \mathbf{0.6275}$ |
| 22 | 7  | 1 | 0 | $0.6275 * \frac{6}{7} = \mathbf{0.5378}$   |
| 23 | 6  | 1 | 5 | $0.5378 * \frac{5}{6} = \mathbf{0.4482}$   |

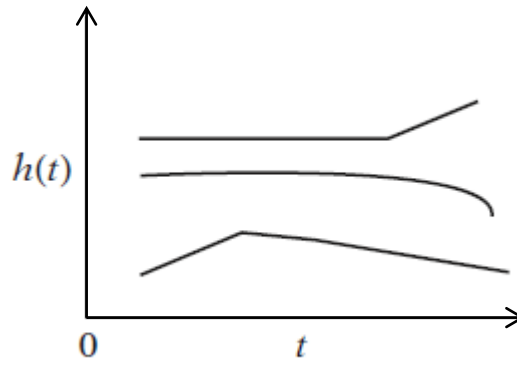
A Kaplan-Meier modell sikeresen alkalmazható például orvosi vizsgálatok során a különböző megfigyelési csoportokba sorolt egyedek kezelés utáni várható élettartamának becslésére.

### 2.1.2 A hazard ráta

A túlélési függvény mellett gyakran használt másik eszköz túlélés elemzés során a hazard függvény, mellyel egy adott időpillanatbeli kockázat határozható meg. Egy  $h(t)$  hazard függvény a 4. egyenlet alapján adható meg.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (4)$$

A hazard függvény tehát egy potenciális választ ad arra, hogy mekkora eséllyel következik be egy esemény  $\Delta t$  idő alatt ha tudjuk, hogy  $t$  időpillanatig az nem következett be. A feltételes valószínűség alkalmazásából adódóan a hazard függvényt szokás még hazard rátának is nevezni. Mivel a határértékben szereplő szám már lényegében nem egy valószínűség, hanem egy ráta, és az értéke sem 0 és 1 között található, hanem 0 és végtelen között, nagysága a megválasztott időintervallum hosszától függ. Látható, hogy a túlélési függvénnyel ellentétben a hazard függvény nem a túlélésre keres valószínűséget, hanem az esemény bekövetkezésére, ezáltal a két függvény egymás ellentétéként is felfogható. Különböző hazard függvények ismeretében hazard grafikonok is megállapíthatóak, a túlélési függvényekhez hasonlóan. A 3. ábra három különböző lehetséges hazard grafikon látható.



3. ábra Különböző hazard függvények

Látható, hogy a túlélési grafikonokkal ellentétben a hazard függvények grafikonjainak nem feltétlenül kell  $t=0$  pillanattól indulniuk, és a monoton csökkenés sem feltétlen jellemző rájuk.

Az  $S(t)$  túlélési és  $h(t)$  hazard függvények között szoros összefüggés áll fenn, egyik ismeretében a másik könnyen meghatározható. Kiszámításukat az 5. egyenlet és a 6. egyenlet által szemléltetett összefüggések segítségével történhet.

$$S(t) = \exp \left[ - \int_0^t h(u) du \right] \quad (5)$$

$$h(t) = - \left[ \frac{dS(t)/dt}{S(t)} \right] \quad (6)$$

Különböző megfigyelt egyének csoportjainak túlélési/hazard grafikonjainak ismeretében azok különböző viselkedése figyelhető meg a kijelölt vizsgálati időszakban. Például gyógyszer-hatékonysági vizsgálatok során két csoportot különíthetünk el, akik megkapták a vizsgált gyógyszert és akik csak placebót kaptak (kontroll csoport). Ezen csoportok hazard ráta és túlélési függvényének eltérése alkalmas arra, hogy kimutassa a gyógyszer hatását, igazolja annak hatékonyságát.

### 2.1.3 Log-Rank teszt

Annak vizsgálatára, hogy két csoport túlélési görbéje statisztikailag megegyezik-e a log-rank teszt [15] alkalmazható. A módszer lényege, hogy két vizsgált,  $n_{1j}$  és  $n_{2j}$  egyeddel rendelkező tesztcsoport mindegyikéhez megállapít egy várható „halálozási” értéket, majd azokat összeveti az eredeti értékkel. A várható értékek számítása a 7. és 8. egyenlet alapján történik.

$$e_{1j} = \left( \frac{n_{1j}}{n_{1j} + n_{2j}} \right) * (m_{1j} + m_{2j}) \quad (7)$$

$$e_{2j} = \left( \frac{n_{2j}}{n_{1j} + n_{2j}} \right) * (m_{1j} + m_{2j}) \quad (8)$$

ahol  $e_{1j}$  az első,  $e_{2j}$  a második megfigyelt csoportra jellemző várható érték a  $j$ . időpontban. Az értékek tehát megkaphatók a veszélyeztetett egyének arányának, és az

elhalálozott egyedek összegének szorzatából a  $j$ -edik időpontban. A várható értékek ismeretében meghatározható azok eltérése a tényleges „halálozási értéktől”  $m_{1j} - e_{1j}$ , és  $m_{2j} - e_{2j}$  alakban, a  $j$ -edik vizsgálati időpontban.

Az ismertett összefüggések azonban nem alkalmazhatók teljes túlélési görbék összehasonlítására, ahhoz a megfigyelési időszakok összes eredménye szükséges. A túlélési görbékre jellemző, log-rank teszthez szükséges adatok meghatározása a  $n=n_{1j}+n_{2j}$  megfigyelés esetén a 9. egyenletben látható módon történik.

$$O_i - E_i = \sum_{j=1}^n (m_{ij} - e_{ij}), \quad i = 1, 2 \quad (9)$$

ahol  $O_i$  jelöli az  $i$ -edik csoport megfigyelt „halálozásainak” számát,  $E_i$  pedig ugyanezen csoport várható halálozási számát. Különbségük ismeretében az  $X^2$  log-rank statisztikai mutató megállapítható  $|O_1 - E_1| = |O_2 - E_2|$  felhasználásával, a 10. és 11. egyenlet szerint.

$$X^2 = \frac{(O_2 - E_2)^2}{Var(O_2 - E_2)} \quad (10)$$

$$Var(O_i - E_i) = \sum_j \frac{n_{1j}n_{2j}(m_{1j}+m_{2j})(n_{1j} + n_{2j} - m_{1j} - m_{2j})}{(n_{1j} + n_{2j})^2(n_{1j} + n_{2j} - 1)}, i = 1, 2 \quad (11)$$

Természetesen a log-rank teszt alkalmazható abban az esetben is, amikor kettőnél több megfigyelt csoport túlélési görbét szeretnénk összehasonlítani [16]. Ha  $k$  számú megfigyelt csoporttal rendelkezünk, akkor  $X^2$  statisztikai mutató megadását a 12. egyenlet szemlélteti.

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (12)$$

Mindezek mellett számtalan egyéb módszer is létezik az irodalomban túlélési grafikonok összehasonlítására. Ezek mindegyike más megközelítéssel dolgozza fel az adatokat. A Wilcoxon- és Breslow teszt például súly értékeket is alkalmaz oly módon, hogy a korai események bekövetkezését nagyobb súllyal tekinti [11].

#### 2.1.4 Cox regresszió

Túlélés analízisre számtalan további eszközt említ meg az irodalom. Ezek közül a Kaplan-Meier módszer mellett a másik leggyakrabban használt eljárás a Cox regressziós modell [14], mely a túlélés analízis problémáját úgy közelíti meg, hogy a vizsgált csoport egy kockázatmentes csoporthoz hasonlítja. A modell értelmében  $h(t)$  kockázati változó  $\underline{x} = [x_1, x_2, \dots, x_n]$  magyarázó változók mellett  $\beta_i$  regressziós koefficiens segítségével meghatározható a 13. egyenlet alapján.

$$h(t, \underline{x}) = h_0(t)e^{\sum_{i=1}^n \beta_i x_i} \quad (13)$$

A modell lehetőséget nyújt annak vizsgálatára, hogy a túlélés mértéke egyénenként milyen mértékben függ a különböző kockázati tényezőktől. A  $\beta_i$  regressziós

koefficiensük ismeretében a különböző magyarázó változók  $RR$  relatív kockázata is meghatározható, a 14. egyenlet segítségével.

$$RR_i = e^{\beta_i} \quad (14)$$

A Cox-modellt gyakran alkalmazzák az exploratív vizsgálatok során hipotézisek felállítására. Ilyen lehet például a tüdőrák rizikófaktorainak vizsgálata is.

A Cox modellnek több továbbfejlesztett változata is fellelhető az irodalomban. Ezen változatokat a különböző problémák követelik meg. Ilyen például a többszörösen megjelenő események esete, melyeknél egy megfigyelt egyénél egy esemény többször is bekövetkezhet. Ilyen típusú túlélés analízisre alkalmas például a Cox PH modell [11], vagy annak egy változata a Counting Process [17] megközelítés.

A Cox modellen kívül léteznek további, paramétereizhető modellek is, mint például a lineáris regressziós, logisztikai regressziós vagy a Poisson regressziós modellek, melyek gyakran használatosak egészségügyi elemzések során. Ezen modellek lényege, hogy kimenetük valamilyen eloszlás szerint alakul, például normális, binomiális vagy Poisson eloszlás szerint. Az eloszlás ismeretlen mindaddig, míg a paraméterek a rendelkezésre álló adatok alapján nincsenek meghatározva.

## 2.2 Hálózatelemzés

A túlélés analízist követően a hálózatelemzés alapjait ismertetjük. Először a hálózatok matematikai reprezentálását mutatjuk be, majd hálózatokon végezhető elemzésekbe nyújtunk betekintést.

### 2.2.1 Gráfelméleti alapok

A hálózatok elemzésére, vizsgálatára nagyon jó eszköz áll rendelkezésre a matematikában. Ez az eszköz a gráf. A gráf csúcsokból és ezeket a csúcsokat összekötő élekből áll. Formális alakját a 15. egyenlet szemlélteti.

$$G = (V, E) \quad (15)$$

ahol  $G$ -vel a gráfot,  $V$ -vel a csúcsokat (vertices), míg  $E$ -vel az élek halmazát (edges) jelöljük. A gráfok különböző tulajdonsággal ruházhatóak fel. Ilyen tulajdonság az élek súlyozása és irányítottsága. Irányított egy gráf, amennyiben fontos az, hogy egy adott él melyik csúcs felől tart a melyik felé, vagyis egy  $A$  és  $B$  csúcspontot összekötő él esetében nem mindegy, hogy  $A$ -ból indul az él  $B$ -be, vagy fordítva,  $B$ -ből  $A$ -ba. Ekkor az éleket formálisan a 16. egyenletnek megfelelően adhatjuk meg.

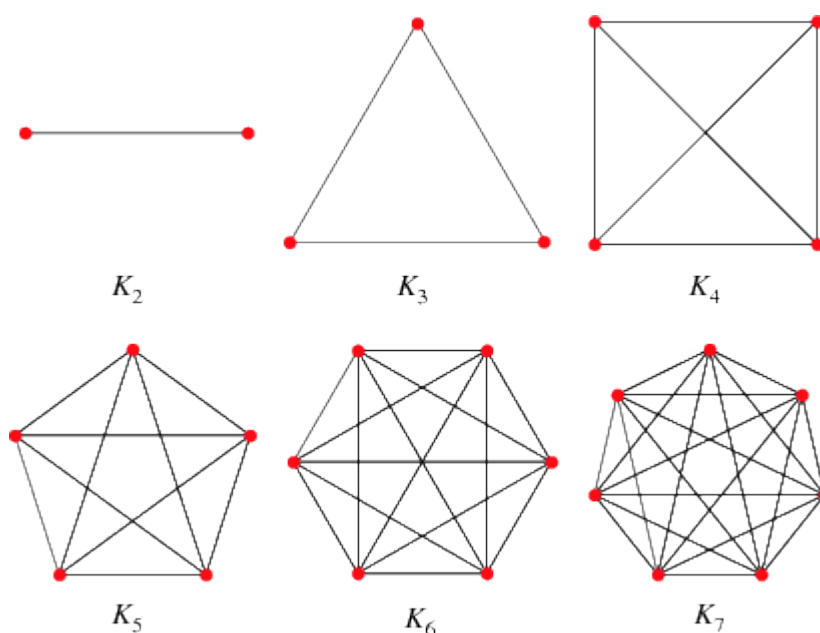
$$E \rightarrow V \times V \quad (16)$$

Amennyiben nem ismert, vagy nem lényeges az élek irányítottsága, irányítatlan gráfról beszélünk. Ez esetben az előző példában szereplő  $A$  és  $B$  csúcsokat összekötő él esetében nem beszélünk irányultságról, csupán arról, hogy van-e él a kettő csúcs között, vagy nincs. Ily módon az élek formális megadása a 17. egyenlet szerint történik.

$$E \rightarrow \{ \{x, y\} : x, y \in V \} \quad (17)$$

A gráfban szereplő élek mindegyikéhez súly is rendelhető. A súlyozott gráf definiálásakor nem elég megadni azt, hogy  $A$  és  $B$  csúcs között él húzódik, meg kell

adnunk ennek az élnek a súlyát is. Az irányított tulajdonsághoz hasonlóan a súlyozottság használata se kötelező. A súlyozást nem használó gráfokat súlyozatlan gráfnak nevezzük és ekkor minden él egyenrangúnak számít. Az irányítottság és súlyozottság megléte egy gráf esetében teljesen független, vagyis mind az irányított és irányítatlan gráfok lehetnek súlyozottak és súlyozatlanok is. A gráfokra további két elnevezés is használatos. Egyik ilyen a teljes gráf, másik pedig a részgráf. Teljes gráfnak nevezzük azokat a gráfokat, melyeknél bármely két csúcsot kiválasztva található él azok között (4. ábra). Egy  $i$  darab csúcsból álló teljes gráfot  $K_i$ -vel szokás jelölni. Egy teljes gráf csúcsainak és éleinek részhalmazából képzett gráfot nevezzük részgráfnak.



4. ábra Példa teljes gráfokra

A gráfok definiálására két eljárás terjedt el. Az első az él listás tárolás. Ez esetben egy állományban az élek az őket meghatározott csúcspárral vannak soronként felsorolva. Minden sor egyetlen élet ír le. Ha irányított gráfról van szó, akkor az adott sorban megjelenő csúcspárból az első jelenti az él kezdőpontját, míg a második az él végpontját. A másik tipikus tárolási mód az adjacencia mátrixon alapuló tárolás. Ebben az esetben egy  $N$  csúcsból álló gráfhoz egy  $N \times N$ -es mátrix tartozik. Ilyen módon a gráf formálisan a 18. egyenlet alapján adható meg.

$$G = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \quad (18)$$

A mátrixban szereplő  $a_{ij}$  értéke jelzi, hogy van-e él a gráfban az  $i$ . és  $j$ . csúcs között, illetve súlyozott gráf esetén  $a_{ij}$  az él súlyát jelöli. Irányítatlan gráfok esetében  $a_{ij} = a_{ji}$ , irányított esetben  $a_{ij}$  él  $i$ . csúcsból indul és  $j$ . csúcsban végződik.

### 2.2.2 Kapcsolati hálók és elemzésük

Kapcsolati hálóknak [18] nevezzük az olyan gráfokat, amelyek a valós életből származó adatokból épülnek fel. Ezek a hálózatok létrejöhetnek a személyek és a közöttük lévő ismertségek alapján, cégek közötti pénzmozgásoknak megfelelően, de az internet is egy hálózatot alkot a honlapok és rajtuk található linkek segítségével. Látható, hogy az élet szinte minden területén találhatunk hasonló kapcsolatrendszereket, így a példák száma végtelen.

Ahogy már említettük, a hálózatokat a matematikában jól tudjuk reprezentálni a gráfok segítségével. A hálózatokban szereplő objektumokat a gráfok csúcsai alkotják. Ezen objektumok között fennálló kapcsolatokat, interakciókat, mint például az ismeretség vagy pénzmozgás, a gráfok élei reprezentálják. A gráfoknál bemutatott tulajdonságok analóg módon érvényesek a hálózatok által alkotott gráfok esetében is. Például az irányultság tekintetében az emberek közötti ismeretségi viszony irányítatlan, mivel ha A ismeri B-t, akkor B is ismeri A-t. Azonban egy banki tranzakciónál számít az irány, és nagyon fontos is, mivel nem mindegy, hogy A utalta át a pénzt B-nek, vagy fordítva. Ezen két példával ismertethetjük a súlyozást is. Egy ismertségnek általában nincs súlya, vagy ismerünk valakit, vagy nem. A banki tranzakció viszont megint kritikus ebből a szempontból, mivel szükséges tudni, hogy az átutalás mekkora összegre vonatkozott.

A hálózatok elemzésénél a hálózat struktúrájára tudunk elsősorban támaszkodni. Ezek sokszor a gráfelméletben alkalmazott egyszerűbb tulajdonságok, úgy mint egy csúcs fokszáma, vagy két csúcs minimális távolsága. A csúcsok fokszáma alapján irányított gráf esetében megkülönböztetjük a befokot és a kifokot. A befok egy gráf adott csúcsába érkező élek száma, melynek kiszámítása a 19. egyenlet alapján történik.

$$d^+(v) = \sum_{i=1}^N a_{iv}. \quad (19)$$

Egy csúcs kifoka analóg módon a csúcshoz tartozó azon élek száma, amely a csúcsból indulnak ki. Kiszámítását a 20. egyenlet szemlélteti.

$$d^-(v) = \sum_{i=1}^N a_{vi}. \quad (20)$$

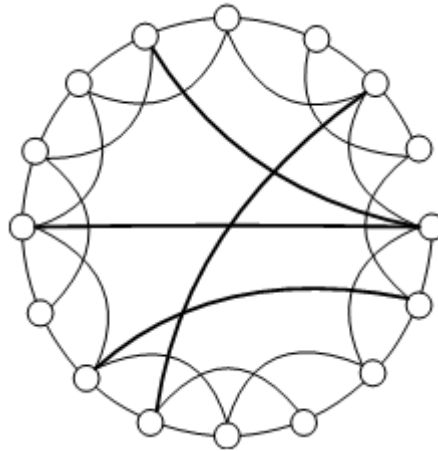
Amennyiben egy csúcs fokszámáról beszélünk, nem pedig ki és befokról, akkor a ki és befokok összegét értjük alatta (21. egyenlet).

$$d(v) = d^-(v) + d^+(v). \quad (21)$$

Ezen felül használhatunk összetettebb számolásokat igénylő értékeket is. Ilyen a sűrűség, ami megadja, hogy milyen arányban áll a létező kapcsolatok száma az összes lehetséges kapcsolatok számával. Ez az érték teljes gráfokra 1. Továbbá ilyenek a központisági (centralitás) mértékek és a közöttiségi (betweenness) mértékek [19].

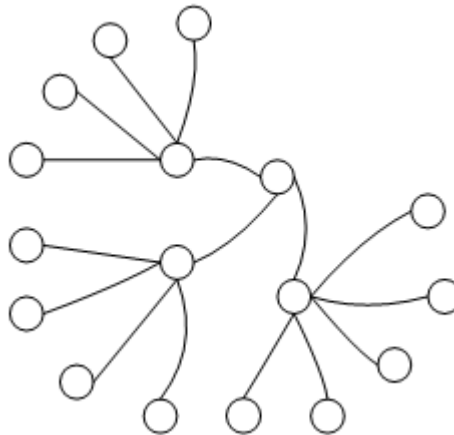
Léteznek továbbá olyan tulajdonságok is, amelyeket valós hálózatokat kutató tudósok állapítottak meg. Ezek a tulajdonságok a hálózat felépítésére vonatkoznak. Az első a kisvilág tulajdonság. Azokról a hálózatokról, amelyek ezzel a tulajdonsággal bírnak, elmondható, hogy a csúcsok fokszáma kicsi a hálózat méretéhez viszonyítva, azonban bármelyik két csúcs között található legrövidebb út hossza rövid (5. ábra).

Ezzel a tulajdonsággal rengeteg hálózat rendelkezik, többek között a szociális hálózatok, vagy a Wikipédia. A kisvilág tulajdonság létezésére vonatkozó első bizonyító erejű kísérleteket Travers és Milgram végezte el 1969-ben [20].



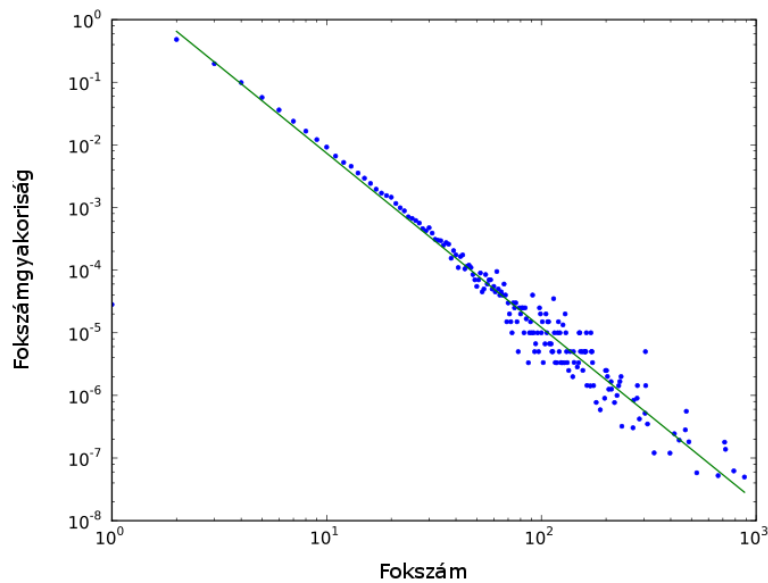
5. ábra Kisvilág tulajdonságú gráf

A másik hálózatokra vonatkozó tulajdonságot Barabási Albert-László és munkatársai nevéhez fűződik, akik a skálafüggetlen hálózatokat [21] definiálták. A skálafüggetlen hálózatok olyan hálózatok, amelyekben léteznek hubok, vagyis olyan csomópontok, amelyeknek rengeteg kapcsolatuk van, és sok olyan kis csomópont, amelyek csak néhány kapcsolattal rendelkeznek (6. ábra). Amikor egy új objektum jön létre a hálózatban, akkor valamelyik központi elemhez akar kapcsolódni, vagyis egy új objektum kapcsolódása során a központibb szerepben lévő csomópontot preferálja. A skálafüggetlen hálózatok általában rendelkeznek a kis világ tulajdonsággal is, de fordítva ez nem mondható el.



6. ábra Skálafüggetlen hálózat

További érdekesség, hogy míg az életben szereplő objektumok eloszlása általában a Gauss-görbét követi, addig a kisvilág tulajdonságú és skálafüggetlen hálózatokra a hatványgörbe a jellemző. Az egyik példa erre az eloszlásra a foksámgyakoriság. Nagy méretű skálafüggetlen kisvilág hálózatoknál jellemző, hogy a foksámeloszlás a fordítottan arányos a foksám hatványával, pontosabban köbével [22]. Az ilyen hálózatokra jellemző foksámgyakoriságot a 7. ábra szemlélteti.



7. ábra Fokszámgyakoriság log-log skálája skálafüggetlen hálózatok esetén

A hálózatok elemzésekor lényeges a különálló komponensek meghatározása is. A komponensek meghatározására több mód is van, amely függ attól, hogy a vizsgált hálózat irányított-e, vagy sem. Ha irányítatlan hálózatot vizsgálunk, akkor csak egy lehetőségünk van a komponensek keresésére, mivel az éleknek nincs iránya. Ebben az esetben azok a csúcsok tartoznak egy komponensbe, amelyekre igaz, hogy egy tetszőleges csúcsot kiválasztva abból el tudunk jutni a többi csúcs bármelyikébe. Irányított hálózat esetében beszélhetünk gyengén és erősen összefüggő komponensekről. A gyengén összefüggő komponensek esetén a hálózatot irányítatlannak tekintjük, és úgy alakítjuk ki a komponenseket, ahogy az irányítatlan hálózatok esetében. Erősen összefüggő komponensek esetén hasonló az eljárás, vagyis azok a csúcsok tartoznak egy komponensbe, amelyekből el lehet érni a többi, komponensbe tartozó csúcsot, azonban ez esetben nem elég csak az élek meglétét figyelni, annak iránya is fontos. Azon kívül, hogy a komponensek megadják az összetartozó elemeket, tovább is egyszerűsíthetjük a vizsgált problémakört, mivel egy időben elegendő lehet csupán a teljes hálózat egy részével foglalkozni. Ez roppant előnyös valós életből származó hálózatok esetén, melyek könnyen állhatnak több százezer, vagy akár még több csúcsból is.

A komponensek meghatározása mellett fontos szereppel bír a csúcsok csoportosítása is, melynek segítségével meghatározhatjuk az azonos tulajdonságokkal rendelkező, hasonló módon viselkedő, esetleg egymáshoz közel álló objektumokat. Tulajdonképpen a komponensek feltárása is megfelelhet egyfajta csoportosításnak, azonban a csoportok feltárása a hálózaton belül csupán a strukturális tulajdonságok alapján nagyon összetett feladat is lehet. Hálózat elemzés terén szintén nagy jelentőséggel bír a fertőzésvizsgálat is, mellyel a csomópontok egy jövőbeli állapota becsülhető meg. A hálózatokon való fertőzésvizsgálat is sok jelentőséggel bír, mivel az ilyen célú kutatások eredményeként a jövőre vonatkozó becsléseket tudunk megadni.

## 3 Kidolgozott megközelítésmód és alkalmazásának eredményei

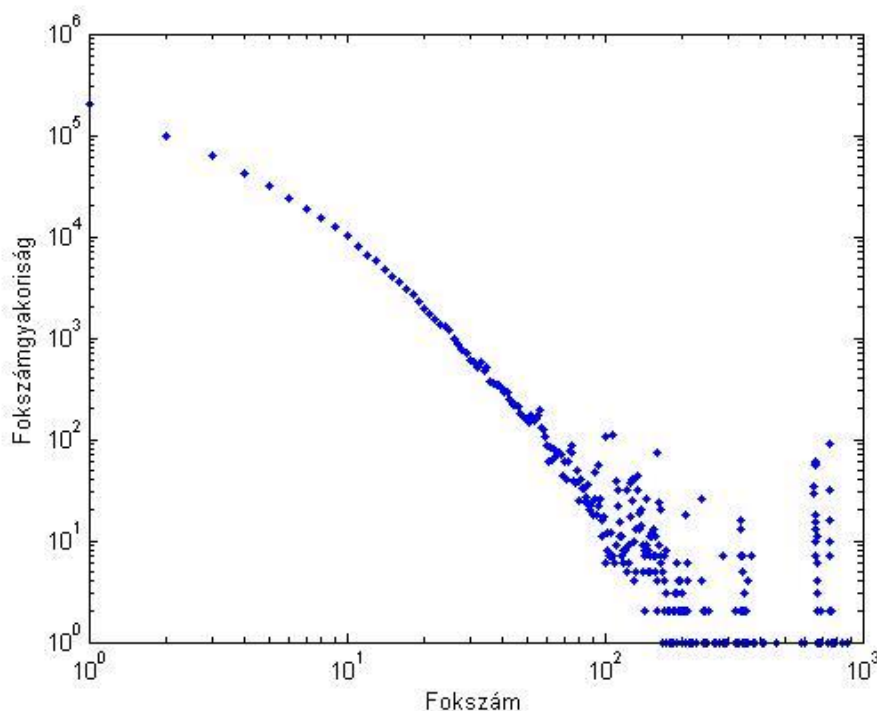
### 3.1 Célkitűzés

Munkánk során egy olyan eszköztárat alkottunk, melyben integráltuk a túlélés analízis és a hálózat elemzés által szolgáltatott technikákat. Az eszköztár megalkotását motiválták, továbbá hatékonyságát igazolják azon párhuzamosan végzett vizsgálataink, amelyek gazdasági szereplőkből történő hálózat felépítésével és azon végzett túlélés elemzéssel, vagyis a gazdasági vállalkozások felszámolás alá kerülésének valószínűségének meghatározásával történtek. Vizsgálataink során az összes magyar gazdasági szereplőt átfogó Dun & Bradstreet országos cégbázison dolgoztunk. Az adatbázis 2009 elejétől tartalmazza a cégekkel kapcsolatos információkat egészen 2013 elejéig. Az új javasolt eszköztárunk bemutatását e példán keresztül tesszük meg. A túlélés vizsgálat esetünkben az említett felszámolás alá kerülés valószínűségének előrejelzését jelenti, és az ennek az eseménynek való kitettséget vizsgálja. Munkánkban arra a kérdésre kerestük a választ a megalkotott eszköztárral, hogy a felszámolás alá kerülés mutat-e valamilyen dinamikát, van-e kapcsolat azon cégek között, amelyek esetében bekövetkezett ez az esemény. A vizsgálatok során cégek alkotta hálózatot építettünk. Ezen hálózati struktúrán végeztünk túlélés analízist, melyek során alkalmaztuk a Kaplan-Meier valamint a Cox regressziós modellt annak érdekében, hogy átfogó képet kapjunk a céghálózatban található vállalkozások sérülékenységről, illetve a bedőlési folyamatok terjedésének dinamikájáról. A vállalatokat kapcsolatszámaik szerint csoportosítva, a létrehozott cégcsoportokra külön-külön megállapított túlélési valószínűségeiket összehasonlítottuk, valamint log-rank teszt alá vetettük annak céljából, hogy a kapott eredmények által valós képet kaphassunk a partnerek számának felszámolás alá kerülés valószínűségére gyakorolt hatásáról.

### 3.2 Hálózat felépítése

Első feladatunk volt a hálózat létrehozása. A feladat nem triviális, hiszen az adatbázis rengeteg, több mint 800 ezer vállalkozás adatait tartja számon. A hálózat felépítése során a cégek közötti közvetlen fennálló kapcsolatokon kívül közvetett kapcsolatokat is elemeztünk, így például a személyeken keresztüli megvalósuló közvetett kapcsolatokat is. Az adatbázis közel 1,4 millió személyről tartalmaz adatokat, melyek közül közel 300 ezer olyan személyt azonosítottunk, amelyek potenciális kapcsolatokat hoztak létre cégek között. Hasonló módon a közös címre bejelentett cégeket is elemeztük. Az adatbázis közel 950 ezer címet tartalmazott, melyből a vizsgálatokhoz szükséges 100 ezer címet identifikáltunk. A hálózat elkészítése után néhány vizsgálat elvégzésével arra az eredményre jutottunk, hogy a hálózatot elegendő oly módon vizsgálni, hogy a cégek között fennálló kapcsolatok közül csupán a tulajdonosi kapcsolatokat vesszük figyelembe. Ezen kapcsolatok létrejöhetnek oly módon, hogy a két cég között közvetlen tulajdonosi kapcsolat van, de akár úgy is, hogy egy személy két különböző cégben is tulajdonosi jelzővel ellátott pozíciót töltött be.

A létrehozott hálózatban a fokszámgyakoriságot vizsgálva a 8. ábraán látható eredményeket kaptuk. Elmondható, hogy az hálózat fokszámgyakorisága követi a skálafüggetlen hálózatokra jellemző fokszámgyakoriságot, így belátható, hogy cégek tulajdonosi kapcsolatrendszeréből épített hálózat is skálafüggetlen.

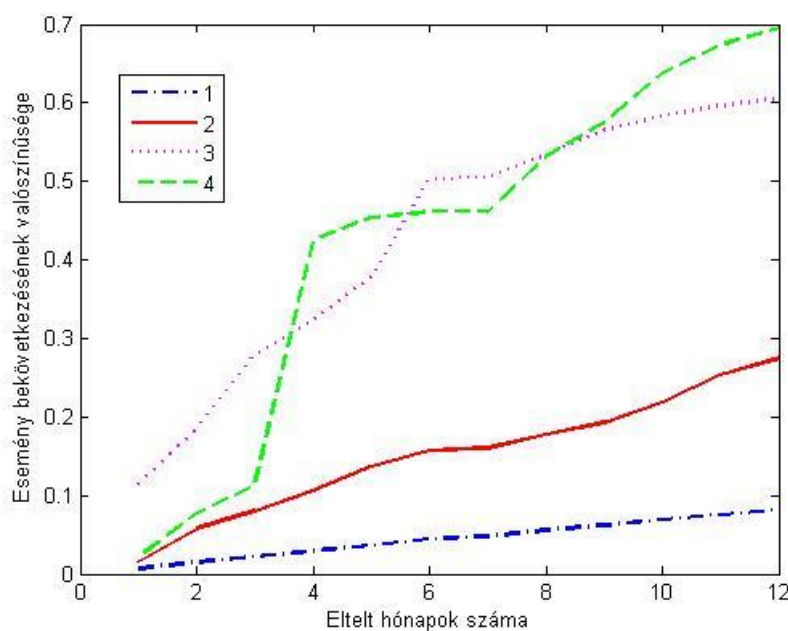


8. ábra Dun & Bradstreet adatbázisból alkotott hálózat fokszámgyakoriság eloszlása log-log skálán

### 3.3 Szomszédok hatásának vizsgálata

Szakértők véleménye szerint a gazdasági élet szereplői minél kevesebb kapcsolatszámmal rendelkeznek, annál sérülékenyebbek. Mivel mi a gazdasági szereplők esetében szeretnénk ezt a nem kívánt eseményt előre meghatározni, meg kell vizsgálnunk, hogy ez az állítás valóban helytálló, vagy esetleges feltételek mellett igaz csak.

Ez alapján a hálózat elkészítése utáni első vizsgálatunk arra irányult, hogy megállapítsuk, hogy a cégek felszámolás alá kerülésének, vagyis a negatív esemény bekövetkezésének valószínűsége függ-e a hálózat struktúrájától, pontosabban attól, hogy az adott cégnek van-e olyan szomszédja, akinél ez az esemény bekövetkezett már, továbbá befolyásolja-e, hogy hány ilyen szomszéddal rendelkezik az adott cég. A kérdés megválaszolásához első lépések egyikeként egy egyszerű vizsgálatot végeztünk el. Ennek során kiválasztottunk egy időpontot, majd azt vizsgáltuk, hogy adott felszámolás alá került szomszéddal rendelkező cégek esetében mekkora valószínűséggel következik be az esemény a kezdeti időpont utáni 12 hónapban. Kiküszöbölve az időpont hatását, a vizsgálatot több különböző kezdeti időpontra is elvégeztük. Az eredmények minden mérés esetében ugyan azt a karakterisztikát mutatták. Egyértelműen kijelenthető, hogy egy cég jobban kitett az eseménnyel szemben, amennyiben felszámolás alatt lévő céggel valamilyen tulajdonosi kapcsolatban áll, továbbá minél több ilyen szomszéddal rendelkezik, annál nagyobb ez a kitettség (9. ábra). A grafikonok az eltelt hónapok függvényében mutatják az események bekövetkezésének valószínűségét különböző számú felszámolás alá került szomszéddal való kapcsolatok esetén.



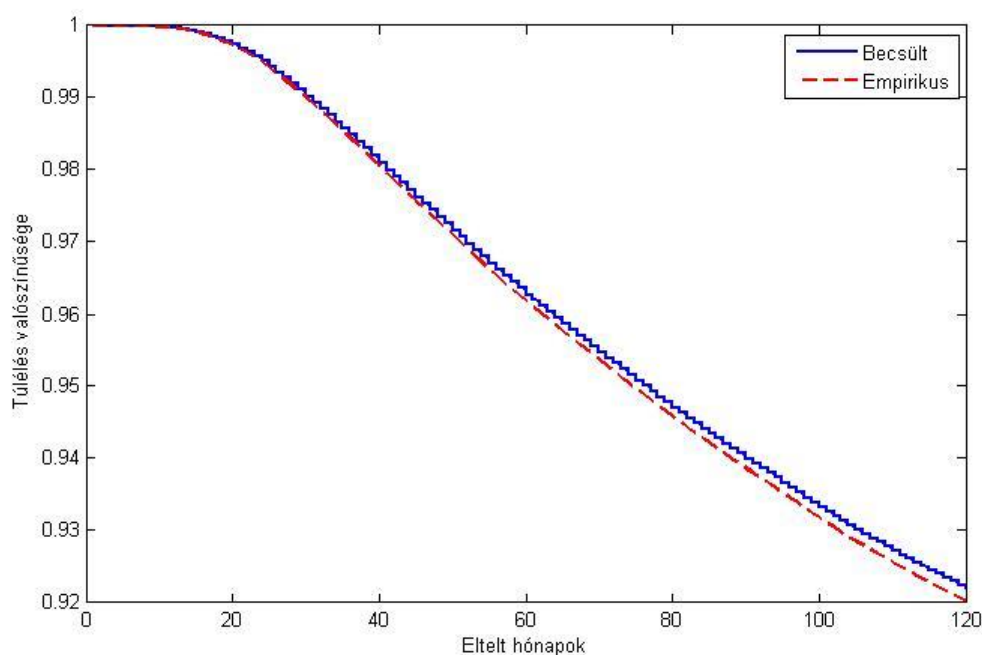
9. ábra Esemény bekövetkezési aránya 12 hónapos vizsgálat során

Azonban a vizsgálat kezdetlegessége miatt nem mutat teljesen valós képet. Amennyiben egy cég eseménnyel rendelkező szomszédjainak száma változott, úgy a céget cenzoráltuk. Mivel a hozzá kapcsolódó felszámolás alá került cégek száma változott, ezért nem tarthattuk abban a halmazban, amiben szerepelt, továbbá másik halmazba se helyezhettük.

Bár az eredmények nem felelnek meg teljesen a valóságnak, az eredmények meglepőek, hiszen a szakértők szerint egy cég annál jobban ki van téve az esemény bekövetkezésének, minél különállóbb, minél elszigeteltebb. Ezzel szemben mi arra a következtetésre jutottunk, hogy egy cég akkor kitettebb, minél több felszámolás alá került szomszédal rendelkezik. További vizsgálataink ennek az ellentmondásnak a felderítésére irányultak.

#### 3.4 Vállalkozások túlélés analízise

A túlélés analízis megkezdése előtt definiálnunk kell, mit tekintünk esemény bekövetkezésének. Jelen esetünkben ez egy cég felszámolás alá kerülését jelenti. A hálózat struktúrájának ismeretében első lépésként a teljes vállalati kapcsolatrendszerre végeztünk túlélés analízist. A vizsgálati időszak meghatározásakor nem egy konkrét időszakaszt vizsgálunk. Minden céghez meghatároztuk tevékenységük idejét, létrejöttük és bedőlésük vagy cenzorálásuk ismeretében. Ezáltal egy közös kiindulási pontból elvégezve a túlélés analízist egy általános képet kaphatunk a vizsgálandó cégek sérülékenységről, illetve várható tevékenyedési idejükről. Az általános túlélés analízis eredményeként kapott grafikon a 10. ábra látható.



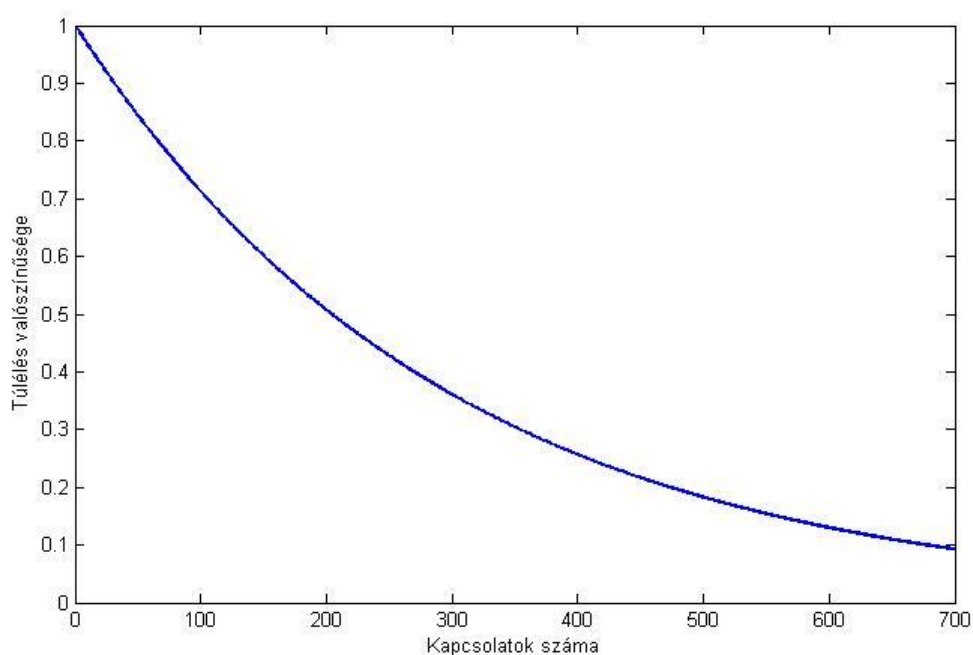
**10. ábra** Empirikus és Cox regresszió által becsült túlélési görbék az adatbázis összes cégének vizsgálatára

A túlélési görbe vizsgálatával látható, és egy általános becslésként elmondható, hogy egy cég a 10 éves ( $t = 120$  hónap) élettartamot  $\hat{S}(t) \sim 0.92$  valószínűséggel éli meg. A teljes adathalmazra a Cox-regressziós módszert alkalmazva meghatározható, hogy milyen mértékben befolyásolja a kapcsolatok száma a felszámolás alá kerülés valószínűségét. Ennek értelmében  $\underline{x}$  magyarázó változó csak a fokszámokat jelölik. Így elvégezve a Cox regressziós eljárást, eredményképpen  $\beta = 0.0034$  regressziós koefficienszt állapítottunk meg. A módszer használatával meghatározható egy becsült túlélési görbe, mely a 10. ábraán tekinthető meg. Jól látható, hogy a Cox regressziós modell alkalmazásával kapott becsült túlélési grafikon jó közelítése a Kaplan-Meier módszer alkalmazásával kapott túlélési görbének.

A  $\beta = 0.0034$  regressziós koefficiens ismeretében megállapíthatunk egy olyan parametrikus túlélési modellt is, melyben a kapcsolatok száma ( $d$ ) ismeretlenként szerepel. Ehhez valamilyen ismert sűrűségfüggvény használata szükséges, például az exponenciális sűrűségfüggvény. Az  $S = e^{-\beta d}$  képlet által meghatározott függvény az 11. ábraán látható.

### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására



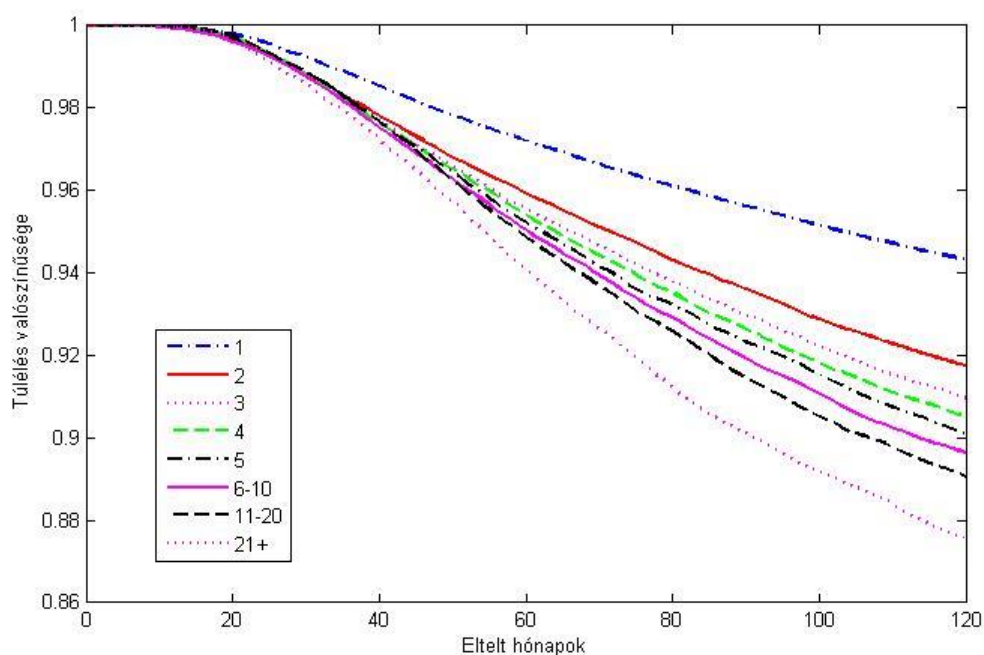
**11. ábra** Túlélés valószínűsége a kapcsolatok számának függvényében

A kapott függvénygörbét vizsgálva megállapíthatunk egy adott kapcsolatszámú vállalkozáshoz tartozó becsült valószínűségi értéket a felszámolás alá kerülés elkerüléséről.

A korábban is megállapított, szakértők véleményével ellentétes megállapítás vizsgálatához szükséges a teljes adathalmaz cégeinek foksám szerinti túlélés analízise. Ennek érdekében kapcsolatok száma szerinti csoportokat alkottunk, ezáltal minden ilyen cégcsoportra megállapíthatunk egy túlélési valószínűséget, melyek segítségével a sérülékenysége és kapcsolatszám közötti összefüggést vizsgálhatjuk. Az így kapott eredményeket mutatja a 12. ábra. Ezt vizsgálva ismét a szakértők véleményével szembeni megállapítást tudunk tenni, mely szerint egy cég sérülékenyebb akkor, ha több kapcsolattal rendelkezik.

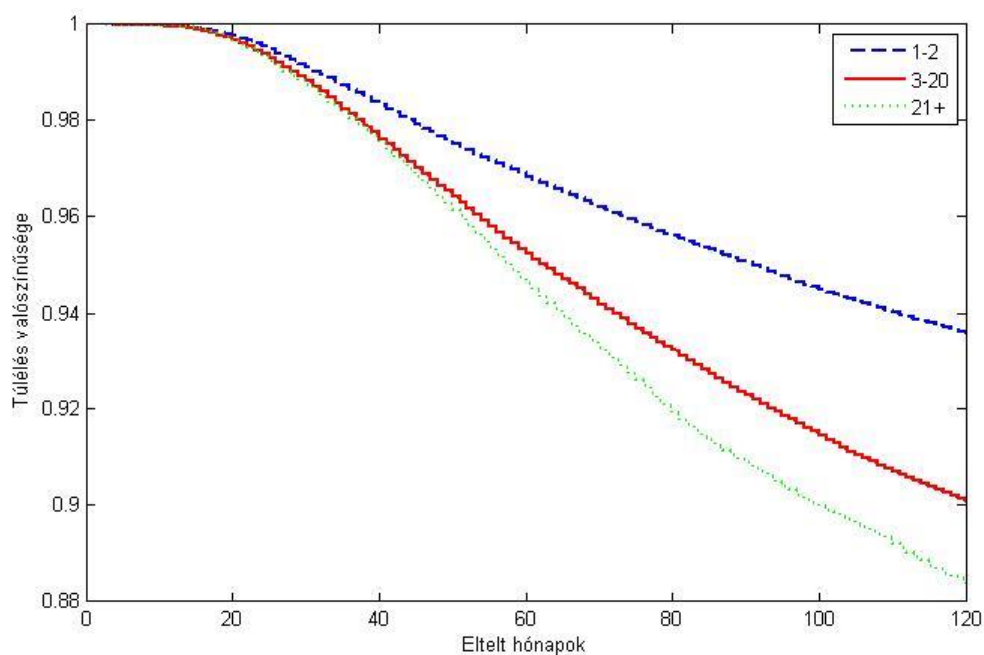
### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására



12. ábra A teljes adathalmaz túlélés grafikonjai, foksámok szerint

Látható, hogy a teljes adatbázis vállalkozásainak túlélési valószínűsége annál nagyobb, minél kevesebb kapcsolattal rendelkeznek. Kevesebb fokszámcsoportot létrehozva az eredmények a Cox regresszió alkalmazását követően sem tértek el. Szintén azt az eredményt kaptuk, mint a korábbi vizsgálatok során. A kapcsolatok száma szerinti  $\beta$  regressziós koeficiens eredmények a 3. táblázatban láthatóak, a kapott túlélési grafikonok pedig a 13. ábraán láthatóak.



13. ábra Cox regresszió által becsült túlélési görbék a teljes adathalmazra, foksámok szerint

### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

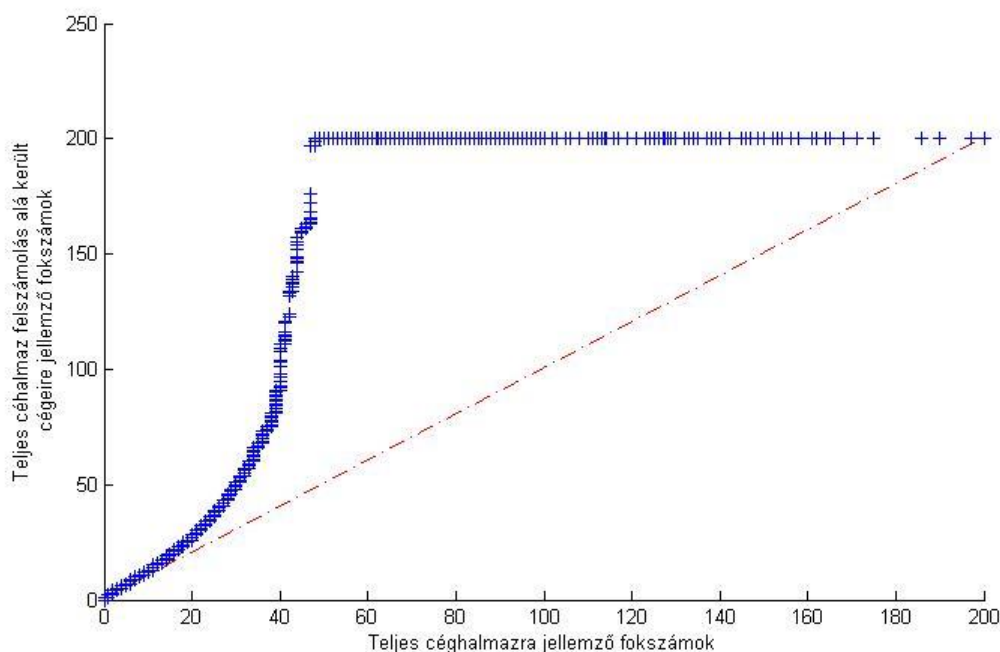
Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására

3. táblázat A cox regresszió alkalmazásával kapott eredmények

| Fokszámok          | $\beta$ |
|--------------------|---------|
| $1 \leq d \leq 2$  | 0,3898  |
| $3 \leq d \leq 20$ | 0,0155  |
| $21 \leq d$        | 0,0033  |

A 14. ábraán látható kvantilis-kvantilis diagramjának vizsgálata is megerősíti azon megállapításunkat, mely szerint bármiféle eseménysorbeli elhelyezkedéstől függetlenül a magasabb fokszámú cégek túlélési valószínűsége alacsonyabb. Jól láthatóan a magasabb fokszámok felé haladva a teljes céghalmaz és a felszámolás alá került cégek eloszlása egyre jobban eltér.

Az eseménysorok struktúrájának és dinamikájának megismerésére azonban az összes cégre elvégzett vizsgálat nem alkalmas, szükséges több további túlélés elemzés elvégzése, melyeket érdekesebb a bedőlési folyamatok kialakulásának irányából megközelíteni. Ezen megközelítéssel választ kaphatunk a korábban már jelentkezett és ismételten felmerülő ellentmondásra is.



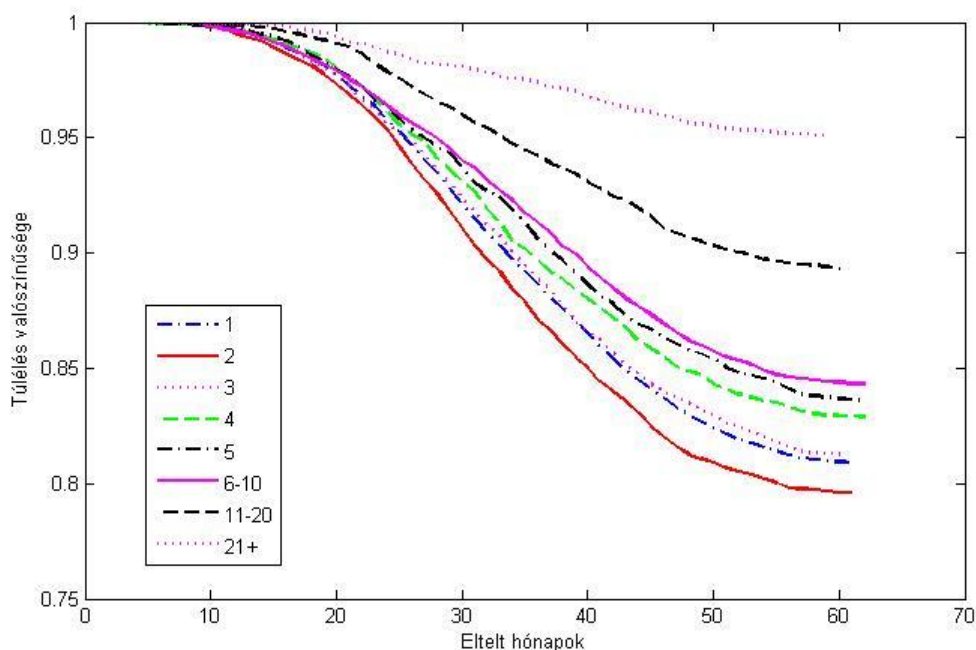
14. ábra Kvantilis-kvantilis diagram a teljes adathalmaz fokszámaina

A további túlélés analízisek elvégzése során két különböző módon kezeltük a céghálózatban szereplő vállalkozásokat. Külön vizsgáltuk azokat, melyek nem rendelkeztek olyan kapcsolattal, mely korábban felszámolás alá került partnerhez fűződtek volna, és külön azokat, akik már rendelkeztek ilyen, korábban felszámolás alá került cég felé irányuló kapcsolattal. Ez az elkülönítés szükséges, mivel a bedőlés terjedési folyamatának elemzéséhez meg kell különböztetnünk az eseménysor különböző helyein előforduló cégeket, különösképpen kiemelve az eseménysorok indításáért felelőseket. Ezenkívül a korábban kapott ellentmondás vizsgálatának érdekében a cégeket kapcsolatok szerint is csoportosítottuk mindkét további túlélés analízis elvégzése során is.

### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására

Azon vállalkozásoknál, melyek nem rendelkeztek korábban bedőlt partnerrel való kapcsolattal, megfigyelési időszaknak a cégek létrejötte és bedőlésük vagy cenzorálási idejük közötti eltelt hónapokat tekintettük, egy közös kiindulási ponttól indítva a vizsgálatokat. A kapott túlélési görbék kapcsolatok száma szerint csoportosítva a 15. ábraán láthatóak.



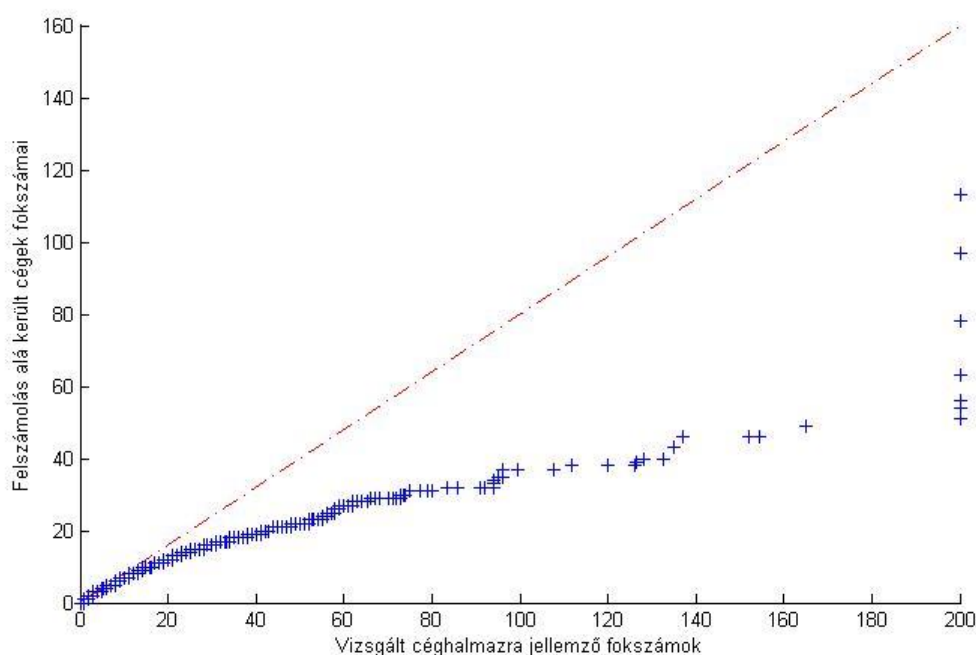
**15. ábra** Korábban bedőlt szomszédal nem rendelkező vállalkozások túlélési görbéi fokszámok szerint, 60 hónapos időintervallumot figyelembe véve, kapcsolatok száma alapján

Látható, hogy azon cégek esetén, melyek nem rendelkeztek felszámolás alá került céggel való kapcsolattal igazolódtak a szakértői vélemények is, ugyanis egy ilyen cég annál sérülékenyebb, minél kevesebb kapcsolattal rendelkezik.

Az általános túlélési görbe meghatározásához hasonlóan érdemes megvizsgálni a teljes vizsgálati adathalmaz, valamint ennek felszámolás alá került cégeinek részhalmazának fokszámain, kvantilis-kvantilis diagram segítségével. Ennek eredményei a 16. ábraán láthatóak.

### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására



16. ábra Kvantilis-kvantilis diagramon a korábban bedőlt szomszédokkal nem rendelkező cégek kapcsolatainak vizsgálatára

A 14. ábra és 16. ábra kvantilis-kvantilis diagramjai eltérő eredményeket mutatnak. Jelen esetben, a 16. ábra látható, hogy a felszámolás alá került, potenciálisan a bedőlési folyamat indításáért felelős cégek fokszám eloszlása nagy mértékben eltér nagyobb kapcsolatszámok esetén a teljes adathalmazétól. Ebből egyértelműen adódik, hogy érdemes a felszámolás alá kerülés dinamikáját vizsgálni és megkülönböztetni azon cégeket, melyek bedőlési folyamat kialakulásáért felelősek lehetnek. Ezen vállalkozások túlélési görbéinek meghatározása után összehasonlításukra ad módszert a log-rank teszt. Az adatokat a tesztnek alávétve a 4. táblázatban látható  $X^2$  eredményeket kaptuk.

4. táblázat Log-rank teszt eredményeként kapott, túlélés görbéket összehasonlító adatok

|                     | $1 \leq d \leq 2$ | $3 \leq d \leq 5$ | $6 \leq d \leq 10$ | $11 \leq d \leq 20$ | $21 \leq d$ |
|---------------------|-------------------|-------------------|--------------------|---------------------|-------------|
| $1 \leq d \leq 2$   | 0                 | 4,6588            | 6,0893             | 16,805              | 33,4963     |
| $3 \leq d \leq 5$   | 4,6588            | 0                 | 0,2946             | 5,8527              | 20,4757     |
| $6 \leq d \leq 10$  | 6,0893            | 0,2946            | 0                  | 3,3852              | 16,308      |
| $11 \leq d \leq 20$ | 16,805            | 5,8527            | 3,3852             | 0                   | 5,9046      |
| $21 \leq d$         | 33,4963           | 20,4757           | 16,308             | 5,9046              | 0           |

A kapott  $X^2$  értékekből is látható, hogy a magasabb foksámokhoz tartozó függvények egyre jobban eltérnek egymástól. A túlélési görbéket tekintve megállapítható tehát, hogy az esetek nagy többségében egy cég minél több kapcsolattal rendelkezik, annál nagyobb valószínűséggel marad stabil, ellentétben a kevesebb kapcsolattal rendelkező cégekkel. Formálisan,  $d_1$ ,  $d_2$  foksámokkal rendelkező, korábban felszámolás alá kerülő céggel kapcsolatban nem lévő cégek esetén, ha a 22. egyenletben foglaltak fennállnak, akkor a 23. egyenletben szereplő egyenlőtlenség állapítható meg.

$$\hat{S}_1(t) = P(T > t | d_1), \quad \hat{S}_2(t) = P(T > t | d_2), \quad d_1 < d_2 \quad (22)$$

$$\hat{S}_1(t) < \hat{S}_2(t) \quad (23)$$

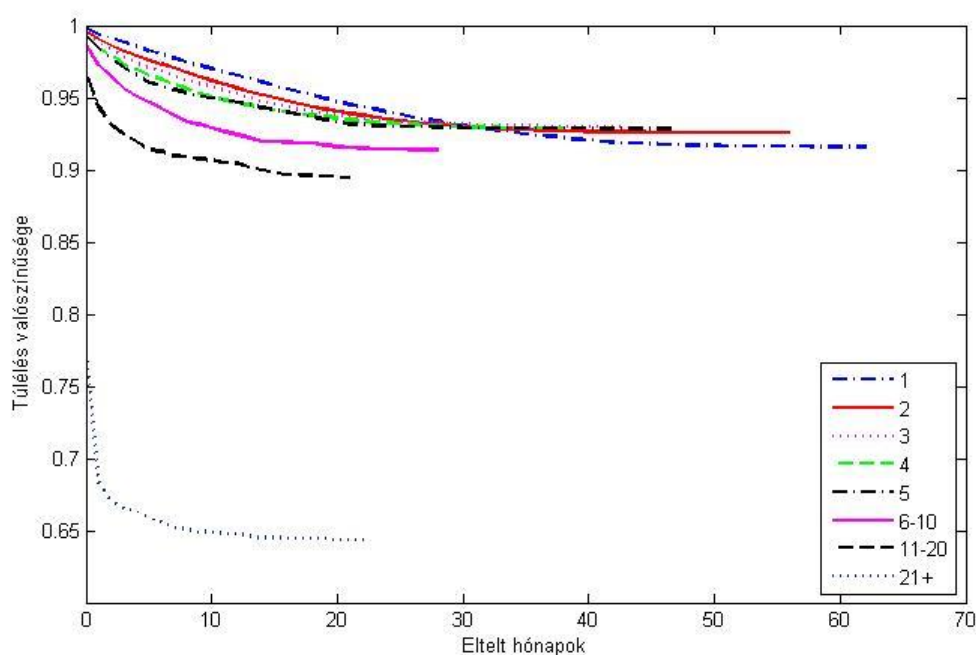
Példának okáért, ha tekintjük az 1-2 céggel kapcsolatban álló ( $1 \leq d_1 \leq 2$ ) cégekhez, és a 21-nél több céggel kapcsolatban álló ( $d_2 \geq 21$ ) cégek túlélési görbéit utóbbi céghalmaz esetén az túlélés valószínűsége  $t = 60$  hónapos megfigyelési időszak esetén magas,  $\hat{S}_2(t) \sim 0.95$ , az alacsonyabb kapcsolatszámúakkal rendelkezők esetén ez a valószínűség ugyanezen időszakaszt figyelembe véve csupán  $\hat{S}_1(t) \sim 0.8$  körüli. Ebből kifolyólag, a kevesebb kapcsolattal rendelkező, ám korábban felszámolás alá nem került céggel kapcsolatban nem lévő vállalkozások sérülékenyebbek.

Kimutattuk tehát, hogy bedőlési folyamatok kiindulásáért az alacsonyabb kapcsolatszámúakkal rendelkező cégek sokkal inkább felelősek, mint a nagyobb foksámúakkal rendelkezők. Ez a megállapítás következhet abból, hogy a több stabil kapcsolattal rendelkező cégek sokkal stabilabb társulást alkotnak, ezáltal a negatív esemény bekövetkezésének is ellenállóbbak. Globálisan tekintve a hálózatot viszont az alacsonyabb kapcsolatszámúakkal rendelkezők kevesebb eséllyel kerülnek felszámolás alá.

Ezen megállapítások következményeként adódik, hogy szükséges elkülönítenünk azon vállalkozásokat, melyek bedőlési folyamatok indítói lehetnek, azaz nem rendelkeznek felszámolás alá került céggel lévő kapcsolattal. Azonban ezen megkülönböztetés megköveteli a korábban felszámolás alá került partnerrel kapcsolatban álló cégek vizsgálatát is, melyek felszámolás alá kerülésük esetén egy létező bedőlési folyamatba kapcsolódhatnak bele. Ezen cégeknek a bedőlési folyamatsorba való bekerülése meghatározhatja a folyamat dinamikáját. A vizsgálatok első lépéseként azon cégeket határoztuk meg, melyek bedőlési láncok kialakulásáért felelősek lehetnek, azaz nem rendelkeznek olyan partnerrel, mely felszámolás alá került. Mivel második lépésként a bedőlési folyamatok alakulását és dinamikáját szeretnénk megismerni, egy cég vizsgálati időtartamának megválasztását célszerű a korábbiaktól eltérően meghatározni. Ezen időszakaszokat az egyes cégeknél a velük kapcsolatban lévő vállalkozások közül legkésőbb bedőlt bedőlési időpontjától a saját bedőlési időpontjáig eltelt idő határozza meg. Az korábban felszámolás alá került vállalatokkal kapcsolatban lévő cégek túlélési grafikonjai a 17. ábraán láthatóak.

### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására



**17. ábra** Korábban bedőlt szomszédokkal rendelkező cégek túlélési görbéi  
korábban felszámolás alá került partnerek szerint

A grafikonokat tekintve látható, hogy a kevesebb korábban bedőlt céggel való kapcsolódás esetén a felszámolás alá kerülés valószínűsége is kisebb. Ezen állítást igazolandó log-rank teszt segítségével elvégeztük a kapott függvénygörbék összevetését. A tesztek eredményeiként kapott  $X^2$  értékek az 5. táblázatban olvashatóak.

**5. táblázat** Log-rank tesztek eredményei korábban felszámolás alá került céggel lévő  
kapcsolattal rendelkező cégek összehasonlítására

|                     | $1 \leq d \leq 2$ | $3 \leq d \leq 5$ | $6 \leq d \leq 10$ | $11 \leq d \leq 20$ | $21 \leq d$ |
|---------------------|-------------------|-------------------|--------------------|---------------------|-------------|
| $1 \leq d \leq 2$   | 0                 | 14,3486           | 0,8266             | 0,0731              | 146,3498    |
| $3 \leq d \leq 5$   | 14,3486           | 0                 | 0,8978             | 1,9465              | 157,097     |
| $6 \leq d \leq 10$  | 0,8266            | 0,8978            | 0                  | 0,4919              | 104,2403    |
| $11 \leq d \leq 20$ | 0,0731            | 1,9465            | 0,4919             | 0                   | 55,1702     |
| $21 \leq d$         | 146,3498          | 157,097           | 104,2403           | 55,1702             | 0           |

A grafikonok összehasonlítása során kapott eredmények azt mutatják, hogy nehéz a kevés korábban felszámolás alá került cégkapcsolattal rendelkező vállalkozások sérülékenységbeli különbségeit megállapítani. Ez adódhat abból, hogy a különböző túlélési grafikonok nem ugyanakkora értelmezési tartománnyal rendelkeznek. Ebből kifolyólag a bedőlési láncok közepének vizsgálata nem triviális annak összetettsége miatt. A nagyobb kapcsolatszámok esetében azonban látható, hogy a korábbiakkal

### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására

ellentétben amikor azon cégeket vizsgáljuk, melyek kapcsolatban álltak olyan céggel, mely korábban felszámolás alá került, a kapcsolatok számának függvényében máshogy alakulnak a túlélések valószínűségei. Formálisan, korábban  $f_1, f_2$  felszámolás alá került céggel kapcsolatban lévő cégek esetén, ha a 24. egyenletben foglaltak fennállnak, akkor a 25. egyenletben szereplő egyenlőtlenség állapítható meg.

$$\hat{S}_1(t) = P(T > t|f_1), \quad \hat{S}_2(t) = P(T > t|f_2), \quad f_1 < f_2 \quad (24)$$

$$\hat{S}_1(t) > \hat{S}_2(t) \quad (25)$$

Példának okáért  $1 \leq f_1 \leq 2$  vizsgált foksám esetén a túlélés valószínűsége  $\hat{S}_1(t) \sim 0,95$  körüli, míg a magasabb,  $f_2 \geq 21$  kapcsolattal rendelkező cégek esetén  $\hat{S}_2(t) \sim 0,65$  körüli a vizsgált időszakaszban. Ezt megfordítva azt kapjuk, hogy a vizsgált halmazon a több felszámolás alá került céggel kapcsolatban álló vállalkozások sérülékenyebbek.

A több felszámolás alá került céggel kapcsolatban lévő cégek túlélési valószínűségének függvénygörbéje a túlélés analízis eredményeként kapott grafikonokat tekintve (17. ábraa) láthatóan eltér a többi görbétől, nem csak azért, mert a vizsgált cégek sérülékenyebbek. Észrevehető, hogy a megfigyelési időszak kezdete után kevéssel máris közel  $\hat{S}(t_1) \sim 0,7$ -re csökken a túlélés valószínűsége, majd pár hónap elteltével ez a valószínűség felveszi a végleges,  $\hat{S}(t_2) \sim 0,65$  körüli értékét. Ez azt is jelentheti, hogy a több felszámolás alá került céggel kapcsolatban lévő vállalkozásoknál az esemény bekövetkezése hamarabb fog megtörténni, ám a függvénygörbe  $t = 0$  időpontban  $\hat{S}(t) = 1$ -ről  $\hat{S}(t) \sim 0,75$ -re csökkenése inkább azt jelzi, hogy ezen sok felszámolás alá került céggel kapcsolatban álló cég közel egyszerre dől be azokkal.

A túlélési görbén kívül használt másik függvény a bedőlési ráta, amelyhez a túlélési rátát határozzuk meg először a feltételes valószínűség segítségével. Esetünkben a feltétel, hogy egy következő megfigyelt rövid időszakban mekkora eséllyel kerül el az esemény bekövetkezését egy adott cég, amennyiben az addig eltelt időszakban az még nem következett be. Másképp fogalmazva a  $(t, t + \Delta t)$  szakaszban történő túlélés valószínűségét akarjuk kiszámolni, amit a 26. egyenlet alapján tehetünk meg.

$$P(t, t + \Delta t) = \frac{R(t + \Delta t)}{R(t)} = \frac{1 - F(t + \Delta t)}{R(t)} \quad (26)$$

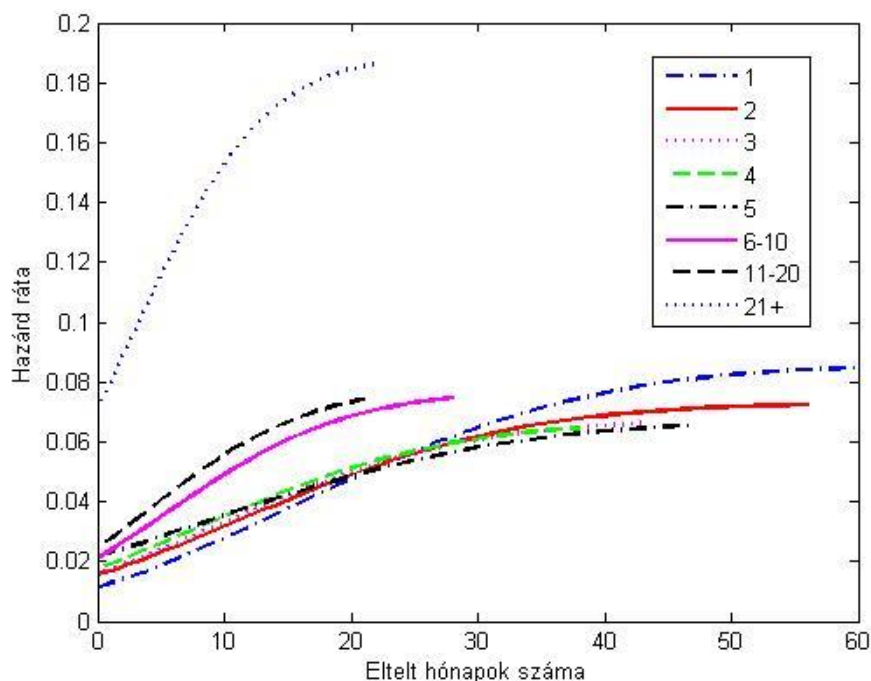
ahol  $R(t)$  a túlélési függvény,  $F(t)$  pedig az esemény bekövetkezésének eloszlásfüggvénye. A túlélési rátából kiszámíthatjuk a bedőlési rátát. A  $(t, t + \Delta t)$  szakaszban történő felszámolás alá kerülés valószínűsége, feltéve, hogy a cég a  $(0, t)$  időtartamban élt a 27. egyenlet alapján történik.

$$1 - P(t, t + \Delta t) = \frac{F(t + \Delta t) - F(t)}{R(t)} \quad (27)$$

Ha  $\Delta t$  értéke nullához tart, akkor a  $\lambda(t)$  bedőlési ráta értelmezése a 28. egyenletben szereplő összefüggés szerint lehetséges.

$$\lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t R(t)} = \frac{f(t)}{R(t)} = \lambda(t) \quad (28)$$

A  $\lambda(t)$  függvény minden  $t$  időpontban lényegében annak a valószínűségét adja meg, hogy a  $t$  időpontig túlélő szereplőn a következő időpontban bekövetkezik az esemény. Az eredményül kapott 18. ábraán jól látható, hogy több eseménnyel rendelkező szomszéd esetében adott pillanatban a felszámolás alá kerülésnek kitettsége nagyobb azoknál, akiknek kevesebb eseménnyel rendelkező szomszédja van.



18. ábra Korábban bedőlt szomszédal rendelkező cégek hazárd rátái fokszám szerint

Látható tehát, hogy nem elég egy vállalkozás sérülékenységeinek megállapításához pusztán a kapcsolatainak számát megvizsgálni. Fontos a korábban felszámolás alá került céggel való kapcsolatok száma is, mivel ez egyértelműen befolyásolja egy cég sérülékenységet. Azonban egy bedőlési folyamat alakulására pusztán ezen kapcsolatokra kihegyezett túlélés analízis nem alkalmazható teljes egészében. A folyamat dinamikájának megismeréséhez feltétlenül szükséges a felszámolás alá került cégek alkotta, úgynevezett bedőlési láncok feltárása és vizsgálata, mely olyan kérdésekre adhat választ, miszerint a bedőlési folyamat sebessége milyen módon változik az eseménysorban újabb és újabb cégeket érintve, vagy egy cég felszámolás alá kerülése közvetetten mely másik cégek bedőlésének hatására következett be. A következőkben ezen elemzéseket mutatjuk be.

### 3.5 Bedőlési folyamatok vizsgálata

A bedőlési folyamat dinamikájának vizsgálatához olyan eseménysorok kialakítása szükséges, melyekben a felszámolás alá került cégek szerepelnek, méghozzá oly módon, hogy két cég sorrendi kapcsolatot alkot az eseménysorban akkor, ha közöttük létezett korábban definiált kapcsolat, és mindketten bedőltek. A sorrendiséget az eseménysorban a vállalatok felszámolás alá kerülésének ideje határozza meg.

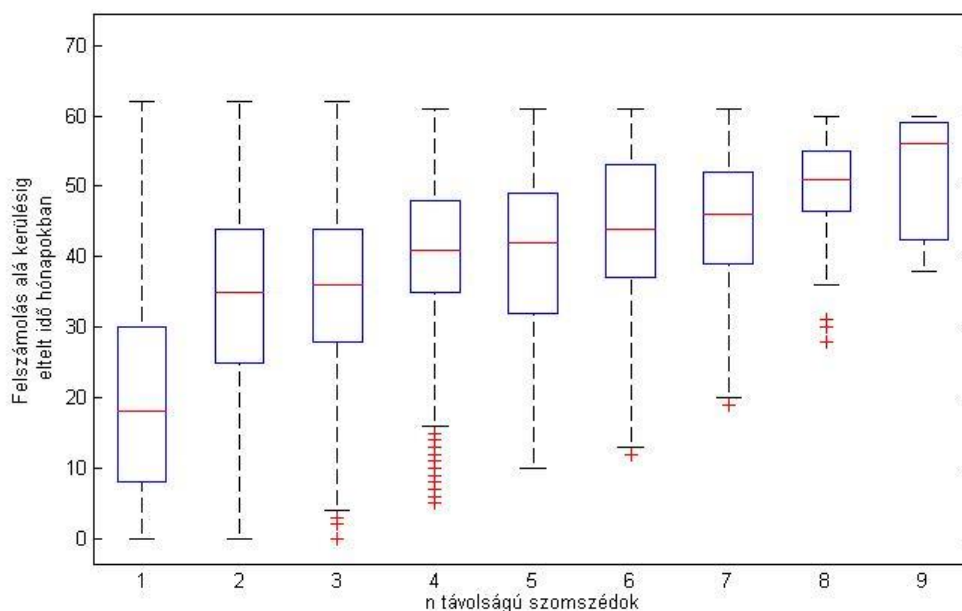
Az eseménysorok kialakításához a túlélés analízishez hasonlóan első lépésként szükséges megállapítani azon cégeket, melyek a bedőlési sorok legelején helyezkednek el, vagyis nincsenek kapcsolatban korábban felszámolás alá került partnerrel. Ezen vállalatok meghatározzák az eseménysorban soron következő cégeket. A vizsgálatok

### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására

során 11354 ilyen céget definiáltunk, ebből adóan ennyi bedőlési lánckezdetet állapíthatunk meg.

A dinamika vizsgálata az eseménysor hosszából, illetve az eseménysor különböző mélységében elhelyezkedő cégekhez tartozó eltelt idők meghatározásából épül fel. A kapott eredmények a 19. ábra láthatóak, dobozdiagramok formájában.

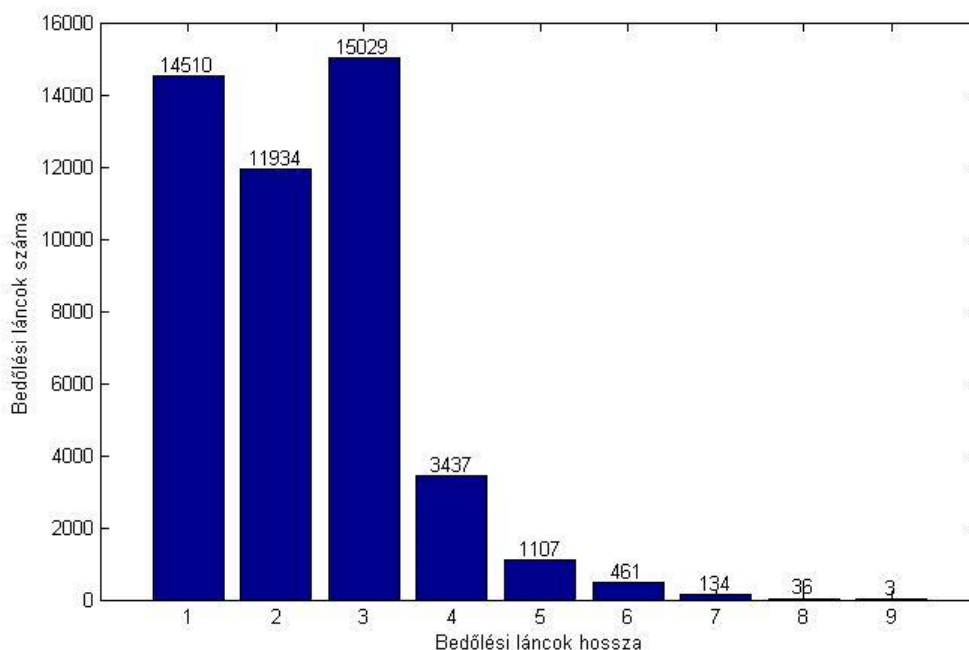


19. ábra Az eseménysorokban elhelyezkedő cégek bedőléséig eltelt idők eredményei

Az eredményeket tekintve látható, hogy felszámolás alá került vállalatok hosszú, akár 9 kapcsolatból álló eseménysort is alkothatnak. Ehhez azonban hozzátartozik az is, hogy ugyan 11354 bedőlési folyamat indításáért felelős céget állapítottunk meg, ilyen hosszal azonban csupán 3 bedőlési lánc rendelkezik. A bedőlési folyamatok számának hossz szerinti alakulását mutatja a 20. ábra. Látható, hogy főleg a rövidebb láncok dominálnak, a kialakult bedőlési folyamatok több mint fele 1 vagy 2 hosszúságú.

### 3. Kidolgozott megközelítésmód és alkalmazásának eredményei

Túlélés-elemzés hálózatokban - új eszköz a komplex rendszerek dinamikájának feltárására



20. ábra Bedőlési folyamatok hosszainak eloszlása

Az eseménysorok kezdetéért felelős cégek felszámolás alá kerülése és a láncon belül különböző távolságra elhelyezkedő cégek felszámolás alá kerülése között eltelt idők átlaga hónapokban mérve szintén a dobozdiagramokat tartalmazó 19. ábra tekinthető meg. A kialakult sorrendi kapcsolatok eltelt idő szerinti vizsgálataiból megállapítottuk, hogy egy cég felszámolás alá kerülése közvetve hatással lehet egy olyan másik cégre is, amelyik akár 50 hónap elteltével kerül dől be.

Az eseménysorokban különböző távolságokban elhelyezkedő vállalkozások felszámolás alá kerüléséig eltelt hónapok vizsgálata azt mutatja, hogy az adatok viszonylag nagy tartományban szórnak. Tekintve az adatsorban egy, kettő vagy három távolságra lévő szomszédait a folyamat indításáért felelős cégeknek, a kapcsolódó vállalkozások bedőlése 0-tól 60 hónapig terjedhet, az átlagos eltelt hónap viszont folyamatosan növekszik.

A kapott bedőlési folyamatra jellemző adatokat megvizsgálva választ adhatunk azon kérdésre, mely azt célozza meg, hogy az eseménysorban a felszámolás alá kerülés gyorsasága hogyan változik. Ezáltal vizsgálható a bedőlési folyamatban különböző mélységeiben elhelyezkedő cégek sérülékenységbeli eltérése. A 19. ábra eredményeiből látható, hogy nem állapítható meg sem az események bekövetkezésének gyorsulása, sem a lassulása, ebből következik az is, hogy a sorozatban elhelyezkedő cégek túlélési valószínűségét nem befolyásolja az, hogy a bedőlési folyamatokban hol helyezkednek el.

A bedölések dinamikájának meghatározásához a túlélés analízis és a bedőlés folyamatának együttes vizsgálata volt szüksége. A kapott eredményeink alapján megállapítható, hogy el kell különítenünk a bedőlési folyamatok kialakításáért potenciálisan felelős, és egy már létező folyamatba bekerülésre esélyes cégeket. A dinamika vizsgálatával megállapítottuk, hogy nem lényeges a cégek folyamatsoron belül való elhelyezkedése, csupán a kapcsolatainak számossága. Ezzel is igazoltuk, hogy a hasonló folyamatok vizsgálatánál sikeresen alkalmazhatóak a Markov modellek [23].

### 4 Összefoglalás

Munkánk során elsődleges célunk egy olyan új eszköztár kifejlesztése volt, amely integrálja a túlélés analízis és a hálózat elemzés eredményeit. A javasolt módszer segítségével komplex módon elemezhetők a vizsgált egyedek túlélési esélyei oly módon, hogy figyelembe vesszük azok kapcsolatait és egymásra gyakorolt hatását is. Munkánk során sikeresen alkalmaztuk mind a Kaplan-Meier, mind a Cox regressziós modellt hálózatban értelmezett adatokon. Mindkét módszert hálózatban számítható paraméterek mentén értelmeztük, beleértve egy adott elem fokszámát, az eseménnyel rendelkező szomszédok számát, arányát, és a szomszédokon bekövetkezett utolsó esemény óta eltelt időt.

Az általunk fejlesztett eszköztárt egy gazdasági szereplők által alkotott hálózaton teszteltük, ahol a cégek felszámolás alá kerülése jelentette a vizsgált kockázati eseményt. A hálózatot az összes magyar gazdasági szereplőt tartalmazó Dun & Bradstreet adatbázisból alkottuk, amely több mint 800000 cégre vonatkozóan tartalmaz információt 2009 eleje és 2013 eleje közötti időintervallumra vonatkozóan. A fejlesztett eszköztár segítségével fontos megállapításokat tettünk vállalkozások sérülékenységét illetően a hálózatban elfoglalt pozíciója alapján. Kutatásunk során megmutattuk, hogy kapcsolatok számától függetlenül egy vállalkozás közel 92% valószínűséggel marad stabil. Azonban a kapcsolatok száma szerinti, vizsgálatok által megfogalmazott megállapításaink, mely szerint egy vállalkozás annál sérülékenyebb, minél több partnerrel áll kapcsolatban ellentmondást mutattak szakértői véleményekkel. Az ellentmondás vizsgálatának, és bedőlési folyamatok analízisének céljából megvizsgáltuk a folyamatok indításáért felelős vállalkozások felszámolás alá kerülésének valószínűségeit. Megállapítottuk, hogy a szakértői álláspont csupán akkor igaz egy vállalkozás élettartamával kapcsolatban, ha a cég potenciális elindítója egy bedőlési folyamatnak. Ezen cégek minél több kapcsolattal rendelkeznek, annál kisebb valószínűséggel vannak kitéve a negatív esemény bekövetkezésének. A bedőlési folyamatokat tovább vizsgálva megmutattuk, hogy a felszámolás alá került partnerek egyértelműen befolyásolják egy vállalkozás bedőlési folyamatba való bekerülésének valószínűségét. Megállapítottuk, hogy minél több bedőlt céghez kapcsolódik egy működő cég, annál nagyobb a felszámolás alá kerülésének veszélye. A felszámolás alá kerülés dinamikájának vizsgálatával megállapítottuk, hogy a folyamat előrehaladása nem befolyásolja a negatív esemény bekövetkezésének valószínűségét, a folyamat se nem gyorsul, se nem lassul. A felszámolás alá kerülés tehát nem függ az eseménysorban elfoglalt pozíciótól, csupán a kapcsolatok száma és partnerek állapota a meghatározó.

Munkán folytatásaként a Cox regresszió további alkalmazásával, új paraméterek bevezetésével olyan elemzéseket kívánunk végrehajtani, amely még jobban árnyalja a negatív esemény bekövetkezésének valószínűségét. A kapott eredményekből meghatározható egy olyan általános dinamikus modell, mely különböző hálózatokon történő sorrendi események vizsgálatát teszi lehetővé. Az így elkészült modell segítségével előre rá lehet majd mutatni a gazdasági élet során veszélyeztetetté váló vállalatokra, mielőtt azokon a negatív esemény valóban bekövetkezne. Ezek az információk rendkívül hasznosak lehetnek például a bankok számára a hitelszolgáltatásaik finanszírozása és menedzselése terén.

## Jelölésjegyzék

|                 |                                                                                 |
|-----------------|---------------------------------------------------------------------------------|
| $\delta$        | cenzorálási változó                                                             |
| $S(t)$          | túlélési függvény                                                               |
| $\hat{S}(t)$    | empirikus túlélési függvény                                                     |
| $h(t)$          | hazard függvény                                                                 |
| $t_j$           | megfigyelési időpont                                                            |
| $m_j$           | „halálozások” száma $t_j$ és $t_{j+1}$ időpontok között                         |
| $q_j$           | cenzorált egyének száma $t_j$ és $t_{j+1}$ időpontok között                     |
| $R(t_j)$        | veszélyeztetett egyének halmaza, melyek $t_j$ időpontig túléltek                |
| $n_{ij}$        | az $i$ -edik vizsgált csoport mérete a $j$ -edik időpontban                     |
| $e_{ij}$        | az $i$ -edik vizsgált csoport várható halálozási száma a $j$ -edik időpontban   |
| $m_{ij}$        | az $i$ -edik vizsgált csoport tényleges halálozási száma a $j$ -edik időpontban |
| $O_i$           | az $i$ -edik csoport megfigyelt halálozásainak száma                            |
| $E_i$           | az $i$ -edik csoport várt halálozásainak száma                                  |
| $X^2$           | log-rank statisztikai mutató                                                    |
| $\underline{x}$ | Cox regressziónál alkalmazott magyarázó változók                                |
| $\beta$         | regressziós koefficiens                                                         |
| $h_0(t)$        | kiindulási hazard függvény                                                      |
| $RR$            | relatív kockázat                                                                |
| $G$             | gráf                                                                            |
| $V$             | csúcsok halmaza                                                                 |
| $E$             | élek halmaza                                                                    |
| $K_i$           | $i$ csúcsból álló teljes gráf                                                   |
| $N$             | Csúcsok számossága                                                              |
| $a_{ij}$        | $i$ . és $j$ . csúcs között található él                                        |
| $d^-(v)$        | $v$ csúcs kifoka                                                                |
| $d^+(v)$        | $v$ csúcs befoka                                                                |
| $F(t)$          | meghibásodás valószínűsége $t$ időpont előtt                                    |
| $d$             | kapcsolatok száma                                                               |
| $f$             | felszámolás alá került cégekkel lévő kapcsolatok száma                          |

## Irodalomjegyzék

- [1] R. M. Anderson, „Population Dynamics of Infectious Diseases: Theory and Applications,” *Chapman and Hall*, 1982.
- [2] M. R. P. Keeling, „Modeling Infectious Diseases: In Humans and Animals,” *Princeton University Press*, 2007.
- [3] J. Golbeck, *Analyzing the Social Web*, 2013.
- [4] M. K. Jiawei Han, *Data Mining. Concepts and Techniques*, 2001.
- [5] C. Berge, *Graphs*, 1985.
- [6] B. Bollobás, *Graph Theory and Combinatorics*, 1988.
- [7] D. M. M. Centola, „Complex Contagion and the Weakness,” 2005.
- [8] G. Kossinets és D. J. Watts, „Empirical Analysis of an Evolving Social Network,” *Science* vol. 311, pp. 88-90, 2006.
- [9] R. W. Butler és D. A. Bronson, „Multistate Survival Models as Transient Electrical Networks,” *Scandinavian Journal of Statistics*, vol. 41, p. 167–186, 2014.
- [10] S. E. Christodoulou, „Water Network Assessment and Reliability Analysis by Use of Survival Analysis,” *Water Resour Management*, p. 1229–1238.
- [11] D. G. Kleinbaum és M. Klein, *Survival Analysis A Self-Learning Text*, Springer, 2005.
- [12] J. Berkson és R. P. Gage, „Survival curves for cancer patients following treatment,” *Journal of the American Statistical Association*, 47, pp. 501-515, 1952.
- [13] E. L. Kaplan és P. Meier, „Nonparametric estimation from incomplete observations,” *J. Amer. Statist. Assn.* 53 (282), p. 457–481, 1958.
- [14] D. R. Cox, „Regression Models and Lifet-Tables,” *Journal of the Royal Statistical Society*, pp. 187-220, 1972.
- [15] N. Mantel, „Evaluation of survival data and two new rank order statistics arising in its consideration,” *Cancer Chemotherapy Reports* 50 (3), p. 163–70, 1966.
- [16] J. Wiley, *The Statistical Analysis of Survival Time Data*, pp 223-224, 1980.
- [17] P. K. Andersen és R. D. Gill, „Cox's regression model for counting processes: A large sample study,” *The Annals of Statistics Volume 10, Number 4*, pp. 1100-1120., 1982.

- [18] S. Wasserman és K. Faust, „Social Network Analysis in the Social and Behavioral Sciences. Social Network Analysis: Methods and Applications,” *Cambridge University Press*, pp. 1-27, 1994.
- [19] M. Girvan és M. E. J. Newman, „Community structure in social and biological networks,” *In Proceedings of the National Academy of Sciences. Vol. 99. National Academy of Sciences*, 2002.
- [20] J. Travers és S. Milgram, „An experimental study of the Small World problem,” *Sociometry* 32, 4, pp. 425-443, 1969.
- [21] A.-L. Barabási és R. Albert, „Emergence of scaling in random networks,” *Science*, 286, pp. 509-512, 1999.
- [22] R. Albert és A.-L. Barabási, „Statistical mechanics of complex networks,” *Reviews of Modern Physics*, volume 74, 2002.
- [23] F. A. Sonnenberg és J. R. Beck, „Markov models in medical decision making: A practical guide,” *Medical Decision Making*, pp. 322-338, 1993.