

# ALKALMAZOTT MATEMATIKAI LAPOK

## A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI TUDOMÁNYOK OSZTÁLYÁNAK KÖZLEMÉNYEI

ALAPÍTOTTÁK

KALMÁR LÁSZLÓ, TANDORI KÁROLY, PRÉKOPA ANDRÁS, ARATÓ MÁTYÁS

FŐSZERKESZTŐ

PÁLES ZSOLT

FŐSZERKESZTŐ-HELYETTESEK

BENCZÚR ANDRÁS, GERENCSÉR LÁSZLÓ, SZÁNTAI TAMÁS

FELELŐS SZERKESZTŐ

BOZÓKI SÁNDOR

FELELŐS SZERKESZTŐ-HELYETTES

CSATÓ LÁSZLÓ

TÖRDELŐSZERKESZTŐ

MOCZÁR KÁROLY

A SZERKESZTŐBIZOTTSÁG TAGJAI

Arató Miklós, Baran Sándor, Csáji Balázs Csanád, Csendes Tibor, Csirik János, Fazekas István, Forgó Ferenc, Frank András, Fridli Sándor, Friedl Katalin, Friedler Ferenc, Galántai Aurél, Garay Barna, Gazdag-Tóth Boglárka, Gyimóthy Tibor, Hajdu András, Hartung Ferenc, Hatvani László, Heppes Aladár, Horváth Zoltán, Illés Tibor, Járai Antal, Jelasity Márk, Katona Gyula, Király Tamás, Kis Tamás, Kovács Gergely, Krisztin Tibor, Lovász László, Maksa Gyula, Maros István, Michaletzky György, Miklós István, Molnár-Sáska Gábor, [Pap Gyula], Pituk Mihály, Rásonyi Miklós, Recski András, Rónyai Lajos, Röst Gergely, Simon Péter, Szabó Péter Gábor, Szeidl László, Szilágyi Brigitta, Tasnádi Attila, Temesi József, Tusnády Gábor, Vizi Zsolt, Vizvári Béla

37. kötet

Szerkesztőség és kiadóhivatal: 1055 Budapest, Falk Miksa u. 12.

Az Alkalmazott Matematikai Lapok változó terjedelmű füzetekben jelenik meg, és olyan eredeti tudományos cikkeket publikál, amelyek a gyakorlatban, vagy más tudományokban közvetlenül felhasználható új matematikai eredményt tartalmaznak, illetve már ismert, de színvonalas matematikai apparátus újszerű és jelentős alkalmazását mutatják be. A folyóirat közöl cikk formájában megírt, új tudományos eredménynek számító programokat, és olyan, külföldi folyóiratban már publikált dolgozatokat, amelyek magyar nyelven történő megjelenítése elősegítheti az elért eredmények minél előbbi, széles körű hazai felhasználását. A szerkesztőbizottság bizonyos időnként lehetővé kívánja tenni, hogy a legjobb cikkek nemzetközi folyóiratok külön-számaként angol nyelven is megjelenhessenek.

A folyóirat feladata a Magyar Tudományos Akadémia III. (Matematikai) Osztályának munkájára vonatkozó közlemények, könyvismertetések stb. publikálása is.

A kéziratok a főszerkesztőhöz, vagy a szerkesztőbizottság bármely tagjához beküldhetők. A főszerkesztő címe:

Páles Zsolt, főszerkesztő

1055 Budapest, Falk Miksa u. 12.

A folyóirat e-mail címe: [aml@renyi.hu](mailto:aml@renyi.hu)

A folyóirat honlapja: <http://aml.math.bme.hu>

Közlésre el nem fogadott kéziratokat a szerkesztőség lehetőleg visszajuttat a szerzőhöz, de a beküldött kéziratok megőrzéséért vagy továbbításáért felelősséget nem vállal.

Az Alkalmazott Matematikai Lapok előfizetési ára évfolyamonként 1200 forint. Megrendelések a szerkesztőség címén lehetségesek.

A Magyar Tudományos Akadémia III. (Matematikai) Osztálya a következő idegen nyelvű folyóiratokat adja ki:

1. Acta Mathematica Hungarica,
2. Studia Scientiarum Mathematicarum Hungarica.

## ELŐSZÓ

Az *Alkalmazott Matematikai Lapok* 4 év szünet után 2004-ben jelent meg ismét. Ez volt az első olyan szám, amit az akkor új szerkesztők jelentettek meg (főszerkesztő: Páles Zsolt, felelős szerkesztő: Vizvári Béla, technikai szerkesztő: Kovács Gergely).

Azóta ez a szerkesztőség 16 évfolyam 25 számának elkészültében vett részt. Megjelent számos méltatás, tudománytörténeti cikk, illetve a számok már az interneten is elérhetőek, letölthetőek.

2019-ben Vizvári Béla és Kovács Gergely sokadik megbízása után a megújulás reményében lemondott. A szerkesztőbizottság felelős szerkesztőnek Bozóki Sándort, helyettesének Csató Lászlót választotta meg. (Nekik köszönhető, hogy azóta a cikkek már DOI-azonosítóval is rendelkeznek.) Az új szerkesztőknek jó munkát kívánunk!

Ugyan a kötet borítóján már az új szerkesztők neve szerepel, de a jelen szám cikkei még zömében 2019-ben érkeztek be, ezért ez a szám még jelentős részben a régi szerkesztők munkája.

Vizvári Béla, Kovács Gergely

STOYAN GISBERT EMLÉKÉRE  
(1942-2018)



**Stoyan Gisbert tudományos életútja**

Stoyan Gisbert 1942-ben született Berlinben. Az egyetem elvégzése után a berlini Alkalmazott Matematika és Mechanika Intézetben dolgozott, majd 1967 és 1971 között Moszkvában, a Lomonoszov Egyetemen volt aspiráns. Szamarszkij professzor, a világhírű szovjet numerikus matematika kiemelkedő képviselője volt a témavezetője. Ő a későbbiekben is figyelemmel kísérte Stoyan Gisbert pályafutását, és mindig nagy elismeréssel szólt tudományos eredményeiről. A sikeres fokozatszerzés után 1983-ig Berlinben, a Weierstrass Intézetben dolgozott. Az Eötvös Loránd Tudományegyetemen 1983-ban az egyetemi számítóközpont munkatársaként kezdte több mint három évtizedes karrierjét. Ottani munkája a számítógépes alkalmazások egy fontos területéhez, a differenciálegyenletek numerikus megoldásához kapcsolódott. A számítóközpontban töltött tíz év után, 1993-tól közel három évtizedig, nyugdíjba vonulásáig a Numerikus Analízis Tanszék egyetemi tanára, utána pedig professzor emeritusa volt.

Stoyan Gisbert professzor az őt felületesen ismerők számára zárkózottnak tűnhetett, de ő egyáltalán nem volt elzárkózó. Mindenkinek szívesen segített, aki matematikai kutatási, oktatási problémával felkereste. Hozzáállására leginkább a szigorú szakmai következetesség és igényesség volt a jellemző. Ennek köszönhető az alkalmazott matematikai körökben kivívott szakmai megbecsülését, hitelességét. Véleményét oktatási kérdésekben is röviden, egyértelműen és nyíltan fogalmazta meg. Kollégáival jó viszonyt ápolt, közelebbi munkatársaival szorosabb személyes kapcsolatban is volt. Magatartása mintaként szolgált környezetének, tanítványainak.

### **Kutatói munkássága**

Stoyan Gisbert professzor magas színvonalú kutatási tevékenységével nemzetközi elismertséget szerzett a numerikus matematika és annak gyakorlati alkalmazásai területén. Különösen szép eredményeket ért el a parciális differenciálegyenletekkel kapcsolatban. Eredményeit jó meglátások és az elegáns megoldások jellemezték.

Kutatási érdeklődése széles körű volt, a numerikus analízis, a differenciálegyenletek numerikus megoldásai témaköréhez, azon belül is leginkább a parciális differenciálegyenletek numerikus megoldási módszereihez kapcsolódott. A differenciálegyenletek numerikus megoldásának megmaradási tételeivel, például a pozitívítással, monotonitással kapcsolatos vizsgálatok során is szép eredményeket ért el. Paraméterbecslési módszerei, az úgynevezett inverz feladat megoldására vonatkozó eredményei is tanúsítják az alkalmazásokhoz való szoros kapcsolatát, elkötelezettségét. A multigríd módszerekhez kötődő kutatásaiból is számos színvonalas publikáció született.

A közel elliptikus parabolikus egyenletek integrálására adott módszere olyan problémakört tárgyalt, ami sokáig szinte megoldhatatlannak tűnő feladat volt. Számos cikkben foglalkozott a Stokes-egyenlet véges elemes megoldásával. Késői publikációi, amelyek a Crouzeix–Velte-felbontás területén születtek, mind fontos és jelentős dolgozatok.

Tudományos munkássága, eredményei Stoyan Gisbertet nemzetközileg ismert és elismert kutatóvá tették.

### **Egyetemi oktatói munkássága**

Az itthoni matematikai közösség olyan széles és mély matematikai műveltséggel rendelkező szakemberként tisztelte, aki számos témakör művelését honosította meg itthon. Nemzetközi kapcsolatait kihasználva a numerikus matematika területén kutatási együttműködést hozott létre neves külföldi egyetemek, kutatóközpontok és egyetemünk kutatócsoportjai között. Tudományos kutatásaihoz kapcsolódóan doktori témákat hirdetett meg tehetséges hallgatók számára. A kutatói utánpótlás nevelése terén végzett tevékenysége példaértékű. Több sikeresen védett doktori ösztöndíjasa ma már a numerikus matematika elismert szakembere. A hazai numerikus matematikát oktató, művelő szakemberek jó része az ő közvetlen, vagy közvetett tanítványának tekinthető.

Széles körű egyetemi oktatási, tananyagfejlesztési tevékenysége két kiemelkedő példájának egyike az intenzív tankönyvírói munkássága, a másik az ELTE-n a matematikus képzés keretében folyó alkalmazott matematikus szak létrehozásában betöltött kezdeményező és megvalósító szerepe. Kitartó szervezői munkájának döntő szerepe volt abban, hogy a színvonalas képzést nyújtó szak elindulhatott. Nemcsak a szak létrehozásában, hanem annak működtetésében is aktívan részt

vett. Ő vezette be és dolgozta ki a „Nemlineáris problémák alkalmazott feladatokban, esettanulmányok” című tárgy anyagát. Emellett a Numerikus analízis tárgy anyagának kidolgozója, a tárgy felelőse és hosszú ideig oktatója volt.

Különösen fontosnak tartotta, hogy a hallgatók valódi gyakorlati problémákkal találkozzanak. A diplomamunka védések alkalmával alaposan meggyőződött arról, vajon a hallgató igazi, az ipari gyakorlatban felmerülő problémával foglalkozott-e, és hogy az általa adott megoldás valóban alkalmazható-e a gyakorlatban. Intenzív tankönyvírói tevékenységének ikonikus példája a társszerzővel írt háromkötetes Numerikus módszerek monográfia. A méretét tekintve is lenyűgöző, 1000 oldalt meghaladó mű hiánypótló és meghatározó a magyar nyelvű numerikus matematika oktatásban. Korszerű, modern tartalma és tárgyalásmódja kiállja az összehasonlítást a külföldi szakirodalomban megjelent hasonló témájú művekkel. Az első sorban nem matematikusok számára írt Numerikus matematika mérnököknek és programozóknak című könyvének elkészítette az angol nyelvű változatát is, ami a Birkhäuser Kiadónál jelent meg. A 2016-os első kiadása olyan sikeresnek bizonyult, hogy a kiadó rögtön megkereste a második kiadás ügyében.

Egyetemi tanári tevékenységén belül a tudományos kutatást és az oktatást végig egymással összhangban művelte. Kiemelkedő évtizedes munkássága során számos elismerésben részesült. Kollégái megbecsülését jelzi a tiszteletére rendezett jubileumi konferencia. A magyar állam 2012-ben a Magyar Érdemrend Lovagkeresztjével tüntette ki. A HU-MATHS-IN, Magyar Ipari és Innovációs Matematikai Szolgáltatási Hálózat, elindította a róla elnevezett Stoyan Gisbert szemináriumot.

Szakmai hagyatékát a hazai numerikus matematikai közösség magáénak érzi, azt a jövőben is gondozni fogja.

Fridli Sándor és Gergő Lajos

### Stoyan Gisbert publikációi

- [1] STOYAN GISBERT: *On some economic additive difference schemes for the solution of manydimensional partial differential equations of parabolic type*, Soviet Journal of Numerical Mathematics and Mathematical Physics, Vol. **10** No. **3**, pp. 644-653 (1970). (in Russian)
- [2] STOYAN GISBERT: *On the stability of additive difference schemes with respect to boundary values*, Soviet Journal of Numerical Mathematics and Mathematical Physics, Vol. **11** No. **4**, pp. 934-947 (1971). (in Russian)
- [3] STOYAN GISBERT: *Zur Genauigkeit eines ökonomischen additiven Differenzschemas*, Math. Nachrichten, Vol. **58**, pp. 247-255 (1973).
- [4] STOYAN GISBERT: *On the stability of the two-dimensional Janenko scheme with respect to boundary values*, in: Theory of Nonlinear Operators: Proceedings of a Summerschool held at Neuchtel (Hiddensee) in October, 1972, Akademie-Verlag, Berlin, pp. 241-246 (1974).
- [5] STOYAN GISBERT: *Higher order difference schemes for the first and third boundary value problem to  $1/rd/dr(rdu/dr) + f(r) = 0$* , ZAMM-Z. Angew. Math. Me., Vol. **55**, pp. 635-645 (1975).
- [6] STOYAN GISBERT: *Numerical experiments on the identification of heat conduction coefficients*, in: Theory of Nonlinear Operators: Proc. Fifth Internat. Summer School, Central Inst. Math. Mech. Acad. Sci. GDR, Berlin, 1977, Akademie-Verlag, Berlin, pp. 259-268 (1978).
- [7] STOYAN GISBERT: *Some results of numerical experiments on identification of a spatially varying heat conduction coefficient*, in: Summerschool of KAPG 5.2, Freiberg (1978).
- [8] STOYAN GISBERT: *On the identification of diffusion coefficients*, in: Mathematical models and numerical methods (Papers, Fifth Semester, Stefan Banach Internat. Math. Center, Warsaw, 1975), Banach Center Publ., Vol. **3**, PWN, Warsaw, pp. 367-377 (1978).
- [9] STOYAN GISBERT: *On a maximum norm stable, monotone and conservative differenceapproximation of the one-dimensional diffusion-convection equation*, in: Simulation der Migrationsprozesse im Boden- und Grundwasser, TU Dresden, pp. 139-160 (1979).
- [10] STOYAN GISBERT: *Identification of a spatially varying coefficient in a parabolic equation. A report on numerical experiments*, in: Inverse and Improperly Posed Problems For Partial Differential Equations, Akademie-Verlag, Berlin, pp. 249-258 (1979).
- [11] STOYAN GISBERT: *Monotone difference schemes for diffusion-convection problems*, ZAMM-Z. Angew. Math. Me., Vol. **59** No. **8**, pp. 361-372 (1979).
- [12] STOYAN GISBERT: *Modelling and computation of water quality problems in river networks*, in: Lecture Notes in Control and Information Science 23, Springer, Berlin, pp. 482-491 (1980). (with H. Baumert, P. Braun, E. Glos and W. Müller)
- [13] STOYAN GISBERT AND D. STOYAN: *Über die Formen-Maxima-Regel von A.H. Müller, Teil 1*, in: Freiburger Forschungshefte C357, Leipzig, pp. 105-110 (1980).
- [14] STOYAN GISBERT: *Über eine monotone Differenzenapproximation einer partiellen Differentialgleichung*, in: Seminar on Numerical Methods for Solving Balance Equations: Papers presented at the Seminar held in Berlin, October 20-25, 1980, Akademie der Wissenschaften der DDR, Berlin, pp. 83-94 (1980).

- [15] STOYAN GISBERT: On the asymptotic stability of some economic difference schemes, *Soviet Journal of Numerical Mathematics and Mathematical Physics*, Vol. **20** No. **2**, pp. 350-358 (1980). (in Russian)
- [16] STOYAN GISBERT: *Ein Fortran-Programm zur Lösung von Randwertproblemen für Systeme aus zwei partiellen Differentialgleichungen mit konstanten Koeffizienten*, in: *Numerische Behandlung mathematischer Modellgleichungen*. Report 09/80, ZIMM der AdW der DDR, Berlin (1980).
- [17] STOYAN GISBERT: *Zu einigen Arbeiten über monotone Differenzenschemata*, *Wiss. Beiträge IHS Zwickau*, Vol. **7** No. **2**, pp. 67-68 (1981).
- [18] STOYAN GISBERT: *Mathematical modelling of a class of paleontological evolution processes*, *Biometrical J.*, Vol. **23** No. **8**, pp. 811-822 (1981).
- [19] STOYAN GISBERT: *Towards a general-purpose difference scheme for the linear one-dimensional parabolic equation*, in: *Nonlinear Analysis: Theory and Applications : Proceedings of the seventh international summer school; Berlin, August 27 - September 1, 1979*, Akademie-Verlag, Berlin, pp. 297-314 (1981).
- [20] STOYAN GISBERT AND H. BAUMERT, L. LUCKNER AND W. MÜLLER: *A generalized programme package for the simultaneous simulation of transient flow and mattertransport problems in river networks*, in: *Proceedings of the Conference Numerical Modelling of River, Channel and Overlandflow for Water Resources and Environment Applications*, Bratislava (1981).
- [21] STOYAN GISBERT AND H. BAUMERT: *Parameter identification in transverse mixing models of rivers - an inverse problem for a parabolic equation*, *ZAMM-Z. Angew. Math. Me.*, Vol. **61** No. **12**, pp. 617-627 (1981).
- [22] STOYAN GISBERT AND D. STOYAN: *Über die Formen-Maxima-Regel von A.H. Müller, Teil 2*, in: *Freiberger Forschungshefte C 366*, Leipzig, pp. 97-102 (1982).
- [23] STOYAN GISBERT: *On the monotone difference approximation of one-dimensional partial differential equations*, *Soviet Diff. Equations*, Vol. **18** No. **7**, pp. 1257-1270 (1982). (in Russian)
- [24] STOYAN GISBERT: *On maximum principles for matrices, and on conservation of monotonicity. With applications to discretization methods*, *ZAMM-Z. Angew. Math. Me.*, Vol. **62** No. **8**, pp. 375-381 (1982).
- [25] STOYAN GISBERT: *Identification of parameters in systems of spatially one-dimensional partial differential equations*. in: *Conference on Math. Models in the Theory of Heat and Mass Transfer, Proceedings*, Minsk, pp. 137-144 (1982).
- [26] STOYAN GISBERT, H. BAUMERT AND W. MÜLLER: *Modelle von Oberflächengewässern*, *Spectrum*, Vol. **4**, pp. 10-11 (1983).
- [27] STOYAN GISBERT: *Explicit error estimates for difference schemes solving the stationary constant coefficient diffusion-convection-reaction equation*, *ZAMM-Z. Angew. Math. Me.*, Vol. **64** No. **3**, pp. 173-191 (1984).
- [28] STOYAN GISBERT: *On monotone difference schemes for weakly coupled systems of partial differential equations*, in: *Computational mathematics (Warsaw, 1980)*, Banach Center Publ., PWN, Warsaw, Vol. **13**, pp. 33-43 (1984).

- [29] STOYAN GISBERT: *On a difference scheme for the spatially one-dimensional diffusion-convection equation in several coordinate systems*, in: Mathematical models in physics and chemistry and numerical methods of their realization: Proceedings of the Seminar Held in Visegrád, 1982, Teubner Verlag, Leipzig, pp. 142-150 (1984).
- [30] STOYAN GISBERT: *On maximum principles for monotone matrices*, Linear Algebra Appl., Vol. **78**, pp. 147-161 (1986).
- [31] STOYAN GISBERT, H. BAUMERT AND W. MÜLLER: *Numerische Simulation von wind- und durchflussinduzierten Strömungen in Flachgewässern auf der Basis des Ekman-Modells*, Acta Hydrophysica, Vol. **30** No. **1**, pp. 51-67 (1986).
- [32] STOYAN GISBERT: *A programme system for the computation of free-surface flows and of pollution transport*, Hidrol. Közlöny, Vol. **4/5**, pp. 260-266 (1986). (in Hungarian)
- [33] STOYAN GISBERT: *Numerical solution of pipeline system problems by monotone difference approximations*, in: Proceedings ECML, Oberwolfach 1987, Teubner Stuttgart, pp. 195-209 (1988).
- [34] STOYAN GISBERT AND I. MERSICH: *Local scale pollution transport model. Part I. A model for air flow over an inhomogeneous surface*, Időjárás, Vol. **91** No. **6**, pp. 347-360 (1988). (in Hungarian)
- [35] STOYAN GISBERT AND H. BAUMERT: *Operational forecasting of toxic waves in rivers*, Acta Hydrochim. Hydrobiol., Vol. **18** No. **4**, pp. 449-458 (1990).
- [36] STOYAN GISBERT AND A. KÉKESI: *On the programming of the multigrid algorithm*, in: Proceedings of the Computing Center of the Moscow State University, pp. 90-107 (1990). (in Russian)
- [37] STOYAN GISBERT: *Numerical aspects of an environment pollution problem in rivers*, in: Proceedings Conf. Numer. Methods, Sofia 1988, Publ. House Bulg. Acad.Sci., Sofia, pp. 473-481 (1989).
- [38] STOYAN GISBERT: *On the monotone approximation of a two-dimensional equation with non-negative characteristic form*, in: Teubner Series in Mathematics: Numerical treatment of differential equations: Selection of papers presented at the Fifth International Seminar "NUMDIFF-5" held at the Martin-Luther-University Halle-Wittenberg, Stuttgart, Teubner, Leipzig, pp. 259-266 (1991).
- [39] STOYAN GISBERT AND J. NYERS: *Analysis of a dynamical model of the dry evaporator of refrigerators and heat pumps*, Alk. Mat. Lapok, Vol. **683**, pp. 279-285 (1990). (in Hungarian)
- [40] STOYAN GISBERT: *Numerical Solution of Partial Differential Equations*, editor, coauthor of chapters: Parabolic Equations; Nonlinear Equations; The Multigrid Method, Tankönyvkiadó, Budapest (1990). (in Hungarian)
- [41] STOYAN GISBERT AND R. STOYAN: *Colouring the discretization graphs arising in the multigrid method*, Computers & Math. with Appls., Vol. **22** No. **7**, pp. 55-62 (1991).
- [42] STOYAN GISBERT AND J. NYERS: *A hőpumpa elpárolgatójának numerikus szimulációja a fázishatár explicit meghatározásával*, Alk. Mat. Lapok, pp. 143-147 (1991).
- [43] STOYAN GISBERT AND J. NYERS: *The discrete evaporator model's solution for heat pump by means of Gauss-Newton method*, Alk. Mat. Lapok, pp. 86-91 (1992).



- [44] STOYAN GISBERT AND J. NYERS: *A dynamical model adequate for controlling the evaporator of a heat pump*, Intern. J. Refrigeration, Vol. **17** No. **2**, pp. 101-108 (1994).
- [45] STOYAN GISBERT AND H. BAUMERT, B. HELLMANN AND K. PFEIFFER: *Erstellung eines Rechenmodells zum thermischen Schichtungsverhalten in Baggerseen*, in: Forschungsbericht, Hydromod, Hamburg (1996).
- [46] STOYAN GISBERT AND L. GERGÓ: *On a mathematical model of a radiating, viscous, heat conducting fluid: remarks on a paper by J. Förste*, ZAMM-Z. Angew. Math. Me., Vol. **77** No. **5**, pp. 367-375 (1997).
- [47] STOYAN GISBERT AND L. GERGÓ L. AND GY. MOLNÁRKA: *Bevezetés a MATLAB-ba: Programozás, lineáris algebra, grafika*, lecture notes, ELTE, Budapest (1997).
- [48] STOYAN GISBERT: *Convergence and nonnegativity of numerical methods for an integrodifferential equation describing batch grinding*, Computers & Math. with Appls., Vol. **35** No. **12**, pp. 69-81 (1998). (with Cs. Mihálykó and Zs. Ulbert)
- [49] STOYAN GISBERT: *Bevezetés a MATLAB-ba: Numerikus módszerek, grafika, statisztika, eszköztárak (Introduction to Matlab - Numerical Methods, Graphics, Statistics, Toolboxes)*, lecture notes (ed.), Typotex, Budapest (1999).
- [50] STOYAN GISBERT: *Towards discrete Velte decompositions and narrow bounds for inf-sup constants*, Computers & Math. with Appls., Vol. **38** No. **7-8**, pp. 243-261 (1999).
- [51] STOYAN GISBERT: *Optimal iterative Stokes solvers in the harmonic Velte subspace*, in: Report des SFB F013, Universität Linz (1999).
- [52] STOYAN GISBERT:  $-\Delta = -\text{grad div} + \text{rot rot}$  for matrices, with application to the finite element solution of the Stokes problem, East-West J. Numer. Math., Vol. **8** No. **4**, pp. 323-340 (2000).
- [53] STOYAN GISBERT: *On inhomogeneous boundary conditions in the Förste model of a radiating, viscous, heat conducting fluid*, Annales Univ. Sci. Budapest., Sec. Math., Vol. **43**, pp. 125-138 (2000).
- [54] STOYAN GISBERT: *Iterative Stokes solvers in the harmonic Velte subspace*, Computing, Vol. **67** No. **1**, pp. 13-33 (2001).
- [55] STOYAN GISBERT AND L. SIMON: *On the existence of a generalized solution to a three-dimensional elliptic equation with radiation boundary condition*, Appl. of Mathematics, Vol. **46** No. **4**, pp. 241-250 (2001).
- [56] STOYAN GISBERT AND M. DOBROWOLSKI: *Algebraic and discrete Velte decompositions*, BIT Numerical Mathematics, Vol. **41** No. **3**, pp. 465-479 (2001).
- [57] STOYAN GISBERT AND G. STRAUBER AND Á. BARAN: *Generalizations to discrete and analytical Crouzeix-Velte decompositions*, Numer. Linear Algebra Appl., Vol. **11** No. **5-6**, pp. 565-590 (2004).
- [58] STOYAN GISBERT AND H. BURCHARD AND E. DELEERSNIJDER: *Some numerical aspects of turbulence-closure models*, in: Marine Turbulence: Theories, Observations, and Models. Results of the CARTUM Project (eds. H. Baumert, J.Simpson, J. Sündermann), Cambridge Univ. Press, Cambridge, pp. 197-206 (2005).
- [59] STOYAN GISBERT AND Á. BARAN: *Crouzeix-Velte decompositions for higher-order finite elements*, Computers & Math. with Appls., Vol. **51**, pp. 967-986 (2006).

- [60] STOYAN GISBERT AND Á. BARAN: *Gauss-Legendre elements: a stable, higher order non-conforming finite element family*, Computing, Vol. **79**, pp. 1-21 (2007).
- [61] STOYAN GISBERT: *Numerikus Matematika mérnököknek és programozóknak*, Typotex, Budapest (2007). (in Hungarian)
- [62] STOYAN GISBERT: *A Stokes-feladat és a Crouzeix-Velte felbontás*, Alk. Mat. Lapok, Vol. **26**, pp. 179-191 (2009).
- [63] STOYAN GISBERT AND W. HOFMANN: *Wide angle absorbing boundary conditions by minimizations*, Mitt. Math. Gesellschaft Hamburg, Vol. **29**, pp. 143-151 (2010).
- [64] STOYAN GISBERT: *Obituary on Aleksandr Andrejevich Samarskij (1919-2008)*, Annales Univ. Sci. Budapest., Sec. Comp., Vol. **32**, pp. 3-11 (2010).
- [65] STOYAN GISBERT: *On a numerical model for the pianoforte*, Annales Univ. Sci. Budapest., Sec. Comp., Vol. **39**, pp. 415-438 (2013).
- [66] STOYAN GISBERT: *Numerical Methods I*, 1st ed. Typotex, Budapest, 1993, 2nd, corr. ed. 2002, 3rd corr. and extended ed., p. 452 (2012). (in Hungarian, with programs by G. Takó) [numanal.inf.elte.hu/~stoyan/nm1ujkeret.pdf](http://numanal.inf.elte.hu/~stoyan/nm1ujkeret.pdf)
- [67] STOYAN GISBERT: *Numerical Methods II*, 1st ed. Typotex, Budapest, 1995, 2nd, corr. and reworked ed., p. 411 (2012). (in Hungarian, with programs by G. Takó) [numanal.inf.elte.hu/~stoyan/nm2ujkeret.pdf](http://numanal.inf.elte.hu/~stoyan/nm2ujkeret.pdf)
- [68] STOYAN GISBERT: *Numerical Methods III*, 1st ed. Typotex, Budapest, 1996, 2nd ed. 2010, 3rd extended ed., p. 672 (2011). (in Hungarian, with programs by G. Takó) [numanal.inf.elte.hu/~stoyan/nm3ujkeret.pdf](http://numanal.inf.elte.hu/~stoyan/nm3ujkeret.pdf)
- [69] STOYAN GISBERT: *MATLAB 2013-2014: Bevezetés használatába, lineáris algebra, grafika, optimalizálás*, lecture notes, ELTE, Budapest, p. 154 (2014) (rev. 2016). [www.inf.elte.hu/dstore/document/313/mljkeret.pdf](http://www.inf.elte.hu/dstore/document/313/mljkeret.pdf)
- [70] STOYAN GISBERT AND Á. BARAN: *Elementary Numerical Mathematics for Programmers and Engineers*, Springer Cham, Germany, p. 220 (2016).

## A FARKAS GYULA EMLÉKDÍJ 2019. ÉVI DÍJAZOTTJA: GYÖRGYI PÉTER



A Bolyai János Matematikai Társulat Farkas Gyula Emlékdíját 2019-ben Györgyi Péter kapta.

Györgyi Péter 1989-ben született. 2013-ban szerzett alkalmazott matematikusi diplomát az Eötvös Loránd Tudományegyetem Természettudományi Karán, „kitüntetéses” minősítéssel. 2013 és 2016 között az ELTE TTK Matematika Doktori Iskola ösztöndíjas hallgatója, 2016-tól pedig a SZTAKI tudományos munkatársa. 2018-ban summa cum laude minősítéssel védte meg PhD-értekezését, témavezetője Kis Tamás volt. 2016 és 2019 között az MTA Fialat Kutatói Ösztöndíjában részesült.

2018-ban elnyerte a SZTAKI publikációs díját, és ugyanebben az évben a Rómában rendezett 16th International Conference on Project Management and Scheduling *c.* konferencián második helyezést ért el a doktoranduszi cikkekre kiírt pályázaton. 2019-ben Bolyai János Kutatási Ösztöndíjat nyert.

Eddigi eredményeit főleg az ütemezéselmélet területén érte el. Elsősorban nem megújuló erőforrásos ütemezési problémák bonyolultságát és approximálhatóságát vizsgálta, ebben a témában jelent meg a legtöbb tudományos dolgozata. Ezen túlmenően részt vett egy új, egzakt módszer kidolgozásában a feladat egy variánsának megoldására. Egy másik témája az online jármű hozzárendelési feladat, valamint a konfliktusmentes irányítás gráfokon. Eredményei felhasználásra kerültek a SZTAKI győri, Ipar 4.0 kísérleti laborjában, az autonóm járműflotta irányításában.

Eddig összesen 9 tudományos dolgozata jelent meg – témavezetőjével mint egy cikk kivételével egyedüli társszerzőjével – nemzetközi, lektorált szakfolyóiratban, amelyek közül 2 cikk Q2-es, 2 cikk Q1/D1-es, 5 cikk pedig D1-es besorolású, ezen belül egy dolgozata egyszerűs. 2019-ben meghívott előadó volt a leideni Lorentz Center „Scheduling Meets Fixed-Parameter Tractability” megnevezésű workshopján.

Részt vett a 2016-ban Budapesten megrendezett European Chapter on Combinatorial Optimization Conference ECCO 2016 nemzetközi konferencia szervezésében (<http://ecco2016.weebly.com/committees.html>). Számos rangos nemzetközi szakfolyóirat számára bírál rendszeresen kéziratokat.

A díjazott öt legfontosabb publikációja:

### Hivatkozások

- [1] P. GYÖRGYI, T. KIS: *Approximation schemes for parallel machine scheduling with non-renewable resources*, European Journal of Operational Research, Vol. **258** No. **1**, pp. 113-123 (2017). DOI: [10.1016/j.ejor.2016.09.007](https://doi.org/10.1016/j.ejor.2016.09.007)
- [2] P. GYÖRGYI: *A PTAS for a resource scheduling problem with arbitrary number of parallel machines*, Operations Research Letters, Vol. **45** No. **6**, pp. 604-609 (2017). DOI: [10.1016/j.orl.2017.09.007](https://doi.org/10.1016/j.orl.2017.09.007)
- [3] P. GYÖRGYI, T. KIS: *Minimizing the maximum lateness on a single machine with raw material constraints by branch-and-cut*, Computers and Industrial Engineering, Vol. **115**, pp. 220-225 (2018). DOI: [10.1016/j.cie.2017.11.016](https://doi.org/10.1016/j.cie.2017.11.016)
- [4] P. GYÖRGYI, T. KIS: *Minimizing total weighted completion time on a single machine subject to non-renewable resource constraints*, Journal of Scheduling, Vol. **22** No. **6**, pp. 623-634 (2019). DOI: [10.1007/s10951-019-00601-1](https://doi.org/10.1007/s10951-019-00601-1)
- [5] P. GYÖRGYI, T. KIS: *A probabilistic approach to pickup and delivery problems with time window uncertainty*, European Journal of Operational Research, Vol. **274** No. **3**, pp. 909-923 (2019). DOI: [10.1016/j.ejor.2018.10.031](https://doi.org/10.1016/j.ejor.2018.10.031)

GYÖRGYI PÉTER

Számítástechnikai és Automatizálási Kutatóintézet  
Mérnöki és Üzleti Intelligencia Kutatólaboratórium  
1111 Budapest, Kende u. 13-17.  
gyorgyi.peter@sztaki.hu

## KREDITRENDSZERŰ KÉPZÉSEK MINTATANTERVEINEK ÉS ELŐTANULMÁNYI HÁLÓINAK ELEMZÉSE A HAZAI MATEMATIKA ALAPSZAKOK PÉLDÁJÁN

BERGMANN JÚLIA, MOLONTAY ROLAND,  
SZABÓ MIHÁLY, SZEKRÉNYES DÓRA LAURA

Dolgozatunkban egyrészt a kreditrendszerű képzések előtanulmányi hálóinak jellemzésére és összehasonlíthatóságára mutatunk néhány módszert, majd ezeket a magyarországi matematika alapképzéseken szemléltetjük. Áttekintjük a tantervi hálók elemzésének irodalmát, illetve bemutatunk egy teljesítési adatokra alapuló valószínűségi modellt is. Ezen módszerek segítségével olyan kérdésekre kaphatunk választ, hogy egy adott tantervben melyek a legfontosabb, leghangsúlyosabb tantárgyak, illetve az előtanulmányi háló topológiája hogyan hat a képzés várható teljesítési idejére. Másrészt ismertetünk néhány felvételi statisztikát az elmúlt évekből, többek között a matematika alapszakokra felvett hallgatók létszámára, illetve felvételi pontszámaira vonatkozóan.

### 1. Bevezetés

„Életünk legszebb évei”, így emlékezünk vissza az egyetemi évekre, ugyanakkor a diákok izgatottan várják felsőfokú tanulmányaik befejezését és a diploma megszerzését. A diploma megszerzéséig vezető utat a képzések mintatanterve, előtanulmányi rendje jelöli ki. Ebben a cikkben egyetemi képzések előtanulmányi hálóját elemezzük gráfelméleti és valószínűségszámítási eszközökkel. Az előtanulmányi rend egy irányított gráffal írható le, ahol a csúcsok a tantárgyaknak felelnek meg, és irányított él fut két csúcs között, ha az egyik tantárgy előkövetelménye a másiknak.

Az oktatásszervezési kérdések kvantitatív módszereken alapuló és adatvezérelt támogatására egyre nagyobb figyelem irányul úgy az oktatási kormányzat és a felsőoktatási intézmények vezetői részéről, mint a kutatók részéről itthon és külföldön egyaránt [4, 9, 11, 12, 18].

Az elmúlt években több cikk is született előtanulmányi hálózatok elemzéséről, melyről jó áttekintést ad Wigdahl munkája [27]. Slim és szerzőtársai felismerték,

hogy a hallgatók egyetemi előrehaladásában az előtanulmányi rend nagy szerepet játszik, ők a mintatantervek elemzésére a komplex hálózatok elméletéből és a gráfelméletből ismert módszereket használják [21].

Ugyanezen szerzők elemzik a mintatantervek hatékonyságát és tanulmányi előmentelre vonatkozó hatásukat gráfelméleti módszerekkel [28]. Az előtanulmányi rendből kiindulva lineáris regresszió és Markov-hálózatok segítségével prediktív analitikai elemzést végeznek a hallgatók teljesítményére vonatkozóan [20], továbbá hálózatelméleti eszközökkel vizsgálják a kurzusok<sup>1</sup> előtanulmányi rendben betöltött szerepét az Új-mexikói Egyetem képzéseiben [21]. Szintén hálózatelméleti módszertannal vizsgálja a mintatanterveket Aldrich [2] és Lightfoot [17].

Több tanulmány született a mintatantervek egyszerű és hatékony vizualizációját, illetve az előtanulmányi rendben való eligazodást, a kurzusfelvétel megkönnyítését segítő rendszerek tervezésével kapcsolatban is [1, 3, 15].

A mintatantervek és előtanulmányi hálók vizsgálatának másik megközelítési módja a hallgatói folyamatok modellezése. Számos tanulmányban vizsgálták hallgatói áramokat szimulálva, hogy a mintatantervek átszervezésének milyen hatása lenne a hallgatók előrehaladására vonatkozóan [19, 23, 26]. A mintatanterv hallgatói teljesítményre vonatkozó hatását vizsgálta Jansen és van der Hulst is [14, 25], a hallgatói folyamatok modellezésekor figyelembe véve a tanulók korát, nemét, tanulmányi átlagát és a középiskolás eredményeit. Rahim és szerzőtársai Markov-lánc alapú modellezést használnak posztgraduális képzések hallgatói áramainak elemzésére [24]. A hallgatók előrehaladásának és végzési rátájának jobb nyomonkövethetősége érdekében a hallgatók áramlását vizualizálták Sankey-diagramok segítségével friss cikkükben Horváth és szerzőtársai [13].

## 2. Mintatantervi gráfok elemzése

Ebben a fejezetben ismertetjük a mintatantervek elemzéséhez általunk használt matematikai eszközöket. A továbbiakban feltételezzük, hogy az olvasó tisztában van az alapvető gráfelméleti és valószínűségszámítási fogalmakkal.<sup>2</sup>

### 2.1. Mintatantervi gráf

A kreditrendszerű képzésben a tantárgyak felvételének feltétele, hogy az adott tantárgy (általában korábbi félév(ek)ben szereplő) előzetes követelményeit (döntően másik tantárgya(ka)t) a hallgató már teljesítse. A kreditrendszerű mintatanterv ábrázolásának legszemléletesebb módja az ún. előtanulmányi háló. A képzések előtanulmányi rendje természetes módon jellemezhető gráfokkal, pontosabban irányí-

<sup>1</sup>Dolgozatunkban a tantárgy és kurzus megnevezéseket szinonimaként használjuk.

<sup>2</sup>Ha mégsem lenne így, [5, 16] forrásokat ajánljuk.

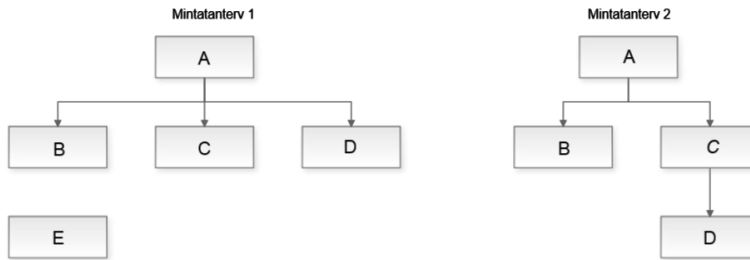
tott körmentes gráfokkal (DAG). A gráf csúcsai a tantárgyaknak felelnek meg, az irányított élek pedig az előkövetelményi rendszert jelölik. Jelölje  $M_{i,j}$   $n \times n$ -es mátrix a gráf szomszédossági mátrixát, ahol  $n$  a vizsgált kurzusok száma, és  $M_{i,j} = 1$  pontosan akkor, ha az  $i$  kurzus előfeltétele  $j$  kurzusnak, vagyis  $j$  tantárgy nem vehető fel  $i$  sikeres teljesítése nélkül. Ez a megközelítés nem veszi figyelembe az ún. gyenge előkövetelmény lehetőségét. További egyszerűsítő feltételezéseinkről a 4. és 5. fejezetekben írunk bővebben.

Egy tantárgy fontossága, képzésben betöltött szerepe nagyon sokféleképpen definiálható, és számos tényezőtől függhet. Mi a következőkben ezt a „fontosságot” az előtanulmányi hálóban betöltött szerepek alapján kvantifikáljuk. Mérőszámokat mutatunk mind az egyes tantárgyakra, mind az előtanulmányi háló egészének struktúrájára, sűrűségére vonatkozóan. Először az előtanulmányi háló topológiájából származtatott mutatókat ismertetjük, később bevezetünk néhány új, valószínűségi alapokon nyugvó mérőszámot.

## 2.2. Topológikus mutatók

### Késleltetési tényező (delay factor)[22]

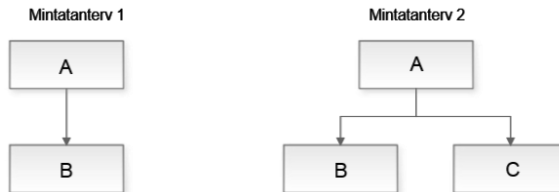
A késleltetési tényező megmutatja, hogy egy kurzus sikertelen teljesítése biztosan maga után vonja-e a képzési idő megnövekedését. A késleltetési tényező értéke legyen a reciproka annak a számnak, ahányadszori bukás már biztosan félévvesztést okoz a képzés elvégzését illetően. Tehát egy adott kurzus késleltetési faktora 1, ha a kurzus nem teljesítése egy félévnyi csúszással jár, és  $1/3$ , ha kétszer „büntetlenül” bukhatunk, de a harmadik nem teljesítés már biztosan félévvesztést okoz. A fogalmat az 1. ábrán látható két rövid, három féléves mintatanterv segítségével szemléltetjük. Vegyük észre, hogy míg a *Mintatanterv 1* esetén az A tantárgy sikertelen teljesítése után még mindig pozitív valószínűséggel végezhető el a program időben, azaz három félév alatt (feltételezve, hogy minden tantárgy meghirdetésre kerül minden félévben), addig a második képzésen a B tantárgy kivételével minden sikertelenség automatikus csúszást eredményez. Ilyen esetekben mondhatjuk, hogy *A, C és D jobban késleltetnek, mint B*, hiszen A, C és D tantárgy késleltetési tényezője 1, B tantárgyé pedig  $1/2$ .



1. ábra. Késleltetési tényező.

### Blokkoló tényező (blocking factor)[22]

Természetes gondolat, hogy egy tantárgy minél több kurzusnak előkövetelménye, annál fontosabb a képzésen. Formálisan azt mondhatjuk, hogy egy kurzus blokkoló tényezőjének értéke  $n \in \mathbb{N}$ , ha pontosan  $n$  darab leszármazottja van a gráfban, azaz a csúcsból pontosan  $n$  másik csúcsba vezet irányított út. Vizsgáljuk a 2. ábrán feltüntetett mintatanterveket. Vegyük észre, hogy az A tantárgy különböző fontossággal bír a két mintatanterv esetén. Az első esetben csak egy, míg a másodikban két tantárgynak is előkövetelménye, tehát utóbbiban fontosabbnak számít.



2. ábra. Blokkoló tényező

### Keveredés (dissimilar mixing)[22]

A kurzusokat számos szempont szerint csoportosíthatjuk, például témaköre vagy a tantárgyat oktató tanszék szerint. Ezen csoportosítás ismeretében vizsgálhatjuk, hogy az egyes tantárgycsoportok mekkora mértékben építenek más csoportok kurzusaira, azaz mennyi él fut egy csoportba a többiből. A gráf irányítottságából adódóan ez a mérőszám nem szimmetrikus. A keveredés mértékének



meghatározására vezessük be a következő mennyiséget:

$$e_{rs} = \frac{1}{m} \sum_{i,j} M_{ij} \delta(c_i, r) \delta(c_j, s),$$

ahol  $r$  és  $s$  tantárgytípusok (tanszékek),  $c_i$  jelöli az  $i$  tantárgy típusát,  $m$  az élek száma és  $\delta$  a Kronecker-delta függvény. Azt mondhatjuk, hogy minél nagyobb az  $e_{rs}$  érték, annál szorosabban kapcsolódnak az  $s$  típusú tantárgyak az  $r$  típusú tantárgyakhoz.

### **Kölcsönös központiség (betweenness centrality)[17]**

Az ún. kölcsönös központiség segít megtalálnunk a gráf különböző részei között *gócpon*tokat formáló csúcsokat. Egy  $v$  csúcs (azaz esetünkben kurzus) kölcsönös központiségén a következőt értjük:

$$b_v = \sum_{s \neq v, t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

ahol  $\sigma_{st}$  az összes  $s$ -ből  $t$ -be menő legrövidebb utak száma,  $\sigma_{st}(v)$  pedig azon ilyen utak, amelyek áthaladnak  $v$  csúcson. Irányított gráf lévén irányított utakról beszélünk.

A kölcsönös központiség szofisztikált mérték, és különböző módosításait a hálózatelméletben széleskörben alkalmazzák [6], [8]. Az előtanulmányi hálók ugyanakkor speciális gráfok, a DAG-tulajdonságuk miatt a fenti mérték kevésbé jól használható fontosságot rendel a kurzusokhoz. A kölcsönös központiség minden első féléves tantárgyhoz triviálisan 0-t fog rendelni, függetlenül attól, hány másik tantárgynak előkövetelménye.

### **Összefüggő komponensek (connected components)[2]**

Az összefüggő komponens olyan tovább nem bővíthető részgráf, melynek bármely két csúcsát út köti össze. A mintatantervek irányított gráfok, a továbbiakban a gyengén összefüggő komponenseket tekintjük (ezek az irányítatlan gráfban összefüggő komponensek). Mintatantervi gráfoknál az ilyen komponensek egy-egy elkülönített tudományterületet jelölnek. Érdekes kérdés, hogy a különböző témák hogyan kapcsolódnak egymáshoz, mennyire építenek egymásra, ezt hivatott mérni a már említett keveredés.

### **Utak (paths)[28]**

Egy mintatantervi gráfban a leghosszabb út az előkövetelmények leghosszabb láncát jelöli. A hosszú utak olyan kurzusokat tartalmaznak, amik nagy hatással lehetnek a végzési időre. Ezen csúcsok késleltetési faktora tipikusan magas (azaz 1

vagy ahhoz közeli), hiszen ezek azok, amelyek sikertelen teljesítése könnyen okozhat csúszást a hallgató életében. Egy hatfélèves képzés esetén egy 4-hosszú út, ami 5 csúcsból és 4 élből áll, már hosszúnak tekinthető.

### Üvegnyakak (bottlenecks)[28]

Vizsgálva az egyes csúcsok ki- és befokát (jelölés:  $\Delta^+(v), \Delta^-(v)$ ), előfordulhatnak kiemelkedően sok szomszéddal rendelkező kurzusok. Egy  $i$  kurzusra akkor mondjuk, hogy üvegnyak, ha  $\Delta^+(v) > a$ , vagy  $\Delta^-(v) > a$ , vagy  $\Delta^+(v) + \Delta^-(v) > b$  valamilyen  $a, b \in \mathbf{N}$  rögzített konstansra. Formálisan, egy programban pontosan

$$b_n(G) = \sum_{v \in V(G)} \mathbb{1} [(\Delta^+(v) > a) \vee (\Delta^-(v) > a) \vee (\Delta^+(v) + \Delta^-(v) > b)],$$

darab üvegnyak van, ahol  $a$  és  $b$  értékét a képzés hosszától és a gráf sűrűségétől függően határozzuk meg. Ezen kurzusok sikertelen teljesítése könnyen félévvesztést eredményezhet.

### 2.3. Sztochasztikus mutatók

Az előbbi mérőszámok hiányossága, hogy csupán a gráf topológiáján alapsznak, és nem veszik figyelembe a csúcsok áteresztő képességét, azaz, hogy a kurzusok elvégzési valószínűsége között nagy különbségek lehetnek. Az alábbiakban bevezetünk olyan új mérőszámokat, melyek az előbbi problémákat hivatottak áthidalni.

#### Várható végzési idő

Egy tantervet az jellemez legjobban, hogy mennyi idő alatt lehet (várhatóan) elvégezni. Ezt alapvetően két dolog határozza meg: az előtanulmányi háló struktúrája és az egyes tantárgyak elvégzési valószínűségei. Az alábbiakban a gráfot a – modellünk szerinti – végzési idejével, mint valószínűségi változóval, és annak várható értékével jellemezzük. Célunk, hogy meghatározzuk a végzési idő ( $X$ ) valószínűségi változójának súlyfüggvényét:

$$p(x) = \mathbb{P}(X = x).$$

Ennek alapján a várható végzési idő könnyen kiszámítható:

$$\mathbb{E}(X) = \sum_x x \cdot p(x).$$

#### Áteresztő hatás

A várható végzési idő tárgyalása után természetesen adódik a kérdés, hogy az egyes tantárgyak milyen hatással vannak a végzési időre. A kérdés megválaszolásához az alábbi egyszerűsített modellt tekintjük. Legyen adott egy virtuális reprezentatív hallgató, aki az előtanulmányi rend szerint igyekszik haladni

tanulmányaival, és az  $i$ -edik tantárgyat  $p_i$  valószínűséggel végzi el (függetlenül a múlttól). Egy tantárgy elvégzését csak akkor kísérheti meg, ha az összes előkövetelményét már sikeresen teljesítette, illetve a tantárgy mintatanterv szerinti féléve nem előzi meg az aktuális félévet. A modellünk reprezentatív hallgatója az aktuális félévben az összes olyan tantárgy elvégzését megkísérli, amit a mintatanterv és az előtanulmányi rend megenged számára.

Ebben a modellben a reprezentatív hallgató várható végzési ideje a tantárgyak teljesítési valószínűségeitől, illetve a mintatanterv topológiájától függ. Tétélezzük fel, hogy ez az  $\mathbb{E}(X) = f(p_1, \dots, p_n)$  függvénnyel írható fel.

A függvényformát ismerve meghatározhatjuk, hogy az egyes tantárgyak teljesítési valószínűségeiben fellépő kis változás milyen hatással van a képzés várható teljesítési idejére, azaz tekinthetjük a következő parciális deriváltat:

$$d_i = \frac{\partial f(p_1, \dots, p_n)}{\partial p_i}.$$

Hasonlóan érdekes érték a képzés teljesítésének kurzusok elvégzési ideje szerinti rugalmassága (elaszticitása):

$$d_i^e = \frac{\partial \log f(p_1, \dots, p_n)}{\partial \log p_i}.$$

A reprezentatív hallgató tantárgyteljesítési valószínűségeit valós historikus adatokból a következőképpen tudjuk becsülni:

$$p_i = \mathbb{P}(\text{az } i \text{ kurzus sikeres teljesítése}) \approx \frac{\text{sikeresen teljesítő hallgatók száma}}{\text{kurzust felvett hallgatók száma}}.$$

A 3. fejezetben még részletesebben visszatérünk a fent ismertetett valószínűségi modellre, az áteresztő hatásra és annak kiszámítására.

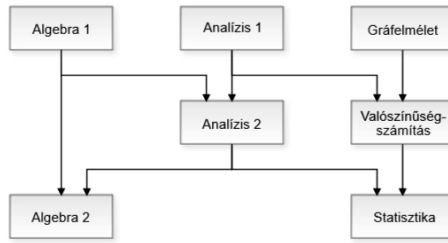
### 3. Végzési idők kiszámítása

A tantervek végzési idejének kiszámítása bár látszólag egyszerű és könnyen érthető, gyorsan meglehetősen bonyolulttá válik. Egy nagyon rövid mintatanterven fogjuk bemutatni a problémát (3. ábra).

Modellünkben egy tantárgy első felvételétől számított, félévekben mért elvégzési idejét egy  $p$  paraméterű (optimista<sup>3</sup>) geometriai valószínűségi változó adja meg. Ez azt jelenti, hogy ebben a megközelítésben egy tantárgy többedszeri felvétele nem változtatja annak elvégzési valószínűségét. Továbbá feltesszük, hogy az egyes tantárgyak végzési idejei is függetlenek egymástól.

---

<sup>3</sup>matematikai értelemben



3. ábra. Egy képzeletbeli mintatanterv.

Tantárgy	Valószínűségi változó	Eloszlás
Algebra 1	$X_1$	$\sim \text{Geom}(p_1)$
Analízis 1	$X_2$	$\sim \text{Geom}(p_2)$
Gráfelmélet	$X_3$	$\sim \text{Geom}(p_3)$
Analízis 2	$Y_1$	$\sim \text{Geom}(q_1)$
Valószínűség-számítás	$Y_2$	$\sim \text{Geom}(q_2)$
Algebra 2	$Z_1$	$\sim \text{Geom}(r_1)$
Statisztika	$Z_2$	$\sim \text{Geom}(r_2)$

1. táblázat. A képzeletbeli mintatanterv tárgyai.

Ebben az esetben könnyű megállapítani, hogy az első féléves kurzusok elvégzésének várható ideje rendre  $\frac{1}{p_1}$ ,  $\frac{1}{p_2}$  és  $\frac{1}{p_3}$ . Tekintsük most az Analízis 2 c. kurzust! Ennek előkövetelménye az Algebra 1 és az Analízis 1. Könnyen látszik, hogy ekkor az Analízis 2 beiratkozástól számított elvégzési ideje a következő valószínűségi változóval írható le:  $\max\{X_1, X_2\} + Y_1$ . Ezt továbbgondolva, minden tantárgyhoz felírhatjuk a tantárgy beiratkozásától számított elvégzési idejét, amit a 2. táblázatban foglaltunk össze. A fenti valószínűségi változók súlyfüggvényé-

Tantárgy	Elvégzési idő valószínűségi változója
Algebra 1	$X_1$
Analízis 1	$X_2$
Gráfelmélet	$X_3$
Analízis 2	$\max\{X_1, X_2\} + Y_1$
Valószínűség-számítás	$\max\{X_2, X_3\} + Y_2$
Algebra 2	$\max\{X_1, \max\{X_1, X_2\} + Y_1\} + Z_1$
Statisztika	$\max\{\max\{X_1, X_2\} + Y_1, \max\{X_2, X_3\} + Y_2\} + Z_2$

2. táblázat. A képzeletbeli mintatanterv kurzusaihoz tartozó valószínűségi változók.

nek és várható értékének analitikus meghatározása még kis mintatanterv esetén is problémás. Általánosan igaz, hogy az  $X_1, \dots, X_n$  független  $p$  paraméterű geometriai változók maximumának várható értéke  $\sum_{k=0}^{\infty} \left(1 - (1 - q^k)^n\right)$ , ahol  $q = 1 - p$  [10].

Különböző valószínűségeket feltéve ez az érték  $\sum_{k=0}^{\infty} \left(1 - \prod_{i=1}^n (1 - q_i^k)\right)$ . Tovább lépve a harmadik félévre, még nehezebb probléma elé kerülünk, hiszen itt már nem azonos eloszlású valószínűségi változók maximumának várható értékét keressük. Az analitikus megközelítés helyett ezért inkább Monte Carlo-szimulációval érdemes dolgozni [7].

#### 4. Egy előtanulmányi hálózat vizsgálata

Az előző fejezetben ismertettünk néhány szempontot, ami jó alapul szolgál ahhoz, hogy egy egyetemi előtanulmányi gráfot vizsgáljunk. Ebben a fejezetben az Eötvös Loránd Tudományegyetem (továbbiakban: ELTE) matematika BSc képzését fogjuk közelebbről szemügyre venni.

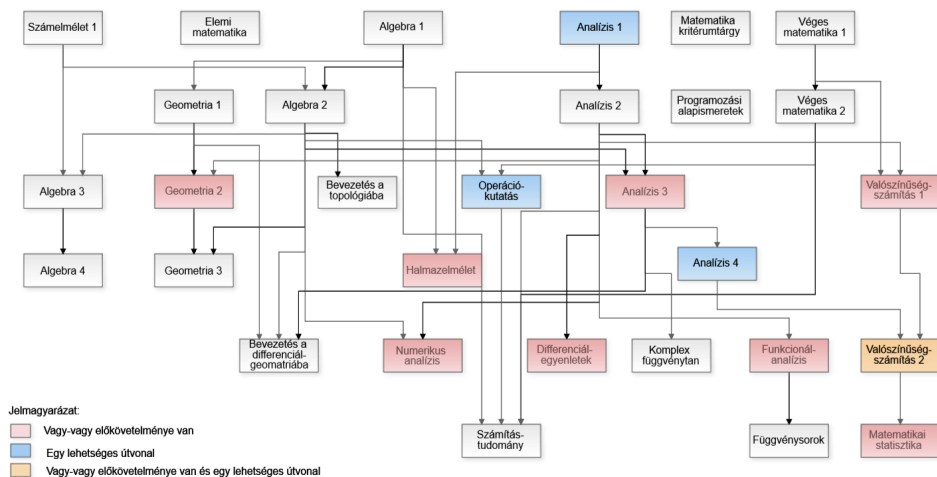
A 4. ábra mutatja az ELTE matematika BSc képzés elméleti specializációjának előtanulmányi gráfját. Néhány egyszerűsítő feltétellel éltünk a vizsgálat során a könnyebb kezelhetőség és jobb átláthatóság érdekében. Csak a képzés kötelező kurzusait vizsgáltuk, a kötelezően, illetve a szabadon választható tantárgyakat elhagytuk, emellett nem vettük figyelembe az egyetem úgynevezett *vagy-vagy* előkövetelményi rendszerét. Ez utóbbi azt jelenti, hogy némely tantárgynak nincs egyértelmű előkövetelménye, hanem két különböző kurzus bármelyikének teljesítése esetén felveheti a hallgató az adott tantárgyat. A következő elemzésben kijelöltünk egy lehetséges utat, amivel teljesíthető a képzés, az ábrán késsel jelöljük azon tantárgyakat, amelyek helyett más is választható lenne, pirossal azokat, amelyek *vagy-vagy* előkövetelményűek.

Már első ránézésre is látszik, hogy az előtanulmányi háló elég sűrű, azaz sok az előkövetelményi kapcsolat a tantárgyak között. Láthatjuk, hogy nincsenek nagy különálló összefüggő komponensek, és izolált pontból is viszonylag kevés van. A két leginkább üvegynek szerepet betöltő tárgynak az Algebra 2 és az Analízis 2 című kurzusok bizonyulnak, befokuk rendre 2 és 1, míg mindkét tantárgy kifoka 6. Számos hosszú út található a gráfban, sőt, van az egész képzést végigkísérő öt hosszú út is: Analízis 1 - Analízis 2 - Analízis 3 - Analízis 4 - Valószínűségszámítás 2 - Matematikai statisztika. Tehát ha ezen tantárgyak bármelyikét sikertelenül végzi egy hallgató, az feltétlen megnöveli a képzés teljesítéséhez szükséges félévek számát.

A csúcsokat vizsgálva a kölcsönös központosság szempontjából a legmagasabb értékeket a 3. táblázatban látjuk.

Kurzus	Kölcsönös központiság
Algebra 2	1,00
Analízis 3	0,64
Analízis 2	0,54

3. táblázat. Az ELTE legnagyobb kölcsönös központiságú kurzusai.



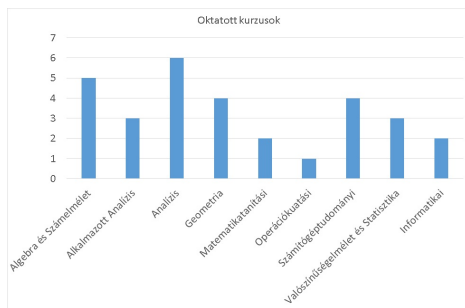
4. ábra. Az ELTE elméleti matematika specializáció képzésének mintatantervi gráfja.

A leghosszabb útra, valamint a ki- és befokokra tett megállapításainkat összegezve kijelenthetjük, hogy *kizárólag* az előtanulmányi háló struktúráját vizsgálva az Algebra 2, az Analízis 3 és az Analízis 2 kurzusok számítanak a legfontosabb tantárgyaknak ezen a képzésen.

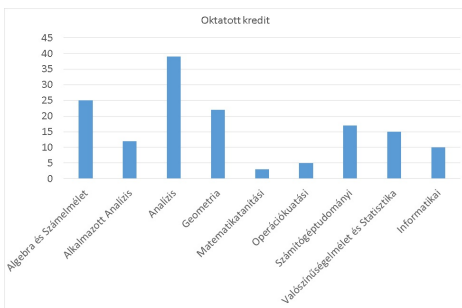
A tantárgyak áteresztő hatását tekintve – valós teljesítési arányok hiányában – az összes tantárgynál azonos, 0,8-as teljesítési arányt feltételezve, szimulációnk alapján a következő tantárgyak teljesítési arányának (*ceteris paribus*) növelése csökkenti leginkább a képzés várható elvégzési idejét: Analízis 3 (0,15), Analízis 4 (0,14), Valószínűség-számítás 2 (0,14). A zárójelbe írt szám azt jelenti, hogy ha az adott tantárgy teljesítési valószínűségét 0,8-ról 0,9-re növeljük (*ceteris paribus*), akkor a képzés félévekben mért várható elvégzési ideje a zárójelbe írt számértékkel csökken a 10 000 reprezentatív hallgatóra futtatott szimuláció alapján. Felhívjuk a figyelmet, hogy ezeknek a tantárgyaknak a késleltetési tényezője is magas (1 értékű).

Pontosabb képet akkor kaphatnánk, ha a kurzusok valós teljesítési arányait ismerve végeznénk el az elemzést.

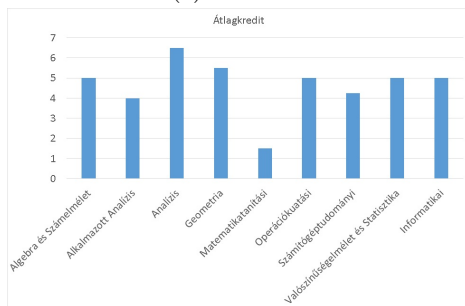
Megvizsgáljuk az egyes tanszékek képzésen betöltött szerepét, súlyát három különböző megközelítésből. Az 5(a) ábra mutatja, hogy összesen az egyes tanszéknek mennyi kötelező órájuk van, míg az 5(b) ábrán az oktatott kreditek száma látszik. Az 5(c) ábra szemlélteti az oktatott órák átlagos kreditértékét.



(a) kurzusok



(b) összkredit



(c) kreditérték

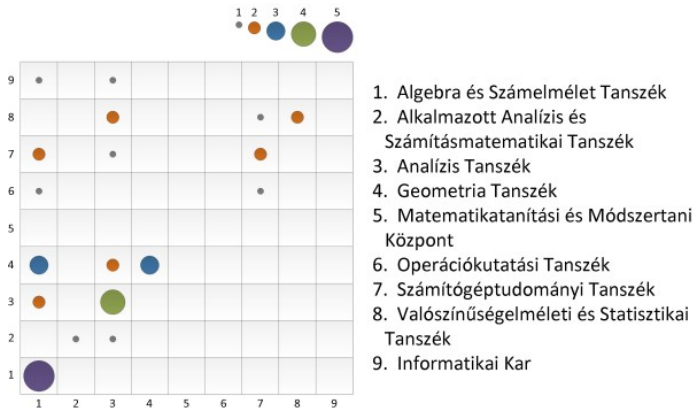
A megjelenített tanszékek balról jobbra:

1. Algebra és Számelmélet Tanszék
2. Alkalmazott Analízis és Számításmatematikai Tanszék
3. Analízis Tanszék
4. Geometria Tanszék
5. Matematikatanítási és Módszertani Központ
6. Operációkutatási Tanszék
7. Számítógéptudományi Tanszék
8. Valószínűségelméleti és Statisztikai Tanszék
9. Informatikai Kar

5. ábra. Tanszékek megoszlása oktatott kurzusok száma (a), összkreditek száma (b), illetve átlagos kreditérték (c) alapján.

Az előző fejezetben definiáltuk az ún. keveredést, amely megállapítja a tanszékek közti kapcsolatot abban az értelemben, hogy melyik tanszék tantárgyai mennyire építenek előkövetelményként a másik tanszék tantárgyaira. Ezt a 6. ábra szemlélteti, a tengelyeken a nyolc tanszék és egy "külsős" kar szerepel, a metszéspontokban lévő körök mérete azzal arányos, hogy a vízszintes tanszék tantárgyai hány esetben előkövetelményei a függőleges tanszéknek. (Tulajdonképpen ez a korábban említett  $e_{r-s}$  érték a normálástól eltekintve.)

Megjegyezzük, hogy a Matematikatanítási és Módszertani Központban tartó sor, illetve oszlop ürességének oka a központ által oktatott tantárgyak (Elemi matematika és Matematika kritériumtárgy) izoláltsága az előtanulmányi hálóban (4. ábra).



6. ábra. Tanszékek kapcsolata.

## 5. Hazai matematika alapszakok tanterveinek összehasonlítása

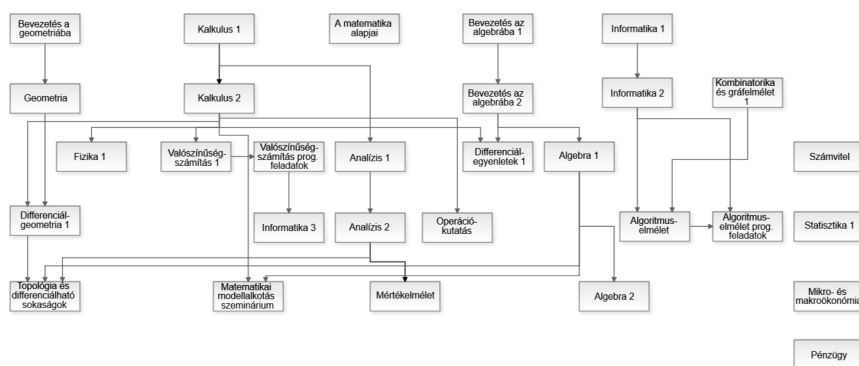
Ebben a fejezetben a Magyarországon akkreditált matematika alapképzések tantervét és előtanulmányi hálóját hasonlítjuk össze különböző szempontok alapján. A Budapesti Műszaki és Gazdaságtudományi Egyetem (BME), az Eötvös Loránd Tudományegyetem (ELTE), a Debreceni Egyetem (DE), a Pécsi Tudományegyetem (PTE) és az Eszterházy Károly Egyetem (EKE) matematika alapszakjait vizsgáljuk. A Szegedi Tudományegyetem (SZTE) nyilvánosan elérhető mintatanterv hiányában nem került az általunk elemzett egyetemek közé. Az elemzéseink során elsősorban a képzések elméleti specializációjára összpontosítottunk a könnyebb összehasonlíthatóság érdekében. A képzések nyújtotta további specializációs lehetőségekről a 7 függelékben lesz bővebben szó. Továbbra is csak a kötelező tantárgyakat vizsgáljuk, eltekintünk a vagy-vagy kapcsolatoktól, helyette egy lehetséges utat vizsgálunk, nem teszünk különbséget gyenge és erős előkövetelmények<sup>4</sup> között, és nem tekintjük külön kurzusnak az adott tantárgyból az előadást és a gyakorlatot.

### 5.1. Előtanulmányi gráfok

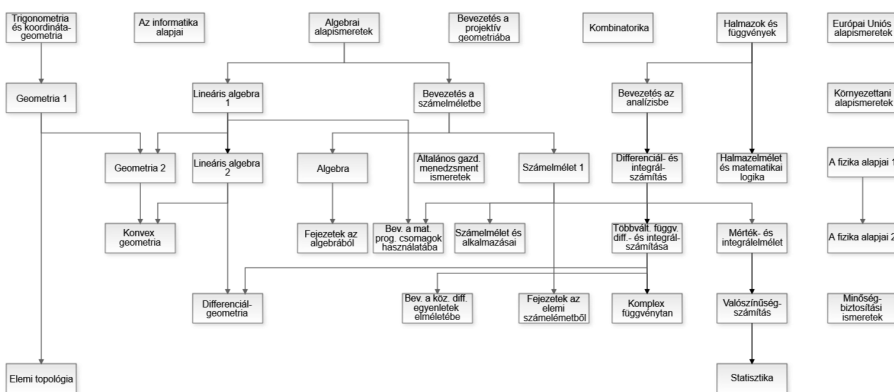
Ebben a részben szemléltetjük az egyes képzések gráfjait. Az ELTE előtanulmányi hálóját már korábban a 4. ábrán bemutattuk.

<sup>4</sup>Gyenge előkövetelmény, ha csupán a gyakorlat vagy labor teljesítése elég a továbbhaladáshoz az előadás sikeres teljesítésétől függetlenül, illetve ha egyazon félévben végezhető az előkövetelmény és a ráépülő tantárgy.





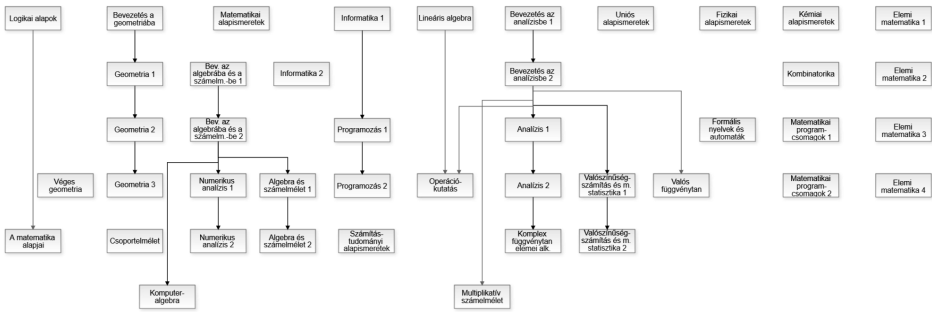
7. ábra. A BME elméleti matematika specializáció képzésének mintatantervi gráfja.



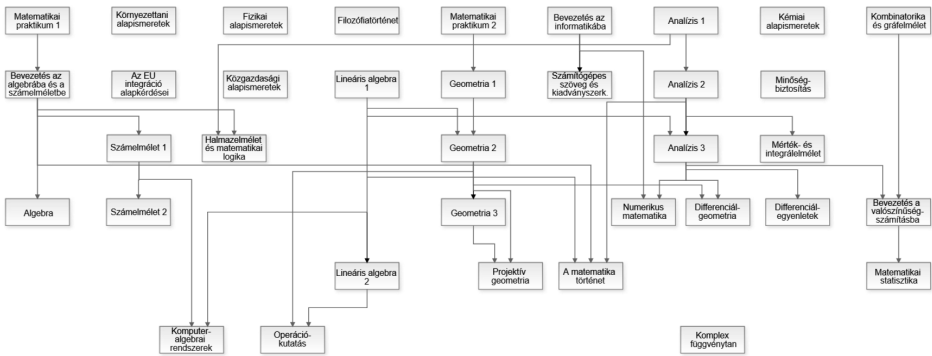
8. ábra. A DE elméleti matematika specializáció képzésének mintatantervi gráfja.

## 5.2. A végzés várható ideje

Mind a hallgatók, mind az egyetemek vezetősége érdekeltek abban, hogy megbecsüljék, a diákok várhatóan mikor szereznek diplomát. Ennek elméleti hátterét a 2. fejezetben, gyakorlati megvalósításának módját pedig a 3. fejezetben tárgyaltuk. Szimulációnkat 10 000 virtuális reprezentatív hallgatóra végeztük. Valós statisztikai adatok hiánya miatt magunk választunk valószínűségeket a kurzusokhoz, jelen vizsgálatban minden tantárgy sikeres teljesítésének valószínűsége 0,8.



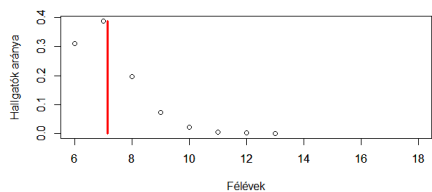
9. ábra. A PTE elméleti matematika specializáció képzésének mintatantervi gráfja.



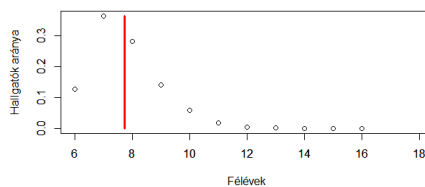
10. ábra. Az EKE elméleti matematika specializáció képzésének mintatantervi gráfja.

Amennyiben historikus adatokból a tényleges becslt valószínűségek rendelkezésre állnak, a módszer sokkal pontosabb eredményre vezet. További egyszerűsítő fel-tételünk, hogy minden tantárgy minden félévben indul, a különböző tantárgyak teljesítései függetlenek egymástól, a hallgatóknak nincs passzív félévük, a hallgatók hamarabb nem veszik fel a tantárgyat, mint ahogy a tantervben szerepel, illetve nem teszünk különbséget az egyes hallgatók képességei között.

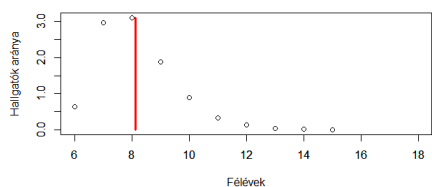
A következő ábrákon a hallgatók végzési idejének a modell által szimulált elosz-lása látható a vizsgált egyetemek előtanulmányi hálói alapján, ahol a piros vonal az átlagos végzési időt jelöli. A függőleges tengely mutatja a végzett diákok arányát, továbbá a vízszintes tengelyen látható, hogy hányadik szemeszterben abszolváltak a szimulált hallgatók. Tehát, ha a 10. félévben a végzettek aránya 0,3, az azt jelenti, hogy 10 000 szimulált hallgatóból 3 000-en fejezték be a tanulmányaikat a 10. félévben. A 10 000 virtuális hallgató végzésének átlagos időpontját a piros vonal jelöli.



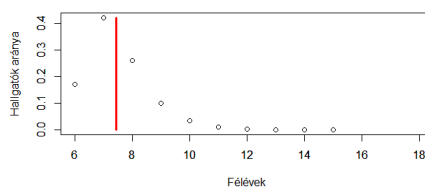
(a) BME



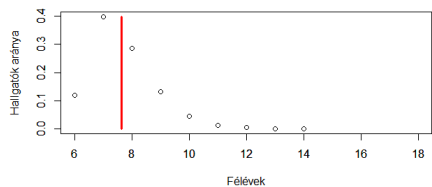
(b) DE



(c) ELTE



(d) PTE



(e) EKE

11. ábra. A hallgatók végzési idejének szimulációval kapott eloszlása az egyes elméleti matematika specializáció képzéseken.

A fentiek alapján azt mondhatjuk, hogy az Eötvös Loránd Tudományegyetem képzésének a legsűrűbb az előtanulmányi rendje, mivel várhatóan itt szerzik meg legkésőbb a hallgatók az abszolutóriumukat. A szimulációnk szerint leggyorsabban a BME képzésén végeznek a hallgatók. Ez indokolható azzal, hogy a BME-n az alapképzés utolsó évében az elméleti specializáció hallgatói nagy szabadsággal rendelkeznek a tantárgyak összeállításával kapcsolatban. Fontos ismét hangsúlyozni, hogy ezek az eredmények a gráf topológiája alapján egyenlő teljesítési valószínűségekkel születtek. A valós teljesítési valószínűségek nagy mértékben megváltoztathatják ezeket az értékeket.

	Átlag	Szórás	Standard hiba	CI(95)
ELTE	8,1308	1,3334	0,0133	(8, 105; 8, 157)
BME	7,1222	1,0823	0,0108	(7, 101; 7, 143)
DE	7,7348	1,2371	0,0124	(7, 711; 7, 759)
PTE	7,4698	1,1259	0,0113	(7, 448; 7, 492)
EKE	7,6487	1,1403	0,0114	(7, 626; 7, 670)

4. táblázat. A várható végzési idők átlaga, szórása, standard hibája és 95%-os konfidenciaintervalluma.

### 5.3. Keresztfélévben is induló tantárgyak feltételének elhagyása

A keresztféléves kurzusok elindítása részben függ attól, hogy az milyen jelentőséggel bír a hallgatók végzési idejére vonatkozóan. Az 5. táblázat azon szimuláció eredményét mutatja képzésenként, amikor minden kurzust évente egyszer indítotunk, tavaszi vagy őszi félévben, azaz a keresztfélév lehetősége nélkül. Jól látszik, hogy keresztféléves tantárgyindítások hiányában jelentősen megnő a képzés elvégzésének ideje.

	Átlag	Szórás	Standard hiba	CI(95)
ELTE	11,3372	2,55	0,0255	(11, 287; 11, 387)
BME	10,2765	2,3846	0,0238	(10, 230; 10, 323)
DE	10,4996	2,3572	0,0236	(10, 454; 10, 546)
PTE	10,4013	2,2738	0,0227	(10, 357; 10, 446)
EKE	10,4224	2,396	0,0240	(10, 376; 10, 469)

5. táblázat. A várható végzési idők átlaga, szórása, standard hibája és 95%-os konfidenciaintervalluma keresztfélév nélkül.

Az 5. táblázat alapján átlagosan a diákok több mint 10 félév alatt végeznek az alapszakon. Természetesen valós adatok esetén ezek az eredmények is módosulnak.

### 5.4. További mutatók

A várható végzési időn kívül a többi ismertetett mérőszám alapján is összehasonlítjuk a hazai matematika alapképzéseket. Ezen mutatók kizárólag az előtanulmányi hálók topológiáján alapszanak.

Megvizsgálva az öt mintatanterv leghosszabb útjait, azt kapjuk, hogy a BME-n csupán három lépés hosszú a leghosszabb út (6. táblázat), igaz ilyen hosszú utakból

Egyetem	Leghosszabb út hossza	Leghosszabb utak száma
BME	3	10
DE	5	1
EKE	4	2
ELTE	5	1
PTE	4	1

6. táblázat. Leghosszabb utak.

tíz különböző (de nem diszjunkt) út is van. A többi egyetemen a leghosszabb út négy vagy öt hosszú (azaz némelyik a teljes képzésen átível).

Egyetem	Legalább kételemű összefüggő komponensek száma	Izolált pontok száma
BME	1	4
DE	2	8
EKE	1	8
ELTE	1	3
PTE	5	15

7. táblázat. Összefüggő komponensek.

A Pécsi Tudományegyetem és a Debreceni Egyetem kivételével mindegyik hazai matematika alapszak gráfja egy legalább kételemű komponensre épül, sőt, Debrecenben is csak egy kételemű csoport adja a többkomponensűséget (7. táblázat). Pécssett öt darabra hullik szét az előtanulmányi háló: logika, geometria, algebra és analízis, számelmélet, illetve informatika blokkokra. Ez a feldaraboltság segítheti a képzés elvégzését. Talán nem annyira meglepő eredmény, hogy korábbi megfigyelésünk alapján a PTE képzése büszkélkedhet a második legrövidebb várható végzési idővel (4. táblázat).

Térjünk át a csúcokra vonatkozó topológiai mérőszámokra. A számszerű eredmények táblázatai a Függelékben találhatóak. Összességében elmondható, hogy minden képzésen kiemelten fontosnak számítanak az analízis, illetve az algebra tudományterületekhez köthető kurzusok a késleltetési tényező alapján. Hasonló mintázat látszódik az üvegnyak szerepet betöltő tantárgyak meghatározásakor. A kölcsönös központiságra vonatkozóan szintűgy ezen két tudományterület a kiemelendő, hiszen az ELTÉ-n egy algebra tárgy kapta a legnagyobb értéket (Algebra 2), míg a többi egyetemen az analízis témakörébe tartozó kurzusok nyertek (BME: Kalkulus 2, DE: Differenciál- és integrálszámítás, EKE: Analízis 3, PTE: Bevezetés az analízisbe 2). Ezen eredmények összecsengenek azon életből vett tapasztalata-

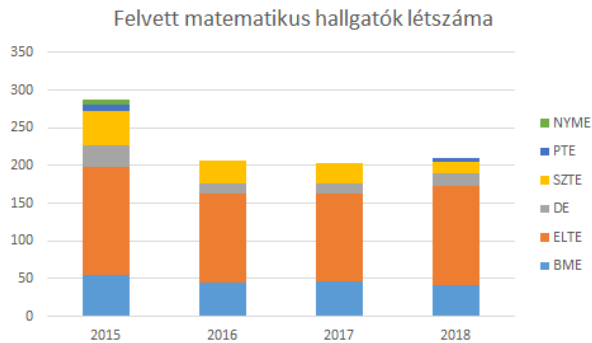
inkkal is, miszerint az analízishez és az algebrahoz kötődő kurzusok okozzák a legnagyobb fejfájást a hallgatók jelentős részének.

### 5.5. További lehetőségek

Elemzésünk további adatok esetén bővíthető azzal, hogy tantárgyanként vizsgáljuk meg a keresztfélévben meghirdetett kurzusok hatását. Továbbá ezzel a modellezési keretrendszerrel vizsgálható az a kérdés is, hogy milyen hatása van többszöri sikertelen tantárgyteljesítésnek mint elbocsátási kritériumnak a végzetek arányára.

## 6. Felvételi statisztikák

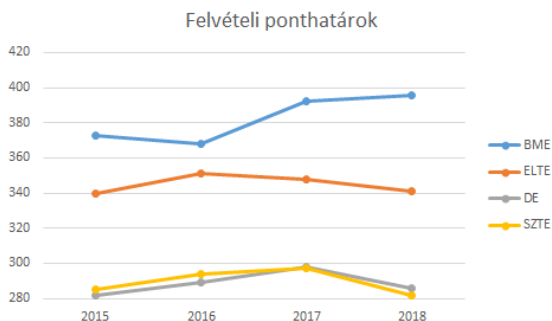
Ebben a fejezetben a nappali tagozatos matematika alapképzések (a képzések részletes leírását lásd a Függelékben) felvételi statisztikáit elemezzük az elmúlt néhány évre vonatkozóan. Vizsgáljuk a felvett hallgatók számát, az önköltséges és államilag támogatott hallgatók arányát, a felvételi ponthatárokat és a felvettek átlagpontszámát. Most itt csak néhány ábrát közlünk, további ábrák, a forrásául szolgáló táblázatok, illetve egyéb elemzések megtalálhatóak a Függelékben.



12. ábra. 2015 és 2018 között a felvételt nyert matematika BSc-s hallgatók létszáma egyetemenként.

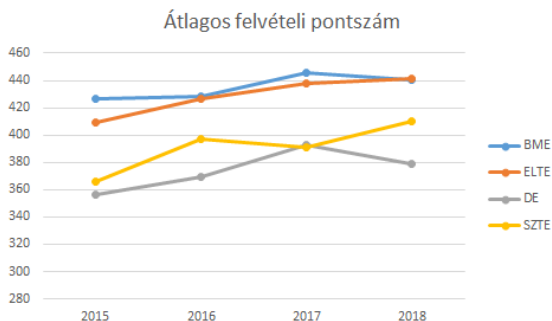
A 12. ábrán megfigyelhető, hogy évről évre egyre kevesebb a matematika BSc szakra felvettek létszáma, bár 2018-ban ez a trend megfordulni látszik. Itt említjük meg, hogy 2017 óta a BME angol nyelven is elindítja a matematika alapszakot, amin félévente tíz-húsz fő körüli Stipendium Hungaricum ösztöndíjas hallgató kezdi meg tanulmányait. Az angol nyelvű képzés azonban nem képezi jelen vizsgalódásunk tárgyát.

Fontos mérőszám a felvételi ponthatár. Az érettségizők mindig nagy figyelmet fordítanak az általuk megjelölt intézmény felvételi pontszámaira. A felvi.hu adatai alapján hasonlítottuk össze a ponthatárokat. Ahol nem indult az elmúlt három év mindegyikében a szak, azokat kihagytuk a jellemzésből (13. ábra). Megfigyelhető, hogy a felvételi ponthatárok növekvő tendenciát mutatnak, de átlagosan a ponthatár nem túl magas más képzésekhez viszonyítva.



13. ábra. Ponthatárok változása 2015-2018 között.

Annak érdekében, hogy pontosabb képet kapjunk a felvételt nyert hallgatók pontjairól, bemutatjuk a hallgatók átlagpontszámát is.



14. ábra. Átlagos pontszám egyetemenként.

Elemzésünk kiterjedt a matematika alapszakra felvett hallgatók összpontszámának egyetemek közötti megoszlására is (lásd a Függelékben).

## 7. Konklúzió, kitekintés

Az egyetemi képzések előtanulmányi hálóinak jellemzésére és összehasonlíthatóságára mutattunk néhány módszert, és ezeket a magyarországi matematika alapképzéseken szemléltettük. Elemzésünk az irodalomban használatos gráfelméleti módszertanon alapszik, de újdonságként bemutattunk egy teljesítési adatokra alapuló valószínűségi modellt is. Ezen módszerek segítségével olyan kérdésekre kaphatunk választ, hogy egy adott tantervben melyik a legfontosabb, leghangsúlyosabb tantárgy, illetve az előtanulmányi háló topológiája hogyan hat a képzés várható teljesítési idejére. Továbbá ismertettünk néhány felvételi statisztikai adatot az elmúlt néhány évre vonatkozóan: többek között a matematika alapszakokra felvett hallgatók létszámának és felvételi pontszámainak alakulását.

Új módszereket mutattunk az egyetemi képzések előtanulmányi hálóinak jellemzésére, illetve az egyes tantárgyaknak a képzés várható végzési idejére vonatkozó hatását illetően. Elemzésünk újdonsága, hogy nemcsak az előtanulmányi háló struktúráját veszi figyelembe, hanem a tantárgyak teljesítési valószínűségeit is. A módszertant hazai matematika alapszakok mintatantervének összehasonlításával szemléltettük. Fontos megemlíteni, hogy valós tantárgyteljesítési adatok hiányában a modellünk eredményei csak nagyon korlátozottan értelmezhetőek. Az ismertett módszerek azonban könnyen alkalmazhatóak valós teljesítési adatok ismeretében, mint ahogyan azt megtettük a BME esetében [7].

Elemzéseink során a könnyebb kezelhetőség érdekében számos egyszerűsítő feltétellel éltünk. Ezen feltételek elhagyása, azaz a modell finomítása pontosabb következtetések levonását teszi lehetővé. A továbbiakban célunk a különböző tantárgyak teljesítését nem egymástól független eseményeknek tekinteni, és az ehhez kapcsolódó feltételes valószínűségeket valós teljesítési adatokon feltérképezni. A reprezentatív hallgató feltevésünket finomítani szeretnénk a jövőben, beépítve a modellbe, hogy a valóságban a hallgatók szorgalma és képességei között nagy különbségek vannak. A feltételek elhagyásával azonban egy bonyolultabb, összetettebb modellt kapunk, ami a könnyű értelmezhetőség rovására mehet. Célunk, hogy a meglévő módszert hatékonyabbá és pontosabbá tegyük.

## Köszönetnyilvánítás

A kutatás részben az EFOP-3.4.4-16. pályázat keretében valósult meg. Montolay Roland kutatása az Innovációs és Technológiai Minisztérium ÚNKP-19-3 kódszámú Új Nemzeti Kiválóság programjának támogatásával készült. Köszönettel tartozunk Lángné Lázi Mártának és Csabay Bálintnak a cikk elkészülte során nyújtott támogatásukért és hasznos tanácsaikért.



## Hivatkozások

- [1] AKBAŞ, M. I., BASAVARAJ, P. AND GEORGIPOULOS, M.: *Curriculum GPS: an adaptive curriculum generation and planning system*, Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), (2015).
- [2] ALDRICH, P. R.: *The curriculum prerequisite network: Modeling the curriculum as a complex system*, Biochemistry and Molecular Biology Education, Vol. **43** No. **3** pp. 168-180 (2015). DOI: [10.1002/bmb.20861](https://doi.org/10.1002/bmb.20861)
- [3] AUVINEN, T., PAAVOLA, J. AND HARTIKAINEN, J.: *STOPS: a graph-based study planning and curriculum development tool*, Proceedings of the 14th Koli Calling International Conference on Computing Education Research. ACM, pp. 25-34 (2014). DOI: [10.1145/2674683.2674689](https://doi.org/10.1145/2674683.2674689)
- [4] BAKER, R. S. AND INVENTADO, P. S.: *Educational data mining and learning analytics*, Learning analytics, Springer New York, pp. 61-75 (2014). DOI: [10.1007/978-1-4614-3305-7\\_4](https://doi.org/10.1007/978-1-4614-3305-7_4)
- [5] BALÁZS M. AND TÓTH B.: *Valószínűségi számítás 1. jegyzet matematikusoknak és fizikusoknak*, (2011).
- [6] BARTHELEMY, M.: *Betweenness centrality in large complex networks*, The European physical journal B, Vol. **38** No. **2**, pp. 163-168 (2004). DOI: [10.1140/epjb/e2004-00111-4](https://doi.org/10.1140/epjb/e2004-00111-4)
- [7] BERGMANN J. AND SZEKRÉNYES D. L.: *A probabilistic approach to the analysis of curriculum prerequisite networks*, TDK, BME, (2017).
- [8] BONACICH, P.: *Some unique properties of eigenvector centrality*, Social networks, Vol. **29** No. **4**, pp. 555-564 (2007). DOI: [10.1016/j.socnet.2007.04.002](https://doi.org/10.1016/j.socnet.2007.04.002)
- [9] DEMCSÁKNÉ ÓDOR ZS. AND HUSZÁRIK P.: *A felvételi eredmények és a tanulmányi előrehaladás összefüggései*, Neptun-konferencia, (2017).
- [10] EISENBERG, B.: *On the expectation of the maximum of IID geometric random variables*, Statistics & Probability Letters, Vol. **78** No. **2**, pp. 135-143 (2008). DOI: [10.1016/j.spl.2007.05.011](https://doi.org/10.1016/j.spl.2007.05.011)
- [11] Fokozatváltás a felsőoktatásban középtávú szakpolitikai stratégia (1785/2016. (XII. 16.) Korm. határozat), (2016).
- [12] HÁMORI Á.: *A 2017. évi keresztféléves eljárásban jelentkezők bejutási eredményessége regionális összevetésben*, Felsőoktatási elemzési jelentések, Oktatási Hivatal, Felsőoktatási Elemzési Főosztály, (2017).
- [13] HORVÁTH D. M., MOLONTAY R. AND SZABÓ M.: *Visualizing student flows to track retention and graduation rates*, Proceedings of the 22nd International Conference on Information Visualisation, IEEE, pp. 338-343 (2018). DOI: [10.1109/iV.2018.00064](https://doi.org/10.1109/iV.2018.00064)
- [14] JANSEN, E.: *The influence of the curriculum organization on study progress in higher education*. Higher education, Vol. **47** No. **4**, pp. 411-435 (2004). DOI: [10.1023/B:HIGH.0000020868.39084.21](https://doi.org/10.1023/B:HIGH.0000020868.39084.21)
- [15] KABICHER, S. AND MOTSCHNIG-PITRIK, R.: *Coordinating curriculum implementation using Wiki-supported graph visualization*, Advanced Learning Technologies, (2009), ICALT (2009), Ninth IEEE International Conference on IEEE, pp. 742-743 (2009). DOI: [10.1109/ICALT.2009.54](https://doi.org/10.1109/ICALT.2009.54)
- [16] KATONA GY., RECSKI A. AND SZABÓ CS.: *A számítástudomány alapjai*, 2. kiadás, Typotex, (2007).

- [17] LIGHTFOOT, J. M.: *A Graph-Theoretic Approach to Improved Curriculum Structure and Assessment Placement*, Communications of the IIMA, Vol. **10** No. **2**, p. 5 (2010).
- [18] NAGY M. AND MOLONTAY R.: *Predicting Dropout in Higher Education based on Secondary School Performance*, Proceedings of 22nd IEEE International Conference on Intelligent Engineering Systems, IEEE, pp. 389-394 (2018). DOI: [10.1109/INES.2018.8523888](https://doi.org/10.1109/INES.2018.8523888)
- [19] SALTZMAN, R. M. AND ROEDER, T. M.: *Simulating student flow through a college of business for policy and structural change analysis*. Journal of the Operational Research Society, Vol. **63** No. **4**, pp. 511-523 (2012). DOI: [10.1057/jors.2011.59](https://doi.org/10.1057/jors.2011.59)
- [20] SLIM, A., KOZLIC, J., HEILEMAN, G. L. AND ABDALLAH, C. T.: *Employing Markov Networks on Curriculum Graphs to Predict Student Performance*, 13th International Conference on Machine Learning and Applications, (2014). DOI: [10.1109/ICMLA.2014.74](https://doi.org/10.1109/ICMLA.2014.74)
- [21] SLIM, A., KOZLICK, J., HEILEMAN, G. L., WIGDAHL, J. AND AVDALLAH, C. T.: *Network Analysis of University Courses*, International Worldwide Web Conference Committee, (2014). DOI: [10.1145/2567948.2579360](https://doi.org/10.1145/2567948.2579360)
- [22] SLIM, A., KOZLIC, J., HEILEMAN, G. L. AND ABDALLAH, C. T.: *The complexity of university curricula according to course cruciality*, Complex, Intelligent and Software Intensive Systems (CISIS), Eighth International Conference, pp. 242-248 (2014). DOI: [10.1109/CISIS.2014.34](https://doi.org/10.1109/CISIS.2014.34)
- [23] PLOTNICKI, W. J. AND GARFINKEL, R. S.: *Scheduling academic courses to maximize student flow: A simulation approach*. Socio-Economic Planning Sciences, Vol. **20** No. **4**, pp. 193-199 (1986). DOI: [10.1016/0038-0121\(86\)90010-8](https://doi.org/10.1016/0038-0121(86)90010-8)
- [24] RAHIM, R., IBRAHIM, H., KASIM, M. M. AND ADNAN, F. A.: *Projection model of postgraduate student flow*. Appl. Math, Vol. **7** No. **2L**, pp. 383-387 (2013). DOI: [10.12785/amis/072L01](https://doi.org/10.12785/amis/072L01)
- [25] VAN DER HULST, M. AND JANSEN, E.: *Effects of curriculum organisation on study progress in engineering studies*. Higher Education, Vol. **43** No. **4**, pp. 489-506 (2002). DOI: [10.1023/A:1015207706917](https://doi.org/10.1023/A:1015207706917)
- [26] WEBER, A. C.: *Simulating the Flow of Students Through Cal Poly's Undergraduate Industrial Engineering Program for Policy Analysis*, (2013).
- [27] WIGDAHL, J.: *Assessment Of Curriculum Graphs With Respect To Student Flow And Graduation Rates*, Albuquerque, New Mexico: The University of New Mexico, Thesis, (2013).
- [28] WIGDAHL, J., HEILEMAN, G. L., SLIM, A. AND ABDALLAH, C. T.: *Curricular efficiency: What role does it play in student success?*, Proceedings of the 121st ASEE Annual Conference and Exposition, IEEE, (2014).

## A. FÜGGELÉK: Üvegnyakak és késleltetési tényezők

A táblázatból kiolvasható, mely tantárgyak töltenek be üvegnyak szerepet az egyes képzések előtanulmányi hálóiban. Korábbi jelöléseinkkel élve (2.2),  $a = 2$  és  $b = 5$  paraméterekkel számoltunk.

BME:
Kalkulus 2
Bevezetés az algebraba 2
Algebra 1
Valószínűségszámítás 1
Informatika 2
Topológia és differenciálható sokaságok
ELTE:
Analízis 2
Algebra 1
Algebra 2
Analízis 3
Bev. a diff.geometriába
Számítástudomány
PTE:
Bevezetés az algebraba és a számelméletbe 2. ea.
Bevezetés az analízisbe 2. ea.
DE:
Lineáris algebra 1
Diff. és integrálszámítás
Többvált. függv. diff- és intsám.
EKE:
Analízis 2
Bevezetés az algebraba és a számelméletbe
Lineáris algebra 1
Analízis 3
Geometria 2
A matematika története

8. táblázat. Üvegnyakak.

A késleltetési tényezők kiszámításánál feltételeztük, hogy minden tantárgy indul minden félévben, továbbá azt is, hogy az adott kurzus leszármazottjai első próbálkozásra sikerülnek. Eredményeinket a 9. táblázat mutatja.

BME kurzusai:	Késleltetési tényező
Pénzügy	1,00
Mikroökonómia+Makroökonómia	0,50
Analízis 1	0,50
Matematikai modellalkotás szeminárium	0,50
Analízis 2	0,50
Differenciálgeometria 1	0,50
Mértékelmélet	0,50
Algebra 2	0,50
Topológia és differenciálható sokaságok	0,50
ELTE kurzusai:	Késleltetési tényező
Analízis 1	1,00
Analízis 2	1,00
Algebra 1	1,00
Algebra 2	1,00
Számelmélet 1	1,00
Analízis 3	1,00
Analízis 4	1,00
Valószínűségyszámítás 2	1,00
Matematikai statisztika	1,00
Funkcionálanalízis 1	1,00
Függvény sorok	1,00
Számítástudomány	1,00
Bev. a diff.geometriába	0,50
Valószínűségyszámítás 1	0,50
Differenciálegyenletek	0,50
Komplex függvénytan	0,50
Numerikus analízis	0,50
DE kurzusai:	Késleltetési tényező
Halmazok és függvények	1,00
Bevezetés az analízisbe	1,00
Diff. és integrálszámítás	1,00
Mérték- és integrálmélet	1,00
Valószínűségyszámítás	1,00
Statisztika	1,00
Elemi topológia	1,00
Többvált. függv. diff- és intsám.	0,50
Bev. a köz. diff.egyenletek elm.	0,50
Differenciálgeometria	0,50
Komplex függvénytan	0,50
Fej. az elemi számelméletből	0,50

Minőségbiztosítási ismeretek	0,50
PTE kurzusai	Késleltetési tényező
Komputeralgebra	1,00
Multiplikatív számelmélet ea	1,00
Bev. az algebrába és a számelméletbe 1. ea.	0,50
Bev. az algebrába és a számelméletbe 2. ea.	0,50
Bevezetés az analízisbe 1. ea.	0,50
Bevezetés az analízisbe 2. ea.	0,50
Algebra és számelmélet 1. ea.	0,50
Algebra és számelmélet 2. ea.	0,50
Analízis 1. ea.	0,50
Analízis 2. ea.	0,50
A matematika alapjai	0,50
Valószínűségszámítás és mat. statisztika 1. ea.	0,50
Valószínűségszámítás és mat. statisztika 2. ea.	0,50
Operációkutatás ea.	0,50
Számítástudományi alapismeretek	0,50
Komplex függvénytan elemei alkalmazásokkal ea.	0,50
Csoportelmélet ea.	0,50
EKE kurzusai:	Késleltetési tényező
Lineáris algebra 2	1,00
Komputeralgebrai rendszerek	1,00
Operációkutatás	1,00
Komplex függvénytan	1,00
Matematikai praktikum 2	0,50
Analízis 1	0,50
Analízis 2	0,50
Geometria 1	0,50
Lineáris algebra 1	0,50
Analízis 3	0,50
Geometria 2	0,50
Bevezetés a valószínűségszámításba	0,50
Geometria 3	0,50
Matematikai statisztika	0,50
Projektív geometria	0,50
A matematika története	0,50

9. táblázat. Késleltetési tényezők.

A nemnulla kölcsönös központiségű tárgyak a 10. táblázatban olvashatók.

BME kurzusai:	Kölcsönös központiség
Kalkulus 2	8,50
Bevezetés az algebrába 2	6,00

Analízis 1	2,50
Algebra 1	6,00
Geometria	2,00
Valószínűségszámítás 1	8,00
Informatika 2	4,00
Analízis 2	3,50
Differenciálgeometria 1	3,50
<hr/>	
ELTE kurzusai:	Kölcsönös központiság
<hr/>	
Analízis 2	13,00
Algebra 2	22,50
Geometria 1	1,50
Véges matematika 2	2,00
Analízis 3	18,00
Analízis 4	8,00
Algebra 3	3,00
Geometria 2	3,00
Valószínűségszámítás 1	6,00
Valószínűségszámítás 2	9,00
Funkcionálanalízis 1	2,00
Operációkutatás 1	2,00
<hr/>	
PTE kurzusai:	Kölcsönös központiság
<hr/>	
Bevezetés az algebra és a számelméletbe 2. ea.	5,00
Bevezetés az analízisbe 2. ea.	8,00
Algebra és számelmélet 1. ea.	2,00
Analízis 1. ea.	4,00
Analízis 2. ea.	3,00
Geometria 1. ea.	2,00
Geometria 2. ea.	2,00
Valószínűségszámítás és matematikai statisztika 1. ea.	2,00
Programozás 1. gyakorlat	1,00
Numerikus analízis 1. ea.	2,00
<hr/>	
DE kurzusai:	Kölcsönös központiság
<hr/>	
Lineáris algebra 1	5,00
Bevezetés az analízisbe	9,00
Bev. az alg. és számelméletbe	5,00
Számelmélet 1	4,00
Diff. és integrálszámítás	16,00
Geometria 1	3,00
Geometria 2	3,00
Lineáris algebra 2	3,00
Algebra	2,00
Többvált. függv. diff- és intsám.	9,00
Mérték- és integrálemélet	6,00
Valószínűségszámítás	4,00
<hr/>	

EKE kurzusai:	Kölcsönös központiság
Analízis 2	9,00
Bevezetés az algebrába és a számelméletbe	6,00
Geometria 1	6,00
Analízis 3	17,50
Geometria 2	11,00
Számelmélet 1	4,00
Bevezetés a valószínűségszámításba	5,00
Lineáris algebra 2	0,50

10. táblázat. Kölcsönös központiságok.

## B. FÜGGELÉK: Hazai matematika alapképzések felépítése

Ebben a fejezetben a képzés felépítése, az induló specializációk és sávok szerint hasonlítjuk össze a hazai matematika alapszakokat az egyetemek honlapján található információk alapján.

### B.1. ELTE

Hazánk egyik legnagyobb egyetemén a hallgatók három specializáció közül választhatnak a matematika BSc képzésen. Ezek:

- matematika,
- alkalmazott matematika,
- matematikai elemző.

A hallgatók a második félévig közös képzésben vesznek részt, majd a harmadik szemeszterben választanak specializációt. Az idejáró diákok tantárgyaikat, saját érdeklődési körük szerint, különböző szinteken sajátíthatják el. Választhatnak a normál, intenzív és haladó szintek között. A haladó szint csak a véges matematika témakörébe tartozó kurzusok esetén indul. Ezen a szinten csak az anyag tárgyalásának sebességében és mélységében különböznek, de adminisztratív szempontból ekvivalensek. Ezek a tantárgyváltozatok szabadon választhatók és szabadon átjárhatók.

A képzés során 180 kreditet kell teljesíteni, melyből 54 kreditet tesz ki az első évben a közös képzés. Részletesebben a tantárgycsoportok megoszlását a következő táblázat mutatja be:

	Kredit
Kötelező	149
Kötelezően választható	12
Szabadon választható tantárgy	9
Szakdolgozat	10
Összesen	180

### B.2. BME

2015-től új lehetőségeket kínál BME a specializációkat illetően. Az idejárók a következő két specializáció és további négy sáv közül választhatnak:

- elméleti matematika,
- alkalmazott matematika,  
adattudomány,



mérnök matematika,  
operációkutatás,  
sztochasztika.

A specializáció kurzusai az 5. félévtől kezdődnek.  
Az egyes tantárgycsoportok krediteit ismerteti az alábbi táblázat.

	Kredit
Alapozó ismeretek	19
Szakmai törzsanyag	21
Differenciált szakmai ismeretek	70
Specializáció tantárgyak	29
Kötelezően választható tantárgyak	22
Szabadon választható tantárgyak	9
Szakdolgozat	10
Összesen	180

### B.3. DE

Debrecenben két specializáció van:

- matematika,
- alkalmazott matematika.

A képzés tantárgycsoportonkénti kreditmegoszlása a következő:

	Kredit
Törzsanyag	50
Differenciált szakmai anyag	47
Specializáció kötelező szakmai anyag	38
Specializáció választható tantárgy	13
Környezettani, Európai Unió, minőségbiztosítási ismeretek	5
Természettudományi alapismeretek	8
Szabadon választható tantárgy	9
Szakdolgozat	10
Összesen	180

### B.4. PTE

Pécsett két specializáció van, az egyik a tanári szakirány, a másik az úgynevezett szakirány nélküli képzés. A szakirányok választására az első év után van lehetőség, habár 120 kredit kötelező mindkét specializáció számára. A kreditek tantárgycsoportonkénti megoszlása a következő:

	Kredit
Alapozó modul	22
Szakmai törzsmodul	68
Kötelezően választható tantárgyak modulja	10
Szabadon választott tantárgyak modulja	10
Szakedolgozat	10
Szakirány nélküli tantárgyak modulja	60
Összesen	180

### B.5. EKE

Egerben nincs külön szakirányválasztás. A többi képzéshez hasonlóan itt is 180 kreditet kell teljesíteni a diploma megszerzéséhez, melyből 10 kredit a szakdolgozatíráásért jár. A kreditek megoszlása a következőképpen néz ki:

	Kredit
Értelmiségi modul	11
Természettudományi ismeretek modul	4
Alapozó szakmai modul	22
Szakmai törzsmodul	32
Differenciált szakmai ismeretek	81
Választható szakmai tantárgyak modulja	11
Szabadon választott tantárgyak modulja	9
Szakedolgozat	10
Összesen	180

Az első négy modul más szakokkal közös, de a további modulok a matematikus specializációhoz tartozó ismeretek.

### B.6. SZTE

A Szegedi Tudományegyetem matematika alapszakának előtanulmányi hálóját nyilvánosan nem elérhető, ezért kimaradt korábbi elemzésünkől, ám néhány adatot sikerült találni a struktúrájáról. A következő ösvények közül lehet választani:

- alkalmazott matematikus specializáció,
- gazdasági specializáció,
- informatikai specializáció,
- matematikus specializáció,
- specializáció nélküli.

Az egyes tantárgycsoportokhoz rendelt kreditszámokat a következő táblázat mutatja. Szegeden egyik tantárgycsoporthoz sincs pontos kreditszám meghatározva, ezek az értékek minimum elvégzendő krediteket jelölnek. Ebből adódóan viszonylag nagy szabadságot kapnak a hallgatók a kurzusválasztást illetően, hiszen 51 kreditet szabadon oszthatnak el a képzés alatt.

	Kredit (min.)
Alapozó ismeretek	19
Szakmai törzsanyag	24
További kötelező matematika tantárgyak	39
Kötelezően választható tantárgyak	30
Egyéb tantárgyak	51
Szakmai gyakorlat	3
Szakdolgozat	14
Összesen	180

A hallgatók a BME és a SZTE képzésén választhatnak a legtöbbször specializáció közül. Az ELTE képzésén a diákok szabadon dönthetnek, hogy a kurzust milyen szinten szeretnék hallgatni. A legnagyobb szabadságot adó képzés a BME képzése, ahol a hallgatóknak a legkevesebb kötelező kreditet kell elvégezni, a hiányzó krediteket a kötelezően választható tantárgyakból pótolják.

Különösen érdekes lenne a matematika alapszakok felépítésének és mintatantervének vizsgálata nemzetközi kitekintésben, de ezen elemzés kimutatna jelen tanulmányunk keretei közül.

**C. FÜGGELÉK: Felvételi statisztikák**

A következő táblázatokban évenkénti bontásban közöljük a hazai matematika alapszakok felvételi statisztikáit 2015-től 2018-ig.

	BME	ELTE	PTE	DE	EKE	SZTE	NYME
Felvett hallgatók létszáma önköltséges képzésen	2	6	0	1	0	3	1
Felvett hallgatók létszáma állami ösztöndíjas képzésen	51	132	9	26	0	39	4
Felvett hallgatók létszáma összesen	53	138	9	27	0	42	5
Felvételi ponthatár (ANA)	373	340	324	282		285	
Átlagos felvételi pontszám	426,84	409,4	377,22	356,78		365,76	346,2
Minimális létszám	20	15	5	10	5	10	5
Maximális létszám	70	140	20	35	20	45	25

11. táblázat. 2015-ben nappali matematika alapképzésre felvettek adatai.

	BME	ELTE	PTE	DE	EKE	SZTE	NYME
Felvett hallgatók létszáma önköltséges képzésen	0	9	0	2	0	0	0
Felvett hallgatók létszáma állami ösztöndíjas képzésen	45	109	0	11	0	30	0
Felvett hallgatók létszáma összesen	45	118	0	13	0	30	0
Felvételi ponthatár (ANA)	368	351	n. i.	289	n. i.	294	n. i.
Átlagos felvételi pontszám	428,37	426,88	0	369,54	0	396,83	0
Minimális létszám	20	15	5	5	5	5	5
Maximális létszám	70	140	20	50	20	70	15

12. táblázat. 2016-ban nappali matematika alapképzésre felvettek adatai.

Megjegyezzük, hogy Szegeden és Egerben levelező tagozat indult.

	BME	ELTE	PTE	DE	EKE	SZTE
Felvett hallgatók létszáma önköltséges képzésen	3	6	0	2	0	1
Felvett hallgatók létszáma állami ösztöndíjas képzésen	43	111	0	12	0	26
Felvett hallgatók létszáma összesen	46	117	0	14	0	27
Felvételi ponthatár (ANA)	392	348	n. i.	298	n. i.	297
Átlagos felvételi pontszám	445,49	437,95	0	392,71		391,33
Minimális létszám	20	15	5	10	5	5
Maximális létszám	70	140	20	50	20	70

13. táblázat. 2017-ben nappali matematika alapképzésre felvettek adatai.

Megjegyezzük, hogy 2017-től a BME angol nyelven is elindítja az alapképzést, de angol nyelvű képzés nem képi jelen vizsgálódásunk tárgyát.

Láthatjuk, hogy az ELTE rendelkezik a legnagyobb létszámú matematika képzéssel. Vidéken a Szegedi Tudományegyetem a legjelentősebb.

	BME	ELTE	PTE	DE	EKE	SZTE
Felvett hallgatók létszáma önköltséges képzésen	0	5	0	2	0	0
Felvett hallgatók létszáma állami ösztöndíjas képzésen	41	127	5	15	0	15
Felvett hallgatók létszáma összesen	41	132	5	17	6	15
Felvételi ponthatár (ANA)	396	341	282	286	329	282
Átlagos felvételi pontszám	440,02	440,94	358,00	379,76	n.i.	410,33
Minimális létszám	20	15	4	5	5	5
Maximális létszám	70	140	20	50	20	70

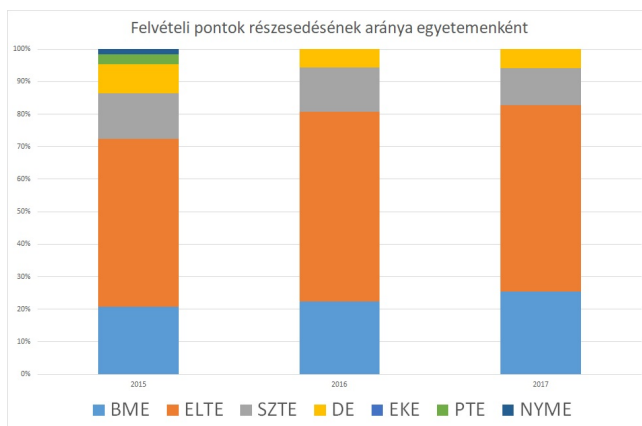
14. táblázat. 2018-ban nappali matematika alapképzésre felvettek adatai.

A 15. ábrán azt szemléltetjük, hogy hogyan oszlanak meg a diákok által hozott összpontszámok az egyes egyetemek között. Egy egyetemre felvett hallgatók létszámát megszoroztuk a hozott átlagponttal, és ennek arányát néztük az összes

felvetthez képest. Formálisan,

$$(\text{részesedési arány})_i = \frac{n_i \cdot \mu_i}{n \cdot \mu},$$

ahol  $n$  az összes felvett hallgató száma,  $\mu$  az összes hallgató átlagpontszáma,  $n_i$  az  $i$  egyetemre felvett hallgatók száma,  $\mu_i$  pedig ezen hallgatók átlagpontszáma.



15. ábra. Felvételi pontszámok részesedése.

Ezen részesedési arány mutatja, hogy a matematika szakra felvett hallgatók összpontszáma hogyan oszlik el az egyes egyetemek között. Egy sáv növekedése magyarázható az adott egyetemre felvett hallgatók számának a növekedésével vagy a felvett hallgatók átlagpontszámának emelkedésével.

(Beérkezett: 2018. február 26.)



Bergmann Júlia 1992-ben született Szekszárdon. A 2011-ben tett érettségije után a Budapesti Műszaki és Gazdaságtudományi Egyetem hallgatója lett, ahol 2015-ben alapszakos, majd 2018-ban mesterzakos matematikus diplomát szerzett analízis specializációval. 2017-ben tantervi hálók elemzéséről szóló tudományos diákköri dolgozatával I. díjat ért el. Diplomaszerezés után az SZTAKI Mérnöki és Üzleti Intelligencia Kutatólaboratóriumában helyezkedett el adatelemzőként, ahol

adatalapú ipari digitalizációs megoldások fejlesztésével és folyamatoptimalizálással foglalkozik, továbbá a Fraunhofer Társaság különböző intézeteivel együtt végez kutatómunkát.

**BERGMANN JÚLIA**

Természettudományi Kar

Budapesti Műszaki és Gazdaságtudományi Egyetem

H 1111 Budapest, Műegyetem rkp. 3-9.



Molontay Roland 1991-ben született Budapesten. A Városmajori Gimnáziumban érettségizett, majd a Budapesti Műszaki és Gazdaságtudományi Egyetemen folytatta tanulmányait, ahol 2015-ben kitüntetéses alkalmazott matematikus diplomát szerzett. Doktori tanulmányait a BME Matematika és Számítástudományi Doktori Iskolájában folytatta hálózatalméleti témában Simon Károly témavezetésével. 2016-ban az egyesült államokbeli Brown Egyetemen volt vendégdoktorandusz. 2012 óta oktat a Műegyetem Sztochasztika Tanszékén, 2019-től pedig az

Aquincum Institute of Technologyban is. 20 tudományos közlemény szerzője, nemzetközi konferenciák rendszeres előadója. 2018-ban a BME Sztochasztika Tanszék magas színvonalú oktatói-kutatói munkájáért Innovációs Díjjal tüntette ki, 2019-ben pedig a hallgatók javaslata alapján a BME TTK Kari Tanácsa a Kar Kiváló Oktatója kitüntetéssel jutalmazta, kiterjedt tehetséggondozó munkájáért a Pro Progressio Alapítvány Oktatói TDK Különdíjában részesült. Jelenleg az MTA-BME Sztochasztika Kutatócsoport munkatársa, a HU-MATHS-IN Hálózat tagja, az Új Nemzeti Kiválósági Program ösztöndíjasa.

**MOLONTAY ROLAND**

Matematika Intézet, Sztochasztika Tanszék

Természettudományi Kar

Budapesti Műszaki és Gazdaságtudományi Egyetem

H 1111 Budapest, Műegyetem rkp. 3-9.



Szabó Mihály 1956-ban született, a Budapesti Műszaki Egyetem Vegyészmérnöki Karán diplomázott 1979-ben, majd 1987-ben általános kémiai technológia "szaktudományból" műszaki doktori címet szerzett. 1993-1997 között a Vegyészmérnöki Kar dékáni hivatalvezetője, majd irányításával hozták létre 2004-ben az egyetem Központi Tanulmányi Hivatalát, amit azóta is vezet. Részt vett a kreditrendszerű képzést kiszolgáló első tanulmányi rendszer kidolgozásában. 40 éve oktat a BME Vegyészmérnöki és Biomérnöki Karának Kémiai és Környezeti Folyamatmérnöki Tanszékén.

#### SZABÓ MIHÁLY

Központi Tanulmányi Hivatal  
Budapesti Műszaki és Gazdaságtudományi Egyetem  
H 1111 Budapest, Műegyetem rkp. 3-9.



Szekrényes Dóra 1993-ban született Budapesten. 2013-ban érettségizett a Premontrei Szent Norbert Gimnáziumban, majd a Budapesti Műszaki és Gazdaságtudományi Egyetemen folytatta tanulmányait, ahol 2016-ban matematika alapszakos, 2018-ban pedig alkalmazott matematikus mesterszakos diplomát szerzett. 2017-ben tantervi hálók elemzéséről szóló tudományos diákköri dolgozatával I. díjat ért el. 2018 óta az okos város gyakorlati megvalósításait kutatja a CHAOS Architects gépi tanulási algoritmusok fejlesztőjeként.

#### SZEKRÉNYES DÓRA LAURA

Természettudományi Kar  
Budapesti Műszaki és Gazdaságtudományi Egyetem  
H 1111 Budapest, Műegyetem rkp. 3-9.



ANALYZING THE CURRICULA AND PREREQUISITE NETWORKS  
OF HUNGARIAN MATHEMATICS BSC PROGRAMS

JÚLIA BERGMANN, ROLAND MOLONTAY, MIHÁLY SZABÓ, DÓRA LAURA SZEKÉNYES

In this paper, we present several methods for analyzing and comparing prerequisite networks of university programs, we evaluate these methods on the curricula of Mathematics BSc programs in Hungary. Also, a brief literature review on the analysis of prerequisite networks is given, and a new data based probabilistic model is presented. Using these methods one can easily answer various important questions, such as which course is the most crucial, or what effect the topology of a network has on the expected time of graduation. Furthermore, results of enrollment process from previous years are summarized regarding Hungarian mathematics BSc programs.

## HÁLÓZATI MODELL EGYÜTTES CSŐDVALÓSZÍNŰSÉGEK MEGHATÁROZÁSÁRA

BIHARY ZSOLT, NAGY NOÉMI, SIMON L. PÉTER

Bizonyos típusú hitelbiztosítások (BDS) árazásának hálózati modellen alapuló lehetőségét vizsgáljuk. Célunk a BDS veszteségfelületének kiszámítása a biztosítási kosárban szereplő cégek egyedi csődvalószínűségeiből. Bemutatunk egy intenzitás alapú, dinamikus, hálózati modellt, ami jól használható a feladat megoldására. Emellett ismertetésre kerül a gyakran alkalmazott, statikus Gauss-kopula modell is. Megmutatjuk, hogy a két eljárás egymáshoz kalibrálható. A hálózati modell egy közelítő változatával nagyszámú cég esetén is hatékonyan mérhető a rendszerkockázat.

**Kulcsszavak:** hálózati modell, pénzügyi folyamatok, rendszerkockázat, mulasztási csereügylet, BDS

### 1. Bevezetés

A pénzügyi piacokon a befektetők számára lehetőség van arra, hogy hitelderivatívák segítségével kötvénykibocsátó cégek nemfizetési kockázata ellen védjék magukat. Ezekre a hitelbiztosításokra abban az esetben van szükség, ha a kötvénykibocsátó cég csődbe menne a futamidő lejáratá előtt. Az egyik legegyszerűbb a mulasztási csereügylet, azaz CDS (credit default swap), ami védelmet nyújt egy kockázatos kötvény vétele esetén. Ennél bonyolultabb biztosítás a basket CDS, azaz BDS (basket default swap), ami több (tipikusan 4-6) különböző cég által kibocsátott kötvény esetén nyújt fedezetet. Aszerint rangsoroljuk a BDS-eket, hogy a kosárban szereplő cégek közül hanyadik cég csődbe menetele után fizet kártérítést: first-to-default (FTD) esetén az első csődbe kerülést állapítják meg kifizetési időpontnak, míg az  $n$ -th-to-default (NTD) azt jelenti, hogy a hitelbiztosítás kiírója az  $n$ -edik cégbedőlés után fizet csak. A szintetikus CDO (synthetic collateralized debt obligation) termékek esetén a biztosított portfólióban több tucat cég is szerepelhet, és a biztosítás a portfólióvesztés egy szeletére (tranche) vonatkozik. A BDS és CDO tranche ügyleteket korrelációs hiteltermékeknek is nevezik, mivel árazásukhoz elengedhetetlenül fontos modellezni a cégek jövőbeli csődeseményeinek korrelációját.

Cikkünkben BDS ügyletek vizsgálatára koncentrálnak. Az arbitrázsmentes árazás feladata ezeknél a termékeknél tipikusan a következőt jelenti: A kosárban szereplő cégek egyedi csődvalószínűségeit, illetve csődfolyamatait adottnak tekintjük, mivel ezek egyszerűen kalibrálhatóak a cégek piacon megfigyelhető CDS-áraihoz. Az alkalmazott modell korrelációs paramétereit likvid korrelációs termékekből, vagy az árazandó BDS múltbeli áraiból becsüljük. A modell megadja az együttes csődvalószínűségeket. Ezek után kiszámítjuk a BDS úgynevezett veszteségfelületét, azaz meghatározzuk, hogy  $t$  idő elteltével mekkora valószínűséggel lesz pontosan  $i$  cég csődben. A veszteségfelületből és a BDS ügylet feltételes kifizetéseiből a veszteség jelenértéke számolható, és így beárazhatjuk a terméket.

A korrelációs hiteltermékek modellezése fontos részterülete a pénzügyi irodalomnak, és a befektetési bankok gyakorlatában is komoly szerepet játszik. Többféle modellípust különböztethetünk meg. A statikus modellek nem próbálják a csődeseményeket sztochasztikus folyamattal reprezentálni, hanem minden lehetséges lejáratra külön kalibrálják az együttes csődvalószínűségek rendszerét. Ezek közé tartozik a Gauss-kopula modell [15], aminek különböző változatai a mai napig a legelterjedtebbek a pénzügyi gyakorlatban. A következő fejezetben röviden ismertetjük majd a Gauss-kopula modellt, mivel javasolt modellünket ezzel a standard modellel fogjuk összehasonlítani.

A dinamikus modellek sztochasztikus folyamatot feltételeznek a csődeseményekre. Ezekben belül a strukturális modellek [13] megpróbálják a cégek eszközértékét korrelált módon modellezni, csődesemény akkor következik be, ha az eszközérték egy kritikus szint alá esik.

Az intenzitás alapú, vagy redukált modellek [8, 10, 23] sztochasztikus portfolyamatként reprezentálják a csődeseményeket. Ezekben a modellekben az együttes viselkedést a csődesemények intenzitásának korrelációja ragadja meg, amit egy vagy több közös intenzitásfaktor valószínűleg meg. Ennek közgazdasági interpretációja szerint a cégek mindegyike csatolódik különböző makrofaktorokhoz, így piaci válságok idején sok egyszerre megy csődbe. Matematikai szempontból ezekben a modellekben a közös faktorok egy rögzített értéke mellett a csődesemények feltételesen függetlenek.

A pénzügyi gyakorlat és empirikus tesztek [14] egyaránt azt mutatják, hogy a feltételesen független modellek realisztikus paraméterezése mellett a csődkorrelációk meglehetősen gyengék. Ezért, és intuitív okokból több szerző úgynevezett fertőző modelleket vezetett be. Ezekben egy cég csődje egyéb cégek csődjét indukálhatja akár közvetlenül, akár a csődintenzitások nagymértékű növelésén keresztül. A fertőző modellek lehetnek statikusak [7], illetve dinamikus strukturális alapúak [9]. Egyéb munkák a fertőzéseket a cégek hitelértékelés-migrációjával [11], vagy hálózati modellekkel [1] ragadták meg.

Ebben a munkában egy egyszerű intenzitás alapú modellt mutatunk be, amelyben a csődbe menő cég fertőzése egyéb cégek csődintenzitásainak növelésével terjed a pénzügyi hálózaton. A matematikai modell felépítése során a fertőzés és informá-

cióterjedés hálózati modelljeiből indulunk ki, egy olyan  $SI$  (susceptible-infected) dinamikájú modellel, amiben spontán fertőződés is van. A járványterjedés hálózaton széles irodalommal rendelkezik [3, 6, 19], ezen összefoglaló művekben ismertetik a járványterjedés modellezésének lehetőségeit különböző dinamikák mellett. A pontos modell egy sztochasztikus folyamat, amihez kapcsolódóan felírhatóak az ún. alapegyenletek, amelyek egy lineáris differenciálegyenlet-rendszert alkotnak. Ezek az egyenletek kisebb hálózat esetén explicit módon felírhatóak, ezt az eljárást a 3. fejezetben tekintjük át. A klasszikus Gauss-kopula módszerrel szemben megközelítésünk dinamikus és hálózati modellen alapul. Így modellünk betekintést enged az egyes cégek állapotaiba bármely időpillanatban, emellett megvizsgálhatjuk a rendszer viselkedését a hálózat felépítésének függvényében. A 4. fejezetben kidolgozunk egy közelítő eljárást, amelynek segítségével hatékonyan és gyorsan elemezhető sok résztvevős rendszerek viselkedése is. Ennél a módszernél az egyes csúcsok állapotának valószínűségeire írunk fel közelítő differenciálegyenleteket lezárás segítségével. Alkalmazásképpen kiszámítjuk és összehasonlítjuk különböző pénzügyi hálózatok rendszerkockázatát.

## 2. A probléma felvázolása és a Gauss-kopula modell bemutatása

Legyen adva  $N$  darab kötvénykibocsátó cég, és tegyük fel, hogy ismerjük valamely  $T$  időpontban az egyes cégek csődbe jutási valószínűségeit, jelöljük ezeket  $\tilde{p}_q$ -val,  $q = 1, \dots, N$ , illetve adott még a cégek közötti kapcsolat erőssége is. A kérdés, hogy hogyan lehet ez alapján kiszámítani az együttes csődesélyeket, azaz a  $D(i)$  értékeket, amely mennyiségek azt adják meg, hogy mi a valószínűsége, hogy pontosan  $i$  darab cég van csődbe a  $T$  időpontban,  $i = 0, \dots, N$ . Ezen  $D(i)$  értékek alapján történik a BDS-ek beárazása, de ennek az eljárásnak az ismeretése nem képezi a cikk tárgyát, számunkra csak a  $D(i)$  értékek meghatározása a cél. Először a pénzügyi életben használt Gauss-kopula modellt fogjuk röviden bemutatni, majd összevetjük az eredményeit az általunk kínált modellel, és megvizsgáljuk, hogy a két eljárást egymáshoz lehet-e kalibrálni a különböző bemeneti paraméterek megfelelő beállításával.

Reprezentálja  $V_q(T)$  standard normális eloszlású valószínűségi változó a  $q$ -adik cég (normalizált) eszközértékét a  $T$  időpontban, és legyen  $B_q(T)$  egy determinisztikus korlát,  $q = 1, \dots, N$ ,  $T \in \mathbb{R}_0^+$ . Ha az eszközérték a  $B_q(T)$  korlát alá esik, akkor a cég csődbe megy, azaz  $p_q(T)$ , a  $q$ -adik cég csődjenek a valószínűsége  $T$ -ben:

$$p_q(T) = P(V_q(T) < B_q(T)) = \phi(B_q(T)),$$

ahol  $\phi$  a standard normális eloszlás eloszlásfüggvénye.

Az eszközértékek közötti kapcsolat a következőképpen van megadva, rögzített  $\rho \in [0, 1]$  korreláció paraméter mellett:

$$V_q = \sqrt{\rho}Z + \sqrt{1 - \rho}Y_q, \quad q = 1, \dots, N,$$

ahol  $Z$  és  $Y_q$ ,  $q = 1, \dots, N$  független standard normális eloszlású valószínűségi változók. (Az eszközértékek közötti kapcsolat időben nem változik.) Ezek alapján  $V_q \sim N(0, 1)$ , és könnyen látható, hogy bármely két cég közötti korreláció  $\rho$ , tehát a  $\rho$  értéke adja meg a cégek egymásra hatásának erősségét, ami ebben a modellben minden cég között ugyanakkora, tehát a kapcsolati háló homogén.

Adottak valamely  $T$  időpontban az egyes cégek csődjének valószínűségei, azaz a  $\tilde{p}_q = p_q(T)$  értékek,  $q = 1, \dots, N$ , illetve a  $\rho \in [0, 1]$  korreláció. A  $B_q(T)$  korlátok a  $T$  időpontban könnyen visszaszámolhatók a bemeneti adatok alapján:  $B_q(T) := \phi^{-1}(\tilde{p}_q)$ . Rögzítsünk le egy konkrét  $z \in \mathbb{R}$  értéket, és vizsgáljuk  $Z = z$  esetén a különböző feltételes valószínűségeket, ugyanis ebben az esetben a  $V_q$  változók függetlenek. Ekkor  $V_q = \sqrt{\rho}z + \sqrt{1-\rho}Y_q$ , vagyis  $V_q \sim N(\sqrt{\rho}z, \sqrt{1-\rho})$  és

$$p_q(T|Z = z) = P(V_q(T) < B_q(T)|Z = z) = \phi\left(\frac{B_q(T) - \sqrt{\rho}z}{\sqrt{1-\rho}}\right),$$

ahol  $p_q(T|Z = z)$  jelöli a  $q$ -adik cég csődjének a valószínűségét  $t$ -ben, feltéve, hogy  $Z = z$ . Ezen értékek függvényében kiszámíthatók a  $D(i|Z = z)$  mennyiségek, azaz annak a valószínűsége, hogy pontosan  $i$  darab cég van csődben  $T$ -ben, feltéve, hogy a  $Z$  valószínűségi változó értéke a rögzített  $z$  szám. Itt fogjuk kihasználni, hogy mivel a  $V_q$  változók függetlenek, így a különböző cégek csődben levésének eseményei függetlenek, tehát szorzatként felírhatóak. Például annak az esélye a  $Z = z$  feltétel mellett, hogy 0 darab cég van csődben:  $D(0|Z = z) = \prod_{q=1}^N (1 - p_q(T|Z = z))$ , illetve a többi  $D(i|Z = z)$  feltételes valószínűség is hasonlóan számolható.

Ezek után használjuk a teljes valószínűség tételét folytonos esetben, így megkapjuk a keresett  $D(i)$  értékeket, azaz hogy mi a valószínűsége annak, hogy pontosan  $i$  cég van csődben a  $T$  időpontban:

$$D(i) = \int_{-\infty}^{+\infty} D(i|Z = z)\varphi(z)dz, \quad i = 0, \dots, N,$$

ahol  $\varphi$  a standard normális eloszlás sűrűségfüggvénye. Érdekes észrevétel, hogy habár a kimeneti adatokat is abban az időpontban kapjuk meg, mint amiben a bemeneti adatokat adtuk meg, a módszer során nem kellett ismerni a  $T$  időpontot. Ilyen értelemben a Gauss-kopula módszer statikusnak tekinthető.

### 3. A hálózati megközelítésen alapuló modell

Most nézzük meg, hogyan lehet ezt a problémát elhelyezni a hálózati modellek világában. Egy természetesen adódó intenzitás alapú hálózati modellt ajánlunk a cégek együttes csődvalószínűségeinek leírására. Ez a modell alkalmas a Gauss-kopula modellel való összevetésre, illetve a rendszerkockázat vizsgálatára, azonban csak kis  $N$ -re alkalmazható a nagy számítási igény miatt. A 4.1. részben egy másik módszert kínálunk, ami nagy csúcsszámú hálózatra is alkalmazható.

### 3.1. A hálózati modell leírása

Egy súlyozott gráffal fogjuk szemléltetni a cégek közötti kapcsolati hálót. A gráf csúcsait tekintjük a cégeknek, míg az egyes csúcsok közötti éleken lévő súly jelzi az adott cégek közötti kapcsolat erősségét. Legyen  $N$  a csúcsok száma és bármely csúcs két lehetséges állapotban lehet:  $S$  – működő,  $I$  – csődbe ment, azaz járványterjedési folyamatként tekintünk a cégek csődbe jutására. Így egy  $2^N$  állapotterű  $\mathbf{X}(t), t \in \mathbb{R}_0^+$  Markov-lánchoz jutunk, ahol  $\mathbf{X}(t)$  egy olyan valószínűségi változó, mely minden rögzített  $t \in \mathbb{R}_0^+$  időpontban megadja, hogy a folyamat melyik állapotban van. Az állapotteret jelöljük  $\mathcal{S}$ -sel, és vezessük be az  $\mathcal{S}^k$  jelölést a  $k$  darab csődöt tartalmazó állapotok részhalmazára, így  $\mathcal{S}^k$  elemei:  $\{\mathcal{S}_1^k, \mathcal{S}_2^k, \dots, \mathcal{S}_{d_k}^k\}$ , ahol  $d_k = \binom{N}{k}$ ,  $k = 0, \dots, N$ . Az  $\mathcal{S}^k$  halmaz  $i$ -edik eleme esetén az  $\mathcal{S}_i^k$  állapotban az  $l$ -edik csúcs állapotát  $\mathcal{S}_i^k(l)$  jelöli.

Feltesszük, hogy kezdetben minden cég működik, és a csődbe menés végleges állapotváltozás, azaz bármely csúcs esetén csak  $S \rightarrow I$  átmenet történhet, tehát  $SI$  típusú dinamikával dolgozunk. Bármely cég csődbe jutását két tényező befolyásolja: egyrészt minden céghez tartozik egy  $\lambda_q \in \mathbb{R}_0^+$ ,  $q = 1, \dots, N$  spontán csődbe menési ráta, ami a cégek sérülékenységét méri, másrészt a már csődbe jutott cégek hatnak a még működő szomszédaira a gráf  $G = ((\omega_{p,q})) \in \mathbb{R}^{N \times N}$  adjacencia mátrixában tárolt nemnegatív rátákkal, tehát  $\omega_{p,q}$  jelentse a  $p$ -edik cég hatáserősségét a  $q$ -adik cégre, ezt a hatáserősséget élsúlynak is nevezzük a továbbiakban. Az  $S \rightarrow I$  átmenetet egy exponenciális eloszlású valószínűségi változó írja le, aminek a paramétere

$$r_i^k(l) := \lambda_l \left( 1 + \sum_{\{p \in \{1, \dots, N\} \mid \mathcal{S}_i^k(p) = I\}} \omega_{p,l} \right), \quad (1)$$

feltéve, hogy  $\mathcal{S}_i^k(l) = S$ , azaz az  $l$ -edik cég még nincs csődbe. Tehát ezzel a rátával kerül az  $l$ -edik cég csődbe az  $\mathcal{S}_i^k$  állapotban.

A hálózat és a dinamika leírása után felírjuk a folyamatot pontosan modellező  $2^N$  dimenziós alapegyenlet-rendszert, a felírás részletei a [20] cikkben megtalálhatóak. Ehhez vezessük be minden  $\mathcal{S}_i^k \in \mathcal{S}^k$ ,  $i = 1, \dots, d_k$ ,  $k = 0, \dots, N$  esetén a  $H_{\mathcal{S}_i^k}^-$ , az „ $\mathcal{S}_i^k$  múltja” halmazt:  $H_{\mathcal{S}_i^k}^- := \left\{ h \in \{1, \dots, d_{k-1}\} \mid \exists! l_h \in \{1, \dots, N\} : \mathcal{S}_i^k(l_h) = I, \mathcal{S}_h^{k-1}(l_h) = S, \mathcal{S}_i^k(m) = \mathcal{S}_h^{k-1}(m), \forall m \neq l_h, m = 1, \dots, N \right\}$ , és a  $H_{\mathcal{S}_i^k}^+$ , az „ $\mathcal{S}_i^k$  jövője” halmazt:  $H_{\mathcal{S}_i^k}^+ := \left\{ j \in \{1, \dots, d_{k+1}\} \mid \exists! l_j \in \{1, \dots, N\} : \mathcal{S}_i^k(l_j) = S, \mathcal{S}_j^{k+1}(l_j) = I, \mathcal{S}_i^k(m) = \mathcal{S}_j^{k+1}(m), \forall m \neq l_j, m = 1, \dots, N \right\}$ . Tehát az  $\mathcal{S}_i^k$  múltja halmaz tartalmazza az összes olyan  $\mathcal{S}^{k-1}$  halmazbeli állapot indexét, melyek pontosan egy csúcs állapotában térnek el az  $\mathcal{S}_i^k$  állapottól, hiszen csak ilyen állapotok között következhet be közvetlen állapotváltozás. Az  $\mathcal{S}_i^k$  jövője halmaz jelentése is értelemszerűen adódik az előbbi alapján.

Jelölje  $x_{\mathcal{S}_i^k}(t)$  annak a valószínűségét, hogy a  $t$  időpontban a Markov-folyamat az  $\mathcal{S}_i^k$  állapotban van. Ekkor az alageyenlet-rendszer alakja:

$$\dot{x}_{\mathcal{S}_i^k}(t) = \sum_{h \in H_{\mathcal{S}_i^k}^-} r_h^{k-1}(l_h) x_{\mathcal{S}_h^{k-1}}(t) - \left( \sum_{j \in H_{\mathcal{S}_i^k}^+} r_i^k(l_j) \right) x_{\mathcal{S}_i^k}(t), \quad (2)$$

$$i = 1, \dots, d_k, \quad k = 0, \dots, N.$$

Az egyenletrendszerre a kompaktabb  $\dot{\mathbf{x}}(t) = \mathbf{P}\mathbf{x}(t)$  alakban is hivatkozhatunk. Az alageyenletek felírását egy 3 csúcsú gráfon szemléltetjük.

*3.1. Példa.* Legyen  $N = 3$ .  $A = ((\omega_{p,q})) \in \mathbb{R}^{3 \times 3}$  a gráf adjacencia mátrixa és a  $\lambda_q$ ,  $q = 1, 2, 3$  értékek a spontán csődbe kerülés rátái. Ekkor az alageyenletek:

$$\begin{aligned} \dot{x}_{SSS}(t) &= -(\lambda_1 + \lambda_2 + \lambda_3)x_{SSS}(t), \\ \dot{x}_{ISS}(t) &= \lambda_1 x_{SSS}(t) - (\lambda_2(1 + \omega_{1,2}) + \lambda_3(1 + \omega_{1,3}))x_{ISS}(t), \\ \dot{x}_{SIS}(t) &= \lambda_2 x_{SSS}(t) - (\lambda_1(1 + \omega_{2,1}) + \lambda_3(1 + \omega_{2,3}))x_{SIS}(t), \\ \dot{x}_{SSI}(t) &= \lambda_3 x_{SSS}(t) - (\lambda_1(1 + \omega_{3,1}) + \lambda_2(1 + \omega_{3,2}))x_{SSI}(t), \\ \dot{x}_{IIS}(t) &= \lambda_1(1 + \omega_{2,1})x_{SIS}(t) + \lambda_2(1 + \omega_{1,2})x_{ISS}(t) - \lambda_3(1 + \omega_{1,3} + \omega_{2,3})x_{IIS}(t), \\ \dot{x}_{ISI}(t) &= \lambda_1(1 + \omega_{3,1})x_{SSI}(t) + \lambda_3(1 + \omega_{1,3})x_{ISS}(t) - \lambda_2(1 + \omega_{1,2} + \omega_{3,2})x_{ISI}(t), \\ \dot{x}_{SII}(t) &= \lambda_2(1 + \omega_{3,2})x_{SSI}(t) + \lambda_3(1 + \omega_{2,3})x_{SIS}(t) - \lambda_1(1 + \omega_{2,1} + \omega_{3,1})x_{SII}(t), \\ \dot{x}_{III}(t) &= \lambda_3(1 + \omega_{1,3} + \omega_{2,3})x_{IIS}(t) + \lambda_2(1 + \omega_{1,2} + \omega_{3,2})x_{ISI}(t) + \\ &\quad + \lambda_1(1 + \omega_{2,1} + \omega_{3,1})x_{SII}(t). \end{aligned}$$

A rendszer második egyenlete például azt rögzíti, hogy az  $(ISS)$  állapotba az  $(SSS)$  állapotból juthatunk el spontán csődbemenéssel  $\lambda_1$  rátával, emellett az  $(ISS)$  állapotból kimehetünk  $\lambda_2(1 + \omega_{1,2})$  rátával az  $(IIS)$  állapotba úgy, hogy vagy a második cég spontán csődbe megy, vagy az első cég csődbe viszi a másodikat, illetve az  $(ISS)$  állapotból átmehetünk az  $(ISI)$  állapotba úgy, hogy vagy a harmadik cég spontán csődbe megy, vagy az első cég csődbe juttatja a harmadikat  $\lambda_3(1 + \omega_{1,3})$  rátával.

### 3.2. Eljárás a csődvalószínűségek meghatározására a hálózati modell segítségével

Most megmutatjuk, hogy ez a hálózati modell hogyan használható a fent ismertett árazási probléma kezelésére, azaz hogyan lehet eljutni az egyes cégek csődbe menési valószínűségeinek ismeretéből az együttes bedőlési valószínűségekhez. A Gauss-kopula modellben a cégek közötti korrelációt a  $\rho$  paraméter adja meg, míg a hálózati modellnél a kapcsolat erősségét a  $G$  szomszédási mátrix írja le, jelen esetben  $\omega_{q,p} \equiv \omega$ , valamely rögzített  $\omega$  esetén,  $p, q = 1, \dots, N$ , így megtartva a homogén kapcsolati hálót. Tehát adott a cégek közötti kapcsolat erőssége

$\omega$  paraméterrel, illetve adottak a  $T$  időpontban az egyes cégek csődbe jutási valószínűségei, azaz a  $\tilde{p}_q$  értékek,  $q = 1, \dots, N$ . Újdonság, hogy ennél a módszernél a  $T$  időpontot is pontosan meg kell adni, ellentétben a Gauss-kopula modellel.

Tekintsünk a (2) alapegyenlet-rendszert. A  $p_q(t)$  valószínűségek és az egyenlet-rendszer változói között felírható egy egyszerű összefüggés minden  $t \in \mathbb{R}_0^+$  esetén:

$$p_q(t) = \sum_{k=1}^N \sum_{\{i \in \{1, \dots, d_k\} | \mathcal{S}_i^k(q) = I\}} x_{\mathcal{S}_i^k}(t), \quad q = 1, \dots, N. \quad (3)$$

Ezek a képletek azt fejezik ki, hogy a  $q$ -adik cég pontosan akkor van csődbe, ha a folyamat egy olyan állapotban van, ahol a  $q$ -adik cég csődbe van, így ezek diszjunkt uniója kiadja a vizsgált eseményt, azaz a valószínűségeik összege megadja a  $p_q$  értéket.

Fontos látni, hogy a hálózati modellhez szükséges  $\lambda_q$  spontán bedőlési ráták nincsenek megadva, így először szeretnénk megkapni őket az eddigi információk alapján. Ehhez bevezetjük a következő  $\mathfrak{F}$  függvényt:

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{pmatrix} \rightarrow \begin{pmatrix} p_1(T) - \tilde{p}_1 \\ \vdots \\ p_N(T) - \tilde{p}_N \end{pmatrix},$$

ahol  $p_q(T)$  jelöli a (3) összefüggéssel megadott függvényeket a  $T$  időpontban. Először megoldjuk numerikusan a (2) egyenletrendszert tetszőleges  $(\lambda_1, \dots, \lambda_N)$  vektor és a kezdetben megadott  $\omega$  bemeneti paraméter mellett, az  $\mathbf{x}(0) = (1, 0, \dots, 0)$  kezdeti feltétellel, azaz a 0 időpontban minden cég működik még. Ezek után kiszámítjuk az egyenletrendszer megoldásainak felhasználásával a (3)-ban megadott  $p_q$  valószínűségeket a  $T$  időpontban. Azt keressük, hogy mely  $\lambda_q$  értékek mellett vesz fel az  $\mathfrak{F}$  függvény nullát. A numerikus tapasztalatok azt mutatják, hogy ezen  $\lambda_q$  értékek egyértelműen meghatározhatóak, azaz a modellünk jól kalibrálható.

Tehát visszakerestük azokat a  $\lambda_q$ ,  $q = 1, \dots, N$  értékeket, melyek esetén a  $T$  időpontban az egyes cégek bedőlési valószínűségei megegyeznek az előre elvárt  $\tilde{p}_q$  értékekkel,  $q = 1, \dots, N$ . Miután megvannak a megfelelő spontán bedőlési ráták értékei, numerikusan megoldható a (2) alapegyenlet-rendszer és kiszámíthatóak a megoldásfüggvények segítségével az együttes bedölések valószínűségei:

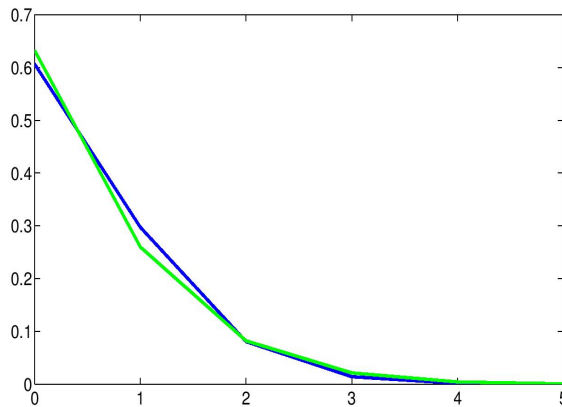
$$D_i(t) = \sum_{j=1}^{d_j} x_{\mathcal{S}_j^i}(t),$$

ahol  $D_i(t)$  jelöli annak a valószínűségét, hogy  $i$  darab cég van csődbe a  $t$  időpontban.

Az eljárás bemutatása után szeretnénk a Gauss-kopula és a hálózatos modell eredményeit összevetni azonos bemeneti paraméterek esetén, azonban ez nehézségbe ütközik, mert a cégek közötti kapcsolatokért felelős mennyiségek, a  $\rho$  és az  $\omega$



viszonya nem ismert. Azonban ezen bemeneti paraméterek megfelelő választásával a két eljárás egymáshoz kalibrálható, azaz a hálózatos modellel kapott  $D_i(T)$  értékek jól illeszkednek a Gauss-kopula modell  $D(i)$  értékeire, amit az 1. ábrán szemléltetünk. Itt az látható, hogy előre megadott  $\tilde{p}_q$  valószínűségek esetén a  $\rho$  paraméterhez megválasztható egy olyan  $\omega$  érték, mely esetén a  $D(i)$  és a  $D_i(T)$  valószínűségek megfelelően közel vannak egymáshoz. Ez azt mutatja, hogy ez az új módszer is jól használható a feladat kezelésére. A mi modellünk fő előnye, hogy dinamikus, azaz megadja a folyamat összes állapotának valószínűségét minden időpillanatban, szemben a Gauss-kopula modellel, ami csak az együttes csődvalószínűségeket határozza meg, és csak a vizsgált  $T$  időpontban. Modellünk így lehetőséget teremt a folyamat mélyrehatóbb vizsgálatára, illetve heterogén kapcsolati hálózattal ellátott hálózat elemzését is lehetővé teszi.



1. ábra. A két módszerrel kapott  $D(i)$  és  $D_i(T)$  értékek,  $i = 0, \dots, 5$ , összehasonlítása (zöld görbe: Gauss-kopula modell, kék görbe: hálózati modell). Bemeneti adatok:  $(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4, \tilde{p}_5) = 0.1014 \cdot (1, 1, 1, 1, 1)$ ,  $\omega = 0.35$ ,  $\rho = 0.2$   $T = 1$ .

#### 4. A rendszerkockázat mérése a hálózati modell segítségével

A továbbiakban egy új problémával foglalkozunk, kihasználva a hálózati modell nyújtotta előnyöket. Az előbbi modell eredményeként kapott  $D_i(t)$  függvények segítségével a pénzügyi élet egy fontos kérdésére is választ találhatunk: hogyan változik a rendszerkockázat az idő függvényében különböző kapcsolati hálókat esetén. Itt megjegyezzük, hogy a rendszerkockázatnak nagyon sokféle értelmezése lehetséges, amelyek számos szakirodalomban megtalálhatóak [18, 5, 4, 21, 16]. Ebben a

munkában a rendszerkockázat alatt a csődök számának várható értékét értjük az idő függvényében, és a következő formulával definiáljuk:

$$K(t) = \sum_{i=1}^N iD_i(t).$$

Sajnos, az ajánlott új eljárás – és így a rendszerkockázat meghatározása – csak kis  $N$  esetén működik, ugyanis beleütközünk a szokásos gondba, hogy a  $2^N$  méretű alapegyenlet-rendszer túl nagy számítási igényű. A következő szakaszban egy másik eljárást kínálunk a rendszerkockázat mérésére, mely nagy  $N$ -re is alkalmazható.

#### 4.1. A csúcsok szintjén felírt egyenletek

A (2) alapegyenlet-rendszer egyenletei alapján szeretnénk felírni egy kisebb dimenziós közelítő differenciálegyenlet-rendszert, az ún. csúcsok szintjén felírt egyenleteket [22, 19]. A redukált rendszer segítségével kiszámoljuk a rendszerkockázatot, illetve azt tanulmányozzuk, hogy hogyan befolyásolja a gráf struktúrája a rendszerkockázat alakulását különböző, nagy méretű hálózatok esetén.

##### 4.1.1. A modell formalizálása

Vezessünk be néhány új jelölést. Legyen  $\langle I_q \rangle(t)$  annak a valószínűsége, hogy a  $q$ -adik cég csődben van a  $t$  időpontban. Tehát

$$\langle I_q \rangle(t) := \sum_{k=1}^N \sum_{i \in G_q^k(I)} x_{\mathcal{S}_i^k}(t), \quad q = 1, \dots, N,$$

ahol  $G_q^k(I) := \{i \in \{1, \dots, d_k\} \mid \mathcal{S}_i^k(q) = I\}$ . Emellett használni fogjuk a következő jelöléseket is:

$$\langle S_q \rangle(t) := \sum_{k=0}^{N-1} \sum_{i \in G_q^k(S)} x_{\mathcal{S}_i^k}(t), \quad q = 1, \dots, N,$$

$$\langle S_q I_r \rangle(t) := \sum_{k=1}^{N-1} \sum_{i \in G_{q,r}^k(S,I)} x_{\mathcal{S}_i^k}(t), \quad q = 1, \dots, N, r = 1, \dots, N,$$

a  $G_q^k(S) := \{i \in \{1, \dots, d_k\} \mid \mathcal{S}_i^k(q) = S\}$ ,  $G_{q,r}^k(S,I) := \{i \in \{1, \dots, d_k\} \mid \mathcal{S}_i^k(q) = S, \mathcal{S}_i^k(r) = I\}$  jelölések mellett. Tehát  $\langle S_q \rangle$  jelöli azon állapotok valószínűségének összességét, melyekben a  $q$ -adik csúcs működő cég, illetve  $\langle S_q I_r \rangle$  annak a valószínűségét, hogy a  $q$ -adik csúcs működik, az  $r$ -edik csődben van a

$t$  időpontban. A (2) alapegyenlet-rendszerből kiindulva szeretnénk  $\langle I_q \rangle$ -ra felírni egy differenciálegyenlet-rendszert.

A (2) egyenletek teljesülése esetén az  $\langle I_q \rangle$  függvényekre fennállnak az alábbi differenciálegyenletek:

$$\langle \dot{I}_q \rangle(t) = \lambda_q \langle S_q \rangle(t) + \sum_{p=1, p \neq q}^N \lambda_q \omega_{p,q} \langle S_q I_p \rangle(t), \quad q = 1, \dots, N. \quad (4)$$

A bizonyítást hosszadalmas technikai részletei miatt itt nem közöljük, az állítást egy későbbi dolgozatban igazoljuk.

Megjegyezzük, hogy ha a kapcsolati háló homogén, és a spontán csődbe jutási ráták is azonosak, azaz  $\omega_{p,q} \equiv \omega$ ,  $\lambda_q \equiv \lambda$ ,  $p, q = 1, \dots, N$ , valamely  $\omega$  és  $\lambda$  mellett, akkor a (4) rendszerből levezethető egy pontos átlagoláson alapuló egyenlet az átlagos csődszámra:

$$\langle \dot{I} \rangle(t) = \lambda \langle S \rangle(t) + (N-1)\lambda\omega \langle SI \rangle(t),$$

amennyiben a dinamika  $SI$  típusú és az (1) függvénycsaláddal adjuk meg az átmeneti rátákat.

A (4) egyenletrendszer alapján megadunk egy közelítő differenciálegyenlet-rendszert is, melynek megoldásai az  $\langle I_q \rangle$  valószínűségeket közelítik.

**4.1. ÁLLÍTÁS.** Felhasználva az  $\langle S_q \rangle = 1 - \langle I_q \rangle$  összefüggést és az  $\langle S_q I_r \rangle \approx \langle S_q \rangle \cdot \langle I_r \rangle$  közelítést az egyenletek lezárására, a (4) egyenlet egy közelítését kapjuk. A közelítő egyenletek a következő alakúak:

$$\dot{x}_q(t) = \lambda_q (1 - x_q(t)) \left( 1 + \sum_{p=1, p \neq q}^N \omega_{p,q} x_p(t) \right), \quad q = 1, \dots, N, \quad (5)$$

ahol az  $x_q$  függvények jelölik a közelítő egyenlet megoldásait.

*Bizonyítás.* Mivel  $\langle I_q \rangle \approx x_q$ ,  $\langle S_q \rangle \approx 1 - x_q$ , és  $\langle S_q I_p \rangle \approx \langle S_q \rangle \cdot \langle I_p \rangle = (1 - x_q)x_p$ , ezért a közelítő differenciálegyenlet-rendszer:

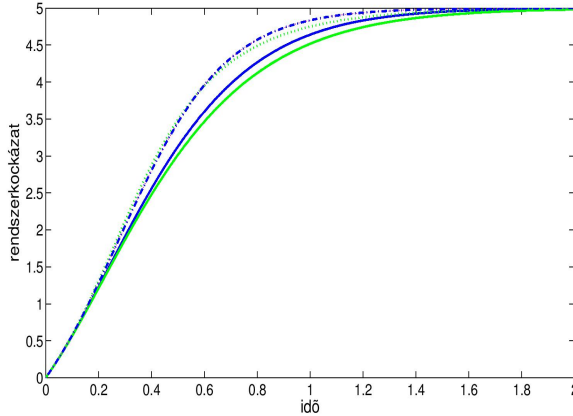
$$\dot{x}_q(t) = \lambda_q (1 - x_q(t)) + \sum_{p=1, p \neq q}^N \lambda_q \omega_{p,q} (1 - x_q(t)) x_p(t).$$

□

Az (5) differenciálegyenlet-rendszer megoldása után a rendszerkockázatot könnyedén számolhatjuk a

$$K(t) = \sum_{q=1}^N x_q(t)$$

képlet segítségével, hiszen  $x_q$  közelítőleg meghatározza a  $q$ -adik csúcs csődjének a valószínűségét. Az (5) modell előnye, hogy tetszőlegesen nagy  $N$  esetén is gyorsan megoldható numerikusan.



2. ábra. Rendszerkockázat az idő függvényében teljes (kék) és csillag (zöld) gráfon a (2) alapegyenletekkel (folytonos) és az (5) modellel (szaggatott görbe),  $N = 5$ ,  $\lambda_q = 1$ ,  $\omega = 2$ .

A 2. ábrán összevetjük az (2) és a (5) modellek eredményeit egyforma, spon-tán csődbe menési ráták mellett különböző gráftípusokon: teljes és csillag gráfon rögzített összélsúly mellett, tehát ha  $\sum_{p=1}^N \sum_{q=1}^N \omega_{p,q}$  adott. Látható, hogy a rendszerkockázati görbék kvalitatíve nagyon hasonlóak, így az (5) modell alkalmas a (2) rendszer helyettesítésére nagy  $N$  esetén. Megjegyezzük, hogy az (5) rendszer ugyanolyan eredményt ad kör- és teljes gráf esetén, míg a (2) modellel különböző rendszerkockázati görbéket kapunk. Ennek magyarázata később található. A következő részben az (5) közelítő rendszer viselkedését vizsgáljuk meg különböző struktúrájú gráfok esetén.

#### 4.1.2. A modell vizsgálata

Ebben a részben egyrészt explicit képletet adunk a rendszerkockázat kiszámítására az (5) rendszer alapján néhány egyszerű esetben, másrészt arra keressük a választ, hogy csökkenthető-e a rendszerkockázat a gráf struktúrájának megfelelő megválasztásával, illetve speciális struktúrájú gráfok esetén hogyan alakul a várhatóan csődbe lévő cégek száma az idő függvényében. Ilyen és ehhez hasonló kérdésekkel gyakorta lehet találkozni az irodalomban, például azzal a felvetéssel is, hogy a hálózat összekapcsoltsága növeli, vagy csökkenti-e a rendszerkockázatot [17]. A tapasztalat azt mutatja, hogy attól függően, hogy a rendszerkockázatot

hogyan definiálják és milyen modellekkel dolgoznak, különböző válaszokat lehet kapni ezekre a kérdésekre [2, 4].

**A közelítő egyenletrendszer analízise egyszerű esetekben** Először néhány egyszerű megállapítást teszünk, ugyanis ha a cégek sérülékenysége azonos, akkor az (5) differenciálegyenlet-rendszer leegyszerűsödik, így megoldható elemi módszerek segítségével. Két lemmát fogalmazunk meg – melyeket egyszerűségükből adódóan bizonyítás nélkül közlünk – az elsőt homogén, míg a másodikat heterogén kapcsolati háló esetén.

4.1. LEMMA. *Ha  $\lambda_q \equiv \lambda$ ,  $\omega_{p,q} \equiv \omega$ ,  $p, q = 1, \dots, N$ , valamely  $\lambda$  és  $\omega$  mellett, és a kezdeti feltétel a szokásos, azaz  $x_q(0) = 0$ ,  $q = 1, \dots, N$ , akkor az összes cég helyzete ugyanúgy fog alakulni, azaz ilyenkor  $x_q(t) = x(t)$  teljesülni fog valamely  $x(t)$  függvény esetén. Felhasználva a feltételeket, az (5) differenciálegyenlet-rendszerből egy szétválasztható változójú differenciálegyenlethez jutunk:*

$$\dot{x}(t) = \lambda(1 - x(t))\left(1 + (N - 1)\omega x(t)\right).$$

Melynek megoldása az  $x(0) = 0$  kezdeti feltétel mellett:

$$x(t) = 1 - \frac{\alpha}{e^{t\lambda\alpha} + \alpha - 1}, \text{ ahol } \alpha = 1 + (N - 1)\omega.$$

Ekkor a rendszerkockázat képlete:  $K(t) = Nx(t) = N \cdot \left(1 - \frac{\alpha}{e^{t\lambda\alpha} + \alpha - 1}\right)$ .

4.2. LEMMA. *Ha  $\lambda_q \equiv \lambda$ ,  $q = 1, \dots, N$ , valamely  $\lambda$  mellett, és a kezdeti feltétel  $x_q(0) = 0$ ,  $q = 1, \dots, N$ , akkor az (5) rendszerből a következő  $N$  változós differenciálegyenlet-rendszerhez jutunk:*

$$\dot{x}_q(t) = \lambda(1 - x_q(t))\left(1 + \sum_{p=1, p \neq q}^n \omega_{p,q}x_p(t)\right), \quad q = 1, \dots, N. \quad (6)$$

Ha a  $\sum_{p=1, p \neq q}^n \omega_{p,q}$  érték, azaz az egy csúcsba befutó súlyok összege független  $q$ -tól, tehát  $\sum_{p=1, p \neq q}^n \omega_{p,q} = B$ , minden  $q = 1, \dots, N$  esetén, akkor  $x_q(t) = y(t)$ ,  $q = 1, \dots, N$  teljesülni fog a (6) rendszer megoldásaira valamely olyan  $y(t)$  függvény esetén, melyre

$$\dot{y}(t) = \lambda(1 - y(t))(1 + By(t)),$$

vagyis minden  $x_q(t)$  függvényre ugyanazt a megoldást kapjuk:  $y(t) = 1 - \frac{1+B}{e^{t\lambda(1+B)} + B}$ . Ezek alapján a rendszerkockázat is kiszámítható:

$$K(t) = Ny(t) = N \cdot \left(1 - \frac{1+B}{e^{t\lambda(1+B)} + B}\right). \quad (7)$$

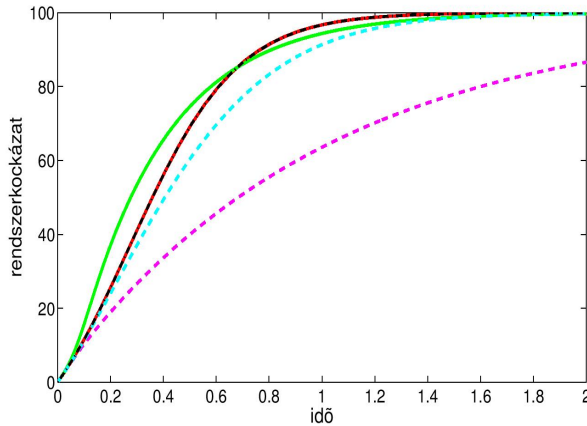
Ezzel magyarázható, hogy ha az összélsúlyt lerögzítjük, akkor például körgráfon és teljes gráfon ugyanazt az eredményt kapjuk, ha az idő függvényében ábrázoljuk a rendszerkockázatot az (5) rendszer alapján.

**Speciális struktúrájú gráfok esete** Ebben a részben a rendszerkockázati görbe változását követjük nyomon a gráfstruktúra függvényében. Nyilván, ahogy az éleken található összsúlyt növeljük, azaz egyre szorosabb kapcsolatot teszünk fel a cégek között, úgy a rendszerkockázat egyre jobban növekszik, tehát adott idő alatt várhatóan egyre több cég kerül csődbe. Ugyanez mondható el a cégek sérülékenységének növelése esetén. Ezért állandó összsúly és rögzített spontán csődbe menési ráták mellett fogjuk tekinteni a különböző gráftípusokon a rendszerkockázat időbeli lefolyását. A kérdés az, hogy hogyan érdemes beállítani a kapcsolati hálót, hogy minél kisebb legyen a várhatóan csődben lévő cégek száma. A folyamatot speciális struktúrájú irányítatlan gráfokon, azaz teljes, kör- és csillag gráfon, emellett véletlen gráfokon: reguláris, bimodális véletlen gráfon, illetve Barabási–Albert-gráfon [12] is vizsgáljuk. Az összsúly legyen:  $\sum_{p=1}^N \sum_{q=1}^N \omega_{p,q} = 2\omega N$ , és  $\lambda_q = 1$ ,  $q = 1, \dots, N$ , a gráfok struktúrájának megfelelően elosztva az egyes éleken, példaként teljes gráf esetén egy élen  $\frac{2\omega N}{N(N-1)} = \frac{2\omega}{N-1}$  súly lesz (mindkét irányban), míg körgráf esetén  $\frac{2\omega N}{2N} = \omega$ , és csillag gráf esetén  $\frac{2\omega N}{2(N-1)} = \frac{\omega N}{N-1}$ . A továbbiakban a vizsgálatok során mindig 0 kezdeti csődből indítjuk a folyamatot.

A numerikus eredményekből azt látjuk, hogy az idő függvényében ábrázolt rendszerkockázati görbe nagyon hasonló a vizsgált gráfokon. A homogén súlyozású esetekben, vagyis a teljes és a reguláris véletlen gráfon a rendszerkockázat előáll a (7) képlet alapján, ahol  $B = 2\omega$ , hiszen ezekben az esetekben az egy csúcsból kiinduló élsúlyok összege állandó. Sőt, még a körgráf rendszerkockázata is egybeesik az előbbiekével, mivel itt is az egy csúcsba befutó súlyok összege független a csúcstól, és ez az érték megegyezik mindhárom gráf esetén. Ezen görbéhez képest elemezzük a többi gráfhoz tartozó rendszerkockázati görbe helyzetét. A folyamat kezdeti szakaszában a heterogén súlyozású gráfok a (7) görbe felett vannak, minél heterogénebb (csillag, majd Barabási–Albert, végül a bimodális gráf), annál inkább felette, majd az idő előrehaladtával a (7) görbe alá mennek (ugyanilyen sorrendben).

A rendszerkockázati görbék hasonlósága arra enged következtetni, hogy irányítatlan gráfok esetén, ha az összsúly fix, akkor a kapcsolati háló speciális beállítással se lehet a rendszerkockázatot látványosan lecsökkenteni. Ez motiválja, hogy megvizsgáljuk ugyanezt a kérdést irányított, speciális struktúrájú gráfok esetén is. Csak néhány jól átlátható esettel foglalkoztunk: körgráf esetén azt vizsgáltuk, amikor a cégek között körbeirányítva mennek az élek, azaz van benne pontosan egy irányított kör. Csillag gráfnál két esetet néztünk meg: az egyik esetben a gráfon minden él egy központi csúcs felé van befelé irányítva (a továbbiakban becsillag gráfként hivatkozunk rá), a másik esetben pedig kifelé irányítjuk az éleket a központi csúcstól (a továbbiakban ki-csillag gráfként említjük). Természetesen itt is feltettük, hogy az összsúly állandó, és a sérülékenységi rátákat megint rögzítettük.

A 3. ábrán az látható, hogy az irányított körgráfon kapott megoldás a (7) görbét adja  $B = 2\omega$  mellett, hiszen itt is az egy csúcsba befutó súlyok összege

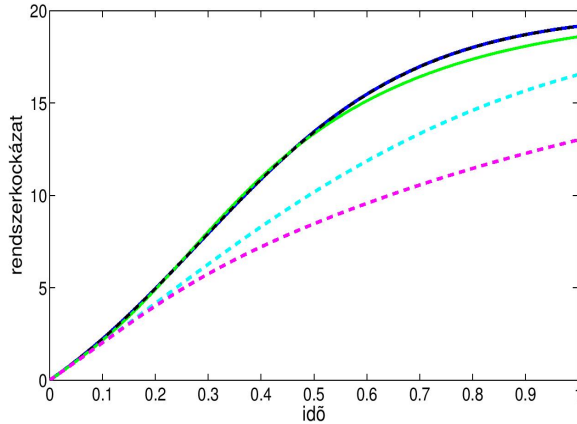


3. ábra. Rendszerkockázat az idő függvényében a (7) egyenletből kapott esetben (piros görbe), csillag gráfon (zöld), irányított kör- (szaggatott fekete), ki-csillag (szaggatott cián) és be-csillag (szaggatott lila) gráfon,  $N = 100$ ,  $\lambda_q = 1$ ,  $q = 1, \dots, N$ ,  $\omega = 2$ .

csúcsfüggetlen. Az aszimmetria csökkenti a rendszerkockázatot a csillag gráf esetében: ha a középponti csúcs csődje van csak hatással a többi csúcs csődjére, az már érzékelhető csökkenést okoz (ki-csillag), ugyanis csak a középponti csúcs csődje esetén kezd el jelentősen összeomlani a rendszer, de ennek a csúcsnak a csődjét csak a saját spontán csődbe jutási rátája befolyásolja. Ha csak a középponti cég csődjére van hatással a többi periférián található cég, akkor lényegében a külső  $N - 1$  cég teljesen független egymástól, csődjüket más cégek összeomlása nem befolyásolja, így ez a gráfstruktúra csökkenti a legjobban a rendszerkockázatot (be-csillag).

Az utolsó vizsgált esetben már nem csak a cégek egymáshoz való viszonya lehet különböző, hanem a különböző cégek sérülékenységi rátája is. Tehát ha egy cég spontán csődbe menési rátája kicsi, az azt jelenti, hogy a cég erős, nehezen kerül csődbe, míg nagy  $\lambda_q$  értékkel szerepelnek a gyenge cégek.

Itt is néhány szélsőséges esetet hasonlítottunk össze: irányítatlan teljes, csillag, illetve irányított kör-, ki-csillag és be-csillag gráfon. A cégek közötti erőviszony: az első cég spontán csődbe menési rátája fele akkora, mint a többi cégnek, tehát az első cég a legerősebb. Ki- és be-csillag gráf esetén az erős cég elhelyezkedése is fontos: ez a cég van közepén. Az összevetésbe néhány irányítatlan gráfot is belevettünk azért, hogy pontosabb kép alakuljon ki az irányított gráfok esetéről. A 4. ábrán látható hogy a teljes, illetve az irányított körgráf majdnem ugyanúgy szerepel, és ők a legrosszabbak rendszerkockázat szempontjából. A ki-csillag gráf esetén a központi erős cég csődje van hatással a periférián lévő gyengébb cégekre, ez tovább csökkenti



4. ábra. Rendszerkockázat az idő függvényében teljes (kék), csillag (zöld), irányított kör- (szaggatott fekete), be-csillag (szaggatott magenta) és ki-csillag (szaggatott cían) gráfon,  $N = 20$ ,  $\lambda_1 = 0.5$ ,  $\lambda_q = 1$ ,  $q \neq 1$ ,  $q = 1, \dots, N$ ,  $\omega = 2$ .

a rendszerkockázatot az előző esetbeli ki-csillag rendszerkockázatához képest, mert az erős cég működése még stabilabb, mint egyforma  $\lambda_q$  értékek esetén. Ezzel szemben a be-csillag gráfon a központi erős cégre vannak hatással a gyenge cégek, amelyek gyorsabban csődbe kerülnek, így ez növeli kicsit a rendszerkockázatot az előző esetbeli be-csillag gráfon mérthez képest, tehát a ki-csillag és be-csillag gráfok rendszerkockázati görbéi egy erős cég esetén közelednek egymáshoz az egyforma  $\lambda_q$  értékeket megadó esethez képest, de még így is a be-csillag gráf az optimális struktúrájú gráf a kockázat szempontjából.

## 5. Összefoglalás

Számos modell ismert a pénzügyi piacokon megjelenő különböző problémák kezelésére. Ebben a munkában a hitelbiztosítások árazása, illetve a rendszerkockázat mérése került a figyelem középpontjába. Kidolgoztunk egy hálózati modellen alapuló eljárást ezen feladatok megoldására. Modellünket összevetettük a gyakran alkalmazott Gauss-kopula modell eredményeivel, amiből látható, hogy ez a dinamikus, hálózati modell is alkalmas a számítások elvégzésére. Továbbá a bemutatott eljárás információt ad a rendszer működéséről és az egyes cégek állapotáról minden időpontban, ellentétben a Gauss-kopula modellel, így lehetőséget teremt további vizsgálatokra is, például választ találhatunk arra a kérdésre is, hogy hogyan változik a rendszerkockázat a hálózat felépítésének függvényében. A modell



egy közelítő változatával pedig sokrésztvevős rendszerek esetén is megfigyelhetjük a cégek csődbejutásának folyamatát.

### Köszönetnyilvánítás

Simon Péter munkáját az NKFIH K 115926. számú pályázata támogatta.

### Hivatkozások

- [1] ACEMOGLU, D., OZDAGLAR, A. AND TAHBAZ-SALEHI, A.: *Systemic risk and stability in financial networks*, American Economic Review, Vol. **105** No. **2**, pp. 564-608 (2015). DOI: [10.1257/aer.20130456](https://doi.org/10.1257/aer.20130456)
- [2] ALLEN, F. AND GALE, D.: *Financial Contagion*, The Journal of Political Economy, Vol. **108** No. **1**, pp. 1-33 (2000). DOI: [10.1086/262109](https://doi.org/10.1086/262109)
- [3] BARRAT, A., BARTHELEMY, M. AND VESPIGNANI, A.: *Dynamical Processes on Complex Networks*, Cambridge University Press, Cambridge (2008). DOI: [10.1017/CBO9780511791383](https://doi.org/10.1017/CBO9780511791383)
- [4] BATTISTON, S., GATTI, D. D., GALLEGATI, M., GREENWALD, B. C. AND STIGLITZ, J. E.: *Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk*, Journal of Economic Dynamics and Control, Vol. **36** No. **8**, pp. 1121-1141 (2012). DOI: [10.1016/j.jedc.2012.04.001](https://doi.org/10.1016/j.jedc.2012.04.001)
- [5] BOSS, M., ELSINGER, H., SUMMER, M. AND THURNER, S.: *Network topology of the interbank market*, Quantitative Finance, Vol. **4** No. **6**, pp. 677-684 (2004). DOI: [10.1080/14697680400020325](https://doi.org/10.1080/14697680400020325)
- [6] DANON, L., FORD, A.P., HOUSE, T., JEWELL, C.P., KEELING, M.J., ROBERTS, G.O., ROSS, J.V. AND VERNON, M.C.: *Networks and the Epidemiology of Infectious Disease*, Interdisciplinary Perspectives on Infectious Diseases 2011:284909 special issue "Network Perspectives on Infectious Disease Dynamics", (2011). DOI: [10.1155/2011/284909](https://doi.org/10.1155/2011/284909)
- [7] DAVIS, M. AND LO, V.: *Infectious defaults*, Quantitative Finance, Vol. **1** No. **4**, pp. 382-387 (2001). DOI: [10.1080/713665832](https://doi.org/10.1080/713665832)
- [8] DUFFIE, D. AND SINGLETON, K. J.: *Modeling term structures of defaultable bonds*, The Review of Financial Studies, Vol. **12** No. **4**, pp. 687-720 (1999). DOI: [10.1093/rfs/12.4.687](https://doi.org/10.1093/rfs/12.4.687)
- [9] EGLOFF, D., LEIPPOLD, M. AND VANINI, P.: *A simple model of credit contagion*, Journal of Banking & Finance, Vol. **31** No. **8**, pp. 2475-2492 (2007). DOI: [10.1016/j.jbankfin.2006.10.023](https://doi.org/10.1016/j.jbankfin.2006.10.023)
- [10] ERRAIS, E., GIESECKE, K., GOLDBERG, L. R.: *Affine Point Processes and Portfolio Credit Risk*, SIAM J. Finan. Math., Vol. **1** No. **1**, pp. 642-665 (2010). DOI: [10.1137/090771272](https://doi.org/10.1137/090771272)
- [11] GIESECKE, K. AND WEBER, S.: *Credit contagion and aggregate losses*, Journal of Economic Dynamics & Control, Vol. **30** No. **5**, pp. 741-767 (2006). DOI: [10.1016/j.jedc.2005.01.004](https://doi.org/10.1016/j.jedc.2005.01.004)
- [12] GUPTA, A., KING, M. M., MAGDANZ, J., MARTINEZ, R., SMERLAK, M. AND STOLL, B.: *Critical connectivity in banking networks*, SFI Complex Systems Summer School Proceedings, (2013), [http://samoa.santafe.edu/media/cms\\_page\\_media/500/Banks\\_FLreport\(1\).pdf](http://samoa.santafe.edu/media/cms_page_media/500/Banks_FLreport(1).pdf)

- [13] HULL, J. C., PREDESCU, M. AND WHITE, A.: *The Valuation of Correlation-Dependent Credit Derivatives Using a Structural Model*, J. Credit Risk, Vol. **6** No. **3**, pp. 99-132 (2010). DOI: [10.21314/JCR.2010.112](https://doi.org/10.21314/JCR.2010.112)
- [14] LANDO, D. AND NIELSEN, M. S.: *Correlation in corporate defaults: Contagion or conditional independence?*, Journal of Financial Intermediation, Vol. **19** No. **3**, pp. 355-372 (2010). DOI: [10.1016/j.jfi.2010.03.002](https://doi.org/10.1016/j.jfi.2010.03.002)
- [15] LI, D. X.: *On Default Correlation: A Copula Function Approach*, Journal of Fixed Income, Vol. **9** No. **4**, pp. 43-54 (2000). DOI: [10.3905/jfi.2000.319253](https://doi.org/10.3905/jfi.2000.319253)
- [16] MARTINEZ-JARAMILLO, S., CASTANON, C. L., PÉREZ, O. P., EMBRIZ, F. A. AND DEY, F. L. G.: *Systemic Risk, Stress Testing and Financial Contagion: Their Interaction and Measurement*. BIS CCA Conference, (2010).
- [17] MOZSÁR, N. AND DR. CSÓKA, P.: *Hálózatelméleti modellek a banki rendszerkockázatra*, szakdolgozat, (2015).
- [18] NEMES, B. AND DR. CSÓKA, P.: *A rendszer összekapcsoltságának hatása a rendszerkockázatra*, szakdolgozat, (2012).
- [19] KISS, I. Z., MILLER, J. C. AND SIMON P. L.: *Mathematics of Epidemics on Networks: From Exact to Approximate Models*, Springer (2017). DOI: [10.1007/978-3-319-50806-1](https://doi.org/10.1007/978-3-319-50806-1)
- [20] SIMON, P. L., TAYLOR, M. AND KISS, I. Z.: *Exact epidemic models on graphs using graph-automorphism driven lumping*, Journal of Mathematical Biology, Vol. **62** No. **4**, pp. 479-508 (2010). DOI: [10.1007/s00285-010-0344-x](https://doi.org/10.1007/s00285-010-0344-x)
- [21] SUMMER, M.: *Banking Regulation and Systemic Risk*, Open Economies Review, Vol. **14** No. **1**, pp. 43-70 (2003). DOI: [10.1023/A:1021299202181](https://doi.org/10.1023/A:1021299202181)
- [22] VAN MIEGHEM, P., OMIC, J. AND KOOLJ, R.: *Virus spread in networks*, Networking, IEEE/ACM Transactions on, Vol. **17** No. **1**, pp. 1-14 (2009). DOI: [10.1109/TNET.2008.925623](https://doi.org/10.1109/TNET.2008.925623)
- [23] WU, J. L. AND YANG, W.: *Pricing CDO tranches in an intensity based model with the mean reversion approach*, Mathematical and Computer Modelling, Vol. **52** No. **5-6**, pp. 814-825 (2010). DOI: [10.1016/j.mcm.2010.05.012](https://doi.org/10.1016/j.mcm.2010.05.012)

(Beérkezett: 2018. február 17.)

BIHARY ZSOLT  
 Budapesti Corvinus Egyetem  
 Befektetések és Vállalati Pénzügy Tanszék  
 1093 Budapest, Fővám tér 8.  
 bihary0@gmail.com



Nagy Noémi 1985-ben született Budapesten. 2010-ben diplomázott az ELTE alkalmazott matematikus szakán, PhD-fokozatát 2016-ban szerezte meg az Eötvös Loránd Tudományegyetemen az alkalmazott matematika programban Simon Péter témavezetésével. Doktori tanulmányai alatt és utána is részt vett az MTA-ELTE Numerikus Analízis és Nagy Hálózatok Kutatócsoport munkájában. 2019-ig tanársegéd az ELTE Alkalmazott Analízis és Számításmatematikai Tanszékén, jelenleg a BME Ana-

lízis Tanszékén adjunktus. Kutatási területe hálózati folyamatok modellezése, illetve járványterjedési modellek analitikus és numerikus vizsgálata differenciálegyenletek segítségével. Négy angol nyelvű folyóiratcikke jelent meg, melyek eredményeit előadás, illetve poszter formájában ismertette több konferencián, például az Epidemics on Networks: Current Trends and Challenges konferencián Gironában (2012), vagy a The 20th European Conference on Mathematics for Industry konferencián Budapesten (2018).

#### NAGY NOÉMI

Eötvös Loránd Tudományegyetem (ELTE)  
Alkalmazott Analízis és Számításmatematikai Tanszék  
1117 Budapest, Pázmány P. sétány 1/C  
MTA-ELTE, Numerikus Analízis és Nagy Hálózatok Kutatócsoport  
nagynoemi0@gmail.com



Simon Péter 1966-ban született. Az Eötvös Loránd Tudományegyetem matematikus szakán 1990-ben szerzett diplomát, majd 1993-ban egyetemi doktori fokozatot. Mintegy 30 éve az Eötvös Loránd Tudományegyetem Természettudományi Kar Alkalmazott Analízis és Számításmatematikai Tanszékének munkatársa, jelenleg tanszékvezető egyetemi tanára. Két évet Nagy-Britanniában, a Leedsi Egyetemen töltött. 2011 és 2014 között az Eötvös Loránd Tudományegyetem Természettudományi Kar Matematikai Intézetének oktatási igazgatóhelyettese, 2017 óta pedig igazgatója. MTA doktori címét 2014-ben szerezte. Kutatási

területe a differenciálegyenletek kvalitatív elmélete, az utóbbi években hálózati folyamatok vizsgálatával foglalkozik. Két könyv és mintegy 90 tudományos közlemény szerzője. 2015 óta a Bolyai János Matematikai Társulat főtitkára.

#### SIMON L. PÉTER

Eötvös Loránd Tudományegyetem (ELTE)  
Alkalmazott Analízis és Számításmatematikai Tanszék  
1117 Budapest, Pázmány P. sétány 1/C  
MTA-ELTE, Numerikus Analízis és Nagy Hálózatok Kutatócsoport  
simonp@math.elte.hu

## NETWORK MODEL FOR JOINED DEFAULT PROBABILITIES

ZSOLT BIHARY, NOÉMI NAGY, PÉTER L. SIMON

We investigate the pricing of multi-name credit derivatives, namely Basket Default Swaps, based on network modeling. The aim of the work is to calculate the loss surface from the individual default probabilities. An intensity-based, dynamical network approach is introduced to handle this problem. We show that the network model yields comparable results to the industry-standard Gauss copula model. Moreover, an approximate network model is formulated to investigate systemic credit risk in large networks.

*Keywords:* network modeling, credit derivatives, systemic risk, credit default swap (CDS), basket default swap (BDS).

*Mathematics Subject Classification (2000):* 37N40, 90B15, 62P20, 91B30.

## EGY GYAKORI ZÁRT MINTÁKAT ÉS GYAKORI GENERÁTOROKAT KERESŐ VERTIKÁLIS ALGORITMUS

SZATHMÁRY LÁSZLÓ

A gyakori minták feltárására az *Apriori* a legismertebb algoritmus. Azonban az érdekes asszociációs szabályok előállításához kellenek még a gyakori zárt minták (GyZM-k) és a gyakori generátorok is (GyG-ok), melyek a gyakori minták részalmazai. A *Zart* egy *Apriori*-szerű algoritmus, mely képes kiszűrni a gyakori minták közül a GyZM-kat és a GyG-okat. Viszont ismeretes, hogy a vertikális mintakereső algoritmusok általában jobban teljesítenek, mint az *Apriori*-szerű szintenkénti algoritmusok. Az *Eclat* egy másik jól ismert vertikális mintakereső, mely ugyanazt a kimenetet állítja elő, mint az *Apriori*, vagyis egy adathalmazban megkeresi az összes gyakori mintát. Ebben a cikkben az *Eclat* egy kiterjesztését, az ún. *Eclog* algoritmust mutatjuk be, amely képes kiszűrni a gyakori minták közül a GyZM-kat és a GyG-okat. Az algoritmusunk egy menetben (utófeldolgozás nélkül) tárja fel egy adathalmaz ekvivalencia-osztályait. A teszteredmények azt mutatják, hogy az *Eclog* nagyon jól teljesít, különösen a sűrű, erősen korrelált adathalmazok esetén.

### 1. Bevezetés

Az adatbányászatban a gyakori minták kinyerése és az asszociációs szabályok feltárása fontos szerepet játszik [1]. A gyakori minták nagy száma miatt számos tömörített reprezentációt ajánlottak, melyek közül a gyakori generátorok (GyG-ok) és a gyakori zárt minták (GyZM-k) a legismertebbek [2, 3, 4]. A szakirodalomban több módszer is létezik a GyG-ok és a GyZM-k egyidejű kinyerésére, de ezen algoritmusok többsége szintenkénti megközelítést alkalmaz [5, 6]. Viszont a tapasztalat azt mutatja, hogy a mélységi elven működő algoritmusok általában jobban teljesítenek, mint a szintenkénti megfelelőik. Ebben a cikkben egy egylépcsős (utófeldolgozást nem igénylő), mélységi elven működő, vertikális algoritmust mutatunk be a GyG-ok és a GyZM-k együttes kinyerésére. Az algoritmusunk kimenetéből triviális módon lehet érdekes asszociációs szabályokat, pl. minimális nem redundáns szabályokat generálni. A minimális nem redundáns szabályokról és azok előállításáról egy korábbi cikkünkben írtunk [7]. Jelen cikkünk egy korábban angolul megjelent közleményünk [8] kibővített változata.

Ezen cikk a következőképpen épül fel. A 2. részben áttekintjük a mintakereséssel kapcsolatos alapvető fogalmakat és definíciókat. A 3. részben az *Eclog* algoritmust mutatjuk be részletesen. A pszeudokód után egy konkrét példán keresztül szemléltetjük az algoritmus működését. A 4. részben az algoritmus futási idejét mérjük össze három másik módszerrel. Végül az 5. részben a jövőbeli tervek következnek, majd a konklúziókkal zárjuk a cikket.

## 2. Fogalmak és definíciók

Vegyük a következő  $4 \times 6$ -os méretű adathalmazt:  $\mathcal{D} = \{(1, ACDE), (2, ABCDE), (3, ABE), (4, BEF)\}$ . A cikk további részében erre a példára  $\mathcal{D}$  adathalmaz néven fogunk hivatkozni.

Legyen  $\mathcal{O} = \{o_1, o_2, \dots, o_m\}$  objektumok (vagy tranzakciók) egy halmaza,  $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$  attribútumoknak egy halmaza,  $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{A}$  pedig egy bináris reláció. Az  $\mathcal{A}$  halmaz egy tetszőleges részhalmazát *mintának* nevezzük. Minden tranzakció rendelkezik egy egyedi azonosítóval (*tid*), s a tranzakció-azonosítók egy tetszőleges halmazát *tidset*nek hívjuk. Ha egy tidset valamennyi tranzakciója tartalmaz egy  $X$  mintát, akkor a tidsetet az  $X$  minta *képének* is nevezzük, s  $t(X)$ -szel jelöljük. Pl. a  $\mathcal{D}$  adathalmazban az  $\{A, B\}$  minta képe a  $\{2, 3\}$ , azaz egyszerűsített jelölést alkalmazva  $t(AB) = 23$ . Egy minta *hossza* alatt a minta kardinalitását értjük, egy  $k$  darab attribútumot tartalmazó mintát pedig  $k$ -hosszúságú mintának nevezzük. Egy  $X$  minta (abszolút) *support* értéke – jelölése  $\text{supp}(X)$  – az  $X$  minta képének a mérete, azaz  $\text{supp}(X) = |t(X)|$ . Vagyis a support azt mutatja meg, hogy egy minta hány tranzakcióban van jelen. Egy  $X$  minta *gyakori*, ha a support értéke nem kisebb, mint egy adott *minimum support* küszöbérték (jelölése  $\text{min\_supp}$ ), azaz  $\text{supp}(X) \geq \text{min\_supp}$ . Két minta  $X, Z \subseteq \mathcal{A}$  ekvivalens ( $X \cong Z$ ), ha a képtük megegyezik, vagyis  $t(X) = t(Z)$ . Egy  $P$  mintával ekvivalens minták halmazát a  $P$  minta ekvivalenciaosztályának nevezzük:  $[P] = \{Q \subseteq \mathcal{A} \mid P \cong Q\}$ . Egy ekvivalenciaosztálynak egyetlen maximális eleme van (zárt minta), s több minimális eleme is lehet (generátorok<sup>1</sup>) [9]. A generátorok a szabad minták egy speciális esetét képviselik [10].

*2.1. Definíció.* Egy minta *zárt*, ha nincs vele azonos support értékű valódi szuperhalmaza.

*2.2. Definíció.* Egy minta (minimális) *generátor*, ha nincs vele azonos support értékű valódi részhalmaza.

<sup>1</sup>A szakirodalomban ezen mintákat többféleképpen is nevezik: kulcsminták, minimális generátorok, szabad minták, kulcsgenerátorok stb.

A *lezárási* operátor egy  $X$  mintához az  $X$  ekvivalenciaosztályában található maximális elemet rendeli hozzá. A hozzárendelt elemet  $X$  *lezártjának* nevezzük, s  $\gamma(X)$ -szel jelöljük. Természetesen egy zárt  $X$  esetén  $X = \gamma(X)$ .

2.1. *Példa.* A  $\mathcal{D}$  adathalmazban  $B$  és  $C$  generátorok, melyek lezártjai:  $\gamma(B) = BE$ , és  $\gamma(C) = ACDE$  (lásd még 1. ábra).

2.1. *Tulajdonság.* Legyen  $X \subseteq \mathcal{A}$ . Ha  $X$  generátor, akkor  $\forall Y \subseteq X$ ,  $Y$  is generátor. Hasonlóképp, ha  $X$  nem generátor, akkor  $\forall Z \supseteq X$ ,  $Z$  sem generátor.

A GyZM-k és a GyG-ok jól ismert tömör reprezentációi [11] a gyakori mintáknak, s együtt alkotják az érvényes asszociációs szabályok nem redundáns bázisait (pl. a *generikus bázist*) [3].

### 3. Generátorok és zárt minták kiszűrése a gyakori minták közül mélységi bejárás alkalmazásával

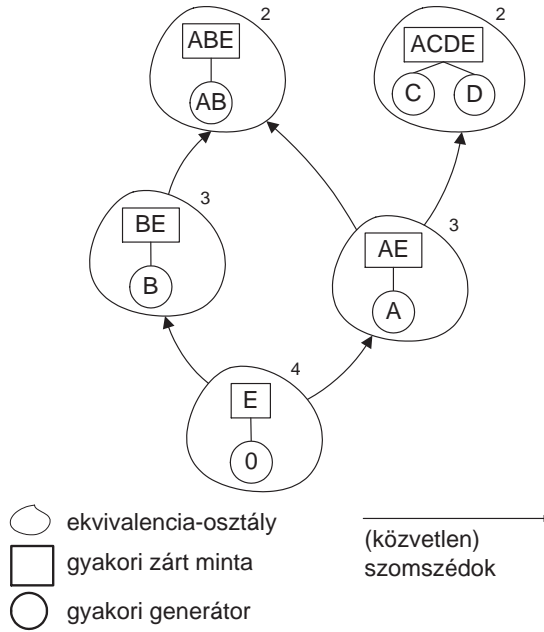
Az *Eclog* algoritmusunk, mely jelen cikkünk központi témája, egy vertikális mintakereső algoritmus, mellyel fel lehet fedezni egy adathalmazban a gyakori ekvivalenciaosztályokat. Ebben a fejezetben először egy áttekintést nyújtunk az algoritmusról. Ezt követően megvizsgálunk néhány vertikális elven működő algoritmust, név szerint az *Eclat*ot [12] és a *Talky*t [13]. Végül részletesen is bemutatjuk az *Eclog* algoritmust.

#### 3.1. Az *Eclog* általános áttekintése

Az *Eclog* a *Talky* [13] algoritmusra épül, míg a *Talky* az *Eclat* [12] egy módosított változata. Az *Eclat* és a *Talky* ugyanazt a kimenetet produkálják, vagyis egy adathalmaz összes gyakori mintáját keresik meg. Azonban a *Talky* egy eltérő bejárési stratégiát alkalmaz, az ún. fordított preorder módszert. Ez a bejárás jobbról balra halad, s rendelkezik egy speciális tulajdonsággal: amikor elérünk egy  $X$  mintához, addigra már felfedeztük  $X$  valamennyi részalmazát. Ennek eredményeképpen ez a bejárás felhasználható a GyG-ok hatékony kiszűrésére. A bejárás során az *Eclog* a GyZM-kat is kiszűri, és hozzárendeli őket a megfelelő GyG-okhoz. Így mire az *Eclog* befejezi a futását, addigra feltárta az összes gyakori ekvivalenciaosztályt. A GyG-ok hatékony módon való kiszűrésére az algoritmus a fordított preorder stratégiára támaszkodik, melyet a következő alfejezetekben mutatunk be.

#### 3.2. Vertikális mintakeresés

A mintakereső algoritmusok – akár az összes gyakori mintát keresik vagy csak a GyZM-kat – általában két osztályba sorolhatók: vannak a szélességi



1. ábra. A  $\mathcal{D}$  adathalmaz ekvivalenciaosztályai  $\min\_supp = 2$  küszöbérték mellett. A support értékek az osztályok jobb felső sarkában láthatók. Az  $E$  minta generátora az üres halmaz.

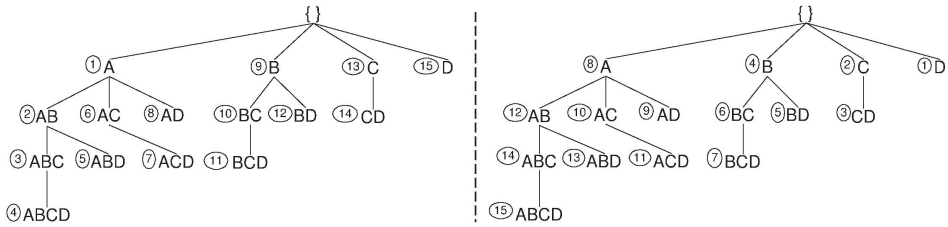
keresők, ill. vannak a mélységi keresők. A szélességi elven működő algoritmusok, egész pontosan az *Apriori*-szerű algoritmusok [1], a keresési teret szintenként járják be, kihasználva a minták gyakoriságának antimonotonitását<sup>2</sup>. A mélységi algoritmusok – pl. *Closet* [14] – ezzel szemben a keresési teret egy prefixába szervezik (lásd 2. ábra). Ezen algoritmusok közül a *vertikális* elven működők az adathalmazt (minta, tidset) párokként reprezentálják ( $\{(i, t(i)) \mid i \in \mathcal{A}\}$ ), mellyel el tudják kerülni az adathalmaz költséges többszöri végigolvasását.

Az *Eclat* [12] volt az első olyan gyakori mintákat kereső algoritmus, amely sikeresen kombinálta a mélységi keresést és az adathalmaz vertikális reprezentációját, s ezt egy fa adatszerkezettel, egy ún. IT-fával oldotta meg, melyben a csúcsok  $X \times t(X)$  párok.<sup>3</sup> Az *Eclat* az IT-fát mélységi módon, preorder stratégiával, balról jobbra haladva járja be [12, 15] (lásd 2. ábra).

<sup>2</sup>Az antimonotonitás elve szerint ha egy minta nem gyakori, akkor annak egyetlen szuperhalmaza sem lehet gyakori [1].

<sup>3</sup>Az IT-fa esetén az IT az *itemset-tidset* rövidítése, ami arra utal, hogy egy csúcsban egy minta és a hozzá tartozó tidset helyezkedik el.





2. ábra. Bal oldal: az *Eclat* preorder bejárása; jobb oldal: a *Talky* fordított preorder bejárása. A bejárési sorrend a körökben van feltüntetve.

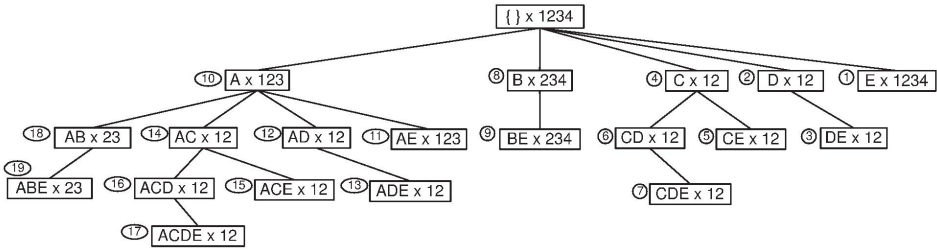
### 3.3. Fordított preorder bejárás

Az *Eclog* a *Talky* algoritmust terjeszti ki oly módon, hogy a gyakori minták közül kiszűri a GyG-okat és a GyZM-kat. Vagyis az *Eclog* megvizsgál minden újonnan talált gyakori mintát, hogy generátor-e. Ahhoz, hogy ezt a tesztet hatékonyan tudjuk elvégezni, egy  $X$  minta esetén már  $X$  összes részalmazát érintenünk kellett. A korábban látott 2.1. tulajdonság alapján egy minta csak akkor lehet generátor, ha az összes részalmazja is generátor. Egy  $X$  minta generátor státuszának megállapításához ezért szükségünk lesz az összes részalmazának előzetes ismeretére. A szélességi keresést alkalmazó keresőalgoritmusok triviális módon biztosítják a minták helyes sorrendben való feltárását. Viszont a kívánt sorrendet egy mélységi keresésen alapuló algoritmussal is el lehet érni, bár ekkor egy speciális bejárásra van szükség.

A kombinatorikus algoritmusok esetén gyakori elvárás, hogy a keresési tér bejárása során egy adott  $X$  mintát az összes részalmazja után dolgozzunk fel. A szintenkénti algoritmusok triviális módon elégítik ki ezt a feltételt. A *fordított preorder bejárást* [16] alapul véve, az adathalmaz attribútumait fordított lexikografikus sorrendbe tesszük ( $E, D, C$  stb.). Ha az attribútumok numerikus megfelelőit növekvő sorrendben követjük, akkor egy mélységi, jobbról balra haladó prefixfa bejárást kapunk, ahol a prefixfa a keresési teret reprezentálja. Mivel a csúcsokat a gyermekcsúcsok előtt dolgozzuk fel, ezért *preorder* (és nem *postorder*) bejárásról beszélünk.

Összefoglalva: a módszerünk az IT-fát preorder módon járja be jobbról balra. Így, ha veszünk egy tetszőleges  $X$  mintát az IT-fában, a bejárás garantálni fogja, hogy az  $X$  csúcs *előtt* már az  $X$  összes részalmazát tartalmazó csúcsot felfedeztük.

3.1. *Példa.* A 2. ábrán látható egy összehasonlítás a kétféle bejárásról. Bal oldalt az *Eclat* preorder bejárása, míg jobb oldalt a *Talky* fordított preorder bejárása szerepel.



3. ábra. A *Talky* algoritmus végrehajtása a  $\mathcal{D}$  adathalmazon  $min\_supp = 2$  küszöbérték mellett. A csúcsok feldolgozási sorrendje a körökben van feltüntetve.

### 3.4. A *Talky* algoritmus

A *Talky* egy vertikális, gyakori mintákat kereső algoritmus, amely a felépített IT-fát mélységi módon, fordított preorder bejárással dolgozza fel (lásd 3. ábra). A  $\mathcal{D}$  adathalmazunkból  $min\_supp = 2$  küszöbérték mellett a *Talky* a következő 19 gyakori mintát fedezi fel ebben a sorrendben<sup>4</sup>:  $E$  (4),  $D$  (2),  $DE$  (2),  $C$  (2),  $CE$  (2),  $CD$  (2),  $CDE$  (2),  $B$  (3),  $BE$  (3),  $A$  (3),  $AE$  (3),  $AD$  (2),  $ADE$  (2),  $AC$  (2),  $ACE$  (2),  $ACD$  (2),  $ACDE$  (2),  $AB$  (2) és  $ABE$  (2).

### 3.5. Az *Eclog* részletes bemutatása

Ebben az alfejezetben az *Eclog* algoritmust mutatjuk be részletesen. Mint már korábban láttuk, az *Eclog* alapja a *Talky*. Az *Eclog* az IT-fát fordított preorder módon járja be (lásd 3. ábra), s a gyakori minták feltárása közben kiszűri a GyG-okat és a GyZM-kat. Az *Eclog* a generátorokat hozzákapcsolja a lezártjukhoz, így kimenetként a gyakori ekvivalenciaosztályok listáját állítja elő (lásd 1. táblázat).

Az *Eclog* egy hash táblázatot épít fel (lásd 1. táblázat). A hash táblázat kulcsa egy tidset, míg a kulcshoz tartozó érték egy sor objektum. A táblázat egy sora egy ekvivalenciaosztályt reprezentál, s a következő mezőkből épül fel: **(1)** tidset (definíció szerint egy ekvivalenciaosztály valamennyi elemének azonos a tidset-je), **(2)** generátorok (egy ekvivalenciaosztály minimális elemei), **(3)** ekvivalenciaosztály tagok (egy ekvivalenciaosztály azon elemei, melyek se nem generátorok, se nem zárt minták), **(4)** lezárt (egy ekvivalenciaosztály maximális eleme; ilyenből minden osztályban egy és csakis egy van), ill. **(5)** support (ez a tidset kardinalitása).

Az algoritmus a következőképpen működik. Amikor az algoritmus feltár egy új gyakori mintát az IT-fában, akkor megnézi, hogy a minta egy már korábban felfedezett ekvivalenciaosztályhoz tartozik-e, vagyis megnézzük, hogy a minta tidsetje szerepel-e a hash táblázatban. Ha nincs benne a hash táblázatban, akkor

<sup>4</sup>A zárójelben a support értékek szerepelnek.

1. táblázat. Az *Eclog* egy ilyen táblázatot épít fel, ami valójában egy hash táblázat, ahol a kulcs egy tidset, az érték pedig a táblázat egy sora.

tidset	generátorok	ekv. osztály tagok (opcionális)	lezárt	support
1234	$\emptyset$	$E$	$E$	4
12	$D, C$	$DE, CE, CD, CDE,$ $AD, ADE, AC, ACE,$ $ACD, ACDE$	$ACDE$	2
234	$B$	$BE$	$BE$	3
123	$A$	$AE$	$AE$	3
23	$AB$	$ABE$	$ABE$	2

egy új ekvivalenciaosztályba tartozik, s így a hash táblázatba egy új sort szúrunk be. Ha a tidsetje benne van a hash táblázatban, akkor két eset lehetséges. Jelöljük  $R$ -rel azt a sort, amelynek tidsetje megegyezik az aktuális minta tidset értékével, vagyis  $R$  azt az ekvivalenciaosztályt reprezentálja, amelybe az aktuális minta tartozik.

**Első eset.** Az aktuális mintának van egy valódi részhalmaza az  $R$  sor „generátorok” mezőjében. Ez azt jelenti, hogy a minta nem generátor, de ebbe az ekvivalenciaosztályba tartozik. Az  $R$  „lezárt” mezőjét frissíteni kell a mintával. Ez a frissítés a következőből áll: ha a minta nagyobb, mint az eddig regisztrált lezárt, akkor a lezártat lecseréljük az új mintára. Mint ismeretes, egy ekvivalenciaosztály lezártja az ekvivalenciaosztály legnagyobb eleme, ami egy egyedi elem (vagyis egy ekvivalenciaosztályban pontosan egy ilyen elem van). Mivel az *Eclog* megkeresi az összes gyakori mintát, ezért az algoritmus befejeződésekor a „lezárt” mezőben garantált módon minden sorban az adott ekvivalenciaosztály lezártja fog szerepelni. Opcionálisan egy ekvivalenciaosztály azon elemeit, amelyek nem generátorok, össze lehet gyűjteni az „ekv. osztály tagok” mezőben. A 1. táblázatban ezt a mezőt a könnyebb érthetőség kedvéért tüntettük fel, de ha csak a GyG-okra és a GyZM-kra van szükségünk, akkor az implementáció során ezt a mezőt üresen lehet hagyni.

**Második eset.** Az aktuális mintának nincs valódi részhalmaza az  $R$  sor „generátorok” mezőjében. Ez azt jelenti, hogy a minta az ekvivalenciaosztály egy új generátora, így hozzáadjuk a „generátorok” mezőhöz. Ha a generátor nagyobb, mint a sor lezártja, akkor a „lezárt” mezőt frissíteni kell (az előző esetben leírtak szerint).

Amikor az algoritmus befejezi a futását, a „lezárt” mezőben lévő minták mindegyike végleges, vagyis az ekvivalenciaosztályok tényleges lezártjait tartalmazzák. Az *Eclog* pszeudokódját az 1. algoritmus szemlélteti.

2. táblázat. Az *Eclog* hash táblázatának inicializálása.

tidset	generátorok	ekv. osztály tagok (opcionális)	lezárt	support
1234	$\emptyset$		$\emptyset$	4

3. táblázat. Az *Eclog* hash táblázatának állapota az  $E$ ,  $D$  és  $DE$  minták beszúrása után.

tidset	generátorok	ekv. osztály tagok (opcionális)	lezárt	support
1234	$\emptyset$	$E$	$E$	4
12	$D$	$DE$	$DE$	2

3.2. *Példa.* A  $\mathcal{D}$  adathalmazunk egy kissé speciális esetnek tekinthető, ui. az adathalmaz  $E$  oszlopa telített. Ez azt jelenti, hogy az  $E$  minta nem generátor, mivel van egy vele azonos support értékű részhalmaza, nevezetesen az üres halmaz. Definíció szerint az üres halmaz valamennyi sorban jelen van, így a support értéke 100%. Ezek alapján a hash táblázatot a 2. táblázatban látható módon inicializáljuk. A hash táblázat valójában bármilyen input adathalmaz esetén inicializálható így, de ha az adathalmazban nincs telített oszlop, akkor az üres halmaz lezárta végig az üres halmaz fog maradni.

Ezután az algoritmus elkezdi felsorolni a  $\mathcal{D}$  adathalmaz 19 gyakori mintáját a *Talky* bejárési stratégiáját alkalmazva (az adathalmaz gyakori mintáit a 3.4-es alfejezetben tüntettük fel). Az első csúcs az  $E \times 1234$ . Az 1234 tidset már egy létező kulcs a hash táblázatban. Az  $E$  mintának van egy vele azonos support értékű részhalmaza (az üres halmaz), így az  $E$  mintát hozzáadjuk az „ekv. osztály tagok” oszlophoz, ill. frissítjük a „lezárt” mezőt az  $E$  mintával. A következő gyakori minta a  $D \times 12$ . Mivel a 12 még nincs benne a hash táblázatban, ezért a táblázatba egy új sort kell beszúrni. A következő csúcs a  $DE \times 12$ . Mivel az 12 tidset már szerepel a hash táblázatban, ezért a  $DE$  egy már felfedezett ekvivalenciaosztályhoz tartozik. Van neki egy valódi részhalmaza, a  $D$ , így a  $DE$ -t hozzáadjuk az „ekv. osztály tagok”-hoz. Mivel a  $DE$  nagyobb, mint a jelenleg regisztrált lezárt ( $D$ ), ezért a „lezárt” mezőt frissítjük a  $DE$  mintával. A hash táblázat aktuális állapota a 3. táblázatban található.

Az algoritmus hátralévő részének részletes bemutatását kihagyjuk. Az *Eclog* kimenete – a végeredmény – az 1. táblázatban látható.

**Algoritmus 1** (az *Eclog* pszeudokódja):

*hashTable*: az 1. táblázatban látható hash táblázat; kezdetben üres

```

1) // inicializálás
2) sor.tidset  $\leftarrow$  {legnagyobb tidset} // a példában: 1234
3) sor.generatorok  $\leftarrow$   $\emptyset$  // üres halmaz hozzáadása
4) sor.lezart  $\leftarrow$   $\emptyset$ 
5) sor.support  $\leftarrow$  |sor.tidset| // kardinalitás
6) hashTable.add(sor)
7)
8) // fő blokk
9) indítsuk el a Talky algoritmust s az aktuális csúcst rendeljük hozzá
    az akt nevű változóhoz
10) {
11)     if akt.tidset nincs benne a hashTable-ben:
12)         sor.tidset  $\leftarrow$  akt.tidset
13)         sor.generatorok.add(akt.minta)
14)         sor.lezart  $\leftarrow$  akt.minta
15)         sor.support  $\leftarrow$  |sor.tidset|
16)         hashTable.add(sor)
17)     else:
18)         sor  $\leftarrow$  hashTable.get(akt.tidset)
19)         if akt.minta-nak van egy valódi részhalmaza a sor.generatorok-ban:
20)             sor.ekv_oszt_tagok.add(akt.minta) // opcionális
21)             if |akt.minta| > |sor.lezart|:
22)                 sor.lezart  $\leftarrow$  akt.minta
23)             endif
24)         else:
25)             sor.generatorok.add(akt.minta)
26)             if |akt.minta| > |sor.lezart|:
27)                 sor.lezart  $\leftarrow$  akt.minta
28)             endif
29)         endif
30)     endif
31) }
```

#### 4. Teszteredmények

Tesztjeink során az *Eclog* algoritmust az *Eclat*, *Zart* és *LCM+GrGrowth* algoritmusokkal mértük össze. Az *Eclog*, *Eclat* és *Zart* algoritmusokat Java-ban implementáltuk a CORON adatbányász platformban [17].<sup>5</sup> Mivel az *Eclog* alapja az *Eclat*, ezért az összehasonlításba betettük az *Eclat*ot is, ui. kíváncsiak voltunk, hogy az *Eclog* extra műveletei okoznak-e érezhető teljesítménycsökkenést. A *Zart* [7] egy szintenkénti algoritmus, mely a *Pascal*-algoritmus [9] kiterjesztése. A *Zart*, köszönhetően az ún. mintaszámláló következtetésének, nagyon hatékonyan tudja megkeresni a gyakori ekvivalenciaosztályokat. Az *LCM* az egyik leggyorsabb GyZM-kat kereső algoritmus [18], míg a *GrGrowth* az egyik leghatékonyabb GyG-okat előállító algoritmus [19]. Az *Eclog*gal való tisztességes összehasonlítás kedvéért összefogtuk a két algoritmust, s az eredményre *LCM+GrGrowth* néven hivatkozunk. Ez a kombinált algoritmus három lépést valósít meg: kigyűjti a GyZM-kat az *LCM*-mel; kigyűjti a GyG-okat a *GrGrowth*-szal; majd végül egy utófeldolgozás során összepárosítja a GyG-okat és a GyZM-kat, így a kimenet a gyakori ekvivalenciaosztályok listája lesz. Az *LCM* és a *GrGrowth* algoritmusok esetén a szerzők eredeti forráskódját használtuk fel. Az *LCM* C nyelven, míg a *GrGrowth* C++ nyelven lett implementálva.

A tesztekét egy Intel Quad Core Xeon 2,33 GHz-es gépen végeztük el Ubuntu GNU/Linux operációs rendszer alatt. Az adott gép 8 GB RAM-mal rendelkezett. Valamennyi válaszidő valós idő, melyet a Unix rendszereken használt *time* paranccsal mértünk le a bemenet és a kimenet között. A tesztekhez a következő adathalmazokat használtuk fel: T20I6D100K, C20D10K, C73D10K és MUSHROOMS. A T20<sup>6</sup> egy ritka adathalmaz, s felépítésében a bevásárlóközpontok adatbázisaira hasonlít, melyek általában gyengén korrelált adatokat tartalmaznak. A C20 és C73 egy népszámlálás részadatait tartalmazza, míg a MUSHROOMS<sup>7</sup> különböző gombok jellemzőit írja le. Ez utóbbi három adathalmaz sűrű és erősen korrelált.

Az említett négy algoritmus futási idejét a 4. táblázatban foglaltuk össze. A táblázatban szintén megtalálható a gyakori minták, a gyakori generátorok, ill. a gyakori zárt minták száma. Az *Eclat* megkeresi, s a kimenetében megjeleníti az összes gyakori mintát. Az *Eclog* és a *Zart* ugyan bejárja a keresési térben az összes gyakori mintát, de a kimenetükben már csak a GyG/GyZM párokat jelenítik meg. Az *LCM+GrGrowth* lecsökkenti a keresési teret a GyZM-kra és a GyG-okra, és az *Eclog*hoz és a *Zarthoz* hasonlóan csak a GyG/GyZM párokat jeleníti meg a kimenetében.

---

<sup>5</sup><http://coron.loria.fr>

<sup>6</sup><http://www.almaden.ibm.com/software/quest/Resources/>

<sup>7</sup><http://kdd.ics.uci.edu/>

4. táblázat. Az *Eclog* futási ideje, ill. egyéb statisztikák (gyakori minták (GyM-k) száma, gyakori generátorok (GyG-ok) száma, gyakori zárt minták (GyZM-k) száma).

min_supp	futási idő (mp.)				# GyM	# GyG	# GyZM
	Eclog	Eclat	Zart	LCM+GrGrowth			
<b>T20I6D100K</b>							
10%	0,97	0,97	0,80	0,36	7	7	7
0.75%	1,26	1,24	9,64	1,64	4710	4710	4710
0.50%	1,79	1,75	16,83	2,32	26 836	26 305	26 208
0.25%	5,25	4,43	41,82	7,31	155 163	149 447	149 217
<b>C20D10K</b>							
30%	0,49	0,57	2,13	0,29	5319	967	951
20%	0,63	0,74	3,85	0,44	20 239	2671	2519
10%	0,83	1,10	7,72	0,67	89 883	9331	8777
5%	1,14	2,49	14,64	0,87	352 611	23 051	21 213
2%	2,63	9,33	45,53	1,41	1 741 883	57 659	50 729
<b>C73D10K</b>							
95%	0,68	0,69	1,84	0,24	1007	121	93
90%	0,88	0,99	13,99	0,39	13 463	1368	942
85%	0,94	1,24	35,56	0,49	46 575	3513	2359
<b>MUSHROOMS</b>							
40%	0,36	0,38	0,72	0,17	505	153	124
30%	0,42	0,46	1,22	0,22	2587	544	425
15%	0,65	1,07	3,83	0,43	99 079	3084	2210
10%	1,23	3,33	13,47	0,56	600 817	7585	4850

A T20 szintetikus adathalmaz a bevásárlóközpontok adatbázisait utánozza, így egy ritka és gyengén korrelált adathalmazról van szó. Ezen adathalmazban kevés a gyakori minta, s majdnem valamennyi gyakori minta GyZM és GyG egyszerre, ami azt jelenti, hogy a legtöbb ekvivalenciaosztály egyelemű. Az *Eclog* és *Eclat* hasonló módon viselkedik, bár az *Eclog* extra műveletei nagyon kis mértékben lassítanak az algoritmuson. Az *Eclog* sokkal jobban teljesít a *Zart*nál, s ez a különbség különösen akkor látványos, amikor a *min\_supp* érték nagyon alacsonyra van állítva. Mivel az *LCM+GrGrowth* algoritmusnak főleg egyelemű ekvivalenciaosztályokkal kell dolgoznia, ezért körülbelül ugyanazt a keresési teret kell bejárnia, mint az *Eclog*nak. A tesztek alapján az *LCM+GrGrowth* egy kicsit lassabban teljesít, mint az *Eclog*.

A C20, C73 és MUSHROOMS adathalmazokban a GyG-ok és a GyZM-k száma sokkal kevesebb, mint az összes gyakori minta száma. Az *Eclog* és *Eclat* hasonlóan teljesít magas *min\_supp* érték esetén. Azonban ahogy csökkentjük a küszöbértéket, úgy nő a két algoritmus futási ideje közti különbség. Ez az *Eclat* nagyobb méretű kimenetével magyarázható. Például a C20 adathalmazon  $min\_supp = 2\%$  esetén

az *Eclat* 1 741 883 gyakori mintát produkál kimenetként, míg az *Eclog* csupán 57 659 GyG-t és 50 729 GyZM-t. Habár a két algoritmus ugyanazt a keresési teret járja be, ill. az *Eclog* még extra szűréseket is végez, az *Eclat* esetén alacsony *min\_supp* mellett nagyon megnő az I/O műveletek száma, ami negatív hatással van az algoritmus teljesítményére. Az *Eclog* és *Zart* esetén a különbség még a korábbi T20-nál is szembetűnőbb. Amint látható, az *LCM+GrGrowth* sűrű adathalmazok esetén egy kicsit jobban teljesít az *Eclog*-nál. Ez azzal a ténnyel magyarázható, hogy az *LCM+GrGrowth*-nak sokkal kisebb keresési teret kell bejárnia.

Összefoglalásképpen azt mondhatjuk, hogy ritka adathalmazok esetén az *Eclog* extra jellemzői nem okoznak szinte semmiféle teljesítménycsökkenést az *Eclat*hoz képest, míg sűrű adathalmazok esetén az *Eclog* jobb teljesítményt nyújt az *Eclat*-nál, köszönhetően a kisebb kimenetének. Az *Eclog* minden esetben felülmúlja a *Zart* teljesítményét, ami összhangban áll azzal az általános nézettel, mely szerint a vertikális algoritmusok általában hatékonyabbak, mint a szintenkénti algoritmusok. Az *Eclog*-ot szintén összehasonlítottuk az egyik leghatékonyabb algoritmussal, az *LCM+GrGrowth*-szal. A ritka adathalmazok esetén az *Eclog* volt gyorsabb, míg a sűrű adathalmazoknál az *LCM+GrGrowth* egy kicsit hatékonyabbnak bizonyult, de az *Eclog* és az *LCM+GrGrowth* közti különbség nem túl jelentős. Az *LCM+GrGrowth* kicsit jobb teljesítménye a következőkkel magyarázható: **(a)** az *LCM* és a *GrGrowth* a mintakereső algoritmusok között a leghatékonyabb megoldások közé tartoznak, **(b)** C/C++-ban lettek implementálva, ill. **(c)** sűrű adathalmazok esetén egy sokkal kisebb keresési teret kell feltárniuk. Mindent egybevetve úgy érezzük, hogy az *Eclog*-nak nincs oka szégyenkeznie az *LCM+GrGrowth* mellett.

## 5. Konklúzió és jövőbeli tervek

Ebben a cikkben egy vertikális, mélységi elven működő algoritmust mutattunk be, mely kimenetként a GyG/GyZM párokat állítja elő, vagyis egy adathalmazban a gyakori ekvivalenciaosztályokat tárja fel. Az *Eclog*-ot az első lépésnek szántuk egy egylépcsős, vertikális elven működő, GyG-okat + GyZM-kat kereső algoritmus tervezéséhez vezető úton. Ehhez egy ismert, gyakori mintákat kereső algoritmust vettünk alapul, amit kiegészítettünk a GyZM-k és a GyG-ok szűrésének a képességével, s azt is sikerült elérnünk, hogy a GyG-ok menet közben rendelődnek hozzá a megfelelő GyZM-khoz, így nincs szükség utófeldolgozásra. Az *Eclog* algoritmus a jövőben könnyen felkészíthető asszociációs szabályok bányászatára is, hiszen a jelenlegi kimenetéből már most is triviális módon előállítható például a generikus bázis [3]. Ha az *Eclog* kimenetét kombináljuk egy olyan algoritmussal, amely képes megállapítani az ekvivalenciaosztályok közti közvetlen és tranzitív kapcsolatokat, akkor egy komplett megoldást kaphatunk az ún. minimális nem redundáns szabályok előállításához [3].

A közeljövőben olyan stratégiákat szeretnénk megvizsgálni, melyekkel csökkenteni lehetne az *Eclog* keresési terét. Ezzel függ össze az az érdekes kérdés,



hogyan lehetne úgy meghatározni egy ekvivalenciaosztály lezártját, hogy nem állítjuk elő az adott ekvivalenciaosztály összes elemét. Egy lehetséges út lehetne a mintaszámláló következtetés [9] alkalmazása, míg egy másik lehetőség az lenne, ha egy ekvivalenciaosztályban csak a kanonikus GyG-okat keresnénk meg, s ezekből állítanánk elő az ekvivalenciaosztály lezártját.

### Köszönetnyilvánítás

A publikáció elkészítését az EFOP-3.6.1-16-2016-00022. számú projekt támogatta. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

### Hivatkozások

- [1] AGRAWAL, R. AND SRIKANT, R.: *Fast Algorithms for Mining Association Rules in Large Databases*, In: Proc. of the 20th Intl. Conf. on Very Large Data Bases (VLDB '94), San Francisco, CA, Morgan Kaufmann, pp. 487-499 (1994).
- [2] BASTIDE, Y., TAOUIL, R., PASQUIER, N., STUMME, G. AND LAKHAL, L.: *Mining Minimal Non-Redundant Association Rules Using Frequent Closed Itemsets*, In: Proc. of the Computational Logic (CL '00). Volume 1861 of LNAI., Springer, pp. 972-986 (2000). DOI: [10.1007/3-540-44957-4\\_65](https://doi.org/10.1007/3-540-44957-4_65)
- [3] KRYSZKIEWICZ, M.: *Concise Representations of Association Rules*, In: Proc. of the ESF Exploratory Workshop on Pattern Detection and Discovery, pp. 92-109 (2002). DOI: [10.1007/3-540-45728-3\\_8](https://doi.org/10.1007/3-540-45728-3_8)
- [4] AGGARWAL, C.C. AND HAN, J., eds.: *Frequent Pattern Mining*, Springer International Publishing (2014). DOI: [10.1007/978-3-319-07821-2](https://doi.org/10.1007/978-3-319-07821-2)
- [5] PASQUIER, N., BASTIDE, Y., TAOUIL, R. AND LAKHAL, L.: *Discovering Frequent Closed Itemsets for Association Rules*, In: Proc. of the 7th Intl. Conf. on Database Theory (ICDT '99), Jerusalem, Israel, pp. 398-416 (1999). DOI: [10.1007/3-540-49257-7\\_25](https://doi.org/10.1007/3-540-49257-7_25)
- [6] STUMME, G., TAOUIL, R., BASTIDE, Y., PASQUIER, N. AND LAKHAL, L.: *Computing Iceberg Concept Lattices with Titanic*, Data and Knowl. Eng., Vol. **42** No. **2**, pp. 189-222 (2002). DOI: [10.1016/S0169-023X\(02\)00057-5](https://doi.org/10.1016/S0169-023X(02)00057-5)
- [7] SZATHMARY, L.: *ZART: egy multifunkcionális mintakereső algoritmus*, Alkalmazott Matematikai Lapok, Vol. **27** No. **2**, pp. 107-122 (2010).
- [8] SZATHMARY, L.: *Finding frequent closed itemsets with an extended version of the Eclat algorithm*. Annales Mathematicae et Informaticae, Vol. **48**, pp. 75-82 (2018).
- [9] BASTIDE, Y., TAOUIL, R., PASQUIER, N., STUMME, G. AND LAKHAL, L.: *Mining frequent patterns with counting inference*, SIGKDD Explor. Newsl, Vol. **2** No. **2**, pp. 66-75 (2000). DOI: [10.1145/380995.381017](https://doi.org/10.1145/380995.381017)
- [10] BOULICAUT, J.F., BYKOWSKI, A. AND RIGOTTI, C.: *Free-Sets: A Condensed Representation of Boolean Data for the Approximation of Frequency Queries*, Data Mining and Knowledge Discovery, Vol. **7** No. **1**, pp. 5-22 (2003). DOI: [10.1023/A:1021571501451](https://doi.org/10.1023/A:1021571501451)
- [11] CALDERS, T., RIGOTTI, C. AND BOULICAUT, J.F.: *A Survey on Condensed Representations for Frequent Sets*, In Boulicaut, J.F., Raedt, L.D., Mannila, H., eds.: Constraint-Based Mining and Inductive Databases, Volume 3848 of Lecture Notes in Computer Science., Springer, pp. 64-80 (2004). DOI: [10.1007/11615576\\_4](https://doi.org/10.1007/11615576_4)

- [12] ZAKI, M.J., PARTHASARATHY, S., OGIHARA, M. AND LI, W.: *New Algorithms for Fast Discovery of Association Rules*, In: Proc. of the 3rd Intl. Conf. on Knowledge Discovery in Databases, pp. 283-286 (1997). DOI: [10.1007/978-1-4615-5669-5\\_1](https://doi.org/10.1007/978-1-4615-5669-5_1)
- [13] SZATHMÁRY, L., VALTCHEV, P., NAPOLI, A. AND GODIN, R.: *Efficient Vertical Mining of Frequent Closures and Generators*, In: Proc. of the 8th Intl. Symposium on Intelligent Data Analysis (IDA '09). Volume 5772 of LNCS., Lyon, France, Springer, pp. 393-404 (2009). DOI: [10.1007/978-3-642-03915-7\\_34](https://doi.org/10.1007/978-3-642-03915-7_34)
- [14] PEI, J., HAN, J. AND MAO, R.: *CLOSET: An Efficient Algorithm for Mining Frequent Closed Itemsets*, In: ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, pp. 21-30 (2000).
- [15] ZAKI, M.J.: *Scalable Algorithms for Association Mining*, IEEE Transactions on Knowledge and Data Engineering, Vol. **12** No. **3**, pp. 372-390 (2000). DOI: [10.1109/69.846291](https://doi.org/10.1109/69.846291)
- [16] CALDERS, T. AND GOETHALS, B.: *Depth-first non-derivable itemset mining*, In: Proc. of the SIAM Intl. Conf. on Data Mining (SDM '05), Newport Beach, USA. (2005). DOI: [10.1137/1.9781611972757.23](https://doi.org/10.1137/1.9781611972757.23)
- [17] SZATHMÁRY, L.: *Symbolic Data Mining Methods with the Coron Platform*. PhD Thesis in Computer Science, Univ. Henri Poincaré - Nancy 1, France (2006).
- [18] UNO, T., KIYOMI, M. AND ARIMURA, H.: *LCM ver. 2: Efficient Mining Algorithms for Frequent/Closed/Maximal Itemsets*, In Jr., R.J.B., Goethals, B., Zaki, M.J., eds.: FIMI. Volume 126 of CEUR Workshop Proceedings., CEUR-WS.org (2004).
- [19] LIU, G., LI, J. AND WONG, L.: *A new concise representation of frequent itemsets using generators and a positive border*. Knowl. Inf. Syst., Vol. **17** No. **1**, pp. 35-56 (2008). DOI: [10.1007/s10115-007-0111-5](https://doi.org/10.1007/s10115-007-0111-5)

(Béérkezett: 2019. február 22.)



Szathmáry László 1977-ben született Debrecenben. A Debreceni Egyetemen 2001-ben szerzett programtervező-informatikus / angol-magyar szakfordító diplomát. Informatikai PhD-fokozatát a francia Henri Poincaré Egyetemen szerezte meg 2006-ban, Nancyban. Ezután posztdoktori ösztöndíjat nyert Kanadába a montreáli UQAM egyetemre. Magyarországra 2012-ben tért vissza, s azóta a Debreceni Egyetem Informatikai Karán dolgozik. Jelenlegi beosztása: habilitált egyetemi docens. Főbb kutatási területei: szimbolikus adatbányászat, mesterséges intelligencia, tudásreprezentáció.

SZATHMÁRY LÁSZLÓ

Debreceni Egyetem, Informatikai Kar  
 Információ Technológia Tanszék  
 4002 Debrecen, Pf. 300  
 szathmary.laszlo@inf.unideb.hu

A VERTICAL ALGORITHM FOR FINDING FREQUENT CLOSED ITEMSETS AND  
FREQUENT GENERATORS

LÁSZLÓ SZATHMÁRY

*Apriori* is the most well-known algorithm for finding frequent itemsets (FIs) in a dataset. However, for generating interesting association rules, we also need the so-called frequent closed itemsets (FCIs) and the frequent generators (FGs), where FCIs and FGs form a subset of FIs. *Zart* is an *Apriori*-like algorithm that can filter FCIs and FGs among FIs. However, it is known that vertical itemset mining algorithms outperform the *Apriori*-like levelwise algorithms. *Eclat* is another well-known vertical miner that can produce the same output as *Apriori*, i.e. it also finds the FIs in a dataset. Here we propose an extension of *Eclat*, called *Eclog* that can filter FCIs and FGs among FIs. The proposed algorithm is a single-pass algorithm and it explores the frequent equivalence classes in a dataset. Experimental results show that *Eclog* performs very well, especially on dense, highly-correlated datasets.

## KÉZZEL ÍRT SZÁMJEGYEKET FELISMERŐ NEURONHÁLÓ ROBUSZTUSSÁGI VIZSGÁLATA

BISCHOF BARBARA HAJNALKA, KISS ATTILA ELEMÉR

A cikkben egy speciális adatbányászati algoritmust, nevezetesen a kézzel írt számjegyeket felismerő neurális hálót vizsgáljuk, miközben az adathalmazt egyre zajosabbá tesszük véletlen torzítások hozzáadásával. A tanító és a teszt adatok zajossá tételéhez többféle módszert is alkalmazunk. Részletesen elemezzük, hogyan hat a zaj az osztályozó algoritmusra. Összefüggést találunk a felismerés pontossága és aközött, hogy a teszt és a tanító adathalmazok milyen mértékben és milyen módon tartalmazzanak zajt.

### 1. Bevezetés

Az adatbányászat olyan technológia, amely képes arra, hogy elemezze a nyers adatokat információ szerzés céljából. Az elnevezés megtévesztő, hiszen nem adatot, hanem számunkra hasznos információt, új és eddig rejtett összefüggéseket keresünk egy nagy adathalmazban.

Manapság adatok millióit tároljuk különböző adatbázisokban, melyeknek egy igen jelentős részét soha nem használjuk. Emiatt jelentősen megnőtt az igény mind a piaci élet résztvevői, mind a kutatók felől, a hatalmas adatbázisokból való információ keresésére. Ennek két fő oka van: egyrészt a növekvő versenyhelyzet miatt az üzleti szféra szereplőinek szüksége van az adatbázisokban megbújó hasznos információkra, így ez a fokozódó igény növekvő kutatói beruházásokat indukált. Másrészt az adatbányászat a maga multidiszciplináris (több tudományágat érintő) voltával attraktív terület számos kutató számára.

A sikeres adatbányászat alapfeltételei közt említhetjük értelemszerűen a nagy mennyiségű adatot, hiszen minél nagyobb az adatmennyiség, annál biztosabban tudjuk kizárni bizonyos összefüggések esetiségét, azaz annál kisebb az esélye, hogy a talált összefüggés csupán a véletlen eredménye.

További alapfeltétel az adatok tisztasága. A zajok, illetve hibás bejegyzések jobb esetben csak nehezítik az adatbányászatot, rosszabb esetben azonban hamis eredményekhez vezetnek. Tekintsünk most el azoktól az esetektől amikor az adatokat szándékosan torzítjuk, például személyes adatok védelmének érdekében.

A cikkben egy egyszerűbb kézírásfelismerő program segítségével mutatjuk be, hogy az adatok különböző módon való torzítása esetén a neuronháló mennyire ismeri fel az adott karaktert. Összefüggéseket mutatunk a felismerés pontosságára az alapján, hogy a teszt és a tanító adathalmaz milyen mértékben és módon tartalmaz zajt.

## 2. Kapcsolódó munkák

A tíz legnépszerűbb adatelemzéssel, klaszterezéssel és statisztikával foglalkozó algoritmus leírását a [8] publikációban találhatjuk meg. További algoritmusok részletesebb leírásával és egymástól eltérő adatbázisra való tesztelésével, illetve ezen eredmények összehasonlításával és elemzésével [3] foglalkozik.

A kézírásfelismerő programokról általánosságban, illetve az ehhez kapcsolódó kérdésekről [6]-ban olvashatunk részletesebben. [5] egy olyan új algoritmust mutat be, mely kézzel írott számok offline felismerésére alkalmas egy egyszerű többrétegű neurális hálózat felhasználásával, a hálózat a hasonló számok hatékony osztályozására alkalmas. Az összetett mintázatfelismerési problémák megoldására a [2] cikkben három összetett neuronhálózati osztályozót mutatnak be. A beszéd-, illetve kézírásfelismerésben alkalmazott mély neurális hálókról [7]-ben olvashatunk, a cikk bemutat egy olyan módszert, melynek segítségével elérhetjük, hogy némi zaj hozzáadásával a program rosszul osztályozzon adatokat.

Részletesebb és átfogóbb magyar nyelvű szakirodalom [4], az adatbányászat alapvető fogalmaival és főbb területeivel foglalkozik.

## 3. Elméleti háttér

A mesterséges neuronhálózat egy biológiai ihletésű program, ami a biológiai neurális háló néhány tulajdonságát modellezi. Ezen modelleket természetesen nemcsak a biológiában, hanem számos más területen alkalmazzák főként tanító rendszerként. Leggyakoribb példája a képfelismerés, vagyis kézírásos vagy digitális szöveg szkennelésétől egészen az arcfelismerésig.

A tanulási technika szempontjából megkülönböztetünk ellenőrzött, illetve nemellenőrzött típusú tanulást. A kutatásunk során is alkalmazott ellenőrzött (felügyelt tanulás) esetében a rendszer számára nagy számú tanító mintapont párok (be- és kimeneti értékek) állnak rendelkezésre és a tanítás ezeken az ismert összerendelt mintapárokon alapul. Míg nemellenőrzött tanításnál címkézetlen tanító-pontjaink vannak, így a hálózatnak kívánt válaszok ismerete nélkül kell valamilyen viselkedést kialakítania, a környezetből azonban nincs semmiféle visszajelzés, ami a hálózat viselkedésének helyességére utalna.

Karakterfelismerő rendszereknél megkülönböztetünk online és offline felismerést, ez a tulajdonsága arra utal, hogy a feldolgozás azonnal, közvetlenül a betűk, vagy szó beírása után, vagy passzív módon, jelentősen később történik. Az online írásfelismerők jellemzően egy úgynevezett digitális tinta (digital ink) technikát alkalmaznak, ahol a beviteli eszköz mozgásának folyamata kerül feldolgozásra, vagyis rendelkezésre áll az írás képzésének módja is. Ezzel ellentétben a cikkben is tárgyalt offline technikát használó karakterfelismerők, az írás befejeztével kapott betűk képét használják fel, azon felül nem rendelkeznek további információval.

A karakterfelismerő algoritmusnak két alapvető eleme van, az úgynevezett tulajdonságkinyerő (feature extractor) és az osztályozó (classifier). A tulajdonság analízis meghatározza azon jellegzetességeket, amikkel a karakter rendelkezik, majd ezt küldi el az osztályozónak. Az egyik leggyakoribb osztályozó eljárás a mintafelismerés, ebben az esetben az egyes pixelek az adott karakterkép sajátosságai. Az osztályozás során az algoritmus képkockáinként összehasonlítja a bemeneti képet a különböző karakter-osztályok mintáival. Ennek eredményeként egy mérőszámot kapunk, amely megadja, hogy mennyire hasonló a bemenet és az adott minta, az eredménye az a karakter lesz, amihez a minta a legjobban hasonlított.

#### 4. Kísérletek

A vizsgálatainkhoz olyan neurális hálózatot építettünk ki, amely képes a kézírásos számok helyes azonosítására. Ehhez a széles körben elterjedt MNIST adatbázist [1] használtuk, mely kézírásos számjegyeket tartalmaz.

Az MNIST adatbázisban a tanító adathalmaz (train) 60 ezer mintát, míg a tesztkészlet 10 ezer képet foglal magába. A képek halmaza tulajdonképpen egy nagyobb adatbázis (NIST) része, amely közel 250 író példáját tartalmazza (a képet 8 biten ábrázoljuk, vagyis legfeljebb 256-féle árnyalatot látunk). Ezen képeket méretnormalizálták, továbbá a számjegyeket a rögzített kép közepére helyezték. Az általunk használt képadatok  $28 \times 28$  szürkeárnyalatos képpont (összesen 784 képpont) formájában vannak rögzítve, címkével együtt a kép helyes azonosításához.























A karakterfelismerő programhoz importálnunk kell a Kereas-t, mely egy a Python programozási nyelvhez elérhető könyvtár, ami Tensorflow-ra, Theano-ra vagy CNTK-ra épül és kifejezetten mély tanuláshoz és neurális hálózatok gyors definíciójához használható. Használata előnyös, mert így nem kell manuálisan kódolni a lineáris algebrát, valamint a szükséges aktivációs függvényeket (activation function) és optimalizálókat.

Egy neuronhálózat elrendezése véletlenszerű, mivel annyi rejtett réteggel rendelkezik, amennyire szükség van, és az egyes rétegeken is eltérő számú neuronok lehetnek. Az általunk felépített neuronhálózatba 784 képpont fog bemenni, ezen pixeleket egy 512 neuronból álló rejtett rétegnek adjuk át, amely ezt 10 neuronnak adja kimenetként (minden számjegyre egyet).

Kutatásunkhoz szükségünk volt különböző módon és mértékben zajosított képek előállítására, ehhez alapvetően 5 különböző módszert használtunk, melyek közös jellemzője, hogy a zaj mértéke paraméterezhető volt, így segítségükkel több adathalmazt is elő tudtunk állítani. A következőkben ezen zajgeneráló technikákat mutatjuk be részletesebben.

#### 4.1. Első módszer - véletlenszerű zaj

Első esetben minden kép esetén adott számú képpontot választottunk ki (ezt a számot az összes képpont számának és az adott adathalmaz zaj százalékának a szorzata adja), véletlenszerűen, egyenletes eloszlással, majd a képpontot értékétől függően, 0-ra vagy 255-re állítottuk át azt. Ha a képpont értéke nagyobb volt, mint 128, akkor 0-ra, ellenkező esetben pedig 255-re, így biztosítva, hogy a képpont mindig változzon (legfeljebb 255-tel és legalább 128-cal nő vagy csökken az értéke). A módosított képpontok száma minden kép esetén megegyezik (elhelyezkedésük azonban eltérő) egy adathalmazon belül. Összesen 50 adathalmazt állítottunk elő ezzel a módszerrel, a zaj mértéke pedig 1-től 50 százalékig terjed, ami minimum 8, maximum 392 zajos képpontnak felel meg.

zaj mértéke	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
	0 pixel	39 pixel	78 pixel	118 pixel	157 pixel	196 pixel	235 pixel	274 pixel	314 pixel	353 pixel	392 pixel
tiszta kép											
zajosított kép											

1. ábra. Zajosított kép - véletlenszerű zaj hozzáadásával

Az 1. ábra mutatja, hogy egy-egy kép mennyire tér el az eredetitől, ha adott százaléknyi zajt adunk hozzá. A karakter 25 és 30 százalékos zaj esetén is könnyedén felismerhető, de nagy valószínűséggel még 50 százaléknyi torzításnál is meg tudjuk mondani, hogy milyen szám szerepel a képen.

#### 4.2. Második módszer - sorok cseréje

A második módszernél szintén véletlenszerűen egyenletes eloszlással választottunk két sorindexet (1 és 28 között), majd a kiválasztott indexek alapján megcseréltük a kép két sorát. Minden képnél más sorindexet jelöltünk ki, de egy adatkészleten belül a megcserélt sorpárok száma mindig azonos. Továbbá fontos megjegyezni, hogy a cserék egymás után hajtottak végre (így kis valószínűséggel, de lehetséges, hogy valójában nem változott a kép). A 2. ábrán látható, hogy hogy néz ki az eredeti, illetve adott számú sorpár értékeinek felcserélése után a kép.

cserék száma	0 sorpár	1 sorpár	2 sorpár	3 sorpár	4 sorpár	5 sorpár	6 sorpár
tiszta kép	9	5	0	1	1	9	/
zajosított kép	9	5	0	1	1	9	/

cserék száma	7 sorpár	8 sorpár	9 sorpár	10 sorpár	11 sorpár	12 sorpár	13 sorpár
tiszta kép	6	3	8	5	9	0	5
zajosított kép	6	3	8	5	9	0	5

2. ábra. Zajosított kép - sorok felcserélésével

### 4.3. Harmadik módszer - oszlopok cseréje

Harmadik esetben a másodikhoz hasonló technikát választottunk, azzal a különbséggel, hogy nem a sorokat, hanem az oszlopokat cseréltük fel a képeken, ennek eredményét a 3. ábrán láthatjuk.

cserék száma	0 oszloppár	1 oszloppár	2 oszloppár	3 oszloppár	4 oszloppár	5 oszloppár	6 oszloppár
tiszta kép	3	3	/	2	4	7	4
zajosított kép	3	3	/	2	4	7	4

cserék száma	7 oszloppár	8 oszloppár	9 oszloppár	10 oszloppár	11 oszloppár	12 oszloppár	13 oszloppár
tiszta kép	2	0	6	7	1	8	7
zajosított kép	2	0	6	7	1	8	7

3. ábra. Zajosított kép - oszlopok felcserélésével

Amint azt a 2. és a 3. ábra jól mutatja, a számok aránylag jól felismerhetők vagy kikövetkeztethetők. Ebben jelentős szerepet játszik, hogy a képen belül a szám középre van igazítva, ezáltal a kép szélén elhelyezkedő sorok és oszlopok nem befolyásolják nagy mértékben a képek olvashatóságát, így ezeket egymással megcserélve a karakter továbbra is könnyedén felismerhető.



### 4.4. Negyedik módszer - fény változtatása

A negyedik módszer esetében egy adathalmazon belül minden képet egységesen világosítottuk vagy sötétítettük. Minden pixelhez hozzáadtunk, egy előre megadott értéket (-200, -150, -100, -50, 0, 50, 100, 150, 200) és ha az így kapott érték a [0, 255] intervallumon kívülre esett, akkor azt az intervallum megfelelő végpontjával helyettesítettük. Az így generált képeket a 4. ábra szemlélteti (minden kép bal felső pixelét fehérre (0), míg a jobb alsó pixelt feketére állítottuk (255), jobban szemléltetve a fény változását).

hozzáadott fény	-200	-150	-100	-50	0
tiszta kép					
zajosított kép					

hozzáadott fény	0	50	100	150	200
tiszta kép					
zajosított kép					

4. ábra. Zajosított kép - fény értékének változtatásával

Általánosságban elmondható, hogy a -100 és 100 között szinte biztosan és egyértelműen felismerhető a szám, míg -200, -150, 150 és 200-as érték hozzáadása esetén nehezebben tudjuk csak beazonosítani.

### 4.5. Ötödik módszer - színek számának változtatása

Az ötödik technika a kép intenzitásának változtatásán alapul. Az eredeti képet 8 biten ábrázoltuk, így 256 különböző árnyalatot tudtunk megkülönböztetni. A következőkben ezt módosítjuk oly módon, hogy minden pixelt annak értékétől függően hozzárendelünk egy csoporthoz és az egy csoporthoz tartozó pixeleket ugyanarra a színre állítjuk be (az intervallum középső elemének színére), ezáltal 7 új tanuló és teszt adatkészletet hozunk létre, ahol az adathalmazokban a színek száma: 128, 64, 32, 16, 8, 4 és 2.

színek száma	256 szín	128 szín	64 szín	32 szín	16 szín	8 szín	4 szín	2 szín
tiszta kép								
zajosított kép								

5. ábra. Zajosított kép - intenzitás változtatással

Az 5. ábra szemlélteti az eredeti, illetve a torzított képet. Ahogy látjuk, 16 szín esetén nem feltétlen tudjuk az eredeti és a zajosított képet megkülönböztetni egymástól, ezenfelül elmondhatjuk, hogy még 2 szín esetén sem romlott jelentősen az olvashatóság, és a karakter szépen kivehető.

### 5. Eredmények

Az előzőleg bemutatott zajgeneráló technikák segítségével különböző tanuló és teszt adathalmazokat készítettünk, majd egy-egy tanuló adathalmaz segítségével létrehozott neurális hálót minden (az adott módszer segítségével előállított) teszt adathalmazzal teszteltünk.

#### 5.1. Első módszer - véletlenszerű zaj

		teszt adatokban a zaj											
		db	0	39	78	118	157	196	235	274	314	353	392
		%	0	5	10	15	20	25	30	35	40	45	50
train adatokban a zaj	0	0	98,2	88,8	65,6	48,0	37,7	30,0	25,5	21,6	19,3	17,1	16,1
	8	1	98,0	89,6	67,1	51,0	39,6	32,4	27,0	23,0	19,8	17,5	16,2
	16	2	97,6	93,7	71,8	47,5	31,5	22,7	17,5	14,9	12,8	11,6	11,0
	24	3	98,0	95,6	83,6	61,7	43,1	31,8	24,9	21,7	18,5	17,2	15,9
	31	4	97,6	96,4	89,1	73,1	53,9	39,8	29,5	23,8	20,3	18,1	16,7
	39	5	97,2	96,5	91,5	79,6	62,0	48,0	36,4	29,8	24,2	20,2	18,4
	78	10	96,5	96,5	94,5	91,1	84,3	74,1	61,7	50,7	41,2	33,3	28,0
	118	15	96,8	96,5	95,1	93,3	90,0	84,9	78,1	70,2	60,9	52,6	46,3
	157	20	95,5	95,6	94,7	93,2	91,3	87,6	83,1	76,9	69,9	61,6	55,1
	196	25	95,7	95,2	94,7	93,1	91,3	88,5	85,1	80,0	73,8	66,6	60,8
	235	30	94,4	93,9	93,0	92,1	90,2	87,7	84,2	80,1	74,5	68,3	63,1
	274	35	94,5	93,6	92,8	91,9	90,2	87,6	84,6	80,9	77,0	70,7	66,1
314	40	93,7	93,0	92,4	91,3	89,6	87,7	85,3	80,8	77,0	71,2	66,3	
353	45	91,3	90,7	90,3	89,1	87,9	85,8	83,5	80,2	76,5	70,6	66,4	
392	50	91,9	91,2	90,1	89,2	87,3	86,0	83,1	79,1	75,5	70,4	65,5	

1. táblázat. Felismerés pontossága a véletlenszerűen hozzáadott zaj hatására

Az 1. táblázatban találjuk az első zajgeneráló módszerrel készült képek esetén a felismerés pontosságának eredményeit (jobb olvashatóság érdekében a táblázat nem tartalmaz minden eredményt). A sorok megadják, hogy a tanító adathalmazban egy kép esetén hány százalék a zaj (illetve hogy ez hány darab pixelt jelent), míg az oszlopok a teszt adatokra vonatkoznak. Egy adott sor egy adott neuronhálót jelent, melyet különböző mértékben zajosított adatokkal teszteltünk.

A legjobb eredményt (98,2 százalék) értelemszerűen abban az esetben értük el, amikor a tanító adathalmaz és a teszt adathalmaz sem tartalmazott zajt. Míg a legrosszabb értékeket (11,0 százalék) akkor kaptuk, amikor a modell által betanult képek minimális (2-3 százalék) zajt tartalmaztak és a teszt adatoknál pedig minden kép esetében a pixelek felének eltért a színe az eredetitől.

Azt mondhatjuk, hogyha a tanító adathalmaz képeinek zajossága  $X$ , a teszt adathalmaz képeinek zajossága  $Y$ , akkor igaz az alábbi összefüggés: ha  $X \leq 25$  és  $X - Y \geq 0$ , vagy ha  $X > 25$  és  $X + Y \leq 50$ , ebben az esetben igaz az, hogy a neuronháló legalább 90 százalékos valószínűséggel felismeri az adott karaktert.

## 5.2. Második és harmadik módszer - sorok és oszlopok cseréje

Az oszlopok, illetve sorok cseréjével előállított képek esetében, hasonló eredményeket kaptunk, ezeket a 2. és a 3. táblázatban látjuk. A legjobb érték a zaj nélküli adatokban keletkezett, ahogy azt az előző esetben is láttuk, míg a legrosszabb értéket akkor kaptuk, ha a teszt adathalmaz minden képében felcseréltünk 13 sort/oszlopot és a tanuló adatokat pedig nem zajosítottuk. Az így kapott legrosszabb értékek (40,6 százalék az oszlop és 32,6 százalék a sor cserék esetén) jelentősen jobbák, mint az első (százalékos zaj) módszerrel zajosított képeknél (11,1 százalék).

		teszt adatokban a cserék száma													
		0	1	2	3	4	5	6	7	8	9	10	11	12	13
train adatokban a cserék száma	0	98,3	89,7	83,8	77,3	70,2	63,6	58,5	52,3	47,9	43,8	40,7	37,2	34,5	32,6
	1	97,9	96,4	95,0	93,0	91,3	87,6	84,9	80,8	77,0	72,9	70,2	65,8	63,1	60,5
	2	97,6	96,5	95,5	94,0	92,2	89,6	87,0	83,6	79,7	76,2	73,2	69,0	66,3	63,5
	3	97,7	96,8	95,7	94,7	93,2	91,1	88,8	85,7	82,6	78,6	76,2	71,9	69,6	66,5
	4	97,3	96,4	95,6	94,8	93,7	91,9	90,0	87,6	84,6	81,8	78,9	75,2	72,4	69,8
	5	96,9	96,4	95,6	94,7	93,8	92,4	90,9	88,5	86,5	83,7	81,1	77,7	75,6	73,0
	6	96,8	95,9	95,2	94,6	93,6	92,4	90,6	88,6	86,0	83,9	81,6	78,4	75,1	73,4
	7	96,7	95,8	95,5	94,3	93,9	92,6	91,5	89,7	87,7	85,6	83,8	80,5	78,5	76,0
	8	96,3	95,5	94,8	94,4	93,9	92,4	91,4	89,9	87,5	85,8	84,4	81,5	80,2	77,3
	9	96,0	95,5	94,8	94,2	93,5	92,2	91,7	90,1	88,0	86,4	85,4	82,8	81,1	79,1
	10	95,7	95,0	94,3	93,8	93,4	92,3	91,0	89,7	88,0	86,9	85,6	83,2	81,5	80,0
	11	95,3	94,5	93,9	93,4	92,8	92,2	90,8	89,8	88,2	86,9	85,9	83,9	82,0	80,1
	12	95,3	94,6	94,1	93,2	92,5	91,7	91,1	89,9	88,6	87,5	86,4	83,6	82,6	81,1
13	94,9	93,9	93,3	92,8	92,1	91,4	90,2	89,4	87,3	86,8	85,2	83,9	82,0	81,0	

2. táblázat. Felismerés pontossága adott számú sorpár felcserélésének hatására

Sorok esetén az alábbi összefüggés adja meg, hogy a felismerés pontossága hol nagyobb mint 90 százalék: ha a tanító adathalmaz képeiben a felcserélt sorok száma  $X$ , a teszt adathalmaz képeiben felcserélt sorok száma  $Y$ , akkor ha  $X \leq 7$  és  $X - Y \geq 0$ , vagy ha  $X > 7$  és  $X + Y \leq 13$ .

Oszlopok cseréje esetén az előzőhöz hasonlóan  $X$  a tanító adathalmazra, míg  $Y$  a tesztkészlet képeire vonatkozik, ekkor ha igaz, hogy  $X - Y \geq 0$ , akkor a neuronháló felismeri a karaktert 90 százalékos valószínűséggel.

		teszt adatokban a cserék száma													
		0	1	2	3	4	5	6	7	8	9	10	11	12	13
train adatokban a cserék száma	0	98,3	91,2	86,4	80,6	75,2	69,8	64,3	59,5	55,8	51,8	48,7	45,9	43,7	40,6
	1	98,1	96,9	96,1	94,9	93,5	91,7	90,0	87,3	85,9	83,9	81,8	79,3	77,5	74,9
	2	97,8	96,9	96,3	95,3	94,2	93,0	92,0	89,5	88,2	86,5	84,6	82,8	81,5	78,7
	3	97,6	96,8	96,3	95,6	94,7	94,0	92,7	91,2	89,8	88,7	87,2	85,2	84,0	81,9
	4	97,6	96,7	96,5	95,5	94,9	93,9	93,0	91,5	90,5	89,4	88,0	86,4	85,2	83,7
	5	97,5	96,7	96,2	95,8	94,9	94,2	93,3	92,4	91,5	90,3	89,2	87,9	86,5	85,2
	6	97,4	96,8	96,4	95,9	95,0	94,7	93,8	93,0	92,1	91,6	90,0	88,4	87,8	86,2
	7	97,0	96,4	96,2	95,8	95,1	94,5	94,0	93,3	92,6	91,6	90,7	89,6	88,7	87,6
	8	96,8	96,2	95,9	95,5	95,2	94,5	94,2	93,3	92,6	91,5	91,0	89,6	89,3	88,0
	9	96,4	95,9	95,3	95,0	94,6	94,3	93,7	92,5	91,9	91,6	91,1	89,7	88,7	88,0
	10	96,6	96,2	95,6	95,6	95,2	94,9	94,0	93,4	92,8	92,2	91,9	90,4	90,0	88,7
	11	96,1	95,9	95,5	95,1	94,8	94,3	93,7	93,1	92,8	91,9	91,7	90,5	90,0	89,1
	12	96,3	95,9	95,2	95,2	94,4	94,7	94,0	93,3	92,9	92,5	91,4	90,7	90,4	89,5
	13	96,2	95,4	95,3	95,2	94,6	94,3	93,8	93,2	92,8	92,4	91,8	90,8	90,4	89,5

3. táblázat. Felismerés pontossága adott számú oszloppár felcserélésének hatására

Amint az a 2. és a 3. táblázatból is kiolvasható, adott számú oszlop felcserélése esetén számottevően jobb a karakter felismerésének pontossága (átlagosan 85,3 százalék), mint ugyanannyi sorpár felcserélése esetén (átlagosan 90,3 százalék). Ennek oka valószínűleg a karakterek elhelyezkedéséből adódik, hiszen a képek nagy részénél a tényleges betű egy kisebb téglalapban helyezkedik el a kép közepén. Így a tőle jobbra, illetve balra levő „szinte” fehér oszlopok felcserélése nem ront jelentősen az olvashatóságon.

### 5.3. Negyedik módszer - fény változtatása

A 4. táblázatban azt láthatjuk, milyen eredményeket kaptunk a felismerés pontosságára abban az esetben, amikor a fény erejét állítottuk az egyes képeken. Az előző kísérletekhez hasonlóan a legjobb eredményt akkor értük el, ha a tanuló adathalmaz és a teszt adathalmaz sem tartalmazott zajt, míg a legrosszabb értéket (10,28 százalék) abban az esetben kaptuk, ha a tanuló adathalmaz nem tartalmazott zajt, és a teszt adathalmaz képei pedig szemmel láthatóan sötétebbek (pixelek értékét 200-zal növeltük).

hozzáadott fény értéke a teszt adatokban

		-200	-150	-100	-50	0	50	100	150	200
hozzáadott fény értéke a train adatokban	-200	96,7	97,3	97,5	97,4	97,1	55,4	32,4	17,4	11,6
	-150	96,5	97,6	97,9	97,9	97,7	51,9	31,0	18,8	11,2
	-100	94,8	97,4	97,9	98,0	98,0	47,3	21,3	11,4	10,3
	-50	92,8	97,1	97,8	98,0	98,1	47,0	20,5	10,5	10,3
	0	91,6	96,9	97,6	98,0	98,1	46,3	16,9	10,5	10,3
	50	95,0	96,2	96,6	96,9	97,1	97,5	85,2	39,8	13,6
	100	93,9	95,0	95,5	95,7	95,8	96,8	96,7	89,4	36,7
	150	90,1	90,6	90,8	90,7	90,3	92,0	93,0	92,1	64,7
	200	80,6	80,1	79,6	79,1	78,2	79,7	81,3	83,8	82,5

4. táblázat. Felismerés pontossága a fény változtatásának hatására

Ha a tanító adathalmaz képeinek zajosságát  $X$ -szel és a teszt adathalmaz képeinek zajosságát  $Y$ -nal jelöljük, akkor igaz az alábbi összefüggés: ha  $-200 \leq X \leq 150$  és  $-200 \leq Y \leq 0$ , vagy ha  $50 \leq X, Y \leq 150$  és  $Y - X \leq 0$ , ilyenkor a karakter felismerésének pontosságának valószínűsége legalább 90 százalék.

#### 5.4. Ötödik módszer - színek számának változtatása

teszt adatokban a színek száma

		256	128	64	32	16	8	4	2
train adatokban a színek száma	256	98,2	98,1	98,1	97,9	97,5	93,1	59,2	21,2
	128	98,2	98,2	98,2	97,9	96,2	79,1	23,6	8,9
	64	98,2	98,2	98,1	98,1	97,2	88,5	47,1	9,8
	32	98,1	98,1	98,2	98,2	97,9	93,8	51,7	9,3
	16	97,7	97,8	97,9	97,9	97,9	96,9	78,3	18,2
	8	97,3	97,3	97,4	97,5	97,6	97,8	95,2	30,1
	4	97,3	97,4	97,4	97,5	97,6	97,8	97,7	82,7
	2	96,4	96,4	96,4	96,5	96,6	96,7	97,1	96,2

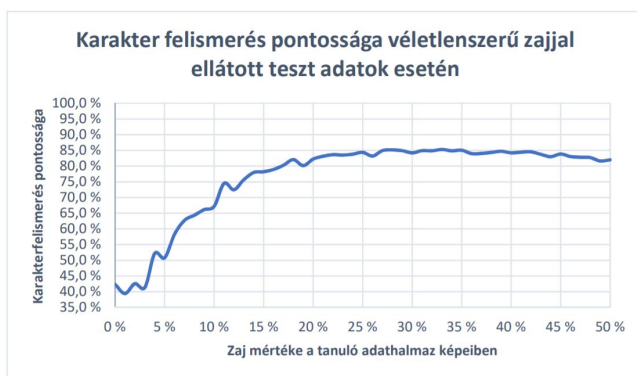
5. táblázat. Felismerés pontossága az intenzitás változtatásának hatására

A színek számának változtatására kapott eredményeket az 5. táblázat tartalmazza. Amint azt az 5. táblázaton is láthattuk, hogy nem romlott nagy mértékben az olvashatóság, így ennek megfelelően a karakter felismerésének pontosságára kapott eredmények is magasak (átlagukat tekintve 85,5 százalék - összehasonlításképp az első zajgeneráló módszerrel kapott eredmények átlaga 67,7 százalék).

Mivel a csupán 16 színt tartalmazó képet szabad szemmel szinte meg sem tudjuk különböztetni a 256 színt tartalmazó képtől, emiatt az 5. táblázatban látható legmagasabb értéket (98,2 százalék) több esetben is elértük. A felismerés pontosságára kapott legrosszabb eredményt (8,9 százalék) abban az esetben kaptuk, ha a tanuló adathalmaz képei 64 színből állt, míg a teszt készlet képei csupán 2-ből.

## 6. Következtetések

Az előző fejezetben tárgyalt kísérletek alapján azt mondhatjuk, ha egy adott mértékig zajos adatokat szeretnénk felismerni, akkor a legjobb módszer, ha a tanító adathalmazt is hasonló módon és mértékben zajosítjuk. Hiszen így tudjuk elérni a felismerés pontosságára a legjobb értéket. Azonban fontos megjegyezni, hogy ezekben az esetekben mind a teszt és mind a tanító adathalmaz elemei azonos mennyiségű zajt tartalmaztak, ami egy nagyon speciális, a valóságtól igencsak eltérő eset.

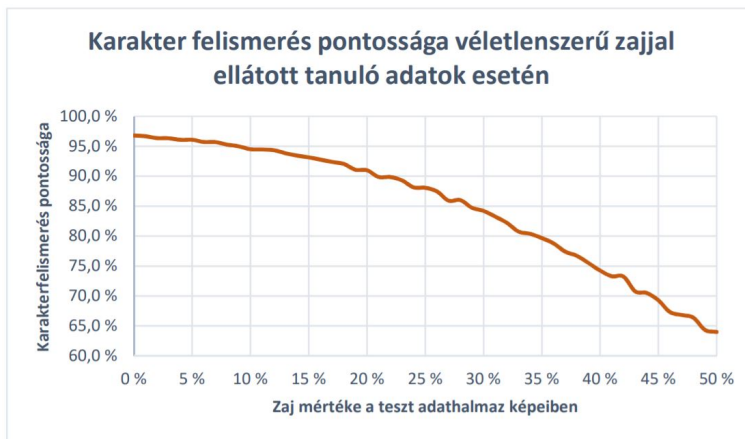


6. ábra. Felismerés pontossága véletlenszerűen zajosított tesztkészlet esetén

Mindezek miatt nézzünk egy olyan esetet, ahol több tanító adathalmazt hozunk létre, oly módon, hogy egy-egy halmazon belül azonos százaléknyi zajt tartalmazó képek szerepelnek (ezen tanító adathalmazok megegyezhetnek az 5. fejezetben tárgyalt tanító adathalmazokkal). Az így létrejött karakterfelismerő neuronhálók mindegyikét ugyanarra az egy teszt adatkészletre teszteljük, melyre az teljesül, hogy minden egyes kép esetén meghatároztunk egy véletlen számot (1 és 50 között) és ezzel a véletlenszámmal generáltunk zajt (az első - százalékos - zajgeneráló módszer segítségével) külön-külön mindegyik kép esetén, ahol a véletlenszám adta meg a zaj százalékos értékét. Ezáltal a tesztadathalmaz a valóságnak megfelelően eltérő mértékben tartalmaz zajt az egyes képeken.

Az így elért eredményeket a 6. ábrán láthatjuk, ahol a vízszintes tengely megadja, hogy a neuronháló milyen mértékű zajt tartalmazó képeken tanult, míg a függőleges tengely a karakter felismerésének valószínűségét adja meg. Amint azt láthatjuk, a legrosszabb esetben nagyjából 40 százalék valószínűséggel ismeri fel a képet a program, azonban egyetlen esetben sem éri el a 90 vagy annál nagyobb százalékot, átlagosan azt mondhatjuk, hogy a felismerés pontossága 75-77 százalék között mozog. Ezen értékek megegyeznek az 1. táblázatban a 25%-hoz tartozó oszloppal, vagyis ha a tesztkészlet minden képét azonosan rontottuk el 25 százalékkal.

Ezenfelül tekintsünk egy további lehetőséget, amikor egy tanító adathalmazunk van, melynek minden képe más mennyiségű zajt tartalmaz és az ezen képek által felépített neuronhálót teszteljük különböző tesztkészletekkel, ebben az esetben a tesztkészleten belül a képek egyformán zajosak.



7. ábra. Felismerés pontossága véletlenszerűen zajosított tanító adathalmaz esetén

A 7. ábra diagramja mutatja, hogy ha egy neuronhálót építünk és azt különböző tesztkészletekre teszteljük (vízszintes tengely), akkor milyen pontossággal ismeri fel a program a karaktert. Az itt kapott értékek (legrosszabb esetben: 64,3 legjobb esetben: 96,8 átlagosan: 85,3 százalék) hozzávetőlegesen megegyeznek az az esettel, amikor a tanító adathalmaz minden képét azonosan zajosítottuk 25 százalékban, ez az 1. táblázatban a 25%-hoz tartozó sor.

Az azonos mértékben zajosított tesztkészlet, és a véletlenszerűen zajt tartalmazó között talált összefüggés miatt alátámasztást nyert az a megállapítás, hogy egy karakterfelismerő neuronháló felismerésének pontosságát javíthatjuk azáltal, hogy a tanító adathalmaz képeit zajosítjuk.

## 7. Kitekintés

A cikkben végzett kísérletek során a zaj mértékét és módját változtattuk, ehhez alapvetően 5 különböző módszert használtunk, ezenfelül hasznos lenne további technikák tesztelése, melyek életszerűbbek és jobban tükrözik a valóságot. Ilyen lehet például az elmosódott, homályos képek vagy a rossz fény beállítással készített fényképek (egyenletlen megvilágítás, erőteljes vaku) vagy elnyújtott/összezsugorított képek (nem megfelelő szögben tartott fényképezőgép esetén).

A továbbiakban még érdemes vizsgálnunk, hogy mivel tudjuk még jobban javítani a felismerés pontosságát, ez lehet esetleg a neuronháló alap beállításainak módosítása, vagyis a neuronok és idegrendszerek számának változtatása, vagy a tanuló adathalmaz méretének módosítása, továbbá a Keras által nyújtott különböző optimalizálók használata.

Ezek mellett azt is célszerű vizsgálni, hogy hogyan tudjuk eldönteni egy megadott képről, hogy milyen mértékben tartalmaz zajt, hiszen ennek az értéknek az ismeretében könnyen tudjuk úgy kalibrálni a neuronhálónkat, hogy minél nagyobb valószínűséggel ismerje fel az adott karaktert.

## Köszönetnyilvánítás

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg (EFOP-3.6.3-VEKOP-16-2017-00002).

Továbbá köszönjük szépen az anonim bírálóknak a hasznos és értékes észrevételeiket, javaslatokat és megjegyzéseiket.

## Hivatkozások

- [1] THE MNIST DATABASE OF HANDWRITTEN DIGITS: <http://yann.lecun.com/exdb/mnist/>.
- [2] CHO, SUNG-BAE: *Neural-network classifiers for recognizing totally unconstrained handwritten numerals*, IEEE Transactions on Neural Networks, Vol. **8** No. **1**, pp. 43-53 (1997). DOI: [10.1109/72.55419](https://doi.org/10.1109/72.55419)
- [3] DOGAN, NESLIHAN, AND ZUHAL TANRIKULU: *A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness*, Information Technology and Management, Vol. **14** No. **2**, pp. 105-124 (2013). DOI: [10.1007/s10799-012-0135-8](https://doi.org/10.1007/s10799-012-0135-8)
- [4] FERENC, BODON: *Adatbányászati algoritmusok (2002)*.  
<http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/adatbanyaszat.pdf>
- [5] LEE, SEONG-WHAN: *Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **18** No. **6**, pp. 648-652 (1996). DOI: [10.1109/34.506416](https://doi.org/10.1109/34.506416)



- [6] SUEN, CHING Y., ET AL.: *Handwriting recognition-the last frontiers*, Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. IEEE, Vol. 4 (2000). DOI: [10.1109/ICPR.2000.902853](https://doi.org/10.1109/ICPR.2000.902853)
- [7] SZEGEDY, CHRISTIAN, ET AL.: *Intriguing properties of neural networks (2013)*, arXiv preprint [1312.6199](https://arxiv.org/abs/1312.6199) (2014).
- [8] WU, XINDONG, ET AL.: *Top 10 algorithms in data mining*, Knowledge and information systems, Vol. 14 No. 1, pp. 1-37 (2008). DOI: [10.1007/s10115-007-0114-2](https://doi.org/10.1007/s10115-007-0114-2)



Bischof Barbara Hajnalka 1998-ban született Körmenden. 2016-ban tett érettségét a körmendi Kölcsey Ferenc Gimnáziumban, majd az Eötvös Loránd Tudományegyetem Informatikai Karán folytatta tanulmányait Programtervező Informatikus szakon, ahol 2019-ben alapszakos diplomát szerzett. A mesterképzést 2019-ben kezdte Információs Rendszerek szakirányon szintén az ELTE-n. Emellett 2018-tól az SAP Hungary Kft.-nél dolgozik junior fejlesztő munkatársként.

Bischof Barbara Hajnalka

ELTE Eötvös Loránd Tudományegyetem Informatikai Kar  
1117 Budapest, Pázmány Péter sétány 1/C  
[bisbarbi@caesar.elte.hu](mailto:bisbarbi@caesar.elte.hu)



Kiss Attila Elemér 1985-ben matematikusként végzett az Eötvös Loránd Tudományegyetemen. 1991-ben lett a matematikai tudomány kandidátusa. 2010-ben habilitált az informatikai tudományokból. 2010 óta az Eötvös Loránd Tudományegyetem Információs Rendszerek Tanszékének vezetője. Több mint 140 publikációja jelent meg, elsősorban adatbázisok, adatbányászat, mesterséges intelligencia, bioinformatika témakörökben. Doktori hallgatói közül eddig heten szereztek meg a doktori fokozatot. A kutatás mellett számos sikeres kutatás-fejlesztési, illetve ipari projektet vezetett.

Kiss Attila Elemér

ELTE Eötvös Loránd Tudományegyetem Informatikai Kar  
1117 Budapest, Pázmány Péter sétány 1/C  
[kiss@inf.elte.hu](mailto:kiss@inf.elte.hu)

ROBUSTNESS TESTING OF NEURAL NETWORK FOR HANDWRITTEN DIGIT  
RECOGNITION

BARBARA HAJNALKA BISCHOF, ATTILA ELEMÉR KISS

The paper examines a special data mining algorithm, namely the neural network that recognizes digits, while making the data set increasingly noisy by adding random distortions. We analyze in detail how noise affects the classification algorithm. Using a simpler handwriting recognition program, we show how the neural network recognizes a given character when it distorts data in different ways (we used five different methods to noise the data). We find correlations for recognition accuracy based on the extent and way in which the test and train data sets contain noise.

*Keywords:* data mining, neural network, robustness, handwriting recognition.

*Mathematics Subject Classification (2000):* 68T05, 68T35.

Az Alkalmazott Matematikai Lapok megjelenését támogatja  
a Magyar Tudományos Akadémia Könyv- és Folyóiratkiadó Bizottsága.

A kiadásért felelős a BJMT főtítkára

Szedte és tördelte: Moczár Károly

Nyomta a Coradix Kft., Budapest

Felelős vezető: Szűcs Ernőné

Budapest, 2020

Megjelent 18 (A/5) ív terjedelemben

100 példányban

HU ISSN 0133-3399

## ÚTMUTATÁS A SZERZŐKNEK

Az Alkalmazott Matematikai Lapok csak magyar nyelvű dolgozatokat közöl. A közlésre szánt dolgozatokat és mellékleteiket egy tömörített zip fájlban e-mailen az `aml@renyi.hu` címre kérjük elküldeni.

A részletes instrukciók és a LaTeX minta fájl megtalálhatók az AML honlapján, a cikkek benyújtása [http://aml.math.bme.hu/?page\\_id=7](http://aml.math.bme.hu/?page_id=7) címén. Mivel az Alkalmazott Matematikai Lapokban megjelenő cikkek 2019 óta DOI azonosítót kapnak, az irodalomjegyzékben szereplő minden forrás adatai között szerepeltetni kell a DOI azonosítót is, amennyiben rendelkezik ilyen-  
nel.

Kérjük a szerzőket, hogy a cikk benyújtása előtt ellenőrizték az alábbi elemek meglétét:

### **A teljes cikk, benne:**

magyar nyelvű kivonat

a cikk szövege, valamint ábrák és táblázatok, ha vannak

irodalomjegyzék, DOI azonosítókkal együtt

szerzők bemutatása és fotói, munkahelyei és e-mail címei

angol nyelvű cím és kivonat

### **Mellékelte fájlok:**

a cikk LaTeX forráskódja, és a teljes cikk pdf formátumban

ábrák, képek fájljai

szerzők fotói

A dolgozatok után szerzői díjat az Alkalmazott Matematikai Lapok nem fizet.

## TARTALOMJEGYZÉK

Előszó .....	1
Stoyan Gisbert emlékére (1942-2018) .....	3
A Farkas Gyula emlékdíj 2019. évi díjazottja: Györgyi Péter .....	11
<i>Bergmann Júlia, Molontay Roland, Szabó Mihály, Szekrényes Dóra Laura</i> , Kreditrendszerű képzések mintatanterveinek és előtanulmányi hálóinak elemzése a hazai matematika alapszakok példáján .....	13
<i>Bihary Zsolt, Nagy Noémi, Simon L. Péter</i> , Hálózati modell együttes csődvalószínűségek meghatározására .....	51
<i>Szathmáry László</i> , Egy gyakori zárt mintákat és gyakori generátorokat kereső vertikális al- goritmus .....	71
<i>Bischof Barbara Hajnalka, Kiss Attila Elemér</i> , Kézzel írt számjegyeket felismerő neuronháló robosztussági vizsgálata .....	87

## INDEX

Preface .....	1
In memoriam Gisbert Stoyan (1942-2018) .....	3
Gyula Farkas Prize recipient in 2019: Péter Györgyi .....	11
<i>Júlia Bergmann, Roland Molontay, Mihály Szabó, Laura Dóra Szekrényes</i> , Analyzing the curricula and prerequisite networks of hungarian mathematics bsc programs .....	13
<i>Zsolt Bihary, Noémi Nagy, Péter L. Simon</i> , Network model for joined default probabilities .....	51
<i>László Szathmáry</i> , A vertical algorithm for finding frequent closed itemsets and frequent generators .....	71
<i>Hajnalka Barbara Bischof, Elemér Attila Kiss</i> , Robustness testing of neural network for handwritten digit recognitions .....	87