Haladvány Kiadvány 17-06-15

Mely merev körű gráfok és hogyan használhatók valószínűségi becslésekhez?

Hujter Mihály hujter.misi@gmail.com

Ajánlás.
Takács Lajos (1924–2015) és Prékopa András (1929–2016) emlékére.

Köszönet.
BME DET-nek konferencia-részvétel támogatásáért,
Kovács Edith inspiráló konzultálásaiért.

Kivonat.

A merev körű gráfok vizsgálatát *Hajós György, Gallai Tibor, Surányi János, Hajnal András* kezdték el. A valószínűségi becslések tanulmányozása *Prékopa András* és tanítványai munkáiban élénkült fel. *Boros* és *Veneziani*, illetve tőlük függetlenül *Dohmen* rátaláltak arra a fontos kapcsolatra, mely a merev körű gráfok és az eseményúniók valószínűségének felső becslése között van.

Előadásunkban röviden áttekintjük a fent idezett eredmények történetét. Rámutatunk majd, hogy a leghasznosabb merev körű gráfok tulajdonságainak feltárása mennyire indokolt.

A merev körű gráfok színezéseinek kérdéskörét is érinteni fogjuk.

Legyenek $A_1, \cdots, A_n$ valószínűségi események, melyek általában nem függetlenek. A híres-nevezetes Bonferroni-becslés, vagy más néven *Boole–Bonferroni*-egyenlőtlenség a következő:

$$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n) \tag{BB}$$

Történeti megjegyzések. Bonferroni azt javasolta, hogy ha α megbízhatósági szintű döntést szeretnénk hozni, de egymás után n darab tesztet kell végeznünk, melyek egymástól nem függetlenek, akkor α szint helyett α/n szinten döntsünk.

Legfőbb célunk. A (BB) egyenlőtlenség esetében a jobb oldal minusz a bal oldal alsó (algoritmikus) becslése (azaz a (BB) egyenlőtlenség élesítése) felhasználva $1 \leq i < j \leq n$ és $1 \leq r < s < t \leq n$ esetére a $P(A_i \cap A_j)$ és $P(A_r \cap A_s \cap A_t)$ számokat.

Jelölés.

$$\mathsf{GAP} = P(A_1) + \cdots + P(A_n) - P(A_1 \cup \cdots \cup A_n)$$

Tétel (*Hunter* 1976). Jelölje az $\{1, \cdots, n\}$ szögponthalmazon egy tetszőleges fa gráf éleit T. Ekkor

$$\mathsf{GAP} \geq \sum_{ij \in T} P(A_i \cap A_j) \tag{HW}$$

Történeti megjegyzések. (Huntertől függetlenül) *Worsley* is megtalálta a (HW) összefüggést. Az 1982-ben publikált eredménye Hunterénél nagyobb ismertséget nyert, ezért sok kutató a (HW) összefüggést Hunter–Worsley-becslésnek hívja. A (HW) becslés jobb oldala *Kruskal* híres algoritmusával $O(n^2 \log n)$ lépésben maximalizálható.

Mivel az $n = 2$ esetben a (HW) egyenlőtlenség egyenlőséggel áll fenn, a továbbiakban feltesszük, hogy $n \geq 3$.

A következő neghatározások az $n \geq 3$ pontú egyszerű gráfokra értelmezve ugyanazt a gráfcsaládot definiálják:

— 2-fák.

— talpas cseresznyefák.

— 3-színezhető merev körű gráfok $n-2$ darab háromszöggel, $2n-3$ darab éllel.

Mivel a gráfosztályba tartozik a *háromszög* és a *gyémánt* (azaz a 3-pontú teljes és a 4-pontú egy él híján teljes gráf), ezért mi a gráfosztályt *delta-diamond* gráfosztálynak nevezzük.

Egy gráfban *szimpliciális* pontnak nevezünk egy u pontot, ha u és szomszédai teljes részgráfot indukálnak.

Tulajdonságok. Az n pontú delta-diamond gráfokoat a következők jellemzik:

— 2-összefüggők, de $n \geq$ **4** esetén nem **3**-összefüggők.

— Síkba rajzolhatók.

— 3 színnel színezhetők

— Minden indukált részgráfjuk perfekt gráf, nevezetesen legfeljebb 3 színnel színezhető merev körű gráf.

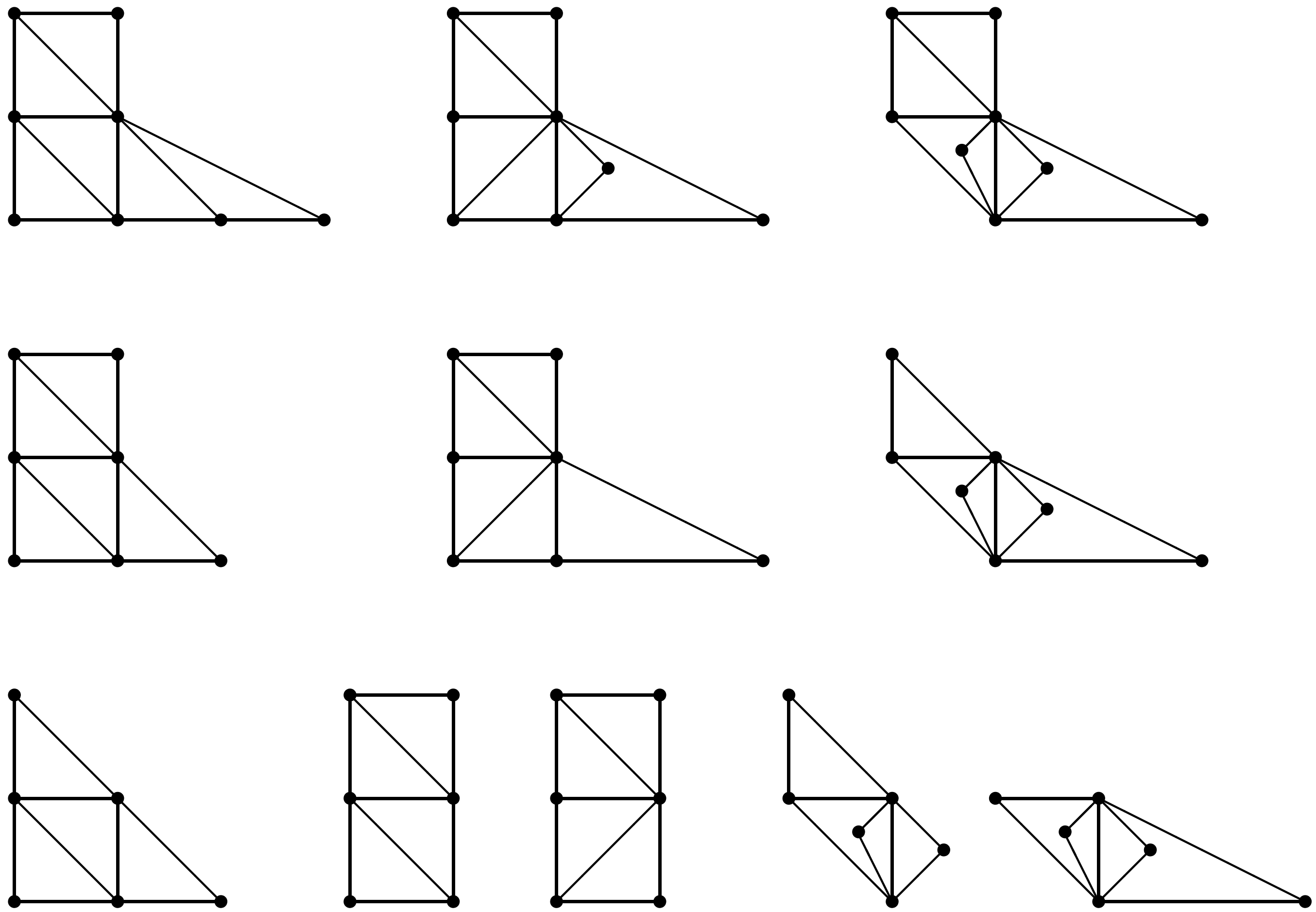
— Minden n pontú, legfeljebb 3 színnel színezhető merev körű gráf részgráfja valemely n pontú delta-diamond gráfnak. (Sőt a kiindulási delta-diamond gráfból egyesével hagyhatók el az élek.)

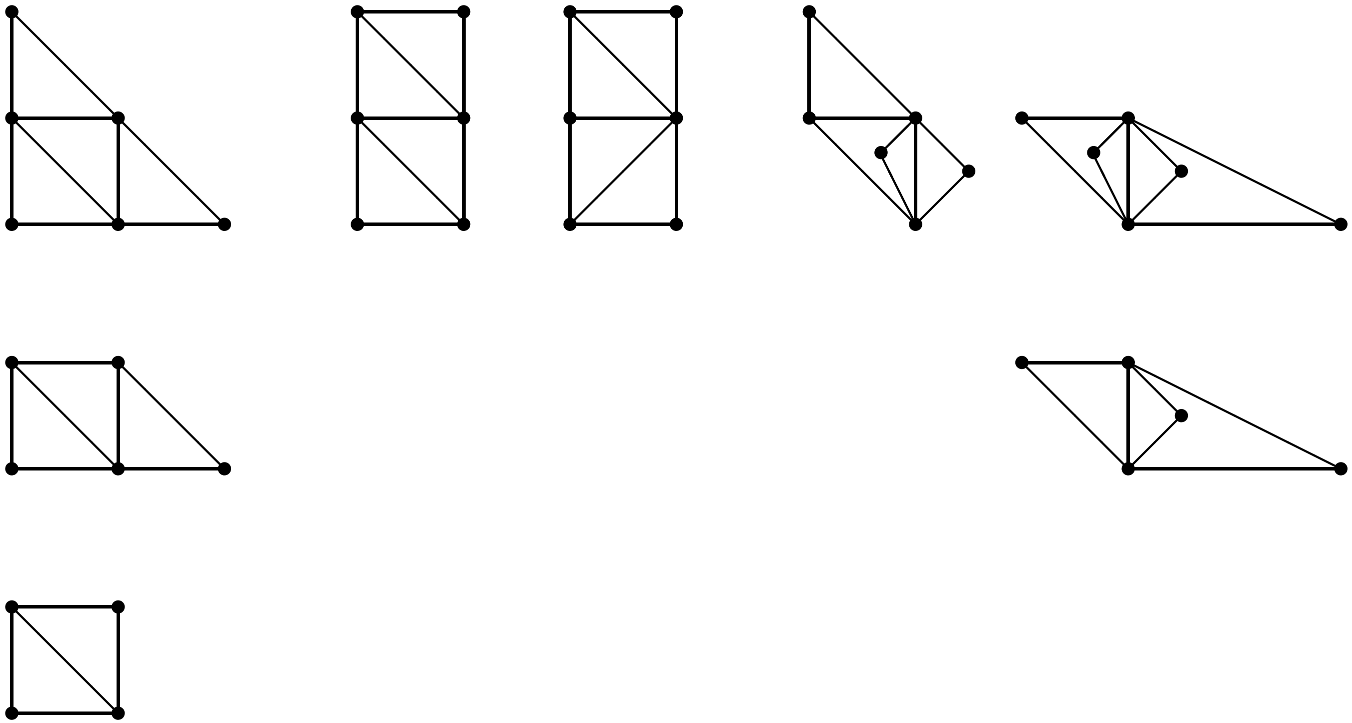
— $n = 3$ esetén mindhárom pont szimpliciális, $n \geq 4$ esetén van legalább két összekötetlen szimpliciális pont.

— $n \geq 4$ esetén bármely szimpliciális pontot elhagyva egy kisebb delta-diamond gráfot kapunk.

— Ha egy n pontú összefüggő, merev körű részgráfjuk d darab háromszöggel bír, akkor a részgráfban az élek száma $n + d - 1$. (Tehát a kiindulási gráfhoz képest pont annyival csökken az élek száma, mint a háromszögeké.)

A következő két oldalon példákat mutatunk $8, 7, 6, 5, 4$ pontú delta-diamond gráfokra. A fent elsorolt tulajdonságok jól megfigyelhetők. A későbbi ábra-oldalon a $6, 5, 4$ pontú delta-diamond gráfok teljes listáját látjuk.





A következőkben rámutatunk a (HW) egyenlőtlenség és a delta-diamond gráfok köztötti legfőbb kapcsolatra, melyet — lényegét tekintve — (legalább) három egymástól független kutatócsopot is megtalált: *Bukszár József* (aki Prékopa András, Szántai Tamás és Hujter Mihály kollégákkal sokat konzultált), *Klaus Dohmen*, továbbá *Pierangela Veneziani* (aki Boros Endre témavezetőjével dolgozott).

Jelölje $n \geq 3$ esetén egy konkrét (de tetszőleges) delta-diamond gráf éleinek halmazát E és háromszögeinek halmazát D. Ekkor fennáll a következő:

$$\text{GAP} \geq \sum_{ij \in E} P(A_i \cap A_j) - \sum_{rst \in D} P(A_r \cap A_s \cap A_t) \qquad \text{(D*)}$$

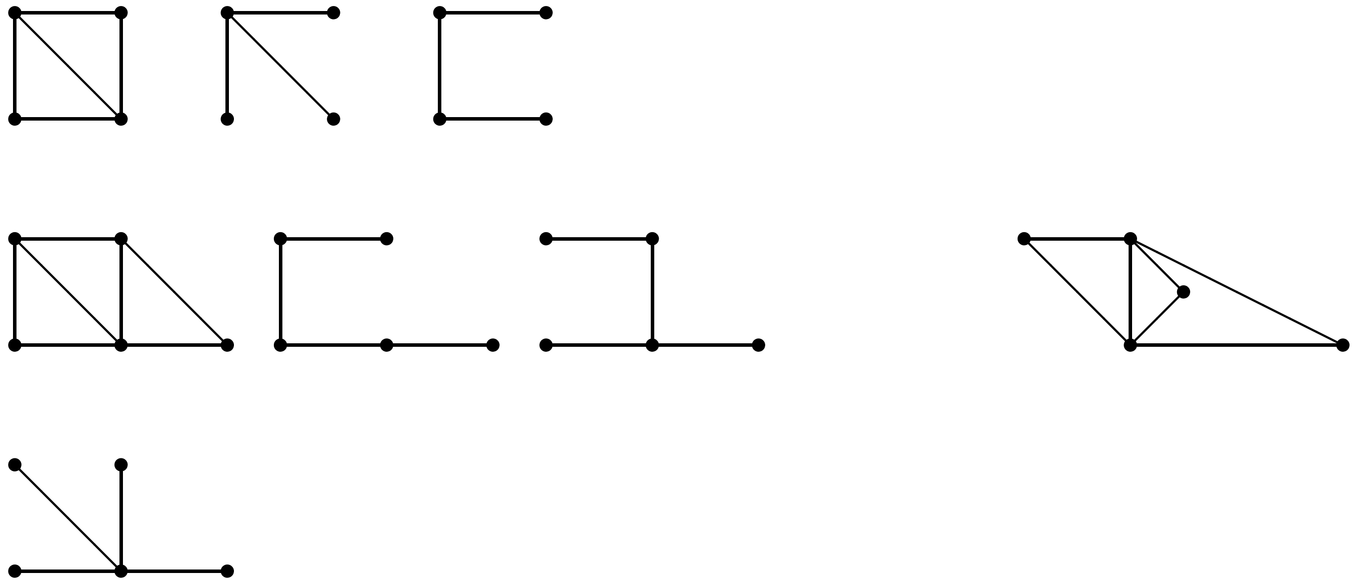
A jelen munka egyik legfőbb eredménye a következő tétel, mely adott $n \geq 3$ szögpontra vonatkozik.

Tétel. *Minden T élhalmazzal adott fához található olyan E élhalmazú és D háromszöghalmazú delta-diamond gráf, melyben $T \subset E$, továbbá létezik $\varphi : (E \setminus T) \to D$ bijekció, melyre minden $e \in E \setminus T$ él az egyik éle a $\varphi(e)$ háromszögnek.*

Következmény. A legjobb (D*) becslés nem rosszabb, mint a legjobb (HW) becslés.

Algoritmikus megjegyzés. A fenti tétel esetében T-ből E és D megkonstruálása nem igényel $O(n^2 \log n)$ lépésnél többet.

A fenti tétel részleges bizonyítása. Az $n = 4$ eset triviális, az $n = 5$ eset is könnyű a következő oldalon látható ábrák alapján.



A következőkben az $n = 4$ esetet vizsgáljuk alaposabban. Kiderül, hogy a (D*) egyenlőtlenség jobb oldala akkor a lehető legnagyobb, ha az E halmazból azt az rs párt hagyjuk ki a 6 lehetséges pár közül, melyre $\{t, w\} = \{1, 2, 3, 4\} \setminus \{s, t\}$ jelöléssel

$$q_{st} = P(A_r \cap A_s \cap A_t) + P(A_r \cap A_s \cap A_w) - P(A_r \cap A_s)$$

értéke a lehető legnagyobb. Abból a célból, hogy jobban megértethessük ezt a esetet, tekintsük a következő konkrét értékeket:

ij	12	13	14	23	24	34
$P(A_i \cap A_j)$	.16	.24	.14	.22	.14	.21

rst	123	124	134	234
$P(A_r \cap A_s \cap A_t)$	.14	.06	.11	.10

Tehát most a következő 6 szám legnagyobbikát kell kikeresni.

st	12	13	14	23	24	34
q_{st}	.04	.00	.03	.01	.02	.00

A legnagyobbat $s = 1$, $t = 2$ esetére kaptuk. Mindazonáltal (D*) ezt az alakot nyeri:

$$\text{GAP} \geq (.24 + .14 + .22 + .14 + .21) - (.11 + .10) \;= .74$$

Érdemes megvizsgálni, hogy vajon ez-e a lehető legnagyobb alsó becslés GAP értékére. Mármost z-vel jelölve

$$P(A_1 \cap A_2 \cap A_3 \cap A_4)$$

értékét ezeket az adatokat nyerjük:

rst, k	123,4	124,3	134,2	234,1
$P((A_r \cap A_s \cap A_t) \setminus A_k)$	$.14 - z$	$.06 - z$	$.11 - z$	$.10 - z$

és így

$$
\begin{aligned}
.16 - z - (.14 - z) - (.06 - z) &= z - .04 \\
.24 - z - (.14 - z) - (.11 - z) &= z - .01 \\
.14 - z - (.06 - z) - (.11 - z) &= z - .03 \\
.22 - z - (.14 - z) - (.10 - z) &= z - .02 \\
.14 - z - (.06 - z) - (.10 - z) &= z - .02 \\
.21 - z - (.11 - z) - (.10 - z) &= z - .00
\end{aligned}
$$

miatt

$ij, \quad kl$	12, 34	13, 24	14, 23
$P\Big((A_i \cap A_j) \setminus (A_k \cap A_l)\Big)$	$z - .04$	$z - .01$	$z - .03$

$ij, \quad kl$	23, 14	24, 13	34, 12
$P\Big((A_i \cap A_j) \setminus (A_k \cap A_l)\Big)$	$z - .02$	$z - .02$	$z - .00$

Mármost ezekkel az adatokkal GAP pontos értéke

$$\begin{aligned} & (z-.04)+(z-.01)+(z-.03)+(z-.02)+(z-.02)+(z-.00) \\ = \;& 6z-.12 \end{aligned}$$

és

$$(.14-z)+(.06-z)+(.11-z)+(.10-z)=.41-4z$$

miatt

$$(2-1)\,(6z-.12)+(3-1)\,(.41-4z)+(4-1)\,z=z+.70$$

Mivel a fenti $z-a$ típusú és $b-z$ típusú valószínűségértékeknek mind nemnegatívnak kell lenni, ezért z értékére fenn kell állni, hogy $0 \leq z \leq .04$. Mindazonáltal a konkrét esetben GAP lehető legnagyobb értéke: $.70+.04=.74$. Szerencsénk volt tehát, mert a (D*) becslés a lehető legerősebb becslésnek bizonyult. A fenti módszer tulajdonképpen lineráris programozás alkalmazása volt.

Most megvizsgálunk egy másik konkrét példát is. Legyen továbbra is $n = 4$, és legyenek $0 \leq p \leq q \leq .5$ paraméterek, melyekre feltesszük, hogy az A_1, A_2, A_3, A_4 események által meghatározott 16 esemény-atom közül a releváns 11 atom valószínűsége úgy alakul, hogy ha egy atom $k \geq 2$ darab eseményben benn van, de $4-k$ eseményben nincs benn, akkor a valószínűsége éppen $p^k q^{4-k}$. Jelen esetben

$$\mathrm{GAP} = 6p^2q^2 + 8p^3q + 3p^4 = p^2\left(3p^2 + 8pq + 6q^2\right)$$

Most a q_{st} számok mindegyikére ugyanazt kapjuk:

$$q_{st} = 2p^3(p+q) - p^2(p^4 + 2p^3q + p^2q^2)$$

Mindazonáltal a (HW) becslés szerint

$$\mathrm{GAP} \geq 3(p^4 + 2p^3q + p^2q^2) = \ 3p^2\,(p+q)^2$$

A (D*) becslés szerint viszont

$$\mathrm{GAP} \geq 5(p^4 + 2p^3q + p^2q^2) - 2p^3(p+q) = p^2\,(p+q)\,(3p+5q)$$

Látható tehát, hogy ez esetben a (D*) becslés ennyivel jobb, mint a (HW) becslés:

$$p^2\,(p+q)\,(3p+5q) - 3p^2\,(p+q)^2 = 2p^2q\,(p+q)$$

Ugyanakkor a (HW) becslés és a (BB) becslés közti különbség:

$$p^2\left(3p^2 + 8pq + 6q^2\right) - 3p^2\,(p+q)^2 = p^2q\,(2p+3q)$$

Néhány konkrét p, q értékre ezeket a számokat táblázatokba foglaltuk:

p	q	HW	D*	GAP
.5	.5	.750	≥ 1	≥ 1
.4	.5	.389	.533	.573
.3	.5	.173	.245	.267
.2	.5	.059	.087	.097
.1	.5	.011	.017	.019

p	q	HW	D*	GAP
.4	.4	.307	.410	.435
.3	.4	.132	.183	.197
.2	.4	.043	.063	.069
.1	.4	.008	.012	.013

p	q	HW	D*	GAP
.3	.3	.097	.130	.138
.2	.3	.030	.042	.046
.1	.3	.005	.007	.008

p	q	HW	D*	GAP
.2	.2	.0192	.0256	.0272
.1	.2	.0027	.0039	.0043
.1	.1	.0012	.0016	.0017

Most rátérünk az $n \geq 5$ esetekre. A legjobb delta-diamond gráf megkeresése nagyon nehéz; vélhetően NP-nehéz probléma. Mindazonáltal a Kruskal-féle algoritmus mintájára javasolhatunk egy hatékony minőségre számot tartó módszert. Ennek a lényege az, hogy mindannyiszor, amikor a Kruskal-algoritmus a (HW) becslés kiszámításához két komponenset egyesít, egy él behúzása helyett néha két, néha három élt húzunk be, de ha 2 élt húzunk be, akkor 1 darab új háromszöget is létrehozunk, ha pedig 3 élt is húzunk be, akkor 2 darab új háromszöget is létrehozunk. Az algoritmus összesen $n-1$ lépésben dolgozik: Kezdetben $n-1$ darab szingleton komponens volt, végül egyetlen delta-diamond gráf lesz. Mind az $n-1$ darab lépés után minden egyes komponens külön-külön vagy egy szigleton, vagy egy él, vagy egy $k \geq 3$ pontú delta-diamond gráf lesz. Minden lépésben azt a két kompenset csatoljuk össze egyetlen komponenssé, amely két komponenset a Kruskal-algoritmus is összekapcsolná. Az összekapcsolás módja azonban 3 féle lehet:

— Ha két szingleton összekapcsolása van napirenden, akkor ugyanazt csináljuk, mind a Kruskal-algoritmus, azaz egyetlen élt húzunk be a két szingleton között.

— Ha egy szingleton és egy nem szingleton összekapcsolása van napirenden, akkor a nem szingleton egyik élével és a szingletonnal alkotunk egy új háromszöget. Olyan új rst háromszög létrehozása történik, ahol s maga a szingleton, rt egy már meglévő él a nem szingletonban, továbbá

$$P(A_r \cap A_s) + P(A_s \cap A_t) - P(A_r \cap A_s \cap A_t)$$

értéke a lehető legnagyobb. Ennek megfelelően kell megválasztani a nem szingletonban az rt élt.

— Ha két nem szingleton összekapcsolása van napirenden, akkor a meglévő gráfunkra ráúniózunk egy olyan diamond gráfot, melynek egyik független élpárja egyik éle az egyik meglévő nem szingleton komponensből legyen, a másik éle a másik meglévő nem szingleton komponensből. Olyan diamond kiválasztása szükséges, melyre ha ij jelenti az egyik régi élt, rt jelenti a másik komponensből a régi élt, akkor

$$P(A_i \cap A_r) + P(A_j \cap A_r) + P(A_j \cap A_t) - P(A_i \cap A_j \cap A_t) - P(A_j \cap A_r \cap A_t)$$

értéke a lehető legnagyobb legyen.

A most javasolt algoritmust *diamond-Kruskal* algoritmusnak nevezzük. Könnyen látható, hogy lépésigénye $O(n^3)$. Gyakorlati alkalmazások által keletkeztetett adathalmazokon további vizsgálatok kívánatosak.

Végezetül kimondjuk a diamond-Kruskal algoritmus működését garantáló tételt:

Tétel. *A fent definiált diamond-Kruskal algoritmus fenntartja minden egyes lépése után azt a helyzetet, hogy az összes összefüggőségi komponens külön-külön vagy egy szingleton, vagy egy él, vagy egy háromszög, vagy egy gyémánt, vagy egy legalább 5 pontú delta-diamond gráf.*

Hivatkozások.

Alajaji,F., Kuai,H., and Takahara,G., *A lower bound for the probability of a finite union of events*, Discrete Appl. Math. 215 (2000)147–158.

Boros,E., and Veneziani,P., *Bounds of degree 3 for the probability of the union of events*, Rutcor Research Report 3-02 (2002).

Bukszár,J., and Prékopa, A., *Probability bounds with cherry trees*, Math. Oper. Res. 26 (2001) 174–192.

Bukszár,J., and Szántai, T., *Probability bounds given by hypercherry trees* [in Hungarian], Alkalmaz. Mat. Lapok 19 (1999) 69–85.

Bukszár, J., Szántai,T., *Probability bounds given by hypercherry trees*, Optimization Methods and Software 17 (2002) 409–422.

Dawson,D.A., and Sankoff,S., *An Inequality for Probability*, Proc. Amer. Math. Soc. 18 (1967) 504–507.

Hunter,D., *An upper bound for the probability of the union*, J. Appl. Prob. 30 (1975) 597–603.

Kounias,S., and Marin,J., *Best linear bonferroni bounds*, SIAM J. Appl. Math. 30 (1976), 301–326.

Kruskal,J.B., *On the shortest spanning subtree of a graph and the travelling salesman problem.* Proc. Am. Math. Soc. 7 (1956) 48–50.

Prékopa,A., *Boole-Bonferoni inequalities and linear programming*, Operations Research 36 (1988) 145–162.

Prékopa,A., *Sharp bounds on probabilities using linear programming*, Operations Research 38 (1990) 227–239.

Prékopa,A., Vizvári, and Regős,G., *Lower and upper bounds on probabilities of Boolean functions of events, Rutcor Research Report*, 36-95 (1995).

Takács,L., *On the method of inclusion and exclusion*, J. Am. Stat. Assoc. 62 (1967) 102–113.

Veneziani,P., *Upper bounds of degree 3 for the probability of the union of events via linear programming*, Discrete Appl. Math. 157 (2009) 858–863.

Vizvári,B., *New upper bounds on the probability of events based on graph structures*, Math. Inequal. Appl. 10 (2007) 217—228.

Worsley,K.J., *An improved Bonferroni inequality and applications*, Biometrika 69 (1982) 297–302.