

Rekurzív függvények és Morgan-fák*

Damir Vukičević¹, Tomislav Došlić² and Ante Graovac^{3,4}

¹Department of Mathematics, University of Split, Nikole Tesle 12, HR-21000 Split, Croatia

²University of Zagreb, Faculty of Civil Engineering, Kačićeva 26, 10000 Zagreb, Croatia

³The Rugjer Bošković Institute, P.O.B. 180, HR-10002 Zagreb, Croatia

⁴Department of Science, University of Split, Nikole Tesle 12, HR-21000 Split, Croatia

1. Bevezetés

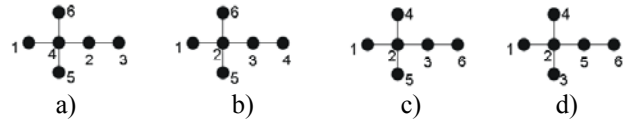
Ebben a közleményben Morgan-fák [4] fogalmát elemezzük és megvizsgáljuk alkalmazásukat adott számú csúccsal rendelkező nem izomorf fa halmazának meghatározására. Munkánkat Lukovits és szerzőtársainak [1-3,5] a munkája inspirálta és dolgozatunkal az [5] közleményben megadott közelítést próbáljuk megjavítani. Bevezetésünk a [6] dolgozat bevezetését követi, majd [5] és [7] numerikus adatait kiterjesztjük és finomítjuk, végül fő eredményünket, egy rekurzív algoritmust mutatunk be, amely felhasználható Lukovits [5] eredményeinek a megjavítására.

Jól ismert hogy a számozott, n csúccsal tartalmazó fák száma n^{n-2} [12]. Ez rendkívüli nagy szám: $n=20$ esetén például, $2.6 \cdot 10^{23}$ számozott fa létezik. A nem izomorf fák száma n csúcsra sokkal kisebb: 4 csúcs esetén például csak 2 nem izomorf fa van, szemben 16 számozott fával. Ezért ha valaki az összes nem izomorf fa iránt érdeklődik, igen rossz határfokú, ha először generálja az összes számozott fát, majd kiküszöböli az izomorfakat. Ez magyarázza azt a jelentős erőfeszítést, amelyet azért tettek, hogy olyan kisebb osztályt találjanak meg fák, amelyek még tartalmazzák az összes nem izomorfát (olyan kevés példányban amennyiben lehetséges). A [8] közlemény arra az ötletre épült, hogy elkészítsék az összes, számozott fizikai fát. Itt a csúcsok számozása során kikötötték, hogy minden új sorszámozott csúcs korábban sorszámozott csúcs szomszédja legyen. Belátható, hogy a fizikai fák olyan valódi részalmazát alkotják a számozott fák, amely tartalmazza az összes nem izomorf fát. További megkövetést bevezetése történt [4] a Morgan-fák fogalmával:

- 1) Jelöljük 1-el az egyik v csúcsot.
- 2) A szomszédjait számozza $2, \dots, d(v)+1$, ahol $d(v)$ a csúcs foka.
- 3) Tekintsük azt a csúcsot, amelynek legkisebb a sorszáma azon csúcsok között, amelyeknek van nem számozott szomszédja. Ezután számozzuk meg a szomszédokat növekvő sorszámmal a következő, még ki nem osztott számokkal.
- 4) Ismételjük ezt az eljárást addig, amíg minden csúcsot meg nem számoztunk.

Ezt a három fogalmat a következő, a [6] –ban megadott ábrával szemléltetjük:

Az 1a) Ábra gráfja számozott fa, de nem fizikai fa, ugyanis a 2-es csúcs nem szomszédja az előző sorszámu csúcsnak (az



1. Ábra. Számozott fák.

1-es sorszámuának). Az 1b) Ábra gráfja fizikai fa de nem Morgan-fa, mivel a 4-es sorszámu csúcs nem lehetne számozható, amíg a 2-es csúcs mindegyik szomszédja sorszámozott. Az 1c) és 1d) Ábra gráfjai Morgan-fák. Az egyszerűség kedvéért az i -vel számozott csúcsot v_i -vel jelöljük. Az 1d) Ábra adjacencia mátrixa a következő:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Vegyük észre, hogy minden oszlopban csak egy 1-es van a fődiagonális felett. Megmutatható [7], hogy ez a tulajdonság érvényes mindegyik Morgan-fára. Minden ilyen mátrixra megadható továbbá egyértelműen egy út a következő módon:

- 1) Az út az A_{11} mátrixelemmel kezdődik és vízszintesen addig folytatódik, amíg 1-eseket talál;
- 2) ezután függőlegesen halad, amíg 1-est nem talál a jobb oldalán;
- 3) majd ismét vízszintesen folytatódik;
- 4) és így tovább...

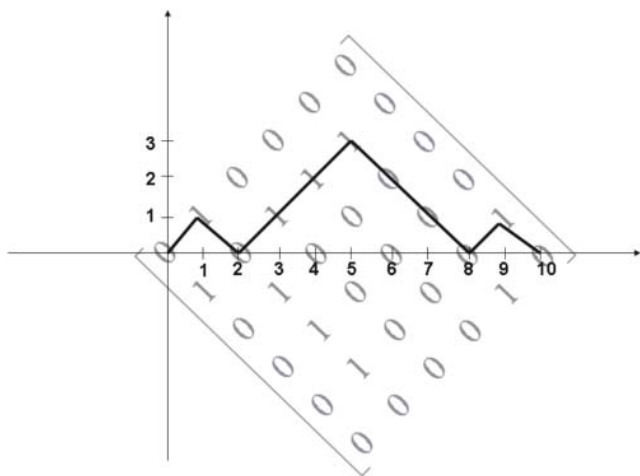
Alább látható az A mátrixhoz rendelt út. (Itt a folytonos vonal megfelel az útnak és a szaggatott vonal, pedig a fődiagonálisnak.)

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

2. Ábra. Utak az A mátrixon.

*fordította László István

Morgan-fák adjacencia mátrixaihoz rendelt utak a rácsokra rajzolt utak egy speciális fajtáját alkotják, a Dyck-utakat. A Dyck-út egy olyan, rácsra rajzolt út (0,0)-ból (2k,0)-be, amely minden lépésben az (1,1) (U lépés) vagy az (1,-1) vektor (D lépés) irányában található szomszédos pontra megy úgy, hogy sohasem keresztezi az x tengelyt. A kölcsönösen egyértelmű megfeleltetés Morgan-fák és Dyck-utak között legjobban akkor látszik, ha 45 fokkal elforgatjuk a mátrixot, és azután rajzoljuk rá az utat. Az A matrix megfeleltetését a következő Ábra mutatja:



3. Ábra. Megfeleltetés Morgan-fa és Dyck-út között

Tehát az A mátrixnak az UDUUDDDDUD út felel meg.

Ezt a megfeleltetést először a [7] közleményben ismertették. Köztudott, hogy a Dyck-utakat Catalan-számok sorozatával számolhatjuk meg [13]. Mivel az n -ik Catalan-számot

$\frac{1}{n+1} \cdot \binom{2n}{n}$ adja, az n csúcsot tartalmazó Morgan-fák

száma $\frac{1}{n} \binom{2n-2}{n-1}$.

A Catalan számsorozat az egyike a jól ismert kombinatorikus sorozatoknak – több mint száz különböző kombinatorikai interpretációja van [13]. Aszimptotikus viselkedését a $C_n \sim \text{const} \cdot 4^n \cdot n^{-3/2}$ összefüggés adja. Ez a mennyiség nagyon gyorsan növekszik n függvényében. A 2. fejezet egyik fő eredményeként fogjuk megmutatni, hogy azon kémiai (azaz olyan Morgan-fák, melyekre a fokszám legfeljebb 4) Morgan-fák száma, melyekre $d(v_1)=1$, asimptotikusan megegyezik $0.9102277 \cdot n^{-3/2} \cdot 3.6107186^{n-1}$ -el, ami sokkal kisebb szám, mint az előző (de még meglehetősen nagy). Azonban ha n nagyon nagy érték felé tart ezeknek a gráfoknak a hányada 0-hoz fog tartani:

$$\lim_{n \rightarrow \infty} \frac{0.9102277 \cdot n^{-3/2} \cdot 3.6107186^{n-1}}{4^n} = 0$$

$$\frac{4n\sqrt{\pi}(n-1)}{4^n}$$

Ezért, ha valaki a kémiai fákra kíváncsi, nagyon fontos, hogy csak a kémiai fák állítsa elő és ne az összes, hogy majd kiválogatja a kémiai fák. Ez utóbbi esetben a

gráfok előállítás, és a kémiai gráfok kiválogatása adná a számítógép idő nagy részét.

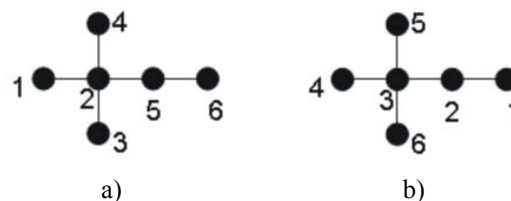
Vizsgáljunk meg néhány olyan javítást, melyeket az [1,2,5] közlemények javasolnak. A Morgan-fák rendelkeznek egy érdekes tulajdonsággal, hogy mindegyik csúcs (kivéve az 1-es sorszámút) csak egy olyan csúccsal van összekötve, amelyiknek kisebb a sorszáma mint az övé [1,2,5]. Ezt a tulajdonságot felhasználhatjuk arra, hogy mindegyik gráfhoz olyan CAM (condensed adjacency matrix) mátrixot rendeljünk, amely n csúcsot tartalmazó gráfra egy olyan $(n-1)$ -es, melyre, $CAM_{i,j}$ ha v_j olyan szomszédja v_{i+1} -nek, hogy teljesül $j < i+1$. Így az 1c) és 1d) Ábrákon bemutatott Morgan-fák CAM mátrixa rendre (1, 2, 2, 2, 3) és (1, 2, 2, 2, 5).

Megjegyezzük, hogy különböző CAM-ek izomorf gráfoknak is megfelelhetnek. Azért, hogy feloldjuk ezt a problémát, szükségünk van a lexikografikus rendezésre. Egész számok n hosszúságú sorozatára akkor és csak akkor létezik az $a < b$ lexikografikus rendezés az a és b sorozat között, ha van olyan $j \in \{1, \dots, n\}$ melyre $a_i = b_i$ minden $i < j$ -re, és $a_j < b_j$. Azért hívják lexikografikus rendezésnek, mert ezt használják a lexikonokban is.

Definíció szerint egy számozatlan T fához tartozó M Morgan-fát kanonikusnak nevezünk, ha minden olyan M' Morgan-fára, amelyik ugyanazon számozatlan fához tartozik, igaz hogy $CAM(M) \geq CAM(M')$. Ekkor azt mondjuk, hogy a $CAM(M)$ kanonikus CAM-ja T -nek. Kanonikus CAM-k egyértelműen meghatározzák a nem izomorf fák [1,2]. Mivel a Morgan-fák száma még sokkal nagyobb, mint a nem izomorf fák száma, érdemes tovább szűkíteni osztályukat.

Az LDF Morgan-fák számozását ugyanúgy végezzük, mint a Morgan-fákét, de még alkalmazzuk azt a szabályt is, hogy először mindig a kisebb fokszámú csúcsot számozzuk meg (LDF = Lowest Degrees First). Vegyük észre, hogy az 1c) Ábra gráfja nem LDF Morgan-fa, mivel a v_2 szomszédjainak a számozása során a 3-as sorszámot magasabb fokszámú csúcsnak adtuk, mint a 4-es és 5-ös sorszámot. Az 1d) Ábra gráfja pedig LDF Morgan-fa. Jegyezzük meg azt is, hogy LDF Morgan-fákra igaz hogy $d(v_1)=1$.

Azt is meg lehet mutatni, hogy minden kanonikus Morgan-fa egyuttal LDF Morgan-fa is [1,2]. Ezért mindig korlátozhatjuk magunkat LDF Morgan-fák vizsgálatára. Azonban, különböző LDF Morgan-fák még ugyanannak a gráfnak felelhetnek meg. A 4. Ábrán izomorf LDF Morgan-fák láthatók:



4. Ábra. Két, egymással izomorf LDF Morgan-fa.

A 4a) és 4b) izomorf gráfok CAM mátrixa rendre (1, 2, 2, 2, 5) és (1, 2, 3, 3, 3). Tehát további szabályokra és algoritmusokra van szükség nem kanonikus LDF Morgan-

fák kiküszöbölésére. Ezeket az algoritmusokat vizsgáljuk a 4. fejezetben.

2. Megszámolási eredmények

Legyen T egy olyan kémiai Morgan-fa, amelyre $d(v_1) \leq 3$. Legyen továbbá A a megfelelő adjacencia mátrix. Az első sorban (természetesen) nincs 1-es a fődiagonális előtt és $d(v_1)$ az 1-esek száma a fődiagonális után. Az összes többi sorban van egy 1-es a fődiagonális előtt és $d(v_i)-1$ az utána következő 1-esek száma. Minden esetben legfeljebb három 1-es van mindegyik sorban. Ezért azon kémiai Morgan-fák, amelyekre $d(v_1) \leq 3$ olyan Dyck-utaknak felelnek meg, melyek nem tartalmaznak "UUUU" sztringet. Természetesen egy Dyck-út akkor és csak akkor nem tartalmaz "UUUU" sztringet, ha nem tartalmaz "UUUUD" sztringet sem. Ilyen utak tanulmányozása található [9]-ben. Ott bebizonyították, hogy ezen utak b_n megszámlálási sorozatának $B(z)$ generátor függvénye kielégíti a következő egyenletet

$$B(z) = 1 + z \cdot B(z) + (z \cdot B(z))^2 + (z \cdot B(z))^3.$$

Ez az egyenlet ekvivalens az

$$U(z) = z(1 + U(z) + U(z)^2 + U(z)^3)$$

egyenlettel, ahol $U(z) = z \cdot B(z)$. Így az $U(z)$ függvényt az $F(z, U(z)) = 0$ egyenlet definiálja implicit módon, ahol $F(x, y) = y - x(1 + y + y^2 + y^3)$. Most használhatjuk a [10]-ben kifejtett eljárást b_n aszimptotikus viselkedésének meghatározására, ami

$$b_n \sim \frac{1}{2\sqrt{\pi}} \sqrt{\frac{1 + s + s^2 + s^3}{1 + 3s}} (1 + s)^{-3/2} r^{-n}$$

ahol r és s a következők

$$F(x, y) = 0 \quad \text{és} \quad \frac{\partial F}{\partial y}(x, y) = 0.$$

egyenletrendszer megoldása, és r a legkisebb megoldás. A pontos r és s értékeket számítógép algebra rendszerrel (például *Mathematica*) oldhatjuk meg. Az s -et a következő kifejezés adja

$$s = \frac{1}{6} \left[\left(53 + 6 \cdot \sqrt{78} \right)^{1/3} + \left(53 - 6 \cdot \sqrt{78} \right)^{1/3} - 1 \right].$$

Az r pontos értékét $r = \frac{s}{1 + s + s^2 + s^3}$ adja, ami elég komplikált és most mellőzzük a többi számítási részlettel. Ezeket összefoglalva, kapjuk a következő tételt.

1. Tétel

$$b_n \sim 0.102277 \cdot (n+1)^{-3/2} \cdot 3.6107186^n$$

Ha olyan CAM-ekre korlátozzuk figyelmünket, melyekre $d(v_1)=1$, akkor az UD sztringgel kezdődő Dyck-utakra szűkítjük vizsgálatunkat, melyek száma megegyezik a $2(n-1)$ hosszúságú Dyck-utakkal. Így a $d(v_1)=1$ feltételű kémiai Morgan-fák c_n számára a kissé módosított 1. Tételt kapjuk.

1'Tétel

$$c_n \sim 0.9102277n^{-3/2} 3.6107186^{n-1}$$

2. Algoritmus

Az [5] munka bemutat egy nem izomorf Morgan-fákat generáló algoritmust majd alkalmazásként, előállítja az összes 13 csúcsot tartalmazó kémiai nem izomorf fát. Ez az algoritmus a következő lépéseket tartalmazza:

- 1a) a [2] cikk szabályainak alkalmazásával keressük meg azt az első két CAM-et, melyek kanonikus fáknek felelnek meg (manuálisan végrehajtott keresés)
- 1b) kezdve (1,2,2,2,5,6,7,7,7,10,10,10) –vel, készítsük el az összes CAM-et
- 2) töröljük azon CAM-eket, melyekre $CAM_i = CAM_{i+1}$ és $d(v_{i+1}) > d(v_{i+2})$;
- 3) töröljük az összes olyan CAM-et, melyekben a számozás nem a leghosszabb oldalláncal kezdődik;
- 4) töröljük az összes olyan CAM-et, mely olyan fát ír le, ahol a maximális fokszám nagyobb mint 4;
- 5) alkalmazzuk [11] további két tételét hogy nem izomorf fákat kapjunk.

Ha ezt az algoritmust, 13 csúcsot tartalmazó fára alkalmazzuk, a következőket kapjuk:

- 1) lépés létrehoz 45958 fát, a 2) lépés ezt 6132-re redukálja és a 3) lépéssel 1872-re jutunk. Végül a 4) és 5) lépések 802 nem izomorf kémiai fát adnak.

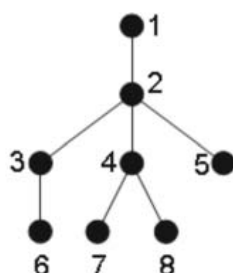
Láthatjuk, hogy az első lépés sokkal nagyobb számú fát hoz létre, mint a 3) lépés utáni fák száma. Az első lépésben közel 25-ször több fa van, mint a második lépésben. Ha az ember növeli a csúcsok számát, ez a hányados jelentősen fog növekedni. Nagy számú csúcs esetén az algoritmus sok időt töltene a 2) és 3) lépésekben a CAM-ek kiküszöbölésére. Ezért itt olyan rekurziós algoritmust javasolunk, amely sokkal kisebb számú CAM-et állít elő (természetesen az összes kanonikus CAM-et tartalmazó halmazt). A mi algoritmusunk közvetlenül generálja ($n=13$ esetén) a 2030 CAM-et (anélkül, hogy használnánk a [2] közleményt hogy megkapjuk az első két CAM-et). Ezután néhány további manuális optimalizáció jelentősen csökkentheti ezt a számot. Itt az $n=13$ esetre nem adunk meg semmi manuális optimalizációt, mivel algoritmusunkat olyan általánosan szeretnénk megadni, amennyire csak lehet. Ez az eljárás szükségtelenné teszi a 2), 3) és 4) lépéseket, és az 5) lépésben megtalálja az összes 802 nem izomorf gráfot.

Az egyszerűség kedvéért feltesszük hogy, $n > 2$, mivel az $n=1,2$ eset triviális. Megkülönböztetünk két fajta CAM-et: az egyikben az első csúcs 2-nél nagyobb fokszámú csúccsal szomszédos. A másikban, pedig az első csúcs szomszédjának foka 2. Mind a két osztályt rekurzív függvények generálják. Hasonló módon történik a generálásuk, csupán az első osztály rekurzív függvénye kiküszöböli azon CAM-eket, amelyekben van legalább kettőhosszúságú oldallánc. E miatt a hasonlóság miatt csak azt az algoritmust írjuk le, amely az első osztály CAM-jeit generálja és az érdeklődő olvasó könnyen megalkothatja a második osztály algoritmusát is.

Az algoritmus három állományt használ:

- 1) *CAM* (a generált *CAM*-ek tárolására)
- 2) *deg* (a csúcsok fokát tárolja)
- 3) *LessDeg* - a *LessDeg(i)* pozitív értéke azt mutatja, hogy $d(v_{LessDeg(i)}) \leq d(v_i)$ és a nulla érték azt jelzi, hogy ilyen egyenlőtlenség nem vizsgálendő.

A *CAM* és *deg* változók jelentése eléggé nyilvánvaló, így a következő példával a *LessDeg* változó jelentését világítjuk meg. Tegyük fel, hogy a rekurziós függvény elhelyezte az első nyolc csúcsot, és a gráf úgy néz ki, ahogy a következő Ábra mutatja:



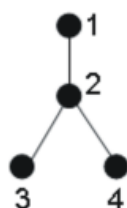
5. Ábra. Az gráf első nyolc csúcsa és azok megfeleltetési viszonyai.

A *LessDeg* értékei a következők (0,0,0,3,4,0,0,7). Ugyanis, a v_3 , v_4 és v_5 csúcsok a v_2 csúcsnak vannak megfeleltetve a számozás során, ezért az LDF szabály miatt $d(v_3) \leq d(v_4) \leq d(v_5)$. A v_7 és v_8 csúcsok a v_4 csúcsnak vannak megfeleltetve, tehát $d(v_7) \leq d(v_8)$.

Most lássuk az algoritmus pszeudo kódját:

```
main()
    deg(2) = deg(3) = 2 and deg(i) = 1
    for all i ≠ 2,3
    LessDeg(4) = 3 and LessDeg(i) = 0,
    for all i ≠ 4 and
    CAM1 = 1, CAM2 = CAM3 = 2
    rec(2,4)
END
```

A main függvény a következő Ábra szerint adja meg az első négy csúcsot.



6. Ábra. Az gráf első négy csúcsa és azok megfeleltetési viszonyai.

Minden ($v_i, i \geq 5$) csúcs fokai-1-nek van véve, kifejezve azt a tényt, hogy minden csúcs pontosan egy kisebb sorszámú csúccsal szomszédos (a magasabb sorszámú csúcsok a rekurziós algoritmus végrehajtása során kerülnek meghatározásra).

```
rec(LastValue, Location)
    int i, j;
    int flag; //indicator if LDF rule is respected and if
              //there are no leaves adjacent to
              //vertices of degree 2
    If (Location == n) //graph is constructed – check
        if it is canonical
            flag = true
        For j = LastValue, ...n
            If
                (deg[j] < deg[LessDeg[j]])
                    then flag = false
                If
                    (deg[j] == 1 AND deg[x[j-1]] == 2)
                        then flag = false
            EndFor
        If (flag) then Go to step 5 of the
            algorithm proposed in [5]
        Else
            If
                (deg[LastValue] < 4 AND LastValue > 1)
                    increase deg[LastValue] by 1
                    LessDeg[Location+1] = Location
                    CAM(Location) = LastValue
                    rec(LastValue, Location+1)
                    decrease deg[LastValue] by 1
            EndIf
        For all i such that LastValue+1 ≤ i ≤ Location
            and that (deg[i] < 4)
                flag = true
                For j = LastValue, ..., i-1
                    if
                        (deg[j] < deg[LessDeg[j]])
                            then flag = false
                            if
                                (deg[j] == 1 AND deg[x[j-1]] == 2)
                                    then flag = false
                            EndFor
                            If (flag)
                                increase deg[i] by 1
                                LessDeg[Location+1] = 0
                                CAM(Location) = i
                                rec(i, Location+1)
                                decrease deg[i] by 1
                            EndIf
                EndFor
            EndIf
        End
    End
```

4. Köszönetnyilvánítások

Hálásan köszönjük a Horvát Köztársaság Oktatási és Sport Minisztériumának (177--0000000-0884 (DV&TD&AG), 037-0000000-2779 (DV&TD) és 098-0982929-2940 (AG) számú) részleges támogatását.

5. Hivatkozások

1. Lukovits, I. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 553-568.
2. Lukovits, I. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 361-366.
3. Lukovits, I. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1147-1150.
4. Morgan, H. L. *J. Chem. Doc.* **1965**, *5*, 107-113.
5. Lukovits, I.; Gutman, I. *Croat. Chem. Acta.* **2002**, *75*, 563-576.

6. Vukičević, D. On the rules for the elimination of the non-canonical Morgan trees, *MATCH Commun. Math. Comput. Chem*, submitted.
7. Došlić, T. *Croat. Chem. Acta.* **2002**, *75*, 881-889.
8. Knop, J. V.; Müller, W. R.; Szymanski, K.; Kroto, H. W.; Trinajstić, N. *J. Comput. Chem.* **1987**, *8*, 549-555.
9. Sapounakis, A.; Tasoulas, I.; Tsikouras, P. *Discrete Math.* **2007**, *307*, 2909-2924.
10. Flajolet, P.; Sedgewick, R. *Analytic Combinatorics*, Cambridge University Press: Cambridge, **2008**.
11. Lukovits, I. in: Rouvray, D. and King, R. B. (Eds), *Chemical Graph Theory* (Proceedings of the Harry Wiener Memorial Conference), Athens, GA, Horwood Publishing Ltd., Westergate, Chichester, **2001**, pp. 327-337.
12. Cayley, A. On the analytical forms called trees, *Amer. J. Math.* **1881**, *4*, 266-268.
13. Stanley, R. P. *Enumerative Combinatorics*, Cambridge University Press: Cambridge, **1999**.

Recursive functions and Morgan trees

This paper is inspired by research of professor Lukovits [1-3,5] in proposing an efficient algorithm that constructs all non-isomorphic chemical trees using Morgan trees [4]. In this paper we optimize the procedure given in [5] by providing a recursive algorithm that

constructs only a small proportion of Morgan trees that contains all non-isomorphic chemical trees. We also determine the asymptotic behavior of the number of all chemical Morgan trees in order to show that this optimization is very important since even the number of all chemical Morgan trees is still very large.