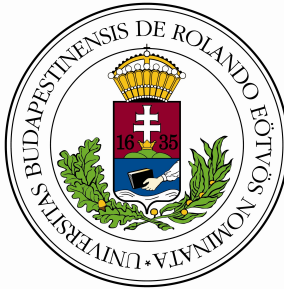


EÖTVÖS LORÁND TUDOMÁNYEGYETEM  
TERMÉSZETTUDOMÁNYI KAR



# A spektroszkópiai hálózatok matematikai elmélete

Írta:

Árendás Péter

matematikus mesterszak

Témavezető:

Dr. Szabó Csaba

Algebra és Számelmélet Tanszék

2013. május 31.

## *Köszönetnyilvánítás*

Ezúton is szeretném megköszönni dr. Szabó Csabának, hogy elindított a kutatási terület felé, és az elmúlt egy év során is kitüntető figyelmével, lelkesítésével és tanácsaival kísérte munkámat.

Köszönöm a segítséget, amit dr. Császár Attila és dr. Furtenbacher Tibor nyújtott a téma tanulmányozása, valamint a TDK-dolgozatok és a szakdolgozat írása során.

Külön köszönet illeti családomat és Ivancsó Veronikát támogatásukért és a tanulmányaimhoz szükséges feltételek optimalitásáért.

Köszönöm továbbá Stark Andrásnak kiváló ötleteit és meglátásait, Fegyverneki Tamásnak a statisztika világában adott útmutatásait, Biszak Elődnek a programozás terén nyújtott segítségét, valamint Jánosi Attilának a precíz lektorálást.

# Tartalomjegyzék

Köszönetnyilvánítás	i
Tartalomjegyzék	ii
Ábrák jegyzéke	iii
<b>1. Bevezetés</b>	<b>1</b>
<b>2. Elméleti kémiai háttér</b>	<b>4</b>
2.1. Bevezetés	4
2.2. Tesztrendszer: $\text{H}_2^{16}\text{O}$	7
<b>3. Gráfelméleti tulajdonságok</b>	<b>9</b>
3.1. Skálafüggetlenség	9
3.2. A skálafüggetlenség következményei	13
3.2.1. A hubok szerepe	13
3.2.2. Robosztusság	14
3.2.3. Ultra kis világ tulajdonság	15
<b>4. Problémák adaptációja</b>	<b>16</b>
4.1. Ismeretlen átmenetek keresése	17
4.2. Ismeretlen energiaszintek keresése	19
<b>5. További kutatási irányok</b>	<b>23</b>
<b>Függelék</b>	<b>25</b>
<b>Irodalomjegyzék</b>	<b>30</b>

# Ábrák jegyzéke

2.1. Összefüggés a spektrum és az SN között . . . . .	5
2.2. A $\text{H}_2^{16}\text{O}$ molekula spektruma adott frekvenciatartományon . . . . .	6
3.1. Ab initio tapasztalati fokszámeloszlás . . . . .	12
3.2. A kapott valószínűségi eloszlás . . . . .	12
3.3. A $\text{H}_2^{16}\text{O}$ SN robusztussága . . . . .	15
4.1. Fontos átmenetek elhelyezkedése az SN-ben . . . . .	18
4.2. Feszítőfa-algoritmus hibalehetősége új energiaszintekre . . . . .	19
4.3. Az új energiaszintek potenciális elhelyezkedése . . . . .	20
4.4. Az iteratív algoritmus szemléltetése . . . . .	21

# 1. fejezet

## Bevezetés

Mind a gráfelmélet, melynek kezdetét a königsbergi hidak problémájának megoldásával 1735-re tehetjük [3], mind a molekulaszpektroszkópia, melynek alapjait először az 1800-as évek elején fedezhetjük fel Fraunhofer és Kirchoff (egyébként königsbergi lakos) munkáiban [4], gazdag közös múltra tekint vissza [5]. A 2000-es évek elején mindkét terület kutatási fókuszja komoly változásokon ment keresztül. A gráfelméleti kutatások előterébe került a nagy hálózatok vizsgálata, illetve az eredmények gyakorlati alkalmazása többek közt a világháló, vagy különféle kapcsolati hálók esetén, melyek mérete általában meghaladta a milliós nagyságrendet is. Ezzel egyidejűleg a változatos mérnöki és tudományos alkalmazások – mint például ismeretlen bolygók és csillagok atmoszférájának szimulációi – egyre nagyobb molekulaszpektroszkópiai adatbázisok létrejöttét eredményezték.

A már néhány atomos molekulák esetén is a legalább milliós nagyságrendű spektroszkópiai adat meghatározása gyakorlati módszerekkel, mérésekkel gyakorlatilag lehetetlen. Ezt nem csak a szükséges idő- és erőforrásigény eredményezi, hanem a jelenleg rendelkezésre álló technológiai fejlettség mellett az összes adat csak egy mérhető tartományba eső hányadának megismerésére nyílik lehetőség.

Egy másik út a teljes spektroszkópiai adatbázis feltérképezésére a kvantumkémiai, *ab initio* számítások révén valósítható meg. Egyértelmű előny ezen az oldalon, hogy ilyen módon a teljes rendszer meghatározható, azonban a mérési adatokkal ellentétben itt csupán közelítő eredményeket kaphatunk.

A molekulaszpektroszkópiában felmerülő új problémák új módszerek életre hívását eredményezték. Császár Attila és Furtenbacher Tibor 2011-ben vezette be a **spektroszkópiai hálózat (SN)** fogalmát [6], mely egy matematikai értelemben vett nagy hálózatot jelent, ahol az egyes csúcsok az adott molekula rezgési-forgási energiaszintjeinek felelnek meg, a köztük futó élek az energiaszintek közötti átmeneteket reprezentálják, a rajtuk értelmezett nemnegatív súlyfüggvény pedig az adott átmenet intenzitásának felel meg. Ekkorra tehető a gráfelméleti alkalmazások kezdete is ezen a területen, és azóta is intenzív és érdekes kutatási témát jelent.

A 2. fejezetben egy rövid elméleti kémiai áttekintést adok a dolgozat megértéséhez szükséges alapfogalmakról és összefüggésekről. Ezt követik a tesztrendszerként használt  $\text{H}_2^{16}\text{O}$  molekula aktuális spektroszkópiai adatai és szerkezeti tulajdonságai.

A  $\text{H}_2^{16}\text{O}$  spektroszkópiai hálózatának vizsgálata során nagy szerepe van annak, hogy az SN-t skálafüggetlen hálózatként kezeljük [1, 7]. A 3. fejezetben egy, a kémiában jelenleg publikálttól eltérő statisztikai módszerrel támasztom alá, hogy a skálafüggetlen gráfok tulajdonságai valóban hatékonyan alkalmazhatóak a hálózatra.

A 4. fejezet főként a 2013. évi OTDK-dolgozatom eredményeit tartalmazza, a kutatási terület két fontos elméleti kémiai problémájára mutatok hatékony megközelítést gráfalgoritmusok felhasználásával [1, 2].

---

Az 5. fejezetben végül olyan további lehetséges kutatási irányokat foglalok össze, melyek esetében amellet, hogy fontos kérdések az elméleti kémiában, célravezető lehet a matematikai megközelítés.

## 2. fejezet

# Elméleti kémiai háttér

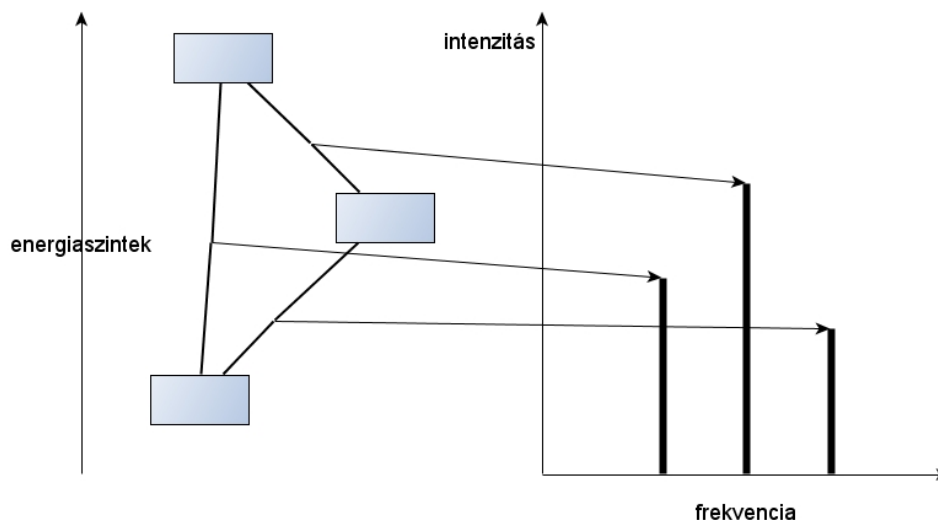
### 2.1. Bevezetés

A bevezetésben említett energiaszintek vizsgálata során a dolgozatban kizárólag a rezgési-forgási energiaszintekkel foglalkozunk, minden állítás és eredmény ezekre vonatkozik.

A kvantummechanika diszkrét energiaszinteket határoz meg. Azt, hogy két adott energiaszint között található-e átmenet, az adott molekulára jellemző ún. kiválasztási szabály deklarálja. További fontos tulajdonság a mérhetőség is: az elméleti úton kiszámított átmenetek intenzitása sok esetben a gyakorlati eszközökkel nem mérhető tartományba esik, tehát a számított, mérhető átmenetek által alkotott hálózatot a kiválasztási szabály által megadott élhalmaz egy vágásaként kapjuk meg.

A 2.1. ábra egy molekula három energiaszintjén keresztül szemlélteti az SN (bal oldalon) és a spektrum (jobb oldalon) közti összefüggést.





2.1. ÁBRA. Összefüggés a spektrum és az SN között

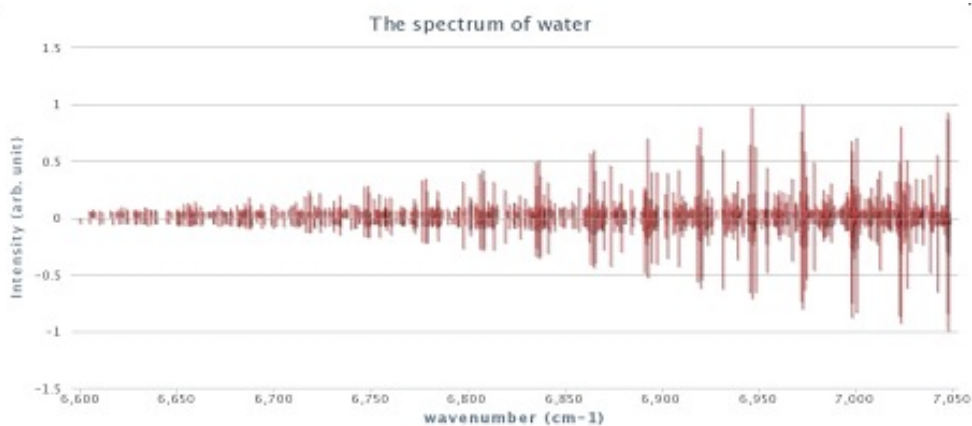
A spektroszkópia az anyag és az elektromágneses energia kölcsönhatását vizsgáló kémiai tudományág. A dolgozat során kémiai szempontból ennek a tudományterületnek a molekulaszpektroszkópiái részével foglalkozunk, ahol a középpontban a vizsgált molekula rezgési-forgási energiaszintjei állnak, melyek között adott intenzitású átmeneteken keresztül változhat a molekula állapota. Ez történhet abszorpció, azaz energiafelvétel során, vagy emisszióban, amely esetben a molekula energiát ad le.

Fontos hangsúlyozni, hogy egy adott molekula esetén két, együttesen jelenlevő spektroszkópiái hálózatot különböztetünk meg:

- az ún. **számított SN-t**, amely *ab initio* kvantumkémiai számításokkal határozható meg, tartalmazhat minden lehetséges energiaszintet és átmenetet, de csak közelítő pontosságú eredményeket szolgáltat mind az energiaszintekre (illetve a belőlük származtatott átmenetekre), mind az intenzitásokra (azaz az éleken értelmezett súlyfüggvényre),

- és az ún. **mért SN-t**, amely a mérésekkel meghatározott átmeneteket és energiaszinteket tartalmazza, és bár pontos átmeneteket tartalmaz, a teljes hálózatnak csak egy töredékét alkotja.

Ennek alapján tehát a mért hálózat (a súlyfüggvényt kivéve) a számított SN részstruktúráját adja mind az éleket, mind a csúcsokat tekintve. A mért hálózatban nem szereplő átmenetek és energiaszintek emellett közelítő értékeket adnak, melyek segítségével további mérési javaslatok fogalmazhatóak meg. A két hálózat együttes jelenlétét egy adott spektrumtartományon a 2.2. ábra szemlélteti [1] alapján (lenn láthatóak a számított hálózat átmenetei, fent pedig a jelenlegi mért adatbázis átmenetei).



2.2. ÁBRA. A  $\text{H}_2^{16}\text{O}$  molekula spektruma adott frekvenciatartományon

Azt, hogy egy adott SN-en belül mely energiaszintek között fut átmenet, a molekulára jellemző ún. kiválasztási szabály írja le. Az egyes energiaszinteket a molekula szerkezetében kvantumszámok azonosítják. Egy energiaszint asszignációja alatt a kvantumszámok hozzárendelését értjük, melyet új energiaszintek bevezetése esetén általában a vele szomszédos, már ismert kvantumszámú energiaszintek ismeretében tudunk elvégezni.

A számított spektroszkópai hálózat meghatározása többatomos molekulák esetén numerikus közelítés segítségével valósítható meg, melyre egy lehetséges, öt lépésből

álló algoritmust [8] tartalmaz. A módszer lényege abban áll, hogy először az energiaszintek kiszámítása történik meg, ezekből a kiválasztási szabály alkalmazásával megadhatóak a köztük futó átmenetek, végül a hozzájuk tartozó intenzitások nagysága.

## 2.2. Tesztrendszer: $\text{H}_2^{16}\text{O}$

A spektroszkópiai hálózatok szemléltetéséhez a  $\text{H}_2^{16}\text{O}$  molekula egy-foton abszorpció, illetve emissziós SN-jeit fogjuk felhasználni.

A  $\text{H}_2^{16}\text{O}$  a legfontosabb háromatomos molekula: ez a leggyakoribb háromatomos molekula a jelenleg ismert univerzumban, kiemelt szerepet játszik a földi üveg-házhatás megértésében, és spektroszkópiájának minél pontosabb leírása fontos asztrofizikai, asztrokémiai és termokémiai vonatkozásokkal bír. Választásunk mellett szól továbbá, hogy rendelkezésünkre áll egyrészt egy majdnem teljes, kvantumkémiai számításokkal meghatározott adatbázis [9], valamint egy nemrég validált, mérési adatbázis is [10]. Az alábbi táblázat a fontosabb különbségeket foglalja össze a két hálózat között.

	<b>Átmenetek száma</b>	<b>Energiaszintek száma</b>
Számított hálózat	505 806 202	221 097
Mért hálózat	99 335	17 690

A  $\text{H}_2^{16}\text{O}$  molekula esetén az előző részben is említett kvantumszámok egy 6 hosszú vektort jelentenek, melyek első három koordinátája a rezgési, második három koordinátája pedig a forgási állapotot kódolja. Ennek alapján minden csúcsot egyértelműen tudunk azonosítani ezzel a vektorral. Munkánk során az egyes energiaszinteket az ún. BT2-indexszel hivatkoztuk, amely forrása, [9], egyértelmű megfeleltetést ad az indexek és a kvantumszámok között.

A  $\text{H}_2^{16}\text{O}$  molekula szerkezetére jellemző továbbá, hogy az SN az ún. elektron-alapállapotban ("ground state") két diszjunkt ágon épül fel, melyek között nem fut él (ún. ortho és para ágak). Ennek köszönhetően találhatóak olyan átmenet-párok, melyek mindkét ágon megtalálhatóak, de a spektrumban ugyanott helyezkednek el. Kvantumkémiai módszerekkel azonban meg lehet különböztetni őket, intenzitásuk hozzávetőleg 70%-30% arányban különbözik. Erre a jelenségre később még figyelemmel kell lennünk.

## 3. fejezet

# Gráfelméleti tulajdonságok

### 3.1. Skálafüggetlenség

Skálafüggetlen gráfról beszélünk, ha annak a valószínűsége, hogy egy tetszőleges csúcs foka  $k$ :

$$P(k) \sim k^{-\gamma},$$

ahol  $\gamma$  az ún. skálaparaméter [13].

Császár és Furtenbacher [7]-ben fogalmazta meg, hogy a vizsgált spektroszkópai hálózatok fokszámeloszlása nagy hasonlóságot mutat a fenti eloszlással, így az SN-ek hasonló karakterizációval bírnak, mint a skálafüggetlen gráfok. A következőkben a  $\text{H}_2^{16}\text{O}$  molekula *ab initio* számított abszorpciós spektroszkópai hálózatának fokszámaira vonatkozó tapasztalati eloszlására mutatunk egy - a korábban publikált számításoktól eltérő - becsléses illeszkedésvizsgálatot [11, 12] alapján.

Mint látható, nem vizsgáljuk ebben a részben a mért spektroszkópai hálózatok paraméterbecslését. Ezt egyrészt az indokolja, hogy a mért hálózat méretét tekintve mind csúcsok, mind élek számában messze elmarad a számított hálózattól, ezért itt még nem érdemes az illeszkedést vizsgálni. Másik szempont emellett, hogy a

mért hálózat jelenlegi felépítettségét olyan külső tényezők is befolyásolják, mint a gyakorlati vegyészek mérési preferenciái, s ezek nem kívánt módon torzítják a tapasztalati fokszámeloszlást is.

A skálafüggetlenség definíciójából kiindulva tehát egy

$$P(X = k) = C * k^{-\gamma}$$

alakú kifejezést keresünk, ahol  $C$  alkalmas konstans szám. A mintaelemekről nem feltételezzük a függetlenséget, ezért  $\gamma$  becslését a Kolmogorov-Smirnov-távolság minimalizálásával végezzük. Azt a kitevőt keressük tehát, melyre a minta a legközelebb áll a fenti eloszláshoz.

Az eloszlás sajátossága, hogy a "power law" viselkedés csak egy adott küszöb felett karakterizálódik. Ebből a megfontolásból ezért a minta minden kezdőszületére elvégezzük a paraméterbecslést, és ezek közül választjuk ki a legjobb illeszkedést. Azt az  $x_{min}$  küszöbértéket keressük tehát a fokszámgyakoriságok között, hogy azokra a fokszámokra, melyek legalább  $x_{min}$ -szer előfordulnak a hálózatban, optimális lesz a paraméter becslése.

Ezután, mivel ahhoz, hogy eloszlást kapjunk, minden lehetséges  $k$ -ra összegezve 1-et kell kapnunk, az alábbi alakban keressük a  $C$  konstanst [11]:

$$C = \frac{1}{\zeta(\gamma, x_{min})},$$

ahol az  $x_{min}$  a minimális  $X$  értéket jelöli,  $\zeta$  pedig a Hurwitz-féle  $\zeta$ -függvény:

$$\zeta(\gamma, x_{min}) = \sum_{n=0}^{\infty} (n + x_{min})^{-\gamma}.$$

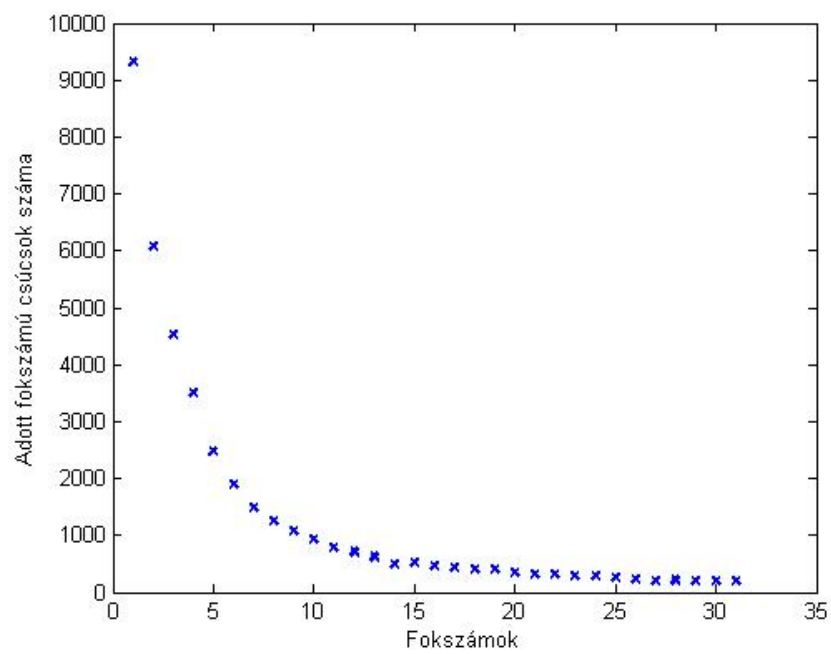
Az alábbi táblázat az SN tapasztalati fokszámeloszlásának első 10 elemét írja le.

Fokszám	Adott fokszámú csúcsok száma
1	9323
2	6068
3	4526
4	3508
5	2468
6	1892
7	1490
8	1255
9	1090
10	925

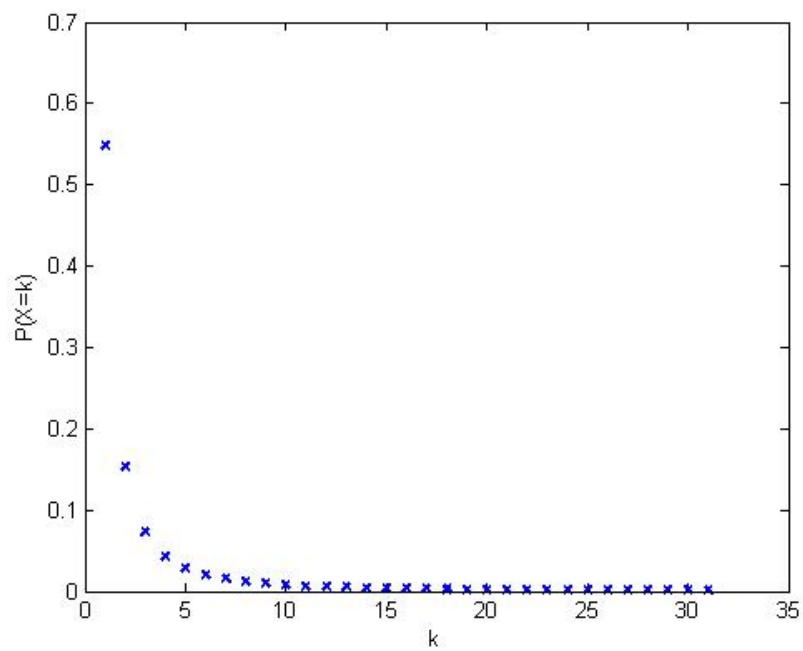
A modellezést [12] segítségével végeztük el az R programnyelven a molekula *ab initio* abszorpció SN-jére, miután intenzitásban a mérhető tartományra vágtuk a hálózatot. Az alábbi táblázat a paraméterbecslés értékeit foglalja össze.

Input SN	$\gamma$ becslése	$x_{min}$	$C$ érték	K-S-próba
Abszorpció, teljes	1,840851	198	1,824131	0,9991

A hálózatban a 31 fokszámú csúcsok közül található pontosan 198 db, tehát ezzel a paraméterbecsléssel a  $j = 1, \dots, 31$  fokú csúcsok gyakoriságát vizsgáljuk. Az  $x_{min} = 198$  érték mellett a tapasztalati fokszámeloszlást a 3.1. ábra szemlélteti, a most kiszámított eloszlásfüggvény pedig a 3.2. ábrán látható.



3.1. ÁBRA. Ab initio tapasztalati fokszámeloszlás



3.2. ÁBRA. A kapott valószínűségi eloszlás



## 3.2. A skálafüggetlenség következményei

A skálafüggetlenség egyik fő következménye, hogy relatív kevés nagy fokszámú csúcsot tartalmaz a hálózat. Ezzel párhuzamos az a megfigyelés, hogy túlnyomórészt kis, esetünkben 1-2 fokszámú csúcs alkotja a gráfot. Ezek alapján néhány fontos következményt is vonatkoztathatunk a spektroszkópiai hálózatra.

### 3.2.1. A hubok szerepe

Egy molekula tisztán mérésekkel történő spektroszkópiai vizsgálata, és az aktuálisan ismert (megmért, validált) energiaszint- és átmenetlistájának bővítése során az egyik fő problémát az jelenti, hogy egy-egy új asszignált átmenet nem ad lényeges információt a molekuláról. Másképp fogalmazva: pontosabb mérési javaslatok segítségével hatékonyabban tudnánk feltérképezni egy adott molekula szerkezetét. Mivel **relatív kevés nagy fokszámú csúcs** van jelen, ezek megismerésével (azaz az energiaszintek asszignálásával) jóval több információt kaphatunk a hálózat szerkezetéről. Kézenfekvőnek tűnik tehát, hogy a kvantumkémiai számítások segítségével meghatározott *ab initio* hálózaton határozzuk meg a nagy fokszámú csúcsokat, az ún. *hubokat*, majd a mérési javaslatok során irányítsuk figyelmünket a hubok elhelyezkedésére. Vizsgálatunk során a hubokat a csúcsok csökkenő sorrendbe rendezett fokszámának felső 10%-ánál határoztuk meg (ha egy adott fokszám-tartományt ezzel kettévágtunk, akkor ez a tartomány még nem került a hubok közé).

Tovább szűrhetjük a kiemelt jelentőségű energiaszintek körét, ha összevetjük a molekula *ab initio* abszorpciós és emissziós spektroszkópiai hálózatát is, és kiválasztjuk azokat az energiaszinteket, melyek mindkét esetben hubnak számítanak.

Az alábbi táblázat a  $\text{H}_2^{16}\text{O}$  abszorpciós és 1650 K emissziós *ab initio* SN-jeinek hub-tulajdonságait foglalja össze, illetve a két hálózat közös hubjait jellemzi.

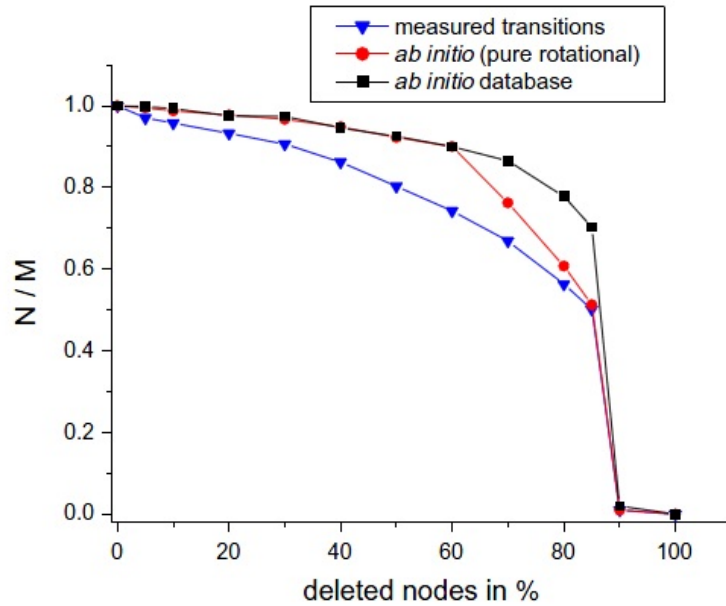
	Hubok sz.	Hubok min. fokszáma
Abszorpció, teljes	4381	32
Egy-foton emisszió, 1650K, teljes	1259	47
Közös	1258	-
Abszorpció, feszítőfa	1189	2
Egy-foton emisszió, 1650K, feszítőfa	856	4
Közös	564	-
Mind a négy hálózatban közös	329	-

Megállapítható tehát, hogy a  $\text{H}_2^{16}\text{O}$  molekula SN-jében található 329 olyan rezgési-forgási energiaszint, melyek megismerésével nagy mennyiségű információhoz juthatunk.

### 3.2.2. Robosztusság

A **kis fokszámú csúcsok nagy száma** okozza a spektroszkópai hálózatok robusztusságát. Az 3.3. ábra [7] alapján azt szemlélteti, hogy a hálózat csúcsainak véletlenszerű törlése mellett miként változik az összefüggőség.

A grafikon az SN összefüggőségét szemlélteti a véletlenszerűen törölt csúcsok %ának függvényében, ahol  $M$  a hálózat csúcsainak számát jelöli,  $N$  pedig a legnagyobb összefüggő komponens csúcsszámát.

3.3. ÁBRA. A  $\text{H}_2^{16}\text{O}$  SN robusztussága

### 3.2.3. Ultra kis világ tulajdonság

Azt mondjuk, hogy egy gráf kis világ tulajdonságú, ha a két tetszőlegesen kiválasztott csúcsa közti átlagos  $L$  távolságra  $N$  csúcs esetén fennáll:  $L \sim \log N$ . R. Cohen és S. Havlin megmutatták, hogy a skálafüggetlen gráfok ultra kis világ tulajdonsággal rendelkeznek, mivel esetükben  $L \sim \log \log N$  [14].

[1], [19] és [20] alapján további tulajdonságokkal is leírhatjuk az SN szerkezetét. Ezek egyike, hogy a hálózat magas fokszámú csúcsai főként alacsony fokszámú csúcsokkal vannak összekötve, emellett igaz az is, hogy nagyon sok él fut az alacsony fokszámú csúcsok között. Ezt igazolja többek közt az is, hogy az 1000 legnagyobb intenzitású átmenet, azaz legnagyobb súlyú él felhasználásával csupán 371 különböző energiaszint érhető el, továbbá összesen 1499 csúcs adja az 5000 legnagyobb súlyú él végpontjait.

## 4. fejezet

# Problémák adaptációja

A spektroszkópiai hálózatok esetén az alapinformációt az energiaszintek hordozzák. Ebből következően ezek minél pontosabb megismerése kiemelten fontos.

A kvantummechanikai számítások lehetővé teszik, hogy a teljes hálózat kiszámítható legyen, de csak közelítő pontossággal. Ennek felhasználásával minél jobb mérési javaslatokat szeretnénk adni új, még nem mért átmenetek formájában. Két esetet kell megkülönböztetnünk aszerint, hogy a javasolt átmenet mindkét végpontja ismert-e, vagy az átmenet egy ismert (azaz MARVEL-beli) és egy még ismeretlen energiaszint között fut.

Két ismert energiaszint közti új, nagy intenzitású átmenet javaslatával stabilizálhatjuk a mért hálózatot, azaz a mérési bizonytalanság szerepét nagyban csökkenthetjük, főként akkor, ha az új él egy új kört hoz létre a hálózatban. További gyakorlati szempont, hogy minél nagyobb az átmenet intenzitása, annál könnyebben mérhető.

A spektroszkópiai hálózatok által nyújtott lehetőségek szemléltetésére két probléma adaptációját mutatom be [2] alapján. Először arra a kérdésre keressük a

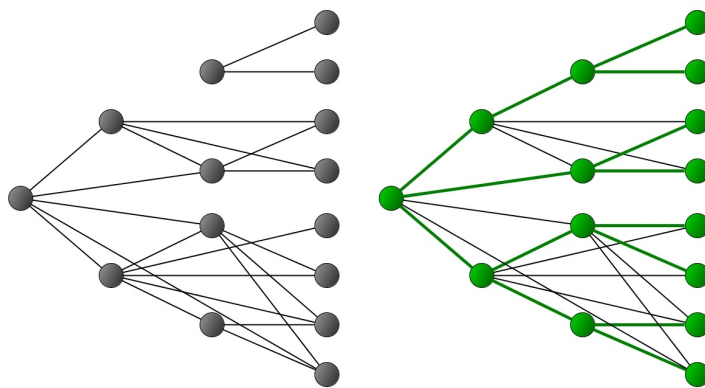
választ, hogy milyen módszerrel találhatunk még meg nem mért, de nagy intenzitású, és emellett szerkezetileg fontos szerepet játszó átmeneteket két ismert energiaszint között. Ezután szeretnénk bővíteni a mért SN-t új, könnyen asszignálható energiaszintekkel.

## 4.1. Ismeretlen átmenetek keresése

A fejezet bevezetésében feltett első kérdést átfogalmazhatjuk arra a problémára, hogy melyek azok az átmenetek az SN-ben, melyek "fontosak", több információt hordoznak a többinél. Ezt a módosított kérdést a hubok jelenléte ihlette, melyek az energiaszintek között számítanak "fontosnak". Az információtartalmat itt az adott átmenet intenzitásának feleltetjük meg, és egy olyan módszert szeretnénk adni, mely nem csak az ismeretlen átmeneteket adja meg, hanem a már ismertek intenzitásainak figyelembe vételével meghatározza a teljes hálózatban a kiemelt jelentőségű átmeneteket.

Fontos megjegyezni, hogy az általunk javasolt módszer nem csak új, nagy intenzitású átmenetek keresésére alkalmas, de járulékos haszonként kezel egy további problémát is: ha ugyanis egy még nem mért átmenet a mért spektroszkópiai hálózat két különböző komponense között teremt kapcsolatot, akkor azzal a mért SN összefüggőségét is javíthatjuk, ami szintén a célok közé tartozik a spektroszkópiai kutatásban.

A keresett, fontos átmenetek potenciális elhelyezkedését a 4.1. ábra szemlélteti.



4.1. ÁBRA. Fontos átmenetek elhelyezkedése az SN-ben

A fenti cél elérése az alábbi algoritmus segítségével megvalósítható:

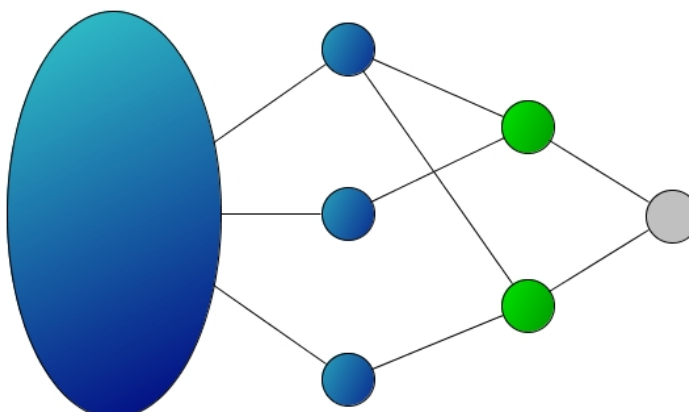
1. Készítsük el az ismert energiaszintek listáját a mért hálózat átmenetlistájából, ezek lesznek a hálózatunk csúcsai.
2. Szűrjük ki azokat az átmeneteket az *ab initio* átmenetlistából, melyek két ismert energiaszint között futnak, ezek adják majd a hálózatunk éleit.
3. Futtassunk le egy Kruskal-féle minimális súlyú feszítőfát létrehozó algoritmust az így kapott hálózatra [17].
4. Válasszunk ki egy tetszőleges vonalat a mért spektrumból, és keressünk hozzá feszítőfaéleket úgy, hogy intenzitásban 30% eltérést (a korábban említett orto-para ágak eltérése miatt), frekvenciában  $\epsilon_2$  (jelen esetben  $0,5 \text{ cm}^{-1}$ ) értékű hibahatárt engedünk meg. A feszítőfaél két végpontjának megfelelő energiaszintekre MARVEL-ben határozzuk meg ki a frekvenciakülönbséget, és ha az ettől számított eltérés kisebb, mint  $\epsilon_2$  (esetünkben  $0,005 \text{ cm}^{-1}$ ), akkor elfogadjuk az asszignációt. Folytassuk ezt a lépést addig, amíg van átvizsgálatlan feszítőfaél.

Az algoritmus segítségével [1]-ben sikerült közel 2000 új, nagy intenzitású átmenetet találnunk.

## 4.2. Ismeretlen energiaszintek keresése

Mint azt a 2. fejezetben is említettük, a  $\text{H}_2^{16}\text{O}$  molekula mért SN-je jóval elmarad még az *ab initio* hálózattól mind az átmenetek, mind az energiaszintek számában. Most arra keressük a választ, hogy milyen módon lehetne új energiaszintekkel bővíteni a mért hálózatot. Szeretnénk egyrészt könnyen asszignálható csúcsokat keresni, másrészt ezt minél kevesebb új átmenet felhasználásával megvalósítani.

Az előző megoldáshoz hasonlóan elsöre jó ötletnek tűnhet itt is feszítőfa-algoritmusok használata. Gyakorlati szempontból ez azonban ellenjavallott, mivel így nagy mennyiségű olyan javaslatot kapnánk új energiaszint elérésére, mely egy másik, nem ismert energiaszinten át érhető csak el.



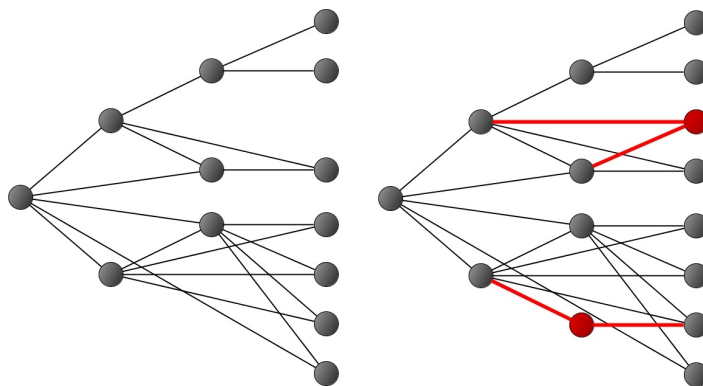
4.2. ÁBRA. Feszítőfa-algoritmus hibalehetősége új energiaszintekre

A problémát a 4.2. ábra szemlélteti. A késsel jelölt csúcsok és csúcshalmaz az ismert energiaszinteket jelölik, a zöld csúcsok a lehetséges új csúcsok javaslatait mutatják. A szürkével jelölt csúcs azonban – pontos mérési adatok hiányában – nem asszignálható még hatékonyan.

Egy gyakorlati szempontot is figyelembe fogunk venni, mely pedig az energiaszint könnyebb asszignációját szolgálja. A  $\text{H}_2^{16}\text{O}$  molekulára vonatkozó kiválasztási

szabály mellett egy adott ismeretlen energiaszint kvantumszámait könnyebb meghatározni, ha nem csak egy átmenetet ismerünk, hanem két ismert energiaszinthez is be tudjuk kötni.

A keresett új energiaszintek potenciális elhelyezkedését az előző részhez hasonló módon a 4.3. ábra szemlélteti.



4.3. ÁBRA. Az új energiaszintek potenciális elhelyezkedése

Ebből a megfontolásból a megfogalmazott célt inkább egy iteratív algoritmus segítségével próbáljuk meg elérni:

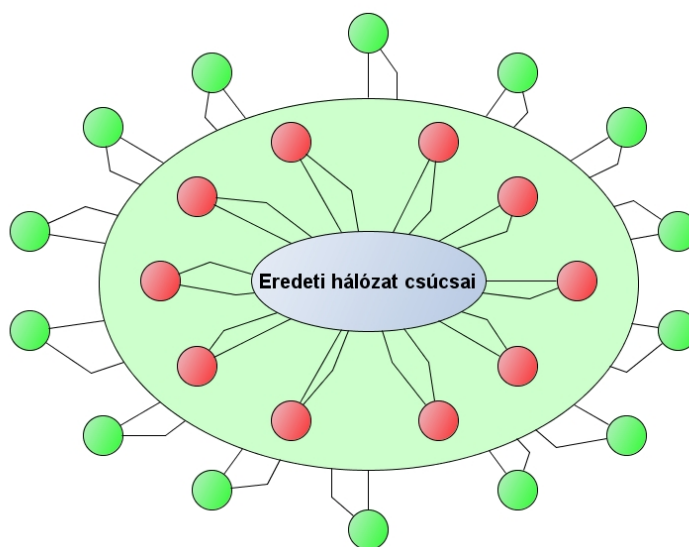
1. Határozzuk meg az *ab initio* átmenetlistából azokat az éleket, melyek pontosan egy mért és egy nem mért energiaszint között futnak.
2. Az így kapott listát vizsgáljuk az ismeretlen energiaszintek szerint, és ha találunk legalább két átmenetet, melyek ugyanabba az ismeretlen energiaszintbe mennek, válasszuk ki őket. Így a mért hálózat csúcsainak egy olyan szűkített szomszédsági halmazát kapjuk az *ab initio* hálózatban, hogy legalább két él megy a mért hálózatból minden "preferált" szomszédba.
3. A lehetséges új energiaszintek körét szűkítsük tovább. Egy adott új csúcshoz számítsuk ki a lehetséges frekvenciáját a két (vagy több) lehetséges átmeneten, az ismert csúcs mért frekvenciája, és a javasolt új átmenet intenzitása



összegeként. Amennyiben az új csúcs frekvenciájára kapott két érték nem tér el jobban, mint  $\epsilon$  (jelen esetben  $0,005 \text{ cm}^{-1}$ ), az új csúcsot írjuk ki a javasolt új energiaszintek listájába.

4. Az elméleti úton kiszámított eredményeket az utolsó lépésben a molekula egy mért spektrumával kell összevetnünk.

Az iteratív algoritmus működését a 4.4. ábra szemlélteti.



4.4. ÁBRA. Az iteratív algoritmus szemléltetése

Az algoritmus 3. lépését az alábbi táblázat szemlélteti, öt átmenetpáron keresztül. Az oszlopok tartalma: A: az új átmenet *ab initio* frekvenciája; B: új energiaszint BT2-indexe; C: a már ismert MARVEL-beli energiaszint BT2-indexe; D: a C-hez tartozó energiaérték; E: a B "jósolt" energiaértéke; F: az új energiaszint újonnan meghatározott kvantumszáma.

A	B	C	D	E	F
6714,2301	4293	1653	4856,2165	11570,4467	0 5 1 3 2 1
6643,5850	4293	3670	4926,8631	11570,4481	0 5 1 3 2 1
6794,5530	6534	4240	5065,3626	11859,9157	0 5 1 4 3 2
6696,6542	6534	7464	5163,2642	11859,9184	0 5 1 4 3 2
6640,2489	6560	7470	6701,8836	13342,1326	0 5 1 4 3 2
6737,8145	6560	4246	6604,3192	13342,1338	0 5 1 4 3 2
6756,6106	7520	6463	5342,1946	12098,8052	0 5 1 4 4 1
6634,5475	7520	9957	5464,2624	12098,8100	0 5 1 4 4 1
6640,8530	8309	9013	6701,4805	13342,3335	0 6 1 4 3 1
6737,9550	8309	5842	6604,3793	13342,3344	0 6 1 4 3 1

## 5. fejezet

# További kutatási irányok

Mivel a spektroszkópiai hálózatok alapvetően egy nagy hálózatként, egy irányítatlan gráfként épülnek fel, vizsgálatuk könnyen megközelíthető matematikai úton. Ebben a fejezetben néhány további problémakört vázolunk, melyek elemzése egyrészt a kémia számára hasznos, új eredményeket jelenthet, másrészt motiváló hatással lehet a matematikai módszerek fejlesztésére is. Ebben a fejezetben további, érdekes kérdéseket vetünk fel a területről.

A jelenleg rendelkezésre álló mérések alapján felépíthető spektroszkópiai hálózatok több komponensből állnak. Fontos lenne a kvantumkémiai és gráfelméleti háttér felhasználásával olyan mérési javaslatok megfogalmazása, melyeken keresztül optimálisan összefüggővé tehetünk egy adott hálózatot.

A molekula spektroszkópiai szerkezetének vizsgálatában nagy szerepet játszik az átmenetek mérési bizonytalansága. Ez a probléma többféle módon is kezelhető. Javítaná egyrészt az adott energiaszint bizonytalanságát, ha nem csak egy úton lenne elérhető a spektroszkópiai hálózaton belül, így szükség lenne egy hatékony kör-keresési algoritmus kifejlesztésére. Fontos lenne emellett megtalálni azokat az utakat, melyek egyik csúcsa sem szerepel egy körben sem a mért hálózatban,

tehát azokat, melyek törlésével több komponensre esik szét a gráf. Ezek bizonytalanságára épülnek ugyanis a velük szomszédos csúcsok, így egy pontatlanabb mérés nagy hatással van a hálózat egy nagyobb szeletére is.

A  $\text{H}_2^{16}\text{O}$  spektroszkópiai hálózatáról belátták, hogy a foksámok eloszlása folytán jól kezelhető egy skálafüggetlen gráfként. Ez három érdekes kérdést vet fel: (a) más többatomos molekulák SN-jére is fennáll-e a skálafüggetlenség, (b) milyen tényezők hatására változik a skálázási paraméter, (c) változik-e a skálafüggetlen viselkedés, ha a kvantumkémiai számításokkal meghatározott átmenetek intenzitását nem vágjuk le a jelenleg mérhető tartományra, hanem tisztán az elméleti hálózatot vizsgáljuk.

A mért spektroszkópiai adatok sok esetben duplikációkat tartalmaznak, azaz egy adott átmenet esetén több, eltérő érték is publikált. Szükség lenne ezért egy gráfelméleti és statisztikai alapú "worst offender" algoritmus kifejlesztésére az adatbázisok önkonzisztenciájának vizsgálatára, és a mérési adatok újrasúlyozására.

Bár a spektroszkópiai hálózatok kutatása új terület, több molekula esetén is vizsgálták már a hálózatok szerkezetét. Érdekes lenne különféle molekulaosztályok esetén összevetni a gráfok felépítését, és vizsgálni annak tulajdonságait a molekulaszerkezet változásának függvényében.

Komoly eredményekre vezethet továbbá egy olyan, tisztán matematikai modell felépítése, mellyel modellezhetőek a spektroszkópiai hálózatok alapvető tulajdonságai. A modellalkotás részeként nagy szerepe lehet az átmenetek intenzitásának (illetve az ezekben szerepet játszó Einstein  $A$  koefficiensnek) esetlegesen tisztán gráfelméleti tulajdonságokból történő meghatározásának.

# Függelék

Itt találhatóak a kutatás során felhasznált fontosabb programkódok. A lista nem teljes, de a kimaradó implementációk vagy csak néhány soros rutinműveleteket végeztek, vagy az egyik itt szereplő kód kisebb módosításaival előállíthatóak.

## C++ programkódok

**kruskal.cpp** – A Kruskal-algoritmus implementációja, mely a LEMON Graph Library [18] és annak segédanyagainak felhasználásával készült.

```
1 #include <iostream>
2 #include <fstream>
3 #include <vector>
4 #include <lemon/kruskal.h>
5 #include <lemon/list_graph.h>
6 #include <lemon/lgf_reader.h>
7
8 using namespace std;
9 using namespace lemon;
10
11 int main() {
12
13     ofstream myfile;
14     myfile.open ("feszitofaelek.txt");
15
16     ListGraph g;
17
18     typedef ListGraph::Node Node;
19     typedef ListGraph::Edge Edge;
20     typedef ListGraph::NodeIt NodeIt;
21     typedef ListGraph::EdgeIt EdgeIt;
22
23     ListGraph::EdgeMap<int> edge_cost_map( g );
24
25     try {
```

```

26 graphReader( g, "graf.lgf")
27     .edgeMap("length", edge_cost_map)
28     .run();
29 }
30 catch( FormatError e ){
31     cout<<"error " <<e.message()<<endl;
32 }
33
34 int count = 0;
35 for( EdgeIt i(g); i != INVALID; i++ ) count++;
36 cout<<count<<endl;
37
38 typedef ListGraph::EdgeMap<bool> EBoolMap;
39 EBoolMap tree_map(g);
40
41 vector<Edge> tree_edge_vec;
42
43 int k=0;
44
45 for(EdgeIt i(g); i!=INVALID; ++i){
46     if (tree_map[i]) {
47         myfile << g.id(i) <<"\n";
48         ++k;
49     }
50 }
51 std::cout << std::endl;
52 myfile.close();
53 return 0;
54 }

```

## Perl programkódok

**BT2conv.pl** – Konverter fájl, mely a C++-ban implementált Kruskal-algoritmus inputját (Lemon Graph Format fájl) állítja elő [16]-beli adatsorokból.

```

1 @ARGV < 2 and die "USAGE: perl.exe BT2conv.pl [infile] [outfile]";
2 open(IN,$ARGV[0]) or die "Unable to open infile...";
3 open(OUT,">$ARGV[1]") or die "Unable to open outfile...";
4
5 @nodes;
6 @edges;
7 $i = 0;
8
9 for( <IN> ){
10     @fields = split(/\s+/);
11
12     $node1 = ($fields[2]);
13     $node2 = ($fields[3]);
14
15     $k=0;
16     $m=0;
17
18     while ($k < $i+1){

```

```

19     if ($node1 eq $nodes[$k]) {
20         $m=1;
21         $k=$i;
22     }
23     else {
24         $k=$k+1;
25     }
26 }
27
28 if ($m == 0) {
29     $nodes[$i]=$node1;
30     $i=$i+1;
31 }
32
33 $k=0;
34 $m=0;
35
36 while ($k < $i+1){
37     if ($node2 eq $nodes[$k]) {
38         $m=1;
39         $k=$i;
40     }
41     else {
42         $k=$k+1;
43     }
44 }
45
46 if ($m == 0) {
47     $nodes[$i]=$node2;
48     $i=$i+1;
49 }
50
51 my $newlabel = $#edges+1;
52
53 $fields[1] = substr($fields[1], -2);
54
55 print "$i\n";
56 push( @edges, "$node1\t$node2\t$newlabel\t$fields[1]");
57 }
58
59 print OUT "\@nodes\nlabel\n";
60 for $i ( sort {$nodes{$a} <=> $nodes{$b}} keys @nodes ){
61     print OUT "$nodes[$i]\n";
62 }
63
64 print OUT "\n\n@edges\n\t\tlabel\tlength\n";
65 for $i ( 0 .. $#edges ){
66     print OUT "$edges[$i]\n";
67 }
68
69 close IN;
70 close OUT;

```

**marvelszinkron.pl** – Egy bemeneti átmenetlistát vet össze a MARVEL adatbázissal, és ad outputként MARVEL energiaszinteket aszerint, hogy az alsó vagy a felső energiaszint ismert jeleneg, vagy mindkettő, vagy egyik sem.

```

1 @ARGV < 2 and die "USAGE: perl.exe marvelszinkron.pl [MARVEL] [
  ellista] [output]";
2 open(IN1,$ARGV[0]) or die "Unable to open infile 1...";
3 open(IN2,$ARGV[1]) or die "Unable to open infile 2...";
4 open(OUT,">$ARGV[2]") or die "Unable to open outfile...";
5
6 %marvel;
7 @feszitofamarvellel;
8 $i=0;
9 $j=0;
10
11 for( <IN1> ){
12   @fields = split(/\s+/);
13   $index=$fields [1];
14   $marvel{$index}="$fields [2]\t$fields [3]\t$fields [4]\t$fields [5]\t$
     fields [6]";
15   $i=$i+1;
16   print "MARVEL csucok beolvasva: $i.\n";
17 }
18
19 for( <IN2> ){
20   @fields = split(/\s+/);
21   $j=$j+1;
22   $node1=$fields [2];
23   $node2=$fields [3];
24   $elsőcsucsbennevan=0;
25   $másodikcsucsbennevan=0;
26   if( exists $marvel{$node1} ) {
27     $elsőcsucsbennevan=1;
28   }
29   if( exists $marvel{$node2} ) {
30     $másodikcsucsbennevan=2;
31   }
32   $szum=$elsőcsucsbennevan+$másodikcsucsbennevan;
33   if( $szum==0 ){
34     push( @feszitofamarvellel , "$fields [0]\t$fields [1]\t$fields [2]\
     \t$fields [3]");
35     }elseif( $szum==1 ){
36     #push( @feszitofamarvellel , "$fields [0]\t$fields [1]\t$fields
     [2]\t$fields [3]\t$marvel{$node1}");
37     }elseif( $szum==2 ){
38     #push( @feszitofamarvellel , "$fields [0]\t$fields [1]\t$fields
     [2]\t$fields [3]\t\t\t\t\t\t$marvel{$node2}");
39     }elseif( $szum==3 ){
40     #push( @feszitofamarvellel , "$fields [0]\t$fields [1]\t$fields
     [2]\t$fields [3]\t$marvel{$node1}\t$marvel{$node2}");
41     }
42   print "Feszitofa elek rendbeteve: $j. Szum=$szum.\n";
43 }
44
45 for $i ( 0 .. $#feszitofamarvellel ){
46   print OUT "$feszitofamarvellel [$i]\n";
47 }
48
49 close IN1;
50 close IN2;

```



```
51 close OUT;
```

**feszitofaelek.pl** – A Kruskal-algoritmus output élindeks-listáját alakítja át [16]-beli stílusú adatsorrá.

```
1 @ARGV < 3 and die "USAGE: perl.exe feszitofaelek.pl [infile1] [
  infile2] [outfile]";
2 open(IN1,$ARGV[0]) or die "Unable to open infile 1...";
3 open(IN2,$ARGV[1]) or die "Unable to open infile 2...";
4 open(OUT,">$ARGV[2]") or die "Unable to open outfile...";
5
6 @indexek;
7 @elek;
8 $i=0;
9 $x=0;
10
11 for( <IN1> ){
12   @fields = split(/\s+/);
13   $indexek[$i]=$fields[0];
14   $i=$i+1;
15   print "$i\n";
16 }
17
18 for( <IN2> ){
19   $x=$x+1;
20   print "$x\n";
21
22   @fields = split(/\s+/);
23
24   for $j ( 0 .. $#indexek ){
25     if ($fields[0] eq $indexek[$j]){
26       push(@elek, "$fields[0]\t$fields[1]\t$fields[2]\t$fields[3]");
27     }
28   }
29 }
30
31 for $k ( 0 .. $#elek ){
32   print OUT "$elek[$k]\n";
33 }
34
35 close IN1;
36 close IN2;
37 close OUT;
```

# Irodalomjegyzék

- [1] T. Furtenbacher, P. Árendás, G. Mellau, A. G. Császár, *Simple molecules as complex systems*, kézirat, 2013.
- [2] P. Árendás, *Gráfelméleti módszerek a molekulaszpektroszkópiában*, Országos Tudományos Diákköri Konferencia, Eger, 2013.
- [3] L. Euler, *The seven bridges of Königsberg*, Comment. Acad. Scient. Petropol. 8 (1741), pp. 128-140.
- [4] A. J. Hinde, *The Development of Modern Chemistry*, Dover, New York, 1984.
- [5] J. J. Sylvester, *Chemistry and algebra*, Nature 17 (1878), p. 284.
- [6] A. G. Császár, T. Furtenbacher, *Spectroscopic networks*, J. Mol. Spectrosc. 266 (2011), pp. 99-103.
- [7] T. Furtenbacher, A. G. Császár, *The role of intensities in determining characteristics of spectroscopic networks*, J. Mol. Struct. 1009 (Boris Galabov Special Issue) (2012), pp. 123-129.
- [8] T. Furtenbacher, *Adatbázis alapú molekulaszpektroszkópia*, doktori értekezés, ELTE, 2008.
- [9] R. J. Barber, J. Tennyson, G. J. Harris, R. N. Tolchenov, *A high-accuracy computed water line list*, Mon. Not. R. Astron. Soc. 368 (2006), pp. 1087-1094.

- 
- [10] T. Furtenbacher, A. G. Császár, J. Tennyson, *MARVEL: Measured Active Rotational-Vibrational Energy Levels*, J. Mol. Spectry. 245 (2007), pp. 115-125.
- [11] A. Clauset, C. R. Shalizi, M. E. J. Newman, *Power-Law Distributions in Empirical Data*, SIAM Rev. 51(4) (2009), pp. 661-703.
- [12] T. Nepusz, G. Csardi, Implementation in R for fitting a power-law distribution, 2012.
- [13] A.-L. Barabási, R. Albert, *Emergence of scaling in random networks*, Science 286 (1999), pp. 509-512.
- [14] R. Cohen, S. Havlin, *Scale-free networks are ultrasmall*, Phys. Rev. Lett. 90(5) (2003), 058701.
- [15] D. J. Watts, S. H. Strogatz, *Collective dynamics of small-world networks*, Nature 393 (1998), pp. 440-442.
- [16] L. Li, D. Alderson, J. C. Doyle, W. Willinger, *Mathematics and the Internet: A Source of Enormous Confusion and Great Potential*, Intern. Math. 2 (2005), pp. 431-523.
- [17] J. B. Kruskal, *On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem*, Proc. Amer. Math. 7(1) (1956), pp. 48-50.
- [18] LEMON Library for Efficient Modeling and Optimization in Networks, ELTE, 2003.