Lecture #4: Spectroscopic networks and active databases Attila G. Császár

Laboratory of Molecular Structure and Dynamics Institute of Chemistry Loránd Eötvös University Budapest, Hungary csaszar@chem.elte.hu http://theop11.chem.elte.hu



Winter School in Theoretical Chemistry Helsinki, Finland, December 17-20, 2012





Introduction

Elements of graph and network theory of simple and complex systems

MARVEL: a Hamiltonian-free, graph-theoretical approach to molecular spectroscopy

Applications of network theory to high-resolution molecular spectroscopy

Summary and outlook

Origin of graph theory (Euler, 1736) Koringsberga

SOLITIO PROSILIMATIS SOLVTIO PROBLEMATIC GEOMETRIAM SITVS Averons: Louis Eulero,

in firm vocanic. Iffe pars ab int ando, finaspot proprietatilars erners tosator; in quo segutio neuto ad quan dano, neque esculo quantizarum vico indi arcen problemata ad hane finas Ge

mont, et quili methodo in lis refoluente neg ficis ell defininam. Quantobrem emunia cantasfam mentio effet fichs, in riam pertinere videbuter, ur ita era m, vt nope determinationem quantitation reret, neque folationem calculi quantitation ope terret, id ad geometriam fitas referre hand d pricterin and to can fountione folus fitus in to prefertin quot a tenha vero sollina prorina fa Methodam ergo meam quam ad huins general fre

100.00 10

130 64

240.00

AD GEOMETRIAM SITE'S PERTINENTIS, 119

tolande interé, taspon facines Gromeriae espirare confirm.

met dateret interne, enquem freenene enterne intergreene content.
S. Foldene anne ker, end målt atte nor reine der pehitekene, engene engenenen internet, engene engene anne fore ender engene en

mills hac methodo, in aliam inquilai, quie plus non Ton. VIII.



Chemical graph theory



Figure 1. Examples of the first chemical graphs used by Cullen and Black in 1758 to represent the interactions of chemical substances. The supposed forces between pairs of substances are indicated either in terms of numbers or symbols.

Chemical graph theory



Figure 9. Reproductions of the graphical formulas employed by Crum Brown (1864) to represent the topological positions of atoms in a variety of molecules. 2000: New direction in graph theory Networks within complex systems (building blocks + interactions = nodes + links)

- Society (human interactions)
- Culture (language)
- Economy (world economy)
- Internet
- Telephones
- Power stations and electric networks
- Brain
- Cells (biological networks, immune system, transport networks, muscle network)

2000: New directions in chemistry Networks within complex chemical systems

- appearance of molecular linelists, containing hundreds of millions of data per isotopologue of a molecule determined via sophisticated quantum chemical computations
- extensive databases assembled for modeling work (*e.g.*, greenhouse effect, radiation transfer, atmospheric modeling) in thermochemistry, reaction kinetics, and high-resolution molecular spectroscopy
- use of active databases and other IT techniques

Spectroscopic databases









High-resolution Transmission Molecular Absorption Database (Harvard-Smithonian) Atmospheric Radiation Analysis (GEISA) Quantitative Infrared Database (NIST) Cologne Database for Molecular Spectroscopy (CDMS)

Molecules as complex systems?

- Size of most molecules studied by quantum chemistry: 2-1000 nuclei (for us 5-6 nuclei...)
- Perhaps 10 times this many electrons
- Few hundred thousand different energy level per electronic state and isotopolugue
- Few billion *allowed* transitions for each case
- Different experimental techniques: transitions between greatly different energy level sets

Network theory

- graphs (vertices, edges, distributions)
- scale dependence
- scale free
- "small world"(Karinthy(1929))
- embedding (principal and subnets)
- weak and strong "forces"
- network stability (attack/error tolerance)
- overall network characteristics (random, scale free, star, disintegrated)
- stochasticity and preferential attachment

Network (graph) theory

- graph G: ordered pair, G = (V,E)
- *degree*: no. of edges that connect to a vertex
- *loop*: edge connecting to the same vertex
- *simple* graph: contains neither loops nor multiple edges
- *undirected* graph: no direction of edges
- *connected* graph: there is a path between any pair of vertices
- *root*: a distinguished vertex
- *cycle*: returning path going through at least three edges
- *tree (forest*): connected (disconnected) graph without cycles
- *hub*: a vertex with a large number of edges

Network (graph) theory





(a) Random network (Erdős-Rényi)

(b) Scale-free network(Barabási)



Questions to be answered

- What kind of graphs (networks) would quantum mechanics build (*e.g.*, dense or sparse, stochastic or deterministic, with one or with several components)?
- What is the degree distribution (Poisson or scale free)?
- Is there a significant difference between measured and computed spectroscopic networks?
- Can a graph theoretical approach help spectroscopy and especially improving spectroscopic databases?
- Can one find similarities between spectroscopic (quantum mechanical) and human networks and their behavior (are social networks unique in their characteristics or does ,,nature" produce similar networks)?
- Etc.

Spectroscopic networks (SN)

- energy levels: vertices (nodes, with given, theoretical or experimental, uncertainties)
- allowed transitions: edges (links)
- transition intensities: weights
- SNs are weighted, undirected graphs
- first-principles SNs are simple graphs
- experimental SNs contain a large number of components (few roots, several floating networks (FSNs) and orphans)

A. G. Császár, T. Furtenbacher, J. Mol. Spectrosc. 2011, 266, 99-103.

Measured SN of para-H₂D⁺



Active databases

Definition:

An active database is a database with an event monitoring scheme for detecting when certain data is INSERTED, DELETED, UPDATED, or SELECTED, and (automatically) executing the actions in response when certain events happen and particular conditions are met.

Inverse problems in physical chemistry

Measured quantities ↓ ("fitting" procedure) ↓ parameters of the physical model(s)

Spectroscopy

from measured transitions to "experimental" energy levels

Thermochemistry

from measured enthalpies of reaction to **enthalpies of formation**

Kinetics

from measured *T*-dependent reaction rates to **Arrhenius parameters**

Spectroscopic databases









High-resolution Transmission Molecular Absorption Database (Harvard-Smithonian) Atmospheric Radiation Analysis (GEISA)

Quantitative Infrared Database (NIST) Cologne Database for Molecular Spectroscopy (CDMS)

IUPAC database of water isotopologues: JQSRT 110 (2009) 573, 111 (2010) 2160 and 113 (2012) in press.

Rovibrational energy levels / transitionsObservedComputed

- $H_2^{16}O: 18\ 152/184\ 300$ $H_2O: 165\ 000/700\ 000\ 000$
- H₂¹⁷O: 2687/ 4839
- H₂¹⁸O: 8 614/ 29 364
- HD¹⁶O: 8 818/ 36 690
- HD¹⁷O: 162/ 485
- HD¹⁸O: 1 864/ 8 729



Spectroscopic networks are scale free





MARVEL prediction

Number of energy levels and transitions as a function of -log(intensity)



Minimal weight spanning tree algorithm









Minimal weight spanning tree algorithm







minimal weight spanning tree

Spanning tree for complex network



Attack tolerance of SNs



MARVEL: An inverse, Hamiltonian-free approach to highly accurate rovibrational energy levels

> Measured Active Rotational-Vibrational Energy Levels

T. Furtenbacher, A. G. Császár, J. Tennyson, J. Mol. Spectrosc. 2007, 245, 115T. Furtenbacher, A. G. Császár, J. Quant. Spectr. Rad. Transfer 2012, in press.



Solve for **E** (in a least-squares sense, taking into account the experimental uncertainties of the v_{ij}) to obtain experimentally derived term values E_i , E_j ,

New Journal of Physics

The open-access journal for physics

New Journal of Physics 12 (2010) 103006 (10pp)

Frequency metrology in the near-infrared spectrum of H₂¹⁷O and H₂¹⁸O molecules: testing a new inversion method for retrieval of energy levels

A Gambetta¹, E Fasci², A Castrillo², M Marangoni¹, G Galzerano³, G Casa², P Laporta³ and L Gianfrani^{2,4}

vibrational state. The measurement of the sub-Doppler line-center frequencies, also extended to the $H_2^{17}O$ spectrum, has been performed with an overall uncertainty of ~ 30 kHz, i.e. about three orders of magnitude lower than the HITRAN data set. The retrieved energy separations agree, by less than 80 kHz, with recent findings provided by the so-called MARVEL procedure for spectral data inversion, thus yielding a very stringent test of its accuracy.

Recalibration of 83Guelachv





IUPAC water database I. H₂¹⁷O and H₂¹⁸O

	H ₂ ¹⁷ O	$H_2^{18}O$
No. of transitions collected	8 463	25 367
Maximum J	17	20
Highest VBO (cm ⁻¹)	16 876	16 855
No. of energy levels	2 687	4 849
No. of sources	33	48

J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, M. R. Carleer, A. G. Császár, R. R. Gamache, J. T. Hodges, A. Jenouvrier, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, R. A. Toth, A. C. Vandaele, N. F. Zobov, L. Daumont, A. Z. Fazliev, T. Furtenbacher, I. F. Gordon, S. N. Mikhailenko, and S. V. Shirin, IUPAC Critical Evaluation of the Rotational-Vibrational Spectra of Water Vapor. Part I. Energy Levels and Transition Wavenumbers for H₂¹⁷O and H₂¹⁸O, *J. Quant. Spectr. Rad. Transfer* **2009**, *110*, 573-596.

MARVEL prediction



IUPAC water database II. HD¹⁶O, HD¹⁷O and HD¹⁸O

	HD ¹⁶ O	HD ¹⁷ O	HD ¹⁸ O
No. of transitions collected	54 740	485	8 728
Maximum J	30	11	18
Highest VBO (cm ⁻¹)	22 625	1 399	9 930
No. of energy levels	8 819	162	9 627
No. of sources	74	3	1 864

J. Tennyson, P. F. Bernath, L. R. Brown, A. Campargue, A. G. Császár, L. Daumont, R. R. Gamache, J. T. Hodges, O. V. Naumenko, O. L. Polyansky, L. S. Rothman, R. A. Toth, A. C. Vandaele, N. F. Zobov, S. Fally, A. Z. Fazliev, T. Furtenbacher, I. F. Gordon, S.-M. Hu, S. N. Mikhailenko, and B. Voronin, IUPAC Critical Evaluation of the Rotational-Vibrational Spectra of Water Vapor. Part II. Energy Levels and Transition Wavenumbers for HD¹⁶O, HD¹⁷O, and HD¹⁸O, *J. Quant. Spectr. Rad. Transfer* **2010**, *111*, 2160-2184.

MARVEL prediction



Utility of a network view of highresolution molecular spectroscopy

- "weak interactions" (in fact, weak transitions) help to stabilize SNs
- comparison of experimental and theoretical hubs helps to design new experiments: maximum gain with minimal (experimental) effort
- identification of "weakest links" (in fact nodes, whether they are hubs or not)
- which would be the most useful experiments
- special role of temperature-dependent emission spectra

Conclusions

- (1) Network-based MARVEL analyses provide excellent opportunities to understand high-resolution experimental spectra of molecules and turn scattered information into focused knowledge.
- (2) Spectroscopic networks (SN) are weighted, undirected graphs, whereby energy levels are the nodes, allowed transitions are the links, and weights are provided by intensities.
- (3) Graph theory helps to advance our understanding of high-resolution molecular spectra and design new experiments.