Simple molecules as complex systems

Tibor Furtenbacher1,2, Péter Árendás2,3, Georg Mellau4 & Attila G. Császár1,2

Received 5 December 2013
Accepted 27 March 2014
Published 11 April 2014

Correspondence and requests for materials should be addressed to A.G.C. (csaszar@chem.elte.hu)

High-resolution molecular spectroscopy is one of the high-end analytical tools which can be used to obtain detailed chemical information about complex natural systems. These systems include the earth’s atmosphere, where spectroscopy helps to understand the greenhouse effect, and astronomical bodies of our universe, where spectroscopy helps, among other things, to answer principal questions concerning life on earth. The extensive spectroscopic data required by related modelling efforts have been consolidated into information systems1–10. The data deposited in these information systems traditionally come from a large number of high-resolution experimental investigations. Experiments are usually done by different groups employing different techniques in different regions of the spectrum, resulting in a broad range of data accuracy. The relative accuracy of transition frequencies detected in the lab ranges from 10−3 to 10−8, while for transition intensities it is only 10−2. As to theory, in the fourth age of quantum chemistry12 it is possible to determine accurate high-resolution spectroscopic data and spectra11,12. To satisfy the demand of modellers, for a number of small molecules nearly complete first-principles linelists have been computed13. These lists contain from thousands to millions of entries in the form of rotational-vibrational-electronic energies and transitions and their most important characteristics (e.g., quantum numbers, symmetries, and intensities).

Although high-resolution spectroscopic experiments yield highly accurate data, at the same time these data are highly incomplete. For example, the 5 000 experimental eigenenergies reported by Mellau14–16 are complete up to 7 000 cm−1 above the HCN ground state, yet they cover only 98 vibrational states. The 25 000 rovibrational states determined in these high-resolution infrared emission studies correspond only to 15% of the vibrational states up to isomerization. When compared with experimental data, ab initio linelists show the following important characteristics: while the relative accuracy of the ab initio energy levels is 10 to 10 000 times worse than that of typical experimental data, most of the transition intensities have accuracies similar to experimental data. The striking disparity between the accuracy and the number of first-principles computed and experimentally measured energy levels and transitions and the fact that in many cases ab initio intensities may directly be used for high-resolution analyses leads to the conclusion that for the foreseeable future one should consider the combination of experimental and ab initio information to simplify the assignment of measured spectra, and how to build the most dependable information systems containing line-by-line spectroscopic data.

We believe that to obtain the best answers to these questions one should consider the energy levels and the spectroscopic transitions of a molecule from the point of view of graph theory. Thus, earlier we introduced the...
concept of spectroscopic networks (SN)\textsuperscript{20–24}, where quantized energy levels are the nodes (vertices) and allowed transitions among the levels are the links (edges) of a graph (see Fig. 1). SNs are considered to be an intrinsic property of molecular systems, though characteristics of SNs can be slightly different based on how we actually probe these systems experimentally (e.g., in absorption or in emission). SNs provide a convenient representation of the experimental and theoretical data and ways for their most advantageous unification, as well.

In this paper we extend the network-theoretical analysis of SNs and, furthermore, develop novel tools for high-resolution spectroscopy research based on the concept of SNs. We use $\text{H}_2^{16}\text{O}$ as the model system of our present investigation. The SN of the $\text{H}_2^{16}\text{O}$ molecule is chosen for several reasons. Water is the most abundant polyatomic molecule in the Universe. It is present in many different environments and at many different temperatures. Detailed characterization of the spectroscopic properties of this triatomic molecule is needed to understand and predict the greenhouse effect on earth and its spectroscopy is of high astrophysical and astrochemical relevance. Furthermore, $\text{H}_2^{16}\text{O}$ was the subject of a large number of experimental high-resolution spectroscopic studies validated recently\textsuperscript{25}. This experimental dataset of $\text{H}_2^{16}\text{O}$, one of the spectroscopically most thoroughly studied molecules, contains 14 319 nodes (energy levels) and 97 868 unique links (transitions)\textsuperscript{25}. A high-quality first-principles linelist\textsuperscript{26}, including energy levels, assignments, transitions, and Einstein $A$ coefficients, is also available for $\text{H}_2^{16}\text{O}$. This computed, so-called BT2 linelist contains altogether 221 097 nodes and 505 806 255 links. Based on the number of nodes and links and the underlying structure one can conclude that even this simple triatomic molecule corresponds to a very complex system if the allowed one-photon transitions among its quantized energy levels are considered.

**Spectroscopic networks**

A graph $G$, corresponding to an SN of a molecule, say $\text{H}_2^{16}\text{O}$, is an ordered pair, $G = (L,T)$, where $L$ is the set of energy levels (vertices) and $T$ is a set of transitions (edges), the edges being 2-element subsets of $L$ (see Fig. 1). The number of transitions that emanate from an energy level is called the degree of the level. SNs do not contain loops and since different experiments may measure the same transitions, SNs corresponding to experiments are in fact multigraphs. First-principles SNs are, on the other hand, simple graphs. SNs contain a large number of cycles of widely differing size. In SNs non-negative transition intensities, different for different experimental techniques, are assigned to edges as weights. In summary, SNs are large, finite, weighted, and rooted graphs.

Construction of a first-principles SN goes through the following steps: (1) take all (available) energy levels for the given molecule as nodes; (2) use the quantum chemical selection rules appropriate for the molecule and the experiment to link the nodes; and (3) add the intensities as weights to the links based on the type of experiment and the chosen temperature. The number of links in the graph built is naturally much smaller than all the possible links between the nodes. Consequently, the corresponding adjacency matrix is extremely sparse. In the particular case of $\text{H}_2^{16}\text{O}$, consideration of nuclear spins results in two distinct connection schemes. In the language of graph theory these are components of a network. The two principal components (PC) correspond to the two nuclear spin isomers (usually called “ortho” and “para”) of $\text{H}_2^{16}\text{O}$ and both have unique roots. Selection rules cause the two PCs of the SN of $\text{H}_2^{16}\text{O}$ to be bipartite graphs. This interesting fact explains why only even-numbered cycles exist in the SN of $\text{H}_2^{16}\text{O}$ and of molecules of a similar nature\textsuperscript{27}.

Measurements map only a very limited part of an SN and yield a graph called $A_{m}$. The intensity of the transitions is responsible for the incompleteness of $A_{m}$ as below a certain intensity it is impossible to detect a transition in a given type of experiment. Using the intensity as a cut-off parameter, a series of model networks can be constructed from the complete SN built upon the BT2 linelist\textsuperscript{26}. We used the following cut-off parameters to construct model networks for the examination of the evolution of one-photon absorption SNs: $10^{-20}$, $10^{-22}$, $10^{-24}$, $10^{-26}$, and $10^{-28}$ cm molecule$^{-1}$ (see Fig. 1 for a visual

| Table 1 | General properties of the spectroscopic networks considered for $\text{H}_2^{16}\text{O}$ |
|-------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Quantity    | $A_{20}$        | $A_{22}$        | $A_{24}$        | $A_{26}$        | $A_{28}$        | $A_{m}$         |
| intensity cut-off | $10^{-20}$   | $10^{-22}$   | $10^{-24}$   | $10^{-26}$   | $10^{-28}$   | measured         |
| number of nodes    | 547            | 1 952         | 5 815          | 15 603         | 44 843          | 18 572          |
| number of links     | 1 238          | 7 288         | 31 283         | 115 886        | 397 147         | 98 927          |
| $S(G)$    | 0.617          | 0.429          | 0.284          | 0.182          | 0.091           | 0.130           |
| $r(G)$    | −0.199         | −0.356         | −0.420         | −0.443         | −0.469          | −0.202          |
| diameter, $d$ | 19             | 23             | 28             | 31             | 34             | 44              |
| average path length | 5.9           | 6.2           | 6.5           | 6.8           | 7.1           | 10.7            |
representation of three of the first-principles model SNs and Table 1
for details about these SNs, including the number of nodes and links
they possess). To emphasize that these SNs belong to absorption, the
corresponding graphs are called $A_{20} - A_{28}$.

Floating components (FC), those which do not connect to the
roots of PCs, arise frequently in measurements. Since no known
transitions exist between the two PCs of the rovibrational SN of
$\text{H}_2^{16}\text{O}$, the absolute energy of the higher-energy root, set to a relative
energy of zero by definition, can be determined only from an outside
source, hindering the high-accuracy absolute determination of all
measured energy levels. Artifical transition energies connecting
roots of SNs may be called “magic numbers”. The traditional route
to obtain them is provided by highly accurate model Hamiltonians. A
network-theoretical possibility is to take advantage of omnipresent
degeneracies of certain higher-energy rovibrational levels in the two
PCs, which can be identified straightforwardly by fourth-age $v^3$
varia-
tional nuclear-motion computations. These degeneracies are able to
connect the distinct components via zero-energy artificial transitions.
This was done in Ref. 25 for $\text{H}_2^{16}\text{O}$ in Ref. 28 for $\text{D}_2^{16}\text{O}$
with the comforting result that the network-theoretical and model
Hamiltonian approaches yield the same magic number.

Degree distributions
For many observables there is a typical mean value they cluster
around. As to SNs, where the number of experimentally measured
links is about an order of magnitude larger than the number of
nodes $25,27,29-32$, the question is whether there is a mean value for the
number of transitions that an “average” energy level has. To answer
this question one needs to investigate the distribution of the links
among the nodes.

Fig. 2 depicts the size–frequency $[\log k - \log P(k)]$ plots for the $A_{\text{m}}$
and $A_{28}$ SNs of $\text{H}_2^{16}\text{O}$. One can find a very broad distribution and,
apart from the very low and very high $k$ part, a reasonably linear
relationship in both cases. As detailed in the Methods section, an
elaborate search has been performed to estimate the form of the
underlying discrete degree-distribution functions of these and the
other model SNs. The search included a power-law form of $P(k) \propto k^{-\gamma}$, where $\gamma$ is the scaling index, as well as exponential and
log-normal forms. The analyses indicate a definitely heavy-tailed and,
heavy-tailed distribution is one of the most important overall char-
acteristics of SNs and it seems to be generally valid for the PCs of
SNs$^{23}$.

Whether the degree distribution follows a power law or it is just
simply top heavy, the degree distribution functions obtained suggest
that SNs are characterized by hubs, i.e., a small number of nodes with
a large number of connections. As expected, the most important hubs
in a room-temperature absorption spectrum are on the ground
vibrational state, $(0 0 0)$, where $(v_1, v_2, v_3)$ are approximate vibrational
quantum numbers corresponding to symmetric stretch, bend, and
antisymmetric stretch, respectively. For $A_{\text{m}}$ the hubs are as follows:
$J_{K_{\text{JC}}} = 6_{14}, 5_{23},$ and $4_{23}$, with 458, 455, and 447 links, respectively$^{23}$,
where $J_{K_{\text{JC}}}$ is the standard rigid-rotor-type quantum number nota-
tion applied for asymmetric top molecules, such as $\text{H}_2^{16}\text{O}$. In the $A_{28}$
SN the energy levels with the largest number of transitions are
$6_{14}(1487), 5_{23}(1433)$, and $6_{23}(1431)$, where the number of links is
given in parentheses. Remarkably, the two largest hubs coincide,
proving how extensive the experimental investigations are for
$\text{H}_2^{16}\text{O}$. Note that the most important hub for $\text{HD}^{16}\text{O}$ in absorption
is also the $(0 0 0)6_{14}$ level$^{23}$.

To investigate the hubs of SNs further we determined an SN cor-
responding to emission created from the first-principles BT2 linelist
with an intensity cut-off of $10^{-27}$ cm molecule$^{-1}$ at 1650 K, which
could be called $E_{20}$. In emission the hubs with the largest number
of connections belong to different vibrational states, they are the
$(0 2 0)9_{63}, (0 0 1)6_{33},$ and $(0 1)10_{38}$ levels with 102, 101, and 100
links, respectively. The most important hubs in absorption appear to
be important hubs in emission but the reverse is obviously not true.

Detailed comparison of the connectivity of measured and first-
principles hubs helps to determine the “weakest”, least well deter-
determined hubs within $A_{\text{m}}$. This allows the design of new experiments

![Figure 2](https://example.com/image2.png)

**Figure 2** Distribution of links among nodes given as log-log size–frequency $[\log k - \log P(k)]$ plots for the measured ($A_{\text{m}}$, left panel) and a first-principles ($A_{28}$, right panel) spectroscopic network of one-photon absorption transitions for $\text{H}_2^{16}\text{O}$.

| network | scaling index | $k_{\text{min}}$ | $p(\text{KS})$ |
|---------|---------------|----------------|--------------|
| $A_{22}$ | 2.11 | 4 | 0.1060 |
| $A_{24}$ | 2.13 | 8 | 0.1867 |
| $A_{25}$ | 2.15 | 10 | 0.2853 |
| $A_{26}$ | 2.16 | 14 | 0.0460 |
| $A_{28}$ | 2.10 | 17 | 2.56e-09 |
| $A_{30}$ | 2.47 | 6 | 2.46e-17 |
| $A_{40}$ | 2.83 | 54 | -0 |

Table 2 | Parameters for the best power-law models fitted to the SNs of $\text{H}_2^{16}\text{O}$

For many observables there is a typical mean value they cluster
around. As to SNs, where the number of experimentally measured
links is about an order of magnitude larger than the number of
nodes $25,27,29-32$, the question is whether there is a mean value for the
number of transitions that an “average” energy level has. To answer
this question one needs to investigate the distribution of the links
among the nodes.

Fig. 2 depicts the size–frequency $[\log k - \log P(k)]$ plots for the $A_{\text{m}}$
and $A_{28}$ SNs of $\text{H}_2^{16}\text{O}$. One can find a very broad distribution and,
apart from the very low and very high $k$ part, a reasonably linear
relationship in both cases. As detailed in the Methods section, an
elaborate search has been performed to estimate the form of the
underlying discrete degree-distribution functions of these and the
other model SNs. The search included a power-law form of $P(k) \propto k^{-\gamma}$, where $\gamma$ is the scaling index, as well as exponential and
log-normal forms. The analyses indicate a definitely heavy-tailed and,
heavy-tailed distribution is one of the most important overall char-
acteristics of SNs and it seems to be generally valid for the PCs of
SNs$^{23}$.

Whether the degree distribution follows a power law or it is just
simply top heavy, the degree distribution functions obtained suggest
that SNs are characterized by hubs, i.e., a small number of nodes with
a large number of connections. As expected, the most important hubs
in a room-temperature absorption spectrum are on the ground
vibrational state, $(0 0 0)$, where $(v_1, v_2, v_3)$ are approximate vibrational
quantum numbers corresponding to symmetric stretch, bend, and
antisymmetric stretch, respectively. For $A_{\text{m}}$ the hubs are as follows:
$J_{K_{\text{JC}}} = 6_{14}, 5_{23},$ and $4_{23}$, with 458, 455, and 447 links, respectively$^{23}$,
where $J_{K_{\text{JC}}}$ is the standard rigid-rotor-type quantum number nota-
tion applied for asymmetric top molecules, such as $\text{H}_2^{16}\text{O}$. In the $A_{28}$
SN the energy levels with the largest number of transitions are
$6_{14}(1487), 5_{23}(1433)$, and $6_{23}(1431)$, where the number of links is
given in parentheses. Remarkably, the two largest hubs coincide,
proving how extensive the experimental investigations are for
$\text{H}_2^{16}\text{O}$. Note that the most important hub for $\text{HD}^{16}\text{O}$ in absorption
is also the $(0 0 0)6_{14}$ level$^{23}$.

To investigate the hubs of SNs further we determined an SN cor-
responding to emission created from the first-principles BT2 linelist
with an intensity cut-off of $10^{-27}$ cm molecule$^{-1}$ at 1650 K, which
could be called $E_{20}$. In emission the hubs with the largest number
of connections belong to different vibrational states, they are the
$(0 2 0)9_{63}, (0 0 1)6_{33},$ and $(0 1)10_{38}$ levels with 102, 101, and 100
links, respectively. The most important hubs in absorption appear to
be important hubs in emission but the reverse is obviously not true.

Detailed comparison of the connectivity of measured and first-
principles hubs helps to determine the “weakest”, least well deter-
determined hubs within $A_{\text{m}}$. This allows the design of new experiments

| network | scaling index | $k_{\text{min}}$ | $p(\text{KS})$ |
|---------|---------------|----------------|--------------|
| $A_{22}$ | 2.11 | 4 | 0.1060 |
| $A_{24}$ | 2.13 | 8 | 0.1867 |
| $A_{25}$ | 2.15 | 10 | 0.2853 |
| $A_{26}$ | 2.16 | 14 | 0.0460 |
| $A_{28}$ | 2.10 | 17 | 2.56e-09 |
| $A_{30}$ | 2.47 | 6 | 2.46e-17 |
| $A_{40}$ | 2.83 | 54 | -0 |
which help to determine a more accurate and robust experimental description of the SN with a minimum amount of effort.

One can also ask the question whether the hubs with the largest number of links take part in the most intense transitions. The answer is a clear no. The 634, 523, and 423 pure rotational energy levels take part in the 16th, 18th, and 13th most intense rovibrational absorption transitions, respectively. Vice versa, the two energy levels taking part in the most intense transition are only 69th and 89th in the list of hubs based on the number of connections.

**Complexity measures**

Complexity of a graph G can be assessed by several metrics[^35-39]. Three of them, C(G), S(G), and r(G) have been investigated in this study (see Table 1).

The local clustering coefficient, C(G)[^39], quantifies how close local graphs are to being a complete graph. This metric cannot be used for the bipartite PCs of the model SNs of H$_2$O as bipartite graphs do not contain odd-numbered cycles such as triangles.

A second metric is the structural metric (s-metric) with the corresponding S(G) value[^39] (see the Methods section for details). The S(G) values of the different networks investigated are collected in Table 1.

As shown by Newman[^39], social networks seem to show “assortative mixing”, i.e., their high-degree vertices preferentially attach to other high-degree vertices. On the contrary, technological and biological networks tend to show “disassortative mixing”, i.e., their high-degree vertices attach to low-degree ones. A graph assortativity measure is the Pearson correlation coefficient, r(G)[^39]. The r(G) values for the first-principles and measured SNs investigated are given in Table 1. For details see the Methods section.

Ordinarily[^36,37], one expects a large value of S(G) to be associated with a large positive r(G) value. As seen in Table 1, the S(G) and r(G) values decrease when the intensity cut-off parameter of the first-principles SNs is decreased. This unusual behavior can be rationalized once the evolution of the underlying SNs is understood. If we examine the smallest model SN, A$_{20}$ (see the leftmost panel of Fig. 1 for its visual representation), we find that it contains only two components (it would not be surprising if the energy levels involved in the largest intensity lines would produce several components but this is not the case here). In these two components, containing the most intense transitions, the likelihood of connections among high-degree nodes (hubs) is high; in other words, their eigenvalue centrality[^39] is high. This is the reason why the S(G) value is relatively large, while r(G) is close to zero. While the r(G) value of A$_{20}$ is negative, the corresponding large S(G) value indicates that this graph is disassortative with hubs showing an assortative behavior. This means that in A$_{20}$ hubs do like to connect to each other but each hub has many connections to low-degree nodes. Investigating the other SNs we can make another interesting and important observation: the nodes characterized as hubs do not change with the cut-off parameter. Of the first 100 hubs of the model A$_{20}$ and A$_{28}$, SNs 98 are common, meaning that the hubs already appear in the smallest SN and hubs remain hubs when the SN is enlarged. When increasing the size of the SN by decreasing the intensity cut-off parameter, the number of low-degree nodes increases substantially and the ratio of the connections among high-degree nodes to that of high-low connections decreases. This is the reason why the S(G) values show a decreasing tendency when going from A$_{20}$ to A$_{28}$ and the SNs become increasingly disassortative. Note also how nicely the experimental SN, A$_{458}$, fits this picture, supporting these findings about SNs.

**Network vulnerability**

A spectroscopic network becomes larger either via new measurements (for an experimental SN) or by a decrease in the intensity cut-off (for a first-principles SN). In either case, the number of transitions increases substantially faster than the number of energy levels, in complete accord with the degree distribution observed. The number of cycles within the network also increases drastically. As a result, SNs appear to be extremely robust.

Robustness of SNs can be ascertained by random removal of nodes[^41]. In scale-free networks removal of nodes leads to an increase in the diameter[^41]. In SNs, after random removal of 10 to 90% of the nodes, d(G) reflects how the graph fragments and thus provides useful characteristics about SNs. The original diameter of the largest first-principles graph investigated, A$_{280}$, is 34 (Table 1), and this value does not change until we randomly remove some 95% of the nodes. Then the diameter suddenly drops to 22. The observed robustness of the SN of H$_2$O can be explained by the nature of the selection rules leading to a bipartite graph and the presence of an assortative core of interconnected hubs. To prove the latter we note that in A$_{280}$ the first 448 hubs, 1% of the nodes, own almost 40% of the links. On one hand, the probability of random removal of hubs is small, on the other hand, if we remove such hubs, another hub “takes over” in the graph, as hubs are ‘well connected’. The situation is quite different when we attack the graph, i.e., we remove the high-degree nodes systematically. If we delete the first 200 hubs, 0.45% of the nodes, which have 20.45% of the links, the diameter reduces to 18. The extreme error tolerance is another characteristic property of this and this property is somewhat similar to that observed in other complex networks.

**Data reduction via SNs**

Since high-resolution spectroscopic measurements yield an extreme amount of information, the reduction of the data to manageable size is a basic challenge for the theory of spectroscopy. The standard solution is to use model Hamiltonians with a small number of parameters and least-squares optimize these parameters to represent all the measured data[^42]. In a way this means that spectroscopic transitions are converted to parameters yielding energy levels. These parameters allow excellent interpolation but they may fail drastically when used to extrapolate beyond the measured range.

SNs offer another data reduction facility via an inversion of transitions to energy levels. For example, the 500 million transitions of the BT2 listelin can be converted back to about 200 thousand energy levels. This feature of SNs has been exploited in the MARVEL (Measured Active Rotational-Vibrational Energy Levels) procedure[^11,12] used, among other applications, to derive the IUPAC spectroscopic database of water isotopologues[^23,28,31,32].

The best way to reduce the information content of SNs is through the use of weighted spanning trees. By using weighted spanning trees[^44], see the Methods section, one can reduce the information contained in the huge number of measured transitions of the
complex $A_m$ network to a relatively small set of energy levels. Each link of $A_m$ has a widely different uncertainty. The network-theoretical view allows to appreciate how cycles, containing a lot of extra information compared to, for example, minimum weight spanning trees, within a component of an SN help to fix the energy levels and tighten their uncertainties.

**Assignment of spectra**

High resolution spectroscopy is also a science (and art) of quantum number assignment of measured lines and levels. The traditional way of analysing high-resolution experimental spectra is the a priori assignment of lines with good and approximate quantum numbers followed by a fitting of the levels via a small number of spectroscopic parameters of a well-designed model Hamiltonian$^{42}$. This type of assignment procedure fails in the case of highly excited rovibrational states and in general when the number of rovibrational transitions exceeds a limit corresponding to an acceptable analysis time. A combined microwave to visible spectrum of any polyatomic molecule is converted to a list of labelled eigenenergies$^{16-18}$ in a high-resolution study.

Hereby we advocate a novel protocol for the assignment of spectra based on SNs: detect the lines in a measured high-resolution spectrum leading to the largest number of new energy levels via an investigation of a suitable first-principles SN and assign the transitions with quantum numbers by mapping the ab initio line-list onto experimental spectra using graph theory. Taking the negative logarithm of the intensity of the transitions as the weight function for the transitions of the SN, the minimum-weight spanning tree displays the transitions with the largest intensities; thus, it readily identifies the most intense and thus the practically most useful spectral features. An illustration of the concept is provided in Fig. 3.

The proposed method based on graph theory allows the automated and fast conversion of very large experimental datasets into complete eigenenergy lists. These lists are the starting points for the development of theoretical models connecting our physical and chemical view on molecules$^{19}$.

Finally, let’s create an artificial spectrum, in order to show the utility of the weighted spanning-tree approach. The complete set of 1 916 H$_2$O rovibrational energy levels up to 7 000 cm$^{-1}$ is known with high-resolution accuracy from a MARVEL study$^{25}$. Based on these energy levels a simulated room temperature absorption spectrum is obtained containing 45 266 allowed transitions with intensities larger than 10$^{-28}$ cm$^2$ molecule$^{-1}$. The corresponding minimum-weight spanning tree contains 1 914 transitions, the minimum number of intense transitions needed to convert the spectrum back to an energy list. This represents a significant, more than 20-fold reduction in the data. In other words, analysis of only 1 914 intense transitions yields the maximum number of energy levels that can be determined from this spectrum. It is worth adding that out of the 45 266 lines 19 482, an order of magnitude more than minimally needed, have indeed been measured and assigned$^{25}$, which is a likely unusually high degree of completeness.

**Conclusions**

Driven by the need of scientific and engineering applications, complex spectroscopic networks, perhaps as part of active databases$^{20-24}$, are expected to become an intrinsic part of the description of the high-resolution spectra of molecules. A good opportunity to advance the field of high-resolution molecular spectroscopy and to turn data into knowledge, as emphasized in the article defining the fourth age of quantum chemistry$^{28}$ and confirmed here, is offered via the joint use of accurate experiments, accurate first-principles computations, and efficient mathematical and numerical algorithms provided by, for example, graph and database theory.
we can define the normalized s-metric used in the text as
\[ S(G) = s / s_{\text{max}}. \] (3)

The graph assortativity, \( r(G) \), is defined by the Pearson coefficient,
\[
r(G) = \frac{\sum_{\langle ij \rangle} d_i d_j - \left( \sum_{\langle ij \rangle} d_i \right)^2}{\sum_{\langle ij \rangle} d_i d_j - \left( \sum_{\langle ij \rangle} d_i \right)^2},
\] (5)
where \( I \) is the number of edges in the graph.

To build a minimum-weight spanning tree from the SNs, we implemented Kruskal's algorithm\(^{44}\). For the weight function, the negative logarithm value of the intensities on the edges were used. Admittedly, a more accurate result can be achieved by multiplying the base intensity values by \(-1\) to obtain a weight function. Nevertheless, the differences are within the same order of magnitude and are negligible for practical considerations; therefore, we believe the weight function employed is adequate.

1. Rothman, L. S. The evolution and impact of the HITRAN molecular spectroscopic database. J. Quant. Spectrosc. Radiat. Transfer 111, 1565–1567 (2010).
2. Rothman, L. S. et al. The HITRAN 2008 molecular spectroscopic database. J. Quant. Spectrosc. Radiat. Transfer 110, 533–572 (2009).
3. Rothman, L. S. et al. HITEMP, the high-temperature molecular spectroscopic database. J. Quant. Spectrosc. Radiat. Transfer 111, 2139–2150 (2010).
4. Jacquinet-Husson, N. et al. The 2003 edition of the GEISA/IASI spectroscopic database. J. Quant. Spectrosc. Radiat. Transfer 95, 429–467 (2005).
5. Landi, E., Young, P. R., Dede, K. P., Del Zanna, G. & Mason, H. E. CHIANTI – An atomic database for emission lines. XIII. Soft X-ray improvements and other changes: Version 7.1 of the database. Astrophys. J. 763, 86 (2013).
6. Müller, H. S. P., Schöder, F., Stutzki, J. & Winnewisser, G. The Cologne database for molecular spectroscopy, CDMS: A useful tool for astronomers and spectroscopists. J. Mol. Struct. 742, 215–227 (2005).
7. Müller, H. S. P., Thorwirth, S., Roth, D. A. & Winnewisser, G. The Cologne database for molecular spectroscopy, CDMS. Astron. Astrophys. 370, L49–L52 (2001).
8. Pickett, H. M. et al. Submillimeter, millimeter and microwave spectral line catalog. J. Quant. Spectrosc. Radiat. Transfer 60, 883–890 (1998).
9. Jacquinet-Husson, N. et al. The 2009 edition of the GEISA spectroscopic database. J. Quant. Spectrosc. Radiat. Transfer 112, 2393–2445 (2011).
10. Dubernet, M. et al. Virtual Atomic and Molecular Data Centre. J. Quant. Spectr. Radiat. Transfer 111, 2151–2159 (2010).
11. Tashkun, S. A., Pervolov, V. I., Telfo, L. J., Bykov, A. D. & Lavrentieva, N. N. CCDSD-1000, the high-temperature carbon dioxide spectroscopic databank. J. Quant. Spectrosc. Radiat. Transfer 82, 165–196 (2003).
12. Császár, A. G. et al. The fourth age of quantum chemistry: Molecules in motion. Phys. Chem. Chem. Phys. 14, 1085–1106 (2012).
13. Polyanysky, O. L. J. et al. High-accuracy ab initio rotation-vibration transitions for water. Science 299, 539–542 (2003).
14. Pavanello, M. et al. Precision measurements and computations of transition energies in rotationally cold triatomic hydrogen ions up to the mid-visible spectral range. Phys. Rev. Lett. 108, 023002 (2012).
15. Tennyson, J. & Yurchenko, S. N. ExoMoL: molecular line lists for exoplanet and other atmospheres. Mon. Not. R. Astron. Soc. 425, 21–33 (2012).
16. Mellau, G. Ch. Complete experimental rovibrational eigenenergies of HNC up to 3743 cm\(^{-1}\) above the ground state. J. Chem. Phys. 133, 164303 (2010).
17. Mellau, G. Ch. Complete experimental rovibrational eigenenergies of HCN up to 6880 cm\(^{-1}\) above the ground state. J. Chem. Phys. 134, 234303 (2011).
18. Mellau, G. Ch. Rovibrational eigenenergy structure of the [H\(_2\)CN] molecular system. J. Chem. Phys. 134, 194302 (2011).
19. Barber, R. et al. ExoMoL line lists III: An improved hot rotation-vibration line list for HCN and HNC. Mon. Not. Royal Astron. Soc. 437, 1828–1835 (2014).
20. Császár, A. G., Czako, G., Furtenbach, T. & Matyus, E. An active database approach to complete spectra of small molecules. Annu. Rep. Comp. Chem. 3, 155–176 (2007).
21. Furtenbach, T., Császár, A. G. & Tennyson, J. MARVEL: measured active rotational-vibrational energy levels. J. Mol. Spectrosc. 245, 115–125 (2007).
22. Furtenbach, T. & Császár, A. G. MARVEL: measured active rotational-vibrational energy levels. II. Algorithmic improvements. J. Quant. Spectr. Radiat. Transfer 113, 929–935 (2012).

Acknowledgments
This project was supported by the Hungarian Scientific Research Fund (OTKA NK83583) and by an ERA-Chemistry grant.

Author contributions
A.G.C., T.F. and P.A. conceived and designed the research described. A.G.C. and G.M. performed the experiments. A.G.C. and G.M. analyzed the data and wrote the manuscript.

Additional information
Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Furtenbach, T., Arendás, P., Mellau, G. & Császár, A.G. Simple molecules as complex systems. Sci. Rep. 4, 6454; DOI:10.1038/srep04654 (2014).