A network approach to atomic spectra

David Wellnitz$^{1,2,7}$, Armin Keki$^{2,3,4,7}$, Julian Heiss$^1$, Michael Gertz$^5$ and Matthias Weidemüller$^{2,*}$
and Andreas Spitz$^6,*$

$^1$ ISIS (UMR 7006) and IPCMS (UMR 7504), University of Strasbourg and CNRS, and icFRC, 8 Rue Gaspard Monge, 67000 Strasbourg, France
$^2$ Physikalisches Institut, Universität Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany
$^3$ École Normale Supérieure, 45 Rue d’Ulm, 75005 Paris, France
$^4$ Max Planck Institute for Intelligent Systems, Max-Planck-Ring 4, 72076 Tübingen, Germany
$^5$ Institute of Computer Science, Universität Heidelberg, 69120 Heidelberg, Germany
$^6$ Department of Computer and Information Science, University of Konstanz, 78464 Konstanz, Germany
$^7$ These authors contributed equally.
$^{*}$ Authors to whom any correspondence should be addressed.
E-mail: weidemueller@uni-heidelberg.de and andreas.spitz@uni-konstanz.de

Keywords: complex networks, community detection, link prediction, atomic spectra, stochastic blockmodel

Abstract

Network science provides a universal framework for modeling complex systems, contrasting the reductionist approach generally adopted in physics. In a prototypical study, we utilize network models created from spectroscopic data of atoms to predict microscopic properties of the underlying physical system. For simple atoms such as helium, an a posteriori inspection of spectroscopic network communities reveals the emergence of quantum numbers and symmetries. For more complex atoms such as thorium, finer network hierarchies suggest additional microscopic symmetries or configurations. Furthermore, link prediction in spectroscopic networks yields a quantitative ranking of yet unknown atomic transitions, offering opportunities to discover new spectral lines in a well-controlled manner. Our work promotes a genuine bi-directional exchange of methodology between network science and physics, and presents new perspectives for the study of atomic spectra.

1. Introduction

Network science [1, 2] provides a holistic framework for studying complex systems, allowing for a deeper understanding of their structure and dynamical behavior [3]. Thus, network science can help identify patterns in the interactions of the systems’ components. Applications include social [4, 5], communication [6, 7], ecological [8, 9], physiological [10, 11], and epidemiological [12, 13] networks. In contrast, complex physical systems such as atoms [14–16], molecules [17, 18], materials [19], or even the Universe [20] are generally approached from a microscopic, genuinely reductionist point of view, starting from fundamental sets of equations. While computing the solutions to these equations represents a compelling challenge in computational physics, the task often becomes intractable with increasing system size.

For instance, while the hydrogen spectrum can be computed analytically from solving the Schrödinger equation (and restricting solely to the Coulomb interaction between the proton and the electron), the inclusion of electron-electron interactions renders even moderately sized atoms containing several electrons intractable. Already the helium atom with just two electrons needs to be treated numerically [21]. The state space of quantum many-body systems grows exponentially with the number of particles [22, 23], making the treatment of atoms with large numbers of electrons highly challenging, and generally necessitating approximate solutions [14–16]. In this context, the perspective offered by network science becomes particularly valuable for the predicting features of complex physical systems [24–26]. Alternatively, physical
systems with known ground truth can serve as a complementary testbeds for evaluating algorithms in network science, in contrast to commonly used social benchmarks with unknown ground truth [27, 28].

Network models of data often reveal underlying patterns in their structure, frequently manifesting as communities with strong internal links. Such patterns are commonly observed in social networks [4, 29, 30]. Stochastic block models in particular provide a method to generalize this notion by identifying generic patterns in the network's connectivity matrix, extending beyond the concept of associative communities [31–33]. By associating nodes with communities, predictions can be made about the structure underlying the system, such as identifying agents with similar properties or functionalities. Similarly, the task of link prediction, another fundamental task in network science [34, 35], is focused on the inference of connections in complex networks, in an effort to either model a network's natural evolution, or to suggest missing links that were not (yet) experimentally measured.

The structural information gained from network science has been applied to the analysis of spectroscopic data by mapping molecular spectra to networks [36]. Such mappings can be used to improve data accuracy and assignment [36, 37], identify errors [38, 39], and design efficient measurements [37, 38]. However, previous applications of network modeling to molecular spectroscopic data have focused on efficiently analyzing the data, rather than making predictions about the underlying molecular structure.

Here, we investigate the potential of network science to make predictions about atomic spectra as prototypical complex physical systems, which are closely related to the underlying quantum mechanical structure [21]. This reverses the reasoning as compared to prior works [36–39], which drew conclusions about spectroscopic networks based on knowledge about the molecular structure. Using a stochastic blockmodel approach, we find that communities reveal the underlying quantum mechanical properties of the atom. For the simpler system of helium I, we show that known physical symmetries can be recovered from the network structure, while for the more complex thorium II, we uncover additional structure beyond the known selection rules [21]. Furthermore, we find that link prediction enables a direct prediction of the atoms’ spectroscopic properties without any intermediate connection to the underlying physics. While standard algorithms that are based on intuition gained from social networks perform poorly due to fundamental differences in the networks' semantics, we identify approaches that provide accurate predictions even for complex spectra.

The remainder of the paper is structured as follows. We first introduce a mapping of spectroscopic data to networks and discuss their basic properties in section 2. We proceed with identifying and discussing the communities of the spectroscopic networks in section 3. In section 4, we demonstrate how link prediction can be used to predict yet unknown atomic transitions with great accuracy. In section 5, we conclude with a discussion of the implications of our findings.

2. Spectroscopic networks

The spectroscopic data of an atom are a set of transitions between its internal energy levels (states) that characterize the atom's structure [41]. This spectroscopic data can be naturally represented as a network, where energy levels correspond to nodes (vertices) and transitions between states correspond to links (edges) between nodes, as illustrated in figure 1(a).

Spectroscopic data can be obtained either from a solution of the Schrödinger equation for small systems such as hydrogen or helium, or from the empirical observation of transitions, as compiled, e.g. in the National Institute of Standards and Technology atomic spectra database [42]. In the following, helium I (neutral helium) serves as an example of a numerically solvable system, while we investigate thorium II (singly ionized thorium) as a more complex system without adequate ab initio solutions. For helium I, 193 energy levels and 2300 transitions are available, while the database contains 516 energy levels and 6502 transitions for thorium II.

The thorium II network forms a single connected component, while the helium I network contains two disconnected components with 191 nodes and two nodes, respectively. We restrict our analysis to the largest connected component only (i.e. we exclude two outliers), since many link prediction algorithms are designed only for connected networks and thus cannot predict links between disconnected components and the two-node outlier component can be regarded as internally fully connected.

One prevalent characteristic that is observed in various network science domains, including in molecular spectroscopic networks, is a scale-free or at least long-tailed degree distribution [36, 43]. In contrast, we find that the spectroscopic networks of helium I and thorium II do not exhibit a long-tailed degree distribution (see appendix A). In fact, it can be argued that typical emergence mechanisms in complex networks, such as preferential attachment (PA) [43], cannot accurately describe spectroscopic networks due to their unique structure governed by microscopic laws and symmetries [21]. These symmetries impose strict constraints on
Figure 1. Helium I spectroscopic network. (a) Mapping of spectra to networks. Top: Grotrian diagram of helium I. Colored horizontal lines indicate electronic energy levels, arranged by their energy on the y-axis (not to scale), and their orbital angular momentum on the x-axis. Arrows indicate transitions between the energy levels, which are spectroscopically measured and identified by their wavelength. Bottom: Corresponding network representation. Energy levels are mapped to nodes (circles with corresponding colors) and transitions are identified with links, as indicated by the grey arrows. (b) Full helium I network. Circles denote network nodes, while light grey lines indicate links between the nodes. The hierarchical community structure as identified by the nested stochastic block model [33] is highlighted by horizontal and vertical lines as labeled. Spatial arrangement marks the lowest hierarchy level. The colors denote the orbital angular momentum $L$, the symbols encode the spin $S$, and the total angular momentum $J$. The black nodes correspond to doubly excited states (helium I), respectively. Plotted using Gephi [40].

the possible transitions, known as selection rules [21, 44]. For instance, electric dipole transitions always change the atomic state's parity, which gives rise to an (approximate) bipartite structure of the spectroscopic network [36], broken only by weaker electric quadrupole or magnetic dipole transitions. Additional selection rules stemming from angular momentum conservation dictate that the state's total angular momentum $J$ changes by at most $|\Delta J| \leq 1$. In the subsequent sections, we unveil patterns in the network structure that correspond to these underlying symmetries.

3. Community detection

To identify patterns with predictive power in spectroscopic networks, we employ community detection, which groups nodes according to their link structure [30]. Community detection based on modularity maximization has been applied to spectroscopic networks in previous work [36]. However, it failed to reveal the full underlying community structure. Modularity maximization builds on intuition from social networks, where communities arise through triadic closure [4, 45] (people who share friends are more likely to become friends as well), and is thus not suitable for bipartite networks in general, and spectroscopic networks in particular. Instead of relying on modularity maximization, we should therefore expect an emergent anti-community structure in spectroscopic networks that can be detected by applying modularity...
minimization, which is computationally expensive [28]. Alternatively, the nested stochastic blockmodel (NSBM) offers an attractive and well-established method for community detection in non-associative networks, without relying on an associative community structure [33]. The NSBM generates a hierarchical partitioning of nodes based on their connectivity patterns, assuming that nodes belong to different communities. In spectroscopic networks, such communities may arise for states with similar physical properties, and thus similar transitions. Here, we use a degree-corrected version of the NSBM, which we heuristically found to perform best among NSBM variants. We conjecture that the performance boost over the non-corrected variant stems from the fact that degree-correction reduces the bias that may arise due to the large number of transitions measured for particularly interesting states, such as the ground state [32, 46]. For a detailed discussion of the implementation, refer to appendix section B.1 and reference [47].

3.1. Helium I
In the case of helium I, which serves as an example of a numerically solvable system, the community detection algorithm reveals a hierarchical structure consisting of four different hierarchy levels, as shown in figure 1(b). On the first (coarsest) hierarchy level, an almost exact bipartite structure is identified with only approximately 5% of the links connecting nodes within each of the two communities. This bipartite structure reflects the aforementioned parity symmetry. On the next lower hierarchy level, both communities are divided further into clusters that reflect the underlying quantum mechanical structure of the states. In particular, the second and third hierarchy level correspond to a successively finer separation into ranges of total angular momentum \( J \), except for the \( ^1S_0 \) and \( ^1P_1 \) states. For example, for the even parity states, we observe one community with \( 0 < J \leq 2 \), another with \( 2 < J \leq 4 \), and a third with \( 4 < J \leq 7 \). The final, fourth hierarchy level groups states based on their quantum numbers for spin, orbital and total angular momentum.

The correspondence between the community structure and the underlying microscopic physical principles stems from the above-mentioned selection rules, which classify transitions as either allowed or forbidden depending on the quantum numbers [21, 44]. Although a transition between any pair of states is generally possible, allowed transitions are typically substantially stronger than forbidden transitions, and thus more likely to be experimentally observed and included in the spectroscopic network. Interestingly, a mixing of states with different quantum numbers only occurs for high angular momentum states, for which fewer states are registered in the NIST database. This suggests that the identification of these quantum numbers is mostly limited by the finite resolution of the NSBM, i.e. by a minimal required number of nodes per community [33].

3.2. Thorium II
As an example of applying community detection to a more complex system that is not numerically solvable, we investigate the spectrum of thorium II, for which a large body of spectroscopic measurement data is available. Due to its complex electronic structure as an actinide with three valence electrons, many symmetries are broken, rendering an \textit{ab initio} solution of the Schrödinger equation impossible. In fact, not even basic quantum numbers such as the total angular momentum can be assigned to every state [48, 49].

Three hierarchy levels are identified by the NSBM as shown in figure 2(a). At the first hierarchy level, we find a perfect bipartite separation of nodes, reflecting parity as a good quantum number. The second hierarchy level essentially corresponds to the states with a given total angular momentum quantum number \( J \). Nodes shown in black in figure 2(a) had no \textit{a priori} assignment in the database of experimental results to date, but based on the underlying community structure, we argue that a corresponding \( J \) value might be associated \textit{a posteriori}.

For example, consider the unassigned state (black) in the bottom left community (pink) in figure 2(a), which entirely consists of states with \( J = 1/2 \). Based on the selection rule \(|\Delta J| \leq 1 \), this node could be assigned \( J = 1/2 \) or \( J = 3/2 \). However, based on the community assignment, we can reasonably suggest this unassigned state should also be assigned \( J = 1/2 \). The NSBM community assignment thus offers an attractive alternative based on statistical connectivity patterns when predictions based purely on selection rules remain ambivalent. A more principled method for predictions including averaging multiple likely hierarchical partitions is presented in reference [50].

For the third hierarchy level however, we find no obvious quantum mechanical correspondence, except for a finer resolution of a few high angular momentum states \( J > 9/2 \). To gain further insight into the microscopic interpretation of this hierarchy level, we arrange the thorium II nodes according to the energies of the corresponding states (see figure 2(b)). Interestingly, the different communities form clusters within a contiguous energy range, indicating that states represented by a community share common physical properties. One might speculate that the orbital structure of these states bares some similarities, an
assumption which is supported by correlations between communities and occupied electron orbitals (see appendix C). Therefore, these communities might indicate hidden symmetries that could be exploited for applying advanced methods in electronic structure calculations, such as Configuration Interaction [48, 51].

In addition to thorium, we also investigate communities in the spectroscopic network of iron, for which we find similar structures (see appendix D).

4. Link prediction

Link prediction is a network analysis technique used to infer missing links based on the network's structure [34]. It often emulates natural densification processes that involve implicit community structure, such as triadic closure in social networks [4, 29]. A link prediction algorithm assigns a ranking to absent links according to the estimated probability that the link should be present in the network. In the following, we demonstrate how such methods can be utilized to predict previously unobserved atomic transitions without relying on quantum mechanical models.

Numerous link prediction methods have been developed in the literature [34], from which we select four representative methods. The Adamic-Adar (AA) index [53] and the PA index [35] are standard methods designed for predicting evolving friendships in social networks. The AA index uses triadic closure, while the PA index is based on the generation of hubs with strong connectivity. The nested stochastic block model,
Figure 3. Receiver operating characteristic (ROC) for link prediction on (a) the helium I and (b) the thorium II network. Four different link prediction methods are compared: the structural perturbation method (SPM, green), the nested stochastic block model (NSBM, orange), the preferential attachment index (PA, purple), and Adamic Adar index (AA, pink). For each spectroscopic network, 10% of links are removed uniformly at random and used as a ground truth for subsequent prediction (dropout method) \[52\]. The ROC is constructed by iterating over the computed ordered list of predictions. Each incorrect prediction moves the curve one step to the right, and each correct prediction one step up. The shaded area indicates the standard error over multiple realizations.

which we used for community detection, also supports link prediction \[33\]. Lastly, the structural perturbation method (SPM) is a state-of-the-art method that derives link probabilities from perturbations of the network’s adjacency matrix \[54\]. For implementation details of these algorithms, see appendix section B.2.

For a quantitative evaluation, we simulate missing links through dropout (see appendix section B.2 for details). In figure 3, we show the receiver operating characteristic curve \[52\] for the helium I and thorium II networks. Traditional link prediction methods that are commonly employed in social networks, such as PA and AA, perform poorly in spectroscopic networks due to the networks’ anti-community structure. In contrast, SPM and NSBM perform well for the networks of both atoms. While SPM shows slightly better results when aiming for the discovery of new transitions, NSBM provides better fidelity in retrieving the complete set of missing transitions. In addition to helium I and thorium II, we also evaluate link prediction on carbon and iron atoms, which yields similar results (see appendix D).

Encouraged by the excellent performance of link prediction in the dropout approach, we apply these methods to predict yet unobserved transitions, which we list in appendix E. We suggest that such predictions can be used to guide future high precision spectroscopic measurements looking for yet unmeasured transitions and thus reduce the associated experimental cost.

5. Conclusion

Through modeling atomic transitions as spectroscopic networks, we have demonstrated the potential of utilizing network science for predicting atomic properties. By analyzing the network structure, we are able to identify communities of states that align well with underlying physical symmetries, which suggests that these communities may be used to make predictions about the underlying states and their quantum numbers. These results may also aid in developing a deeper understanding of complex atomic spectra. A better understanding of the thorium spectrum in particular has direct practical implications for the development of nuclear clocks \[55–57\]. Furthermore, we have shown that the network structure can also be used to directly make predictions about new, yet unobserved transitions and thereby guide experimental observation.

More generally, our results highlight the holistic character of network science, offering a promising path towards gaining further insights into the structure of complex physical systems. This complements previous applications in spin and Hubbard models \[24, 26\]. Furthermore, complex physical systems such as atomic spectra, for which the solutions of the microscopic equations serve as a ground truth, can be considered a complementary and novel paradigm for testing algorithmic methods in network science \[28\], in contrast to, e.g. commonly used social benchmarks \[27\], where determining an \textit{ab initio} ground truth is challenging or impossible. Finally, the approach that we present here can readily be generalized to other discrete spectra, such as molecular or nuclear spectra \[36, 58\]. Other promising future applications of concepts from network science to spectroscopy include the derivation of transition strengths from weighted networks \[59\], or the
prediction of yet unobserved atomic states by emerging node prediction methods [50, 60], providing valuable guidance to experimental and theoretical studies of atomic structure.

**Data availability statement**

All data that were used in our experiments as well as our Python implementation are available at https://doi.org/10.5281/zenodo.8059282 [47].

**Acknowledgments**

The authors thank Sebastian Lackner, Kathinka Gerlinger, Guido Pupillo and Johannes Schachenmayer for discussions and comments. This work is supported in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster) and the Heidelberg Center for Quantum Dynamics. This work was supported in part by the LabEx NIE (‘Nanostructures in Interaction with their Environment’) under Contract ANR-11-LABX0058 NIE as part of the ’Investments for the future program’, and the Agence Nationale de la Recherche (Grant ANR-17-EURE-0024 EUR QMat). Armin Kekić would like to thank the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes) for their support.
Appendix A. Degree distributions

Figure A1 shows the degree distributions of the networks analyzed in the main text. Many networks have a long-tailed degree distribution or follow a power-law, which would show as a straight line in a log-log plot. As is obvious from the figure, neither of the networks analyzed here exhibit this type of scale-invariance.

Appendix B. Methods

B.1. Community detection

Finding the absolute best partition by brute-force is intractable for the network sizes analyzed in this paper. Therefore, we use a stochastic Markov chain Monte Carlo algorithm implementation from graph-tool [62] to identify a partition with small entropy. We estimate the variance of this search over 100 independent executions of the algorithm. The resulting partitions are then compared using the adjusted rand index, with 0 corresponding to random correlations and 1 to identical partitions [61]. On the lowest hierarchy level, corresponding to the finest structure, the different identified partitions have high similarity, as shown in figure B1. In the remainder of this paper, we focus our analysis on the partition with the absolute lowest entropy (marked by an arrow in figure B1). Note that using the other partitions, we obtain identical conclusions.

B.2. Link prediction

For link prediction based on the AA index [53] and the PA index [35], we use the networkx [63] implementation. For the SPM we use our own implementation following the description in [54].

To improve numerical efficiency, we slightly adapt the principled approach described in [33]. Here, the probability of each undetected link $l$ is given by $p(l) \propto \sum_b p(l|b) e^{-S(b)}$, where the $b$ is a hierarchical partition of the network, $p(l|b)$ is the probability of link $l$ in a given partition $b$, and $S(b)$ is the entropy of $b$. 
To approximate the sum, we first sample 100 partitions with a Markov chain Monte Carlo algorithm (as implemented in [62]) that returns a set of likely partitions, starting from a low entropy partition. In contrast to the original proposal, for each partition, we compute its entropy $S(b)$, and sum all unique partitions with $S(b) - S_{\text{min}} \leq 1$, where $S_{\text{min}}$ is the smallest entropy. The addition of the entropy cutoff increases the computational efficiency, while we confirm convergence with number of samples and entropy cutoff.

For a quantitative evaluation on a known network structure, we simulate missing links through dropout, that is, by removing a randomly selected subset of links from the network and subsequently attempting to recover them. In the dropout procedure used to evaluate the link prediction, the network can become disconnected. In this case, the link prediction was performed on the largest connected component. Since less than one disconnected node is removed from the networks in this way in our experiments on average, the impact on the results is negligible.

Appendix C. Thorium II configurations

Figure C1 shows the contribution of the different orbitals to the states. In the NIST database, at most the two dominant configurations are registered, and for many states no orbitals were registered.

For even parity states, we find a clear separation between states with and without contributions from the $5f$ orbital at the lowest hierarchy level. For odd parity states, we find no single important orbital, but the $6d^27p$ and the $5f6d7s$ orbitals are generally assigned to different communities.
Appendix D. Results for other elements

In this section, we discuss the generalization of our results to other spectroscopic networks generated from the spectra of iron I and carbon I, made up of 846 nodes connected by 9897 links, and 180 nodes connected by 1377 links, respectively. Figure D1 shows a comparison of the communities identified for iron to the quantum numbers of each state. We find perfect correspondence between the first hierarchy level and the parity quantum number of the corresponding states, as indicated by the even and odd label. Furthermore, we find that on the fourth level, nodes in the same community are likely to share the same total angular momentum quantum number, analogous to our analysis of thorium II in the main text. A more detailed discussion can be found in [50].

Figure D2 shows a benchmark for the link prediction of carbon and iron, analogous to figure 3 of the main text. The results mirror our findings for helium I and thorium II in section 4. The Adamic Adar index performs worst in all cases, followed by PA. The SPM performs best for the top ranked predictions in all cases, whereas the NSBM performs more accurate for low ranked predictions, as discussed in section 4 of the paper. The prediction was run after a random dropout of 10% of the links. Due to computational cost, only one run was performed with the NSBM on iron. In all other cases the mean and standard deviation of the mean computed from 100 different dropouts are displayed. See [64] for more methods and a more detailed discussion.

Appendix E. Link prediction results

Tables E1–E4 show the top ranked prediction for missing links when taking the full spectroscopic network as a base for prediction. All predictions in the tables are dipole-allowed, showing that the link prediction methods do not contradict known selection rules.
Figure D2. Link prediction for carbon and iron. The receiver operating characteristics for link prediction in the network generated from a neutral carbon and b neutral iron with the methods discussed in the paper are shown, analogous to figure 3 in the main text.

Table E1. Predicted transitions by the NSBM for helium I. The top ten predictions from the full network are given, ranked by the prediction score. The states are identified by their energies in cm\(^{-1}\). Additionally, the term and electron configuration for both states is given.

| Score      | Energy 1   | Energy 2     | Term 1 | Term 2 | Conf 1 | Conf 2 |
|------------|------------|--------------|--------|--------|--------|--------|
| 13.304 399 | 196 955.226 641 | 196 955.946 185 | 3D2    | 3F2    | 1s.9d  | 1s.9f  |
| 9.977 080  | 195 262.726 142 | 197 213.352 292 | 3G3    | 3F4    | 1s.6g  | 1s.10f |
| 8.910 412  | 196 071.415 767 | 197 213.420 506 | 3H4    | 3G5    | 1s.7h  | 1s.10g |
| 8.252 403  | 197 213.437 274 | 197 213.420 506 | 1H5    | 3G5    | 1s.10h | 1s.10g |
| 7.463 494  | 195 262.726 142 | 195 262.795 043 | 3G3    | 3H4    | 1s.6g  | 1s.6h  |
| 4.935 619  | 197 213.432 784 | 196 071.415 767 | 1G4    | 3H4    | 1s.7g  | 1s.7h  |
| 4.935 619  | 196 071.368 738 | 197 213.420 506 | 3G4    | 3H4    | 1s.7g  | 1s.7h  |
| 4.715 659  | 196 956.061 082 | 196 956.067 880 | 1H5    | 3F2    | 1s.7h  | 1s.6h  |
| 4.438 541  | 197 213.420 506 | 197 213.436 916 | 3G5    | 3F4    | 1s.10g | 1s.10f |

Table E2. Top 10 predicted transitions by the SPM for helium I. The top ten predictions from the full network are given, ranked by the prediction score. The states are identified by their energies in cm\(^{-1}\). Additionally, the terms and electron configurations of both states are given.

| Score      | Energy 1   | Energy 2     | Term 1 | Term 2 | Conf 1 | Conf 2 |
|------------|------------|--------------|--------|--------|--------|--------|
| 0.445 071  | 195 262.795 750 | 195 262.724 684 | 1H5    | 3G5    | 1s.6h  | 1s.6f  |
| 0.430 666  | 191 451.897 461 | 191 444.482 131 | 1F3    | 3D2    | 1s.4f  | 1s.4f  |
| 0.421 943  | 196 069.676 490 | 196 071.178 730 | 3D1    | 3F2    | 1s.7d  | 1s.7f  |
| 0.416 554  | 195 262.794 095 | 195 262.724 684 | 3H6    | 3G5    | 1s.6h  | 1s.6g  |
| 0.411 815  | 193 921.125 753 | 193 917.151 287 | 3F2    | 3D3    | 1s.5f  | 1s.5d  |
| 0.397 299  | 193 921.617 719 | 193 921.121 343 | 3G5    | 3F4    | 1s.5g  | 1s.5f  |
| 0.384 420  | 196 069.673 050 | 196 071.178 730 | 3D1    | 3F2    | 1s.7d  | 1s.7f  |
| 0.377 473  | 196 595.062 365 | 196 596.080 442 | 3D2    | 3F2    | 1s.8d  | 1s.8f  |
| 0.373 079  | 196 955.946 185 | 196 955.228 258 | 3F2    | 3D1    | 1s.9f  | 1s.9d  |
| 0.367 857  | 195 262.724 684 | 195 262.793 051 | 3G5    | 3H5    | 1s.6g  | 1s.6h  |

Table E3. Predicted transitions by the NSBM for thorium II. The top ten predictions from the full network are given, ranked by the prediction score. The states are identified by their energies in cm\(^{-1}\). Additionally, the total angular momenta of both states are given.

| Score      | Energy 1   | Energy 2     | J 1 | J 2 |
|------------|------------|--------------|----|----|
| 7.641 832  | 1859.93 843 | 40 472.45 100 | 3/2| 3/2|
| 7.587 684  | 40 472.45 100 | 22 106.43 260 | 3/2| 3/2|
| 4.870 402  | 17 121.62 038 | 33 730.93 510 | 3/2| 5/2|
| 4.701 044  | 7828.56 081 | 40 472.45 100 | 1/2| 3/2|
| 4.661 393  | 44 807.93 520 | 10 673.13 832 | 7/2| 5/2|
| 4.618 527  | 44 807.93 520 | 8378.85 915 | 7/2| 7/2|
| 4.613 255  | 41 46.67 708 | 40 706.81 310 | 7/2| 7/2|
| 4.552 611  | 1521.89 632 | 40 472.45 100 | 5/2| 3/2|
| 4.542 269  | 15 453.03 596 | 33 730.93 510 | 7/2| 5/2|
| 4.515 303  | 15 710.84 204 | 42 336.82 990 | 3/2| 5/2|
Table E4. Top ten predicted transitions by the SPM for thorium II. The top ten predictions from the full network are given, ranked by the prediction score. The states are identified by their energies in cm$^{-1}$. Additionally, the total angular momenta of both states is given.

| Score   | Energy 1     | Energy 2     | J 1   | J 2   |
|---------|--------------|--------------|-------|-------|
| 0.200 714 | 21 297.41 708 | 4146.57 708 | 5/2   | 7/2   |
| 0.188 276 | 17 460.62 698 | 1521.89 632 | 5/2   | 5/2   |
| 0.178 387 | 12 570.49 403 | 21 682.74 765 | 7/2   | 7/2   |
| 0.176 379 | 12 902.37 757 | 34 543.55 740 | 3/2   | 5/2   |
| 0.172 929 | 4146.57 708 | 24 982.44 580 | 7/2   | 7/2   |
| 0.170 751 | 30 452.72 533 | 9202.26 506 | 9/2   | 7/2   |
| 0.169 594 | 9585.40 432 | 31 259.29 670 | 5/2   | 5/2   |
| 0.168 869 | 8460.35 308 | 21 131.79 959 | 3/2   | 3/2   |
| 0.166 028 | 37 063.30 650 | 14 275.57 661 | 9/2   | 9/2   |
| 0.164 764 | 9720.29 776 | 37 787.87 910 | 7/2   | 7/2   |

ORCID iDs

David Wellnitz Ⓥ https://orcid.org/0000-0003-4978-8938
Armin Kekić Ⓥ https://orcid.org/0000-0002-1940-2523
Michael Gertz Ⓥ https://orcid.org/0000-0003-4530-6110
Matthias Weidemüller Ⓥ https://orcid.org/0000-0001-5639-5126
Andreas Spitz Ⓥ https://orcid.org/0000-0002-5282-6133

References

[1] Newman M E J 2010 Networks: An Introduction (Oxford University Press)
[2] Barabási A-L’o 2016 Network Science (Cambridge University Press)
[3] Albert R, Jeong H and Barabási A-L’o 2000 Error and attack tolerance of complex networks Nature 406 378
[4] Granovetter M S 1973 The strength of weak ties Am. J. Soc . 78 1360–80
[5] Newman M E J 2001 The structure of scientific collaboration networks Proc. Natl Acad. Sci. 98 404–9
[6] Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, Tomkins A and Wiener J 2000 Graph structure in the web Comput. Netw. 33 309–20
[7] Faloutsos M, Faloutsos P and Faloutsos C 1999 On power-law relationships of the internet topology ACM Sigcomm Computer Communication Review vol 29 (ACM) pp 251–62
[8] Montoya J M, Pimm S L and Sole R V 2006 Ecological networks and their fragility Nature 442 259
[9] Proulx S R, Promislow D E L and Phillips P C 2005 Network thinking in ecology and evolution Trends Ecol. Evol. 20 345–53
[10] Jeong H, Tombor B, Albert R, Oltvai Z N and Barabási A-L’o 2000 The large-scale organization of metabolic networks Nature 407 651
[11] Rual J-F et al 2005 Towards a proteome-scale map of the human protein–protein interaction network Nature 437 1173
[12] Keeling M J and Eames K T D 2005 Networks and epidemic models J. R. Soc. Interface 2 295–307
[13] Ancel Meyers L, Kerr B and Kolek K 2012 Network perspectives on infectious disease dynamics Interdiscip. Perspect. Infectious Dis. 2011 1
[14] Froese Fischer C, Godefroid M, Brage T, Jönsson P and Gaigalas G 2016 Advanced multiconfiguration methods for complex atoms: I. Energies and wave functions J. Phys. B: At. Mol. Opt. Phys. 49 182004
[15] Lehtola S 2019 A review on non-relativistic, fully numerical electronic structure calculations on atoms and diatomic molecules Int. J. Quantum Chem. 119 e25968
[16] Barca G M J et al 2020 Recent developments in the general atomic and molecular electronic structure system J. Chem. Phys. 152 154102
[17] Hebre W J 1976 Ab initio molecular orbital theory Acc. Chem. Res. 9 399–406
[18] Marx D and Hutter J 2009 Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods (Cambridge University Press)
[19] Ohno K, Esfarijani K and Kawoaze Y 2018 Computational Materials Science: From ab initio to Monte Carlo Methods (Springer)
[20] Springel V, Frenck C S and White S D M 2006 The large-scale structure of the Universe Nature 440 1150–3
[21] Braendsen B H and Joachain C J 2003 Physics of Atoms and Molecules (Pearson Education. Prentice Hall)
[22] Feynman R 1982 Simulating physics with computers Int. J. Theor. Phys. 21 467–88
[23] Georgescu I M, Ashhab S and Nori F 2014 Quantum simulation Rev. Mod. Phys. 86 153–85
[24] Valdez M A, Jaschke D, Vargas D L and Carr L D 2017 Quantifying complexity in quantum phase transitions via mutual information complex networks Phys. Rev. Lett. 119 223301
[25] García-Pérez G, Rossi M A C, Sokolov B, Borrelli E M and Maniscalco S 2020 Pairwise tomography networks for many-body quantum systems Phys. Rev. Res. 2 023393
[26] Sokolov B, Rossi M A C, García-Pérez G and Maniscalco S 2022 Emergent entanglement structures and self-similarity in quantum spin chains Phil. Trans. R. Soc. A 380 20200421
[27] Zachary W W 1977 An information flow model for conflict and fission in small groups J. Anthropol. Res. 33 452–73
[28] Lackner S, Spitz A, Weidemüller M and Gertz M 2018 Efficient anti-community detection in complex networks Proc. 30th Int. Conf. on Scientific and Statistical Database Management, SSDBM ‘18 (ACM) pp 1–12
[29] Simmel G 1950 The Sociology of Georg Simmel vol 92892 (Simon and Schuster)
[30] Fortunato S 2010 Community detection in graphs Phys. Rep. 486 75–174
[31] Holland P W, Blackmond Laskey K and Leinhardt S 1983 Stochastic blockmodels: first steps Soc. Netw. 5 109–37
[32] Peixoto T P 2014 Hierarchical block structures and high-resolution model selection in large networks Phys. Rev. X 4 011047
