BNL-103473-2013-IR

Data mining the EXFOR database

David A. Brown[*a], John Hirdt[b], Michal Herman[a]

[a] National Nuclear Data Center. Brookhaven National Laboratory
Upton, NY 11973

[b]Saint Joseph's College
Patchogue, NY USA

December 13, 2013

Energy Sciences & Technology Department
National Nuclear Data Center
Brookhaven National Laboratory
P.O. Box 5000
Upton, NY 11973-5000
www.bnl.gov

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Data mining the EXFOR database

David Brown,*[a] John Hirdt[b] Michal Herman[a]
[a]National Nuclear Data Center, Brookhaven National Laboratory,
Upton, NY, USA
[b]Saint Joseph's College
Patchogue, NY, USA

Abstract

The EXFOR database contains the largest collection of experimental nuclear reaction data available as well as this data's bibliographic information and experimental details. We created an undirected graph from the EXFOR datasets with graph nodes representing single observables and graph links representing the connections of various types between these observables. This graph is an abstract representation of the connections in EXFOR, similar to graphs of social networks, authorship networks, etc. Analysing this abstract graph, we are able to address very specific questions such as 1) what observables are being used as reference measurements by the experimental community? 2) are these observables given the attention needed by various standards organisations? 3) are there classes of observables that are not connected to these reference measurements? In addressing these questions, we propose several (mostly cross section) observables that should be evaluated and made into reaction reference standards.

Introduction

The EXFOR nuclear experimental database provides the data which underpins nearly all evaluated neutron and charged particle evaluations in the ENDF-formatted nuclear data library (e.g. ENDF/B, JEFF, JENDL, ...). Therefore, EXFOR is in many ways the "mother library" which leads to the data used in all applications in nuclear power, security, nuclear medicine, etc. The EXFOR database includes a complete compilation of experimental neutron-induced, a selected compilation of charged-particle-induced, a selected compilation of photon-induced reaction data, and assorted high-energy and heavy-ion reaction data. The EXFOR library is the most comprehensive collection of experimental nuclear data available so it is the best place to look for an overview of what the applied and basic experimental community feels are valuable experimental reactions and quantities.

The basic unit of EXFOR is an ENTRY. An ENTRY corresponds to one experiment and contains the numerical data along with the related bibliographic information and a brief description of the experimental method. An entry is typically divided in several SUBENTs containing the various data tables resulting from the experiment. Each SUBENT contains within it a REACTION field which encodes what reaction was studied (e.g. ^{1}H(n,el)) and what quantity was measured (e.g. cross-section). A SUBENT may also contain a MONITOR fields which encodes a second well characterized reaction and quantity used to reduce or eliminate systematic experimental errors. Often the measured data encoded in the REACTION field is measured relative to the reaction/quantity encoded in the MONITOR field. There is usually a straightforward mapping between the reactions/quantities measured in EXFOR and the evaluated reactions/quantities stored in the ENDF libraries.

Several specific reaction/quantities are important enough, usually because of one or more specific applications, that the nuclear data community has elevated them to the level of an international reference standard. References [1-3] provide details of the well known neutron-induced, charged-particle and photonuclear standard reaction/quantities. There is also a new standards-level effort just beginning known as CIELO pilot project [4] which promises to entire generate standards-level evaluations including all reactions/quantities needs for the ENDF-formatted libraries for neutron-induced reactions on ^{1}H, ^{16}O, ^{56}Fe, ^{235}U, ^{238}U and ^{239}Pu.

In this work, we take a somewhat abstract view of the EXFOR database and generate an undirected graph describing all the connections between reactions/quantities in the EXFOR database. From just these connections, we can infer what reactions/quantities the nuclear data community collectively (and somewhat unconsciously) views as important. This set of reactions/quantities does not exactly match our previous expectations. We will provide a series of recommendations for reactions/quantities that should also be elevated to the level of the standards in references [1-4] and possibly included in a follow-on CIELO project.

Building the graph

As each EXFOR SUBENT corresponds to one measured dataset and each SUBENT contains exactly one REACTION field, we will focus on the REACTION fields. Each SUBENT may also contain a MONITOR field which we also note. Both REACTION and MONITOR fields have essentially the same format and contain much the same information. The MONITOR field may also contain other free-text information detailing how the monitor was used and we ignore this information. An example of a simple measurement is

$$(1\text{-}H\text{-}1(N,TOT),,SIG,,MXW) \hspace{2cm} (1)$$

This REACTION field tells us that the ^{1}H(n,tot) Maxwellian (MXW) averaged cross section (SIG) was measured in the associated SUBENT. In our graph, we consider each of these elementary reaction/quantities in REACTION or MONITOR fields to be nodes. Table 1 lists all of the types of nodes. In this table, the nodes are colorized by whether they correspond to one of the standards in the various standards efforts [1-4].

Table 1: Types of nodes

Description	Example
Regular node	⚪
CIELO isotope	🟢
ENDF/B-VII.1 Standards	⚫
Standards proposed at IAEA Technical Meeting, July '13	🔵
Standards proposed in the past / Proposed by us	🟠 🔵
Mughabghab, S. F., Atlas of Neutron Resonances	🟢
Diagnostic radioisotopes and monitor reactions	🟣
Isomer target	🔴
Elemental target	🟢

The nodes in our graph are connected by edges. The types of edges we consider are listed in Table 2. By far the most common type of edge in our graph is the MONITOR-REACTION connection. However, the EXFOR format provides several other connections between elementary nodes. REACTION and MONITOR fields may also contain mathematical relations, e.g.

$$(3\text{-}LI\text{-}6(N,T)2\text{-}HE\text{-}4,,SIG,,SPA)/(92\text{-}U\text{-}235(N,F),,SIG,,SPA) \hspace{1cm} (2)$$

Here, this field tells us that the ratio of 6Li(n,t) and 235U(n,f) spectrum averaged cross sections was measured. Any relation using +,-,*,/,//,= are allowed in the REACTION and MONITOR fields. EXFOR also allows what we call "isomer math":

$$(72\text{-}HF\text{-}177(N,G)72\text{-}HF\text{-}178\text{-}M/T,,SIG/RAT) \hspace{2cm} (3)$$

Here, what was measured was the ratio of ^{177}Hf(n,g)^{178m}Hf cross section to the total of ^{177}Hf(n,g)^{178m}Hf and ^{177}Hf(n,g)^{178g}Hf cross sections. There are several other EXFOR quantities which have special meanings and are actually simply mathematical relations:

- ALF: capture-to-fission ratio

- ETA: ave. neutron yield per nonelastic event for n-induced reactions

- RI: resonance integral

- NON, INEL, SCT: all obey sum rules

We also consider all reactions/quantities covered by the CIELO pilot project and link all reaction/ quantities corresponding to a CIELO isotope. Finally, as an element is an abundance weighted sum of the isotopes that make up the element, we link any reaction/quantity on an elemental target to the corresponding isotopic reaction/quantities.

Table 2: Types of edges

Edge type	Description	Example
MONITOR	Typically a, well characterized reaction used to reduce or eliminate systematic experimental errors.	
Mathematical relation (e.g. "isomer math"; sum rules; math is REACTION string; ALF, ETA, etc)	Connections representing a simple ratio or a more complex mathematical equation.	
Neutron Standards/CIELO	All evaluated simultaneously and therefore are linked.	
Elemental	Data on a elemental target is connected to every stable isotope of the element for the same measurement.	

We used the x4i code [5] to read the EXFOR database and parse the REACTION and MONITOR strings. We then built up the undirected graph within x4i and stored the resulting graph in a graphml formatted file. The full graph has 87925 nodes and 276 852 edges. We then studied this graph with the NetworkX [6] and graph-tool [7] codes. With graph-tool, we were able to visualize portions of the graph and this is shown in Figure 1. The final graph is too large and fully connected to visualize with the tools we currently have available. Unfortunately, we were not able to visualize the portion of the graph that contains the majority of the standards and CIELO nodes.

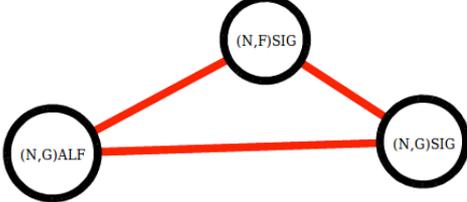**Figure 1: A portion of the second largest cluster in the graph.**
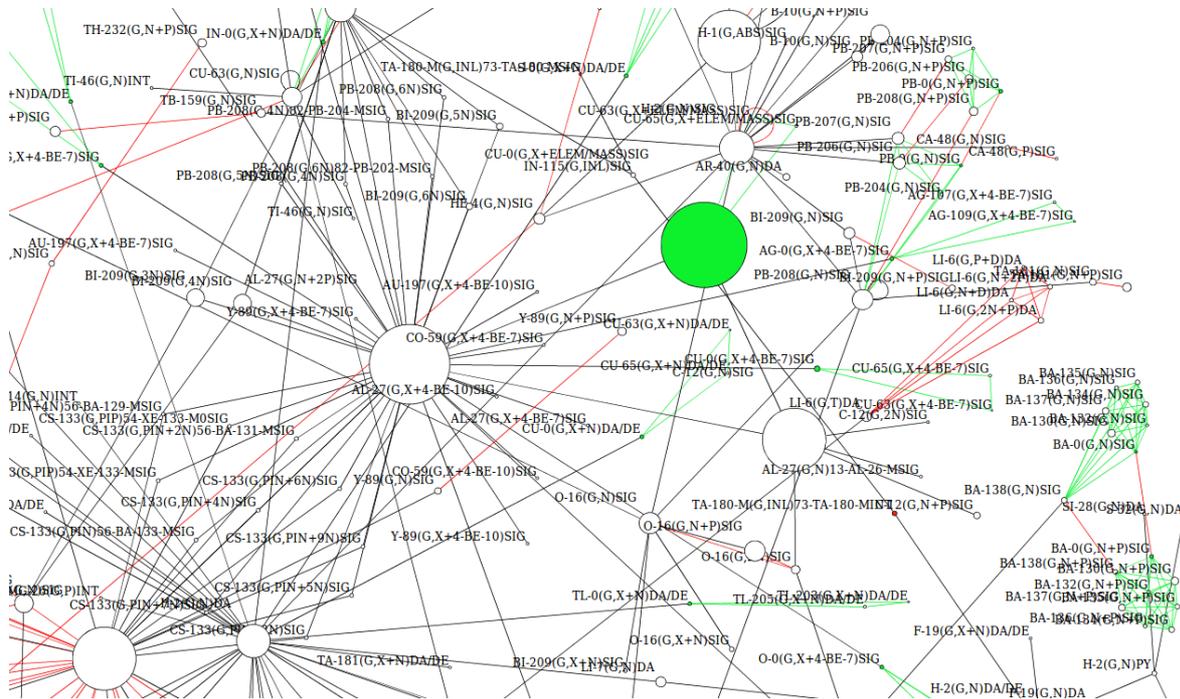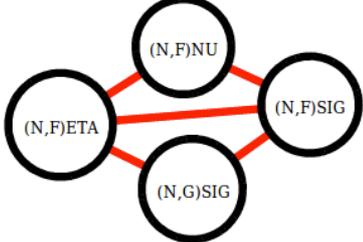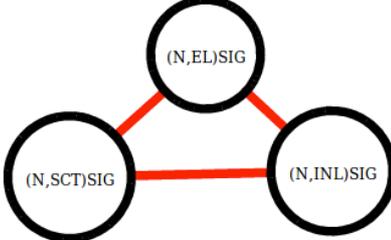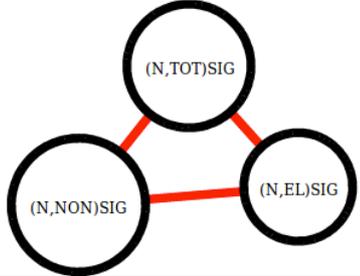

Analysing the graph

Gross features of the graph

The graph is not fully connected and the probability that any 2 nodes connected is 7.162e-05. There are 23196 isolated nodes. These nodes correspond to experiments that purport to be absolute measurements. Aside from these nodes, the graph is arranged in series of clusters. Figure 1 is a close-up of the second largest cluster which contains 488 nodes and 977 edges and contains mostly photonuclear data. The largest cluster contains 35 410 nodes including all of our standards nodes and is clearly less than 1/2 of all nodes in graph. Two other large clusters of note are the 3rd largest with 164 nodes and 460 edges and consists of thick-target yield data and the 4th largest with 149 nodes and 294 edges and consists of charged pion scattering. We wonder if all of the disconnected nodes and clusters are as disconnected as our simple analysis implies. Also, if they are, should they really remain disconnected? A reaction/quantity pinned to a standard is often in better shape than an absolute measurement.

There are many repeating patterns in our graph. In graph theory literature these patterns are known as motifs. We tabulate the motifs in our graph in Table 3. In all cases, they correspond to clusters of nodes we imposed on the graph either to encode the special EXFOR quantities or elemental data.

We also note that our graph has several self-loops. These are cases where a node connects to itself. This is caused by a bug in our graph generation code and will be removed before final publication. These self-loops are small in number and do not affect our central results.

Table 3: Reoccurring motifs in the graph

EXFOR Quantity	Definition	Example
ALF	σ_γ / σ_f	
ETA	$\overline{\nu}\sigma_f / (\sigma_\gamma + \sigma_f)$	
SCT	$\sigma_{el} + \sigma_{inel}$	
NON	$\sigma_{tot} - \sigma_{el}$	
RI	$\int_0^\infty dE \dfrac{\sigma(E)}{E}$	
Elemental	$\sum_i w_i \sigma_i(E)$	
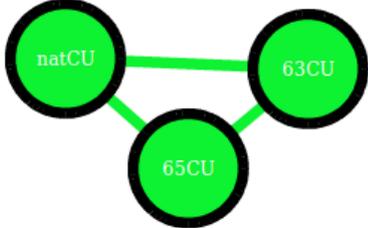
Discerning the important nodes

We now discuss which nodes are most important. There are several ways we could rank the elementary reaction/quantities by importance and the most obvious is just counting the number of measurements of each reaction/quantity. We could also rank the elementary reaction/quantities by number of connections to an elementary reaction/quantity. The number of connections to a node is known as the node degree. In table 4, we list the most important nodes rank ordered by degree. Interestingly, the top four nodes by degree or number of measurements are the same, although the rank order is different. Also interesting is the fact that seven of the top twenty reactions/quantities are not addressed by any standards effort.

Table 4: Top twenty nodes in graph ranked by degree

Node	# Measurements	Degree	Note
27Al(p, X+22Na): σ	3806	2276	IAEA Charged-particle Monitor
27Al(p, X+24Na): σ	3626	2122	IAEA Charged-particle Monitor
27Al(p, n+3p): σ	2316	1535	
27Al(n, α): σ	5049	1281	
1H(n, el): σ	2903	1207	ENDF Neutron Standard/CIELO
197Au(n, γ): σ	4106	1073	ENDF/Atlas Neutron Standard
1H(n, el): dσ/dΩ	2601	953	ENDF Neutron Standard/CIELO
56Fe(n, p)56Mn: σ	2272	833	CIELO
235U(n, f): σ	3707	774	ENDF Neutron Standard/CIELO
93Nb(n, 2n)92mNb: σ	2465	710	
natCu(p, X+65Zn): σ	1899	627	IAEA Charged-particle Monitor
27Al(12C, X+24Na): σ	1060	610	
natMo(p, X+96Tc): σ	1109	600	
natMo(p, X+97Ru): σ	547	594	
59Co(n, γ): σ	1332	578	Atlas Neutron Standard
27Al(n, p)27Mg: σ	1902	544	
238U(n, f): σ	1394	511	ENDF Neutron Standard/CIELO
27Al(d, X+24Na): σ	990	507	IAEA Charged-particle Monitor
197Au(n, γ): RI	1961	440	ENDF/Atlas Neutron Standard
10B(n, α): σ	860	432	ENDF Neutron Standard

We attempted to confirm this with several other measure of node importance including node centrality, betweenness and eigenvalue centrality [5,6]. All of these measures fail for our graph because the graph is too large. All of these measures rely on performing complex linear algebra on the adjacency matrix of the graph. The adjacency matrix is the matrix that one constructs by associating each node with a column/row. For node i, the adjacency matrix A_{ij} is defined as $A_{ij} = 1$ if node i is connected to node j and 0 otherwise. For our graph, the adjacency matrix is a 87925 x 87925 sparse matrix and is a challenge for most off-the-shelf linear algebra packages.

There is one measure, Google's PageRank [5,6], which can be used on our graph. PageRank is an iterative process to determine what the probability is that a given node is connected. The exact algorithm is given in many places and is implemented in the codes we used for our analysis (see references [5,6]). This algorithm is robust and simple and can work on graphs as large as the entire Internet. Using PageRank, we confirm the importance of the Al, Mo, and Nb reactions/quantities noted in Table 4.

Conclusion

It is clear from the analysis of our graph that the following reaction/quantities have out-sized importance and they are not considered in any standards effort:

- Aluminium reaction/quantities:
  - n+^{27}Al: (n,α), (n,p) cross sections
  - p+^{27}Al: (n,n+3p) cross section and ^{22}Na and ^{24}Na production cross sections
  - ^{12}C+^{27}Al: ^{24}Na production cross section
- Molybdinum and Niobium also very important structural materials:
  - n+^{93}Nb: ^{93}Nb(n,2n)^{92m}Nb cross section
  - p+^{nat}Mo: ^{96}Tc and ^{97}Ru production cross section

We recommend that at the very least that 27Al, 93Nb and all of the Mo isotopes be considered as a target material in any CIELO follow-on project.

Acknowledgements

References

[1] A.D. Carlson, *et al.* (2009), "International Evaluation of Neutron Cross Section Standards", *Nucl. Data Sheets*, 110.12, pp. 3215-3324.

[2] S.F. Mughabghab (2006), *Atlas of Neutron Resonances*, Elsevier Science.

[3] P. Oblozinsky, International Atomic Energy Agency IAEA, IAEA-TECDOC-1211 http://www-nds.iaea.org/medical/.

[4] M. Chadwick, *et al.* (2013), "CIELO: A Future Collaborative International Evaluated Library", *Proc. of the International Conference of Nuclear Data for Science and Technology (ND2013)*.

[5] D.A. Brown, (2011), *x4i: the EXFOR interface*, https://ndclx4.bnl.gov/gf/project/x4i/.

[6] A. Hagberg, *et al.* (2013), *NetworkX version 1.8*, http://networkx.github.io/.

[7] T. P. Peixoto (2013), *graph-tool version 2.2.26*, http://graph-tool.skewed.de/.