Network Analysis of Nuclear Databases

John Anthony Hirdt[1]

[1] *Department of Mathematics and Computer Science, St. Joseph's College, Patchogue, NY 11772, USA*

(Dated: April 15, 2014)

The EXFOR database contains the largest collection of experimental nuclear reaction data available as well as the 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 various types of connections between these observables. This graph is an abstract representation of the connections in EXFOR, similar to graphs of social networks, authorship networks, etc. By analyzing 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 nuclear science community? 2) are these observables given the attention needed by various nuclear data evaluation projects? 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.

I. PROBLEM AND MOTIVATION

In the early 1950's, the Department of Energy's Brookhaven National Laboratory began compiling and archiving nuclear reaction experimental data in the SCISRS database [1]. Over the years, this project has grown and evolved into the EXFOR project [2]. EXFOR is an international experimental nuclear data collection and dissemination project now led by the IAEA nuclear data section, coordinating the experimental nuclear data compilation and archival work of several nations.

The EXFOR nuclear experimental database provides the data which underpins nearly all evaluated neutron and charged particle evaluations in ENDF-formatted nuclear data libraries (e.g. ENDF/B, JEFF, JENDL, ...). Therefore, EXFOR is in many ways the "mother library" which leads to the nuclear 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, and assorted high-energy and heavy-ion reaction data. The EXFOR library is the most comprehensive collection of experimental nuclear data available. Therefore, it is the best place to look for an overview of what the applied and basic experimental nuclear science communities feel are valuable experimental reactions and quantities [1, 2]. In this work, we take an 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 science community collectively (and somewhat unconsciously) views as important.

The basic unit of EXFOR is an `ENTRY`. An `ENTRY` corresponds to one experiment and contains the numerical data, the related bibliographic information and a brief description of the experimental method. What an EXFOR `ENTRY` really represents is the results of work that was performed at a given laboratory at a given time. An `ENTRY` does not necessarily correspond to one particular publication, but very often corresponds to several publications. The EXFOR compiler takes the essential information from all available sources, and if needed, contacts the author for more information.

Our goal is to be able to view the EXFOR database in a way previously unseen. We wish to take the data, compiled by experimental physicists and analyzed by theoretical physicists and draw mathematical connections between the two groups. We will utilize Python to transform the database into our graph and run data analysis using our network theory metrics. We intend to compare our graph to other real life networks and look for similarities and differences that may be able to help us draw upon patterns and graph motifs in our analysis. It is our hope that we will be able to either verify existing nuclear data measurements or draw conclusions for future data measurements.

II. BACKGROUND AND RELATED WORK

An entry is typically divided in several `SUBENT`s containing the data tables resulting from the experiment. Each `SUBENT` contains a `REACTION` field which encodes what reaction was studied (e.g. ^{1}H(n,el)) and what quantity was measured (e.g. cross-section or angular distribution). A `SUBENT` may also contain a `MONITOR` field which encodes one or more well characterized reactions and quantities used to reduce or eliminate systematic experimental errors. Thus, reaction monitors are an important part of experimental data reduction. 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. Many experimenters use these reaction/quantities as monitors in their experiments. References [3–6] provide details of well known neutron-induced, charged-particle and photonuclear standard reaction/quantities. We divided these references into two different classes. We define Tier 1 observables as the product of sustained evaluation efforts, with periodic refinement. Our Tier 1 standards include the evaluations from the ENDF/B Neutron Standards [3] project and the *Atlas of Neutron Resonances* [4]. Our second tier encompasses standards that are of very high quality but are not performed as part of a sustained project. There may be follow ups or limited refinements. This second tier includes medical and dosimitry evaluations in Ref. [5] and the results of the IRDFF project [6]. There is also a new Tier 1 standards-level effort just beginning known as CIELO pilot project [7]. CIELO promises to generate entire standards-level evaluations including all reactions/quantities needed for the ENDF-formatted libraries for neutron-induced reactions on ^{1}H, ^{16}O, ^{56}Fe, ^{235}U, ^{238}U and ^{239}Pu.

When undertaking this project we specifically set out with the goal of answering some important questions.

- What are the most connected targets? What are the most connected reactions/quantities?
- Are there reactions/quantities that are so connected that they should be a standard?
- What is the connection number distribution for targets and reactions? What is the link number distribution between any two targets?
- Can we use this information to inform new measurements that decrease the distance between important targets and standards?
- Are there "bottlenecks" along the pathways from a given reaction to reaction standards that are not well measured?
- What elements and isotopes of reactions are not linked? Are any of them important for specific applications?

In order to attempt to resolve these questions we utilized graph theory as a tool to examine the connections between measurements in EXFOR.

Although we take an abstract view of the EXFOR database, from just the connections in our undirected graph we can infer what reactions/quantities the nuclear science 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 be considered for elevation to the level of the standards in references [3–6] or possibly included in a follow-on CIELO project. We also find that our graph is disconnected, implying there are large numbers of reactions/quantities that are not pinned to any monitor. In many cases, this is due to poor coding of the EXFOR `ENTRY`s. Although this is a serious problem in our study, there is no easy fix. Even if additional information is given, it is often given in one of the free text fields in EXFOR which are difficult, if not impossible, to parse. Finally, we use phenomenology from other model graphs (i.e. "scale free", "small world" and "random" graphs) to inform our discussion.

III. APPROACH AND UNIQUENESS

The goal of this project was to take the EXFOR database, transform the entries into a graph, and apply graph theory metrics to said graph, in order to draw conclusions about the database. Entries in the EXFOR data base as stated earlier, date back to the Manhattan Project and contain all pertinent information from the experiment including a free text field, in which experimental physicists can add any additional information they please about their experiment. Each EXFOR `SUBENT` corresponds to one or more measured datasets and each dataset in the `SUBENT` is associated with exactly one reaction/quantity in the `REACTION` field. Multiple reaction/quantities and datasets are denoted with EXFOR pointers. 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 [9]. 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-H-1(N,TOT),,SIG,,MXW)
```

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 assign each of these elementary reaction/quantities in `REACTION` or `MONITOR` fields to a node.

While building our graph, we associate the number of occurrences of each elementary reaction/quantity with the corresponding node. In our graph we did not make any distinction between variations in observerables. For example, "PAR,SIG" coding for partial cross sections, "SIG" for integrated cross sections and "CN,SIG" for compound nuclear cross sections are all treated as "SIG".

We make two notes on double counting of nodes. First, when the EXFOR compiler flags a reaction/quantity in the `MONITOR` field and also compiles the reaction/quantity as a ratio we count both the occurrences of the monitor reaction/quantity separately. This is straightforward to fix and will be done in future incarnations of this project. Second, it often happens that an experimenter who makes a ratio measurement will publish both the ratio values and the unfolded absolute value of a measured reaction/quantity. An EXFOR compiler will then compile both values as if they are independent datasets and provide an explanation of the sets in one of the EXFOR

free text fields. As one must parse the free text descriptions of the experiment in order to discern this, we have no simple workaround.

The nodes in our graph are connected by edges. The types of edges we consider are listed in Table I. 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.

```
((94-PU-239(N,F),,NU,,MXW)/
    (92-U-235(N,F),,NU,,MXW))
```

In this example, the measurement was the ratio of Maxwellian averaged $\overline{\nu}_p$'s from ^{239}Pu(n,f) and ^{235}U(n,f). Any relation using `+`,`-`,`*`,`/`, `//`, and `=` are allowed in the `REACTION` and `MONITOR` fields (here `//` means double ratio). EXFOR also allows what we call "isomer math":

```
(72-HF-177(N,G)72-HF-178-M/T,,SIG/RAT)
```

Here, what was measured was the ratio of ^{177}Hf$(n,\gamma)^{178m}$Hf cross section to the total of ^{177}Hf$(n,\gamma)^{178m}$Hf and ^{177}Hf$(n,\gamma)^{178g}$Hf cross sections.

We link all of the Neutron Standards [3] reactions/quantities together because they are evaluated simultaneously. We also consider all reactions/quantities covered by one isotope in the CIELO pilot project [7] to be linked together since they to are evaluated together.

Although we are not the first to take a database, and transform it into a graph, and analyze its properties in new ways, we are the first to do this in the field of Nuclear data to the extent of looking at an entire database. We generated our graph using `x4i` [8] and saved the results as a GraphML [10] file. A preliminary versions of this graph were presented in Ref. [11]. The full graph has 87,925 nodes and 276,852 edges. We then studied this graph with the `networkx` [12] and `graph-tool` [13] Python packages. With `graph-tool`, we were able to visualize portions of the graph. Examples are shown in Figures 1. We utilized `networkx` [12] for our analysis of the graph, as well as for testing purposes throughout our experimentation.

IV. RESULTS AND CONTRIBUTIONS

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 Tier 1 and 2 standards and CIELO nodes.

We summarize the graph's properties here:

- Number of nodes: 87,925
- Number of edges: 276,852
- Number of isolated nodes ("isolates"): 23,196
- Number of clusters with 40 or more nodes: 7
- Average degree $\langle k \rangle$: 6.2975
- Variance of degree distribution $\langle k^2 \rangle$: 39.6584
- Probability two nodes are connected p: 7.162e-5
- Average cluster coefficient $\langle C \rangle$: 0.5958

In November, at the annual meeting of the United States Nuclear Data Program, a part of National Nuclear Data Week, we had the privilege to present our results in a conference style setting. Having placed mathematical significance behind our choices(found below in the Conclusion) this will be the first time, since EXFOR's origin's, that mathematics will be used in the decision making process of future Nuclear Data Evaluation's. It is our hope, that this project will be able to save the nations involved in the EXFOR project, millions if not billions of dollars, by allowing evaluator's to make better decisions and choices then ever before.

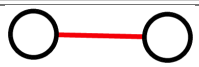
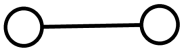
V. CONCLUSION

In this project, we created an undirected graph from the `REACTION` and `MONITOR` strings from datasets in the EXFOR database. This graph is a large, nearly scale-free network composed of disconnected clusters. The largest clusters have a "small-world" character. Our graph is in many ways typical for real world graphs.

With our graph, we identify what reactions and quantities the nuclear science community views as important enough to directly measure or measure relative to. We do this in a relatively objective fashion. Clearly the various standards projects in Refs. [3–6] have a good handle on what is important. Also, the clustering coefficients allowed us the ability to demonstrate how connected the CIELO nodes actually are. Verifying, choices made for the CIELO project. However, it is clear from the analysis of our graph that the following reaction/quantities have out-sized importance and are not considered in any standards effort:

- Aluminum reaction/quantities:
  - n+^{27}Al: the $(n, p+^{27}\mathrm{Mg})$ cross section
  - p+^{27}Al: the $(p, n+3p)$ cross section and the ^{22}Na and ^{24}Na production cross sections
  - ^{12}C+^{27}Al: the ^{24}Na production cross section
- Molybdinum also a very important structural material:
  - p+^{nat}Mo: the ^{96}Tc production cross section
  - α + ^{nat}Mo: the ^{97}Ru production cross section

TABLE I. Types of edges in our graph.

Type	Description	Example
Mathematical relations	These types of connections can be a simple ratio or a more complex mathematical relations between two or more other nodes. These include "isomer math" and the special quantities and sum rules.	
Monitor	Typically a second, well characterized target used to reduce or eliminate systematic experimental problems during data analysis.	
Elemental	Data from a natural element is connected to every stable isotope of the element for the same measurement.	
Neutron Standards/CIELO	All reactions/quantities are evaluated simultaneously and therefore are linked.	

We recommend that at the very least that ^{27}Al and all of the Mo isotopes be considered as a target material in either a follow-on CIELO or IRDFF project. In addition, a standards level study of fission product yields of the major actinides as suggested in the discussions at the recent Working Party on Evaluation Cooperation Subgroup 37 meeting [14] would improve the connectivity of all fission product yield data.

The results of our project have yielded numerous results with direct implications on the future. First, nuclear data is a field of extreme importance. Without this data, we cannot run devices such as Linear Accelerators, the Large Hadron Collider or Nuclear Reactors all over the world. The major results of this project include the first definitive, logical proof based reasoning behind the choice of possible standards. With our proposed standards, there are possibilities down the road for improvements in Nuclear Reactors/Nuclear Technologies. This includes but is not limited to improving Medical Dosimetry Reactions, material analysis for Nuclear Reactors and improved decision making tools now available to the United States Nuclear Data Program.

Further information and a full paper can be found at `http://arxiv.org/abs/1312.6200`

ACKNOWLEDGMENTS

Many Thanks go out to M. Herman (BNL) and J. Fritz (St. Joseph's College) for their support of this project and acknowledge the useful discussions with N. Otsuka (IAEA), A. Carlson (NIST), A. Plompen (IRMM) and R. Capote (IAEA). The work at Brookhaven National Laboratory was sponsored by the Office of Nuclear Physics, Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-98CH10886 with Brookhaven Science Associates, LLC. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI).

[1] Holden, Norman. "A short history of CSISRS", National Nuclear Data Center, Brookhaven National Laboratory, BNL Report BNL-75288-2005-IR (2005).
[2] International Network of Nuclear Reaction Data Centres (NRDC), "Compilation of experimental nuclear reaction data (EXFOR/CSISRS)", `http://www-nds.iaea.org/exfor/` and `http://www.nndc.bnl.gov/exfor/` (2012).
[3] A.D. Carlson, *et al.* "International Evaluation of Neutron Cross Section Standards", Nucl. Data Sheets, 110.12 (2009) 3215-3324.
[4] S.F. Mughabghab, *Atlas of Neutron Resonances*, Elsevier Science, (2006).
[5] P.Oblozinsky, "Charged-particle cross section database for medical radioisotope production; Diagnostic radioisotopes and monitor reactions", International Atomic Energy Agency IAEA, IAEA-TECDOC-1211 `http://www-nds.iaea.org/medical/` (2003).
[6] E.M. Zsolnay, R. Capote Noy, H.J. Nolthenius, A. Trkov, "Summary description of the new International Reactor Dosimetry and Fusion File (IRDFF release 1.0)" International Atomic Energy Agency IAEA, INDC(NDS)-0616, `https://www-nds.iaea.org/IRDFF/` (2012).
[7] M. Chadwick, *et al.* "CIELO: A Future Collaborative International Evaluated Library", Proceedings of the International Conference of Nuclear Data for Science and Technology (ND2013) (2013).
[8] D.A. Brown, *x4i: the EXFOR interface, version 1.0*, `https://ndclx4.bnl.gov/gf/project/x4i/` (2011).
[9] O. Schwerer, "EXFOR Formats Description for Users (EXFOR Basics)", Documentation Series for the IAEA Nuclear Data Section, Vienna (2008).
[10] GraphML File Format, `graphml.graphdrawing.org`
[11] D.A. Brown, "Visualizing the Connections in the EXFOR Database", Proceedings of the International Conference of Nuclear Data for Science and Technology (ND2013)

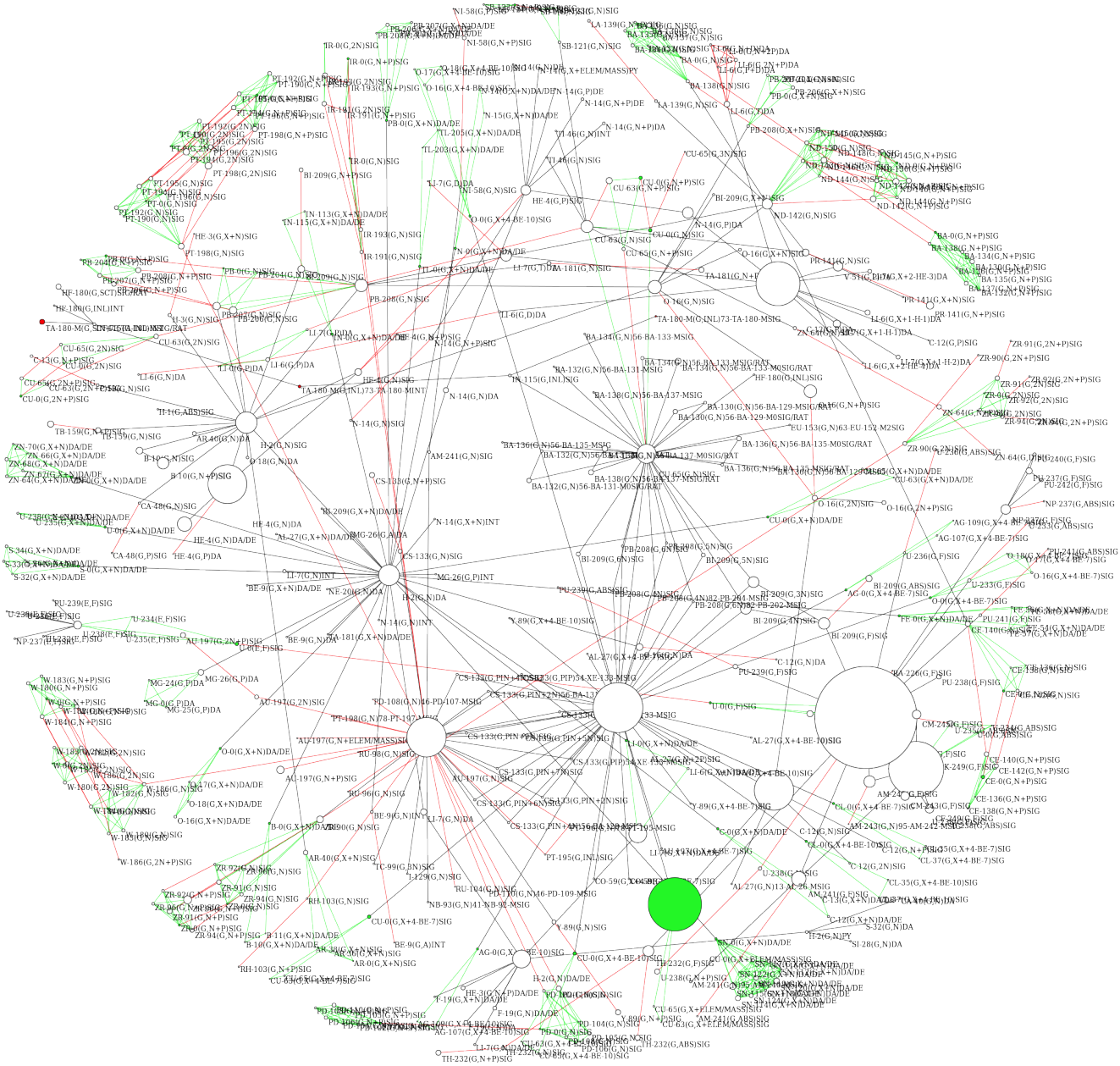


FIG. 1. The second largest cluster in the graph. This cluster is mainly composed of photonuclear data. The size of the nodes is proportional to the number of occurrences of each reaction/quantity in the EXFOR database.

(2013); D.A. Brown, J. Hirdt, M. Herman "Data mining the EXFOR database," NEMEA-7 Workshop, Geel, Belgium, BNL Report BNL-103473-2013-IR (2013).

[12] A. Hagberg, D. Schult, P. Swart, *Networkx Version 1.8.1* (2012)

[13] T.P. Peixoto, *graph-tool version 2.2.27*, `http://graph-tool.skewed.de/` (2013).

[14] WPEC-SG37 Meeting, "Improved fission product yield evaluation methodologies", NEA Headquarters, Paris, France 22 May (2013); `https://www.oecd-nea.org/science/wpec/sg37/Meetings/2013_May/`.