

# Modeling complex networks of nuclear reaction data for probing their discovery processes\*

Xiaohang Wang(王小航) Long Zhu(祝龙) Jun Su(苏军)<sup>†</sup>

Sino-French Institute of Nuclear Engineering and Technology, Sun Yat-sen University, Zhuhai 519082, China

**Abstract:** Hundreds of thousands of experimental data sets of nuclear reactions have been systematically collected, and their number is still growing rapidly. The data and their correlations compose a complex system, which underpins nuclear science and technology. We model the nuclear reaction data as weighted evolving networks for the purpose of data verification and validation. The networks are employed to study the growing cross-section data of a neutron induced threshold reaction ( $n,2n$ ) and photoneutron reaction. In the networks, the nodes are the historical data, and the weights of the links are the relative deviation between the data points. It is found that the networks exhibit small-world behavior, and their discovery processes are well described by the Heaps law. What makes the networks novel is the mapping relation between the network properties and the salient features of the database: the Heaps exponent corresponds to the exploration efficiency of the specific data set, the distribution of the edge-weights corresponds to the global uncertainty of the data set, and the mean node weight corresponds to the uncertainty of the individual data point. This new perspective to understand the database will be helpful for nuclear data analysis and compilation.

**Keywords:** modeling complex networks, neutron induced threshold reaction, photoneutron reaction, nuclear database

**DOI:** 10.1088/1674-1137/ac23d5

## I. INTRODUCTION

Nuclear reaction data underpin nuclear science and technology [1, 2]. One of the most important works in nuclear physics is to measure the nuclear reactions and reduce the data uncertainties in order to achieve the accuracy requirements in fundamental research fields, such as nuclear astrophysics [3, 4], and many application fields, such as transmutation of nuclear waste [5] and design of future nuclear reactors [6]. Since 1935, tens of thousands of experiments have been performed and hundreds of thousands of experimental data sets have been systematically collected to develop the experimental nuclear reaction databases, among which the most important and complete one is EXFOR [7]. In addition to this, evaluation and prediction using the historical data have also attracted attention, resulting in various international nuclear data libraries, such as ENDF [8], CENDL [9], JEFF [10], JENDL [11], and BROND [12].

Among the methods for data evaluation and prediction, nuclear reaction models combined with the generalized least squares method have gained prominence [13]. Successful applications of Monte Carlo evaluation meth-

ods based on the Bayesian statistical inference have also been reported [14, 15]. To solve ill-posed regression problems for prediction with uncertainty quantification, machine learning, which is a very powerful tool to learn complex big data, has been applied [16, 17]. In the past decade, investigations on advanced reactor systems have defined new accuracy and scope requirements for nuclear reaction data [1]. To this end, huge investments have been devoted to salaries, equipment, and working hours for measuring new data. Meanwhile, statistical verification and validation is an effective means for quality improvement of the rapidly growing data [18].

The growth of the nuclear reaction data is actually an innovation process, which is intrinsic for the human experience. In this respect, fruitful research on complex networks [19–22] has made it possible to model and quantify the innovation. In fact, the abstract process of innovation has been investigated in different domains including linguistics [23], biology [24], economics [25], knowledge [26], and science [27]. Many empirical analyses of real-world discovery processes have demonstrated that the basic signature of the innovation process is the Heaps

Received 7 July 2021; Accepted 6 September 2021; Published online 11 October 2021

\* Supported by the National Natural Science Foundation of China (11875328, 12075327)

<sup>†</sup> E-mail: sujun3@mail.sysu.edu.cn

©2021 Chinese Physical Society and the Institute of High Energy Physics of the Chinese Academy of Sciences and the Institute of Modern Physics of the Chinese Academy of Sciences and IOP Publishing Ltd

law [28-30], which was originally introduced to describe the number of distinct words in a text document [31]. Various models have proven that the Heaps law well describes the pace at which scientists discover concepts or users collect new items [32, 33]. In the case of science, different networks have been extracted from scholars, projects, papers, patents, ideas, and/or academic positions [27, 34]. Investigation of the key network properties offers a quantitative understanding of the interactions among scientific agents and quantitative insight into the evolution of individual scientific impact [35, 36].

This work is devoted to modeling the discovery process of the nuclear reaction data as a weighted evolving network. It is expected to map the salient features of the database to network properties so that the global uncertainty of the specific data set and the uncertainty of the individual data point can be quantified. The paper is organized as follows. In Sec. II, we describe the Bayesian Gaussian CANDECOMP/PARAFAC (BGCP) tensor decomposition model and the method to compose the networks. In Sec. III, we present both the results and discussion. Finally, a summary is given in Sec. IV.

## II. METHOD

The cross-section in a specific reaction channel, such as  $(n,2n)$  reaction, depends on the change and neutron numbers of the target and the incident energy. Let us discretize the energy degree of freedom using the energy gap  $dE$ ; the data can then be represented by a three order missing tensor. Nowadays, tensor completion is widely used in image inpainting [37] and data imputation [38]. There are various types of non-Bayesian tensor completion techniques; for example, Liu *et al.* proposed an algorithm for missing values completion in visual data [39]. The beginning of the Bayesian model in the matrix completion area was introduced by Salakhutdinov and Mnih, who built the models of Bayesian matrix factorization [40]. Kolda and Bader gave a comprehensive review of tensor decomposition [41]. Xiong *et al.* combined the tensor decomposition with Bayesian inference considering the time dependency of each tensor [42]. Chen *et al.* recently proposed a Bayesian Gaussian CANDECOMP/PARAFAC (BGCP) tensor decomposition model without the time structure [43], which is taken into account in this work. We denote by  $\mathcal{S} \in \mathbb{R}^{I \times J \times K}$  the actual value of the tensor, where the entry  $\sigma_{ijk}$  is the physical reality of the cross section in the reaction of the target  ${}^{2i+j}X_{i+j}$  at the incident energy  $E = k \cdot dE$ . Here,  $j$  denotes the isospin degrees of freedom, expressed as the difference between neutron and charge numbers  $j = N - Z$ . Indeed, we do not have the values of  $\sigma_{ijk}$  but some observations with uncertainties. Considering the multiple measurements, let  $\sigma_{ijk}^{(p)}$  ( $p = 1, \dots, P_{ijk}$ ) represent the  $p$ -th observation of the

cross-section and  $P_{ijk}$  the total number of observations. A list of missing tensors  $\bar{\mathcal{S}} = \{\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(P)}\}$  could be used to describe all the observations, where  $P = \max(P_{ijk})$  for all possible  $ijk$ . The observations are arranged in the missing tensors according to their superscripts, such as  $\sigma_{ijk}^{(1)}$  in  $\mathcal{S}^{(1)}$ ,  $\sigma_{ijk}^{(2)}$  in  $\mathcal{S}^{(2)}$ , and so on. The entries for no observations are missing. Being different from Ref. [43], the multiple observations for the physical reality  $\sigma_{ijk}$  are considered. In the following, we expound the resulting changes of the Bayesian framework of parameters.

It is assumed that the uncertainty of each observed value follows an independent Gaussian distribution,

$$\bar{\sigma}_{ijk} \sim \mathcal{N}(\sigma_{ijk}, \tau_\epsilon^{-1}), \quad (1)$$

where  $\tau_\epsilon$  is the precision. In real-world applications the expectation  $\sigma_{ijk}$  is unknown and replaced with the estimated value  $\hat{\sigma}_{ijk}$ , which is the entry of the estimated tensor  $\hat{\mathcal{S}}$ . The CP decomposition is applied to calculate the estimation  $\hat{\mathcal{S}}$ :

$$\hat{\mathcal{S}} = \sum_{l=1}^L \mathbf{z}_l \circ \mathbf{d}_l \circ \mathbf{e}_l, \quad (2)$$

where  $\mathbf{z}_l \in \mathbb{R}^I$ ,  $\mathbf{d}_l \in \mathbb{R}^J$ , and  $\mathbf{e}_l \in \mathbb{R}^K$  are respectively the  $l$ -th column vector of the factor matrices  $\mathbf{Z} \in \mathbb{R}^{I \times L}$ ,  $\mathbf{D} \in \mathbb{R}^{J \times L}$ , and  $\mathbf{E} \in \mathbb{R}^{K \times L}$ . The symbol  $\circ$  represents the outer product.

The prior distribution of the row vectors of the factor matrix  $\mathbf{Z}$  is the multivariate Gaussian

$$\mathbf{z}_i \sim \mathcal{N}[\boldsymbol{\mu}_i^{(z)}, (\boldsymbol{\Lambda}_i^{(z)})^{-1}], \quad (3)$$

where the hyper-parameter  $\boldsymbol{\mu}^{(z)} \in \mathbb{R}^L$  expresses the expectation, and  $\boldsymbol{\Lambda}^{(z)} \in \mathbb{R}^{L \times L}$  indicates the width of the distribution. The likelihood function can be written as

$$\mathcal{L}(\sigma_{ijk}^{(p)} | \mathbf{z}_i, \mathbf{d}_j, \mathbf{e}_k, \tau_\epsilon) \propto \exp\left\{-\frac{\tau_\epsilon}{2} [\sigma_{ijk}^{(p)} - (\mathbf{z}_i)^T (\mathbf{d}_j \otimes \mathbf{e}_k)]^2\right\}, \quad (4)$$

where  $\otimes$  is the Hadamard product. The posterior values of the hyper-parameters  $\boldsymbol{\mu}^{(z)}$  and  $\boldsymbol{\Lambda}^{(z)}$  are given as

$$\begin{aligned} \widehat{\boldsymbol{\Lambda}}_i^{(z)} &= \tau_\epsilon (\mathbf{d}_j \otimes \mathbf{e}_k) (\mathbf{d}_j \otimes \mathbf{e}_k)^T + \boldsymbol{\Lambda}_i^{(z)}, \\ \widehat{\boldsymbol{\mu}}_i^{(z)} &= (\widehat{\boldsymbol{\Lambda}}_i^{(z)})^{-1} [\tau_\epsilon \sigma_{ijk}^{(p)} (\mathbf{d}_j \otimes \mathbf{e}_k) + \boldsymbol{\Lambda}_i^{(z)} \boldsymbol{\mu}_i^{(z)}]. \end{aligned} \quad (5)$$

The contribution of the observations to the hyper-parameter is equivalent, being independent of which missing tensor it is arranged in.

The likelihood function of all observations is

$$\mathcal{L}(\tilde{\mathcal{S}}|\mathbf{Z}, \mathbf{D}, \mathbf{E}, \tau_\epsilon) \propto \prod_{p=1}^P \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K (\tau_\epsilon)^{1/2} \times \exp\left[-\frac{\tau_\epsilon}{2} b_{ijk}^{(p)} (\sigma_{ijk}^{(p)} - \hat{\sigma}_{ijk})^2\right]. \quad (6)$$

where  $b_{ijk}^{(p)}$  is 1 for the measured entry and 0 for the missing entry. Placing a conjugate  $\Gamma$  prior to the precision  $\tau_\epsilon$ ,

$$\tau_\epsilon \sim \Gamma(a_0, b_0), \quad (7)$$

The posterior values of the hyper-parameters  $a_0$  and  $b_0$  are given as

$$\hat{a}_0 = \frac{1}{2} \sum_{p=1}^P \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K b_{ijk}^{(p)} + a_0, \\ \hat{b}_0 = \frac{1}{2} \sum_{p=1}^P \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\sigma_{ijk}^{(p)} - \hat{\sigma}_{ijk})^2 + b_0. \quad (8)$$

Based on Eq. (8), each observation contributes to the increase of  $\frac{1}{2}$  in  $\hat{a}_0$ , and  $\frac{1}{2}(x_i^{(p)} - \hat{x}_i)^2$  in  $\hat{b}_0$ .

Equation (5) shows that the observations with the same subscript  $i$  contribute to the posterior values of the hyper-parameter  $\mu_i^{(z)}$ . Changing the second formula in Eq. (5) to

$$\hat{\mu}_i^{(z)} = \mu_i^{(z)} + \Delta\mu_i^{(z)}, \\ \Delta\mu_i^{(z)} = (\hat{\Lambda}_i^{(z)})^{-1} (\mathbf{d}_j \otimes \mathbf{e}_k) \tau_\epsilon [\sigma_{ijk}^{(p)} - (\mathbf{d}_j \otimes \mathbf{e}_k)^T \mu_i^{(z)}], \quad (9)$$

the relative deviation between two observations can be defined as

$$\delta(\sigma_{ij_1k_1}^{(p_1)}, \sigma_{ij_2k_2}^{(p_2)}) = \left[ \frac{\Delta\mu_i^{(z)}(\sigma_{ij_1k_1}^{(p_1)}) - \Delta\mu_i^{(z)}(\sigma_{ij_2k_2}^{(p_2)})}{\mu_i^{(z)}} \right]^2. \quad (10)$$

The weighted network can be built, where nodes are the observations, and the weight of the link between two observations with the same subscript is defined as

$$w(\sigma_{ij_1k_1}^{(p_1)}, \sigma_{ij_2k_2}^{(p_2)}) = \exp[-\delta(\sigma_{ij_1k_1}^{(p_1)}, \sigma_{ij_2k_2}^{(p_2)})], \quad (11)$$

Cases for the subscripts  $j$  and  $k$  are similar.

In Fig. 1, the above method is illustrated. In brief, as with the image inpainting, the cross-section in a specific reaction channel is represented by a three order tensor  $\mathcal{S}$ . According to the CP decomposition, the tensor  $\mathcal{S}$  is expressed as the outer product of the factor matrixes  $\mathbf{Z}$ ,  $\mathbf{D}$ , and  $\mathbf{E}$ . The prior distributions of the factor matrixes are assumed to be multivariate Gaussians. With the observed data of the cross-section, the posterior values of the factor matrixes and their distributions can be calculated using Bayesian inference and iteration. Finally, the predicted cross-section is reconstituted with the factor matrixes  $\mathbf{Z}$ ,  $\mathbf{D}$ , and  $\mathbf{E}$ , while the network is built with the hyper-parameter  $\mu$  and  $\Delta\mu$ .

### III. RESULTS AND DISCUSSION

The pioneers in measuring the  $(n,2n)$  reaction were Fowler and Slye, who measured the cross-sections in  $^{63}\text{Cu}(n,2n)^{64}\text{Cu}$  reactions near the threshold [44]. After decades, measurement technology and deuterium-tritium (D-T) neutron generators became widely used, resulting in the linear growth of the data points (mainly around the incident energy 14 MeV). By 1980, a good deal of data measured between the threshold and 20 MeV using the

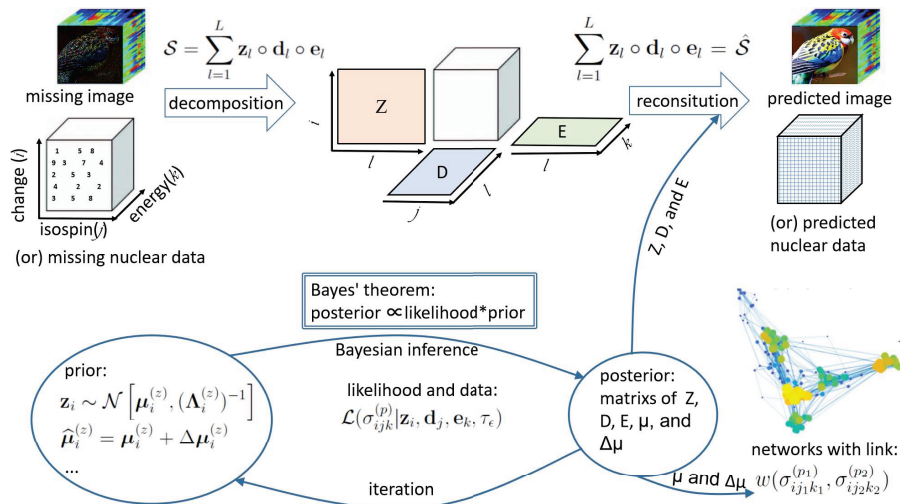


Fig. 1. (color online) Illustration for visualizing the definitions of the involved quantities and their relations.

pulsed neutron source at Bruyères-le-Châtel were published [45]. After that, owing to continuing investment in salaries, equipment, and working hours, the data points grew linearly. To date, 7671 cross-section data points of  $(n,2n)$  reaction (including 98 derived data) are recorded in the EXFOR database. Their annual growth is shown in Fig. 2(a).

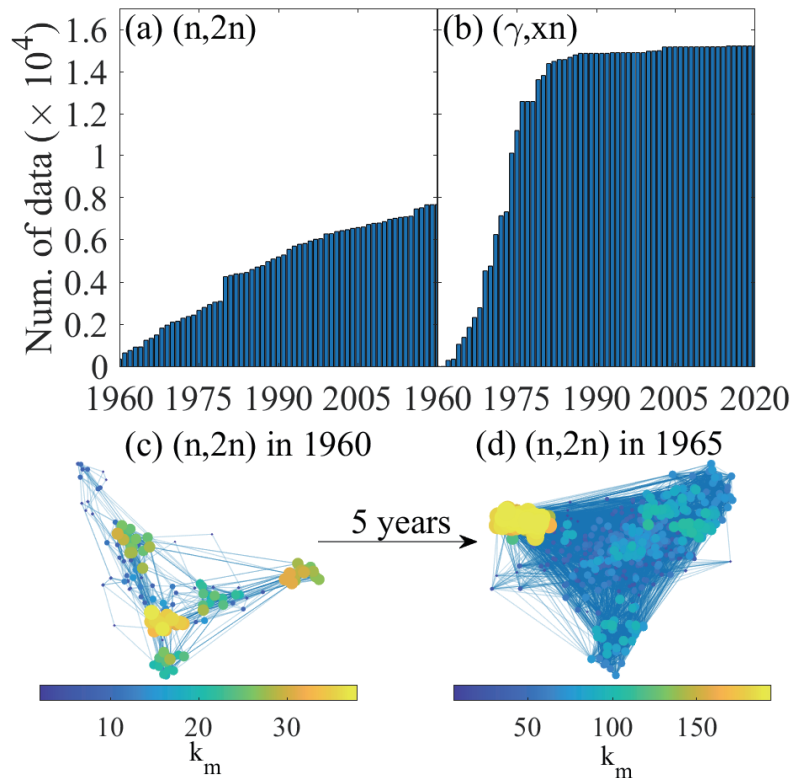
The discovery of the photoneutron reaction may date back to 1956, when the cross-section of a  ${}^6\text{Li}(\gamma,x)n$  reaction was measured by Edge [46]. Subsequently, the published data of this reaction grew rapidly. Due to the proven technology to provide a monoenergetic photon beam, the cross-section data for most of the stable nuclei have been measured within 20 years of its discovery. After 1980, only a few data were published because the scientific interest moved to more subdivided channels, such as  $(\gamma,n)$ ,  $(\gamma,2n)$  and so on. Twelve thousand data points have been collected in the EXFOR database. Their annual growth is shown in Fig. 2(b).

Those two sets of nuclear reaction data are modeled respectively as the weighted evolving networks using the BGCP approach. The nodes in the networks are the data points and weights of the links are computed from the relative deviation between the data points. An example of the network is displayed in Fig. 2(c), where 362 data

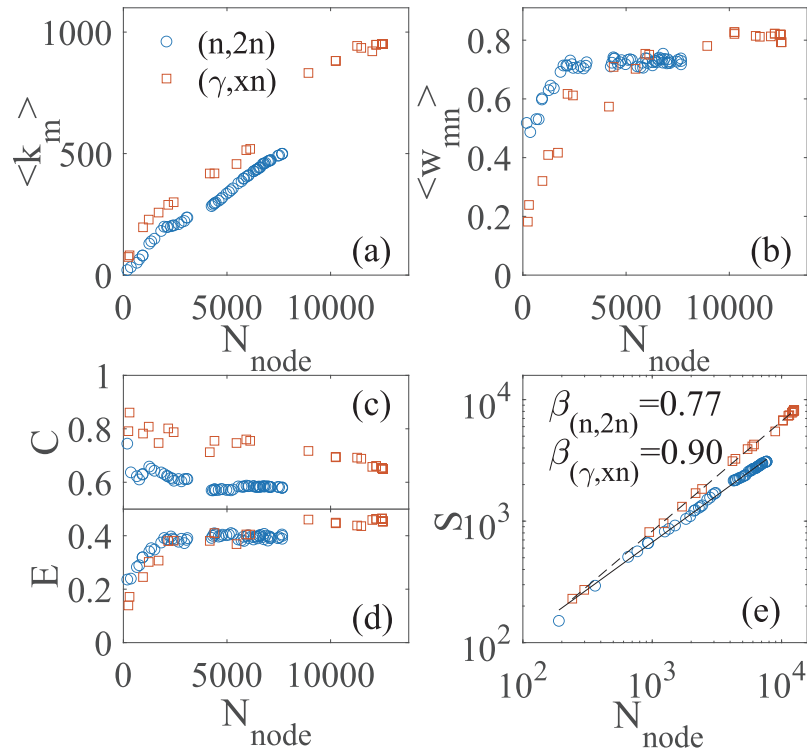
points published before 1960 are considered. The nodes, links, and numbers of neighbors are visualized clearly. The complexity of the network increases with increasing node number. Another example is shown in Fig. 2(d), where 1251 nodes are included, but the links are too numerous to recognize in the figure.

The degree  $k_m$  of a node  $m$  is the number of edges linked with the node (the number of neighbors of node  $m$ ). As shown in Fig. 3(a), the mean degrees and the node number in both networks grow approximately proportionally. It is an inevitable consequence of the discovery processes, where an experiment usually probes new data on the basis of verifying existing data. The slope reveals the innovativeness. A decrease in the slope indicates the emergence of new technologies for detecting new isotopes or a new incident energy region. A representative example is the slowdown of the mean degree  $\langle k_m \rangle$  in the  $(n,2n)$  network near  $N_{\text{node}} = 4000$ , which results from the extension of the neutron beam from 14 MeV to the threshold.

In a weighted graph, the node strength  $s_m$  is defined as the summation of the edge-weights linked with a node. It integrates the information on the number and the weights of links. The mean strength in both networks also increases linearly with increasing node number. In con-



**Fig. 2.** (color online) Growth of nuclear reaction data and their evolving networks. Annual growth cross-section data of (a)  $(n,2n)$  reaction and (b)  $(\gamma,xn)$  reaction in the EXFOR database. Network generated from the  $(n,2n)$  cross-section data published before (c) 1960 and (d) 1965. The size and color of a node represent the degree  $k_m$  of a node  $m$ .



**Fig. 3.** (color online) Properties of the evolving small-world networks. (a) Mean degree of a node  $\langle k_m \rangle$ , (b) mean weight of edges  $\langle w_{mn} \rangle$ , (c) clustering coefficient  $C$ , (d) global efficiency  $E$ , and (e) number of novelties  $S$  of the graph as a function of node number.

trast, a more interesting property is the mean weight of the edges in the graph, as shown in Fig. 3(b). It is found that not only the number (mean degree  $\langle k_m \rangle$ ) but also the weights (mean weight  $\langle w_{mn} \rangle$ ) of links increase during the growth of both networks. However, for the  $(n,2n)$  reaction, an approximate platform for the mean weight  $\langle w_{mn} \rangle$  appears for  $N_{\text{node}} > 3000$ .

The weight  $w_{mn}$  describes the linking strength between two nodes, while its reciprocal  $l_{mn} = 1/w_{mn}$  naturally expresses their relative distance in the network. Once  $\{l_{mn}\}$  is given, it can be used to calculate the matrix of the shortest path length  $d_{mn}$  between two generic nodes  $i$  and  $j$ . The so-called clustering coefficient  $C$  and the global efficiency  $E$  of the graph can then be calculated [22]. The clustering coefficient  $C$  is also considered as the first approximation of the local efficiency. By using those efficiencies, both the local and global behaviors in small-world networks can be studied [19, 20]. The large  $C$  values and small  $E$  values in the early network indicate that the early data are locally clustered [see example in fig. 2(c)]. It is proven that the local verification and validation of the data are effective, where data of same isotope and similar incident energy are compared. This local method has been widely applied to date. With the growth of the networks, the clustering coefficients  $C$  decrease, while the global efficiencies  $E$  increase. The networks generated by the latest experimental data sets have  $C = 0.58$  and  $E = 0.40$  in the  $(n,2n)$  network, and  $C = 0.65$

and  $E = 0.45$  in the  $(\gamma,xn)$  network. These results indicate that each region in the data-network is intermingled with the others, and hence, the verification and validation of the data can be performed either locally or globally.

Figure 3(e) is presented with the goal to investigate the Heaps law. The Heaps law was originally introduced to describe the number of distinct words in a text document [31]. Thereafter, this statistical property was observed in other real data of innovation processes by empirical analyses [28–30]. Various models have proven that the Heaps law well describes the pace at which scientists discover concepts or users collect new items [32, 33]. In this work, the three order missing tensor is applied to represent our knowledge of the cross-sections in the nuclear reaction. The number of novelties  $S$  corresponds to the number of the entries in the tensor that has been observed. The Heaps law is then  $S \propto N_{\text{node}}^\beta$ , where  $N_{\text{node}}$  is the node number in the network. As shown in Fig. 3(e), the Heaps laws are found in discovery processes of nuclear reaction data, with the Heaps exponents  $\beta_{(n,2n)} = 0.77$  for the  $(n,2n)$  reaction and  $\beta_{(\gamma,xn)} = 0.90$  for the  $(\gamma,xn)$  reaction. A higher value of the Heaps exponent  $\beta$  denotes a faster exploration of the adjacent possibilities in the measurements of the cross-section in the  $(\gamma,xn)$  reaction. Being limited by the monoenergetic neutron source, the exploration of the innovative data (new isotope or new incident energy) for the  $(n,2n)$  reaction is slower than that for the  $(\gamma,xn)$  reaction.

One of the objectives of generating networks in this work is data evaluation. In the traditional methods, to evaluate a new experimental data point, one may compare it with historical measurements, predictions in the data libraries, and calculations by theoretical models for the same reaction. However, the global and hidden uncertainties of the historical measurements may have propagated in the data libraries and theoretical models, which will mislead the evaluation such that the uncertainties will remain undetected in the database. This is a positive feedback process that may last for decades before it is discovered [18]. The data-networks based on the Bayesian statistics approach reveal the relative deviations between data points. The weight  $w_{mn}$  of an edge is calculated by the posterior values of the hyper-parameter  $\hat{\mu}$  recommended by the  $m$ -th and  $n$ -th data points. The  $w_{mn}$  value will be 1 if two data points are the same but close to 0 for two data points with a huge discrepancy. This definition of the relative deviation expands the data range for direct comparison. Traditionally, only data points for the same isotope and similar incident energy can be compared.

In the evolution of weight distribution, shown in Fig. 4, not only are new edges linked to the network, but the weights of the original edges will also change due to the appearance of new nodes. This is a universal law in weighted evolving networks deduced by the coupling topology and weight dynamics [21]. These dynamics are naturally reproduced by the iterative computation in the Bayesian statistics-based approach. This may be very meaningful, since a few abnormal data points mean a large uncertainty in the measuring technique, but many abnormal data points reveal a novel mechanism.

The distribution of the weights in the network is an appropriate quantity to evaluate the global uncertainty of the data set, while the mean node weight  $w_m = s_m/k_m$  can

be applied to estimate the uncertainty of the individual data. It is noted here that the mentioned uncertainty is beyond the systematic and statistical errors, which are published with the data. Even the experimental data points with small published errors may be quite different from other data for the same nuclear reaction. This means that there is a global and hidden uncertainty, which comes from the limitations of the measurement technology and theoretical knowledge.

As shown in Fig. 4, both networks show peaks near  $w_{mn} = 0.9$  at different periods except for the  $(\gamma, xn)$  reaction in 1970. However, different shapes are observed for the two reactions. In the case of the  $(n, 2n)$  reaction, wide distributions are observed. For the  $(\gamma, xn)$  reaction, narrow distributions near  $w_{mn} = 0.9$  are observed after 1973. These results correspond to the fact that the global uncertainty in the database of the  $(n, 2n)$  reaction is larger than that of the  $(\gamma, xn)$  reaction.

The correlations between mean node weight  $w_m$  and node degree  $k_m$  are shown in Fig. 5. In the region with a good deal of existing data, the data points can be mutually verified before publishing. Hence, in the  $w_m v.s. k_m$  map, the nodes with large numbers of neighbors are concentrated in the region with large  $w_m$  values, while those with a few neighbors are distributed in a wide region from  $w_m = 0.3$  to 0.8. Figure 6 shows the distribution of the  $w_m$  values for the  $(n, 2n)$  and  $(\gamma, xn)$  reactions. The uncertainty in the  $(n, 2n)$  database is larger than that for the  $(\gamma, xn)$  reaction. A more visual estimator for the individual data is the subgraph, four examples of which are shown in the embedded illustrations in Fig. 6. In the subgraph, the data point to be evaluated is shown as the center node, and its neighbors are displayed around it. The distance from the center node to its neighbor is defined as  $\ln(1/w_{mn})$ . The color of the node indicates its number of neighbors. The subgraph for a data point with a small  $w_m$

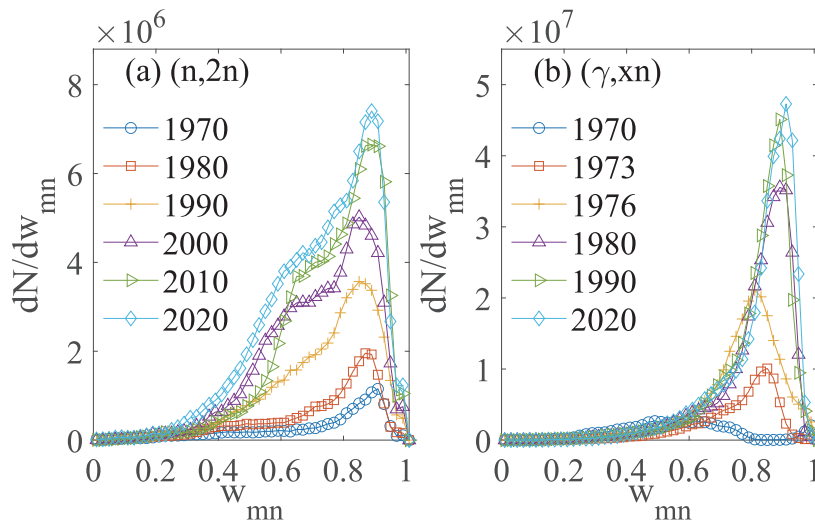
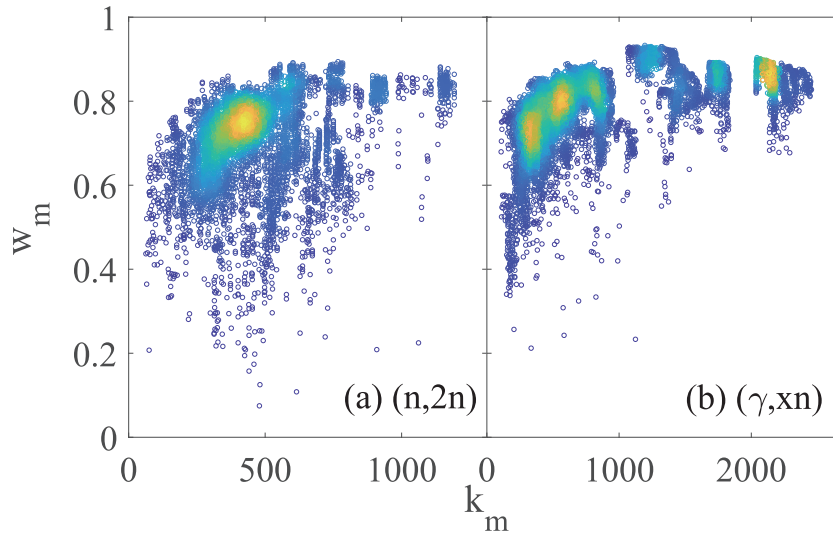
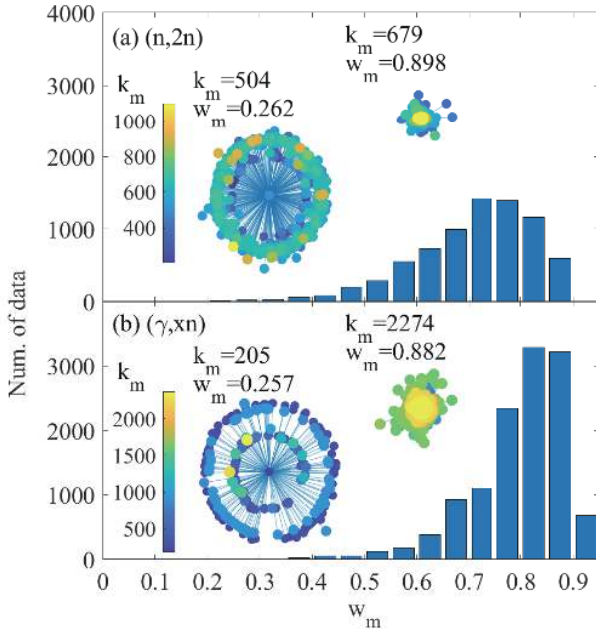


Fig. 4. (color online) Distribution of the edge-weight  $w_{mn}$  in the networks.



**Fig. 5.** (color online) Correlation between mean node weight  $w_m$  and node degree  $k_m$ . The color is a visual guide for the point-density.



**Fig. 6.** (color online) Distribution of the  $w_m$  value for the (a)  $(n,2n)$  and (b)  $(\gamma,xn)$  reactions. The embedded illustrations are examples of the estimator, where the data to be evaluated are shown as the center node, and its neighbors are displayed around it. The distance from the center node to its neighbor is defined as  $\ln(1/w_m)$ . The color of the node indicates its number of neighbors.

value, and hence large uncertainty, is like a blooming flower, as its distances to other nodes in the network are huge. In contrast, a data point with small uncertainty will huddle in the network.

#### IV. CONCLUSION

In summary, based on a Bayesian statistics-based ap-

proach, a model is developed to build networks for discovery processes of nuclear reaction data. After the incident energy degree of freedom is discretized, the data are recorded by a list of three order missing tensors and the data evaluation is constructed as a problem of tensor decomposition and imputation with multiple observations on an entry. To solve this problem, the Bayesian tensor decomposition approach by Chen *et al.* [43] is extended. Case studies of cross-sections in the neutron induced threshold reaction  $(n,2n)$  and photoneutron reaction  $(\gamma,xn)$  are presented to build the weighted evolving networks, where the nodes are the historical data, and the weights of the links are the relative deviation between the data points. It is found that the networks exhibit small-world behavior, and their dynamics are well described by the Heaps law, which has been widely observed in other real networks of innovation processes. What makes the networks novel is the mapping relation between the properties in the network and the salient features of the database. It is this relation that makes it possible to (i) quantify the exploration efficiency of the specific data set by the Heaps exponent, (ii) evaluate the global uncertainty of the data set by the distribution of the edge-weights, and (iii) visualize and quantify the uncertainty of the individual data point by the mean node weight.

The network built in this work is a new perspective to understand the database, which is helpful for nuclear data analysis and compilation as well as quality improvement of the database. Future works can focus on studying the effect of the uncertainty distribution that is similar for the data measured in the same period but changes with the development of the experimental technique. The idea of noise modeling in Ref. [47] is an enlightening way of feeding the extracted uncertainty back to the likelihood function.

## DECLARATION OF INTEREST

The authors declare that they have no known compet-

ing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] M. Salvatores, A Report by the Working Party on International Evaluation Cooperation of the Nuclear Science Committee, (2008)
- [2] N. Otuka, E. Dupont, V. Semkova *et al.*, *Nuclear Data Sheets* **120**, 272-276 (2014)
- [3] S. Goriely, S. Hilaire, and A. J. Koning, *Astronomy & Astrophysics* **487**, 767-774 (2008)
- [4] A. Tumino, C. Spitaleri, M. L. Cognata *et al.*, *Nature* **557**, 687-690 (2018)
- [5] G. Aliberti, G. Palmiotti, M. Salvatores *et al.*, *Nuclear Science and Engineering* **146**, 13-50 (2004)
- [6] G. Aliberti, G. Palmiotti, M. Salvatores *et al.*, *Annals of Nuclear Energy* **33**, 700-733 (2006)
- [7] V. Semkova, N. Otuka, M. Mikhaïliukova *et al.*, *Exfor—a global experimental nuclear reaction data repository: Status and new developments*, in: EPJ Web of Conferences, volume 146, EDP Sciences, 2017, p. 07003
- [8] D. A. Brown, M. Chadwick, R. Capote *et al.*, *Nuclear Data Sheets* **148**, 1-142 (2018)
- [9] Z. Ge, R. Xu, H. Wu *et al.*, *Cendl-3.2: The new version of chinese general purpose evaluated nuclear data library*, in: EPJ Web of Conferences, volume 239, EDP Sciences, 2020, p. 09001
- [10] A. Plompen, O. Cabellos, C. De Saint Jean *et al.*, *The European Physical Journal A* **56**, 1-108 (2020)
- [11] K. Shibata, O. Iwamoto, T. Nakagawa *et al.*, *Journal of Nuclear Science and Technology* **48**, 1-30 (2011)
- [12] A. Blokhin, E. Gai, A. Ignatyuk *et al.*, *Yad. Reak. Konst* **2**, 62 (2016)
- [13] M. Salvatores, G. Palmiotti, G. Aliberti *et al.*, *Nuclear Data Sheets* **118**, 38-71 (2014)
- [14] C. De Saint Jean, P. Archier, E. Privas *et al.*, *On the use of bayesian monte-carlo in evaluation of nuclear data*, in: EPJ Web of Conferences, volume 146, EDP Sciences, 2017, p. 02007
- [15] E. Alhassan, D. Rochman, A. Vasiliev *et al.*, *Iterative bayesian monte carlo for nuclear data evaluation*, arXiv preprint arXiv: 2003.10827 (2020)
- [16] Z. Niu and H. Liang, *Physics Letters B* **778**, 48-53 (2018)
- [17] Z.-A. Wang, J. Pei, Y. Liu *et al.*, *Physical review letters* **123**, 122501 (2019)
- [18] A. Koning, *Statistical verification and validation of the exfor database: (nn), (n, 2n), (n, p), (n, a) and other neutron-induced threshold reaction cross-sections*, 2014
- [19] D. J. Watts and S. H. Strogatz, *nature* **393**, 440-442 (1998)
- [20] V. Latora and M. Marchiori, *Physical review letters* **87**, 198701 (2001)
- [21] A. Barrat, M. Barthélemy, and A. Vespignani, *Physical review letters* **92**, 228701 (2004)
- [22] S. Boccaletti, V. Latora, Y. Moreno *et al.*, *Physics reports* **424**, 175-308 (2006)
- [23] R. D. Gray and Q. D. Atkinson, *Nature* **426**, 435-439 (2003)
- [24] D. H. Erwin and D. C. Krakauer, *Science* **304**, 1117 (2004)
- [25] F. Saracco, R. Di Clemente, A. Gabrielli *et al.*, *PloS one* **10**, e0140420 (2015)
- [26] M. Andjelković, B. Tadić, M. Mitrović Dankulov *et al.*, *PloS one* **11**, e0154655 (2016)
- [27] B. Hofstra, V. V. Kulkarni, S. M.-N. Galvez *et al.*, *Proceedings of the National Academy of Sciences* **117**, 9284-9291 (2020)
- [28] C. Cattuto, V. Loreto, and L. Pietronero, *Proceedings of the National Academy of Sciences* **104**, 1461-1464 (2007)
- [29] F. Tria, V. Loreto, V. D. P. Servedio *et al.*, *Scientific reports* **4**, 1-8 (2014)
- [30] B. Monechi, A. Ruiz-Serrano, F. Tria *et al.*, *PloS one* **12**, e0179303 (2017)
- [31] H. S. Heaps, *Information retrieval, computational and theoretical aspects*, Academic Press, 1978
- [32] I. Iacopini, S. Milojević, and V. Latora, *Physical review letters* **120**, 048301 (2018)
- [33] I. Iacopini, G. Di Bona, E. Ubaldi *et al.*, *Physical Review Letters* **125**, 248301 (2020)
- [34] A. Rzhetsky, J. G. Foster, I. T. Foster *et al.*, *Proceedings of the National Academy of Sciences* **112**, 14569-14574 (2015)
- [35] R. Sinatra, D. Wang, P. Deville *et al.*, *Science*, 354 (2016)
- [36] S. Fortunato, C. T. Bergstrom, K. Börner *et al.*, *Science*, 359 (2018)
- [37] X. Li, Y. Ye, and X. Xu, *Low-rank tensor completion with total variation for visual data inpainting*, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 31, 2017
- [38] B. Ran, H. Tan, J. Feng *et al.*, *Computational intelligence and neuroscience*, 2015 (2015)
- [39] J. Liu, P. Musialski, P. Wonka *et al.*, *IEEE transactions on pattern analysis and machine intelligence* **35**, 208-220 (2012)
- [40] R. Salakhutdinov and A. Mnih, *Bayesian probabilistic matrix factorization using markov chain monte carlo*, in: *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 880-887
- [41] T. G. Kolda and B. W. Bader, *SIAM review* **51**, 455-500 (2009)
- [42] L. Xiong, X. Chen, T.-K. Huang *et al.*, *Temporal collaborative filtering with bayesian probabilistic tensor factorization*, in: *Proceedings of the 2010 SIAM international conference on data mining*, SIAM, 2010, pp. 211-222
- [43] X. Chen, Z. He, and L. Sun, *Transportation research part C: emerging technologies* **98**, 73-84 (2019)
- [44] J. L. Fowler and J. Slye Jr, *Physical Review* **77**, 787 (1950)
- [45] J. Frehaut, A. Bertin, R. Bois *et al.*, CEA Centre d'Etudes de Bruyeres-le-Chatel, (1980)
- [46] R. Edge, *Australian Journal of Physics* **9**, 429-435 (1956)
- [47] X. Chen, Z. Han, Y. Wang *et al.*, *A general model for robust tensor factorization with unknown noise*, arXiv preprint arXiv: 1705.06755 (2017)