Editors' Suggestion | Featured in Physics

Efficient inference of rankings from multibody comparisons

Jack Yeung, Daniel Kaiser, and Filippo Radicchi*
Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana 47408, USA

(Received 1 May 2025; accepted 20 May 2025; published 9 July 2025)

Many of the existing approaches to assess and predict the performance of players, teams, or products in competitive contests rely on the assumption that comparisons occur between pairs of such entities. There are, however, several real contests where more than two entities are part of each comparison, e.g., sports tournaments, multiplayer board and card games, and preference surveys. The Plackett-Luce (PL) model provides a principled approach to infer the ranking of entities involved in such contests characterized by multibody comparisons. Unfortunately, traditional algorithms used to compute PL rankings suffer from slow convergence, which limits the application of the PL model to relatively small-scale systems. Here, we present an alternative implementation that allows for significant speed-ups and validate its efficiency in both synthetic and real-world sets of data. Further, we perform systematic cross-validation tests concerning the ability of the PL model to predict unobserved comparisons. We find that a PL model trained on a set composed of multibody comparisons is more predictive than a PL model trained on a set of projected pairwise comparisons derived from the very same training set, emphasizing the need of properly accounting for the true multibody nature of real-world systems whenever such information is available.

DOI: 10.1103/64h7-z39p

I. INTRODUCTION

Competitive contests in the real world often involve multiple comparisons between pairs of entities, with each comparison characterized by an outcome in favor of one or the other entity. Examples include tournaments in sports like tennis or boxing and competitions in board games like Scrabble or chess. A natural representation for this type of contest is a directed and weighted network or graph [1]. Entities (e.g., players, teams) are represented by the nodes of the graph. Each edge denotes a comparison (e.g., game) between two nodes. The direction of an edge represents a specific outcome of the corresponding comparison, whereas its integer-valued weight quantifies the occurrence of that specific outcome in the contest. There exist several graph-based approaches that leverage the graph representation of a competitive contest, to assess past performance and predict the future performance of individual entities [2–5].

Among the existing approaches to asses and predict performance in pairwise competitive contests, the Bradley-Terry (BT) model plays a significant role [6]. This is a principled probability model for the outcome of pairwise comparisons between entities. Given a set of observed comparisons between pairs of entities, the BT model allows to derive scores associated with individual entities. Those scores can be used to rank entities as well as predict the outcomes of nonobserved comparisons. The graph representation of the contest is essential to derive the properties of the BT model. For example, inferring BT scores is possible through an iterative algorithm whose convergence is guaranteed only if the graph representing the contest is strongly connected [7]. The BT model has been the subject of many papers and has been applied in the analysis of a large variety of systems [8,9]. However, only very recently, Newman has shown that the computational efficiency of inferring the BT scores from observed data can be greatly improved if the equations used in the original formulation by Zermelo are slightly re-arranged [8]. Typical speed-up factors range between 3 and 100 depending on the set of data considered [8].

Although competitive contests based on pairwise comparisons are abundant, there are also many real competitive contests where more than two entities are involved in each comparison. In football competitions, such as the final tournament of the FIFA World Cup or the UEFA Champions League, there is an initial group-stage round, and only the top teams in a group qualify for the following direct elimination rounds [10,11]. In a Texas Hold'em Poker tournament, rounds consist of tables composed of several players [12]. Many online video games are played by multiple players simultaneously [13]. It is often the case that multiple products (e.g., songs, movies, books, and videos) compete at the same time on a market [14]. A scientific paper, where the order of the authorship reflects the contribution of each individual to the paper, is often written by more than two authors [15].

Such systems based on multibody comparisons are naturally represented by hypergraphs, where hyperedges are vectors denoting the outcome of individual comparisons eventually weighted by the frequency of their occurrence [16]. Existing graph-based methods can be adapted to deal with such systems upon prior reduction of the hypergraph to a properly devised graph. However, such an operation is not

*Contact author: f.radicchi@gmail.com

only responsible for the loss of potentially useful information, but also not specifically tailored for the prediction of the outcome of multibody comparisons. Based on our knowledge, the Plackett-Luce (PL) model is one of the few principled approaches existing in the literature that can fully account for the ground-truth multibody structure of the underlying competition. The PL model is a probabilistic model that represents the natural extension of the BT model from pairwise to multibody comparisons. The model was independently introduced by Luce [17] and Plackett [16]. A large literature on the PL model exists, including spectral algorithms to approximate solutions of the PL model [18], alternative approaches to evaluating the PL model such as a generalized method-of-moments [19] and other iterative approaches [20], and shortcomings of the model in multimodal rank spaces [21]. Foundational work on choice theory and decision-making processes and their connection to the PL model has also been conducted; we refer the reader to the classic collection of essays edited by A. A. J. Marley [22]. The PL model is also central in many existing machine learning approaches to the so-called learning-to-rank tasks [23–25].

In this paper, we first present the generalization of the computationally efficient algorithm proposed by Newman for the BT model to the PL model. We present a detailed derivation of the iterative algorithm and provide evidence that such an algorithm displays significant convergence speed-up compared to the usual iteration schema for solving the PL model. Second, we apply the PL model to the analysis of several real-world systems characterized, at least in principle, by multibody comparisons. These sets of data include: (i) results of national football teams in the FIFA World Cup, (ii) results of football club teams in the UEFA Champions League, (iii) lists of authors of papers in the area of Network Science, (iv) elections of members in the American Psychological Association, (v) two surveys concerning sushi preferences, and (vi) a survey concerning the preference for courses at the AGH University of Science and Technology. For elections, we find that projecting multibody comparisons into pairwise comparisons does not lead to any loss in the predictive power of the PL model. The finding poses some doubts on the authenticity of the multibody nature of this set of data. For the other datasets, we find instead that their multibody representation is genuine, as the descriptions/predictions based on multibody comparisons are generally more accurate than their projected pairwise counterparts. Finally, we use the scores inferred by the PL model to generate top-10 rankings of football teams and network scientists.

II. METHODS

A. Ordered and weighted hypergraphs

We consider systems of N elements/entities where M interactions (e.g., games, rounds, collaborations, and comparisons) occur between an arbitrary number $2 \leqslant K \leqslant N$ of elements, and the simultaneous presence of interactions of different size K is allowed. We denote with $K_{\min}$ and $K_{\max}$ respectively the minimum and the maximum number of elements involved in a single interaction. These systems can be considered as hypergraphs [26], with the peculiarity that their hyperedges are ordered and weighted. An ordered hyperedge is a vector of nodes' labels, i.e., $\vec{\omega} = (\omega_1, \omega_2, \ldots, \omega_K)$, representing for example the rank positions of teams in a group-stage round of a soccer tournament or the list of authors in a scientific paper. The weight $z(\vec{\omega})$ denotes the number of occurrences of the specific hyperedge $\vec{\omega}$, e.g., the number of scientific articles with identical lists of authors. Denoting with Ω the set of all distinct hyperedges, we can write $M = \sum_{\vec{\omega}\in\Omega} z(\vec{\omega})$. Note that if all hyperedges are of size $K = 2$, then the corresponding ordered and weighted hypergraph reduces to a directed and weighted graph.

Different from standard hypergraphs, two ordered hyperedges involving the very same sets of nodes are considered different if the involved nodes are listed in a different order in the two vectors. Hypergraphs with vector representations of hyperedges were previously considered, among others, in the study of folksonomies [27,28]. There, each hyperedge was composed of exactly $K = 3$ components, namely a user, an object, and a tag. In our case, however, we do not distinguish types of nodes. Also, as we already mentioned, we do not require vectors to have the same number of components.

B. Plackett-Luce model

We consider the so-called Plackett-Luce (PL) model, representing the immediate generalization of the Bradley-Terry (BT) model from directed graphs to ordered hypergraphs

$$P(\vec{\omega}|\vec{\pi}) = \prod_{r=1}^{K-1} \frac{\pi_{\omega_r}}{\sum_{q=r}^{K} \pi_{\omega_q}}. \tag{1}$$

Equation (1) represents the probability of observing the hyperedge $\vec{\omega}$ given the vector of nodes' scores $\vec{\pi} = (\pi_1, \ldots, \pi_i, \ldots, \pi_N)$. Here, $\pi_i = e^{s_i} > 0$ is a numerical value associated with the element i that quantifies the importance/strength of the element i with respect to all other elements in the system. The scores $\vec{\pi}$ are further subject to the normalization condition

$$\prod_{i=1}^{N} \pi_i = 1. \tag{2}$$

Equation (1) reduces to the BT model for $K = 2$, i.e., $P(\omega_1, \omega_2|\vec{\pi}) = \pi_{\omega_1}/(\pi_{\omega_1} + \pi_{\omega_2})$. Note that the model of Eq. (1) is invariant to any transformation $\pi_i \to a\pi_i$ for $i = 1, \ldots, N$ and $\forall a > 0$, and the normalization of Eq. (2) serves to fix the value of the constant a.

The nodes' scores $\vec{\pi}$ must be learned from an observed hypergraph. Given a set of observed hyperedges Ω, the best estimates of the nodes' scores $\vec{\pi}$ are obtained by maximizing the likelihood

$$P(\Omega|\vec{\pi}) = \prod_{\vec{\omega}\in\Omega} [P(\vec{\omega}|\vec{\pi})]^{z(\vec{\omega})}. \tag{3}$$

Such an optimization can be performed by actually maximizing the logarithm of the likelihood, i.e.,

$$\mathcal{L}(\Omega|\vec{\pi}) = \log P(\Omega|\vec{\pi}) = \sum_{\vec{\omega}\in\Omega} z(\vec{\omega}) \log [P(\vec{\omega}|\vec{\pi})]. \tag{4}$$

Indicate with $\Omega_s^{K,r}$ the set of vectors of length K where node s appears at position r, i.e., $\Omega_s^{K,r} = \{\vec{\omega} \in \Omega \,|\, \dim(\vec{\omega}) = K \wedge \omega_r = s\}$. We can write the portion of the log-likelihood of Eq. (4) that contains all terms that depend on the variable π_s as

$$\mathcal{L}_s(\Omega|\vec{\pi}) = \sum_{K=2}^{N}\sum_{r=1}^{K}\sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})\left[\log(\pi_s) - \sum_{v=1}^{r}\log\left(\sum_{i=v}^{K}\pi_{\omega_i}\right)\right], \tag{5}$$

We note that $\frac{\partial\mathcal{L}}{\partial\pi_s}(\Omega|\vec{\pi}) = \frac{\partial\mathcal{L}_s}{\partial\pi_s}(\Omega|\vec{\pi})$; thus, to maximize the likelihood, we take the derivatives of Eq. (5) with respect to π_s for each $s = 1, \ldots, N$ and set them equal to zero, i.e.,

$$\frac{\partial\mathcal{L}_s}{\partial\pi_s}(\Omega|\vec{\pi}) = \sum_{K=2}^{N}\sum_{r=1}^{K}\sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})\left(\frac{1}{\pi_s} - \sum_{v=1}^{r}\frac{1}{\sum_{i=v}^{K}\pi_{\omega_i}}\right) = 0, \tag{6}$$

which defines a system of N equations and N unknowns. Such a system of equations has no closed-form solution but can be solved by iteration.

The most straightforward iterative algorithm is based on

$$\pi_s' = \frac{\sum_{K=2}^{N}\sum_{r=1}^{K}\sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})}{\sum_{K=2}^{N}\sum_{r=1}^{K}\sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})\sum_{v=1}^{r} 1\Big/\sum_{i=v}^{K}\pi_{\omega_i}}. \tag{7}$$

The above equation represents the generalization to hypergraphs of the iterative algorithm introduced by Zermelo for graphs [7]. Solutions are obtained by starting from some initial vector $\vec{\pi}$ to compute the updated vector $\vec{\pi}'$ using Eq. (7); the vector $\vec{\pi}'$ is then normalized according to Eq. (2). These two operations are iterated until convergence. In the iterative algorithm, we consider cyclic asynchronous updating, where each score π_s is updated in order. One iteration of the algorithm is given by a full update of all N scores π_s. Asynchronous updates are used to minimize the possibility of the system of equations entering a limit cycle; this is a relatively rare situation, and synchronous updating typically leads to the same solution as asynchronous updating. At the end of each iteration, we determine the convergence of the algorithm by testing whether the inequality

$$A = \sqrt{\frac{1}{N}\sum_{s=1}^{N}\left(\frac{\pi_s}{1+\pi_s} - \frac{\pi_s'}{1+\pi_s'}\right)^2} \leqslant \epsilon \tag{8}$$

is satisfied or not. In the above equation, $\frac{\pi_s}{1+\pi_s}$ is the probability that node s beats the average node in a pairwise comparison [8].The score of the average player is 1, as follows directly from the normalization condition of Eq. (2). We estimate this quantity for each node between consecutive iterations, and quantify A as the mean square of such an error across all elements in the system. In all our numerical calculations, we use $\epsilon = 10^{-6}$.

As for the case of the standard BT model on graphs, the iterative procedure of Eq. (7) displays slow convergence, meaning that several iterations are required before the condition of Eq. (8) is met. For the BT model, the issue can be addressed by writing a different iterative algorithm as shown by Newman [8]. We generalize the iterative algorithm by Newman to the PL model as follows. First, we rewrite Eq. (6) as

$$\sum_{K=2}^{N}\sum_{r=1}^{K}\sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})\left(\frac{1}{\pi_s} - \frac{1}{\sum_{i=r}^{K}\pi_{\omega_i}} - \sum_{v=1}^{r-1}\frac{1}{\sum_{i=v}^{K}\pi_{\omega_i}}\right) = 0.$$

Then, we notice that for any $\vec{\omega} \in \Omega_s^{K,r}$ we have by definition that $\omega_r = s$, thus for each non-null summand in the above equation, we can write

$$\frac{1}{\pi_s} - \frac{1}{\sum_{i=r}^{K}\pi_{\omega_i}} = \frac{\sum_{i=r}^{K}\pi_{\omega_i} - \pi_s}{\pi_s\sum_{i=r}^{K}\pi_{\omega_i}} = \frac{\sum_{i=r+1}^{K}\pi_{\omega_i}}{\pi_s\sum_{i=r}^{K}\pi_{\omega_i}}$$

for $r \leqslant K-1$. For $r = K$, we automatically have $\frac{1}{\pi_s} - 1/\sum_{i=r}^{K}\pi_{\omega_i} = 0$. Finally, we write the iterative step of the Newman's algorithm for the PL model as

$$\pi_s' = \frac{\sum_{K=2}^{N}\sum_{r=1}^{K-1}\sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})\sum_{i=r+1}^{K}\pi_{\omega_i}\Big/\sum_{i=r}^{K}\pi_{\omega_i}}{\sum_{K=2}^{N}\sum_{r=1}^{K}\sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})\sum_{v=1}^{r-1} 1\Big/\sum_{i=v}^{K}\pi_{\omega_i}}. \tag{9}$$

The resulting iterative algorithm requires first the application of Eq. (9) and then the imposition of the normalization condition of Eq. (2). Still, convergence is assessed using Eq. (8).

Both iterative procedures of Eqs. (7) and (9) may not converge. The convergence issue arises because they are obtained by maximizing a likelihood function instead of a posterior probability. The necessary condition for convergence of the traditional BT model is that the underlying graph is composed of a single strongly connected component [7]. The same criterion generalizes to the PL model [18], where the graph that must be strongly connected is the one obtained by properly projecting the hypergraph into its so-called "full breaking" graph [19], that is, where every possible directed dyadic edge is included to preserve the strict total ordering implicit in the multibody comparisons. The above condition for convergence requires projecting the higher-order network into a dyadic network, essentially reducing the PL model to the BT model. This allows us to state that if the full-breaking projection of the hypergraph is strongly connected, then the fast iterative schema converges, based on Newman's proof for the BT model [8].

To address any potential convergence issue that might emerge from the absence of the above necessary condition, we adopt the same strategy as in Newman [8], and assume that the prior for the scores $\vec{\pi}$ is the logistic distribution

$$P(\vec{\pi}) = \prod_{i=1}^{N}\frac{\pi_i}{(1+\pi_i)^2}, \tag{10}$$

so that the posterior probability of the model reads

$$P(\vec{\pi}\,|\Omega) \propto P(\Omega|\vec{\pi})\,P(\vec{\pi}) = \prod_{\vec{\omega}\in\Omega}[P(\vec{\omega}|\vec{\pi})]^{z(\vec{\omega})}\prod_{i=1}^{N}\frac{\pi_i}{(1+\pi_i)^2}. \tag{11}$$

The maximization of the posterior probability is performed using a similar procedure as the one shown above for

the likelihood. The analog of the Zermelo's iterative step is

$$\pi_s' = \frac{1 + \sum_{K=2}^{N} \sum_{r=1}^{K} \sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})}{2/(\pi_s+1) + \sum_{K=2}^{N} \sum_{r=1}^{K} \sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega}) \sum_{v=1}^{r} 1/\sum_{i=v}^{K} \pi_{\omega_i}}. \tag{12}$$

The Newman's iterative procedure is instead

$$\pi_s' = \frac{1/(\pi_s+1) + \sum_{K=2}^{N} \sum_{r=1}^{K-1} \sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega}) \sum_{i=r+1}^{K} \pi_{\omega_i}/\sum_{i=r}^{K} \pi_{\omega_i}}{1/(\pi_s+1) + \sum_{K=2}^{N} \sum_{r=1}^{K} \sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega}) \sum_{v=1}^{r-1} 1/\sum_{i=v}^{K} \pi_{\omega_i}}. \tag{13}$$

As already stressed by Newman, the prior of Eq. (10) serves as a regularization term that allows for the convergence of the iterative scheme irrespective of the connectedness of the underlying graph. Once more, this was proven by Newman for the BT model, but it should be valid for the PL model too as the prior probability *de facto* makes the projected dyadic graph strongly connected. Further, imposing the normalization condition of Eq. (2) is no longer required; applying the normalization condition to the scores obtained via Eq. (13) is actually ill-suited as the scores are no longer scale invariant. In our application of the method, however, we still apply it after each iteration. We noticed in fact that this helps speeding up the convergence of the iterative algorithm, see Appendix. Further, numerical estimates of scores obtained by imposing the normalization condition at each iteration of the algorithm are basically indistinguishable from those obtained without imposing the normalization condition.

C. Position 1-breaking Plackett-Luce model

Also, we consider

$$P(\vec{\omega}|\vec{\pi}) = \frac{\pi_{\omega_1}}{\sum_{q=1}^{K} \pi_{\omega_q}}, \tag{14}$$

which is a slight modification of the PL model of Eq. (1) known in the literature as the position 1-breaking PL model [19]. Note that since $\pi_i = e^{s_i}$, the probability appearing in Eq. (14) is often referred to as the multinomial logistic or softmax function. The model of Eq. (14) gives emphasis to the node occupying the first position in the vector hyperedge $\vec{\omega}$ only, irrespective of the order in which the other nodes appear in the hyperedge.

We can repeat the same exact steps as above to generate an efficient algorithm for determining the scores $\vec{\pi}$ of each node in the graph for a given set of observed hyperedges Ω. The Newman's iterative equation for this model is

$$\pi_s' = \frac{1/(\pi_s+1) + \sum_{K=2}^{N} \sum_{\vec{\omega}\in\Omega_s^{K,1}} z(\vec{\omega}) \sum_{i=2}^{K} \pi_{\omega_i}/\sum_{i=1}^{K} \pi_{\omega_i}}{1/(\pi_s+1) + \sum_{K=2}^{N} \sum_{r=2}^{K} \sum_{\vec{\omega}\in\Omega_s^{K,r}} z(\vec{\omega})/\sum_{i=1}^{K} \pi_{\omega_i}}. \tag{15}$$

D. Approximating many-body interactions with pairwise interactions

An immediate question that can arise is whether the hypergraph-based PL model can be well approximated by properly modified versions of the standard graph-based BT model. An approximation of this type requires generating a graph projection of the hypergraph, by mapping each vector hyperedge to a set of directed edges, and then apply the standard BT model to the graph projection. Please note that the projection should properly account for any eventual weight associated with each hyperedge.

The natural approximation of the model of Eq. (1) is

$$\tilde{P}(\vec{\omega}|\vec{\pi}) = \prod_{r=1}^{K-1} \prod_{t=r}^{K} \frac{\pi_{\omega_r}}{\pi_{\omega_r} + \pi_{\omega_t}}, \tag{16}$$

which actually consists of mapping the vector hyperedge $\vec{\omega} = (\omega_1, \ldots, \omega_K)$ to the set of directed edges $\{(\omega_1, \omega_2), \ldots, (\omega_1, \omega_K), (\omega_2, \omega_3), \ldots, (\omega_2, \omega_K), \ldots, (\omega_{K-1}, \omega_K)\}$. We refer to the model of Eq. (16) as the projected Plackett-Luce (pPL) model. The above projection of the hypergraph is named "full breaking" graph in Ref. [19].

For the model of Eq. (14), we have instead

$$\tilde{P}(\vec{\omega}|\vec{\pi}) = \prod_{r=2}^{K} \frac{\pi_{\omega_1}}{\pi_{\omega_1} + \pi_{\omega_r}}, \tag{17}$$

consisting of mapping the vector hyperedge $\vec{\omega} = (\omega_1, \ldots, \omega_K)$ to the set of directed edges $\{(\omega_1, \omega_2), \ldots, (\omega_1, \omega_K)\}$. Also in this case, we refer to the model of Eq. (17) as the position 1-breaking version of the pPL model. The above projection of the hypergraph is named "position-1 breaking" graph in Ref. [19].

III. DATA

A. Synthetic hypergraphs

We consider synthetic ordered and weighted hypergraphs generated as follows. Inputs are the number of nodes N, the total number of multibody interactions M, as well as the range $[K_{\min}, K_{\max}]$ for the size of those interactions. We first assign a score π_s to each node $s = 1, \ldots, N$ by extracting a random value from the logistic distribution of Eq. (10). Then, to generate one interaction, we first determine its size K by extracting a random number at uniform from the interval $[K_{\min}, K_{\max}]$. Then, we extract, at uniform and without replacement, K nodes out of the N available; those nodes are ordered at random according to either Eq. (1) or Eq. (14) to generate the actual hyperedge $\vec{\omega}$; while building the network, we keep track of the weight $z(\vec{\omega})$ associated with each hyperedge $\vec{\omega}$.

B. FIFA World Cup

We analyze data concerning the outcome of the different rounds played by football national teams in all 22 FIFA World Cup (FWC) final tournaments held between 1930 and 2022.

Data are retrieved from the Wikipedia pages of the individual tournaments, see for example [29] for data concerning the FWC tournament of Spain 1982. The dataset consists of $N = 83$ total teams, with the caveat that we merge both "West Germany" and "East Germany" into the single team "Germany" between 1954 and 1990. The dataset is composed of 364 total rounds of rather different types, given that the format of the tournament changed dramatically in the course of the years. The format of the most recent tournaments consists of an initial round played in groups of four teams, with the first two teams of each group proceeding to the following direct elimination rounds. There were, however, tournaments consisting of only direct elimination rounds (e.g., Italy 1934) and with more than one group-stage round (e.g., Spain 1982). Each round is considered as an ordered hyperedge in our analysis. The dataset contains $M = 364$ hyperedges, with size ranging between $K_{\min} = 2$ and $K_{\max} = 4$.

C. UEFA Champions League

We consider data reporting the outcome of all rounds played in the UEFA Champions League (UCL) tournaments between seasons 1992–1993 and 2023–2024. Preliminary/qualification rounds are excluded. Our sources of data are the Wikipedia pages of the individual editions, for data concerning the UCL tournament of season 2022–2023 see [30]. The resulting dataset consists of $N = 174$ teams and $M = 674$ rounds of size between $K_{\min} = 2$ and $K_{\max} = 4$. Also in this case, the tournament structure changed drastically in the course of the years. For example, the 1992–1993 tournament consisted of two direct elimination rounds, a round with groups of size four, and a final direct elimination round. In 2002–2003, there were two rounds with groups of size four, and then direct elimination rounds. In the most recent tournaments, an initial round with groups of four teams was followed by direct elimination rounds. In UCL tournaments, two teams generally play each other twice, with the exception of the final round that consists of a single game. Also here, the outcome of each round is represented as an ordered hyperedge in the hypergraph-based PL model.

D. Sushi preferences

We consider the results of a series of surveys conducted by T. Kamishima asking $M = 5000$ individuals for their preferences about various kinds of sushi [31]. We consider two distinct datasets: one containing $K_{\min} = K_{\max} = 10$ complete strict rank orders of $N = 10$ different kinds of sushi; the other concerning $K_{\min} = K_{\max} = 10$ strict rank orderings for $N = 100$ different kinds of sushi.

E. AGH course selections

We consider a dataset containing the results of a survey for $M = 153$ students at the AGH University of Science and Technology about their course preferences for the year 2004 [32]. Each student provided a rank ordering over all $N = K_{\min} = K_{\max} = 7$ courses.

F. APA elections

We consider a dataset containing the results of the elections of the American Psychological Association (APA) in year 2009 [33]. The voters were allowed to rank any number up to $K_{\max} = 5$ candidates without ties. We exclude data where only one candidate is listed, thus imposing $K_{\min} = 2$. The total number of voters in the dataset is $M = 12078$.

G. Network Science collaborations

Data are from the November 2024 version of OpenAlex. We consider all papers published between 1999 and 2023 in the following journals: *Nature, Nature Communications, Nature Physics, Proceedings of the National Academy of Sciences USA, Physical Review E, Physical Review Letters, Physical Review X, Science and Science Advances*. Only papers with topic "Statistical Mechanics of Complex Networks" (topic id T10064 in the OpenAlex database) are selected. For each paper, we retrieve the ordered list of authors. We rely on the authors' OpenAlex identifiers for disambiguation. We exclude single-author papers as they do generate any hyperedge in our hypergraph. The resulting dataset consists of $M = 2461$ papers authored by $N = 3517$ distinct scholars; the size of the teams involved in the writing of the papers range between $K_{\min} = 2$ and $K_{\max} = 14$. Our main goal in considering this dataset is to quantify the leadership of individual scientists, typically reflected in the order of authorship in this specific scientific discipline (i.e., the senior author appears in the last position in the list of authors). To make the dataset consistent with the others considered in this paper, we actually invert the order of the authors in each paper, so that the first authors appears in the last position in the corresponding hyperedge, and the last author appears as the first element in the corresponding hyperedge.

IV. RESULTS

A. Convergence of the algorithms

First, we report on some tests aimed at demonstrating the improved efficiency of the iteration algorithm *á la* Newman of Eq. (13) compared to the standard Zermelo's schema of Eq. (12). In all our tests, we ensure that the initial guess of the vector $\vec{\pi}$ is the same for both the iterative schemes; also, we use the same exact criterion of Eq. (8) to estimate the convergence of each algorithm. We then compare the total number of iterations required by either one method or the other to converge. Results of this analysis for a selection of two real datasets are displayed in Figure 1 and summarized in Table I for all sets of real and synthetic data. Speed-ups in convergence are apparent: depending on the dataset considered, we roughly observe a 3 to 70 speed-up factor.

B. Prediction of unobserved events

Next, we compare the PL model of Eq. (1) against its graph counterpart pPL of Eq. (16). We do not show any results here, but we systematically verify that the two approaches infer vectors of scores $\vec{\pi}$ whose components are highly correlated. In Figure 2(a), we show results concerning the predictive power of the two models when applied to synthetic hypergraphs

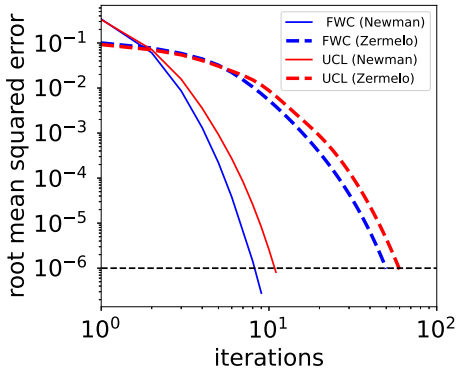

FIG. 1. Convergence of the algorithms. We plot the root mean squared error of Eq. (8) as a function of the number of iterations for the schemas of Eqs. (12) [dashed curves] and (13) [full curves]. We display results for the FIFA World Cup (FWC) [blue] and the UEFA Champions League (UCL) [red] datasets. The horizontal black dashed line denotes $\epsilon = 10^{-6}$, i.e., the accuracy parameter of Eq. (8) used to determine the convergence of the iterative algorithms.

generated according to the PL model of Eq. (1). We specifically focus on cross-validation tests where 80% of the interactions are used to train the models, and the unobserved 20% are used to quantify their predictive powers. For both models, predictive performance is quantified using the log-likelihood of Eq. (4) estimated for the unobserved 20% data; we denote this quantity as $\mathcal{L}_{\rm PL}$ for the PL model and as $\mathcal{L}_{\rm pPL}$ for the pPL model; we also estimate the log-likelihood $\mathcal{L}$ using the ground-truth scores $\vec{\pi}$ used to generate the synthetic hypergraph. The results of Figure 2(a) clearly show that the true model consistently outperforms its approximated version. The same conclusions are also valid for the position 1-breaking versions of PL and pPL, respectively obtained using Eqs. (14) and (17), see Fig. 2(b).

We repeat the same cross-validation tests also for the real-world datasets. Results are summarized in Fig. 3(a) for the standard PL model, and in Fig. 3(b) for its position 1-breaking variant. For real-world datasets, two fundamental observations are in order. First, the actual score of the individual nodes is not known. This fact implies that predictions cannot be judged in an absolute fashion, rather only relatively to other predictive models. Although all real systems are represented as ordered and weighted hypergraphs, there is no guarantee that this is their *bona fide* representation. The very comparison of the predictive power of the multibody PL against the pairwise pPL actually offers the possibility of stating whether interactions are better represented by multibody comparisons rather than the simpler aggregation of pairwise comparisons. Based on the results of Fig. 3(a), we conclude that the multibody-comparison hypothesis is meaningful for all real datasets considered in this paper; the only exception is represented by the APA 2009 dataset, where most of the interactions can be equally well explained as being the aggregation of multiple pairwise comparisons. The full version of the position 1-breaking PL model appears at least as predictive as its projection-based variant for all datasets considered, see Fig. 3(b). For WFC and UCL, we still observe a better performance of the hypergraph-based approach: this is probably due to the fact that the vast majority of the comparisons is between pairs of elements, thus no great difference between the PL and the position 1-breaking PL models is expected. For the NS data, we observe comparable performance, denoting that the full order of the authors is reflecting well their contribution in the collaboration. For survey or election data, the results indicate that respondents/voters identify the first element by comparing it against each of the other alternatives in pairwise comparisons.

C. Ranking of football teams and network scientists

We conclude our study by providing ranked lists of football teams and network scientists using the data at our disposal. In all cases, we generate those lists using the entire datasets at our disposal to infer the nodes' scores $\vec{\pi}$. For FWC and UCL, we use the PL model of Eq. (1). For NS, we rely on the position 1-breaking PL model of Eq. (14). For brevity, we list in Table II only the top-10 elements in the rankings; complete rankings are available on the github repository.

In the WFC dataset, we see that the first four positions are occupied by Brazil, Germany, Italy, and Argentina, i.e., the most successful national teams in the history of the World Cup. All others are also good examples of teams characterized by historically good performance in the competition. The fact

TABLE I. Summary table. Each row of the table corresponds to a specific dataset analyzed. From left to right, we report the name of the dataset, the number of nodes N, the number of interactions M, the minimum $K_{\rm min}$ and the maximum size $K_{\rm max}$ of the interactions, the number of iterations I required for convergence for the Zermelo algorithm and for the Newman algorithm, and the average speed-up factor of using the Newman's version of the algorithm instead of the Zermelo's one. The number of iterations has been estimated over at least 10 runs of the algorithms, each started from a different initial condition. We report the average value and standard deviation over those runs.

Full name	N	M	$K_{\rm min}$	$K_{\rm max}$	I (Zermelo)	I (Newman)	Speed-up
Synthetic 1	1000	10000	2	10	103 ± 1	11.0 ± 0.1	9
Synthetic 2	1000	100000	2	10	169 ± 1	11.0 ± 0.1	15
FIFA World Cup (FWC)	83	364	2	4	50 ± 2	9.0 ± 0.3	6
UEFA Champions League (UCL)	174	674	2	4	60 ± 2	11.1 ± 0.5	5
Sushi 10	10	5000	10	10	14 ± 1	7.6 ± 0.5	1.8
Sushi 100	100	5000	10	10	21 ± 2	6.9 ± 0.3	3
AGH course selection 2004	7	153	7	7	534 ± 4	7.6 ± 0.4	70
APA election 2009	10	12078	2	5	16 ± 1	7.3 ± 0.5	2.2
Network Science (NS)	3517	2461	2	14	144 ± 7	24 ± 1	6

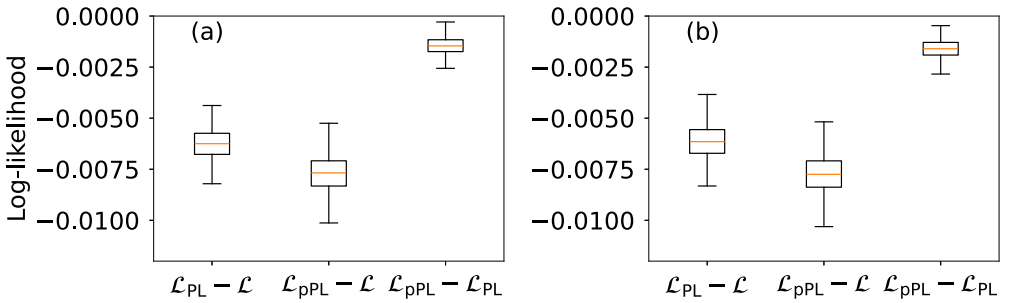

FIG. 2. Model selection for synthetic hypergraphs. (a) We consider synthetic hypergraphs with $N = 1000$ nodes and $M = 100000$ interactions. The size of those interactions ranges between $K_{min} = 2$ and $K_{max} = 10$ nodes. Nodes' scores are preassigned from the model of Eq. (11), and the order of the nodes in each interaction is determined on the basis of Eq. (1). We train the Plackett-Luce (PL) and its projected version (pPL) using 80% of the interactions, and make predictions on the remaining 20%. We estimate the logarithm of the likelihoods of the two models, respectively $\mathcal{L}_{PL}$ and $\mathcal{L}_{pPL}$, using Eq. (4); at the same time, we also estimate the logarithm of the likelihood $\mathcal{L}$ of the unobserved interactions using the ground-truth value of the nodes' scores. We display results over 1000 independent realizations of the prediction task; the orange line identifies the median, the boxes delimit the first and third quartiles, and the error bars denote the 5% and 95% confidence intervals. (b) Same as in panel (a), but for the position 1-breaking PL model as defined in Eq. (14).

that Netherlands, Croatia, Sweden, and Czechoslovakia are in the top 10 even if they did not win any World Cup so far is somehow surprising.

Similar conclusions are valid also for the UCL dataset, where Real Madrid dominates the ranking followed by Bayern Munich, Barcelona, and Liverpool. It is a bit unexpected that Milan appears only at the 8th position in the table in spite of being the second most successful team in the competition. We stress, however, that our dataset covers only the period 1993–2024, when Milan won 3 UCL titles.

For the NS dataset, we first rank scholars using the model of Eq. (14); we then consider only scholars with more than 10 papers in our set of data. A. Vespignani tops the ranking being the senior author in all the 28 papers he co-authored according to our dataset. All other scholars in the top-10 table have very high percentages of papers where they appear as senior authors combined with an overall large number of publications. Specifically, we have that: H.A. Makse has 13 papers with 92% of senior-author positions, L.A.N. Amaral 17 and 88%, R. Pastor-Satorras 49 and 67%, K. Sneppen 11 and 91%, A.L. Barabási 34 and 76%, L. Zdeborová 11 and 91%, S. Havlin 95 and 56%, J.F.F. Mendes 53 and 77%, and H.E. Stanley 70 and 50%. We stress that being the senior author in many papers is not the only important aspect to reach a top rank position; a very important factor is also to be the senior author in papers with other authors with large scores. At the same time, we remark that the contribution of being the senior author of a paper to the score of a scholar is proportional to the number of authors in the paper. Also, scholars with many single-author publications are penalized as these publications are not accounted for in this specific ranking.

V. CONCLUSIONS

We successfully generalized a recent iteration schema for ranking entities from pairwise comparisons to the multibody-comparison setting. Alongside a formal derivation of the iterative algorithm, we provided numerical evidence that the proposed method converges faster than the state-of-the-art iterative procedure in a variety of synthetic and real-world datasets.

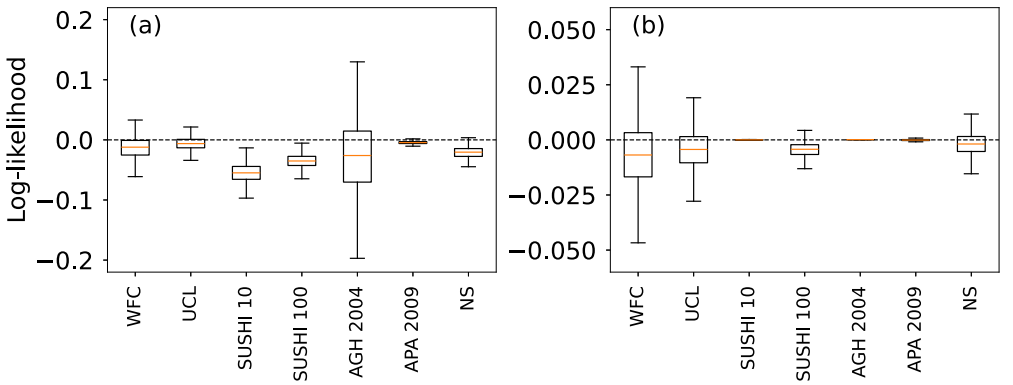

FIG. 3. Model selection for real hypergraphs. (a) Same as in Fig. 2(a), but for the real datasets considered in the paper. Each plot summarizes the statistics of the log-likelihood difference $\mathcal{L}_{pPL} - \mathcal{L}_{PL}$ over 1000 realizations of the prediction task where 80% of the interactions are used to infer the nodes' scores then used to predict the remaining 20% of unobserved interactions. (b) Same as in (a), but for the position 1-breaking PL model.

TABLE II. Top-10 ranking in football and network science. We display the items ranking in top-10 positions in the datasets FIFA World Cup (WFC), UEFA Champions League (UCL), and the Network Science (NS).

Rank	FWC	UCL	NS
1	Brazil	Real Madrid	A. Vespignani
2	Germany	Bayern Munich	H.A. Makse
3	Italy	Barcelona	L.A.N. Amaral
4	Argentina	Liverpool	R. Pastor-Satorras
5	Netherlands	Chelsea	K. Sneppen
6	France	Manchester City	A.-L. Barabási
7	Croatia	Juventus	L. Zdeborová
8	England	Milan	S. Havlin
9	Sweden	Paris Saint-Germain	J.F.F. Mendes
10	Czechoslovakia	Atlético Madrid	H.E. Stanley

We then performed a systematic analysis for predicting the outcome of unobserved comparisons in various real-world systems. For all datasets considered, models that take advantage of the multibody nature of the comparisons display larger predictive power than that of models based on pairwise projections of such multibody comparisons. In election data, both approaches have comparable predictive power, questioning the *bona fide* multibody nature of this dataset.

ACKNOWLEDGMENTS

The authors thank G. Bianconi for comments and suggestions on the paper. This work received partial support from the Air Force Office of Scientific Research (Grants No. FA9550-21-1-0446 and No. FA9550-24-1-0039). The funders had no role in study design, data collection and analysis, the decision to publish, or any opinions, findings, conclusions, or recommendations expressed in the manuscript.

TABLE III. Numerical estimation of the Plackett-Luce scores. Each row of the table corresponds to a specific dataset analyzed. From left to right, we report the name of the dataset, the number of iterations I and $\tilde{I}$ respectively required for convergence for the Newman algorithm with and without normalization after each iteration, the speed-up factor of using the Newman's version of the algorithm relying on normalization of the scores at each iteration compared to the one that does not normalize the scores, the Pearson's, Kendall's and Spearman's correlation coefficients of the scores obtained with and without normalization of the scores after each iteration. The values of the correlation coefficients are not exactly 1, but they have been rounded to the second decimal digit. We compute the scores only once for each dataset starting from random initial conditions.

Full name	I	$\tilde{I}$	Speed-up	Pearson	Kendall	Spearman
Synthetic 1	15	487	32.5	1.00	1.00	1.00
Synthetic 2	15	3921	261.4	1.00	1.00	1.00
FWC	12	77	6.4	1.00	1.00	1.00
UCL	15	83	5.5	1.00	1.00	1.00
Sushi 10	10	26162	2616.2	1.00	1.00	1.00
Sushi 100	9	3738	415.3	1.00	1.00	1.00
AGH 2004	9	1163	129.2	1.00	1.00	1.00
APA 2009	10	43406	4340.6	1.00	1.00	1.00
NS	35	38	1.1	1.00	1.00	1.00

DATA AVAILABILITY

The data that support the findings of this article are openly available [34].

APPENDIX: COMPUTING THE PLACKETT-LUCE SCORES

In Fig. 4, we show the convergence of the algorithm of Eq. (13) in case the normalization condition of Eq. (8) is imposed or not after each iteration. We consider the same networks as in Fig. 1. As the results of panel (a) show,

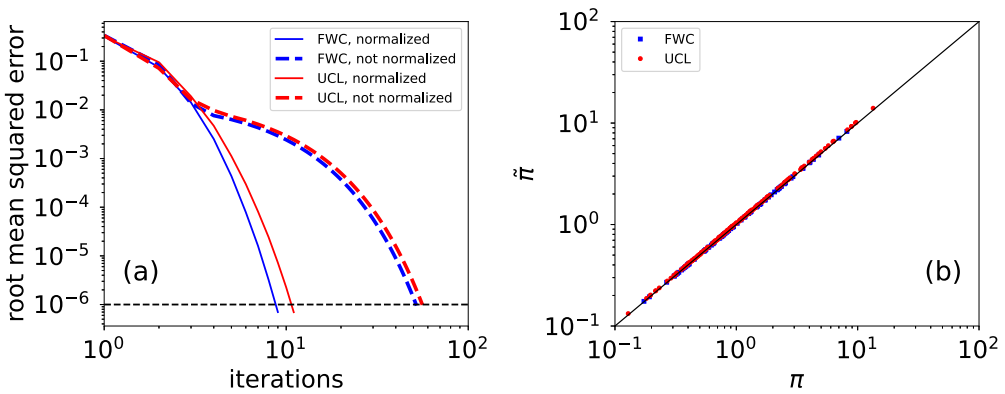

FIG. 4. Numerical estimation of the Plackett-Luce scores. (a) We plot the root mean squared error of Eq. (8) as a function of the number of iterations for the schema of Eq. (13). Full curves are obtained by normalizing the scores at each iteration of the algorithm [i.e., Eq. (2)], whereas dashed curves are obtained without normalization of the scores. We display results for the FIFA World Cup (FWC) [blue] and the UEFA Champions League (UCL) [red] datasets. The horizontal black dashed line denotes $\epsilon = 10^{-6}$, i.e., the accuracy parameter of Eq. (8) used to determine the convergence of the iterative algorithms. (b) Normalized $\tilde{\pi}$ *vs.* non-normalized π scores for the same networks as in panel (a). Each point represents a node in the network. Blue squares are used for FWC and red circles for UCL. The black line denotes equality between the scores.

applying the normalization condition has the beneficial effect of dramatically reducing the number of iterations required to reach convergence; such a gain in convergence time does not significantly affect the accuracy of the numerical estimates, as the results of panel (b) demonstrate.

We systematically compared the two implementations of the algorithm with and without normalization on all datasets considered in this paper. Results are summarized in Table III. Applying the normalization condition allows always to reduce the time to convergence; the speedup factor is in some cases close to 10^0, in some other cases even larger than 10^3. At the same time, the scores estimated with normalization are almost perfectly correlated with those computed without normalization.

[1] M. Newman, *Networks* (Oxford University Press, Oxford, 2010).
[2] F. Radicchi, PloS One **6**, e17249 (2011).
[3] S. Motegi and N. Masuda, Sci. Rep. **2**, 904 (2012).
[4] Ş. Erkol and F. Radicchi, J. Complex Networks **9**, cnab012 (2021).
[5] C. De Bacco, D. B. Larremore, and C. Moore, Sci. Adv. **4**, eaar8260 (2018).
[6] R. A. Bradley and M. E. Terry, Biometrika **39**, 324 (1952).
[7] E. Zermelo, Math. Z. **29**, 436 (1929).
[8] M. Newman, J. Mach. Learn. Res. **24**, 238 (2023).
[9] M. Jerdee and M. Newman, Sci. Adv. **10**, eadn2654 (2024).
[10] FIFA World Cup, https://www.fifa.com/en/tournaments/mens/worldcup.
[11] UEFA Champions League, https://www.uefa.com/uefachampionsleague.
[12] Wikipedia page on Texas hold em poker, https://en.wikipedia.org/wiki/Texas_hold_em.
[13] Wikipedia page on Multiplayer video games, https://en.wikipedia.org/wiki/Multiplayer_video_game.
[14] Wikipedia page on the Netflix Prize, https://en.wikipedia.org/wiki/Netflix_Prize.
[15] S. Bhattacharya *et al.*, Ind. J. Plast. Surg. **43**, 233 (2010).
[16] R. L. Plackett, J. R. Stat. Soc. Ser. C **24**, 193 (1975).
[17] R. D. Luce, Econometrica **26**, 193 (1958).
[18] L. Maystre and M. Grossglauser, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2015), Vol. 28.
[19] H. A. Soufiani, W. Z. Chen, D. C. Parkes, and L. Xia, *Advances in Neural Information Processing Systems* (NIPS, 2013).
[20] S. Negahban, S. Oh, and D. Shah, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2012), Vol. 25.
[21] A. Seshadri, S. Ragain, and J. Ugander, arXiv:2312.15081.
[22] *Choice, decision, and measurement: Essays in honor of R. Duncan Luce*, edited by A. A. J. Marley (1997).
[23] T.-Y. Liu *et al.*, Found. Trends. Inf. Retrieval **3**, 225 (2009).
[24] H. Oosterhuis, in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2022), pp. 2266–2271.
[25] W. Cheng, E. Hüllermeier, and K. J. Dembczynski, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (2010), pp. 215–222.
[26] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno *et al.*, Nat. Phys. **17**, 1093 (2021).
[27] G. Ghoshal, V. Zlatić, G. Caldarelli, and M. E. J. Newman, Phys. Rev. E **79**, 066118 (2009).
[28] V. Zlatić, G. Ghoshal, and G. Caldarelli, Phys. Rev. E **80**, 036118 (2009).
[29] https://en.wikipedia.org/wiki/1982_FIFA_World_Cup.
[30] https://en.wikipedia.org/wiki/2022–23_UEFA_Champions_League.
[31] T. Kamishima, in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (2003), pp. 583–588.
[32] P. Skowron, P. Faliszewski, and A. Slinko, arXiv:1301.6400.
[33] M. Regenwetter, A. Kim, A. Kantor, and M.-H. R. Ho, Psychol. Sci. **18**, 629 (2007).
[34] https://github.com/jackyeung99/higher_order_ranking.