



Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

The incomplete analytic hierarchy process and Bradley–Terry model: (In)consistency and information retrieval

László Gyarmati^{a,*}, Éva Orbán-Mihálykó^a, Csaba Mihálykó^a, Zsombor Szádóczki^{b,c},
Sándor Bozóki^{b,c}

^a Department of Mathematics, University of Pannonia, 8200 Egyetem u. 10., Veszprém, Hungary

^b Research Laboratory on Engineering & Management Intelligence, Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network (ELKH), 1111 Kende u. 13-17., Budapest, Hungary

^c Department of Operations Research and Actuarial Sciences, Institute of Operations and Decision Sciences, Corvinus University of Budapest, 1093 Fővám tér 8., Budapest, Hungary

ARTICLE INFO

Keywords:

Paired comparison
Pairwise comparison
Consistency
Bradley–Terry model
Information retrieval
Graph of graphs

ABSTRACT

Several methods of preference modeling, ranking, voting and multi-criteria decision-making include pairwise comparisons. It is usually simpler to compare two objects at a time, furthermore, some relations (e.g., the outcome of sports matches) are naturally known for pairs. This paper investigates and compares pairwise comparison models and the stochastic Bradley–Terry model. It is proved that they provide the same priority vectors for consistent (complete or incomplete) comparisons. For incomplete comparisons, all filling in levels are considered. Recent results identified the optimal subsets and sequences of multiplicative/additive/reciprocal pairwise comparisons for small sizes of items (up to $n = 6$). Simulations of this paper show that the same subsets and sequences are optimal in the case of the Bradley–Terry and the Thurstone models as well. This somehow surprising coincidence suggests the existence of a more general result. Further models of information and preference theory are subject to future investigation to identify optimal subsets of input data.

1. Introduction

Comparison in pairs is a frequently used method in ranking and rating objects when scaling is difficult due to its subjective nature. From a methodological point of view, two main types of models can be distinguished: the ones based on pairwise comparison matrices (PCMs) and the stochastic models motivated by Thurstone. The aim of this paper is to present some linkages between these approaches. We establish a direct relation between them via the concept of consistency. Evaluating consistent data, we mainly focus on the similarities of the results in the case of incomplete comparisons. In [Bozóki and Szádóczki \(2022\)](#) and [Szádóczki et al. \(2022b\)](#), the authors investigate incomplete pairwise comparisons evaluated via two different weight calculation techniques from an information retrieval point of view. Incompleteness means that the comparisons between some pairs of objects are missing. In the case of consistent data, this absent information is preserved in the results of the compared pairs. However, in real problems, the results of comparisons are usually not consistent, therefore this missing information can cause significant modifications in the evaluations. The main question of the articles ([Bozóki & Szádóczki, 2022](#)) and [Szádóczki](#)

[et al. \(2022b\)](#) is which structure of the comparisons is optimal for fixed numbers of comparisons investigating pairwise comparison matrices via the logarithmic least squares and the eigenvector methods (LLSM and EM). We ask the same question in the case of the Bradley–Terry (BT) model, when the evaluation is performed via maximum likelihood estimation (MLE). What are the similarities and the differences between the optimal comparison arrangements in the case of LLSM and BT? Do the findings of the papers remain valid in the case of a substantially different paired comparison method? Are the conclusions method-specific, i.e., do the results of the paper ([Bozóki & Szádóczki, 2022](#)) apply only for the method LLSM and EM or in general as well, for other models based on paired comparisons?

The rest of the paper is organized as follows. Section 2 presents the closely related literature and the research gap that we would like to consider in the current study. Section 3 describes the preliminary methods, namely the AHP with LLSM, EM and the Bradley–Terry model with MLE. In Section 4 we describe the connection between these models. We pay special attention to consistency, which is deeply investigated in PCM-based methods, but not in stochastic models. Sub

* Corresponding author.

E-mail addresses: gyarmati.laszlo@phd.mik.uni-pannon.hu (L. Gyarmati), orban.eva@mik.uni-pannon.hu (É. Orbán-Mihálykó), mihalyko.csaba@mik.uni-pannon.hu (C. Mihálykó), szadoczki.zsombor@sztaki.hu (Z. Szádóczki), bozoki.sandor@sztaki.hu (S. Bozóki).

<https://doi.org/10.1016/j.eswa.2023.120522>

Received 12 October 2022; Received in revised form 10 May 2023; Accepted 18 May 2023

Available online 24 May 2023

0957-4174/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Section 4.1 contains theoretical results about the connection of the models in consistent cases, while Sub Section 4.2 presents examples demonstrating the differences when the data are inconsistent. Section 5 details the simulation methodology that is used to find optimal solutions concerning information retrieval, while Section 6 contains the main results of the numerical experiments. Finally, Section 7 concludes and discusses further research questions.

2. Literature review

The set of information on the preferences of users or decision-makers serves as input data of information retrieval and decision support systems, including ranking, recommendation, relevance filtering, relation identification etc. (Zhou & Yao, 2010). Many information retrieval tasks are themselves multi-criteria decision problems, where the steps are highly ‘automated’. Explicit preferences are collected via questionnaires, decision support softwares (Ishizaka & Nemery, 2013; Tzeng & Huang, 2011). Implicit preferences can originate from e.g., observed clicks as feedback from user interaction (Hofmann et al., 2013; Jeon & Kim, 2013). Several data such as frequencies (Wu et al., 2023), results of sport matches or competitions etc. can also be transformed to implicit preferences.

Applying the method of paired comparisons is essential in psychology (Thurstone, 1927), sports (Csató, 2021; Orbán-Mihálykó et al., 2022), preference modeling (Choo & Wedley, 2004; Mantik et al., 2022), ranking (Fürnkranz & Hüllermeier, 2011; Shah & Wainwright, 2018), and decision-making methods (Abel et al., 2022; Stewart, 1992).

Nowadays, more and more attention is paid to incomplete comparisons, see for instance (Faramondi et al., 2020; Mondal et al., 2023) and for a detailed literature review (Ureña et al., 2015).

Through the concept of consistency/inconsistency, we compare some recent results (Bozóki & Szádóczki, 2022) gained on the domain of PCMs used by the popular multi-attribute decision-making method Analytic Hierarchy Process (AHP) (Saaty, 1977) to the outcomes provided by the also widely used Bradley–Terry model (Bradley & Terry, 1952).

The latter one is a special case of the more general Thurstone-motivated stochastic models. Both these stochastic methods and the generalization of the AHP can be used on incomplete data, when some of the paired comparisons are missing (Harker, 1987; Ishizaka & Labib, 2011), which is often demonstrated on sports examples (Bozóki et al., 2016; Orbán-Mihálykó et al., 2019a). However, in this regard, several theoretical questions have been investigated in the most recent literature (see for instance Chen et al., 2022; Mazurek & Kułakowski, 2022; Tekile et al., 2023).

AHP-based and stochastic Thurstone-motivated models are significantly different in their fundamental concepts. In the case of two different principles of a problem’s solution, the linkage between them is always motivational: what are the common features and the differences between the methods. As far as the authors know, only few publications are devoted to this question. Researchers usually deal with one of the methods. However, in MacKay et al. (1996), the authors recognize the following: ‘the two branches resemble each other in that both may be used to estimate unidimensional scale values for decision alternatives or stimuli from pairwise preference judgments about pairs of stimuli. The models differ in other respects.’ Nevertheless, in Genest and M’lan (1999) the authors compare the AHP-based methods and the Bradley–Terry model in the case of complete comparisons and they prove that in special cases some different types of techniques provide equal solutions. In Orbán-Mihálykó et al. (2015), the authors compared numerically the AHP and Thurstone methods: evaluating a real data set on a 5-value scale, the numerical results provided by the different methods were very close to each other. Temesi et al. (2023) compare the rankings provided by the AHP and the Bradley–Terry model on a real incomplete data set. Further numerical comparisons for incomplete data can be found in Orbán-Mihálykó et al. (2019a).

Consistency/inconsistency of PCMs is a focal issue in the case of pairwise comparisons (Brunelli, 2018; Duleba & Moslem, 2019) as well as for fuzzy decisions, such as various incomplete linguistic preference relations (Liu et al., 2022, 2021; Wang et al., 2023, 2021), but it is not investigated in stochastic models. The results’ compatibility with real experiences is related to the inconsistency of the PCM: discrepancy may appear even in the case of complete comparisons. The question necessarily raises: what does the consistency mean in the BT model?

In the case of missing comparisons, two further aspects have a crucial effect on the results in every model, namely, the number of known comparisons, and their arrangement. Our approach is strongly relying on the graph representation of incomplete paired comparisons (Gass, 1998). We are determining the best representing graphs (the best pattern of known comparisons) in the Bradley–Terry model for all possible numbers of comparisons (edges) for a given number of alternatives (vertices). The importance of the pattern of known comparisons in PCMs has been investigated for some special cases by Szádóczki et al. (2022b), who emphasized the effect of (quasi-)regularity and the minimal diameter (longest shortest path) property of the representing graphs. Szádóczki et al. (2022a) also examined some additional ordinal information in the examples studied by them. Finally, Bozóki and Szádóczki (2022) have investigated all the possible filling in patterns of incomplete PCMs, and determined the best ones for all possible (n, e) pairs up until 6 alternatives with the help of simulations, where n is the number of items to be compared, while e is the number of compared pairs. Their main findings (besides the concrete graphs) are (i) the star graph is always optimal among spanning trees; (ii) regularity and bipartiteness are important properties of optimal filling patterns.

To the best of the authors’ knowledge, there has been no similar study on the family of stochastic models, thus in this paper we would also like to fill in this research gap.

3. Preliminaries of the applied methods

3.1. Analytic hierarchy process

The AHP methodology is based on PCMs, which can be used to evaluate alternatives according to a criterion or to compare the importance of the different criteria.

Definition 1 (Pairwise Comparison Matrix (PCM)). Let us denote the number of items to be compared (usually criteria or alternatives) in a decision problem by n . The $n \times n$ matrix $A = [a_{ij}]$ is called a pairwise comparison matrix, if it is positive ($a_{ij} > 0$ for $\forall i$ and j), and reciprocal ($1/a_{ij} = a_{ji}$ for $\forall i$ and j).

a_{ij} , the general element of a PCM, shows how many times item i is better/more important than item j . In an ideal case, these elements are not contradicting to each other, thus we are dealing with a consistent PCM.

Definition 2 (Consistent PCM). A PCM is called consistent if $a_{ik} = a_{ij}a_{jk}$ for $\forall i, j, k$. If a PCM is not consistent, then it is said to be inconsistent.

Remark 1. A PCM is consistent if and only if its elements can be written in the form of $a_{ij} = w_i/w_j$ for $\forall i, j$.

In practical problems, the PCMs filled in by decision-makers are usually not consistent, and because of that, there is a large literature on how to measure the inconsistency of these matrices (Brunelli, 2018). Recently even a general framework has been proposed for defining inconsistency indices of reciprocal pairwise comparisons (Bortot et al., 2022).

In the case of consistent PCMs all the different weight calculation techniques result in the same weight (prioritization/preference) vector that determines the ranking of the compared items. However, for inconsistent data, the results of different weight calculation methods can

vary. Two of the most commonly used techniques are the logarithmic least squares method (Crawford & Williams, 1985) and the eigenvector method (Saaty, 1977).

Definition 3 (Logarithmic Least Squares Method (LLSM)). Let A be an $n \times n$ PCM. The weight vector \underline{w} of A determined by the LLSM is given as follows:

$$\min_{\underline{w}} \sum_{i=1}^n \sum_{j=1}^n \left(\ln(a_{ij}) - \ln\left(\frac{w_i}{w_j}\right) \right)^2, \quad (1)$$

where w_i is the i th element of \underline{w} , $0 < w_i$ and $\sum_{i=1}^n w_i = 1$.

Definition 4 (Eigenvector Method (EM)). Let A be an $n \times n$ PCM. The weight vector \underline{w} of A determined by the EM is defined as follows:

$$A \cdot \underline{w} = \lambda_{\max} \cdot \underline{w} \quad (2)$$

where λ_{\max} is the principal eigenvalue of A , and the componentwise positive eigenvector \underline{w} is unique up to a scalar multiplication.

These techniques can be generalized to the case when some elements are absent from the PCM, when we have to deal with an incomplete pairwise comparison matrix (IPCM). For IPCMs the LLSM's optimization problem only includes the known elements of the matrix, while the EM is based on the Consistency Ratio-minimal completion of the IPCM and its principal right eigenvector (Bozóki et al., 2010; Shiraishi & Obata, 2002; Shiraishi et al., 1998).

It is also worth mentioning that the inconsistency indices and their respective thresholds can also be generalized for the IPCM case (Ágoston & Csató, 2022; Kułakowski & Talaga, 2020). A suitable tool to handle IPCMs is to use their graph representation (Gass, 1998).

Definition 5 (Representing Graph of An IPCM). An IPCM A is represented by the undirected graph $G_E = (V, E)$, where the vertex set V of G_E corresponds to the items to be compared (alternatives/criteria) of A , and there is an edge in the edge set E of G_E if and only if the appropriate element of A is known.

The generalized definition of a consistent PCM in the incomplete case is as follows (Bozóki & Tsyganok, 2019):

Definition 6 (Consistent IPCM). An incomplete pairwise comparison matrix (IPCM) A represented by the graph $G_E = (V, E)$ is called consistent, if

$$a_{i_1 i_2} \cdot a_{i_2 i_3} \cdot \dots \cdot a_{i_k i_1} = 1 \quad (3)$$

for any cycle in G_E that is $(i_1, i_2, \dots, i_k, i_1)$, for which $(i_l, i_{l+1}) \in E$ and $(i_k, i_1) \in E$.

All the definitions introduced in this section can be modified applying additive or reciprocal (fuzzy) PCMs instead of the multiplicative ones (Brunelli, 2014).

3.2. Stochastic approach: The Bradley–Terry model

From the stochastic branch, this paper investigates the Bradley–Terry model (BT) for two options, worse and better. It is worth mentioning that the Bradley–Terry model is generalized for ties in Rao and Kupper (1967), but we deal with the model that is allowing only two options. It is a special case of the more general Thurstone-motivated model, in which the performances of the objects are considered to be random variables ξ_i with expected values m_i , $i = 1, 2, \dots, n$. The result of a comparison is connected to the realized value of the difference of the random variables $\xi_i - \xi_j = m_i - m_j + \eta_{i,j}$. If the result of the comparison is that ‘ i is worse than j ’, then the difference is negative. Vice versa, the comparison result that ‘ i is better than j ’ means that the difference of the random variables belonging to the objects i and j is non-negative.

Of course, ‘ i is worse than j ’ means that ‘ j is better than i ’. $\eta_{i,j}$ ($i < j$) are supposed to be independent continuous random variables with a common cumulative distribution function (c.d.f.) F . The probability density function f is supposed to be symmetric, positive on \mathbb{R} , and its logarithm is strictly concave.

The probabilities of the above-mentioned events (worse and better), due to the symmetry, are as follows.

$$p_{i,j,1} = P(\xi_i - \xi_j < 0) = F(0 - (m_i - m_j)) = F(m_j - m_i) = p_{j,i,2} \quad (4)$$

$$p_{i,j,2} = P(0 \leq \xi_i - \xi_j) = 1 - p_{i,j,1} = F(m_i - m_j) = p_{j,i,1} \quad (5)$$

If F is the standard normal cumulative distribution function, then we are dealing with the Thurstone model. In the case, when

$$F(x) = \frac{1}{1 + \exp(-x)}, \quad (6)$$

the c.d.f. of the logistic distribution, then we are applying the BT model. In the case of the BT model, one can check that

$$p_{i,j,1} = \frac{\exp(m_j)}{\exp(m_i) + \exp(m_j)}, \quad (7)$$

and

$$p_{i,j,2} = \frac{\exp(m_i)}{\exp(m_i) + \exp(m_j)}. \quad (8)$$

In the rest of the paper, we focus on the case of the logistic distribution. Let $D_{i,j,1}$ be the number of comparisons when ‘ i is worse than j ’ and let $D_{i,j,2}$ be the number of comparisons when ‘ i is better than j ’. For the sake of simplicity, $D_{i,i,1} = D_{i,i,2} = 0$. Data matrix D contains the elements $D_{i,j,k}$, and its dimension is $n \times n \times 2$. Obviously $D_{i,j,1} = D_{j,i,2}$ and $D_{i,j,2} = D_{j,i,1}$, and the data matrix can also be written in a two-dimensional matrix form as follows: the row number is the number of pairs for which $i < j$, $n(n-1)/2$, while the number of columns is the number of options (better and worse), 2. We use the same notation D for both data forms. If the decisions are independent, the probability of sample D as a function of the parameters is as follows,

$$L(D|\underline{m}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^n p_{i,j,1}^{D_{i,j,1}} \cdot p_{i,j,2}^{D_{i,j,2}}. \quad (9)$$

The maximum likelihood estimation (MLE) of the parameter vector \underline{m} is the n -dimensional argument $\hat{\underline{m}}$ for which function (9) reaches its maximal value. As (9) depends only on the differences of the expected values, one parameter can be fixed (for example $m_1 = 0$). Supposing this, in Ford Jr (1957) the author proves a necessary and sufficient condition that guarantees the existence and uniqueness of the MLE in the BT model: the directed graph of the comparisons has to be strongly connected. In this case, the vertices are the items to compare, as previously, and there is a directed edge from item i to item j if $0 < D_{i,j,2}$ and from item j to item i if $0 < D_{i,j,1}$ ($i < j$).

Instead of (9), frequently the log-likelihood,

$$\ln L(D|\underline{m}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n (D_{i,j,1} \cdot \ln(p_{i,j,1}) + D_{i,j,2} \cdot \ln(p_{i,j,2})) \quad (10)$$

is maximized. Its maximum value is at the same argument, as the maximum of (9). The condition for the existence and uniqueness of the maximal point is the same, as mentioned before. Note, that the matrix D can be multiplied by an arbitrary positive number, and the argument of the maximal value of (9) and (10) does not change. This can be easily seen in (10). This means that the values of $D_{i,j,k}$, $k = 1, 2$ are not necessarily integers or rational numbers. We will use this property during the investigations in the next section.

4. Connection between the BT and PCM models

4.1. Consistent data matrix evaluation

The vector of the estimated expected values \hat{m} can be converted into a weight vector by a strictly monotone transformation as follows:

$$\hat{w} = \left(\frac{\exp(\hat{m}_1)}{\sum_{i=1}^n \exp(\hat{m}_i)}, \dots, \frac{\exp(\hat{m}_n)}{\sum_{i=1}^n \exp(\hat{m}_i)} \right). \tag{11}$$

Similarly, if the exact expected values m were known, a weight (priority) vector could be constructed as follows:

$$w = \left(\frac{\exp(m_1)}{\sum_{i=1}^n \exp(m_i)}, \dots, \frac{\exp(m_n)}{\sum_{i=1}^n \exp(m_i)} \right). \tag{12}$$

This transformation eliminates the effect of the fixed parameter. The weights are positive and their sum equals to 1, as in the case of AHP. Now starting from a normalized vector with positive components

$$\underline{w}^{(PCM)} = \left(w_1^{(PCM)}, \dots, w_n^{(PCM)} \right), \tag{13}$$

we can construct a consistent PCM (Remark 1), containing the ratios of the coordinates

$$a_{i,j} = \frac{w_i^{(PCM)}}{w_j^{(PCM)}}, i = 1, 2, \dots, n, j = 1, 2, \dots, n. \tag{14}$$

Similarly, taking the ratios of the coordinates of (12), a PCM can be constructed, whose elements are the exponential values of the differences of the expected values: $a_{i,j}^{(BT)} = \exp(m_i - m_j)$ $i = 1, 2, \dots, n, j = 1, 2, \dots, n$. In the case of BT, due to (7) and (8), these ratios are exactly the ratios of the probabilities of ‘better’ and ‘worse’.

It is a well-known fact, that, in the case of AHP, either EM or LLSM is applied, supposing a complete and consistent PCM composed by (14), the starting vector $\underline{w}^{(PCM)}$ will be the result of the methods. On the other hand, if the PCM is complete and consistent, its elements can be expressed as the ratios of the components of a weight vector. This priority vector equals the result of the evaluation of the PCM via LLSM or EM. From Bozóki et al. (2010) it is also known that in the case of incomplete comparisons, the existence and uniqueness of the computed priority vector of EM and LLSM holds if and only if the representing graph of the comparisons (see Definition 5) is connected. This implies that even in the case of incomplete comparisons with a consistent IPCM constructed by (14), if the representing graph is connected, then the starting $\underline{w}^{(PCM)}$ would be recovered by LLSM and EM, too.

Accordingly, in the following, we state a theorem that proves that if the coefficients $D_{i,j,k}$ are the exact probabilities of the possible comparisons’ results, then the MLE recovers the exact expected values in both complete and incomplete cases.

First, let $K = \{(i, j) : i \neq j\}$ be the pairs of all possible comparisons. In the case of incomplete comparisons, let $I \subset K$ with pairs $i < j$. Every subset I defines a representing graph G_I , exactly as in Definition 5.

Theorem 1. Let F be the logistic c.d.f. (6). Let our starting vector be $\underline{m}^{(0)} = (m_1^{(0)}, \dots, m_n^{(0)})$, an arbitrary expected value vector with a fixed first coordinate. Let the data belonging to $\underline{m}^{(0)}$ be defined as follows:

$$D_{i,j,1}^{(d)} = F(m_j^{(0)} - m_i^{(0)}), \tag{15}$$

and

$$D_{i,j,2}^{(d)} = F(m_i^{(0)} - m_j^{(0)}). \tag{16}$$

Let the incomplete data matrix $D^{(m^{(0)})}$ be defined as follows: $D_{i,j,1}^{(m^{(0)})} = D_{j,i,2}^{(d)}$ and $D_{i,j,2}^{(m^{(0)})} = D_{j,i,1}^{(d)}$ if $(i, j) \in I$ and zero otherwise. Fix the first coordinate of \underline{m} at 0. If the graph belonging to I is connected, then the log-likelihood function,

$$\ln L(D^{(m^{(0)})} | \underline{m}) = \sum_{i < j, (i,j) \in I} \left(D_{i,j,1}^{(m^{(0)})} \cdot \ln p_{i,j,1} + D_{i,j,2}^{(m^{(0)})} \cdot \ln p_{i,j,2} \right), \tag{17}$$

attains its maximal value at $\hat{m} = \underline{m}^{(0)}$ and the argument of the maximal value is unique.

Proof. First we note that the connectedness of the graph defined by I and its strongly connected property are equivalent in this case, as both $D_{i,j,1}^{(m^{(0)})}$ and $D_{i,j,2}^{(m^{(0)})}$ are positive in the case of $(i, j) \in I$. Recalling that $D_{i,j,1}^{(m^{(0)})} + D_{i,j,2}^{(m^{(0)})} = 1$ and $p_{i,j,2} = 1 - p_{i,j,1}$, (17) has the following form:

$$B(x_1, \dots, x_k) = \sum_{s=1}^k C_s \cdot \ln x_s + (1 - C_s) \cdot \ln(1 - x_s), 0 < x_s < 1, 0 < C_s < 1. \tag{18}$$

where $C_s = D_{i,j,1}^{(m^{(0)})}$, $x_s = p_{i,j,1}$, k is the number of elements (i, j) in the set I for which $i < j$ and we made a one-to-one correspondence Z between these pairs in I and the positive integers $1, 2, \dots, k$, $Z(i, j) = s, (i, j) \in I, i < j$.

If we consider the multivariate function (18), we can see that it takes its maximal value at $\tilde{x}_s = C_s, s = 1, 2, \dots, k$, and the argument of the maximum is unique. Now the question is whether \underline{C} can be expressed by an expected value vector, for example in the form $C_s = F(m_j - m_i)$, $Z(i, j) = s$. It does not hold for an arbitrary vector \underline{C} , but since $C_s = D_{i,j,1}^{(m^{(0)})}$, ($Z(i, j) = s$), then $\underline{m}^{(0)}$ provides a possible solution for the above question, $C_s = F(m_j^{(0)} - m_i^{(0)})$, due to the construction of the coefficients. On the other hand, taking into account the connectedness of the graph defined by I , the condition in Ford Jr (1957) is satisfied, therefore the argument maximizing (17), \hat{m} , is unique. These together imply that $\hat{m} = \underline{m}^{(0)}$. ■

First, we note that Theorem 1 remains true if the coefficients C_s are multiplied by a positive constant value, therefore only the ratios of $D_{i,j,k}^{(m^{(0)})}$ are relevant.

Theorem 1 states that if the coefficients are the exact probabilities belonging to an expected value vector, then the maximum likelihood estimation reproduces the exact expected value vector. In addition, starting with an expected value vector, computing the exact probabilities by (15) and (16), moreover constructing a consistent IPCM matrix from the ratios of the probabilities $p_{i,j,2}/p_{i,j,1}$, LLSM, EM and the priority vector of BT (12) are the same weight vectors. In these special cases the results of the three methods correspond even for incomplete comparisons. The fact that the coefficients are the exact probabilities belonging to an expected value vector means that there is no contradiction in them. On the other hand, the property of ‘free of contradictions’ is described by the concept of consistency in the case of PCM and IPCM matrices. Additionally, in Section 3.2 the data $D_{i,j,1}$ and $D_{i,j,2}$ are integers, not probabilities. Still, if they are all positive, a complete PCM matrix can be composed by taking the ratios $D_{i,j,2}/D_{i,j,1}$ and $D_{i,j,1}/D_{i,j,2}$. If for a certain pair (i, j) $D_{i,j,1} = 0$ and/or $D_{i,j,2} = 0$, the IPCM matrix will not contain the element $a_{i,j}$ and $a_{j,i}$.

The consistency of an incomplete PCM is given in Definition 6. Taking this definition as the basis, we define the concept of consistent data in the case of the BT model as follows. Let I be the subset of those pairs for which there exist comparisons and both data are positive, i.e, $I = \{(i, j) : i \neq j, 0 < D_{i,j,1}, 0 < D_{i,j,2}\} \subseteq K$.

Definition 7. Let $I \subset K$, and let the graph G_I be connected. Introduce $h_{i,j} = \frac{D_{i,j,2}}{D_{i,j,1}}, (i, j) \in I$. Data matrix D is called consistent in I , if for any cycle in G_I , that is $(i_1, i_2, \dots, i_k, i_1)$, with $(i_l, i_{l+1}) \in I$ and $(i_k, i_1) \in I$

$$h_{i_1, i_2} \cdot h_{i_2, i_3} \cdot \dots \cdot h_{i_k, i_1} = 1. \tag{19}$$

Data matrix D is called inconsistent in I if it is not consistent in I .

Data consistency is defined by the consistency of IPCM constructed by the ratios of the appropriate data in D . Now we prove the following statement, which is an equivalent form, but presents another feature of data consistency in the BT model:

Theorem 2. Let I define a connected graph. Data matrix D is consistent in I if and only if there exists an $\underline{m}^{(0)} \in \mathbb{R}^n$, $\underline{m}^{(0)} = (0, m_2^{(0)}, \dots, m_n^{(0)})$ for which

$$\frac{D_{i,j,2}}{D_{i,j,1}} = \frac{F(m_i^{(0)} - m_j^{(0)})}{F(m_j^{(0)} - m_i^{(0)}), (i, j) \in I \tag{20}$$

Proof. As a first step, let I be a spanning tree. As $0 < D_{i,j,1}$ and $0 < D_{i,j,2}$ if $(i, j) \in I$, we know that the graph G_I is strongly connected, the MLE exists and it is unique. Let us investigate the log-likelihood function in the form of a multivariate function, as in [Theorem 1](#), with the coefficients $D_{i,j,1}$ and $D_{i,j,2}$ (see [\(18\)](#)). Taking the partial derivatives with respect to x_s , we can conclude that the maximum is at $\hat{x}_s = D_{i,j,1}/(D_{i,j,1} + D_{i,j,2})$. We have to prove that there exists an n -dimensional vector $\underline{m}^{(0)} = (0, m_2^{(0)}, \dots, m_n^{(0)})$, for which $D_{i,j,1}/(D_{i,j,1} + D_{i,j,2}) = F(m_j^{(0)} - m_i^{(0)})$. For that, consider the following form:

$$m_j - m_i = F^{-1} \left(\frac{D_{i,j,1}}{D_{i,j,1} + D_{i,j,2}} \right). \tag{21}$$

It is easy to see that this system of linear equations can be uniquely solved if $(i, j) \in I$, after fixing $m_1 = 0$. Denoting the solution by $\underline{m}^{(0)} = (0, m_2^{(0)}, \dots, m_n^{(0)})$, we get

$$\frac{D_{i,j,2}}{D_{i,j,1}} = \frac{F(m_i^{(0)} - m_j^{(0)})}{F(m_j^{(0)} - m_i^{(0)})} = \frac{\exp(m_i^{(0)})}{\exp(m_j^{(0)})}. \tag{22}$$

We note that if I is a spanning tree, but $D_{i,j,1} = 0$ or $D_{i,j,2} = 0$ for some $(i, j) \in I$, then the MLE does not exist, as the necessary and sufficient condition given by Ford is not satisfied.

In the second step, let I be a general subset of K with $0 < D_{i,j,1}$ and $0 < D_{i,j,2}$ if $(i, j) \in I$, moreover assume that I defines a connected graph. Take a spanning tree $I_b \subseteq I$. Let $\underline{m}^{(0)} = \hat{m}_{I_b}$, the MLE of the expected value vector belonging to the data set $D_{i,j,k}, (i, j) \in I_b$. Now [\(20\)](#) is satisfied if $(i, j) \in I_b$. If $(i, j) \notin I_b$, there exists a cycle in I , $(i_1, i_2, \dots, i_k, i_1)$, $i_1 = i, i_k = j, (i_1, i_2), (i_2, i_3), \dots, (i_{k-1}, i_k) \in I_b$. Therefore, for $l = 1, 2, \dots, k - 1$,

$$h_{i_l, i_{l+1}} = \frac{D_{i_l, i_{l+1}, 2}}{D_{i_l, i_{l+1}, 1}} = \frac{F(m_{i_{l+1}}^{(0)} - m_{i_l}^{(0)})}{F(m_{i_l}^{(0)} - m_{i_{l+1}}^{(0)})} = \frac{\exp(m_{i_l}^{(0)})}{\exp(m_{i_{l+1}}^{(0)})}. \tag{23}$$

Taking the productions of these quantities along the path from i to j and applying [\(19\)](#), the left-hand side is $\frac{1}{h_{i,j}} = h_{i,j}$, while the right hand side equals

$$\frac{\exp(m_i^{(0)})}{\exp(m_j^{(0)})} = \frac{F(m_i^{(0)} - m_j^{(0)})}{F(m_j^{(0)} - m_i^{(0)})}, \tag{24}$$

which proves [\(20\)](#) for any $(i, j) \in I$.

Conversely, assume that

$$\frac{D_{i,j,2}}{D_{i,j,1}} = \frac{F(m_i^{(0)} - m_j^{(0)})}{F(m_j^{(0)} - m_i^{(0)})} \tag{25}$$

if $(i, j) \in I$. Applying again the equality

$$\frac{F(m_i^{(0)} - m_j^{(0)})}{F(m_j^{(0)} - m_i^{(0)})} = \frac{\exp(m_i^{(0)})}{\exp(m_j^{(0)})}, \tag{26}$$

and taking their product on a cycle $(i_1, i_2, \dots, i_k, i_1)$ with $(i_l, i_{l+1}) \in I, (i_k, i_1) \in I$, after simplification we get [\(19\)](#), namely

$$h_{i_1, i_2} \cdot h_{i_2, i_3} \cdot \dots \cdot h_{i_k, i_1} = 1. \tag{27}$$

■

We note, that if D is consistent in I and G_I is a connected graph, then $\underline{m}^{(0)} = (0, m_2^{(0)}, \dots, m_n^{(0)})$ in [\(20\)](#) is unique. Recalling [Theorem 1](#), $\underline{m}^{(0)}$ equals to the evaluation of the MLE. The conclusion of [Section 4.1](#) is that in the case of consistent comparison data D and connected

representing graph, the evaluations of BT, LLSM and EM provide the same priority vector. In other words, the results of these three methods coincide both in complete and incomplete cases. The ratio $\frac{D_{i,j,2}}{D_{i,j,1}}$ is common in sports results' evaluations: the ratios of the number of wins and defeats often appear in PCMs (for example in [Bozóki et al., 2016](#)). Therefore, in this case, supposing consistent data, the evaluation of BT, LLSM and EM would be the same. As an example, consider

$$D = \begin{pmatrix} D_{1,2,1} & D_{1,2,2} \\ D_{1,3,1} & D_{1,3,2} \\ D_{1,4,1} & D_{1,4,2} \\ D_{2,3,1} & D_{2,3,2} \\ D_{2,4,1} & D_{2,4,2} \\ D_{3,4,1} & D_{3,4,2} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 1 \\ 1 & 1 \\ 2 & 1 \\ 2 & 1 \end{pmatrix}. \tag{28}$$

The PCM constructed based on these data is

$$A = \begin{pmatrix} 1 & \frac{D_{1,2,2}}{D_{1,2,1}} & \frac{D_{1,3,2}}{D_{1,3,1}} & \frac{D_{1,4,2}}{D_{1,4,1}} \\ \frac{D_{1,2,1}}{D_{1,2,2}} & 1 & \frac{D_{2,3,2}}{D_{2,3,1}} & \frac{D_{2,4,2}}{D_{2,4,1}} \\ \frac{D_{1,3,1}}{D_{1,3,2}} & \frac{D_{2,3,1}}{D_{2,3,2}} & 1 & \frac{D_{3,4,2}}{D_{3,4,1}} \\ \frac{D_{1,4,1}}{D_{1,4,2}} & \frac{D_{2,4,1}}{D_{2,4,2}} & \frac{D_{3,4,1}}{D_{3,4,2}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 2 & 1 \\ \frac{1}{2} & 1 & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & 1 & \frac{1}{2} \\ 1 & 2 & 2 & 1 \end{pmatrix}. \tag{29}$$

One can easily check that data matrix D is consistent and so is A . Performing MLE in BT model, the estimated expected value vector equals

$$\hat{m} = (0, -0.693, -0.693, 0) \tag{30}$$

and the priority vector is

$$\underline{\hat{w}}^{(BT)} = \left(\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3} \right). \tag{31}$$

One can check that, applying PCM A , LLSM and EM also provide the same priority vector.

4.2. Inconsistent data matrices' evaluations

In [Section 4.1](#), we proved that in the case of consistent comparison data, LLSM, EM and BT result in the same priority vectors. However, the following example demonstrates that in case of inconsistent data, the three methods provide different priority vectors both for complete and incomplete cases.

Let $n=4$, $\underline{m}^{(0)} = (0, 0.25, 0.75, 1.75)$. The probabilities of the comparisons' results are

$$D^{(m^{(0)})} = \begin{pmatrix} p_{1,2,1} & p_{1,2,2} \\ p_{1,3,1} & p_{1,3,2} \\ p_{1,4,1} & p_{1,4,2} \\ p_{2,3,1} & p_{2,3,2} \\ p_{2,4,1} & p_{2,4,2} \\ p_{3,4,1} & p_{3,4,2} \end{pmatrix} = \begin{pmatrix} 0.562 & 0.438 \\ 0.679 & 0.321 \\ 0.852 & 0.148 \\ 0.622 & 0.378 \\ 0.818 & 0.182 \\ 0.731 & 0.269 \end{pmatrix}. \tag{32}$$

These data are consistent. If we modify $p_{3,4,1}$ and $p_{3,4,2}$ by subtracting/adding 0.2, we get the modified data matrix

$$D^{(mod)} = \begin{pmatrix} p_{1,2,1} & p_{1,2,2} \\ p_{1,3,1} & p_{1,3,2} \\ p_{1,4,1} & p_{1,4,2} \\ p_{2,3,1} & p_{2,3,2} \\ p_{2,4,1}^{(mod)} & p_{2,4,2}^{(mod)} \\ p_{3,4,1}^{(mod)} & p_{3,4,2}^{(mod)} \end{pmatrix} = \begin{pmatrix} 0.562 & 0.438 \\ 0.679 & 0.321 \\ 0.852 & 0.148 \\ 0.622 & 0.378 \\ 0.818 & 0.182 \\ 0.531 & 0.469 \end{pmatrix}. \tag{33}$$

One can check that $D^{(mod)}$ is inconsistent. The corresponding PCM is

$$A = \begin{pmatrix} 1 & \frac{p_{1,2,2}}{p_{1,2,1}} & \frac{p_{1,3,2}}{p_{1,3,1}} & \frac{p_{1,4,2}}{p_{1,4,1}} \\ \frac{p_{1,2,1}}{p_{1,2,2}} & 1 & \frac{p_{2,3,2}}{p_{2,3,1}} & \frac{p_{2,4,2}}{p_{2,4,1}^{(mod)}} \\ \frac{p_{1,3,1}}{p_{1,3,2}} & \frac{p_{2,3,1}}{p_{2,3,2}} & 1 & \frac{p_{3,4,2}^{(mod)}}{p_{3,4,1}^{(mod)}} \\ \frac{p_{1,4,1}}{p_{1,4,2}} & \frac{p_{2,4,1}^{(mod)}}{p_{2,4,2}^{(mod)}} & \frac{p_{3,4,1}^{(mod)}}{p_{3,4,2}^{(mod)}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.779 & 0.472 & 0.174 \\ 1.284 & 1 & 0.607 & 0.223 \\ 2.117 & 1.649 & 1 & 0.883 \\ 5.755 & 4.482 & 1.132 & 1 \end{pmatrix}.$$

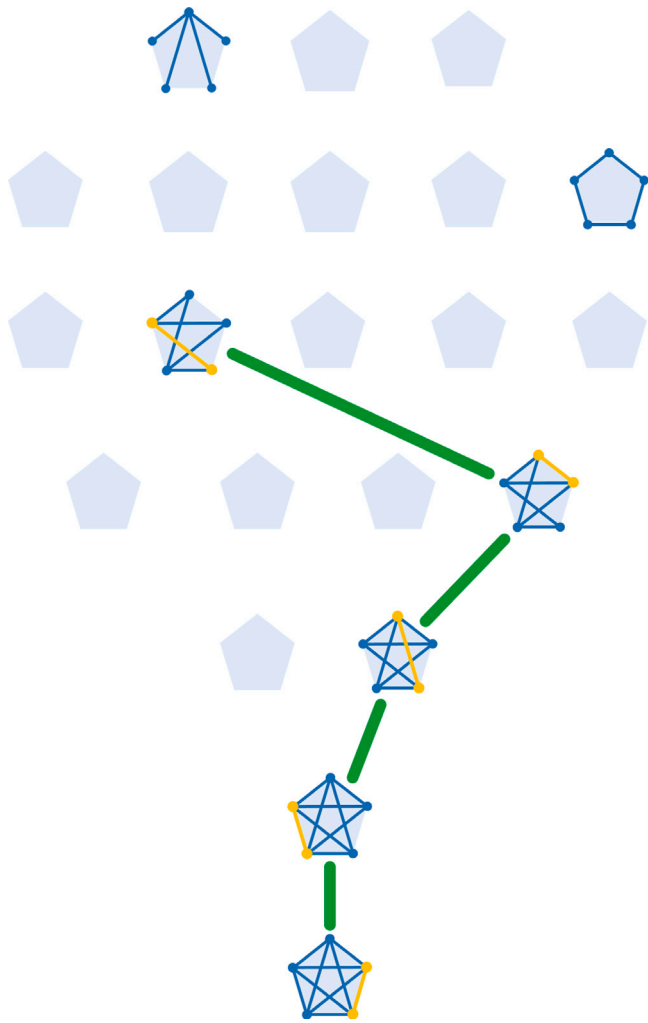


Fig. 1. Optimal graphs for $n = 5, e = 4, 5, \dots, 10$.

that is we evaluate incomplete comparisons.

$$D_I^{(mod)} = \begin{pmatrix} p_{1,2,1} & p_{1,2,2} \\ p_{1,3,1} & p_{1,3,2} \\ p_{1,4,1} & p_{1,4,2} \\ p_{3,4,1}^{(mod)} & p_{3,4,2}^{(mod)} \end{pmatrix} = \begin{pmatrix} 0.562 & 0.438 \\ 0.679 & 0.321 \\ 0.852 & 0.148 \\ 0.531 & 0.469 \end{pmatrix} \quad (40)$$

Now, the incomplete PCM A_I looks as follows,

$$A_I = \begin{pmatrix} 1 & \frac{p_{1,2,2}}{p_{1,2,1}} & \frac{p_{1,3,2}}{p_{1,3,1}} & \frac{p_{1,4,2}}{p_{1,4,1}} \\ \frac{p_{1,2,1}}{p_{1,2,2}} & 1 & * & * \\ \frac{p_{1,3,1}}{p_{1,3,2}} & * & 1 & \frac{p_{3,4,2}^{(mod)}}{p_{3,4,1}^{(mod)}} \\ \frac{p_{1,4,1}}{p_{1,4,2}} & * & \frac{p_{3,4,1}^{(mod)}}{p_{3,4,2}^{(mod)}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.779 & 0.472 & 0.174 \\ 1.284 & 1 & * & * \\ 2.117 & * & 1 & 0.883 \\ 5.755 & * & 1.132 & 1 \end{pmatrix} \quad (41)$$

It is inconsistent for the same reason as it was presented in (35). Its evaluation by LLSM and EM are

$$\underline{w}_I^{(LLSM)} = (0.106, 0.136, 0.301, 0.457), \quad (42)$$

and

$$\underline{w}_I^{(EM)} = (0.107, 0.134, 0.302, 0.458), \quad (43)$$

respectively. Applying the incomplete and inconsistent data matrix $D_I^{(mod)}$, estimating the expected values by BT via MLE and taking the priority vector, the result is different from (42) and also from (43), namely

$$\hat{m}_I = (0, 0.250, 1.017, 1.366) \quad (44)$$

and

$$\hat{w}_I^{(BT)} = (0.112, 0.143, 0.308, 0.437). \quad (45)$$

This incomplete example demonstrates that all three investigated methods, namely BT, LLSM and EM may provide different results during the evaluations.

5. Evaluating information retrieval from incomplete comparisons

The information retrieval problem investigated here is strongly connected to the fact that the comparisons provided by the decision-makers are always costly. Thus, the number of comparisons can easily become a bottleneck. However, choosing the most appropriate set of comparisons helps to retrieve the most information estimating the preferences of the decision-maker, which can be interpreted as the optimal usage of this costly resource. Our approach can be especially attractive in the case, when the number of the provided comparisons is uncertain a priori, e.g., in online questionnaires.

Rather than establishing presumably too complex analytical results, we perform Monte-Carlo simulations to investigate the information retrieval from incomplete comparisons in the case of the BT model. The large number of connected subgraphs (21 in the case of $n = 5$ and 112 in the case of $n = 6$) also suggests an extensive computational approach.

The steps of the simulations with explanations are the following.

1. Generate n random integer values from 1 to 9 with uniform distribution and normalize them. In this way, a random weight (priority) vector is created. This will be the initial priority vector belonging to BT.
2. Take the logarithm of all coordinates and decrease the new components by the first one. A random vector with zero first coordinate is now constructed. This vector serves as the initial random expected value vector in the BT model.
3. Calculate the exact probabilities of the comparison results by (15) and (16) for all possible pairs. A consistent and complete system of the comparisons' results has been generated.

One can easily check that A is not consistent. For example, take

$$\frac{a_{1,3}}{a_{1,4}} = \frac{0.472}{0.174} = 2.713 \neq a_{4,3} = 1.132. \quad (34)$$

The result vectors computed by LLSM and EM are

$$\underline{w}^{(LLSM)} = (0.105, 0.135, 0.276, 0.484), \quad (35)$$

$$\underline{w}^{(EM)} = (0.103, 0.132, 0.279, 0.485). \quad (36)$$

If we evaluate data matrix $D^{(mod)}$ by BT via MLE, the estimated expected values are

$$\hat{m} = (0, 0.246, 0.954, 1.450). \quad (37)$$

Transforming this into a weight vector we get

$$\hat{w}^{(BT)} = (0.109, 0.140, 0.284, 0.466). \quad (38)$$

$\hat{w}^{(BT)}$ differs from $\underline{w}^{(LLSM)}$ and $\underline{w}^{(EM)}$, demonstrating that, in the case of inconsistent data, the evaluations of LLSM, EM and BT provide different results. We can observe the same phenomenon if we restrict the data to the set

$$I = \{(1, 2), (1, 3), (1, 4), (3, 4), (2, 1), (3, 1), (4, 1), (4, 3)\},$$

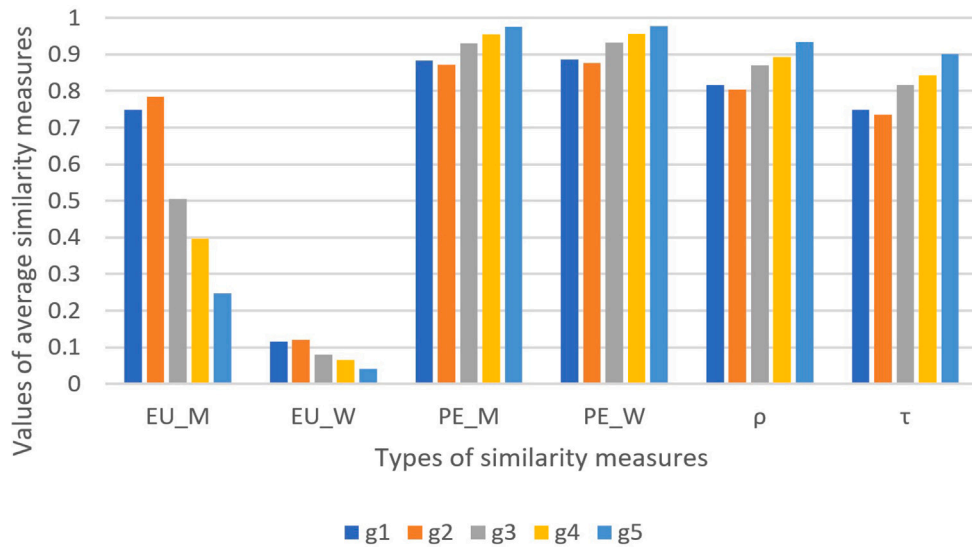


Fig. 2. Average measures of information retrieval in the case of different graph structures: average distances defined by (46), (47), average correlations defined by (48), (49), (50) and (51), respectively, comparing $n = 4$ objects and applying 0.15 perturbation.

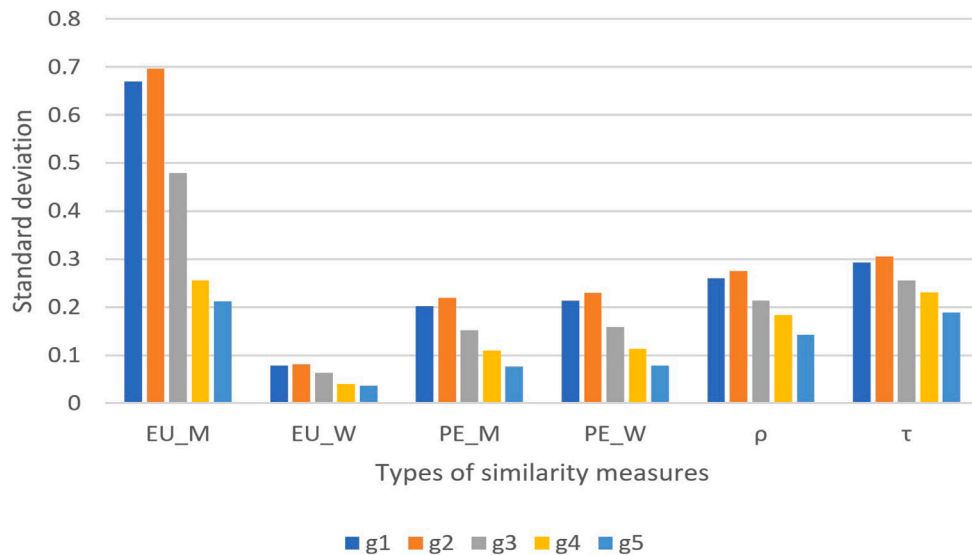


Fig. 3. Standard deviations of the different measures defined by (46),(47),(48),(49),(50) and (51) in the case of different graph structures comparing $n = 4$ objects and applying 0.15 perturbation.

- As previously presented, based on consistent comparison results MLE recovers the initial expected values both in complete and incomplete cases. Because of that, perturbations are made on the consistent results as follows: each value $D_{i,j,1}^{(m(0))}$ are modified by adding to them independent uniformly distributed random numbers on $[-perturb, perturb]$. The value of $perturb$ can be chosen between 0 and 1. The perturbed values $^{pert}D_{i,j,1}^{(m(0))}$ are claimed to be strictly between 0 and 1.
- Compute $^{pert}D_{i,j,2}^{(m(0))}$ by $^{pert}D_{i,j,2}^{(m(0))} = 1 - ^{pert}D_{i,j,1}^{(m(0))}$. As $0 < ^{pert}D_{i,j,1}^{(m(0))} < 1$, so is $^{pert}D_{i,j,2}^{(m(0))}$.
- Use the previously constructed complete comparisons' results as data in the BT model, and evaluate them via MLE. The outcome expected value vector is \hat{m}^K . The outcome priority vector is \hat{w}^K .
- Compute the results belonging to the different graph structures as follows. For each possible connected graph I : apply only the data belonging to the graph structure and perform MLE in the BT model. By omitting a subset of comparisons incomplete and inconsistent data are considered. The existence and uniqueness

of the evaluation are guaranteed by the connectedness and by $0 < ^{pert}D_{i,j,1}^{(m(0))} < 1$ and $0 < ^{pert}D_{i,j,2}^{(m(0))} < 1$. The results of the evaluation in the case of graph structure I are denoted by \hat{m}^I and \hat{w}^I .

- The differences between the results belonging to the complete case and the graph structure I can be defined by several measures. We used Euclidean distance of the estimated expected value vectors

$$EU_M = EU(\hat{m}^K, \hat{m}^I) = \sqrt{\sum_{i=1}^n (\hat{m}_i^K - \hat{m}_i^I)^2}, \tag{46}$$

the Euclidean distances of the estimated weight vectors

$$EU_W = EU(\hat{w}^K, \hat{w}^I) = \sqrt{\sum_{i=1}^n (\hat{w}_i^K - \hat{w}_i^I)^2}, \tag{47}$$

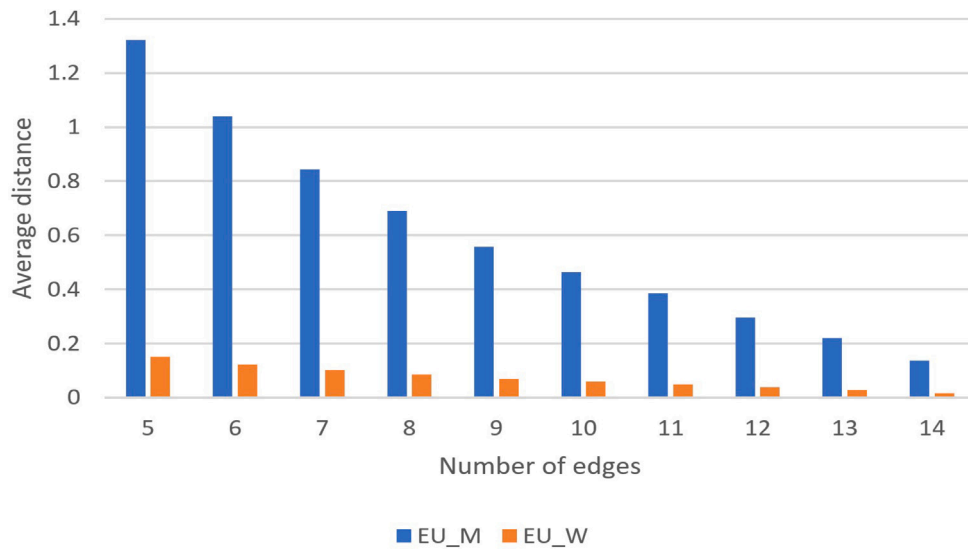


Fig. 4. Average EU_M and EU_W distances defined by (46) and (47) in the function of the number of graphs' edges in the case of $n = 6$ and perturbation 0.15.

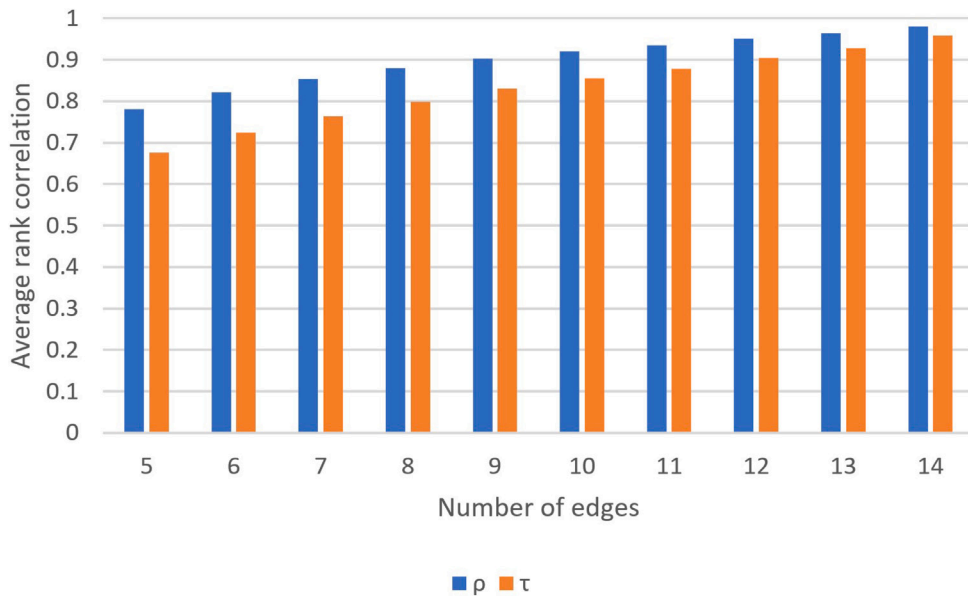


Fig. 5. Average rank correlations Spearman ρ and Kendall τ (defined by (50) and (51)) in the function of the number of graphs' edges in the case of $n = 6$ and perturbation 0.15.

Pearson correlation coefficient belonging to the expected values

$$PE_M = PE(\hat{w}^K, \hat{w}^I) = \frac{\frac{\sum_{i=1}^n \hat{m}_i^K \cdot \hat{m}_i^I}{n} - \bar{\hat{m}}^K \cdot \bar{\hat{m}}^I}{\sqrt{\frac{\sum_{i=1}^n (\hat{m}_i^K - \bar{\hat{m}}^K)^2}{n}} \sqrt{\frac{\sum_{i=1}^n (\hat{m}_i^I - \bar{\hat{m}}^I)^2}{n}}}, \quad (48)$$

and belonging to the priority vectors

$$PE_W = PE(\hat{w}^K, \hat{w}^I) = \frac{\frac{\sum_{i=1}^n \hat{w}_i^K \cdot \hat{w}_i^I}{n} - \bar{\hat{w}}^K \cdot \bar{\hat{w}}^I}{\sqrt{\frac{\sum_{i=1}^n (\hat{w}_i^K - \bar{\hat{w}}^K)^2}{n}} \sqrt{\frac{\sum_{i=1}^n (\hat{w}_i^I - \bar{\hat{w}}^I)^2}{n}}}. \quad (49)$$

To investigate the similarities of the rankings, the Spearman ρ rank correlation

$$\rho = \rho(\hat{m}^K, \hat{m}^I) = 1 - \frac{\sum_{i=1}^n d_i^2}{6n(n^2 - 1)} \quad (50)$$

where d_i is the difference in rankings of the i th object, and the Kendall τ rank correlations

$$\tau = \tau(\hat{m}^K, \hat{m}^I) = \frac{2}{n(n-1)} \sum_{i < j} \text{sign}(\hat{m}_i^K - \hat{m}_j^K) \cdot \text{sign}(\hat{m}_i^I - \hat{m}_j^I) \quad (51)$$

were applied. These rank correlations are the same either the expected value vectors' or the priority vectors' rankings are considered. The above-mentioned measurements characterize the deviations and the similarities in a different manner, hence providing different measures of information retrieval. In the case of Euclidean distances, the smaller the value, the more information is regained, and the less information is lost due to the incomplete comparisons. In the case of the correlation coefficients, the closer the result is to 1, the more information is recovered. We call these measurements similarity measures and compute all of them for the different graph structures.

- Repeat the steps 1–8 N times, where N is the number of simulations. These quantities are, of course, random quantities depending on the random values of the data matrix. Take the

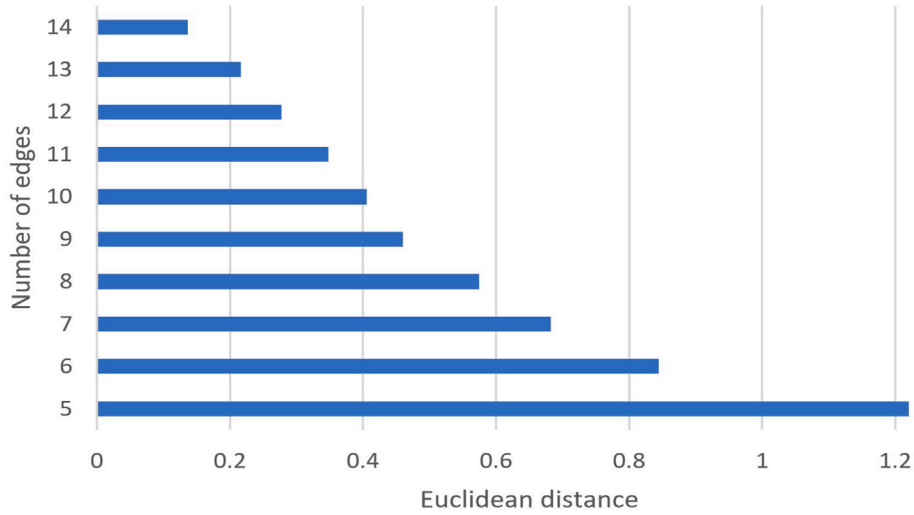


Fig. 6. Euclidean distances between the priority vectors belonging to the optimal graph structure with a given edge number and the complete graph (defined by (46)) in the function of the number of graphs' edges in the case of $n = 6$ objects to compare and perturbation 0.15.

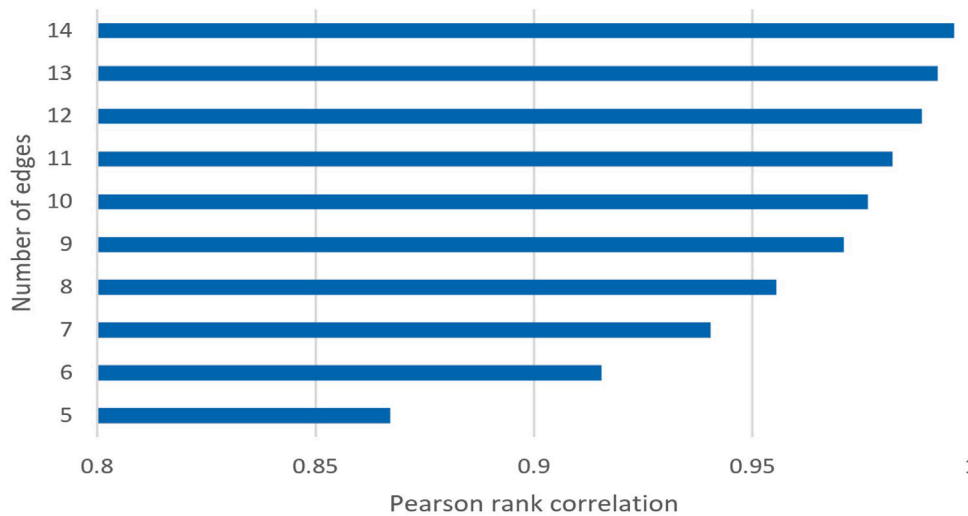


Fig. 7. Pearson correlation coefficients between the optimal graphs and the complete graph in the function of the number of graphs' edges in the case of $n = 6$ objects to compare and perturbation 0.15.

average and the dispersion of the similarity measures over the simulations.

We do not know the bounds for the values \hat{m} but we do for the other measurements. As they are all bounded, therefore their expected values and variances are finite. Consequently, both the law of large numbers and the central limit theorem can be applied to them. The central limit theorem guarantees the $(u_\alpha \cdot \sigma) / \sqrt{N}$ upper bound for the simulation error with reliability $1-\alpha$ if σ is an upper bound for the standard deviation of the random quantity, and $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$. Now, the Euclidean distance of the weight vectors is less than $\sqrt{2}$, the correlation coefficients are between -1 and 1 , therefore the standard deviations of these quantities are less than or equal to $\frac{\sqrt{2}}{2}$ and 1 , respectively. Using the upper bound 1 , $(u_\alpha \cdot \sigma) / \sqrt{N} = (2.58 \cdot 1) / \sqrt{N}$ when the reliability is $1-\alpha = 0.99$. $N = 10^6$ simulations were carried out, thus the theoretical upper bound of the simulation error is 0.00258 . Of course, the estimated standard deviations are much less than the theoretical upper bounds, therefore the true simulation errors are much smaller than the theoretical values. They are usually of different order

of magnitude. Narrow confidence intervals can be given in which the means of the quantities are situated with a given reliability.

6. Simulation results

The examined number of objects to compare are $n = 4, 5$ and 6 . In the case of $n = 4$, together with the complete case, 6 graphs can be distinguished for defining the set of comparisons. Two of them have 3, two of them have 4, one and one have 5 and 6 edges, respectively. In the case of $n = 5$ objects, the number of different graphs is 21. While in the case of $n = 6$, there are 112 different possible graph structures. We can observe that in almost all cases, for a given number of objects and a given number of comparisons each similarity measure marks the same graph to be the best. There are few exceptions for $n = 6$ and large numbers of edges, namely if $e = 12$ and 13 . In the case of $e = 12$ the best structure by ρ differs from the best marked by the other measures. If the number of edges equals 13, then the rank correlations ρ and τ indicate graph g109 to be the best, while the other measures indicate g110. Although the numerical differences are small, this phenomenon appears independently of the value of the perturbation, moreover, in the case

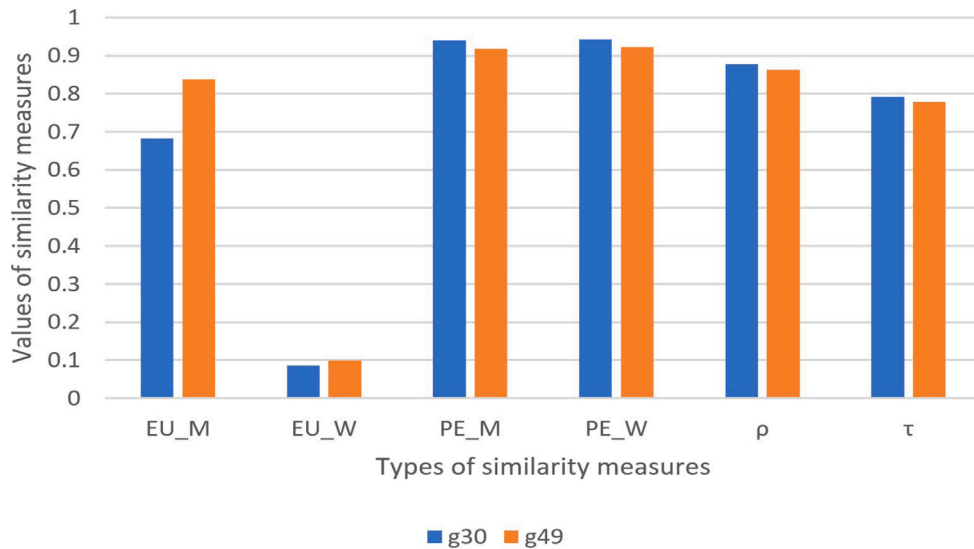


Fig. 8. Comparison of the best graph structure with 7 edges and the worst graph structure with 8 edges in the case of $n = 6$ objects to compare and 0.15 perturbation.

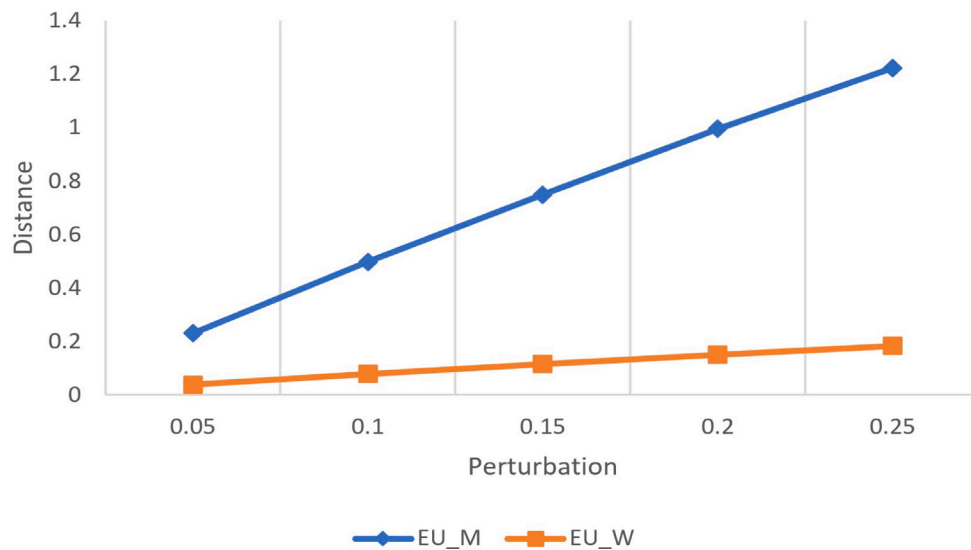


Fig. 9. Euclidean distances EU_M and EU_W (defined by (46) and (47)) between the results belonging to the star graph and to the complete graph in the function of the value of the perturbation in the case of $n = 4$ objects to compare.

of PCM methods, too. The optimal graph structures, sequences and the exceptions are the same as it was presented in Bozóki and Szádóczki (2022). Fig. 1 shows the optimal graphs for the different completions levels. Graphs with e and $e + 1$ ($e = 6, 7, 8$) edges are reachable from each other by the addition/deletion of a single (orange) edge.

This observation supports that the most information retrieval belongs to the structure of the graph. The independence of the rankings from the similarity measures can be seen in Fig. 2: graphs g1, g2, ..., and g5 (the complete graph g6 is omitted) are in the same ranking according to all 6 measurements. It is important to note that the smaller EU_M and EU_W indicate the better information retrieval, while in the case of PE_M, PE_W, ρ and τ the larger value is preferred. Fig. 3 demonstrates that even the standard deviations work the same way in the case of choosing the best graph, keeping in mind that the smaller the standard deviation, the smaller the volatility in every measurement. See, for example, τ defined by (51): the graph structure g2 is the worst and g5 is the best, as they have the smallest and largest average values, respectively. Investigating the dispersion, the largest value belongs to g2 and the smallest one is for the graph structure g5.

Based on the simulations, we can conclude that the larger the number of edges, the better the information retrieval. This observation is true in average and for the best graphs, as well, for all the examined number of objects (4, 5 and 6), for all the perturbation values and for all the investigated measurements. Fig. 4 presents the average distances, when the averages contain the data of the graphs with 5, 6, ..., and 14 edges independently of the graph structures. Fig. 5 represents the averaged rank correlations ρ and τ (averages are taken by the number of edges) in the case of $n = 6$ items and 0.15 perturbation. Figs. 6 and 7 show the values belonging to the optimal graphs in the case of fixed edge numbers. One can see the monotone decreasing features of the Euclidean distances and the monotone increasing feature of the correlations, as expected.

In the case of 4 objects, the weakest graph structure with 4 edges is better than the best graph structure with 3 edges. This is true for 4 and 5 edges, too. In the case of 5 and 6 objects, although the best graphs with k edges are always weaker than the best graphs with $k + 1$ edges, nevertheless one can find such cases when the best graph with k edges is better than the worst graph with $k + 1$ edges. We draw the attention

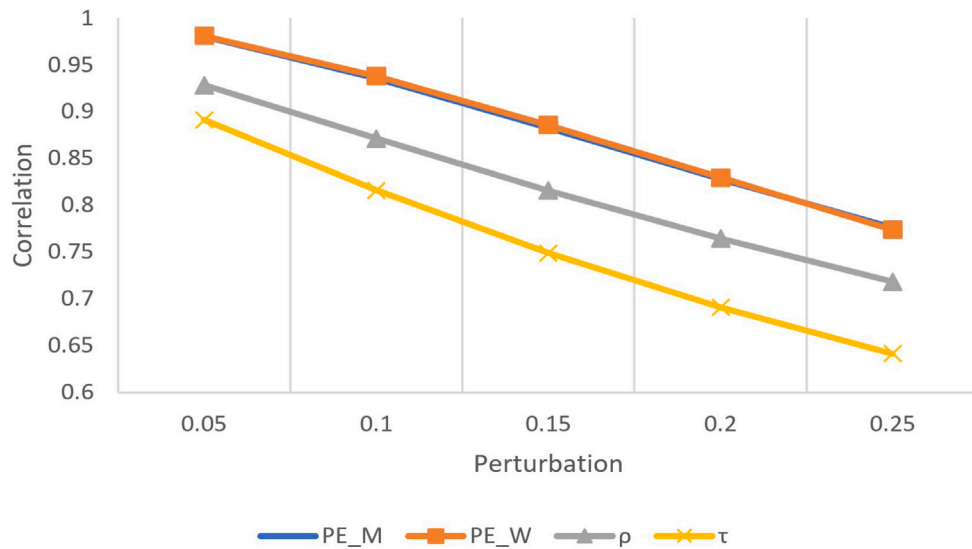


Fig. 10. Correlation coefficients PE_M, PE_W, Spearman ρ and Kendall τ defined by (48),(49),(50) and (51), respectively, in the function of the perturbation values in the case of $n = 4$ objects to compare. PE_M and PE_W are very close to each other.

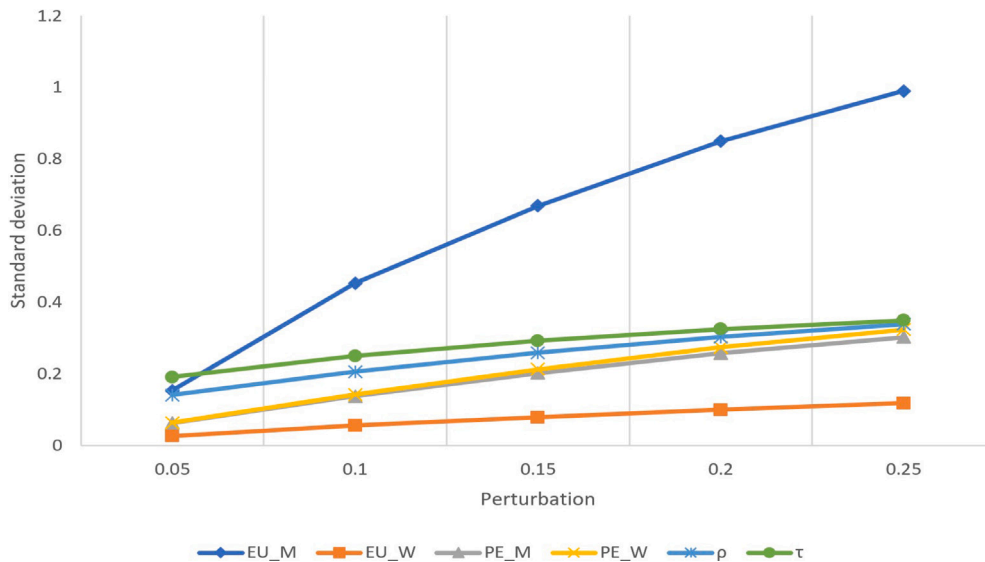


Fig. 11. The standard deviations of the similarity measures EU_M, EU_W, PE_M, PE_W, ρ and τ defined by (46), (47), (48), (49), (50) and (51), respectively, in the function of the perturbation values in the case of star graph with $n = 4$ objects to compare.

that every measure indicates this observation, which is presented in Fig. 8 for g_{30} and g_{49} . The same conclusion can be drawn for all perturbation values. This fact also supports that the reason resides in the structures of the graphs. Therefore it is worth planning the structure of the comparisons to gain the most information that is possible.

Fixing a graph structure, let us investigate the similarity measures in the function of the perturbation. We can realize that by increasing the perturbation, the distances get larger, while the correlations get smaller, as expected. The standard deviations of the similarity measures are also growing in all cases. Presenting the data belonging to the star graph, Figs. 9–11 illustrate this ascertainment, too.

The smallest number of comparisons, if the results can be evaluated, equals $n - 1$, therefore it is especially interesting which comparison graph is worth using to get as much information as possible. According to the simulation results, the star graphs proved to be the best whatever similarity measure and perturbation were applied, independently of

the number of compared items. This result coincides with the observation published in Bozóki and Szádóczi (2022) for the EM and LLSM methods, too. In the case of $n = 6$ objects and 0.15 perturbation, the correlation coefficients are presented in Fig. 12. The first (black) columns belong to the star graph. It is clear that they are the highest.

As in Orbán-Mihálykó et al. (2019b) were proved, the axiomatic properties of the Thurstone-motivated models are similar, independently of the distribution of the random variables. As a robustness analysis, the simulations were also performed for the Thurstone model (standard normal c.d.f. instead of logistic (6)) allowing two options in decisions (worse and better). Based on those, we can conclude that the observations stated in this section are all valid for this model, as well. Once again, this supports the ascertainment that the information retrieval depends on the structure of the comparisons' graph and not on the model itself, and the method of the evaluation. The detailed data of the simulations are available in the online appendix at <https://math.uni-pannon.hu/~orban>.

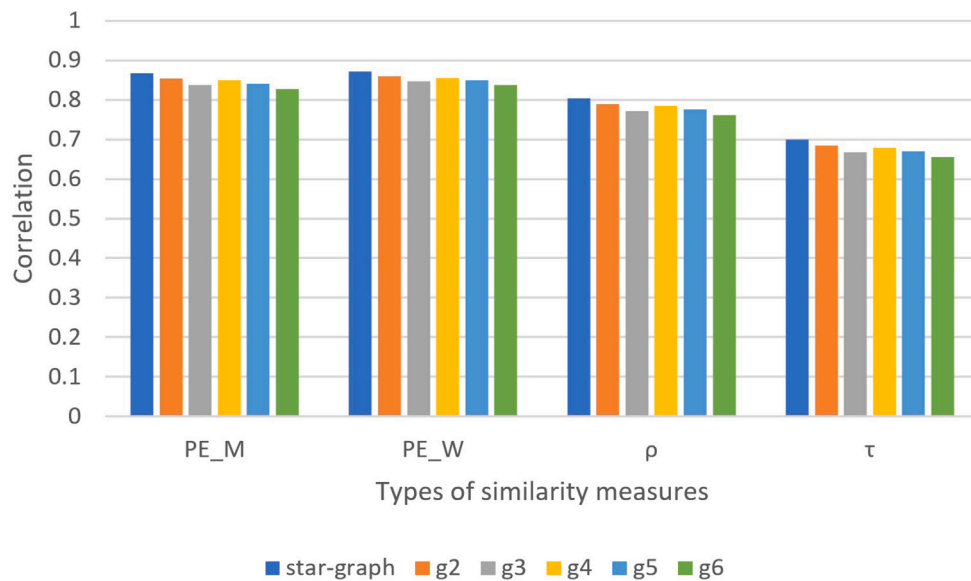


Fig. 12. The correlation coefficients between the results belonging to the different graph structures having 5 edges and the complete case for 6 objects and 0.15 perturbation.

7. Conclusion

In this paper, we investigated some of the connections between the PCM-based decision making techniques and the stochastic models. One of the main contributions of the study is the similar definition to the concept used in the AHP and the deeper investigation of the consistency in the case of the Bradley–Terry model (see [Theorems 1 and 2](#), and [Definition 7](#)).

The information retrieval of the different graphs of comparisons was examined for the case of the Bradley–Terry (and the Thurstone) model(s) with the help of many different measurements. We found that the measures almost always provide the same ranking of the different structures of comparison. For spanning trees the star graph provided the most information (comparing all possible comparison structures to the complete one). It is also true that the best graphs for a given number of objects to compare and a given number of comparisons were always the same as the ones found by [Bozóki and Szádóczki \(2022\)](#), which suggests that the optimality of these graphs is more general, it is not restricted to the AHP model or the domain of PCMs.

Future research could explore whether these graphs are optimal in the case of other models – such as some incomplete fuzzy decision models, e.g., different kinds of incomplete linguistic preference relations –, and what can be the (analytical) reasons behind this. Are the reasons derived from the comparisons' structures? In the current study, we did not label the different objects. One could study the labeled structures of comparisons: what are the best graphs in those cases? In particular, which star graph is optimal, the one centered around the best, the worst or another object? In our view, it would be also important to continue the investigation of the connections between the AHP and the stochastic models, especially in connection with the properties of inconsistency and inconsistency indices.

CRedit authorship contribution statement

László Gyarmati: Software, Writing. **Éva Orbán-Mihálykó:** Conceptualization, Methodology, Writing. **Csaba Mihálykó:** Methodology, Validation, Writing. **Zsombor Szádóczki:** Conceptualization, Writing. **Sándor Bozóki:** Conceptualization, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The authors are grateful to the anonymous reviewers for their constructive comments.

The research was partly supported by the ÚNKP-22-3 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. L. Gyarmati thanks the support.

The research of S. Bozóki and Zs. Szádóczki has been supported in part by the TKP2021-NKTA-01 NRDIO grant.

References

- Abel, E., Galpin, I., Paton, N. W., & Keane, J. A. (2022). Pairwise comparisons or constrained optimization? A usability evaluation of techniques for eliciting decision priorities. *International Transactions in Operational Research*, 29, 3190–3206. <http://dx.doi.org/10.1111/itor.12907>.
- Ágoston, K. C., & Csató, L. (2022). Inconsistency thresholds for incomplete pairwise comparison matrices. *Omega*, 108, Article 102576. <http://dx.doi.org/10.1016/j.omega.2021.102576>.
- Bortot, S., Brunelli, M., Fedrizzi, M., & Marques Pereira, R. A. (2022). A novel perspective on the inconsistency indices of reciprocal relations and pairwise comparison matrices. *Fuzzy Sets and Systems*, <http://dx.doi.org/10.1016/j.fss.2022.04.020>.
- Bozóki, S., Csató, L., & Temesi, J. (2016). An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research*, 248(1), 211–218. <http://dx.doi.org/10.1016/j.ejor.2015.06.069>.
- Bozóki, S., Fülöp, J., & Rónyai, L. (2010). On optimal completion of incomplete pairwise comparison matrices. *Mathematical and Computer Modelling*, 52(1–2), 318–333. <http://dx.doi.org/10.1016/j.mcm.2010.02.047>.
- Bozóki, S., & Szádóczki, Z. (2022). Optimal sequences for pairwise comparisons: the graph of graphs approach. <http://dx.doi.org/10.48550/arXiv.2205.08673>, arXiv preprint [arXiv:2205.08673](http://dx.doi.org/10.48550/arXiv.2205.08673).

- Bozóki, S., & Tsyganok, V. (2019). The (logarithmic) least squares optimality of the arithmetic (geometric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. *International Journal of General Systems*, 48(4), 362–381.
- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324–345. <http://dx.doi.org/10.2307/2334029>.
- Brunelli, M. (2014). *Introduction to the analytic hierarchy process*. Springer.
- Brunelli, M. (2018). A survey of inconsistency indices for pairwise comparisons. *International Journal of General Systems*, 47(8), 751–771. <http://dx.doi.org/10.1080/03081079.2018.1523156>.
- Chen, P., Gao, C., & Zhang, A. Y. (2022). Optimal full ranking from pairwise comparisons. *The Annals of Statistics*, 50(3), 1775–1805. <http://dx.doi.org/10.1214/22-AOS2175>.
- Choo, E., & Wedley, W. (2004). A common framework for deriving preference values from pairwise comparison matrices. *Computers & Operations Research*, 31(6), 893–908. [http://dx.doi.org/10.1016/S0305-0548\(03\)00042-X](http://dx.doi.org/10.1016/S0305-0548(03)00042-X).
- Crawford, G., & Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, 29(4), 387–405. [http://dx.doi.org/10.1016/0022-2496\(85\)90002-1](http://dx.doi.org/10.1016/0022-2496(85)90002-1).
- Csató, L. (2021). *Tournament design: how operations research can improve sports rules*. palgrave pivots in sports economics, <http://dx.doi.org/10.1007/978-3-030-59844-0>.
- Duleba, S., & Moslem, S. (2019). Examining Pareto optimality in Analytic Hierarchy Process on real Data: An application in public transport service development. *Expert Systems with Applications*, 116, 21–30. <http://dx.doi.org/10.1016/j.eswa.2018.08.049>.
- Faramondi, L., Oliva, G., & Bozóki, S. (2020). Incomplete Analytic Hierarchy Process with minimum weighted ordinal violations. *International Journal of General Systems*, 49(6), 574–601. <http://dx.doi.org/10.1080/03081079.2020.1786380>.
- Ford Jr, L. R. (1957). Solution of a ranking problem from binary comparisons. *American Mathematical Monthly*, 64(8P2), 28–33.
- Fürnkranz, J., & Hüllermeier, E. (2011). Preference learning and ranking by pairwise comparison. In *Preference learning* (pp. 65–82). Berlin, Heidelberg: Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-642-14125-6_4.
- Gass, S. (1998). Tournaments, transitivity and pairwise comparison matrices. *Journal of the Operational Research Society*, 49(6), 616–624. <http://dx.doi.org/10.1057/palgrave.jors.2600572>.
- Genest, C., & M'lan, C.-É. (1999). Deriving priorities from the Bradley-Terry model. *Mathematical and Computer Modelling*, 29(4), 87–102.
- Harker, P. (1987). Review of the main developments in the Analytic Hierarchy Process. *Mathematical Modelling*, 9(11), 837–848. [http://dx.doi.org/10.1016/0270-0255\(87\)90503-3](http://dx.doi.org/10.1016/0270-0255(87)90503-3).
- Hofmann, K., Whiteson, S., & de Rijke, M. (2013). Balancing exploration and exploitation in listwise and pairwise online learning to rank for information retrieval. *Information Retrieval*, 16, 63–90. <http://dx.doi.org/10.1007/s10791-012-9197-9>.
- Ishizaka, A., & Labib, A. (2011). Review of the main developments in the Analytic Hierarchy Process. *Expert Systems with Applications*, 38(11), 14336–14345. <http://dx.doi.org/10.1016/j.eswa.2011.04.143>.
- Ishizaka, A., & Nemery, P. (2013). *Multi-criteria decision analysis: methods and software*. Wiley.
- Jeon, J.-J., & Kim, Y. (2013). Revisiting the Bradley-Terry model and its application to information retrieval. *Journal of the Korean Data and Information Science Society*, 24(5), 1089–1099. <http://dx.doi.org/10.7465/jkdi.2013.24.5.1089>.
- Kulakowski, K., & Talaga, D. (2020). Inconsistency indices for incomplete pairwise comparisons matrices. *International Journal of General Systems*, 49(2), 174–200. <http://dx.doi.org/10.1080/03081079.2020.1713116>.
- Liu, P., Li, Y., & Wang, P. (2022). Consistency threshold- and score function-based multi-attribute decision-making with q-rung orthopair fuzzy preference relations. *Information Sciences*, 618, 356–378. <http://dx.doi.org/10.1016/j.ins.2022.10.122>.
- Liu, P., Wang, P., & Pedrycz, W. (2021). Consistency- and consensus-based group decision-making method with incomplete probabilistic linguistic preference relations. *IEEE Transactions on Fuzzy Systems*, 29(9), 2565–2579. <http://dx.doi.org/10.1109/TFUZZ.2020.3003501>.
- MacKay, D. B., Bowen, W. M., & Zinnes, J. L. (1996). A Thurstonian view of the Analytic Hierarchy Process. *European Journal of Operational Research*, 89(2), 427–444.
- Mantik, S., Li, M., & Porteous, J. (2022). A preference elicitation framework for automated planning. *Expert Systems with Applications*, 208, Article 118014. <http://dx.doi.org/10.1016/j.eswa.2022.118014>.
- Mazurek, J., & Kulakowski, K. (2022). On the derivation of weights from incomplete pairwise comparisons matrices via spanning trees with crisp and fuzzy confidence levels. *International Journal of Approximate Reasoning*, 150, 242–257. <http://dx.doi.org/10.1016/j.ijar.2022.08.014>.
- Mondal, A., Roy, S. K., & Pamucar, D. (2023). Regret-based three-way decision making with possibility dominance and SPA theory in incomplete information system. *Expert Systems with Applications*, 211, Article 118688. <http://dx.doi.org/10.1016/j.eswa.2022.118688>.
- Orbán-Mihálykó, É., Koltay, L., Szabó, F., Csuti, P., Kéri, R., & Schanda, J. (2015). A new statistical method for ranking of light sources based on subjective points of view. *Acta Polytechnica Hungarica*, 12(8), 195–214.
- Orbán-Mihálykó, É., Mihálykó, C., & Gyarmati, L. (2022). Application of the generalized thurstone method for evaluations of sports tournaments' results. *Knowledge*, 2(1), 157–166.
- Orbán-Mihálykó, É., Mihálykó, C., & Koltay, L. (2019). A generalization of the Thurstone method for multiple choice and incomplete paired comparisons. *Central European Journal of Operations Research*, 27(1), 133–159. <http://dx.doi.org/10.1007/s10100-017-0495-6>.
- Orbán-Mihálykó, É., Mihálykó, C., & Koltay, L. (2019). Incomplete paired comparisons in case of multiple choice and general log-concave probability density functions. *Central European Journal of Operations Research*, 27(2), 515–532.
- Rao, P., & Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. *Journal of the American Statistical Association*, 62(317), 194–204.
- Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3), 234–281. [http://dx.doi.org/10.1016/0022-2496\(77\)90033-5](http://dx.doi.org/10.1016/0022-2496(77)90033-5).
- Shah, N. B., & Wainwright, M. J. (2018). Simple, robust and optimal ranking from pairwise comparisons. *Journal of Machine Learning Research*, 18(199), 1–38. <http://jmlr.org/papers/v18/16-206.html>.
- Shiraishi, S., & Obata, T. (2002). On a maximization problem arising from a positive reciprocal matrix in AHP. *Bulletin of Informatics and Cybernetics*, 34(2), 91–96. <http://dx.doi.org/10.5109/135111>.
- Shiraishi, S., Obata, T., & M., D. (1998). Properties of a positive reciprocal matrix and their application to AHP. *Journal of the Operations Research Society of Japan*, 41(3), 404–414. <http://dx.doi.org/10.15807/jorsj.41.404>.
- Stewart, T. (1992). A critical survey on the status of multiple criteria decision making theory and practice. *Omega*, 20(5), 569–586. [http://dx.doi.org/10.1016/0305-0483\(92\)90003-P](http://dx.doi.org/10.1016/0305-0483(92)90003-P).
- Szádóczki, Z., Bozóki, S., Juhász, P., Kadenko, S. V., & Tsyganok, V. (2022). Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. *Annals of Operations Research*. <http://dx.doi.org/10.1007/s10479-022-04819-9>.
- Szádóczki, Z., Bozóki, S., & Tekile, H. A. (2022). Filling in pattern designs for incomplete pairwise comparison matrices: (quasi-) regular graphs with minimal diameter. *Omega*, 107, Article 102557. <http://dx.doi.org/10.1016/j.omega.2021.102557>.
- Tekile, H. A., Brunelli, M., & Fedrizzi, M. (2023). A numerical comparative study of completion methods for pairwise comparison matrices. *Operations Research Perspectives*, 10, Article 100272. <http://dx.doi.org/10.1016/j.orp.2023.100272>.
- Temesi, J., Szádóczki, Z., & Bozóki, S. (2023). Incomplete pairwise comparison matrices: Ranking top women tennis players. *Journal of the Operational Research Society*, 1–13. <http://dx.doi.org/10.1080/01605682.2023.2180447>.
- Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286. <http://dx.doi.org/10.1037/h0070288>.
- Tzeng, G.-H., & Huang, J.-J. (2011). *Multiple attribute decision making - methods and applications*. CRC, <http://dx.doi.org/10.1201/b11032>.
- Ureña, R., Chiclana, F., Morente-Moliner, J., & Herrera-Viedma, E. (2015). Managing incomplete preference relations in decision making: A review and future trends. *Information Sciences*, 302, 14–32. <http://dx.doi.org/10.1016/j.ins.2014.12.061>.
- Wang, Y.-M., Jia, X., Song, H.-H., & Martínez, L. (2023). Improving consistency based on regret theory: A multi-attribute group decision making method with linguistic distribution assessments. *Expert Systems with Applications*, 221, Article 119748. <http://dx.doi.org/10.1016/j.eswa.2023.119748>.
- Wang, P., Liu, P., & Chiclana, F. (2021). Multi-stage consistency optimization algorithm for decision making with incomplete probabilistic linguistic preference relation. *Information Sciences*, 556, 361–388. <http://dx.doi.org/10.1016/j.ins.2020.10.004>.
- Wu, J. J., Mazzuchi, T. A., & Sarkani, S. (2023). Comparison of multi-criteria decision-making methods for online controlled experiments in a launch decision-making framework. *Information and Software Technology*, 155, Article 107115. <http://dx.doi.org/10.1016/j.infsof.2022.107115>.
- Zhou, B., & Yao, Y. (2010). Evaluating information retrieval system performance based on user preference. *Journal of Intelligent Information Systems*, 34, 227–248. <http://dx.doi.org/10.1007/s10844-009-0096-5>.