Szádoczki Zsombor

Preference modelling with a graph theoretic approach

Department of Operations Research and Actuarial Sciences

Supervisor:

Bozóki Sándor Ph.D.

© Szádoczki Zsombor

Corvinus University of Budapest

Doctoral School of Economics, Business and Informatics

Preference modelling with a graph theoretic approach

Ph.D. Thesis

Szádoczki Zsombor

Budapest, 2024

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my gratitude to my supervisor, *Sándor Bozóki*, who provided me his time, as well as his extensive professional, academic, and personal support to finish my Ph.D. thesis and also to increase the quality of my publications. I am grateful that he involved me in many exciting research projects as well.

I would like to acknowledge the Research Laboratory on Engineering & Management Intelligence, Institute for Computer Science and Control (SZTAKI), Hungarian Research Network (HUN-REN), as well as the Department of Operations Research and Actuarial Sciences, Institute of Operations and Decision Sciences, Corvinus University of Budapest to ensure the conditions of my research.

I would like to thank my co-authors, *Sándor Bozóki*, *Patrik Juhász*, *Sergii V. Kadenko*, *Hailemariam Abebe Tekile*, *József Temesi*, and *Vitaliy Tsyganok*, who greatly contributed to the studies included in this thesis.

I would also like to acknowledge my co-authors in other papers, *Szabolcs Duleba*, *László Gyarmati*, *Csaba Mihálykó*, *Éva Orbán-Mihálykó*, *Gabriela Cecilia Stănciulescu*, *Zsuzsanna Katalin Szabó*, and *Dalma Szabó*, who extensively helped me in my research.

I am grateful to the members of the industrial projects, in which I participated at SZTAKI, to *Júlia Bergmann*, *Péter Dobrovoczki*, *Dávid Gyulai*, *Zsófia Illényi*, *Dávid Karnok*, *Ádám Szaller*, and especially to *András Kovács*, from whom I have learned a lot in this regard.

I got useful feedback from *László Csató*, *András London*, and *Éva Orbán-Mihálykó* on different MKE (Hungarian Society of Economics) conferences as well as from *Matteo Brunelli*, *Michele Fedrizzi*, *János Fülöp*, *Konrad Kułakowski*, and *Gabriele Oliva* on the online seminar organized by the University of Trento. I am grateful to *Tamás Solymosi*, who provided me the opportunity to present my research at the Corvinus Game Theory Seminar.

The many exciting ideas and constructive feedback of *László Csató* in connection with my work should also be highlighted.

I would like to acknowledge my friends, who always supported me, and even helped with some computational capacity, when it was needed.

I would like to thank my wife, *Vera*, for the discussions, ideas and for her consistent help. I am grateful to my parents and my sister for their support and encouragement through these years, and my son, *Márton*, who waited to start crawling just enough for me to finish my thesis.

1 Dissertation summary

1.1 Research background and basic concepts

1.1.1 Introduction: Decision theory

Decision making, choosing from a number of alternatives according to different criteria, is a common part of human life. We make the majority of these choices without modelling or calculations, based on different habits or heuristics. However, the most important, high-profile decisions are usually handled with a wide range of modelling tools, especially in the case of companies or governments.

As decisions are present everywhere in our life, decision making theory is a large and interdisciplinary field, which contains approaches from the classical utility based ones (Fishburn, 1970) to the ones focusing on the psychological features (Kahneman, 2011). Naturally, here we do not aim to summarize the diverse literature of this field.

Our research focuses on multicriteria decision making (MCDM) problems (Ishizaka and Nemery, 2013), especially methods connected to pairwise comparisons. The aim of MCDM is to select the best, the best few, or to provide a whole ranking of a finite set of alternatives based on a finite number of (usually conflicting) quantitative and/or qualitative criteria.

One of the most popular MCDM methodologies is the Analytic Hierarchy Process (AHP), proposed by Saaty (1977, 1980). It is based on a hierarchical system of criteria, subcriteria, etc., while uses pairwise comparison matrices (PCMs) to evaluate the alternatives according to each criterion separately and to determine the importance weights of the criteria as well. An element of a (multiplicative, ratio scale) PCM shows how many times the alternative (criterion) corresponding to the given row of the PCM is better/stronger/larger/more important than the alternative (criterion) corresponding to the given column of the matrix.

Besides decision modelling, pairwise comparisons are used in many other areas as well, e.g., preference measurement, ranking, sports, and psychometrics (Thurstone, 1927; Davidson and Farquhar, 1976; Csató, 2021). These types of comparisons are placed into a matrix in the case of a PCM. The main idea behind this process is that the decision makers cannot provide their preferences accurately for a complex problem; however, they can estimate their real preferences well between a pair of alternatives according to a single criterion.

The focus of our research is the case when some of the comparisons are missing, thus, we have to deal with a not complete data set, an incomplete pairwise comparison matrix (IPCM). Although we apply the decision modelling point of view throughout the dissertation, as both pairwise comparisons and missing data are common in many different research fields, our results can be useful in a much wider range.

Next, Sections 1.1.2 and 1.1.3 detail the research background in a more focused way, as well as contain the necessary definitions connected to pairwise comparison matrices and their graph representations. Section 1.2 lists the publications included in the Ph.D. thesis, while Section 1.3 presents the frame of our research and the connections between the different studies. Section 1.4 details the exact contributions to the articles, Section 1.5 provides a short overview of the included studies, whereas Section 1.6 discusses several future research directions. Finally, Chapters 2, 3, 4, and 5 present the four original studies of the dissertation.

1.1.2 Research background: Incomplete pairwise comparison matrices

There can be many reasons behind the incompleteness of a pairwise comparison matrix. Some data may have been lost, certain comparisons can be simply impossible, or the decision maker might have no time or willingness to provide all the comparisons, which is a lingering task.

Harker (1987) was among the first to propose IPCMs in order to reduce the number of questions asked from the decision maker in the Analytic Hierarchy Process. It is especially useful in the case of group decision making, when the choice should be made based on the preferences of several decision makers, and all of them have to fill in all PCMs.

If we are dealing with incomplete data, the result—i.e., the ranking of the alternatives calculated from the IPCM—is heavily dependent on the number of known elements and their arrangement. The latter one, the structure of the comparisons can be suitably handled by the representing graph of the IPCM (Gass, 1998). In the representing graph, the vertices correspond to the alternatives (criteria), while there is an edge between two vertices if and only if the comparison between the appropriate two alternatives (criteria) is known.

Although the literature of IPCMs is relatively limited compared to other areas connected to pairwise comparisons, there are many recent studies on theoretical results (Zhou et al., 2018; Kułakowski and Talaga, 2020; Szybowski et al., 2020; Ágoston and Csató, 2022), as well as applications (Bozóki et al., 2016).

A large portion of our research is centered around recommended filling in patterns for incomplete pairwise comparison matrices. What kind of designs of comparisons ensure that the computed results are close to the ones that would be calculated from the complete PCM? This and similar questions can be answered using graph theoretical properties of the representing graphs. The results are not just important from a theoretical point of view, but they can be easily applied in the practice of multicriteria decision making problems as well.

1.1.3 Pairwise comparisons: Basic concepts

In this section, the most important concepts connected to (incomplete) pairwise comparison matrices and their graph representations are defined formally, in order to make it easier to follow the later parts of the dissertation. Most of the definitions listed here are also included in the four original studies of the thesis.

From now on, let us denote the number of criteria (alternatives) in a multicriteria decision making problem by n.

Definition 1 (Pairwise comparison matrix (PCM)) *The $n \times n$ matrix $A = [a_{ij}]$ is called a pairwise comparison matrix if it is positive ($a_{ij} > 0$ for all i and j) and reciprocal ($1/a_{ij} = a_{ji}$ for all i and j).*

When a decision maker fills in a PCM, there are usually some kind of inconsistency among the elements of the matrix. It can occur that alternative A is 2 times better than alternative B, and alternative B is 3 times better than alternative C, but alternative A is not $(2 \times 3 =)6$ times better than alternative C.

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

There are several ways to measure the level of inconsistency (Brunelli, 2018; Kułakowski and Talaga, 2020), however, in practice the most often applied metric is still Saaty's Consistency Ratio (CR) (Saaty, 1977).

Definition 3 (Consistency Ratio (CR)) *The CR of an $n \times n$ PCM A is defined as follows:*

$$CR = \frac{CI}{RI}, \tag{1}$$

where CI stands for Consistency Index, that is:

$$CI = \frac{\lambda_{\max} - n}{n - 1}, \tag{2}$$

where $\lambda_{\max}$ is the principal eigenvalue of matrix A, and RI is the Random Index, which is the average CI obtained from a sufficiently large set of randomly generated PCMs of size n.

To determine the ranking of the alternatives, we need a weight (priority) vector calculation technique. The elements of the computed vector show the performance of the given alternatives. In the case of consistent PCMs, all of these techniques provide the same vector, however, for inconsistent matrices the result can be different.

Probably the two most commonly used techniques to calculate a weight vector are the logarithmic least squares (LLSM) (Crawford and Williams, 1985) and the eigenvector (EV) (Saaty, 1977) methods.

Definition 4 (Logarithmic Least Squares Method (LLSM)) *Let A be an $n \times n$ PCM. The positive weight vector w of A determined by the LLSM is the optimal solution of the following problem:*

$$\min_{w} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\ln(a_{ij}) - \ln\left(\frac{w_i}{w_j}\right)\right)^2, \tag{3}$$

where w_i is the ith coordinate of w.

Definition 5 (Eigenvector (EV) Method) *Let A be an $n \times n$ PCM. The positive weight vector w of A determined by the EV method is defined as follows:*

$$A \cdot w = \lambda_{\max} \cdot w, \tag{4}$$

where λ_{max} is the principal eigenvalue of matrix A.

The solutions of (3) and (4) are only unique up to scalar multiplication, thus, the sum of the components (the weights) are usually normalized to one.

Most of our research focuses on the representing graphs of incomplete pairwise comparison matrices.

Definition 6 (Incomplete pairwise comparison matrix (IPCM)) *An $n \times n$ matrix $A = [a_{ij}]$ is an incomplete pairwise comparison matrix (IPCM) if:*

- $a_{ij} \in \mathbb{R}_+ \cup \{*\}\ \forall\ 1 \le i, j \le n$ *and*
    - $a_{ji} = 1/a_{ij}$ *if* $a_{ij} \in \mathbb{R}_+$,
    - $a_{ji} = *$ *if* $a_{ij} = *$,

where $*$ denotes the missing elements, and $\mathbb{R}_+$ is the set of positive real numbers.

Definition 7 (Representing graph) *An incomplete pairwise comparison matrix A can be represented by an undirected graph $G = (V, E)$, where:*

- *the vertices $V = \{1, 2, \dots, n\}$ correspond to the alternatives,*
- *while the edge set E represents the known elements of A outside the main diagonal:*

$$e_{ij} \in E \iff a_{ij} \neq * \text{ and } i \neq j.$$

Definition 8 (Path) *A path is a finite (or infinite) sequence of edges which joins a sequence of vertices such that all vertices (and therefore all edges) are distinct.*

Both the logarithmic least squares and eigenvector weight calculation methods can be used for incomplete pairwise comparison matrices as well. In the case of LLSM, optimization problem (3) is only applied to the known elements of the matrix, while for the EV method, the priority vector is calculated from the complete matrix where all missing comparisons are replaced by variables such that the Consistency Ratio (CR) is minimized (this technique is sometimes denoted as the CREV method). Both approaches provide a unique weight vector if and only if the representing graph of the IPCM is connected (Bozóki et al., 2010).

Definition 9 (Connected graph) *In an undirected graph, two vertices u and v are called connected if the graph contains a path from u to v. A graph is said to be connected if every pair of vertices in the graph is connected.*

The smallest connected graphs are the so-called spanning trees, which contain $n-1$ edges for n vertices, and play a special role in our research.

Definition 10 (Spanning tree) *Let $G = (V, E)$ be a connected graph. $G' = (V, E')$ is a spanning tree of G if $E' \subseteq E$ is a minimal set of edges that connects all vertices of G.*

Probably the two most important graph properties that we examine in our studies are related to the regularity and the diameter of the representing graph.

Definition 11 (k-regularity) *A graph is called k-regular if every vertex has k neighbours, which means that the degree of every vertex is k.*

Definition 12 (The diameter of a graph) *The diameter (denoted by d) of a graph G is the length of the longest shortest path between any two vertices:*

$$d = \max_{u,v \in V(G)} \ell(u, v),$$

where $V(G)$ denotes the set of vertices of G and $\ell(.,.)$ is the graph distance between two vertices, namely the length of the shortest path between them.

We will use the concepts defined in this section throughout the dissertation.

1.2 List of publications included in the Ph.D. thesis

All of the research we present here have been developed as academic journal articles. The exact contributions of the given studies are detailed in Section 1.4, while a summary of each publication can be found in Section 1.5. We compiled the following studies in the dissertation without re-editing, and present them in the form as they were—or are planned to be—published.

I. Szádoczki, Zs., Bozóki, S., and Tekile, H. A. (2022). Filling in pattern designs for incomplete pairwise comparison matrices: (Quasi-)regular graphs with minimal diameter. *Omega*, 107:102557.

https://doi.org/10.1016/j.omega.2021.102557.

II. Szádoczki, Zs., Bozóki, S., Juhász, P., Kadenko, S. V., and Tsyganok, V. (2023). Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. *Annals of Operations Research*, 326(2):783-807.

https://doi.org/10.1007/s10479-022-04819-9.

III. Szádoczki, Zs., and Bozóki, S. (2023). Optimal sequences for pairwise comparisons: the graph of graphs approach. *Working paper.*

https://doi.org/10.48550/arXiv.2205.08673.

IV. Temesi, J., **Szádoczki, Zs.** and Bozóki, S. (2024). Incomplete pairwise comparison matrices: Ranking top women tennis players. *Journal of the Operational Research Society*, 75(1):145-157. https://doi.org/10.1080/01605682.2023.2180447.

1.3 Research frame

Our dissertation belongs to the same academic research group as Bozóki (2006), Csató (2015), Ábele-Nagy (2019), and Poesz (2022). As mentioned before, the main questions and studied problems are focused around the topic of the graph theoretic properties of representing graphs of incomplete pairwise comparison matrices. The relation of research questions, publications, and results of the studies included in the Ph.D. thesis can be seen in Figure 1.

Studies I., II., and III. are natural (and in some sense linear) continuations of each other. From a methodological point of view, all of them rely on different simulations. Study IV. uses the same tools from the literature of multicriteria decision making and graph theory as the previous publications, however, it focuses on the ranking aspect of pairwise comparisons instead of the question of the optimal filling in designs.

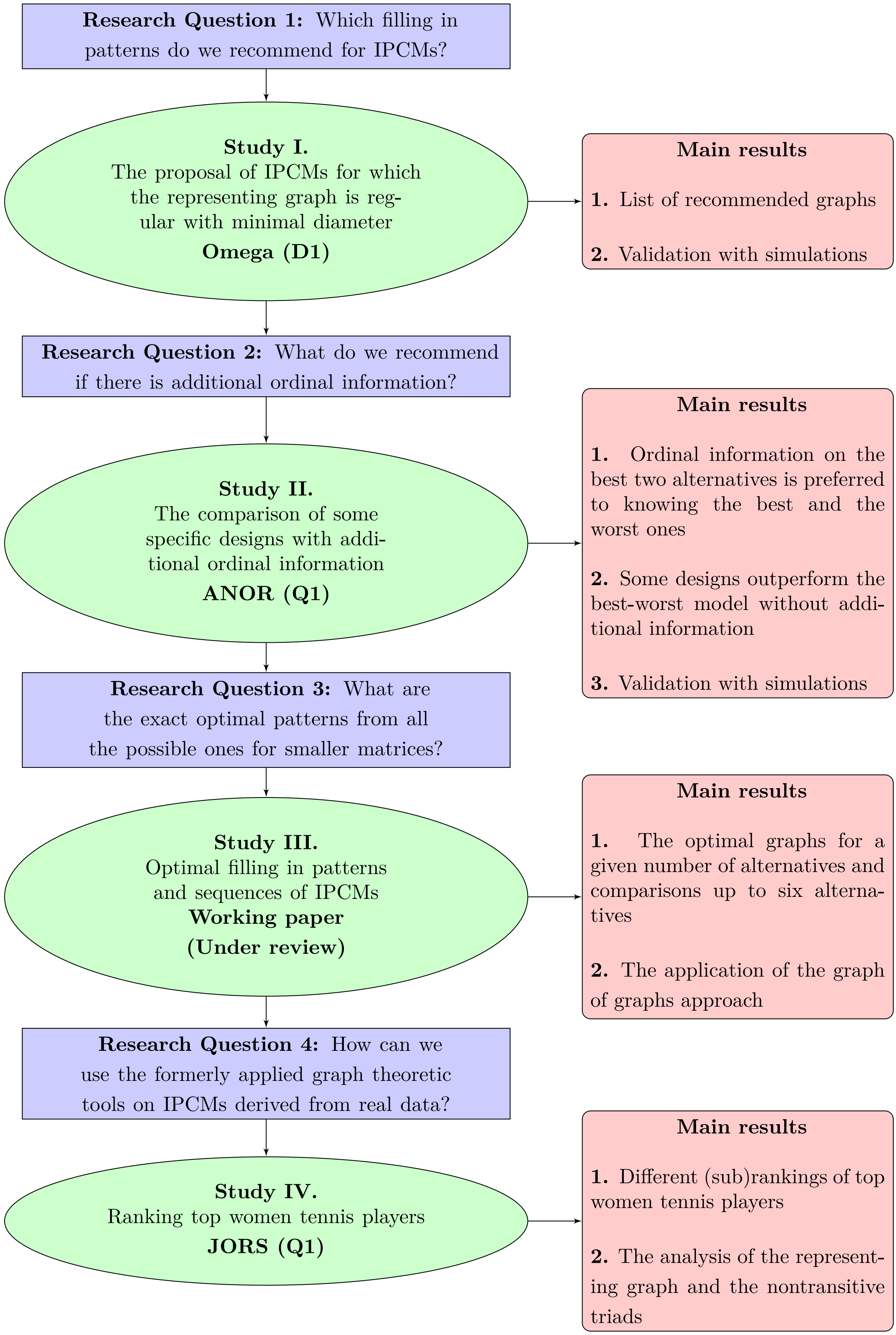


Figure 1: The flow chart of research questions, publications, and main results

All research presented in this Ph.D. thesis started with a conjecture that the diameter (the longest shortest path) of the representing graph of incomplete pairwise comparison matrices can be important to get a reliable and good estimation of the decision maker's preferences. This conjecture came from the master thesis of one of the co-authors of Study I. (Tekile, 2017). It showed an example where the graph generated from the table tennis players' matches included a long shortest path between two vertices (players), and the calculated result appeared to be misleading because of that.

We carried out an extensive literature review on the filling designs of incomplete pairwise comparison matrices, and found that some sense of regularity of the representing graph was detected as an important property (Wang and Takahashi, 1998; Kułakowski et al., 2019), but the diameter was almost entirely missing from the relevant papers.

In Study I., we proposed regular graphs with minimal diameter as a new design of filling in patterns for incomplete pairwise comparison matrices, created a list of proposed graphs and validated our recommendations by simulations. These results led to numerous new research questions.

First, in Study II., we dealt with the most common cases with a few number of alternatives in more detail, as well as focused on the inclusion of additional a priori ordinal information that is often used in multicriteria decision making methods, such as the best-worst method (Rezaei, 2015). It turned out that a Ukrainian research group found the importance of the diameter of the representing graph more or less at the same time, independently from us (Kadenko and Tsyganok, 2020). Thus, we continued the research together in Study II. We were able to compare some popular designs with our proposals, and the usefulness of additional ordinal information was also evaluated.

In the case of Study III., we continued to examine small matrices (with at most six alternatives), which are the most common in multicriteria decision making problems. One of the most important limitations of Study II. was that in some instances the examined designs used a different number of comparisons, thus, the effect of the filling structure and the effect of the number of known comparisons were inseparable. Based on that, in Study III., all possible filling in patterns for incomplete pairwise comparison matrices have been compared with a given number of comparisons, thus, it was possible to select the best one among them. According to Gyarmati et al. (2023), our results seem to be more general and not specific for the domain of pairwise comparison matrices.

As mentioned before, Study IV. uses the same tools as the other included papers, however, it focuses on the ranking aspect of incomplete pairwise comparisons as top women tennis players are ranked with this method. We revisited a former research of Bozóki et al. (2016) with a similar real-world database, however, we extended the results with a deeper analysis of the graph representation of the matches between the players, as well as the detailed investigation of nontransitive triads.

All of the above-mentioned studies opened up several further research questions that will be detailed in Section 1.6.

1.4 Results and contributions

In this section, we list the main results of the included studies, as well as highlight which of them are individual and joint (inseparable) findings.

Besides extending some previous results of other studies, we also would like to emphasize that the papers of this Ph.D. thesis opened up totally new avenues of research directions as well.

Results and contributions of Study I.

Study I. deals with filling in patterns for incomplete pairwise comparison matrices. We draw attention to the diameter of the representing graph as an important property to select a filling pattern of comparisons that estimates the preferences of the decision maker well. We also confirm that regularity of the representing graph is an important property. The proposed graphs are determined and provided in several different formats, while the usefulness of the recommended graphs is demonstrated via simulations.

Individual contributions:

- Implementation of the validating simulations in Scilab;

- Implementation of graph searching methods in Wolfram Mathematica and Python;
- Creating the systematic list of recommended graphs in several formats (adjacency matrix, edge list, graph, and ‘Graph6’ format);
- Creating the (more than 100 pages long) online Appendix;
- Implementing LaTeX codes to create illustrative figures.

Joint (inseparable) **results with our co-authors** (*Sándor Bozóki* and *Hailemariam Abebe Tekile*)**:**

- Extensive literature review connected to the filling designs for incomplete pairwise comparison matrices and the degree/diameter graph theoretical problem;
- Searching for the representing graphs with the proposed properties;
- Finding a motivational example to show the importance of the diameter;
- Running the simulations;
- Editing and writing the article.

Results and contributions of Study II.

Study II. compares some well-known and often applied filling designs for incomplete pairwise comparison matrices to benchmark designs and several models proposed by us up until 10 alternatives. The additional a priori ordinal information of models has a crucial role in this research. With the help of the benchmark methods, we are able not just to compare different designs, but also to evaluate the usefulness of these additional information.

It turns out that the ordinal information about the best and second best alternatives is preferred compared to the often applied best and worst ones. However, some of the designs were able to outperform all the examined models with additional information, among which the design of 2 edge-disjoint spanning trees proved to be the best.

All of the results are equally joint (inseparable) **with our co-authors** (*Sándor Bozóki*, *Patrik Juhász*, *Sergii V. Kadenko*, and *Vitaliy Tsyganok*).

Results and contributions of Study III.

Study III. determines the filling patterns (representing graphs) for incomplete pairwise comparison matrices that provide the closest (LLSM) weight vectors to the ones calculated from the complete matrix for a given number of alternatives and a given number of comparisons up to six alternatives. The proposed filling patterns are in that sense optimal for a given number of comparisons. Moreover, in many cases the optimal graphs are 'reachable' from each other, i.e., adding exactly one comparison to an optimal case with e comparisons results in the optimal pattern with $e+1$ comparisons creating optimal filling in sequences for pairwise comparison matrices. These sequences can be especially useful when the number of questions answered by the decision maker is uncertain (e.g., in online questionnaires).

The results are presented in different formats in order to make them instantly applicable in practice. We are also the first ones to apply the graph of graphs concept to visualize our results in this field of research.

Individual contributions:

- Implementation of the methods that determine all the possible representing graphs in Wolfram Mathematica;
- Implementation of the simulations in Scilab, with which we can determine the optimal cases;
- Implementing LaTeX codes to create visualizations (creating the graphs of graphs).

Joint (inseparable) **results with our co-author** (*Sándor Bozóki*)**:**

- Literature review connected to the filling sequences for pairwise comparison matrices and the graph of graphs concept;
- Running the simulations;
- Editing and writing the article.

Results and contributions of Study IV.

Study IV. demonstrates that incomplete pairwise comparison matrices can be used to create a ranking based on a large real database. We create all-time rankings of top women tennis players. Our methodology is similar to Bozóki et al. (2016), who created rankings of men tennis players. However, we dive deeper into the properties of the representing graph of the data, and study the ordinally nontransitive triads and the submatrices of the pairwise comparisons as well. We find that the results are robust to the application of different methods (LLSM, EV, Bradley–Terry) and modifications, which might be caused by the structure of the representing graph. We apply tests to the number of nontransitive triads, and also demonstrate that the rankings determined based on different submatrices are well-interpretable, and some of them can determine the ranking of some players in the overall order.

Individual contributions:

- Calculating the rankings of top women tennis players based on different methods and submatrices;
- Implementing R, Scilab, and Wolfram Mathematica codes to apply different methods and tests to the data;
- Implementing LaTeX codes to create illustrative figures.

Joint (inseparable) **results with our co-authors** (*József Temesi* and *Sándor Bozóki*)**:**

- Extensive literature review connected to ranking in sports, nontransitive triads, and ranking with incomplete pairwise comparisons;
- Editing and writing the article.

1.5 Overview of the studies

In this section, we summarize the compiled studies on the topic of preference modelling with the tools of multicriteria decision making from a graph theoretic point of view. It is important to emphasize the previous point that, besides the interesting theoretical findings and new research questions, all of the studies provide useful results for practitioners, decision analysts as well.

1.5.1 Study I. Filling in pattern designs for incomplete pairwise comparison matrices: (Quasi-)regular graphs with minimal diameter

Study I. (Szádoczki et al., 2022) served as the basis and inspiration for most of the later research included in the dissertation. Our goal was to propose a new approach of filling in patterns for incomplete pairwise comparison matrices that shows which comparisons should be made. We heavily relied on the graph representations of these matrices (Gass, 1998), and their graph theoretical properties.

We had the following crucial assumptions applied in this article:

- it can be chosen which comparisons should be made (they are not given);
- there is no prior information available on the items that should be compared, we can handle them in a symmetric way and focus on unlabelled graphs (the isomorphic graphs are considered to be equivalent);
- the whole set of comparisons should be determined before the start of the decision making process (the questions are not adaptive, they do not depend on the previous answers).

As mentioned before, we had a conjecture based on a previous result (Tekile, 2017) that the diameter of the representing graph can be an important property to find a filling in pattern that estimates the results of the complete matrix well.

The diameter measures the longest shortest path in a graph. With the minimization of this metric, we can ensure that there is not a pair of alternatives that can only be compared through a long indirect path. In that case, the small errors along this path could sum up causing a significant error in the results.

After an extensive literature review, we found that although some sense of regularity of the representing graph was proposed in several designs (McCormick and Bachus, 1952; McCormick and Roberts, 1952; Wang and Takahashi, 1998), the diameter was not studied broadly.

The regularity of the representing graph is also important as it provides some kind of symmetry to the comparisons, all the alternatives are compared to the same number of items, and there is no pivotal element in the system of comparisons.

It is also true that the regularity definitions used by former designs were rather restrictive. We used a more general concept, and provided a more systematic list of graphs in the online Appendix of the study, which can serve as a 'recipe' for practitioners to determine which comparisons should be asked from the decision maker in a given problem. Editing the online Appendix and creating the systematic list of recommended graphs in several formats—that are key parts of our results—were among our individual contributions to this study.

To find the interesting graphs for our research, the three parameters were the number of vertices (alternatives) n, the level of regularity k, and the diameter of the given pattern d.

It is important to mention that k-regular graphs with n vertices do not exist if both k and n are odd, thus, we defined k-quasi-regular graphs, where one vertex has degree $k+1$, while all the other vertices have k neighbours.

We aimed to keep the number of comparisons (and so parameter k) as low as possible, and minimize the diameter at the same time. From a mathematical point of view, we were looking for the minimal d for a given (n, k) pair.

It turned out that this problem has a strong relation to the degree/diameter problem (see, for instance, Miller and Širáň (2013)), which looks for the largest n for a given (k, d) pair.

Based on the known results of this field and the characteristics of multicriteria decision making problems, it turned out that the parameter values of $k = 3, 4, 5$ and $d = 2, 3$ can be interesting for us ($d = 1$ means a complete PCM), while the number of alternatives were examined between 5 and 24. It is true that finding graphs and running simulations would become difficult above 24 vertices. However, in practical problems the number of alternatives is usually much below 24, and the largest 5-regular graph with diameter 2 has 24 vertices, thus, it is a natural theoretical bound as well. At the same time the problem is not too relevant below five alternatives.

Although some graphs were known from the degree/diameter literature, as mentioned before, most of them had to be determined by us. We used several different methods, for smaller cases the enumeration of all the graphs and the selection of the ones with minimal diameter, while for larger cases different graph extending techniques, graph products, integer linear programming, and so on. This was a long, time-consuming process, the same idea rarely worked twice, and we determined 34 types of graph instances for different parameter combinations. As for our individual contributions, the majority of the graph instances were found by different graph searching methods implemented in Wolfram Mathematica and Python by us.

After the finding of the graphs, we had to validate that these designs provide better results, namely, closer weight vectors to the ones calculated from the complete matrix than other filling patterns. In order to test this, we used simulations, where the CREV and the incomplete LLSM weight calculation techniques were used, while the closeness measures were the Euclidean and Chebyshev (maximum absolute) distances. We determined the means and standard deviations of the distances between the weight vectors calculated from a given design and the complete matrix for all combinations of the parameters. The implementation of these simulations in Scilab was another key individual contribution of ours to this study.

We generated 1000 random, consistent PCMs for every parameter combination, and perturbed those using three different levels of perturbation (inconsistency) to get inconsistent matrices. The weight vectors were determined from these matrices using only the elements included in the examined patterns.

The following designs were compared in our simulations (all of the graphs were connected):

- (i) k-(quasi-)regular graphs with minimal diameter;
- (ii) Random connected graphs with the same number of edges as our recommendations;
- (iii) k-(quasi-)regular graphs, but not of minimal diameter;
- (iv) Randomly generated, minimal diameter, but not regular graphs;
- (v) Modified star graphs with the same number of edges (and minimal diameter).

The simulations confirmed that our recommendations outperformed all other designs in the case of almost all examined parameter combinations, for both measures, both weight calculation techniques, and every level of inconsistency. It was also validated that both properties—regularity and the minimal diameter—are necessary. The latter was also highlighted with a motivational example, where a regular graph had a long diameter, which resulted in large distances in the weight vectors from the complete case.

As mentioned before, the results of this study can be instantly applied in practical problems, while it also generated many further research questions, some of which later on have been addressed in Studies II. and III.

1.5.2 Study II. Incomplete pairwise comparison matrices based on graphs with average degree approximately 3

It occurs in many different multicriteria decision making methods, such as the SMART (simple multi-attribute rating technique) (Edwards, 1977), the Swing (von Winterfeldt and Edwards, 1986), and their generalizations, that some additional ordinal information—usually the best, the worst or both alternatives—are used. One could argue that if the best alternative is already known, then we do not need the given methods. However, the explanation lies in the multicriteria nature of the problems. For a simple example, we can imagine that it is easy to determine which is the cheapest car from a set of alternatives, but to determine which car should we buy based on a number of other criteria, is a much more difficult task.

Accordingly, we continued our research started in Study I., and resolved the assumption that no prior information is available. We were particularly interested in the additional information used by the best-worst method (Rezaei, 2015), which generated an extensive literature in the past few years (Mi et al., 2019). The study (Szádoczki et al., 2023) also focuses on the smaller matrices (with at most 10 alternatives), which are more common in MCDM problems.

We applied a similar simulation-based approach as in Study I. with three different levels of inconsistency, however, several aspects of the methods were improved. Instead of the matrix-wise solution, we used element-wise perturbations, and also improved the handling of different scales. The elements of the consistent PCMs were randomly generated from the interval $[1/9, 9]$, and we used uniformly distributed errors for the perturbation. However, the new elements were uniformly distributed not on the original scale, but around the original value on the scale presented in Figure 2. The reasoning behind this is as follows. If a decision maker is hesitant whether item A is 2 or 3 times as good as item B, then that is the same problem as if the decision maker would be hesitant whether B is $1/2$ or $1/3$ as good as A. Thus, we used a scale where the distance between $1/9$ and 1 is the same as between 1 and 9.

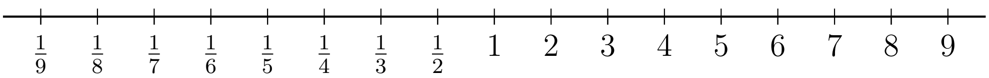


Figure 2: The scale on which the perturbed elements are uniformly distributed around the original value

The list of applied measures was also extended in the simulations. The following types of metrics were included:

- distances (Euclidean, Chebyshev, and Manhattan);
- rank correlations (Kendall's tau and Spearman's rho);
- and compatibility (similarity) indices (Garuti's compatibility index (Garuti, 2017), cosine similarity index (Kou et al., 2021), and Dice similarity index (Ye, 2012)).

We used a sample of 10 000 PCMs for the simulations, and aimed to keep the number of comparisons relatively low, thus, the following designs (representing graphs) were compared:

- (i) Best-worst graph;
- (ii) TOP2 graph;
- (iii) Best-random graph;
- (iv) Random-random graph;

- (v) 3-(quasi-)regular graph with minimal diameter;
- (vi) Union of two random edge-disjoint spanning trees;
- (vii) Random 2-edge-connected graph.

The first four designs are based on the same graph representation (the union of two star graphs) with different additional ordinal information. Thus, there are two pivotal elements, which are different in these models, and all of the other alternatives are only compared to them. We were interested in these kinds of structures because of the best-worst method, where every alternative is compared to the best and the worst items. This method also assumes that it is always possible to determine the best and the worst elements, however, this implies that it is always possible to determine the ranking of the alternatives according to one criterion (i.e., if we remove the former best and worst, we can determine the new best and worst again). Thus, it can make sense to test other pairs of highlighted alternatives.

In TOP2, the pivotal elements are the best and the second best alternatives, for the Best-random graph they are the best and a randomly chosen element, while for the Random-random case both of them are chosen randomly. This way we were able to evaluate these additional information as well, whether comparing to those elements provide weight vectors closer to the one calculated from the complete case. The TOP2 design was proposed by us, as usually if there are more comparisons for a given alternative, then its weight is estimated more accurately, and probably the most important items are the two best ones.

The 3-(quasi-)regular graph design was also proposed by us in Study I., but it does not use any additional ordinal information. Nor do the Union of two random edge-disjoint spanning trees and the Random 2-edge-connected graphs. The latter two are generalizations of the union of two star graphs, as a star graph is a certain type of spanning tree, and a 2-edge connected graph is a graph that remains connected after deleting an edge, which also holds for the union of two star graphs.

According to the results of the simulations, basically every measure that we used provided the same findings, thus, the outcomes do not heavily depend on the used metrics. It turned out that the TOP2 graph is preferred to the Best-worst one, which means that if we have an opportunity in this kind of problems to ask for additional ordinal information, on average it is better to ask the two best alternatives compared to the popular best and worst ones. However, some of the designs that do not use any additional information outperformed all the models using the union of two star graphs. For smaller cases, this is true for the 3-(quasi-)regular graphs, and for basically all examined cases, this is observed for the Union of two random edge-disjoint spanning trees.

Nonetheless, it is important to keep in mind one of the most important limitations of this research. That is, some of these designs use a different number of comparisons in certain cases, which does not make it possible to separate the effect of the filling pattern from the effect of the number of comparisons. We tried to resolve this problem in Study III.

1.5.3 Study III. Optimal sequences for pairwise comparisons: the graph of graphs approach

In Study III. (Szádoczki and Bozóki, 2022), we aimed to elaborate on the smallest cases that most commonly occur in MCDM problems, and also address the limitation of Study II. that the number of comparisons were different, thus, in some cases it was difficult to draw strong conclusions.

We were able to automize our simulations—which is one of our key individual contributions to this study—in order to examine larger samples and check all the possible patterns of comparisons (representing graphs) instead of some special designs as before, only analyzing the structures with the same number of comparisons (edges) this time, up to six alternatives (vertices). At the same time, we also went back to our original assumption that no prior information is available about the elements, as the labelling of graphs results in even more possibilities.

As measures, the Euclidean distance and the Kendall's tau metrics were used, with a sample size of 1 million matrices for a given number of alternatives (n) and a given number of comparisons (edges – e), which resulted in certain margins of error and significance levels for the different means.

It turned out that basically always the same representing graph provided the closest weight vector to the complete case for a given (n, e) pair according to the

Euclidean distance and Kendall's tau, and the results neither depended on the level of inconsistency nor on the weight calculation technique. Thus, in that sense we were able to determine the optimal filling in patterns for a given number of alternatives (up until six) and a given number of comparisons, which is a key contribution of this study to the literature.

Furthermore, we also found that many optimal cases are reachable from each other, namely, if we add an additional comparison to the optimal pattern with e comparisons, it results in the optimal case with $e + 1$ comparisons. This way, we can determine optimal filling in sequences for pairwise comparison matrices, which can be especially useful in the case of group decision making based on online questionnaires, where the number of comparisons provided by the decision makers is uncertain. If we follow an optimal filling in sequence, then we can ensure that the preferences of the decision makers are estimated on average as well as possible whenever they stop making further comparisons.

After that, we conducted a literature review in order to find out what kind of visualization can be used to present our results. Based on that, we applied the graph of graphs approach (see, for instance, Lovász (1977)), where there is a large graph and its vertices are also graphs, in our case the representing graphs of filling patterns. There is an edge between two small graphs if and only if they are reachable from each other, we can get one of them from the other one by adding (or deleting) exactly one comparison (edge). Applying and creating the illustrative figures (the graphs of graphs) in LaTeX is a crucial part of this study, that is also an individual contribution of ours.

As an example, the graph of graphs on five vertices can be seen in Figure 3. The number of comparisons (edges) are denoted on the left side of the chart, while the optimal graphs and partial optimal sequences are highlighted with blue color and boldness. It is important to note that the pattern of results is similar for different number of alternatives as well.

Among the spanning trees (in the first row), the star graph proved to be optimal for all cases examined, and it is expected that this structure retains its optimality for larger cases. Regular and quasi-regular graphs were also optimal, moreover, regularity was important in an even more general way, namely, the degrees of the

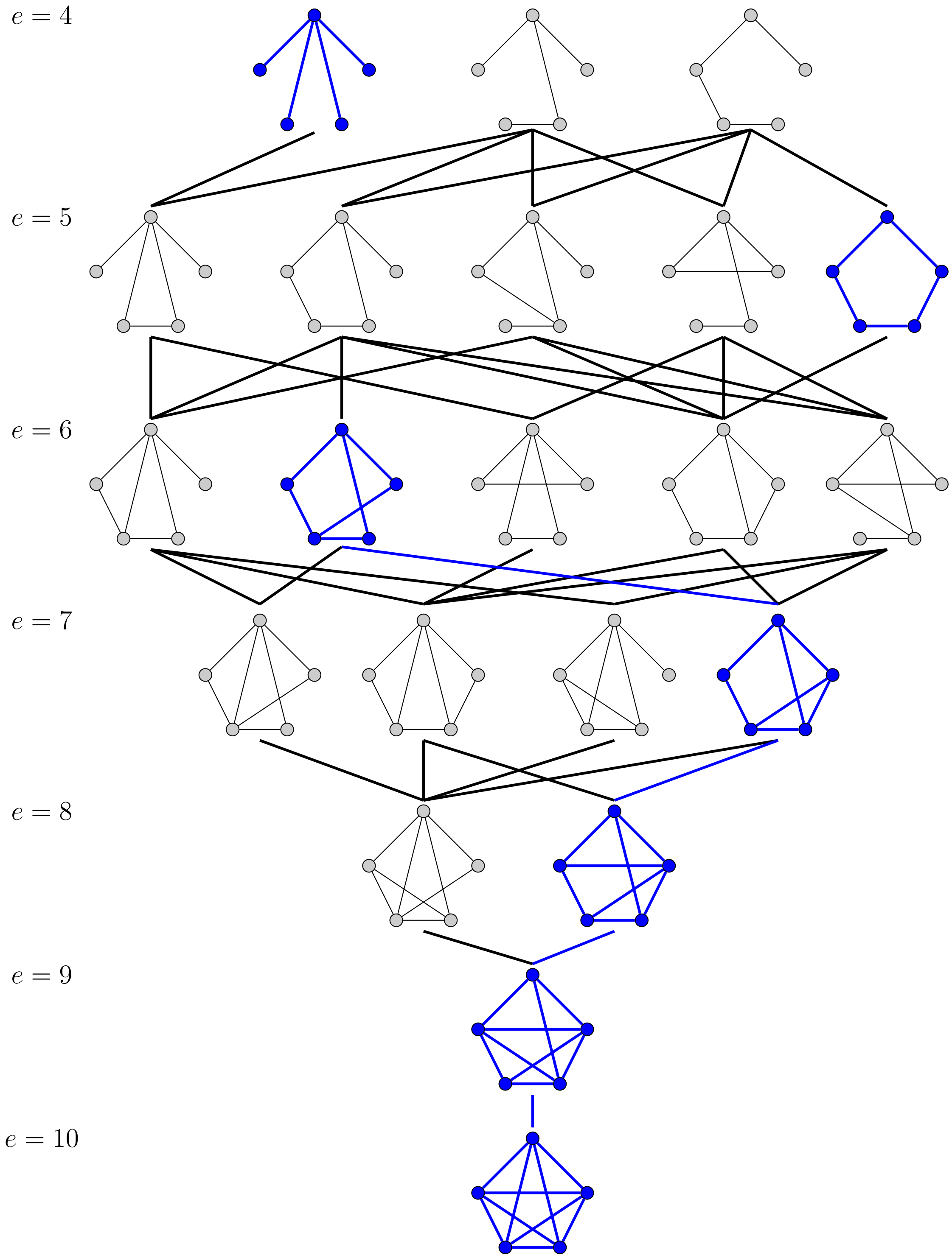


Figure 3: The graph of graphs for $n = 5$, optimal graphs are highlighted by blue and bold

vertices were always as close to each other as possible in the optimal graphs. The bipartiteness of the graphs also turned out to be important, i.e., the vertices were divisible to two independent sets A and B, and all edges were between a vertex from A and one from B.

As one can see in Figure 3, there is not a total optimal filling in sequence for the PCM in this case, however, we can determine a path that contains as many optimal cases as possible, and all the other graphs are close to optimal as well.

Similarly to Study I., our results were presented in different formats that make it possible for practitioners to instantly apply them in multicriteria decision making problems. Moreover, some of our findings can be used to determine the minimal number of questions to be asked from the decision makers in the case of a given problem, if we would like to have a threshold for the metrics examined. Based on our recent results (Gyarmati et al., 2023), which largely rely on Study III., the findings (the proposed graphs and sequences) also seem to be more general, not specific for PCMs, but optimal in the case of other pairwise comparison-based models as well.

1.5.4 Study IV. Incomplete pairwise comparison matrices: Ranking top women tennis players

Study IV. (Temesi et al., 2024) applies the same graph theoretical tools as the previously presented articles on a real database, while in some sense continues the research started by Bozóki et al. (2016).

Our aim was to create all-time rankings of top women tennis players based on an incomplete pairwise comparison matrix derived from their Win/Loss ratios. The alternatives were those players who were at some time between 1973 (the foundation of the Women's Tennis Association (WTA)) and 2022 number one in the official ranking of the WTA. We found 28 such players, and there were, of course, missing elements in our data, as some of them never played with each other during their career.

In order to apply the weight calculation techniques (besides checking whether the representing graph is connected), we had to make several technical adjustments in the data, in which we followed Bozóki et al. (2016).

If only one of the players won all the matches in a pair, then there should have been a 0 in the denominator of the given Win/Loss ratio. In this case, we used the number of wins and added 2 to it as the value in the matrix. After that, we also modified all the values by a power, in order to account for the fact that some pairs only played a few games against each other, while others clashed in several tournaments. The point of this modification was that if a pair had more matches with each other, than their comparison was considered to be more reliable.

We used both the eigenvector and the LLSM weight calculation methods with different parameters in the aforementioned adjustments, as well as applied another ranking technique as a benchmark, the Bradley–Terry model (Bradley and Terry, 1952). The rankings turned out to be similar to each other, the calculated Spearman rank correlations were convincingly high (0.860 was the smallest value) and robust to the modifications as well.

The first two places were occupied by Serena Williams and Steffi Graf, respectively, according to all methods. They were followed by Navratilova, Hingis, Clijsters, and Henin with slight modifications. Somewhat surprisingly, the main difference between the results of the Bradley–Terry model and the IPCM-based models was that some of the earliest players (Goolagong, Evert, and Austin) performed better according to the Bradley–Terry method.

Besides demonstrating that the MCDM methods connected to IPCMs can be applied to create rankings based on large real data sets, the goal of this research was also to thoroughly investigate the structure of the pairwise comparisons. We focused on several submatrices of the data as well as the case when the players entered to the data set one by one.

We found that it is possible to determine relatively large subsets of players with their most active career in the same time period, so that their results against each other practically determine their positions in the overall ranking. Entering players one by one to the data set showed that there are usually only small changes due to an entry, and every player affects the most those ones against whom she played the most.

Looking into the properties of the representing graph, it turned out that if we eliminate the four earliest players (Goolagong, Evert, Navratilova and Austin) along with the most recent world number one (Swiatek), then we get a graph, for which the different connectivity measures are really strong. The diameter (longest shortest path) of this modified graph is only 2, and it also can be interpreted as a union of two star graphs centered around the Williams sisters (i.e., they played with everyone else) complemented with further edges. This kind of structure and the strong connectivity might be the reason behind the robustness of the rankings.

We also analyzed the ordinally nontransitive triads in the data set. In sports, it can happen that A won against B in the majority of their matches, and generally B was successful against C as well. We would expect that A also beat C, however, in reality A lost against C in most occasions.

These kind of triads also have a significant literature, Kendall and Babington Smith (1940) determined the distribution of them for low number of elements and proposed a significance test, which was later extended by several researchers. Connected to PCMs and decision making problems, Iida (2009) investigated different tests and indices connected to the question.

In order to perform the necessary chi-square nontransitivity test, we had to modify the originally used IPCM, as it can be applied only on complete data without ties. The adjusted matrix contained W/L set ratios instead of match ratios for every pair. When we had ties even for the set ratios, the original LLSM ranking was used as a reference to make a precedence relation. If two players have never played against each other during their career (no edge exists between the two vertices), then we used the same LLSM ranking to determine the winner of the pair.

Applying these adjustments, we were able to construct a directed graph, on which the test itself is based on. The number of ordinally nontransitive triads turned out to be insignificant in our database, which can also explain the robustness of the rankings.

It needs to be highlighted that all the calculations and implementations (in R, Scilab, and Wolfram Mathematica) required to determine the different rankings and apply the tests of nontransitive triads are our individual contributions.

Our results can be interesting not only for tennis experts and fans, but also provide empirical evidence that the method of incomplete pairwise comparison matrices is appropriate for producing well-understandable rankings, as well as connect the research of representing graphs to a large real-world database and sports.

1.6 Directions for future research

Conducting the analysis connected to the four original studies of the thesis, we found numerous further questions as well as totally new directions for future research, and also started to work on some of them.

There are some natural continuations of the studies connected to filling in patterns and sequences for incomplete pairwise comparison matrices (Studies I., II., and III.), such as looking for larger cases. However, one could argue that it would be more important to use a wider range of methodology instead of simulations.

We could tackle this problem from at least two different directions. Empirical pairwise comparison matrices can be quite different from simulated ones, thus, it would be interesting to investigate some real-world PCMs as well, which have been used in the literature earlier (Bozóki et al., 2013). We can test which patterns are optimal in this case for given number of comparisons, as the order of filling in is also saved, and it was different for the decision makers.

On the other hand, it would be nice to prove formal propositions about the optimal graphs as well. Based on some recent research (Gyarmati et al., 2023), the results seem to be more general, but we could not characterize the optimal representing graphs yet. A key factor in this can be the fact that we focused on the closeness to the calculated weight vectors, which are the results of different optimization problems themselves. However, if we would focus on the matrices instead of the weight vectors, that could make the formal proofs much easier to conduct.

It is also possible to look into the graph of graphs of the labelled representing graphs. Which ordinal information is the best for the optimal designs of Study III.? Do these structures remain optimal if we can get some additional information? All of these issues can be interesting not just from a decision making point of view, but it could be used in sports as well. Which teams (players) should play against each other if we have some prior information about their strengths?

The concept of graph of graphs also raises some further research questions that can be connected to sports. What happens if, instead of direct paths in the graph of graphs, we are analyzing inclusion relations, i.e., we can make more than one comparisons in one step? A neat example can be the case of sports tournaments, where in every round all team (player) should play with another one, which leads to the graph of regular graphs.

In connection with Study IV., there are several possibilities to extend the data set, or create different rankings for different surfaces, which would be interesting for tennis experts and fans, but probably would not give extra methodological benefits. However, we see the calculations and tests about ordinally nontransitive triads as another important research direction. Currently, we had to modify the data to carry out a nontransitivity test. It would be interesting to extend these kinds of tests to incomplete data, incomplete pairwise comparison matrices as well, and we do believe that the key to this also resides in the tools of graph theory.

Of course, other areas connected to pairwise comparison matrices and graphs that have not been mentioned in the dissertation can be interesting too, such as the case of the Pareto efficient weight vectors (Blanquero et al., 2006; Ábele-Nagy and Bozóki, 2016; Ábele-Nagy et al., 2018; Fernandes and Furtado, 2022).

Consequently, that there are many different directions of research based on the studies included in the thesis. In our opinion, the collaboration with the Ukrainian research group also supports the relevance of these results. We have already started to work on some of these research questions, while we are still keen on to begin other ones.

References

Ábele-Nagy, K. and Bozóki, S. (2016). Efficiency analysis of simple perturbed pairwise comparison matrices. *Fundamenta Informaticae*, 144(3-4):279–289. https://doi.org/10.3233/FI-2016-1335.

Ábele-Nagy, K., Bozóki, S., and Rebák, Ö. (2018). Efficiency analysis of double perturbed pairwise comparison matrices. *Journal of the Operational Research Society*, 69(5):707–713. https://doi.org/10.1080/01605682.2017.1409408.

Ábele-Nagy, K. (2019). *Pairwise comparison matrices in multi-criteria decision* *making*. PhD thesis, Corvinus University of Budapest. In Hungarian. https://doi.org/10.14267/phd.2019036.

Ágoston, K. Cs. and Csató, L. (2022). Inconsistency thresholds for incomplete pairwise comparison matrices. *Omega*, 108:102576. https://doi.org/10.1016/j.omega.2021.102576.

Blanquero, R., Carrizosa, E., and Conde, E. (2006). Inferring efficient weights from pairwise comparison matrices. *Mathematical Methods of Operations Research*, 64(2):271–284. https://doi.org/10.1007/s00186-006-0077-1.

Bozóki, S., Csató, L., and Temesi, J. (2016). An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research*, 248(1):211–218. https://doi.org/10.1016/j.ejor.2015.06.069.

Bozóki, S. (2006). *Weighting with pairwise comparisons and evaluation with utility functions in multicriteria decision making problems*. PhD thesis, Corvinus University of Budapest. In Hungarian. https://phd.lib.uni-corvinus.hu/245/1/bozoki_sandor.pdf.

Bozóki, S., Dezső, L., Poesz, A., and Temesi, J. (2013). Analysis of pairwise comparison matrices: an empirical research. *Annals of Operations Research*, 211(1):511–528. https://doi.org/10.1007/s10479-013-1328-1.

Bozóki, S., Fülöp, J., and Rónyai, L. (2010). On optimal completion of incomplete pairwise comparison matrices. *Mathematical and Computer Modelling*, 52(1):318–333. https://doi.org/10.1016/j.mcm.2010.02.047.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4):324–345. https://doi.org/10.2307/2334029.

Brunelli, M. (2018). A survey of inconsistency indices for pairwise comparisons. *International Journal of General Systems*, 47(8):751–771. https://doi.org/10.1080/03081079.2018.1523156.

Crawford, G. and Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, 29(4):387–405. https://doi.org/10.1016/0022-2496(85)90002-1.

Csató, L. (2015). *Methodological and applicational issues of paired comparison based ranking*. PhD thesis, Corvinus University of Budapest. In Hungarian. https://doi.org/10.14267/phd.2015022.

Csató, L. (2021). *Tournament Design: How Operations Research Can Improve Sports Rules*. Palgrave Pivots in Sports Economics, Palgrave Macmillan. https://doi.org/10.1007/978-3-030-59844-0.

Davidson, R. and Farquhar, P. (1976). A bibliography on the method of paired comparisons. *Biometrics*, 32(2):241–252. https://www.jstor.org/stable/2529495.

Edwards, W. (1977). How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(5):326–340. https://doi.org/10.1109/TSMC.1977.4309720.

Fernandes, R. and Furtado, S. (2022). Efficiency of the principal eigenvector of some triple perturbed consistent matrices. *European Journal of Operational Research*, 298(3):1007–1015. https://doi.org/10.1016/j.ejor.2021.08.012.

Fishburn, P. C. (1970). *Utility Theory for Decision Making*. Wiley, New York.

Garuti, C. E. (2017). Reflections on scales of measurement, not measurement of scales. *International Journal of the Analytic Hierarchy Process*, 9(3). https://doi.org/10.13033/ijahp.v9i3.522.

Gass, S. (1998). Tournaments, transitivity and pairwise comparison matrices. *Journal of the Operational Research Society*, 49(6):616–624. https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2600572.

Gyarmati, L., Orbán-Mihálykó, É., Mihálykó, Cs., Szádoczki, Zs., and Bozóki, S. (2023). The incomplete analytic hierarchy process and Bradley–Terry model: (In)consistency and information retrieval. *Expert Systems with Applications*, 229(Part B):120522. https://doi.org/10.1016/j.eswa.2023.120522.

Harker, P. T. (1987). Incomplete pairwise comparisons in the analytic hierarchy process. *Mathematical Modelling*, 9(11):837–848. https://doi.org/10.1016/0270-0255(87)90503-3.

Iida, Y. (2009). The number of circular triads in a pairwise comparison matrix and a consistency test in the AHP. *Journal of the Operations Research Society of Japan*, 52(2):174–185. https://doi.org/10.15807/jorsj.52.174.

Ishizaka, A. and Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. Wiley. https://doi.org/10.1002/9781118644898.

Kadenko, S. and Tsyganok, V. (2020). An update on combinatorial method for aggregation of expert judgments in AHP. Proceedings of the International Symposium on the Analytic Hierarchy Process, ISAHP-2020. https://doi.org/10.13033/isahp.y2020.012.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan, New York.

Kendall, M. G. and Babington Smith, B. (1940). On the method of paired comparisons. *Biometrika*, 31(3/4):324–345. https://doi.org/10.2307/2332613.

Kou, G., Peng, Y., Chao, X., Herrera-Viedma, E., and Alsaadi, F. E. (2021). A geometrical method for consensus building in GDM with incomplete heterogeneous preference information. *Applied Soft Computing*, 105:107224. https://doi.org/10.1016/j.asoc.2021.107224.

Kułakowski, K., Szybowski, J., and Prusak, A. (2019). Towards quantification of incompleteness in the pairwise comparisons methods. *International Journal of Approximate Reasoning*, 115:221–234. https://doi.org/10.1016/j.ijar.2019.10.002.

Kułakowski, K. and Talaga, D. (2020). Inconsistency indices for incomplete pairwise comparisons matrices. *International Journal of General Systems*, 49(2):174–200. https://doi.org/10.1080/03081079.2020.1713116.

Lovász, L. (1977). A homology theory for spanning trees of a graph. *Acta Mathematica Academiae Scientiarum Hungaricae*, 30(3-4):241–251. https://doi.org/10.1007/bf01896190.

McCormick, E. and Bachus, J. (1952). Paired comparison ratings: 1. The effect on ratings of reductions in the number of pairs. *Journal of Applied Psychology*, 36(3):123–127. https://doi.org/10.1037/h0054842.

McCormick, E. and Roberts, W. (1952). Paired comparison ratings: 2. The reliability of ratings based on partial pairings. *Journal of Applied Psychology*, 36(3):188–192. https://doi.org/10.1037/h0055956.

Mi, X., Tang, M., Liao, H., Shen, W., and Lev, B. (2019). The state-of-the-art survey on integrations and applications of the best worst method in decision making: Why, what, what for and what's next? *Omega*, 87:205–225. https://doi.org/10.1016/j.omega.2019.01.009.

Miller, M. and Širáň, J. (2013). Moore graphs and beyond: A survey of the degree/-diameter problem. *Electronic Journal of Combinatorics*, 20(2):1–92. https://doi.org/10.37236/35.

Poesz, A. (2022). *Inconsistency in multi-attribute decision problems.* PhD thesis, Corvinus University of Budapest. In Hungarian. https://doi.org/10.14267/phd.2022053.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, 53:49–57. https://doi.org/10.1016/j.omega.2014.11.009.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281. https://doi.org/10.1016/0022-2496(77)90033-5.

Saaty, T. L. (1980). *The Analytic Hierarchy Process.* McGraw-Hill, New York.

Szádoczki, Zs. and Bozóki, S. (2022). Optimal sequences for pairwise comparisons: the graph of graphs approach. *Working paper.* https://doi.org/10.48550/arXiv.2205.08673.

Szádoczki, Zs., Bozóki, S., Juhász, P., Kadenko, S. V., and Tsyganok, V. (2023). Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. *Annals of Operations Research*, 326(2):783–807. https://doi.org/10.1007/s10479-022-04819-9.

Szádoczki, Zs., Bozóki, S., and Tekile, H. A. (2022). Filling in pattern designs for incomplete pairwise comparison matrices: (Quasi-)regular graphs with minimal diameter. *Omega*, 107:102557. https://doi.org/10.1016/j.omega.2021.102557.

Szybowski, J., Kułakowski, K., and Prusak, A. (2020). New inconsistency indicators for incomplete pairwise comparisons matrices. *Mathematical Social Sciences*, 108:138–145. https://doi.org/10.1016/j.mathsocsci.2020.05.002.

Tekile, H. A. (2017). Incomplete pairwise comparison matrices in multi-criteria decision making and ranking. Master's thesis, Central European University.

Temesi, J., Szádoczki, Zs., and Bozóki, S. (2024). Incomplete pairwise comparison matrices: Ranking top women tennis players. *Journal of the Operational Research Society*, 75(1):145–157. https://doi.org/10.1080/01605682.2023.2180447.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286. https://doi.org/10.1037/h0070288.

von Winterfeldt, D. and Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge.

Wang, K. and Takahashi, I. (1998). How to select paired comparisons in AHP of incomplete information – strongly regular graph design. *Journal of the Operations Research Society of Japan*, 41(2):311–328. https://doi.org/10.15807/jorsj.41.311.

Ye, J. (2012). Multicriteria decision-making method using the Dice similarity measure based on the reduct intuitionistic fuzzy sets of interval-valued intuitionistic fuzzy sets. *Applied Mathematical Modelling*, 36(9):4466–4472. https://doi.org/10.1016/j.apm.2011.11.075.

Zhou, X., Hu, Y., Deng, Y., Chan, F. T. S., and Ishizaka, A. (2018). A DEMATEL-based completion method for incomplete pairwise comparison matrix in AHP. *Annals of Operations Research*, 271:1045–1066. https://doi.org/10.1007/s10479-018-2769-3.

List of publications

Publications included in the Ph.D. thesis

Szádoczki, Zs. and Bozóki, S. (2022). Optimal sequences for pairwise comparisons: the graph of graphs approach. *Working paper, under review.* https://doi.org/10.48550/arXiv.2205.08673.

Szádoczki, Zs., Bozóki, S., Juhász, P., Kadenko, S. V., and Tsyganok, V. (2023). Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. *Annals of Operations Research*, 326(2):783–807. https://doi.org/10.1007/s10479-022-04819-9.

Szádoczki, Zs., Bozóki, S., and Tekile, H. A. (2022). Filling in pattern designs for incomplete pairwise comparison matrices: (Quasi-)regular graphs with minimal diameter. *Omega*, 107:102557. https://doi.org/10.1016/j.omega.2021.102557.

Temesi, J., Szádoczki, Zs., and Bozóki, S. (2024). Incomplete pairwise comparison matrices: Ranking top women tennis players. *Journal of the Operational Research Society*, 75(1):145–157. https://doi.org/10.1080/01605682.2023.2180447.

Publications that are not included in the Ph.D. thesis

Duleba, Sz. and Szádoczki, Zs. (2022). Comparing aggregation methods in large-scale group AHP: Time for the shift to distance-based aggregation. *Expert Systems with Applications*, 196:116667. https://doi.org/10.1016/j.eswa.2022.116667.

Gyarmati, L., Orbán-Mihálykó, É., Mihálykó, Cs., Szádoczki, Zs., and Bozóki, S. (2023). The incomplete analytic hierarchy process and Bradley–Terry model: (In)consistency and information retrieval. *Expert Systems with Applications*, 229(Part B):120522. https://doi.org/10.1016/j.eswa.2023.120522.

Kadenko, S., Tsyganok, V., Szádoczki, Zs., and Bozóki, S. (2021). An update on combinatorial method for aggregation of expert judgments in AHP. *Production*, 31:e20210045. https://doi.org/10.1590/0103-6513.20210045.

Szabo, Zs. K., Szádoczki, Zs., Bozóki, S., Stănciulescu, G. C., and Szabo, D. (2021). An Analytic Hierarchy Process approach for prioritisation of strategic objectives of sustainable development. *Sustainability*, 13(4). https://doi.org/10.3390/su13042254.

Szádoczki, Zs. (2022). Operations research for the profitability of sports. Review of the book *Tournament Design. How Operations Research Can Improve Sports Rules?* by L. Csató. *Közgazdasági Szemle*, 69(2):283–288. In Hungarian. https://doi.org/10.18414/KSZ.2022.2.283.

Szádoczki, Zs. and Duleba, Sz. (2022). Distance-based aggregation in group AHP. *Journal of Decision Systems*, 31(sup1):98–106. https://doi.org/10.1080/12460125.2022.2070952.

Temesi, J., Szádoczki, Zs., and Bozóki, S. (2022). Incomplete pairwise comparisons: The example of the ranking of top women tennis players. *Szigma*, 53(1):1–32. In Hungarian. https://journals.lib.pte.hu/index.php/szigma/article/view/5709/5492.

Working papers that are not included in the Ph.D. thesis

Csató, L., Kiss, L. M., and Szádoczki, Zs. (2024). The allocation of FIFA World Cup slots based on the ranking of confederations. *Working paper, under review.* https://doi.org/10.48550/arXiv.2310.19100.

Kelemen, A., Szabo, Zs. K., Szádoczki, Zs., and Bozóki, S. (2024). A sensitivity analysis of composite indicators: Min/max thresholds. *Working paper, under review.*

2 Study I. Filling in pattern designs for incomplete pairwise comparison matrices: (Quasi-)regular graphs with minimal diameter

Authors: Zsombor Szádoczki, Sándor Bozóki, Hailemariam Abebe Tekile

Published in Omega, 107: 102557. (2022)

https://doi.org/10.1016/j.omega.2021.102557

Abstract

Pairwise comparisons have become popular in the theory and practice of preference modelling and quantification. In case of incomplete data, the arrangements of known comparisons are crucial for the quality of results. We focus on decision problems where the set of pairwise comparisons can be chosen and it is designed completely before the decision making process, without any further prior information. The objective of this paper is to provide recommendations for filling patterns of incomplete pairwise comparison matrices based on their graph representation. The proposed graphs are regular and quasi-regular ones with minimal diameter (longest shortest path). Regularity means that each item is compared to others for the same number of times, resulting in a kind of symmetry. A graph on an odd number of vertices is called quasi-regular, if the degree of every vertex is the same odd number, except for one vertex whose degree is larger by one. We draw attention to the diameter, which is missing from the relevant literature, in order to remain the closest to direct comparisons. If the diameter of the graph of comparisons is as low as possible (among the graphs of the same number of edges), we can decrease the cumulated errors that are caused by the intermediate comparisons of a long path between two items. Contributions of this paper include a list containing (quasi-)regular graphs with diameter 2 and 3 up until 24 vertices. Extensive numerical tests show that the recommended graphs indeed lead to better weight vectors compared to various other graphs with the same number of edges. It is also revealed by examples that neither regularity nor small diameter is sufficient on its own, both properties are needed. Both theorists and practitioners can utilize the results, given in several formats in the appendix: plotted graph, adjacency matrix, list of edges, 'Graph6' code.

Keywords: pairwise comparison, incomplete pairwise comparison matrix, graph, diameter, regular graph

2.1 Introduction

Pairwise comparisons form the basis of preference measurement, ranking, psychometrics and decision modelling (Davidson and Farquhar, 1976; Thurstone, 1927; Zahedi, 1986). Multicriteria Decision Making is indeed an important tool both at an individual and at an organizational level. We can think about different kind of ranking of alternatives or weighting of criteria, like tenders, selection among schools or job offers, selection among the implementation of different projects in an enterprise, etc.

One of the most commonly used techniques in connection with Multicriteria Decision Making is the method of the pairwise comparison matrices. One can apply this technique both for determining the weights of the different criteria and for the rating of the alternatives according to a criterion. Usually we denote the number of criteria or alternatives by n, which means the pairwise comparison matrix is an $n \times n$ matrix, often denoted by A. In this case the ij-th element of the A matrix, a_{ij} shows how many times the i-th item is larger/better than the j-th element.

Formally, matrix A is called a pairwise comparison matrix (PCM) if it is positive ($a_{ij} > 0$ for $\forall$ i and j) and reciprocal ($1/a_{ij} = a_{ji}$ for $\forall$ i and j) (Saaty, 1980), which also indicates that $a_{ii} = 1$ for $\forall$ i.

Dealing with incomplete data gets more and more attention in the literature. When some elements of a PCM are missing we call it an incomplete PCM. There could be many different reasons why these elements are absent, some data could have been lost or the comparisons are simply not possible (for instance in sports (Bozóki et al., 2016)).

The most interesting case for us is when the decision makers do not have time, willingness or the possibility to make all the $n(n-1)/2$ comparisons.

In this article we would like to study which comparisons should be made, or more precisely what patterns of comparisons are recommended in order to get good approximation of the decision makers' preferences calculated from the whole set of comparisons. The graph representation of the pairwise comparisons is a natural and convenient tool to examine our question, thus we will use this throughout the paper.

In many cases the set of comparisons can be adaptive, i.e., the next questions depend on the answers to the previous ones as in, e.g., Ciomek et al. (2017); Fedrizzi and Giove (2013); Glickman and Jensen (2005). However, we assume in the paper that the whole set of comparisons is designed completely before the decision making process, and we do not have any further prior information about the items to be compared. Thus the 'confidence level' of every single comparison is the same in our problems, the probability of their 'errors' is identical. For instance the (pre-)compilation of questionnaires in connection with decision making problems can be named as an indeed common practical example that satisfies these conditions.

There are already known special structures proposed for incomplete pairwise comparison matrices in the literature, which include:

(i) spanning tree, in particular if one row/column is filled in completely (its associated graph is the star graph)

(ii) two rows/columns are filled in completely (its associated graph is the union of two star graphs) (Rezaei, 2015)

(iii) a method of 2-cyclic designs, the union of two edge-disjoint n-cycles, has been also recommended to select $2n$ paired comparisons from n number of objects (Miyake et al., 2003)

(iv) more or less regular graphs, for example the regularity of the comparisons' graph appears in the designs of McCormick and Bachus (1952) and McCormick and Roberts (1952).

Regularity results in a kind of symmetry that is also desirable in case of sport competitions (Csató, 2013), where the number of matches played equals for every player or team, at least in the first phase (before the knockout stages). This also appears in other sport tournaments, where they use the so-called Swiss system, in which besides a lot of other requirements, every player or team plays the same number of matches (if possible) (Ólafsson, 1990; Biró et al., 2017; Kujansuu et al., 1999). Thus the resulting representing graph of the comparisons is regular (Csató, 2017).

A special type and extension of regular graphs is considered by Wang and Takahashi (1998). They proposed the (quasi-)strongly regular designs based on (quasi-)strongly regular graphs in order to select pairs to be compared within incomplete information. A graph is called strongly regular with parameters (n, k, λ, μ), if each of the n vertices has degree k, and (i) for any pair of adjacent vertices u and v, the number of vertices adjacent to both u and v is λ; (ii) for any pair of not adjacent vertices

u and v, the number of vertices adjacent to both u and v is μ. Since these properties are rather restrictive, a linear algebraic generalization, the so called quasi-strongly regular graphs are also taken into consideration. By simulation, they showed that both designs give better results (based on a logarithmic distance function defined on the weight vectors) than other random designs of the same cardinality.

Kułakowski et al. (2019) create an incompleteness index based on the number of missing pairwise comparisons and their arrangements. Using different kind of Monte Carlo simulations they conclude that inconsistency and incompleteness both have crucial effect on sensitivity, and the regularity of the PCM also has a huge effect both on the quantitative and the qualitative results.

Note that the first three examples above lack regularity. Regularity means that each item is compared to others for the same number of times (if the cardinality of the items to compare is odd, one of the degrees can be smaller or greater – in our analysis, greater – by one), resulting in a kind of symmetry, as we mentioned earlier. Despite the fact that regularity has been recognized as an important property in connection with the representing graph of the comparisons, the above-mentioned examples do not examine it as generally as we do, their definitions on regularity is more restrictive and their instances are less systematic.

Diameter, the other key concept of the paper besides regularity, shows how far items can be from each other in the sense that how many comparisons are needed in order to have an indirect comparison between them. The well known telephone game or effect (Ribeiro et al., 2019), also known as The Whisper Game (Chatburn, 2013) shows small errors are cumulated along a sufficiently long series. If a message passes through a line of people, in a whisper, the original and the final versions differ a lot, despite the neighboring versions are usually quite similar. A classical example for the non-transitivity of indifference (Fishburn, 1970) is the addition of very small portions of sugar to the same cup of coffee. No one can distinguish between two consecutive steps, however, if this sequence is long enough, the indifference disappears (Luce, 1956).

In the set of connected graphs, diameter can be considered as a measure of closeness, or a stronger type of connectedness. It is not properly studied in the literature, however, for instance in Pananjady et al. (2020) the estimation of the matrix of comparison probabilities is investigated for several graph structures and some research questions, e.g., on a possible relation of the graph's diameter and the worst-case approximation error, are raised. One of our notable findings is to determine the diameter of the representing graph as a crucial property for filling in pattern designs of incomplete PCMs.

Note that regular graphs can have large diameter, e.g., a cycle on n vertices is 2-regular and has diameter $d = \lfloor n/2 \rfloor$. The star graph, mentioned among the examples, has minimal diameter 2, but it is far from being regular. Our aim is to find the graphs, among (quasi-)regular ones, with minimal diameter. We are especially interested in the smallest nontrivial values of the diameter, namely $d = 2$ and $d = 3$. Intuition suggests, and it is confirmed by the graphs found, that for a fixed n, higher regularity, i.e., more edges, makes the diameter smaller.

The rest of the paper is structured as follows. Basic mathematical concepts are introduced in Section 2.2. Later on we assume that we know the number n of alternatives or criteria, it is also a key assumption through our paper that the graph representing the MCDM problem is k-(quasi-)regular and we also know (or with the help of the other inputs we can determine) the diameter d of the graph. In Section 2.3 (which is complemented by Appendix A (online)) we provide a systematic collection of suggested incomplete pairwise comparisons' patterns with the help of the above-mentioned inputs and all/some graphs for the examined cases. We would like to emphasize that this list is a major contribution of our paper. Section 2.4 presents a motivational example showing that the diameter of a regular graph can be large and the result can be very sensitive to the errors of the matrix elements. A wide range of numerical simulations, using the distances of the weights computed with different filling in patterns respect to the weights calculated from the complete PCMs, is also provided in order to validate our recommendations. Finally, Section 2.5 concludes and provides further research questions closely connected to the discussed topic. Results of Sections 2.3 and 2.4 are given in more details in the appendices. B (online) includes the recommended graphs themselves. For practitioners, this list might serve as a 'recipe' in designing questionnaires based on pairwise comparisons. Appendix D (online) includes the results of the comparisons of weight vectors calculated from the different graphs.

2.2 Basic concepts of the graph representation

The graph representation of paired comparisons has already been used in the 1940s (Kendall and Babington Smith, 1940). Of course after the widespread application of PCMs and incomplete PCMs it has become a common method in the literature, see for instance Blanquero et al. (2006), Csató (2015) or Gass (1998).

Usually in these articles the authors use directed graphs for the representation, because they distinguish the preferred item from the less preferred one in every pair. In our approach the only important thing is whether there exists a comparison between the two elements. This means that we use undirected graphs, where the vertices denote the criteria or the alternatives. There is an edge between two vertices if and only if the decision makers made their comparison for the two respective items (the appropriate element of the PCM is known). In order to understand the concepts so far, there is a small example below:

Example 1 *Let us assume that there are 4 criteria ($n = 4$) and our decision maker already answered some questions, denoted their locations in the matrix by $\bullet$ and their reciprocal values by $\circ$, which lead to the following incomplete PCM:*

$$A = \begin{bmatrix} 1 & \bullet & \bullet & \\ \circ & 1 & & \bullet \\ \circ & & 1 & \bullet \\ & \circ & \circ & 1 \end{bmatrix}$$

This incomplete PCM is represented by the graph in Figure 4.

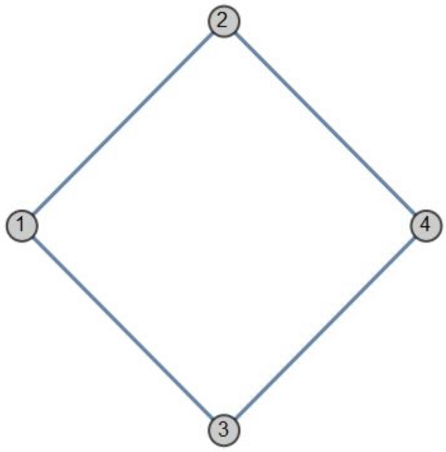


Figure 4: Graph representation example

As we can see there is no edge between the first and the fourth vertices, where the PCM has missing values and there is no edge between the second and third vertices, where the situation is the same. There is an edge between every other pair, where we have no missing values in the PCM.

It is important to emphasize that as the known elements of the PCM determine the representing graph, it is also true in the other way around. Thus, the graph in Figure 4 shows which comparisons are known in the PCM. This is the key property that we use in this paper, as we present the representing graphs that show the filling in patterns, the comparisons that should be made. We assume that the representing graphs are connected and k-(quasi-)regular through our paper, thus we need some definitions to make these concepts clear.

Definition 13 (Connected graph) *In an undirected graph, two vertices u and v are called connected if the graph contains a path from u to v. A graph is said to be connected if every pair of vertices in the graph is connected.*

Definition 14 (k-regular graph) *A graph is called k-regular if every vertex has k neighbours, which means that the degree of every vertex is k.*

Definition 15 (k-quasi-regular graph) *A graph is called k-quasi-regular if exactly one vertex has degree $k+1$, and all the other vertices have degree k.*

The k-regularity basically means that the vertices are not distinguished, there is no particular vertex as, for example, in the case of the star graph, thus we would like to avoid the cases when the elimination of relatively few vertices would lead to the disintegration of the whole comparison system (Tekile, 2017). Besides regularity, the connectedness of the representing graph is indeed important, because to approximate the decision makers' preferences well, we need to have at least indirect comparisons between the different criteria, otherwise we cannot say anything about the relation between certain elements (Bozóki et al., 2010).

However, it is also notable that we would like to avoid the cases when two items are compared only indirectly through a very long path, because this could aggregate the small, tolerable errors of the different comparisons and we could end up with an intolerably large error in the relation between the two elements. Such an example was found in Tekile (2017), where the graph generated from the table tennis players' matches included a long shortest path between two vertices (players), and the calculated result appeared to be misleading. The diameter of the representing graph is a very suitable mathematical tool to measure this problem:

Definition 16 (The diameter of a graph) *The diameter (denoted by d) of a graph G is the length of the longest shortest path between any two vertices:*

$$d = \max_{u,v \in V(G)} \ell(u,v),$$

where $V(G)$ denotes the set of vertices of G and $\ell(.,.)$ is the graph distance between two vertices, namely the length of the shortest path between them.

We also define here the twisted product, a graph construction method that is used by us extensively to find the proposed graphs:

Definition 17 (Twisted product of two graphs) *(Bermond et al. (1982))*

Let $G = (V,E)$ and $G' = (V',E')$ be two undirected graphs, where V and V' are the vertex sets, while E and E' are the edge sets of the respective graphs. Let $\overrightarrow{E}$ denote the set of arcs in an arbitrary orientation of G. For each arc $(i,j) \in \overrightarrow{E}$, let $\pi_{(i,j)}$ be a one to one mapping from V' to itself. The twisted product of graphs G and G', denoted by $G * G'$ is defined as follows: its vertex set is the Cartesian product $V \times V'$, and there is an edge between vertices (i,i') and (j,j') if either [$i = j$ and $(i',j') \in E'$] or [$(i,j) \in \overrightarrow{E}$ and $j' = \pi_{(i,j)}(i')$].

Note that the twisted product with π =identity results in the Cartesian product.

Briefly from now on we will examine graphs representing MCDM problems defined by the following inputs: (n, k, d), where n is the number of vertices (criteria), k shows the level of regularity of the graph and d is the diameter of the graph.

2.3 (Quasi-)regular graphs with minimal diameter

In this section, we present one of the most important findings of the paper, the examined (quasi-)regular graphs themselves. First of all, it is a key step to determine which cases are interesting for us considering our inputs. It is important to emphasize that we deal with unlabelled graphs, because we are trying to find out what kind of patterns are needed in the comparisons for different instances. Thus if we exchange the 'names' of two criteria (for instance '1' and '2' in Example 1) the pattern would be the same.

Then we can consider the regularity parameter k. The $k = 1$ case is possible only when n is even, but they are not connected except for $n = 2$, so this is not interesting for us. When $k = 2$ there is only one connected graph for every n, namely the cycle, for which $d = \lfloor n/2 \rfloor$ as already mentioned in the introduction.

The larger regularity parameters could be interesting, but of course we need a reasonable upper bound for the number of criteria, n, which is also an indirect upper bound for k. In our research we examined the $n = 5, 6, \dots, 24$ cases, because on the one hand for larger n parameters, some computations become very difficult, and on the other hand the largest 5-regular graph with diameter 2 contains 24 vertices, so this is a nice theoretical bound, as well. It is also true that in the majority of the fields of application it is sufficient to examine the number of alternatives (vertices) up until 24.

The smaller the d parameter is, the more stable or trustworthy our system of comparisons is. This means that in an optimal case we would like to minimize this parameter, while the number of the criteria (n) is always a fixed exogenous parameter in our MCDM problems. As we mentioned above, k is crucial to avoid the cases when some criteria (vertices) would be too important in the system, however it also shows us how many comparisons have to be made, because every vertex has a degree of k, which means the number of edges is $nk/2$. Thus if our decision makers would like to spend the shortest time with the creation of the PCM, we should choose a small k parameter. But, of course, as usually happens in these situations, there is a trade off between the parameters, because for many criteria (large n) the smaller regularity (k) will cause a bigger diameter (d), namely, a more fragile system of comparisons.

In this paper we would like to provide a list of graphs which shows the patterns of the comparisons that have to be made in case of different parameters. We used computational and constructing methods to determine the graph(s) with the smallest diameter (d parameter) for a given (n, k) pair. With the help of these results it was easy to determine which k is the smallest that is needed to reach a given d for a given n. We found that, with the chosen upper bound of n (24) the interesting values for the regularity are $k = 3, 4, 5$, while the interesting values for the diameter of the graph are $d = 2, 3$. Of course $d = 1$ would mean a complete graph that is not reachable for many (n, k) pairs, and it represents a complete PCM, thus it is not interesting for us. For a general MCDM problem probably instead of k, it would give more information if we considered an indicator that shows how far we are from the 'extreme' case, when the decision makers have to make all the comparisons. This would mean $n(n-1)/2$ comparisons instead of our $nk/2$ in case of regular graphs or $(nk+1)/2$ in case of quasi-regular graphs, therefore the completion ratio is defined as follows:

$$c = \begin{cases} \frac{nk/2}{n(n-1)/2} & \text{if } n \text{ or } k \text{ is even} \\ \\ \frac{(nk+1)/2}{n(n-1)/2} & \text{if } n \text{ and } k \text{ are odd} \end{cases}$$

that we will calculate for every instance.

Here we will present the graphs with the smallest diameter for a given (n, k) pair, it is important to emphasize that it is recommended to read this section together with Appendix A, as a large part of our list (Tables 2, A1a, A1b, A2, A3, A4a, and A4b) takes place there, because of the length of the tables. The finding for the different graphs in our list consisted of several methods, sources and layers:

1. As a starting reference point, we checked the built in graphs in Wolfram Mathematica (Wolfram Research, 2021), which are even complete catalogues in case of small number of vertices, thus we selected the ones with minimal diameter among them.

2. For smaller and middle-sized graphs, when Mathematica's built in examples cover only a sample of the cases, we used nauty and Traces (McKay and Piperno, 2014) and IGraph/M (Horvát, 2020) to generate all the possible (quasi-)regular graphs and select the needed ones.

3. Our results contain many well known graphs as well, like the Petersen graph (Holton and Sheehan, 1993), that we collected from different kind of articles indicated in the respective tables as 'Source'. We also collected further information, like uniqueness, about those graphs that we got with the help of Mathematica and are well known cases. We cite these information as 'See also' in our tables.

4. For larger graphs we were not able to generate all the possible regular cases, thus we used several construction techniques such as the twisted product, integer linear programming or merging and extending methods with the help of some already known graphs. Many of these cases were challenging and time-consuming to find, the same idea rarely worked twice.

5. It is also important that as k-quasi-regularity was defined by us, all of the quasi-regular graphs are our findings (or at least we are the first to use them in this kind of context), but we do not denote this separately in the tables.

Table 1 presents a table of tables that provides an overview of our list of graphs.

	k		
n	**3**	**4**	**5**
$\boldsymbol{n = 5, \ldots, 10}$	Table 2		
$\boldsymbol{n = 11, \ldots, 15}$	Table A1a	Table A2	
$\boldsymbol{n = 16, \ldots, 20}$	Table A1b		Table A4a
$\boldsymbol{n = 21, \ldots, 24}$		Table A3	Table A4b

Table 1: The summary of our list of graphs: the different sets of graphs based on the regularity level k and the number of vertices n can be found in the indicated tables. Lightgray denotes $d = 2$ and gray denotes $d = 3$.

Table 2 shows the cases when $k = 3$ and $d = 2$ is the minimal value of the parameter. It is important to note that $k = 3$ is only possible when n is even, but when it is odd, we examine 3-quasi-regular graphs, where all vertices have degree 3 except one where it has 4, because these are the closest to 3-regularity.

We can see that with $k = 3$ the minimal diameter can be 2 until we have 10 vertices. Of course for $n \leq 3$ the 3-regularity is not possible, and for $n = 4$ the diameter is 1, because this is a complete graph, that is why we skip those in the table. It is also notable that the completion ratio (c) even reaches 1/3 when we have 10 vertices (it is obviously decreasing in n). We should emphasize the fact that there are only a few graphs for every (n, k) pair with the minimal diameter. Some of them are bipartite graphs, which have special spectral properties (Csató, 2015, Lemma 4, Theorem 2, Proposition 3), and they also indicate that there are two groups which are always compared through the other ones.

If we go on to larger graphs ($n > 10$), then we will find that the smallest reachable diameter changes to $d = 3$, but it is also true that at first we have so many graphs that satisfy these properties. However, as we examine the $n = 18$ or the $n = 20$ cases, we can see that there is only one graph that fulfils our assumptions (Pratt,

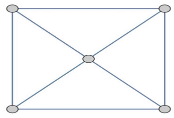
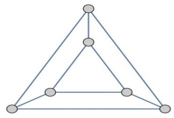
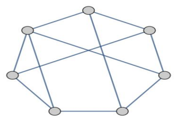
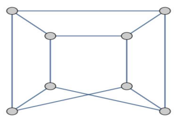
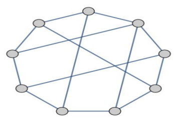
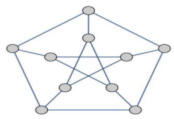
$\boldsymbol{k=3}$	Graph	Further information
$\boldsymbol{n=5}$		• $c = 8/10 = 0.8$ • Unique graph
$\boldsymbol{n=6}$	 3-prism graph ($C_3 \times K_2$)	• $c = 9/15 = 0.6$ • 2 graphs • Source: Pratt (1996) • The other solution is the bipartite graph $K_{3,3}$
$\boldsymbol{n=7}$		• $c = 11/21 \approx 0.524$ • 4 graphs
$\boldsymbol{n=8}$	 Wagner graph	• $c = 12/28 \approx 0.429$ • 2 graphs • See also: Maharry and Robertson (2016) • The other solution is the X_8 graph (Bermond et al., 1982)
$\boldsymbol{n=9}$		• $c = 14/36 \approx 0.389$ • 2 graphs
$\boldsymbol{n=10}$	 Petersen graph	• $c = 15/45 \approx 0.333$ • Unique graph • See also: Hoffman and Singleton (1960)

Table 2: $k = 3$-(quasi-)regular graphs on n vertices with minimal diameter $d = 2$

1996). The results in case of larger graphs, with 3-regularity and 3 as the minimal diameter can be found in Tables A1a and A1b in Appendix A.

As we can see the completion ratio is still decreasing in n and on larger graphs it can be taken below 0.2. It is also true that we do not need to answer for more than 30 questions for an MCDM problem even with 20 criteria, which can be indeed useful.

If we go on to larger graphs, the minimal diameter would change to $d = 4$, however, in this paper we only consider the graphs with $d \leq 3$, so we discussed the interesting cases for $k = 3$. The former results mean that, if we would like to examine the graphs where $k = 4$, it is obvious that the minimal diameter would be 2 until $n = 10$, but it is not so important to make so many comparisons because this property can be reached with $k = 3$, as well. Thus for $k = 4$ the interesting cases start above 10 vertices, and the question is if we can reach a smaller diameter (a more stable system of comparisons) with the rise of the answered questions. We found that with $k = 4$ we can get 2 as the minimal diameter until $n = 15$, but for larger values of n, it will be 3 again, which can be also reached by $k = 3$, thus we would not recommend these combinations of parameters. The results for $(11 \leq n \leq 15, k = 4)$ are shown in Table A2. It is also important to note that $k = 4$ is possible in case of both odd and even values of n, thus now we do not have to pay special attention to this.

As we can see, the completion ratio is increasing in k, so we cannot get so small c values as in Table A1a, however the system of comparisons will be more stable even on many vertices, because the smallest diameter is 2 here. It is also interesting that, for larger graphs and regularity levels, the number of connected graphs increases very rapidly. For instance, when we have 15 vertices, there are 805 491 connected 4-regular graphs (that means 805 491 possible filling patterns of the PCM), and only one has 2 as its diameter. Our results and methodology has a strong relationship with the so-called degree/diameter problem that is well known in the literature of mathematics (Elspas (1964), Dinneen and Hafner (1994), Loz and Širáň (2008)), but they are looking for the largest possible n for a given diameter and a given level of maximum degree. Several construction techniques have been proposed for graphs in connection with the degree/diameter problem (Storwick, 1970; Bermond et al., 1982; Branković et al., 1998), and one can also find extended tables with the known results (Comellas and Gómez, 1994). For an indeed extensive summary of the problem, see Miller and Širáň (2013). The scientific results in this field support our findings, too, because for $(k = 3, d = 2)$ the largest n is 10, while for $(k = 3, d = 3)$ it is 20. In the case of $(k = 4, d = 2)$ the largest n is 15, but for $(k = 4, d = 3)$ it is proven that the largest graph is much above our bound, while the optimal number of the vertices in this case is still an open question.

As we mentioned earlier, there is no point in finding 4-regular graphs when $16 \leq n \leq 20$, thus Table A3 contains the 4-regular graphs for $21 \leq n \leq 24$ for which the diameter is 3. When the tables contain '$\geq \ldots$ graphs', that means we have not checked all the possible cases with minimal diameter, but in connection with decision making problems, it is enough to see that there is one pattern that satisfies the needed properties.

Finally, we can increase the regularity level to 5 in order to find out if we are able to get 2 as the smallest diameter for larger graphs. The answer is yes, actually it is also proven that $d = 2$ is reachable for 5-regular graphs until 24 vertices, but of course we are interested in the specific graphs that could help us determine the adequate comparison patterns. Our results can be found in Tables A4a and A4b. The $k = 5$ parameter is only possible when n is even again, so when it is odd, we let one vertex to have 6 as its degree.

The 5-quasi-regular graph on 21 vertices has been found by us as a twisted product $K_3 * X_7$, where X_7 is a graph with diameter 2 on 7 vertices, in which all vertices have degree 3, except one, where it has 2. The 5-regular graph on 22 vertices has been found by Pratt (2020) with the help of the following integer linear programming problem:

Let $N = \{1, \ldots, 22\}$ be the nodes, and let $P = \{i \in N, j \in N : i < j\}$ be the set of node pairs. For $(i, j) \in P$, let binary decision variable $X_{i,j}$ indicate whether (i, j) is an edge. For $(i, j) \in P$ and $k \in N \setminus \{i, j\}$, let binary decision variable $Y_{i,j,k}$ indicate whether k is a common neighbor of i and j. For $(i, j) \in P$ let binary decision variable $SLACK_{i,j}$ be a slack variable.

$$\min \sum_{(i,j)\in P} SLACK_{i,j} \tag{5}$$

$$\sum_{(i,j)\in P: k\in\{i,j\}} X_{i,j} = 5 \qquad \text{for } k \in N \tag{6}$$

$$X_{i,j} + \sum_{k\in N\setminus\{i,j\}} Y_{i,j,k} + SLACK_{i,j} \geq 1 \qquad \text{for } (i,j) \in P \tag{7}$$

$$Y_{i,j,k} \leq [i < k]X_{i,k} + [k < i]X_{k,i} \qquad \text{for } (i,j) \in P \text{ and } k \in N \setminus \{i,j\} \tag{8}$$

$$Y_{i,j,k} \leq [j < k]X_{j,k} + [k < j]X_{k,j} \qquad \text{for } (i,j) \in P \text{ and } k \in N \setminus \{i,j\} \tag{9}$$

Constraint **(2)** enforces 5-regularity. Constraint **(3)** enforces diameter 2. Constraints **(4)** and **(5)** enforce that $Y_{i,j,k} = 1$ implies k is a neighbor of i and j, respectively. A desired graph exists if and only if the integer linear program has a solution with $SLACK_{i,j} = 0$ for $\forall\ (i, j) \in P$.

The authors of this paper are still looking for a 5-quasi-regular graph on 23 vertices with diameter 2, but managed to find a graph, which has 23 vertices, and its diameter is 2, but it has one more edge than it should, namely three vertices have degree 6 and all the others have 5.

As we can see in Tables A4a and A4b there are higher completion ratios again, and for instance when we have 24 vertices, the decision makers should make 60 comparisons, which in certain situations can be too many. One can also note that in this table we report that there are some graphs with the needed properties, but never indicate the number of them. The reason behind this is simple: the very high number of the potential connected 5-regular graphs (for instance in the case of $n = 24$ there are roughly $2 \cdot 10^{22}$ possibilities).

This means that we have examined all the cases that we previously called interesting. According to our results, if we use the (n, k, d) parameters, then for smaller

MCDM problems the $k = 3$ is enough to get 2 as the diameter of the representing graph, which leads to a small completion ratio and a stable system of the comparisons. In larger problems, when we have more alternatives or criteria, we can choose if we use $k = 3$, when the completion ratio is smaller, but our approximation can be unstable, or choose higher level of regularity (and completion ratio) with more reliable results. We also showed examples and graphs with the needed properties for the different cases, which can help anyone in a MCDM problem to decide which comparisons have to be made. One can find the summary of our results in Table 3, which shows how many graphs we know for given (n, k, d) parameters. It is also true that if there is a graph for (n, k, d) in the table, then, on the one hand, no graph exists with the parameters $(n, k, d - 1)$, and, on the other hand, graphs for (n, k, D), where $D > d$, are not counted, and the corresponding cells are left empty. We omitted the cases when $k = 4$ and $n \leq 10$, because the minimal diameter is the same as it was in the case of $k = 3$. There is the same reasoning behind the emptiness of the table when $k = 5$ and $n \leq 15$. We have not included the cases when $k = 4$ and $16 \leq n \leq 20$, because $d = 3$ can be achieved by 3-regular graphs, but for $d = 2$ at least 5-regularity is needed. We also not included the $k = 3$ and $n \geq 20$ cases, because we were examining graphs with $d = 2$ and 3 only.
All the graphs in Tables 2, A1a, A1b, A2, A3, A4a, and A4b are given in several forms in Appendix B: graph, adjacency matrix (that directly shows which comparisons should be made, which PCM elements are required), list of edges and 'Graph6' format. The list of edges also present the needed comparisons, for instance the graph on 5 vertices in Figure B1 in Appendix B (see it also in Table 2) shows that the decision maker should fill in the following elements of the PCM: a_{12}, a_{13}, a_{14}, a_{15}, a_{23}, a_{24}, a_{35} and a_{45}. Upon request the other graphs of each family are available from the authors in these and other forms, as well.

	k		
n	**3**	**4**	**5**
5	1		
6	2		
7	4		
8	2		
9	2		
10	1		
11	134	37	
12	34	26	
13	353	10	
14	34	1	
15	290	1	
16	14		≥ 3
17	51		≥ 1
18	1		≥ 1
19	4		≥ 1
20	1		≥ 1
21		≥ 3	≥ 1
22		≥ 1	≥ 1
23		≥ 1	?
24		≥ 1	≥ 1

Table 3: The summary of the results: the number of k-(quasi-)regular graphs on n nodes with diameter d. Lightgray denotes $d = 2$ and gray denotes $d = 3$, '$\geq$' means that there are at least as many graphs as indicated, but we could not check all the possible cases.

2.4 Numerical example and simulations

The regularity of the representing graphs has been extensively studied in connection with incomplete pairwise comparisons' designs, while the diameter has only been investigated partially in the literature, as it was mentioned in the introduction. We would like to present what kind of problems can occur even with regular graphs, if we do not take into account the diameter, through a motivational example.

A wide range of simulations has also been performed in order to validate our recommendations, the applied methodology and the gained results are discussed in many details below. We would like to emphasize that, in this section we rely on the framework of the pairwise comparison matrices, though, our recommendations can be adopted in many other fields, as well.

2.4.1 Simulation methodology

It is important to see if the filling in pattern designs recommended by us are truly useful, thus we applied extensive simulations to have a better understanding of the problem. As for the calculation techniques of the weights derived from the PCMs, we used the well-known Logarithmic Least Squares Method (LLSM) and the Eigenvector Method based on the CR-minimal completion (CREV) (Bozóki et al., 2010). We applied two metrics to determine the differences from the weights calculated from the complete PCMs, that is the Euclidean distance (d_{euc}) and the maximum absolute distance (d_{max}, also known as Chebyshev distance), given by the following formulas:

$$d_{euc}(u, v) = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}$$

$$d_{max}(u, v) = \max_{i \in 1,\ldots,n} |u_i - v_i|,$$

where u denotes the weight vector calculated from a certain filling in design, while v is the weight vector calculated from the complete PCM. u and v are normalized by $\sum_{i=1}^{n} u_i = 1$ and $\sum_{i=1}^{n} v_i = 1$, respectively, while v_i and u_i denote the ith element of the appropriate vectors.

The process of the simulation for a given (n, k) pair consisted of the following steps:

1. We generated random $n \times n$ complete and consistent pairwise comparison matrices. The elements of these matrices were given as $a_{ij} = w_i/w_j$, where $w_i \in [1, 9]$ is a uniformly distributed random real number for $\forall i$.

2. Then we perturbed the elements of our consistent matrices three different ways, to get inconsistent PCMs with three distinguishable inconsistency levels. We call these levels weak, modest and strong given with the following formulas:

$$b_{ij} = \max\left(\frac{1}{2},\, a_{ij} + \Delta\right) \qquad \Delta \in [-1, 1] \qquad \text{(weak)}$$

$$b_{ij} = \max\left(\frac{1}{2},\, a_{ij} + \Delta\right) \qquad \Delta \in [-2, 2] \qquad \text{(modest)}$$

$$b_{ij} = \max\left(\frac{1}{3},\, a_{ij} + \Delta\right) \qquad \Delta \in [-3, 3] \qquad \text{(strong)}$$

   Where b_{ij} is the element of the perturbed matrix, a_{ij} is the element of the consistent matrix, $a_{ij} \geq 1$, and Δ is uniformly distributed in the given ranges. The motivation behind this structure is the following, we can get perturbed data even from an ordinal point of view, when $b_{ij} < 1$. However, in order to get meaningful results, we should use a different scale for the range of $(0, 1)$ compared to the range of $(1, 9]$ in connection with PCMs, as Figure 5 suggests. That is why the maximum function and the lower bounds ($1/2, 1/2$ and $1/3$, respectively) appear in the definition. These element-wise perturbation methods correlate with the well known Consistency Ratio (CR), as it is shown in Figure 6. We tested several combinations of parameters, and found that these, more or less balanced perturbations around 1, result in the most relevant levels of inconsistency.

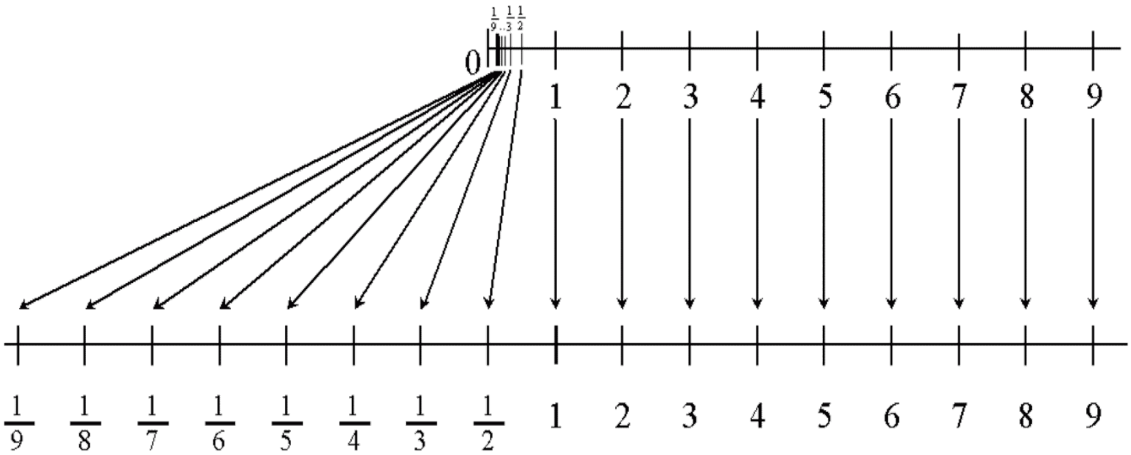


Figure 5: The scaling on different ranges

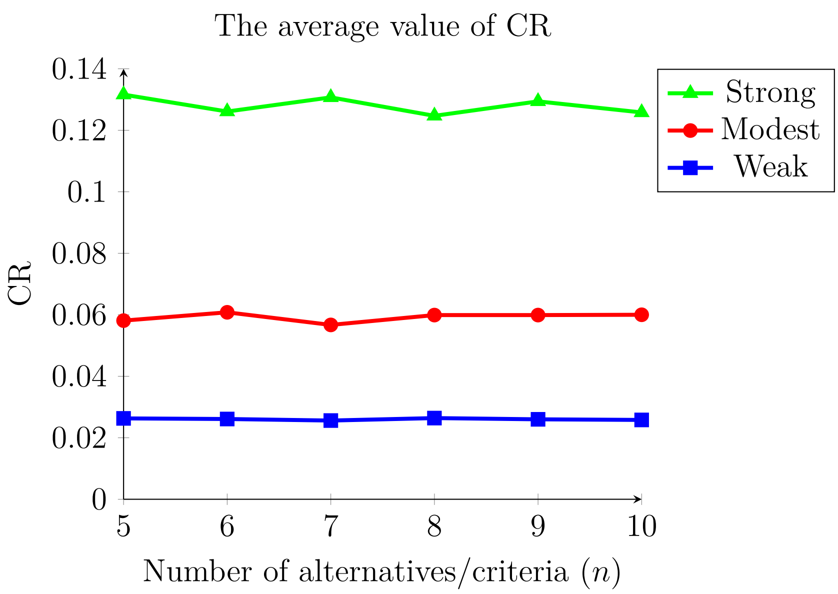


Figure 6: The connections between CR and our element-wise perturbations. Each point shows the average CR of 1000 randomly generated perturbed pairwise comparison matrices.

3. We deleted the respective elements of the matrices in order to get the filling in

pattern that we were examining, and applied the LLSM and the CREV techniques to get the weights. We always computed the certain designs' distances from the weights that we calculated from the complete inconsistent matrices. We used 1000 PCMs for every level of inconsistency and applied the following filling in patterns to compare them with each other:

(i) Our recommendations: k-(quasi-)regular graphs of minimal diameter, detailed in Section 2.3 and Appendix A

(ii) Random connected graphs with the same number of edges as our recommendation (1000 graphs per inconsistency level per simulation)

(iii) Connected k-(quasi-)regular graphs, but not of minimal diameter (1000 graphs per simulation)

(iv) Randomly generated, connected, of minimal diameter, but not regular graphs with the same number of edges (1000 graphs per simulation)

(v) Minimal diameter, modified/extended star graphs with the same number of edges (1000 graphs per inconsistency level per simulation)

4. Finally, we saved the mean and standard deviation of the distances for the different weight calculation methods and filling in designs.

We restricted the connected k-(quasi-)regular graphs to the Hamiltonian ones during the generation. With this we excluded the k-(quasi-)regular graphs with the largest diameters as well. This was also interesting, because all of our recommendations in Section 2.3 and Appendix A are Hamiltonian except the Petersen graph and the Tietze graph, but these two are well-known exceptions (Robinson and Wormald, 1994; Gould, 2003).

In case of (iv), we basically generated random connected graphs and selected the ones with minimal diameter (the same diameter as our recommendation), until we had 1000 such graphs, at least in the cases where we have found so many instances in a reasonably long time.

As for (v), when the diameter of our recommendation was 2, then we generated a random star graph, and complemented it with the needed number of random edges. While in case of diameter 3, we did the same, but at the end, we deleted one edge from the star and replaced it with another one, so that the diameter of the graph became 3.

It is important to note that we considered only the graph presented in Section 2.3 and Appendix A for a given (n, k) pair in (i), and not all the k-(quasi-)regular graphs with minimal diameter. This is due to the fact that in many cases we were able to find one graph with the needed properties, but could not find all of them or even could not determine the exact number of such graphs.

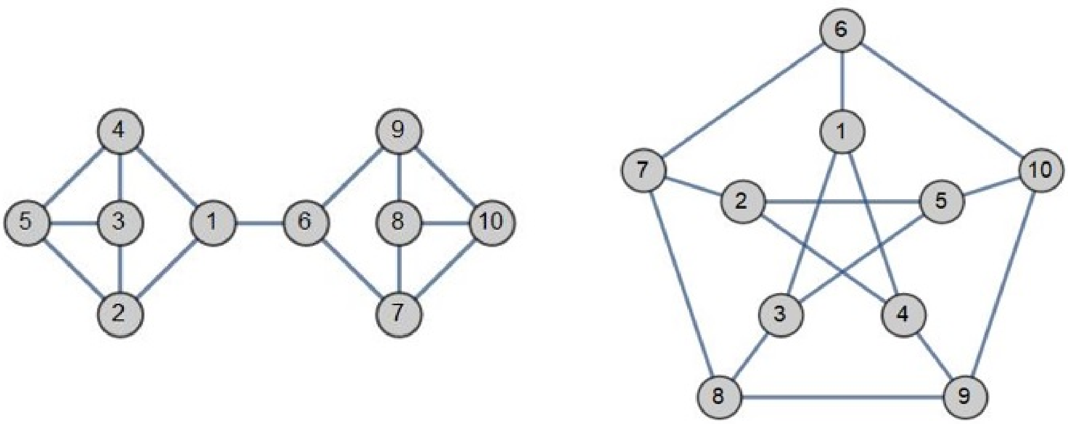
Before the results of the simulations, we show a motivational example, in which we compare two different filling in pattern designs similarly as in the case of the simulations. This numerical instance shows that it is also important to take into account the minimal diameter property, and not just regularity.

2.4.2 Motivational example

Let us demonstrate the simulation process, as well as the importance of the diameter, when we have 10 alternatives, and we examine only two different filling in structures.

We generate 1000 $n \times n$ consistent PCMs with elements $a_{ij} = w_i/w_j$, where $w_i, w_j \in [1, 9]$ are uniformly distributed random real numbers. Then we perturb all of the elements of these PCMs three different ways as described in Equations weak, modest and strong.

We would like to compare the differences of the calculated weights from the ones that we get from these complete perturbed PCMs, when we consider the two filling in patterns represented by the graphs in Figure 7. The filling structures related to these graphs can be seen in Table 4, which means that we delete all the other elements, when we compute the weights according to the given pattern.



(a) An alternative 3-regular graph (b) Our recommendation: the Petersen graph

Figure 7: The graph representation of two 3-regular designs

As for the two representing graphs, the Petersen graph has minimal diameter among 3-regular graphs on 10 vertices, its diameter is 2, while the Alternative 3-regular graph's diameter is 5. As one can see there are common elements of the two filling in patterns, as for instance the bridge-edge between vertices 1 and 6 (a_{16}, bridge set (Csató and Tóth, 2020)), which connects the two symmetric components

	1	2	3	4	5	6	7	8	9	10
1	1	b_{12}		b_{14}		b_{16}				
2	$\frac{1}{b_{12}}$	1	b_{23}		b_{25}					
3		$\frac{1}{b_{23}}$	1	b_{34}	b_{35}					
4	$\frac{1}{b_{14}}$		$\frac{1}{b_{34}}$	1	b_{45}					
5		$\frac{1}{b_{25}}$	$\frac{1}{b_{35}}$	$\frac{1}{b_{45}}$	1					
6	$\frac{1}{b_{16}}$					1	b_{67}		b_{69}	
7						$\frac{1}{b_{67}}$	1	b_{78}		b_{710}
8							$\frac{1}{b_{78}}$	1	b_{89}	b_{810}
9						$\frac{1}{b_{69}}$		$\frac{1}{b_{89}}$	1	b_{910}
10							$\frac{1}{b_{710}}$	$\frac{1}{b_{810}}$	$\frac{1}{b_{910}}$	1

	1	2	3	4	5	6	7	8	9	10
1	1		b_{13}	b_{14}		b_{16}				
2		1		b_{24}	b_{25}		b_{27}			
3	$\frac{1}{b_{13}}$		1		b_{35}			b_{38}		
4	$\frac{1}{b_{14}}$	$\frac{1}{b_{24}}$		1					b_{49}	
5		$\frac{1}{b_{25}}$	$\frac{1}{b_{35}}$		1					b_{510}
6	$\frac{1}{b_{16}}$					1	b_{67}			b_{610}
7		$\frac{1}{b_{27}}$				$\frac{1}{b_{67}}$	1	b_{78}		
8			$\frac{1}{b_{38}}$				$\frac{1}{b_{78}}$	1	b_{89}	
9				$\frac{1}{b_{49}}$				$\frac{1}{b_{89}}$	1	b_{910}
10					$\frac{1}{b_{510}}$	$\frac{1}{b_{610}}$			$\frac{1}{b_{910}}$	1

Table 4: The known elements of the given PCM in case of the two different filling in patterns represented by the graphs in Figure 7. The design related to the Alternative graph can be seen to the left, while the filling structure of the Petersen graph is shown in the PCM to the right.

of the Alternative graph. It is also worth to mention that the special structure of this graph (also highlighted by the two separate parts of the related PCM in Table 4) ensures that the weights of 1 and 6 are always determined exactly by b_{16}.

Table 5 summarizes the mean (denoted by M) and the standard deviation (σ) of distances (d_{euc} and d_{max}) of the weights calculated from the two filling patterns respect to the complete case for the three inconsistency (perturbation) levels (Weak, Modest and Strong).

Weak	**LLSM d_{euc} M**	**CREV d_{euc} M**	**LLSM d_{max} M**	**CREV d_{max} M**	**LLSM d_{euc} σ**	**CREV d_{euc} σ**	**LLSM d_{max} σ**	**CREV d_{max} σ**
Petersen	0.0424	0.0422	0.0275	0.0274	0.0285	0.0283	0.0193	0.0191
Alternative	0.0605	0.0604	0.0370	0.0369	0.0468	0.0467	0.0286	0.0285
Modest								
Petersen	0.0673	0.0669	0.0450	0.0445	0.0378	0.0376	0.0278	0.0274
Alternative	0.0956	0.0956	0.0604	0.0602	0.0610	0.0611	0.0400	0.0399
Strong								
Petersen	0.0967	0.0952	0.0665	0.0652	0.0527	0.0519	0.0402	0.0390
Alternative	0.1318	0.1314	0.0881	0.0877	0.0825	0.0826	0.0590	0.0592

Table 5: The average distances and their standard deviation for the different designs. The following notations are used: M-mean, σ-standard deviation, 'Weak', 'Modest' and 'Strong' refer to the level of perturbation.

One can see that there are significant contrasts between the outcomes of the examined filling in patterns. In case of both the Euclidean and maximum absolute (Chebyshev) metrics, the distances of the weights computed from the Alternative graph respect to the ones we got from the complete PCM are approximately 1.5 times larger, than the same for the Petersen graph, however both the relative and absolute differences are smaller in case of the absolute maximum distance. The same results are true when we consider the standard deviation of the distances. This means that the Petersen graph tends to provide small errors and a consistent performance (small standard deviation) depending on the perturbations, compared to the filling pattern represented by the Alternative graph.

We think that this example can give a deeper understanding of the simulation method. Besides that, the main message of this sub-section is that, one should consider the diameter of the graph as an important parameter in these designs, because even among regular graphs, there can be large differences.

2.4.3 Simulation results

The results of the simulations seem to mainly depend on the value of k, and barely on n, as well as the patterns of the outcomes seem to be the same for every case.

The tables for all parameters (n, k, d) calculated are available in Appendix D, while we have chosen to visualize only the following representative examples: $(n = 16, k = 3, d = 3)$, $(n = 11, k = 4, d = 2)$ and $(n = 24, k = 5, d = 2)$. The first one is the largest 3-regular case, where we could apply (iv), and it is the only one that can be found in the main text due to the length of the figures. The second one is the smallest 4-regular, and the last one is the largest 5-regular case that we examined. The results of the simulations for them are shown in Figures 8, C1 and C2, respectively, and it is also recommended to read this section together with Appendix C, as the latter two cases are presented there. The figures show the mean of the different metrics (M) and the standard deviation (σ) as well. We refer to the different levels of the perturbation as 'Weak', 'Modest' and 'Strong', as before.

It is clear from the outcomes of the simulations that the stronger perturbation causes larger distances, and the higher regularity level leads to smaller differences. As one can see, our recommendations have the smallest means and standard deviations among the different designs in case of both metrics and both weight calculation methods for every (n, k) pair, which suggests that the results are not solely dependent on the used techniques and parameters. The smallest mean shows that the k-(quasi-)regular graphs with minimal diameter provide the closest weights to the complete PCM on an average level. On the other hand, the smallest standard deviation also implies that our recommendations are commonly not connected to huge errors, and that these filling in pattern designs perform at a very consistent level regarding the deviations from the results of the complete PCMs. It is also true that the randomly generated minimal diameter graphs (denoted by (iv)) tend to have

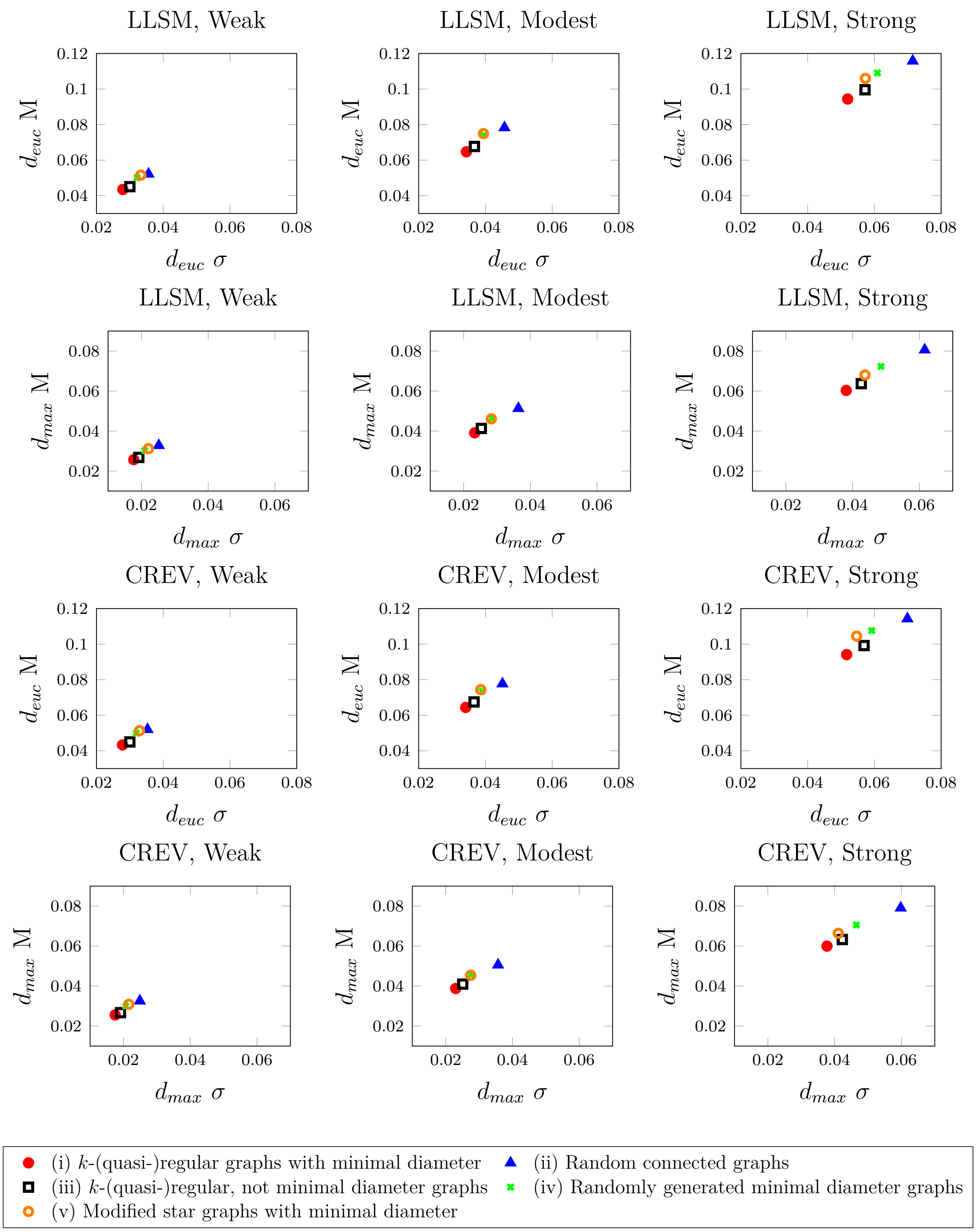


Figure 8: The results of the simulation for $(n = 16, k = 3, d = 3)$
The following notations are used: M–mean, σ–standard deviation, d_{euc}–Euclidean distance, d_{max}–maximum absolute distance, ‘Weak’, ‘Modest’ and ‘Strong’ refer to the level of perturbation. See Table D13 in Appendix D for numerical details.

smaller means and standard deviations compared to the simple random graphs. Again, this suggests that, the diameter of the representing graph is relevant. The k-(quasi-)regular graphs (denoted by (iii)) always have the second smallest means and standard deviations in their distances, thus the already known fact, that regularity is a key property, confirmed here as well. It is also important to note that we have excluded the k-(quasi-)regular cases with the largest diameters, because of the Hamiltonian construction as we mentioned earlier, thus we expect random (quasi-)regular graphs to have a bit even 'worse' results compared to our recommendations. The case of the modified star graphs (denoted by (v)) is interesting. In case of $k = 3$, they always have smaller means and standard deviations compared to the simple random graphs, but for $k = 4$ they always have larger means, and in some cases even their standard deviations are higher. For $k = 5$ the modified star graphs tend to have the largest means and standard deviations among the examined designs. This also suggests that considering only the diameter is not sufficient in these problems. Finally, we would like to emphasize that these patterns and findings, are the very same for all studied (n, k) pairs, especially regarding the dominance of the k-(quasi-)regular graphs, thus our recommendations seem to perform indeed well in the framework of pairwise comparison matrices.

2.5 Conclusions and further research

2.5.1 Summary

The main contribution of the paper is a systematic collection of recommended filling patterns of incomplete pairwise comparisons' using the graph representation of the PCMs. The proposed (quasi)-regular graphs with minimal diameter have not only pure graph theoretical relevance, but their importance in multicriteria decision making is also demonstrated via the comparisons to other incomplete filling in patterns of the same cardinality.

Graphs are included in several formats in Appendix B, which can show practitioners the comparisons that should be made, i.e. the PCM elements to be filled in. We presented our results using the number n of criteria or alternatives, regularity level k and diameter d of the representing graph as parameters. We identified the diameter, that was missing from the relevant literature of decision theory and preference modelling, as an important parameter in these problems. It has been shown that relatively small diameters $d = 2, 3$ can be achieved with relatively small completion ratios, and examples has been provided for every case up until 24 vertices.

We also validated our recommendations with the help of numerical simulations. 1000 perturbed PCMs were used in case of 3 different inconsistency (perturbation) levels to compare several filling in patterns with the proposed ones for every examined parameter combinations (all in all 34). The recommended filling structures provided the closest weight vectors to the complete case on average, with the smallest standard deviation, according to 2 distances (Euclidean and Chebyshev), in case of both the incomplete LLSM and Eigenvector weight calculation techniques (for detailed results of all parameter combinations, see Appendix D). Examples also show that neither regularity nor small diameter is sufficient on its own, both of these properties are needed.

2.5.2 Limitations and further research

Simulations show that the proposed (quasi)-regular graphs with minimal diameter are better, in the sense of the metrics we considered, than e.g., the random ones, or the ones having only one of the two properties, regularity and minimal diameter, instead of both. However, it certainly does not mean that other, yet undiscovered or unidentified structures could not be even better.

The investigation of the robustness of the results, namely what is 'between' the different regularity levels (when the degrees of different vertices are not the same), could be the topic of a further research, as well as the cases with larger minimal diameters. Similarly, what is between diameters $d+1$ and d, in particular 2 and 1 (i.e. the complete graph)? According to Tables 2, A2, A4a and A4b, diameter 2 is achieved at relatively low completion ratios, especially for larger n parameters, so the *game* of having better weight vectors by adding more comparisons is continuing rather than ending at $d = 2$, as the values in Tables D2–D7, D18–D22 and D27–D35 show.

It is also an interesting problem to concentrate directly on the completion ratio as a parameter instead of the regularity of the representing graph. If the (n, c) pair is given (and $(n-1)c$, the average degree is not necessarily integer), then which comparisons should be made?

Our approach definitely has a strong connection with other metrics based on the lengths of shortest paths (e.g. their average) as well as centrality measures (Chebotarev and Gubanov, 2020). When there are several graphs with the needed properties, we can reduce their number based on some chosen centrality measures. We would like to deal with these questions in our future works.

Group decision making (Oliva et al., 2019) is a potential application area of our results, as we may assume that the individual preferences can be colorful enough, so we cannot suppose any prior information. In other words: we treat the items to be compared in a symmetric way, therefore our recommended graphs can be applied.

Although our results were presented within the framework of pairwise comparison matrices, they are applicable in a wider range. A lot of other models based on pairwise comparisons can utilize our findings. For example ranking of sport players or teams based on their matches leads to the problem of tournament design: which pairs should play against each other (without the use of prior knowledge or estimation of their strength)?

Acknowledgments

The authors thank the valuable comments and suggestions of the anonymous Reviewers. The comments of János Fülöp, László Csató, Gabriele Oliva, Michele Fedrizzi, Matteo Brunelli and Konrad Kułakowski are greatly acknowledged. Special thanks to Robert W. Pratt for his help in finding a 5-regular graph on 22 vertices and searching for a 5-quasi-regular graph on 23 vertices, with diameter two. The research of S. Bozóki and Zs. Szádoczki was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) under Grant NKFIA ED 18-2-2018-0006. Zs. Szádoczki was supported by the ÚNKP-21-3-II-CORVINUS-19 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.1016/j.omega.2021.102557.

References

Bermond, J., Delorme, C., and Farhi, G. (1982). Large graphs with given degree and diameter III. In Bollobás, B., editor, *Graph Theory*, volume 62 of *North-Holland Mathematics Studies*, pages 23–31. North-Holland. https://doi.org/10.1016/S0304-0208(08)73544-8.

Biró, P., Fleiner, T., and Palincza, R. P. (2017). Designing chess pairing mechanisms. In Frank, A., Recski, A., and Wiener, G., editors, *Proceedings of the 10th Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications*, pages 77–86. http://real.mtak.hu/80729/7/jXaio4T11ygd57-77-86.pdf.

Blanquero, R., Carrizosa, E., and Conde, E. (2006). Inferring efficient weights from pairwise comparison matrices. *Mathematical Methods of Operations Research*, 64(2):271–284. https://doi.org/10.1007/s00186-006-0077-1.

Bozóki, S., Csató, L., and Temesi, J. (2016). An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research*, 248(1):211–218. https://doi.org/10.1016/j.ejor.2015.06.069.

Bozóki, S., Fülöp, J., and Rónyai, L. (2010). On optimal completion of incomplete pairwise comparison matrices. *Mathematical and Computer Modelling*, 52(1):318–333. https://doi.org/10.1016/j.mcm.2010.02.047.

Branković, L., Miller, M., Plesník, J., Ryan, J., and Širaň, J. (1998). Large graphs with small degree and diameter: A voltage assignment approach. *The Australasian Journal of Combinatorics*, 18:65–76. https://ajc.maths.uq.edu.au/pdf/18/ocr-ajc-v18-p65.pdf.

Chatburn, R. (2013). The whisper game. *Respiratory Care*, 58(11):paper 157. http://rc.rcjournal.com/content/respcare/58/11/e157.full.pdf.

Chebotarev, P. and Gubanov, D. (2020). How to choose the most appropriate centrality measure? https://arxiv.org/abs/2003.01052.

Ciomek, K., Kadziński, M., and Tervonen, T. (2017). Heuristics for selecting pairwise elicitation questions in multiple criteria choice problems. *European Journal of Operational Research*, 262(2):693–707. https://doi.org/10.1016/j.ejor.2017.04.021.

Comellas, F. and Gómez, J. (1994). New large graphs with given degree and diameter. https://arxiv.org/abs/math/9411218.

Csató, L. (2013). Ranking by pairwise comparisons for Swiss-system tournaments. *Central European Journal of Operations Research*, 21(4):783–803. https://doi.org/10.1007/s10100-012-0261-8.

Csató, L. (2015). A graph interpretation of the least squares ranking method. *Social Choice and Welfare*, 44(1):51–69. https://doi.org/10.1007/s00355-014-0820-0.

Csató, L. (2017). On the ranking of a Swiss system chess team tournament. *Annals of Operations Research*, 254(1-2):17–36. https://doi.org/10.1007/s10479-017-2440-4.

Csató, L. and Tóth, Cs. (2020). University rankings from the revealed preferences of the applicants. *European Journal of Operational Research*, 286(1):309–320. https://doi.org/10.1016/j.ejor.2020.03.008.

Davidson, R. and Farquhar, P. (1976). A bibliography on the method of paired comparisons. *Biometrics*, 32(2):241–252. https://www.jstor.org/stable/2529495.

Dinneen, M. J. and Hafner, P. R. (1994). New results for the degree/diameter problem. *Networks*, 24(7):359–367. https://doi.org/10.1002/net.3230240702.

Elspas, B. (1964). Topological constraints on interconnection-limited logic. *Proceedings. 5th Annual IEEE Symposium on Switching Circuit Theory and Logical Design, Princeton, New Jersey, USA*, pages 133–137. https://doi.org/10.1109/SWCT.1964.27.

Fedrizzi, M. and Giove, S. (2013). Optimal sequencing in incomplete pairwise comparisons for large dimensional problems. *International Journal of General Systems*, 42(4):366–375. https://doi.org/10.1080/03081079.2012.755523.

Fishburn, P. (1970). Intransitive indifference in preference theory: A survey. *Operations Research*, 18(2):207–228. https://www.jstor.org/stable/168680.

Gass, S. (1998). Tournaments, transitivity and pairwise comparison matrices. *Journal of the Operational Research Society*, 49(6):616–624. https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2600572.

Glickman, M. E. and Jensen, S. T. (2005). Adaptive paired comparison design. *Journal of Statistical Planning and Inference*, 127(1-2):279–293. https://doi.org/10.1016/j.jspi.2003.09.022.

Gould, R. J. (2003). Advances on the Hamiltonian problem – A Survey. *Graphs and Combinatorics*, 19:7–52. https://doi.org/10.1007/s00373-002-0492-x.

Hoffman, A. J. and Singleton, R. R. (1960). On Moore graphs with diameters 2 and 3. *IBM Journal of Research and Development*, 4:497–504. https://doi.org/10.1147/rd.45.0497.

Holton, D. A. and Sheehan, J. (1993). *The Petersen Graph*. Australian Mathematical Society Lecture Series. Cambridge University Press. https://doi.org/10.1017/CBO9780511662058.

Horvát, S. (2020). IGraph/M. An immediately usable version of this software is accessible from its GitHub repository. https://doi.org/10.5281/zenodo.3739056.

Kendall, M. G. and Babington Smith, B. (1940). On the method of paired comparisons. *Biometrika*, 31(3/4):324–345. https://doi.org/10.2307/2332613.

Kujansuu, E., Lindberg, T., and Mäkinen, E. (1999). The stable roommates problem and chess tournament pairings. *Divulgaciones Matemáticas*, 7(1):19–28. https://www.emis.de/journals/DM/v71/art3.pdf.

Kułakowski, K., Szybowski, J., and Prusak, A. (2019). Towards quantification of incompleteness in the pairwise comparisons methods. *International Journal of Approximate Reasoning*, 115:221–234. https://doi.org/10.1016/j.ijar.2019.10.002.

Loz, E. and Širáň, J. (2008). New record graphs in the degree-diameter problem. *The Australasian Journal of Combinatorics*, 41:63–80. http://ajc.maths.uq.edu.au/pdf/41/ajc_v41_p063.pdf.

Luce, R. (1956). Semiorders and a theory of utility. *Econometrica*, 24(2):178–191. https://www.jstor.org/stable/pdf/1905751.pdf.

Maharry, J. and Robertson, N. (2016). The structure of graphs not topologically containing the Wagner graph. *Journal of Combinatorial Theory, Series B*, 121:398 – 420. https://doi.org/10.1016/j.jctb.2016.07.011.

McCormick, E. and Bachus, J. (1952). Paired comparison ratings: 1. The effect on ratings of reductions in the number of pairs. *Journal of Applied Psychology*, 36(3):123–127. https://doi.org/10.1037/h0054842.

McCormick, E. and Roberts, W. (1952). Paired comparison ratings: 2. The reliability of ratings based on partial pairings. *Journal of Applied Psychology*, 36(3):188–192. https://doi.org/10.1037/h0055956.

McKay, B. D. and Piperno, A. (2014). Practical graph isomorphism, II. *Journal of Symbolic Computation*, 60(0):94–112. https://doi.org/10.1016/j.jsc.2013.09.003.

Miller, M. and Širáň, J. (2013). Moore graphs and beyond: A survey of the degree/diameter problem. *Electronic Journal of Combinatorics*, 20(2):1–92. https://doi.org/10.37236/35.

Miyake, C., Harima, S., Osawa, K., and Shinohara, M. (2003). 2-cyclic design in AHP. *Journal of the Operations Research Society of Japan*, 46(4):429–447. https://doi.org/10.15807/jorsj.46.429.

Ólafsson, S. (1990). Weighted matching in chess tournaments. *Journal of the Operational Research Society*, 41(1):17–24. https://doi.org/10.2307/2582935.

Oliva, G., Scala, A., Setola, R., and Dell'Olmo, P. (2019). Opinion-based optimal group formation. *Omega*, 89:164–176. https://doi.org/10.1016/j.omega.2018.10.008.

Pananjady, A., Mao, C., Muthukumar, V., Wainwright, M., and Courtade, T. (2020). Worst-case versus average-case design for estimation from partial pairwise comparisons. *Annals of Statistics*, 48(2):1072–1097. https://doi.org/10.1214/19-AOS1838.

Pratt, R. W. (1996). The complete catalog of 3-regular, diameter-3 planar graphs. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.9058&rep=rep1&type=pdf.

Pratt, R. W. (2020). Personal communication. https://math.stackexchange.com/questions/3745954/how-to-construct-a-5-regular-graph-with-diameter-2-on-22-vertices.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, 53:49–57. https://doi.org/10.1016/j.omega.2014.11.009.

Ribeiro, M., Gligoric, K., and West, R. (2019). Message distortion in information cascades. In *Proceedings of The World Wide Web Conference - WWW'19*, page 681–692. http://doi.org/10.1145/3308558.3313531.

Robinson, R. W. and Wormald, N. C. (1994). Almost all regular graphs are Hamiltonian. *Random Structures & Algorithms*, 5(2):363–374. https://doi.org/10.1002/rsa.3240050209.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill, New York.

Storwick, R. M. (1970). Improved construction techniques for (d, k) graphs. *IEEE Transactions on Computers*, C-19(12):1214–1216. https://doi.org/10.1109/T-C.1970.222861.

Tekile, H. A. (2017). Incomplete pairwise comparison matrices in multi-criteria decision making and ranking. Master's thesis, Central European University. https://mathematics.ceu.edu/sites/mathematics.ceu.hu/files/attachment/basicpage/29/thesishailemariam.pdf.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286. https://doi.org/10.1037/h0070288.

Wang, K. and Takahashi, I. (1998). How to select paired comparisons in AHP of incomplete information – strongly regular graph design. *Journal of the Operations Research Society of Japan*, 41(2):311–328. https://doi.org/10.15807/jorsj.41.311.

Wolfram Research, I. (2021). Mathematica, Version 12.3. Champaign, IL, 2021. https://www.wolfram.com/mathematica.

Zahedi, F. (1986). The analytic hierarchy process: A survey of the method and its applications. *Interfaces*, 16(4):96–108. https://doi.org/10.1287/inte.16.4.96.

3 Study II. Incomplete pairwise comparison matrices based on graphs with average degree approximately 3

Authors: Zsombor Szádoczki, Sándor Bozóki, Patrik Juhász, Sergii V. Kadenko, Vitaliy Tsyganok

Published in Annals of Operations Research, 326(2): 783-807. (2023)

https://doi.org/10.1007/s10479-022-04819-9

Abstract

A crucial, both from theoretical and practical points of view, problem in preference modelling is the number of questions to ask from the decision maker. We focus on incomplete pairwise comparison matrices based on graphs whose average degree is approximately 3 (or a bit more), i.e., each item is compared to three others in average. In the range of matrix sizes we considered, $n = 5, 6, 7, 8, 9, 10$, this requires from $1.4n$ to $1.8n$ edges, resulting in completion ratios between 33% ($n = 10$) and 80% ($n = 5$). We analyze several types of union of two spanning trees (three of them building on additional ordinal information on the ranking), 2-edge-connected random graphs and 3-(quasi-)regular graphs with minimal diameter (the length of the maximal shortest path between any two vertices). The weight vectors are calculated from the natural extensions, to the incomplete case, of the two most popular weighting methods, the eigenvector method and the logarithmic least squares. These weight vectors are compared to the ones calculated from the complete matrix, and their distances (Euclidean, Chebyshev and Manhattan), rank correlations (Kendall and Spearman) and similarity (Garuti, cosine and dice indices) are computed in order to have cardinal, ordinal and proximity views during the comparisons. Surprisingly enough, only the union of two star graphs centered at the best and the second best items perform well among the graphs using additional ordinal information on the ranking. The union of two edge-disjoint spanning trees is almost always the best among the analyzed graphs.

Keywords: Pairwise comparison, Incomplete pairwise comparison matrix, Graph of comparisons, Filling in pattern

3.1 Introduction

Given n items (in multi-attribute decision making, typically criteria, alternatives, voting powers of decision makers, subjective probabilities, levels of performance with respect to a fixed criterion etc.), the structure of pairwise comparisons is often represented by graphs (Gass, 1998). The minimally sufficient number of comparisons in order to have a connected system of preferences is $n - 1$, and the pairs of items compared can be associated to the edges of a spanning tree on n nodes. This system has no redundance at all, and the calculated weight vector is highly sensitive to the change of any comparison. Observe that the average degree in a spanning tree is $(2n - 2)/n = 2 - 2/n$, i.e., every item is compared to (almost) 2 other items in average. We would like to keep the number of comparisons low, namely around the average degree 3 or a bit more, and compare the weight vectors calculated from several such graphs to determine the best filling in pattern that provides the closest weights to the ones computed from the complete matrix.

In our research predetermined graphs are used, thus we assume that the set of comparisons can be chosen, and the whole questionnaire should be prepared a priori.

Some of the examined models also use additional ordinal information, thus the evaluation of this information is a main contribution of our study as well.

Our aim is to gain as much information from the decision maker's revealed preferences as possible. We would like to find out if there is better ordinal information than the one usually used by multi-attribute decision making models, and whether we can find better solutions even without additional ordinal information. The paper deals with specified numbers of comparisons, however, the key question remains valid for all incomplete pairwise comparison matrices.

The remainder of this paper is organized as follows. Section 3.2 provides a brief summary of the related literature and research gaps. The used methodology is detailed in Section 3.3, where besides the method of pairwise comparisons, as well as the relevant graph theoretical concepts (Section 3.3.1), the applied simulation process to compare the different models is also included (Section 3.3.2). The analyzed models and the associated filling in patterns are presented in Section 3.4, while Section 3.5 shows the obtained results. Finally, Section 3.6 concludes and raises some further research questions.

3.2 Literature review

The main goal of multi-attribute decision making (MADM) is to determine the best or the best few, perhaps the complete ranking of the discrete number of alternatives based on a finite number of (usually conflicting) qualitative and/or quantitative criteria (Triantaphyllou, 2000a).

It is not uncommon that the MADM models use some part of the ordinal information of the ranking of items from the decision maker at the beginning of the process as well, namely the best, the worst or both alternatives. The most popular such methodologies are the SMART (simple multi-attribute rating technique) (Edwards, 1977), the Swing method (von Winterfeldt and Edwards, 1986), the SMARTS (SMART using Swings) and the SMARTER (SMART Exploiting Ranks) (Edwards and Barron, 1994; Mustajoki et al., 2005), and last but not least the best-worst method (Rezaei, 2015). The latter one generated an indeed large literature in the last few years (Mi et al., 2019), with theoretical extensions and studies (Liang et al., 2020; Mohammadi and Rezaei, 2020) as well as real applications (Rezaei et al., 2016). However, the significance of the ordinal information used by these models has barely been studied.

One of the most fundamental concepts of MADM is the method of the pairwise comparisons (Thurstone, 1927). It is also the cornerstone of the indeed popular and widely used methodology of the Analytic Hierarchy Process (AHP) (Saaty, 1977), where these comparisons form a pairwise comparison matrix (PCM).

The incomplete case of PCMs was originally proposed by (Harker, 1987) to reduce the number of questions in the AHP, especially in group decision making. Since that besides their application for different problems (Bozóki et al., 2016), many aspects of incomplete PCMs has been examined in detail from the inconsistency measures (Szybowski et al., 2020) and their thresholds (Ágoston and Csató, 2022) to different optimal completions (Zhou et al., 2018; Fedrizzi and Giove, 2007). As answering all the questions is time-consuming, there were several proposal to solve this problem (Triantaphyllou, 2000b). Based on Revilla and Ochoa (2017) if we ask the respondents, they clearly prefer shorter (with a maximum length of 20 minutes) questionnaires and surveys, while longer questionnaires result in lower data quality as well (Deutskens et al., 2004). These time spans more or less correspond to the number of questions included in our models. We focus on a similar problem as the reduction of comparisons, however, we assume that the whole questionnaire has to be prepared before the decision making process and we cannot ask the decision makers to modify some answers, thus our approach is in some sense similar to Amenta et al. (2021). There are only a few research dealing with the question that which pattern of comparisons should we use in a given problem (Szádoczki et al., 2022), and to our knowledge, none of them regards additional ordinal information as well, thus this paper would like to fill in this research gap. The comparison of different priority vectors has an important role in our study, which has been also used in the different optimizations of aggregation of group preferences (Duleba et al., 2021) and in several weight calculation problems (Kou and Lin, 2014).

3.3 Methodology

3.3.1 Pairwise comparisons and their graph representation

The pairwise comparison matrix technique can be applied in decision problems both to determine the weights of the different criteria and to evaluate the alternatives according to a criterion.

Definition 18 (Pairwise comparison matrix (PCM)) Let us denote the number of criteria (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).

The general element of a PCM a_{ij} shows how many times item i is better/larger/stronger/more important than item j.

There are several techniques to calculate a weight vector from a PCM that shows the importance of compared items. Probably the two most commonly used methods are the logarithmic least squares (LLSM) (Crawford and Williams, 1985) and the eigenvector (Saaty, 1977) techniques, given by the following formulas respectively:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\ln(a_{ij}) - \ln\left(\frac{w_i}{w_j}\right)\right)^2 \rightarrow \min \tag{10}$$

$$A \cdot w = \lambda_{\max} \cdot w \tag{11}$$

Where w denotes the weight vector with the general element w_i, A is an $n \times n$ PCM and λ_{max} is the maximal eigenvalue of A. As we are studying graphs in our research, we should also mention that the combinatorial weight calculation method (Tsyganok, 2000), which is based on the weight vectors gained from different spanning trees provides the same solution as the LLSM if we use the geometric mean (Lundy et al., 2017). The comparison of this method with other weight calculation techniques can be found in Tsyganok (2010).

Another important aspect of PCMs is their inconsistency.

Definition 19 (Consistent PCM) A PCM is said to be consistent if and only if $a_{ik} = a_{ij}a_{jk}\ \forall i, j, k$. If a PCM is not consistent, then it is called inconsistent.

Of course, there are different levels of inconsistency and several inconsistency indices have been proposed to measure this problem (Brunelli et al., 2013), which satisfy different properties (Brunelli, 2017). However, the most popular one is probably still the Consistency Ratio (CR) (Saaty, 1977). In order to define the CR, we need the Consistency Index (CI), which is given as

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{12}$$

Where λ_{max} is the largest eigenvalue of the PCM. The Consistency Ratio can be calculated as

$$CR = \frac{CI}{RI} \tag{13}$$

Where RI is the Random Index, which is the average CI obtained from a large enough set of randomly generated matrices of size n.

In case of incomplete data, namely when some elements of the PCM are missing, we are talking about an incomplete PCM. The absence of these elements can be caused by several different reasons: the loss of data, some comparisons are simply not possible (Bozóki et al., 2016) or the decision maker does not have time, possibility or willingness to fill in all the $n(n-1)/2$ comparisons (it is sufficient to focus only on the elements above the principal diagonal of the matrix because of the reciprocity).

In our research the most important case is the latter one, as we examine different kinds of filling in patterns of incomplete PCMs, thus we assume that the set of pairwise comparisons to be made can be chosen.

The weight calculation techniques can be extended to the incomplete case, but the results depend on both the number of known comparisons and their positioning. We assume that it is possible to find a pattern of pairwise comparisons, that minimizes experts' efforts and accumulated estimation errors, while ensuring estimation stability. In order to compare the effect of different arrangements of known elements (filling in patterns), we use the graph representation of the PCM (Gass, 1998). The representing graph is an undirected graph, where the vertices denote the criteria/alternatives, and there is an edge between any two vertices if and only if the comparison has been made for the two respective items (the associated element of the PCM is known).

There are several different attributes that are studied in connection with decision making problems, among which regularity of comparisons (in some sense) is, indeed, an important one (Wang and Takahashi, 1998; Kułakowski et al., 2019). If we use the graph representation of the PCM, the most common definition of regularity that we use in this paper, can be seen below.

Definition 20 (k-regularity) A graph is called k-regular if every vertex has k neighbours, which means that the degree of every vertex is k.

In decision problems regularity ensures a certain level of symmetry, it means that every item is compared to the same number of elements. This kind of property is also required in the design of some sport tournaments (Csató, 2017).

As k-regularity is not possible, when both the number of vertices (n) and the level of regularity (k) are odd, Szádoczki et al. (2022) defined the graphs that are the closest to k-regularity in this case, as follows.

Definition 21 (k-quasi-regularity) A graph is called k-quasi-regular if exactly one vertex has degree $k+1$, and all the other vertices have degree k.

In case of indirect relations, when there is no direct comparison between two elements, the small errors of the intermediate comparisons can add up (Szádoczki et al., 2020). The diameter of the representing graph can measure this problem indeed naturally.

Definition 22 (Diameter of a graph) The d diameter of a graph G is the length of the longest shortest path between any two vertices:

$$d = \max_{u,v \in V(G)} \ell(u,v), \tag{14}$$

where $V(G)$ denotes the set of vertices of G and $\ell(.,.)$ is the graph distance between two vertices, namely the length of the shortest path between them (in our case the length of every edge is one).

The diameter also seems to be a crucial property in case of the weight calculation method based on spanning trees (Kadenko and Tsyganok, 2020).

It is important to note that the solution of the weight calculation problem is unique in case of incomplete PCMs if and only if the representing graph is connected, thus there are at least indirect comparisons between the pairs of items (Bozóki et al., 2010).

Definition 23 (Connected graph) In an undirected graph, two vertices u and v are called connected if the graph contains a path from u to v. A graph is said to be connected if every pair of vertices in the graph is connected.

It is worth to examine some of the so-called stronger kind of connectedness measurements as well regarding the representing graph of a PCM. From these indicators the following property has special importance in our study.

Definition 24 (*k*-edge-connectivity) A graph G is called k-edge-connected if it remains connected whenever fewer than k edges are removed from G. Formally: let $G = (V, E)$ be an undirected graph, where V is the vertex set, while E is the edge set of G. If $G' = (V, E \setminus H)$ is connected for $\forall H \subseteq E$, where $|H| < k$, then G is k-edge-connected. The edge connectivity of G is the maximum value k such that G is k-edge-connected.

3.3.2 The simulation process

We apply extended numerical experiments to compare different filling in models of pairwise comparisons. In our simulations, the weight vectors are calculated using the natural extension of the two most popular weighting techniques, the eigenvector method and LLSM, to the incomplete PCM case. The former one is based on the CR-minimal completion (CREV), and its principal right eigenvector (Bozóki et al., 2010, Sections 3 and 5). The LLSM's optimization problem includes the known elements of the matrix, the optimal solution can be calculated by solving a system of linear equations (Bozóki et al., 2010, Sections 4), furthermore, it can also be written as the geometric mean of weight vectors calculated from all spanning trees (Bozóki and Tsyganok, 2019). As we mentioned, in both of the CREV and the LLSM models the optimal solution is unique if and only if the graph of comparisons is connected. We compare the weight vectors obtained based on different filling in models to the ones calculated from the complete PCMs. As for the measurements of comparison, we use three types of metrics:

- distances, as the most commonly used cardinal measures, which are represented by the Euclidean distance (d_{euc}), the Chebyshev distance (d_{cheb}), and the Manhattan distance (d_{man}),
- rank correlation coefficients, as the basic ordinal indicators, namely the Kendall rank correlation (Kendall's τ), and the Spearman rank correlation coefficient (Spearman's ρ),
- and last but not least, the so called compatibility (or similarity) indices, which are argued in the literature to be the most important measures to compare priority vectors (Garuti, 2017). In our analysis we include Garuti's compatibility index (G index) (Garuti, 2020), the cosine similarity index (C index) (Kou et al., 2021), and the dice similarity index (D index) (Ye, 2012).

The above-mentioned indicators are given by the following formulas, respectively:

$$d_{euc}(u, v) = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2} \tag{15}$$

$$d_{cheb}(u, v) = \max_{i \in 1,\ldots,n} |u_i - v_i| \tag{16}$$

$$d_{man}(u, v) = \sum_{i=1}^{n} |u_i - v_i| \tag{17}$$

$$\tau(u, v) = \frac{n_c(u, v) - n_d(u, v)}{n(n-1)/2} \tag{18}$$

$$\rho(u, v) = 1 - \frac{6 \sum_{i=1}^{n} R(u_i) - R(v_i)}{n(n^2 - 1)} \tag{19}$$

$$G(u, v) = \frac{1}{2} \sum_{i=1}^{n} \left(\frac{\min(u_i, v_i)}{\max(u_i, v_i)} (u_i + v_i) \right) \tag{20}$$

$$C(u, v) = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} (u_i)^2} \sqrt{\sum_{i=1}^{n} (v_i)^2}} \tag{21}$$

$$D(u, v) = \frac{2 \sum_{i=1}^{n} u_i v_i}{\sum_{i=1}^{n} (u_i)^2 + \sum_{i=1}^{n} (v_i)^2} \tag{22}$$

where u denotes the weight vector gained from a certain filling in pattern and v is the weight vector calculated from the complete matrix. u and v are normalized by $\sum_{i=1}^{n} u_i = 1$ and $\sum_{i=1}^{n} v_i = 1$, respectively, and v_i and u_i denote the ith element of the appropriate vectors. $n_c(u, v)$ denotes the number of concordant pairs and $n_d(u, v)$ the number of discordant pairs of the examined vectors, while $R(.)$ shows the rank of the given element within the appropriate vector. The range of the Kendall's τ and Spearman's ρ is $[-1, 1]$, and considering the notation above, the higher value indicates a better performance of the given filling in pattern. In case of the compatibility indices (G, C, and D) the range corresponds to $[0, 1]$, and similarly, the higher value is the better. On the other hand, here the distances can be interpreted as errors, thus their smaller level is preferred. This way we use the most commonly applied categories of closeness measures in case of priority vectors, and also include several of them to see if the results depend on the given category or may even on the specific metric.

Our simulations are in a sense similar to the ones in Szádoczki et al. (2022), but (besides using other filling patterns and other metrics) we apply elementwise perturbations instead of their matrixwise solution, and the handling of scales is improved in our case.

The process of the simulation for a given n consisted of the following steps:

1. We generated n random weights (the general weight is denoted by w_i), where $w_i \in [1, 9]$ is a uniformly distributed random real number for $\forall i \in 1, 2, \ldots, n$.

We calculated random $n \times n$ complete and consistent PCMs, where the elements of the matrices were given as follows:

$$a_{ij} = w_i / w_j \tag{23}$$

2. Then we used three different special perturbations of the elements of the consistent matrices to get inconsistent PCMs with three well-distinguishable inconsistency levels. We denote these levels by weak, modest and strong given with the following formulas:

$$\hat{a}_{ij}^{weak} = \begin{cases} a_{ij} + \Delta_{ij} & : a_{ij} + \Delta_{ij} \geq 1 \\ \frac{1}{1-\Delta_{ij}-(a_{ij}-1)} & : a_{ij} + \Delta_{ij} < 1 \end{cases} \qquad \Delta_{ij} \in [-1, 1] \tag{24}$$

$$\hat{a}_{ij}^{modest} = \begin{cases} a_{ij} + \Delta_{ij} & : a_{ij} + \Delta_{ij} \geq 1 \\ \frac{1}{1-\Delta_{ij}-(a_{ij}-1)} & : a_{ij} + \Delta_{ij} < 1 \end{cases} \qquad \Delta_{ij} \in \left[-\frac{3}{2}, \frac{3}{2}\right] \tag{25}$$

$$\hat{a}_{ij}^{strong} = \begin{cases} a_{ij} + \Delta_{ij} & : a_{ij} + \Delta_{ij} \geq 1 \\ \frac{1}{1-\Delta_{ij}-(a_{ij}-1)} & : a_{ij} + \Delta_{ij} < 1 \end{cases} \qquad \Delta_{ij} \in [-2, 2] \tag{26}$$

Where $\hat{a}_{ij}^{weak}$, $\hat{a}_{ij}^{modest}$ and $\hat{a}_{ij}^{strong}$ are the elements of the perturbed matrices, a_{ij} is the element of the consistent matrix, $a_{ij} \geq 1$ (thus we only perturb the elements above one and keep the reciprocity of the matrices), and Δ_{ij} is uniformly distributed in the given ranges. This perturbation is able to provide ordinal differences as well (when $\hat{a}_{ij} < 1$). However, we account for the contrast that can be examined above and below 1, thus our perturbed data is uniformly distributed around the original element on the scale presented by Figure 9. Our perturbation method aims to ensure 3 different and meaningful inconsistency levels and it is, indeed, correlated with the Consistency Ratio (CR), as it is shown in Figure 10. We tested several combinations of parameters, and found that they resulted in the most relevant levels of CR.

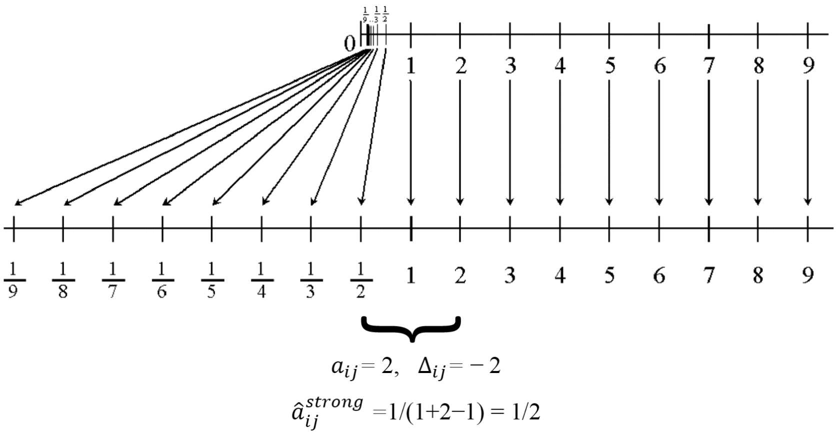


Figure 9: The ratio scale $1/9, \ldots, 9$ and the perturbation of elements

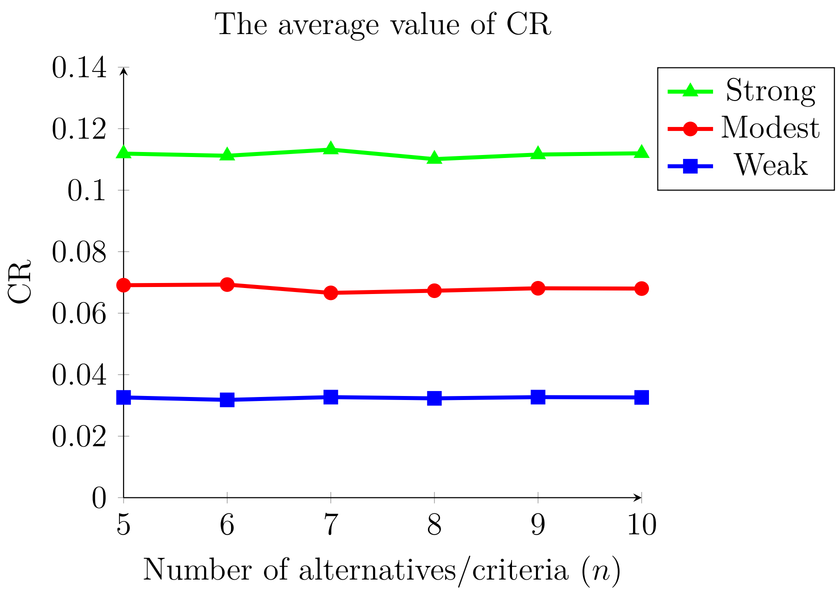


Figure 10: The relation between CR and our element-wise perturbation. Each point shows the average CR of 1000 randomly generated perturbed PCMs.

3. We deleted the respective elements of the matrices in order to get the filling in pattern that we were examining, and applied the LLSM and the CREV techniques to get the weights. In case of the models that use ordinal information, we always chose the needed element according to the examined weight calculation method, based on the complete, perturbed PCM (thus we assume that the decision maker can provide accurate ordinal data). The certain models' distances, rank correlations and compatibility indices were computed with respect to the weights that were calculated from the complete inconsistent matrices. The compared filling in patterns were the ones presented in great detail in Section 3.4, which are represented by the following graphs:

    (i) Best-worst graph

(ii) TOP2 graph

(iii) Best-random graph

(iv) Random-random graph

(v) 3-(quasi-)regular graph with minimal diameter

(vi) Union of two random edge-disjoint spanning trees

(vii) Random 2-edge-connected graph

4. We repeated steps 1-3 for 10000 times for every level of inconsistency (thus altogether we examined 30000 PCMs for a given n). Finally, we saved the mean of the applied metrics for the different weight calculation methods and filling in patterns.

In case of (vi) and (vii) we randomly generated 10000 graphs satisfying the required properties and used them in the simulations. As for (v), when there were more graphs, which met the requirements, we randomly chose one of them at every iteration.

It worth to include a small theoretical analysis of the simulations to see how they work in general.

Remark 1 The distribution of the elements of PCMs in the simulation is independent of n. This property holds for both consistent and perturbed PCM cases.

The reason behind this is as follows. If we analyze our simulation process, we can see that at first the elements of a given matrix are generated independently from n, and then they are placed into the $n \times n$ PCM. The histograms of the matrix elements above 1 in the different perturbation cases, based on large samples containing 1 million elements each are presented in Figure 11.

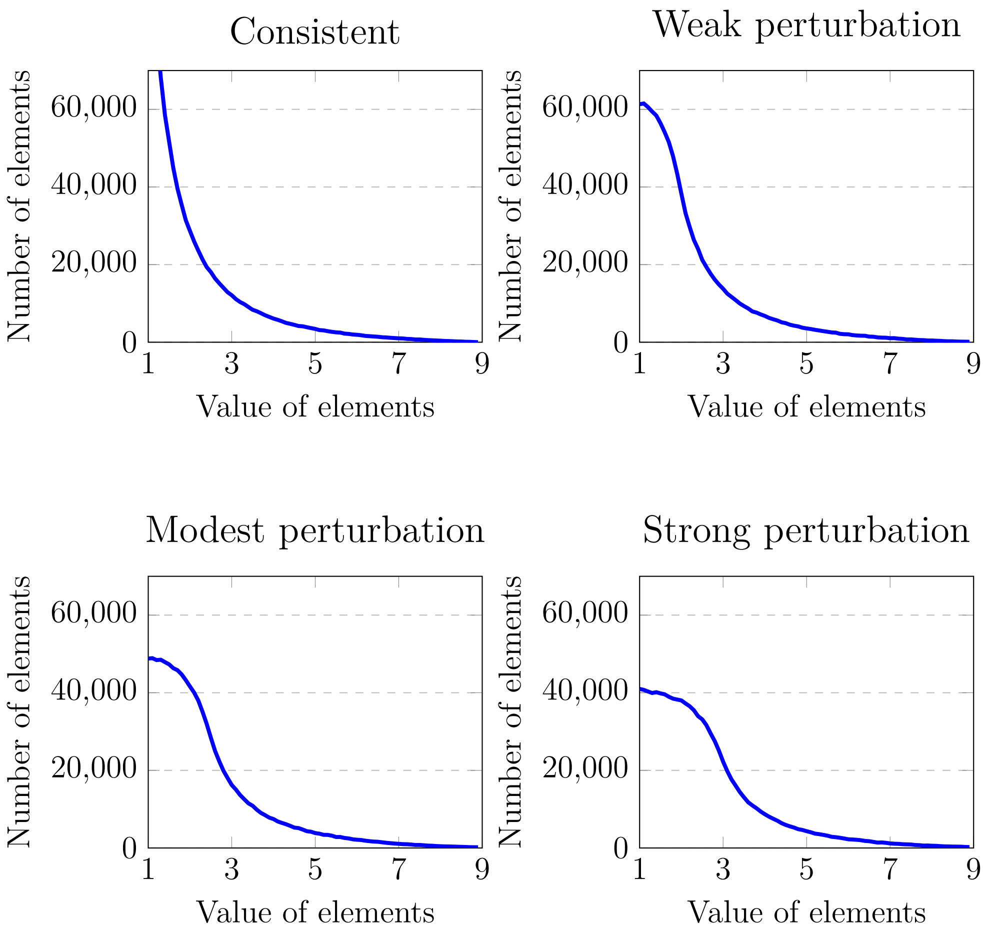


Figure 11: The histograms of the ≥ 1 elements of PCMs in case of different perturbations based on a sample of 1 million elements.

It can be seen that the higher the level of perturbation (inconsistency), the higher the chance to have large (extreme) matrix elements.

It also makes sense to consider the average of the maximal elements of the analyzed PCMs and the average number of ordinal perturbations (when the ordinal preference between two items is changed due to perturbation). These details are presented in Tables 6 and 7, respectively, for our specific simulations.

As one can see, the higher the number of alternatives (criteria), the higher the average maximal element in the matrices. However, this is the case only because in a larger matrix we have a larger sample of elements, thus the maximum has a higher probability to be an extreme element. Naturally, the stronger perturbation also results in larger maximal element (as it is also suggested by the histograms). As for the ordinal perturbations, we can see that the average number of them for all n in case of the weak perturbation level is slightly below 20% (≈19%) of the possible comparisons, while it is slightly above 25% (≈27%) for the strong level.

Size (n)	5	6	7	8	9	10
Perfectly consistent	3.96	4.35	4.67	4.94	5.18	5.38
Weak perturbation	4.21	4.60	4.93	5.19	5.46	5.67
Modest perturbation	4.45	4.85	5.20	5.47	5.71	5.95
Strong perturbation	4.74	5.14	5.49	5.80	6.05	6.25

Table 6: The summary of the mean of largest elements in the PCMs based on the simulations

Size (n)	5	6	7	8	9	10
Weak perturbation	1.94	2.90	4.02	5.32	6.90	8.68
Modest perturbation	2.50	3.64	5.03	6.72	8.62	10.75
Strong perturbation	2.79	4.21	5.77	7.70	9.96	12.46
Possible comparisons	10	15	21	28	36	45

Table 7: The average number of ordinal perturbations in the simulations

3.4 Filling in patterns

As we mentioned earlier we would like to keep the number of comparisons low in the analyzed models, the average degree of the representing graphs should be approximately 3. All in all we chose to examine seven different filling in patterns, from which the first three use additional ordinal information of the ranking as well. These models (and their associated graphs) are as follows.

(i) Best-worst graph

(ii) TOP2 graph

(iii) Best-random graph

(iv) Random-random graph

(v) 3-(quasi-)regular graph with minimal diameter

(vi) Union of two random edge-disjoint spanning trees

(vii) Random 2-edge-connected graph

The detailed description of these models can be found below.

(i) Best-worst graph The name of the model comes from the popular best-worst method (Rezaei, 2015), where only the best and the worst items are compared to all the others. We use exactly this filling in pattern, where only two rows and columns are completed (the ones associated with the best and the worst items), and the related graph is the union of two star graphs, which results in $2n-3$ comparisons. We examine this case, because it is widely used and fits perfectly within the outline of our paper. It is important to note that this model utilizes some additional ordinal information, however, we also examine the same graph representation with other ordinal data, namely when not the best and the worst elements are the highlighted ones.

(ii) TOP2 graph Intuitively in general, the more comparisons are made in connection with one particular item, the more accurate our estimation of its weight becomes. In most of the cases the first few places of the ranking are much more important for the decision maker, than the last few. The TOP2 model suggests to compare all the items only to the best and the second best elements. One could argue that in the best-worst case, it is easier for the decision maker to choose the best and the worst elements compared to the best and the second best ones. However, if we assume that it is always possible for the decision maker to find the best and the worst elements, then we can remove those as the first step and find the new best and worst items, thus we can determine the best and the second best as well. The advantages of the models that use ordinal information are emphasized in the multicriteria nature of the problems (Rezaei, 2015). The associated graph of this pattern is the aforementioned union of two star graphs, of course with different highlighted vertices than before.

(iii) Best-random graph It also makes sense to investigate, whether it is really necessary to ask the decision maker to provide additional ordinal information, beside just naming the best item. Thus, we consider the Best-random model, in which the elements are compared only to the best one and a randomly chosen other item. The associated graph is still the union of two star graphs, however in this case we use less ordinal information.

(iv) Random-random graph As a benchmark for the previous models, we also examine the case, when no additional ordinal information is included, and the two highlighted vertices are both chosen randomly. This is the last case, when the associated graph is the union of two star graphs, for which an example on 6 nodes is shown in Figure 12.

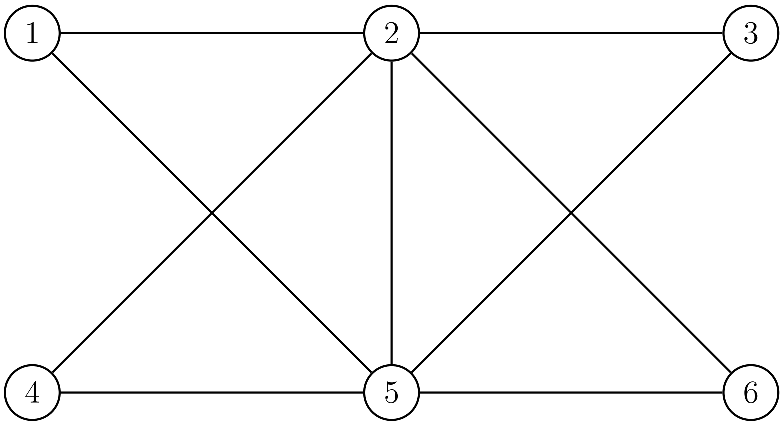


Figure 12: The union of two star graphs on $n = 6$ vertices, which is the associated graph of the (i) Best-worst, (ii) TOP2, (iii) Best-random, and (iv) Random-random models. Here the highlighted vertices are 2 and 5.

(v) 3-(quasi-)regular graph with minimal diameter This filling in pattern was suggested by Szádoczki et al. (2022). No additional ordinal information is needed, and the cardinality of the model fits into the examined cases. The (quasi-)regularity results in some kind of symmetry, while the minimal diameter ensures that the comparisons are close enough to the direct ones (the shortest path between any two vertices is not so long). The number of comparisons for these graphs is $3n/2$ in case of even number of vertices (regularity) and $3n/2 + 1/2$ in case of odd number of vertices (quasi-regularity). Two examples on $n = 5$ and $n = 6$ vertices can be seen in Figure 13.

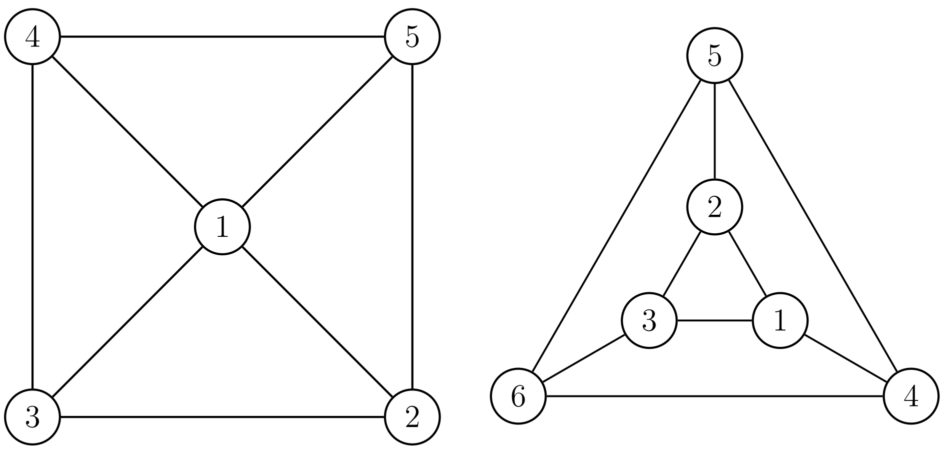


Figure 13: Two examples for 3-(quasi-)regular graphs with minimal diameters ($d = 2$). The left one is the only 3-quasi-regular graph with minimal diameter on $n = 5$ vertices, while the right one is one of the two 3-regular graphs with minimal diameter on $n = 6$ vertices.

(vi) Union of two random edge-disjoint spanning trees Graph based decision making has a special interest in spanning trees as they are the minimal units from which we can calculate weight vectors. The star graph is a special spanning tree as well, thus the union of two random spanning trees of a graph can be seen as a generalization of the union of two star graphs. For the sake of simplicity the model examines only random edge-disjoint spanning trees. This way the gained union contains $2n-2$ edges (comparisons). An example of two edge-disjoint spanning trees on $n=5$ vertices can be seen in Figure 14.

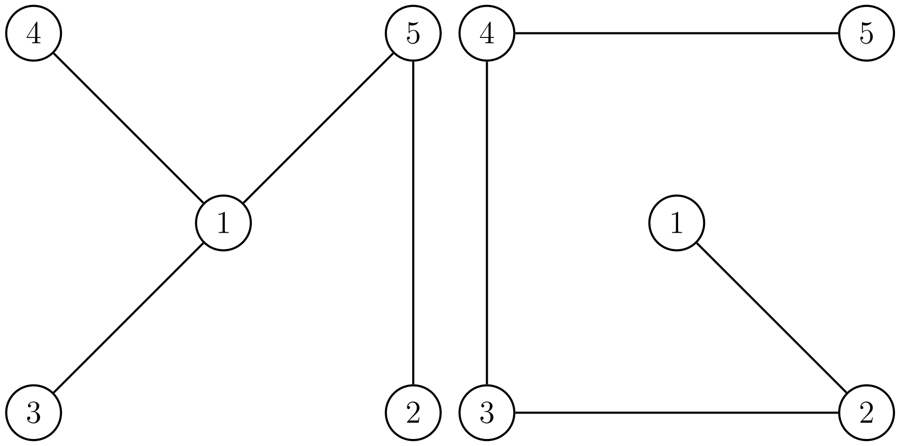


Figure 14: Two edge-disjoint spanning trees on $n=5$ vertices, which, when unified, form the 3-quasi-regular graph with minimal diameter on 5 vertices that can be seen in Figure 13.

(vii) Random 2-edge-connected graph 2-edge-connected graphs remain connected if we remove any one of their edges. As the union of two star graphs satisfies this property, this can be considered (according to the number of graphs) an even more common generalization of the union of two star graphs based on this connectedness measurement. As there are many 2-edge-connected graphs with different cardinalities, the model contains only the graphs with $2n-3$ or $2n-2$ edges. An example that does not fit into any of the previously listed filling in patterns, a 2-edge-connected graph with $2n-3$ edges, is presented in Figure 15.

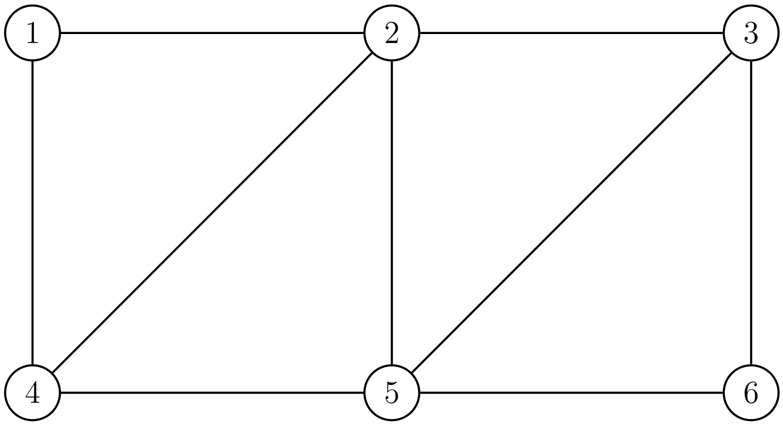


Figure 15: A 2-edge-connected graph on $n = 6$ vertices with $2n-3$ edges that does not satisfy the properties of any other examined model.

The numbers of possible non-isomorphic graphs for the different filling in patterns are presented in Table 8.

Size (n)	(i)-(ii)-(iii)-(iv)	(v)	(vi)	(vii)
5	1	1	2	4
6	1	2	7	21
7	1	4	95	164
8	1	2	1064	1788
9	1	2	17100	26158
10	1	1	327732	478657

Table 8: The number of non-isomorphic graphs for the different cases. We use the former notation of the models: (i) Best-worst graph, (ii) TOP2 graph, (iii) Best-random graph, (iv) Random-random graph, (v) 3-(quasi-)regular graph with minimal diameter, (vi) Union of two random edge-disjoint spanning trees, (vii) Random 2-edge-connected graph.

As one can see the number of graphs for the union of two edge-disjoint spanning trees (vi) and the 2-edge-connected case with $2n-3$ or $2n-2$ edges (vii) are, in a way, outliers from this point of view. In order to count the number of those graphs, we used Wolfram Mathematica (Wolfram Research, 2021), nauty and Traces (McKay and Piperno, 2014), and IGraph/M (Horvát, 2020). The union of two star graphs (i-ii-iii-iv) is on the other end of the spectrum, where there is only one non-isomorphic graph for every n.

The discussed filling in patterns and the associated numbers of comparisons are summarized in Table 9.

Size (n)	Complete	(i)-(ii)-(iii)-(iv)	(v)	(vi)	(vii)
	Number of edges				
	$n(n-1)/2$	$2n-3$	$3n/2$ (+1/2 for odd n)	$2n-2$	$2n-3$ or $2n-2$
5	10	7	8	8	7-8
6	15	9	9	10	9-10
7	21	11	11	12	11-12
8	28	13	12	14	13-14
9	36	15	14	16	15-16
10	45	17	15	18	17-18

Table 9: The summary of the cardinality of the graphs' edges (the number of comparisons). We use the former notation of the models: (i) Best-worst graph, (ii) TOP2 graph, (iii) Best-random graph, (iv) Random-random graph, (v) 3-(quasi-)regular graph with minimal diameter, (vi) Union of two random edge-disjoint spanning trees, (vii) Random 2-edge-connected graph. The third row shows the formula for the number of comparisons for the different patterns.

The cardinality of the different graphs, namely the required number of comparisons is, indeed, similar. The inclusion of models, which utilize additional ordinal data (i-ii-iii), makes it possible to evaluate this information as well, which is an important contribution of our paper.

3.5 Simulation results

The results of the simulations are presented in four types of figures (all of them are sorted by n):

1. Figures 16, 17, 18, 19, 20 and 21 show the Euclidean distances (y axis) and Kendall's taus (x axis),
2. Figures 22, 23, 24, 25, 26 and 27 show the cosine similarity indices (y axis) and Garuti's compatibility indices (x axis),
3. Figures A1, A2, A3, A4, A5 and A6 show the Chebyshev distances (y axis) and Spearman's rhos (x axis),
4. finally, Figures A7, A8, A9, A10, A11, and A12 show the Manhattan distances (y axis) and dice similarity indices (x axis) for the different models in case of the given perturbation level and the given weight calculation technique.

It is important to note that for the distances the smaller value, and in case of the rank correlation coefficients and compatibility indices the higher level indicates the better performance. Thus for the first, third and fourth types of figures (Figures 16, 17, 18, 19, 20, 21, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, A11, and A12) models closer to the bottom right corner are the preferred ones. In case of the second type of figures (Figures 22, 23, 24, 25, 26 and 27) the upper right corner provides the best results. The third and fourth types of figures can be found in the online Appendix A, mainly because we found that the metrics coming from the same category (distances, rank correlations, compatibility indices) tend to provide similar results. Now, let us analyze the results of the first type of figures in more detail.

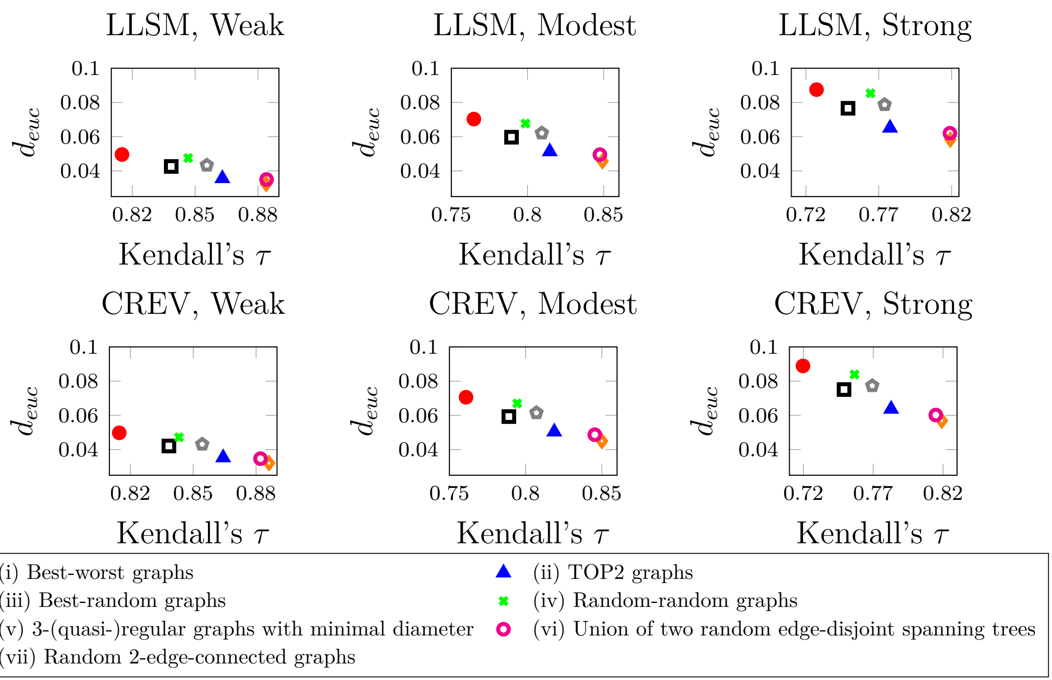


Figure 16: The results of the simulation for $n = 5$. The following notations are used: d_{euc}– Euclidean distance, Kendall's τ–Kendall rank correlation coefficient, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

Naturally, the outcomes suggest that the higher level of perturbation, namely the higher inconsistency leads to higher distances and lower rank correlation coefficients.

Despite of the used additional ordinal information the (i) Best-worst and (iii) Best-random filling patterns were performing according to both the distance and rank correlation measurements more or less the same as the (iv) Random-random case, which is based on an identical graph without any further information. This would suggest that the additional ordinal information does not provide significant improvement.

The outcomes for the (ii) TOP2 model are much better regarding both of our

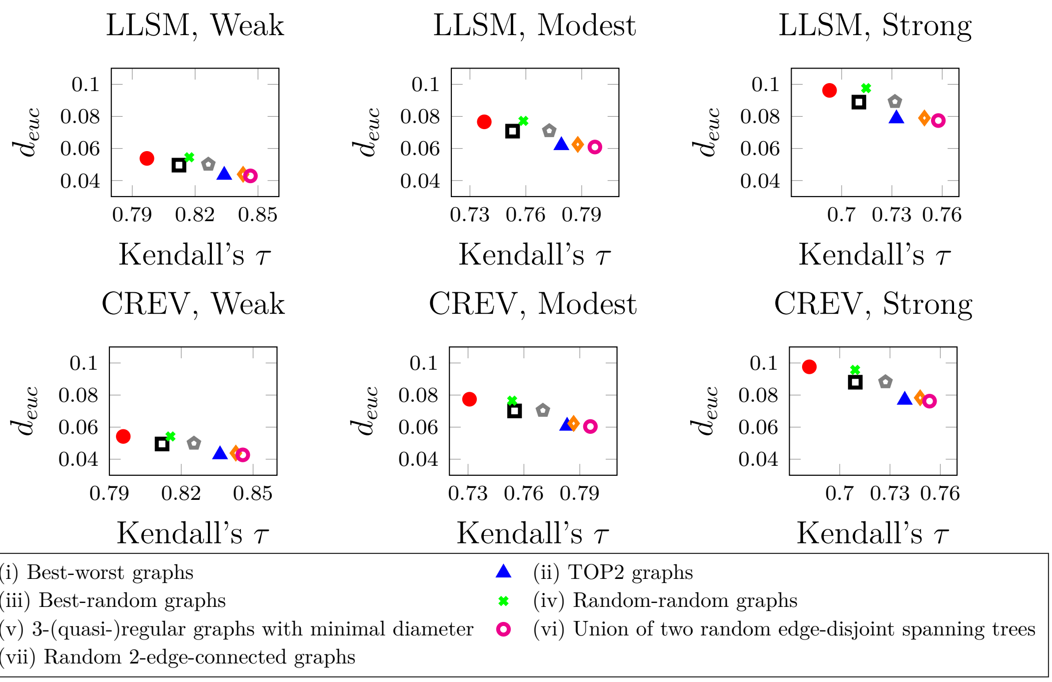


Figure 17: The results of the simulation for $n = 6$. The following notations are used: d_{euc}–Euclidean distance, Kendall's τ–Kendall rank correlation coefficient, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

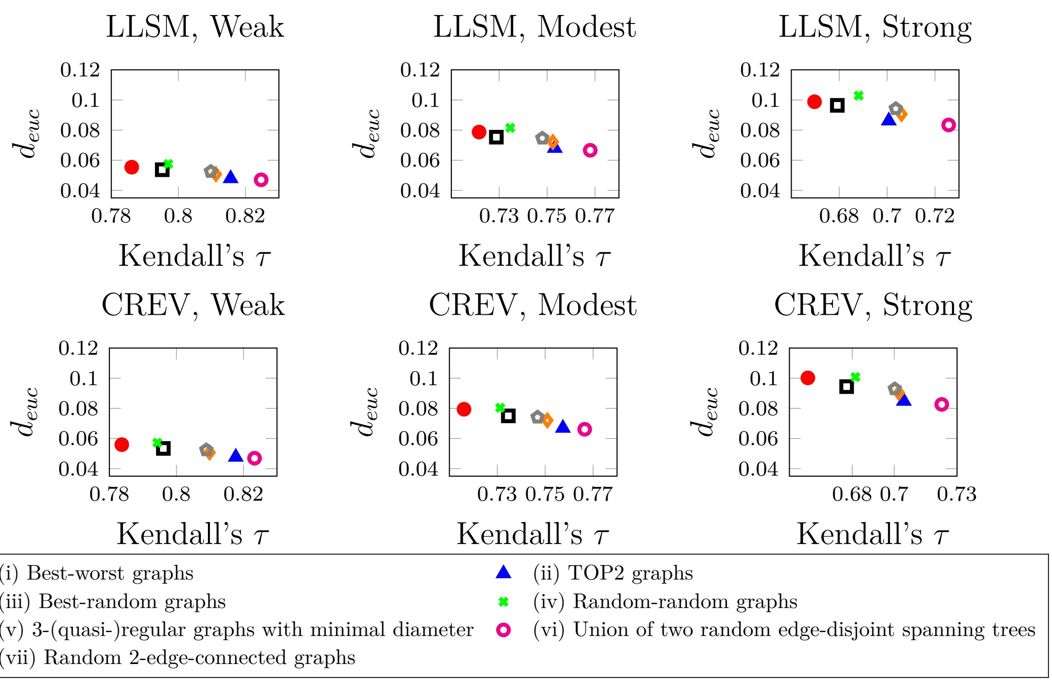


Figure 18: The results of the simulation for $n = 7$. The following notations are used: d_{euc}–Euclidean distance, Kendall's τ–Kendall rank correlation coefficient, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

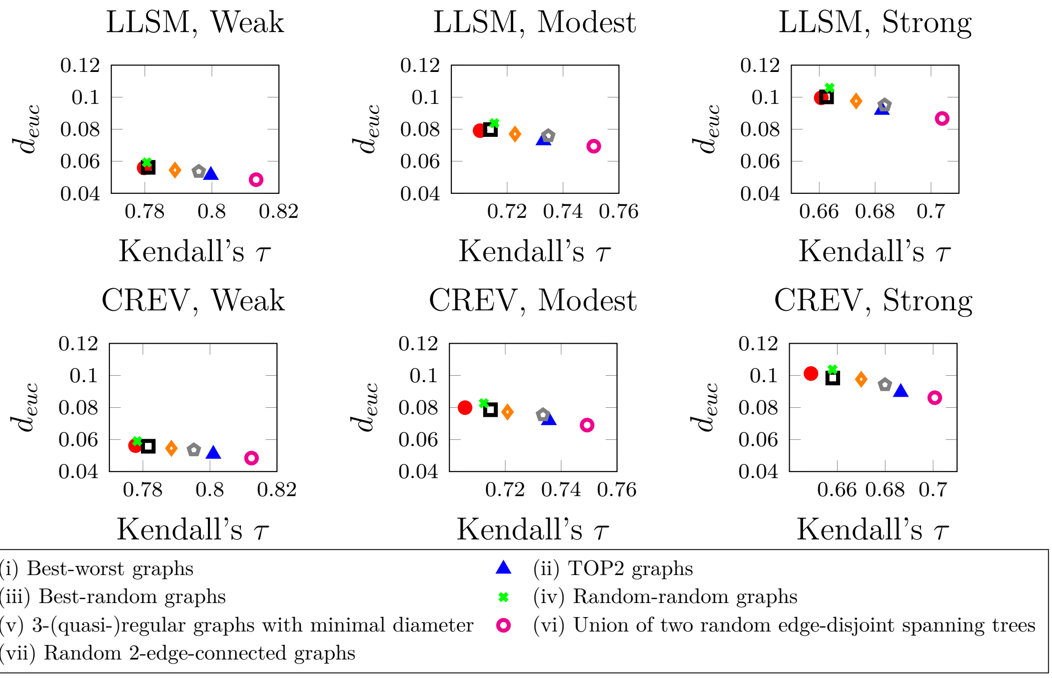


Figure 19: The results of the simulation for $n = 8$. The following notations are used: d_{euc}–Euclidean distance, Kendall's τ–Kendall rank correlation coefficient, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

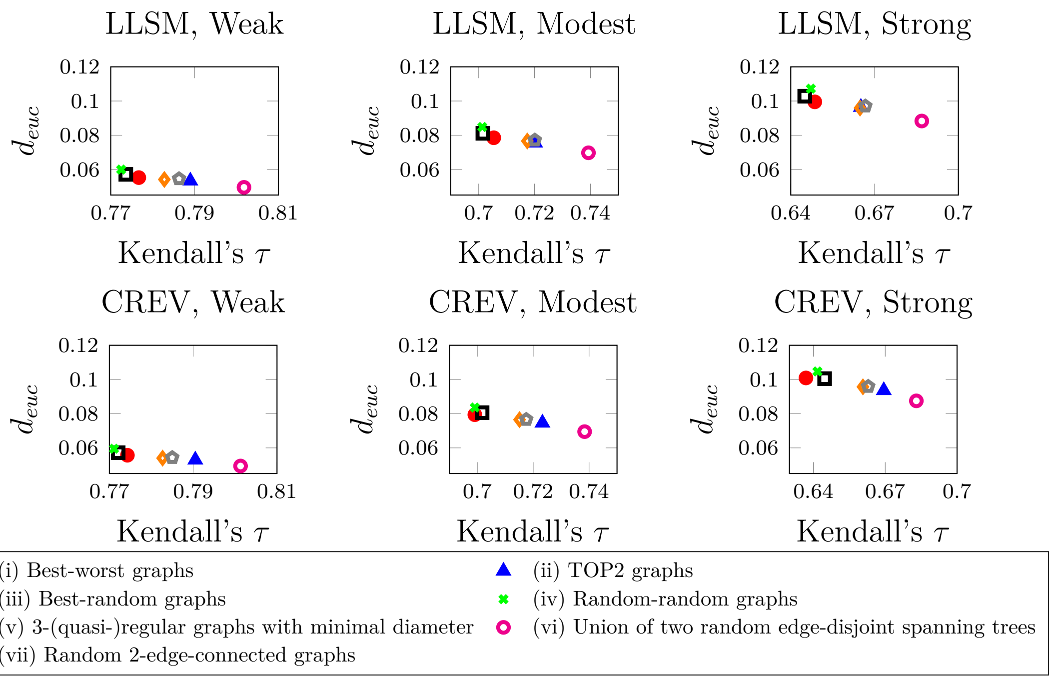


Figure 20: The results of the simulation for $n = 9$. The following notations are used: d_{euc}–Euclidean distance, Kendall's τ–Kendall rank correlation coefficient, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

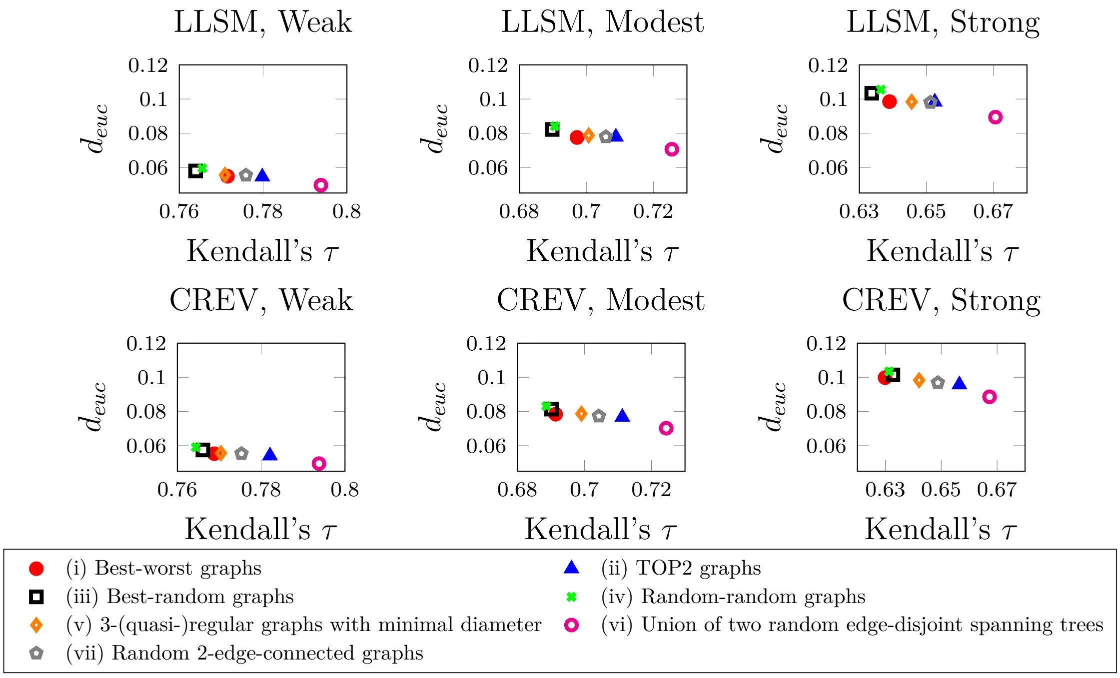


Figure 21: The results of the simulation for $n = 10$. The following notations are used: d_{euc}–Euclidean distance, Kendall's τ–Kendall rank correlation coefficient, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

distance and rank correlation indicators, and it convincingly outperformed the union of two star graphs with other centers (i-iii-iv). However, surprisingly enough the (vi) Union of two random edge-disjoint spanning trees provided the best weight vectors, basically, in every case according to all distances and rank correlations, without any additional ordinal information. Thus, according to our simulations, this filling in pattern results in weight vectors and rankings, closest to the ones, computed from the complete PCMs for all the used weight calculation methods and inconsistency levels. It is important to note that this filling in pattern usually contains one more edge than most other patterns, however, the results are still quite convincing.

As for the (v) 3-quasi-regular graphs with minimal diameter, it performs well in case of smaller problems, but starts to decline as n grows. However, it still outperforms most of the models that use additional information. It is also true that for larger number of vertices this case uses the smallest number of edges (comparisons), which can also contribute to the aforementioned trend.

The (vii) Random 2-edge-connected graph always performs worse than (vi), but better, than (i), (iii) and (iv) and it also has a slight edge advantage, thus it seems to be the medium method.

Now let us focus on the second type of figures (Figures 22, 23, 24, 25, 26 and 27).

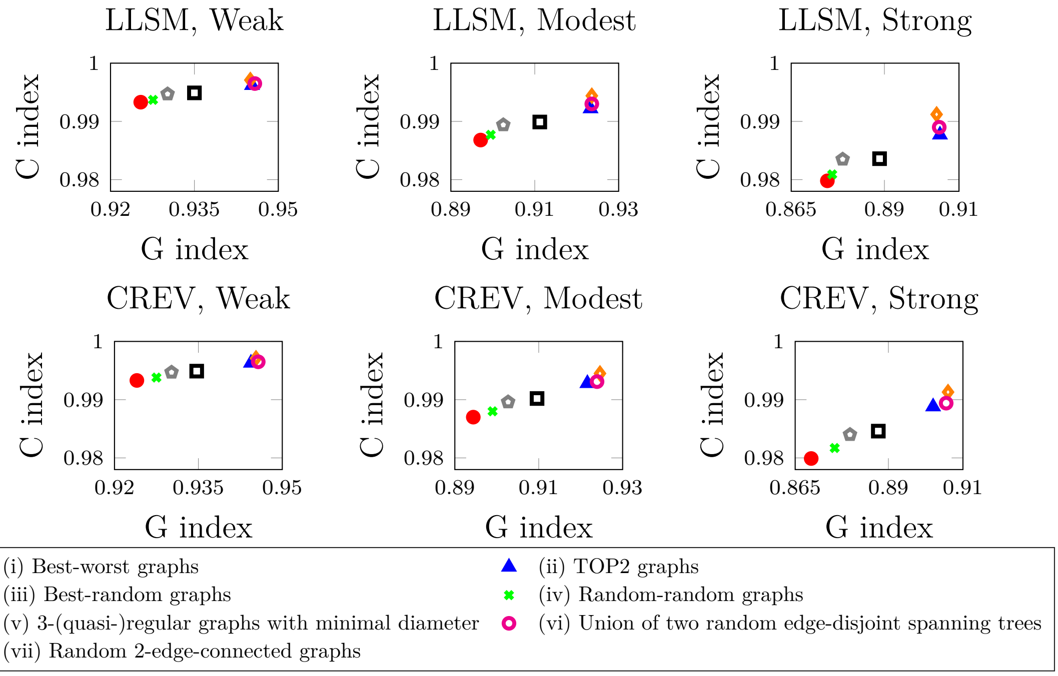


Figure 22: The results of the simulation for $n = 5$. The following notations are used: G index–Garuti's compatibility index, C index–Cosine similarity index, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

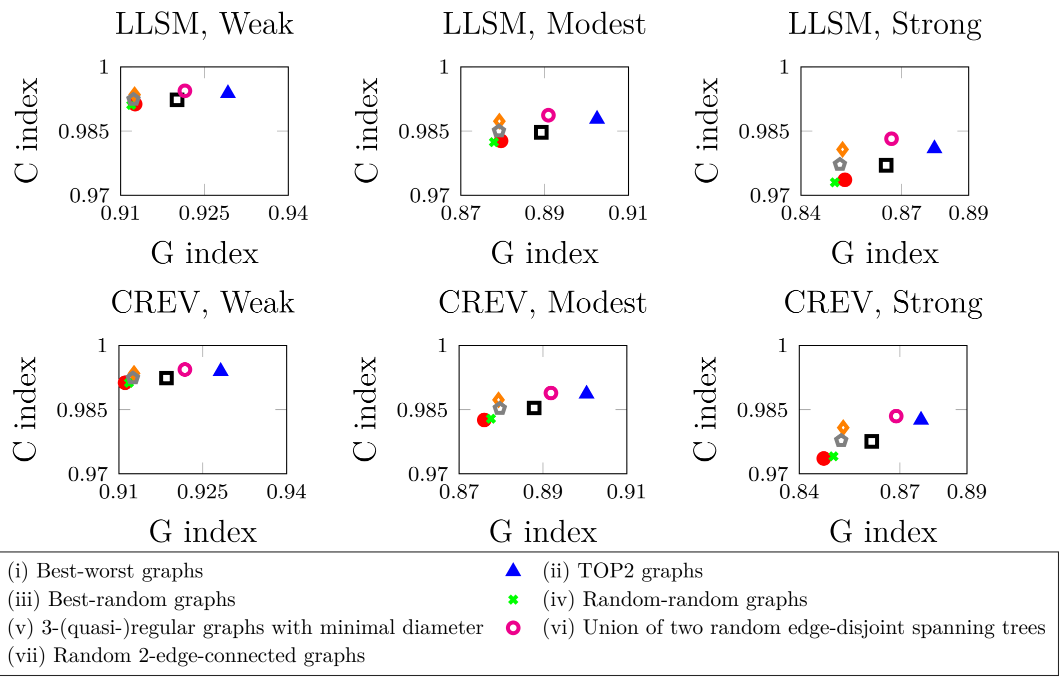


Figure 23: The results of the simulation for $n = 6$. The following notations are used: G index–Garuti's compatibility index, C index–Cosine similarity index, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

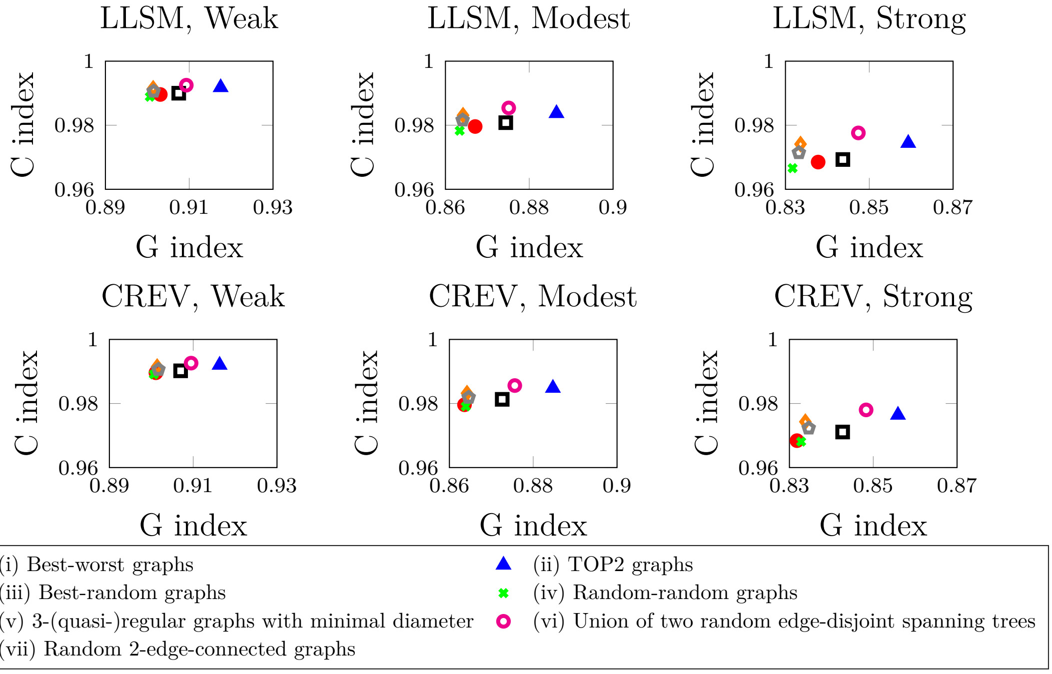


Figure 24: The results of the simulation for $n = 7$. The following notations are used: G index–Garuti's compatibility index, C index–Cosine similarity index, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

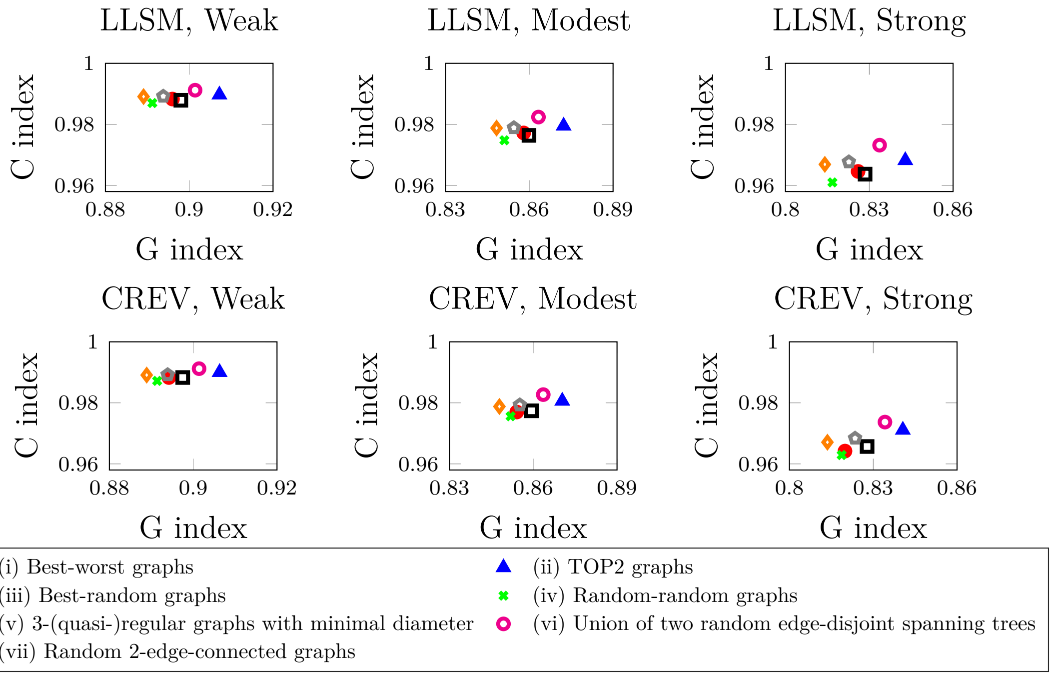


Figure 25: The results of the simulation for $n = 8$. The following notations are used: G index–Garuti's compatibility index, C index–Cosine similarity index, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

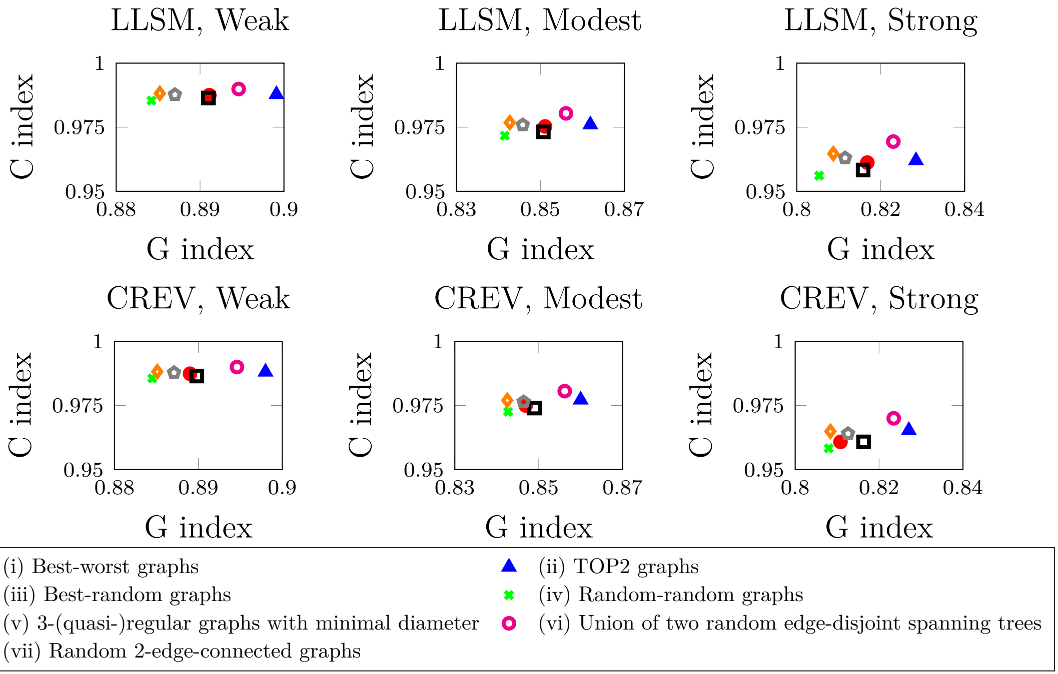


Figure 26: The results of the simulation for $n = 9$. The following notations are used: G index–Garuti's compatibility index, C index–Cosine similarity index, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

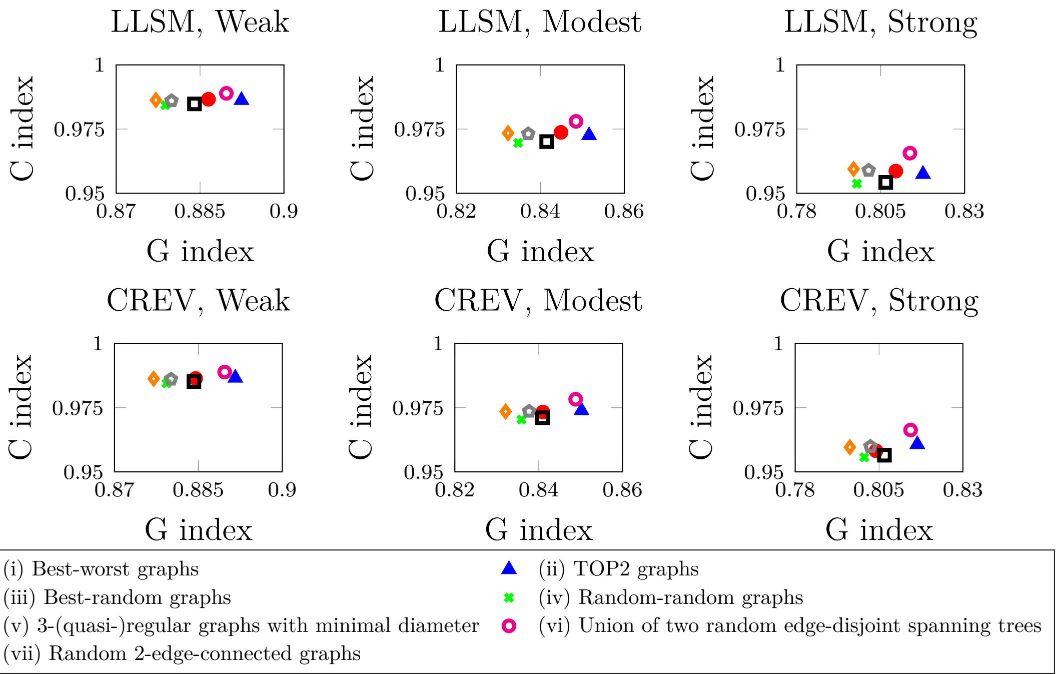


Figure 27: The results of the simulation for $n = 10$. The following notations are used: G index–Garuti's compatibility index, C index–Cosine similarity index, 'Weak', 'Modest' and 'Strong' refers to the level of perturbation. Every subfigure shows the mean computed from 10000 pairwise comparison matrices.

The results provided by the compatibility indices are similar to the previous ones, the main difference is that the models using additional ordinal information tend to perform better compared to the distances and rank correlations, especially for larger problems (n parameters). The most remarkable change is that the (ii) TOP2 model performs even better, and in many cases it provides the best values among the examined patterns. Also, the decreasing performance of the (v) 3-quasi-regular graphs seem to be even stronger for these metrics. However, the (vi) Union of two random edge-disjoint spanning trees model still outperforms all of the other models using additional information as well, except for the (ii) TOP2 model in some cases.

All in all we can conclude that the (vi) Union of two random edge-disjoint spanning trees model provided the best results without additional ordinal information. It is also shown that the best-second best ordinal information is indeed more valuable, than the best-worst or only the best case.

3.6 Conclusions and future research

We compared the weight vectors calculated from incomplete pairwise comparisons, such that the underlying graphs have approximately the same number of edges for each matrix size $n = 5, 6, 7, 8, 9, 10$. Based on the simulations we found that the presumed advantage of additional ordinal information on the ranking is realized only for the union of two star graphs, centered at the best and the second best items ((ii), TOP2). However, the union of two random edge-disjoint spanning trees (vi) outperforms all the other graphs according to both distance and rank correlation measures for every examined parameters, and only fell to second place in a few number of cases according to similarity indices, when the TOP2 model was the first one. This basically means that if there is an opportunity in a decision making problem to gain additional ordinal information, then the best and second best alternatives are preferred to the best and the worst ones. Also, the union of edge-disjoint spanning trees can result in as good as, or even better weight vectors than the ones calculated with additional information. In our view, these results of the paper are major contributions for AHP (and MADM) practitioners.

It is to be further investigated how the differences in the measures we applied, the Euclidean distance, Chebyshev distance, Manhattan distance, Kendall's tau, Spearman's rho, Garuti's compatibility index, cosine similarity index and dice similarity index (in Figures 8-13, 14-19, A1-A6 and A7-A12) can be interpreted in practical decision making.

Can better graphs be found among the ones having the same number of edges?

A future research can investigate the different cardinalities of comparisons once the matrix size is fixed: how much can the result be improved by filling in more elements? Empirical (experimental) PCMs have been tested from this point of view without an emphasis on the graphs' structure (Bozóki et al., 2013).

We considered some of the most intuitively-understandable closeness measures. Although there are infinitely many ways of measuring weight vectors' similarity, it is not at all trivial which ones coincide with the subjective measures of most decision makers.

More dense graphs are subject to investigation, in particular, for larger dimensionalities (n). Does the same type/family of graphs (e.g. union of random edge-disjoint spanning trees) perform the best for incomplete PCMs of low, middle and high density? Can the graphs be constructed layer by layer?

Our analysis covers predetermined graphs only. We assumed that the whole questionnaire should be prepared a priori. However, one would expect that adaptive patterns could perform better compared to static ones. Once the decision maker provides some matrix elements, the remainder of the graph itself can be optimized in a dynamic way in order to maximize the expected information content of further responses.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10479-022-04819-9.

Acknowledgements The authors thank the valuable comments and suggestions of the anonymous Reviewers. The comments of László Csató are greatly ackonwledged. The research of S. Bozóki and Zs. Szádoczki was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) under Grant NKFIA ED_18-2-2018-0006.

Author Contributions All the authors equally contributed in every part of the manuscript.

Funding Open access funding provided by Corvinus University of Budapest.

Availability of data and material The simulation results are added as a supplementary file.

Code availability Upon request from the authors.

Declarations

Conflict of interest Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ágoston, K. Cs. and Csató, L. (2022). Inconsistency thresholds for incomplete pairwise comparison matrices. *Omega*, 108:102576. https://doi.org/10.1016/j.omega.2021.102576.

Amenta, P., Lucadamo, A., and Marcarelli, G. (2021). On the choice of weights for aggregating judgments in non-negotiable AHP group decision making. *European Journal of Operational Research*, 288(1):294–301. https://doi.org/10.1016/j.ejor.2020.05.048.

Bozóki, S., Csató, L., and Temesi, J. (2016). An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research*, 248(1):211–218. https://doi.org/10.1016/j.ejor.2015.06.069.

Bozóki, S., Dezső, L., Poesz, A., and Temesi, J. (2013). Analysis of pairwise comparison matrices: an empirical research. *Annals of Operations Research*, 211:511–528. https://doi.org/10.1007/s10479-013-1328-1.

Bozóki, S., Fülöp, J., and Rónyai, L. (2010). On optimal completion of incomplete pairwise comparison matrices. *Mathematical and Computer Modelling*, 52(1):318–333. https://doi.org/10.1016/j.mcm.2010.02.047.

Bozóki, S. and 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(3-4):362–381. https://www.tandfonline.com/doi/abs/10.1080/03081079.2019.1585432.

Brunelli, M. (2017). Studying a set of properties of inconsistency indices for pairwise comparisons. *Annals of Operations Research*, 248:143–161. https://doi.org/10.1007/s10479-016-2166-8.

Brunelli, M., Canal, L., and Fedrizzi, M. (2013). Inconsistency indices for pairwise comparison matrices: a numerical study. *Annals of Operations Research*, 211(1):493–509. https://doi.org/10.1007/s10479-013-1329.

Crawford, G. and Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, 29(4):387–405. https://doi.org/10.1016/0022-2496(85)90002-1.

Csató, L. (2017). On the ranking of a Swiss system chess team tournament. *Annals of Operations Research*, 254(1-2):17–36. https://doi.org/10.1007/s10479-017-2440-4.

Deutskens, E., de Ruyter, K., Wetzels, M., and Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters*, 15:21–36. https://doi.org/10.1023/B:MARK.0000021968.86465.00.

Duleba, Sz., Alkharabsheh, A., and Gündoğdu, F. K. (2021). Creating a common priority vector in intuitionistic fuzzy AHP: a comparison of entropy-based and distance-based models. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04491-5.

Edwards, W. (1977). How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(5):326–340. https://doi.org/10.1109/TSMC.1977.4309720.

Edwards, W. and Barron, F. (1994). Smarts and smarter: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, 60(3):306–325. https://doi.org/10.1006/obhd.1994.1087.

Fedrizzi, M. and Giove, S. (2007). Incomplete pairwise comparison and consistency optimization. *European Journal of Operational Research*, 183(1):303–313. https://doi.org/10.1016/j.ejor.2006.09.065.

Garuti, C. E. (2017). Reflections on scales of measurement, not measurement of scales. *International Journal of the Analytic Hierarchy Process*, 9(3). https://doi.org/10.13033/ijahp.v9i3.522.

Garuti, C. E. (2020). A set theory justification of Garuti's compatibility index. *Journal of Multi-criteria Decision Analysis*, 27(1-2):50–60. https://doi.org/10.1002/mcda.1667.

Gass, S. (1998). Tournaments, transitivity and pairwise comparison matrices. *Journal of the Operational Research Society*, 49(6):616–624. https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2600572.

Harker, P. T. (1987). Incomplete pairwise comparisons in the analytic hierarchy process. *Mathematical Modelling*, 9(11):837–848. https://doi.org/10.1016/0270-0255(87)90503-3.

Horvát, S. (2020). IGraph/M. An immediately usable version of this software is accessible from its GitHub repository. https://doi.org/10.5281/zenodo.3739056.

Kadenko, S. and Tsyganok, V. (2020). An update on combinatorial method for aggregation of expert judgments in AHP. Proceedings of the International Symposium on the Analytic Hierarchy Process, ISAHP-2020. https://doi.org/10.13033/isahp.y2020.012.

Kou, G. and Lin, C. (2014). A cosine maximization method for the priority vector derivation in AHP. *European Journal of Operational Research*, 235(1):225–232. https://doi.org/10.1016/j.ejor.2013.10.019.

Kou, G., Peng, Y., Chao, X., Herrera-Viedma, E., and Alsaadi, F. E. (2021). A geometrical method for consensus building in GDM with incomplete heterogeneous preference information. *Applied Soft Computing*, 105:107224. https://doi.org/10.1016/j.asoc.2021.107224.

Kułakowski, K., Szybowski, J., and Prusak, A. (2019). Towards quantification of incompleteness in the pairwise comparisons methods. *International Journal of Approximate Reasoning*, 115:221–234. https://doi.org/10.1016/j.ijar.2019.10.002.

Liang, F., Brunelli, M., and Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. *Omega*, 96:102175. https://doi.org/10.1016/j.omega.2019.102175.

Lundy, M., Siraj, S., and Greco, S. (2017). The mathematical equivalence of the "spanning tree" and row geometric mean preference vectors and its implications for preference analysis. *European Journal of Operational Research*, 257(1):197–208. https://doi.org/10.1016/j.ejor.2016.07.042.

McKay, B. D. and Piperno, A. (2014). Practical graph isomorphism, II. *Journal of Symbolic Computation*, 60(0):94–112. https://doi.org/10.1016/j.jsc.2013.09.003.

Mi, X., Tang, M., Liao, H., Shen, W., and Lev, B. (2019). The state-of-the-art survey on integrations and applications of the best worst method in decision making:

Why, what, what for and what's next? *Omega*, 87:205–225. https://doi.org/10.1016/j.omega.2019.01.009.

Mohammadi, M. and Rezaei, J. (2020). Bayesian best-worst method: A probabilistic group decision making model. *Omega*, 96:102075. https://doi.org/10.1016/j.omega.2019.06.001.

Mustajoki, J., Hämäläinen, R., and Salo, A. (2005). Decision support by interval SMART/SWING – incorporating imprecision in the SMART and SWING methods. *Decision Sciences*, 36:317 – 339. https://doi.org/10.1111/j.1540-5414.2005.00075.x.

Revilla, M. and Ochoa, C. (2017). Ideal and maximum length for a web survey. *International Journal of Market Research*, 59(5):557–565. https://doi.org/10.2501/IJMR-2017-039.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, 53:49–57. https://doi.org/10.1016/j.omega.2014.11.009.

Rezaei, J., Nispeling, T., Sarkis, J., and Tavasszy, L. (2016). A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. *Journal of Cleaner Production*, 135:577–588. https://doi.org/10.1016/j.jclepro.2016.06.125.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281. https://doi.org/10.1016/0022-2496(77)90033-5.

Szádoczki, Zs., Bozóki, S., and Tekile, H. A. (2020). Proposals for the set of pairwise comparisons. Proceedings of the International Symposium on the Analytic Hierarchy Process, ISAHP-2020. https://doi.org/10.13033/isahp.y2020.054.

Szádoczki, Zs., Bozóki, S., and Tekile, H. A. (2022). Filling in pattern designs for incomplete pairwise comparison matrices: (Quasi-)regular graphs with minimal diameter. *Omega*, 107:102557. https://doi.org/10.1016/j.omega.2021.102557.

Szybowski, J., Kułakowski, K., and Prusak, A. (2020). New inconsistency indicators for incomplete pairwise comparisons matrices. *Mathematical Social Sciences*, 108:138–145. https://doi.org/10.1016/j.mathsocsci.2020.05.002.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286. https://doi.org/10.1037/h0070288.

Triantaphyllou, E. (2000a). Multi-criteria decision making methods. In *Multi-criteria Decision Making Methods: A Comparative Study*. Applied Optimization, vol 44. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-3157-6_2.

Triantaphyllou, E. (2000b). Reduction of pairwise comparisons in decision making via a duality approach. *Journal of Multi-criteria Decision Analysis*, 8(6):299–310. https://doi.org/10.1002/1099-1360(199911)8:6<299::AID-MCDA253>3.0.CO;2-7.

Tsyganok, V. (2000). Combinatorial method of pairwise comparisons with feedback (in Ukrainian). *Data Recording, Storage & Processing*, 2:92–102. https://doi.org/10.1016/0022-2496(77)90033-5.

Tsyganok, V. (2010). Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method. *Mathematical and Computer Modelling*, 52(3):538–544. https://doi.org/10.1016/j.mcm.2010.03.052.

von Winterfeldt, D. and Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge: Cambridge University Press.

Wang, K. and Takahashi, I. (1998). How to select paired comparisons in AHP of incomplete information – strongly regular graph design. *Journal of the Operations Research Society of Japan*, 41(2):311–328. https://doi.org/10.15807/jorsj.41.311.

Wolfram Research, I. (2021). Mathematica, Version 12.3. Champaign, IL, 2021. https://www.wolfram.com/mathematica.

Ye, J. (2012). Multicriteria decision-making method using the Dice similarity measure based on the reduct intuitionistic fuzzy sets of interval-valued intuitionistic fuzzy sets. *Applied Mathematical Modelling*, 36(9):4466–4472. https://doi.org/10.1016/j.apm.2011.11.075.

Zhou, X., Hu, Y., Deng, Y., Chan, F. T. S., and Ishizaka, A. (2018). A DEMATEL-based completion method for incomplete pairwise comparison matrix in AHP. *Annals of Operations Research*, 271:1045–1066. https://doi.org/10.1007/s10479-018-2769-3.

4 Study III. Optimal sequences for pairwise comparisons: the graph of graphs approach

Authors: Zsombor Szádoczki, Sándor Bozóki

Presented at the 12th Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications and at the International Symposium on the Analytic Hierarchy Process Webconference 2022. Currently under review at a Q1 journal.

https://doi.org/10.48550/arXiv.2205.08673

Abstract

In preference modelling, it is essential to determine the number of questions and their arrangements to ask from the decision maker. We focus on incomplete pairwise comparison matrices, and provide the optimal filling in patterns, which result in the closest (LLSM) weight vectors on average to the complete case for at most six alternatives and for all possible number of comparisons, when the underlying representing graph is connected. These results are obtained by extensive numerical simulations with large sample sizes. Many optimal filling structures resulted in optimal filling in sequences, one optimal case can be reached by adding a comparison to a previous one, which are presented on GRAPH of graphs. The star graph is revealed to be optimal among spanning trees, while the optimal graphs are always close to bipartite ones. Regular graphs also correspond to optimal cases, furthermore regularity is important for all optimal graphs, as the degrees of different vertices are always as close to each other as possible. Besides applying optimal filling structures in given decision making problems, practitioners can utilize the optimal filling sequences in the cases, when the decision maker can abandon the problem at any period of the process (e.g., in online questionnaires).

Keywords: Decision support systems, Pairwise comparison, Incomplete pairwise comparison matrix, Filling in sequence, GRAPH of graphs

4.1 Introduction

The concept of pairwise comparisons (Thurstone, 1927) is fundamental both in preference modelling and Multicriteria Decision Making (MCDM) (Triantaphyllou, 2000). These comparisons are frequently placed into so-called pairwise comparison matrices (PCMs), which are the basis of the Analytic Hierarchy Process (AHP)

(Saaty, 1977, 1980). Incompleteness (the absence of some comparisons) occurs quite often in practical problems (Bozóki et al., 2016), as well as in theoretical questions (Fedrizzi and Giove, 2007; Bozóki et al., 2010; Csató and Rónyai, 2016; Kułakowski and Talaga, 2020). In connection with decision making problems, one major source of missing data is the lack of willingness or time of the decision maker, as completing all comparisons – especially in the case of many different levels, criteria, and alternatives – can be exhausting and lingering (Szádoczki et al., 2022; Fedrizzi and Giove, 2013).

We would like to underline that the aim of our research is not to encourage decision makers to make less comparisons or decision analysts to ask fewer questions. However, we would like to provide the sequence of questions for the analysts, which ensures that whenever the decision maker stops answering the questions, the calculated preferences are in some sense the closest to the decision makers real preferences.

The arrangement of comparisons, which has a crucial effect on the results, is often represented by graphs (Gass, 1998). In this paper, we are the first to provide the optimal filling in patterns of incomplete pairwise comparison matrices, which on average produce the (both cardinally and ordinally) closest weight vectors to the complete case, for at most six alternatives (criteria) (n) for all possible given number of comparisons (e), when the respective graph is connected. These optimal patterns for the examined (n, e) pairs are significant findings of this paper themselves, however they result in (partial) optimal filling in sequences, which can be instrumental in the case of such problems (e.g., online questionnaires), where the decision makers can abandon the problem at any period of the process to always be as close to the decision makers preferences as possible.

These kind of problems are often present in the case of large-scale group decision making (Chao et al., 2021; Duleba et al., 2012; Li et al., 2022), or when several different experts' comparisons should be evaluated from different fields as well (Francis-Oliviero et al., 2021).

In the analysis of filling in sequences, the focus of the paper, but also in structural analysis of graphs and graph sequences in general, GRAPH of graphs is a convenient and efficient tool for research and visualization, too. NODEs of a GRAPH are graphs, and there is an EDGE between two NODEs (=graphs) if the associated graphs are in a specified relation, e.g., they can be drawn from each other by adding or deleting an edge. Depending on the specification of the relation, several GRAPHs of graphs have been investigated, see for instance Lovász (1977). Another remarkable GRAPH of graphs is the Petersen family of seven graphs, including the Petersen graph itself (Hashimoto and Nikkuni, 2013). The GRAPH of graphs by Mesbahi

(2002) is motivated by the evolution of graphs in a dynamic system.

It is worth noting that the term 'neighbouring graphs' in Lovász (1977) is used synonymously for 'there is an EDGE between two graphs'. Analogously, 'reachable' in Mesbahi (2002) means that there is a PATH between two graphs. We use the concept of GRAPH of graphs to visualize our findings throughout the paper.

The rest of the paper is organized as follows. Section 4.2 presents the fundamental concepts and definitions regarding PCMs and their graph representation. The methodology of the applied simulations and the related probability theoretical reasoning are detailed in Section 4.3, while Section 4.4 contains the results, the optimal filling in sequences for the examined cases. Finally, Section 4.5 concludes and raises research questions for the future.

4.2 Basic concepts: PCMs and their graph representation

Pairwise comparisons are the core of ranking, sports competitions, as well as many statistics and decision making techniques (Davidson and Farquhar, 1976; Csató, 2021). We focus on pairwise comparison matrices (PCMs) which are used in the Analytic Hierarchy Process (AHP) MCDM methodology to evaluate alternatives according to a criterion, as well as to determine the importance of the different criteria. However, our results can be beneficial in a wider range.

Definition 25 (Pairwise comparison matrix (PCM)) *Let us denote the number of criteria (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 all i and j) and reciprocal ($1/a_{ij} = a_{ji}$ for all i and j).*

The element a_{ij} of a PCM shows how many times item i is better/stronger/more important than item j. However, when a decision maker fills in all $n(n-1)/2$ elements (the elements above the principal diagonal, because of the reciprocity) there can be some kind of contradiction, a certain inconsistency in the PCM.

Definition 26 (Consistent PCM) *A PCM is said to be consistent if $a_{ik} = a_{ij}a_{jk}$ for all i, j, k. If a PCM is not consistent, then it is called inconsistent.*

Naturally, there are several degrees of inconsistency, which leads to the deeply analyzed problem of different inconsistency indices (Brunelli, 2018), their properties (Brunelli, 2017), and the appropriate recommended thresholds (Amenta et al., 2020). Although, many measures have been proposed, the most widely used one is probably still Saaty's Consistency Ratio (CR) (Saaty, 1977).

Definition 27 (Consistency Ratio (CR)) *The CR of an $n \times n$ PCM A is defined as follows:*

$$CR = \frac{CI}{RI}, \tag{27}$$

where CI stands for Consistency Index, that is:

$$CI = \frac{\lambda_{\max} - n}{n - 1}, \tag{28}$$

where $\lambda_{\max}$ is the principal eigenvalue of the matrix A, and RI is the Random Index, which is the average CI obtained from a sufficiently large set of randomly generated PCMs of size n.

Probably the two most commonly used techniques to calculate a weight vector (prioritization vector) from a PCM that shows the importance of compared items, are the logarithmic least squares (LLSM) (Crawford and Williams, 1985) and the eigenvector (EV) (Saaty, 1977) methods.

Definition 28 (Logarithmic Least Squares Method (LLSM)) *Let A be an $n \times n$ PCM. The weight vector w of A determined by the LLSM is given as follows:*

$$\min_{w} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\ln(a_{ij}) - \ln\left(\frac{w_i}{w_j}\right) \right)^2, \tag{29}$$

where w_i is the ith coordinate of w.

Definition 29 (Eigenvector (EV) Method) *Let A be an $n \times n$ PCM. The weight vector w of A determined by the EV method is defined as follows:*

$$A \cdot w = \lambda_{\max} \cdot w, \tag{30}$$

where the componentwise positive principal eigenvector w is unique up to a scalar multiplication.

These two methods are shown to be indeed similar in their results, however LLSM has significantly lower computational time (Dong et al., 2008).

In several situations some comparisons are absent, which may happen because the decision makers do not have time, willingness or possibility to make all of them, data has been lost, the direct comparison is simply impossible (for instance in sports (Bozóki et al., 2016)), etc. When a PCM has missing elements, it is said to be an incomplete PCM (IPCM).

The LLSM and EV weight calculation methods can be generalized to the incomplete case as well, when the LLSM's optimization problem (Equation 29) includes only the known elements of the matrix, while the EV method is based on the CR-minimal completion (CREV) of the PCM and its principal right eigenvector (Shiraishi et al., 1998; Shiraishi and Obata, 2002).

In this paper we analyze different kinds of filling in structures of IPCMs, thus we assume that the set of pairwise comparisons to be made can be chosen. We also heavily rely on the graph representation of IPCMs (Gass, 1998).

Definition 30 (Representing graph of an IPCM) *An IPCM A is represented by the undirected graph $G = (V, E)$, where the V vertex set of G corresponds to the alternatives (criteria) of A, and there is an edge in the edge set E of G if and only if the appropriate element of A is known.*

We assume that no prior information is available about the items to be compared, thus in the examined filling in patterns we do not distinguish between the isomorphic representing graphs. The optimal solutions of both above-mentioned weight calculation techniques for IPCMs (LLSM and CREV) are unique if and only if the representing graph is connected (Bozóki et al., 2010).

Definition 31 (Connected graph) *In an undirected graph, two vertices u and v are called connected if the graph contains a path from u to v. A graph is said to be connected if every pair of vertices in the graph is connected.*

The smallest connected systems are associated with spanning trees, which contain $n-1$ edges for n vertices.

Definition 32 (Spanning tree) *Let $G = (V, E)$ be a connected graph. $G' = (V, E')$ is a spanning tree of G if $E' \subseteq E$ is a minimal set of edges that connect all vertices of G.*

An IPCM represented by a spanning tree always can be complemented to a consistent PCM, however, the results based on such an IPCM are usually extremely unreliable. The special importance of spanning trees is emphasized by the combinatorial weight calculation method (Tsyganok, 2010), which is built on the weight vectors obtained from all different spanning trees. This technique provides the same prioritization vector as the LLSM, if we use the geometric mean, both for PCMs (Lundy et al., 2017) and IPCMs (Bozóki and Tsyganok, 2019).

The results obtained by any weight calculation methods for IPCMs is strongly dependent on the number of known comparisons, namely the number of edges of the representing graph (e), and the arrangements of these known elements. Several properties have been examined in connection with the positioning of the known items, among which (some sense of) regularity of comparisons seems to be an especially important one (Szádoczki et al., 2020; Wang and Takahashi, 1998; Kułakowski et al., 2019), which can be also described by the representing graph.

Definition 33 (k-regularity) *A graph is called k-regular if every vertex has k neighbours, which means that the degree of every vertex is k.*

When both the number of vertices (n) and the level of regularity (k) are odd, k-regularity is not possible. However, the graphs that are the closest to k-regularity in this case are called k-quasi-regular graphs (Szádoczki et al., 2022).

Definition 34 (k-quasi-regularity) *A graph is called k-quasi-regular if exactly one vertex has degree $k+1$, and all the other vertices have degree k.*

In decision making the (quasi-)regularity of the representing graph ensures a certain level of symmetry, as every item is compared to the (approximately) same number of elements. This kind of property is also required in other fields, for instance, in the design of some sport tournaments (Csató, 2017).

We have only focused on multiplicative PCMs in the above definitions in this section, however, one can make the appropriate transformations to get an additive or a reciprocal (fuzzy) PCM from those (Brunelli, 2014). Thus, we would like to emphasize that all of our findings in the sections below are true for those types of matrices as well.

4.3 Methodology

Our aim is to find the filling structures that provide the closest results to the complete case for a given (n, e) pair, number of alternatives (criteria) and comparisons. As it is assumed that we do not have any prior information, and so, the different items are not distinguished, we used Wolfram Mathematica (Wolfram Research, 2021), nauty and Traces (McKay and Piperno, 2014), and IGraph/M (Horvát, 2020) to generate every non-isomorphic (representing) graph for the examined (n, e) pairs. Our extensive numerical simulations are based on the filling patterns related to these graphs. The used methods are similar to Szádoczki et al. (2023), however they only focus on some special cases based on smaller samples and even compare representing graphs with different number of edges (comparisons), while we in the current paper compare all the possible filling structures for a given (n, e) pair with a more general approach.

In order to measure the differences between the weight vectors, we apply commonly used cardinal and ordinal indicators, the Euclidean distance (d_{euc}) and the Kendall rank correlation coefficient (Kendall's τ), respectively, which are defined as follows.

$$d_{euc}(u, v) = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2} \tag{31}$$

$$\tau(u,v) = \frac{n_c(u,v) - n_d(u,v)}{n(n-1)/2} \tag{32}$$

where u denotes the weight vector obtained from a certain filling structure and v is the weight vector computed from the complete PCM. u and v are normalized by $\sum_{i=1}^{n} u_i = 1$, and $\sum_{i=1}^{n} v_i = 1$, respectively, and v_i and u_i denote the ith element of the appropriate vectors. $n_c(u,v)$ and $n_d(u,v)$ are the number of concordant and discordant pairs of the examined vectors, respectively. The range of the Kendall's τ is $[-1,1]$, and considering the notation in Equation 32, a higher value indicates a better performance of the given filling pattern. However, in this case (Equation 31) the Euclidean distance can be interpreted as an error, thus its smaller level is preferred. It is also worth mentioning that besides these, Szádoczki et al. (2023) used many different kinds of measures for the special cases examined by them, and all of those provided similar results.

An instrumental part of our methodology is to determine the sample size needed in the simulations, which is based on a certain form of Chebyshev's inequality (Steliga and Szynal, 2010; Saw et al., 1984) that leads to the weak law of large numbers.

Proposition 1 (Weak law of large numbers) *Let (ξ_k) be independent and identically distributed random variables with finite standard deviation (σ), and let $E(.)$ denote the expected value operator. Then Equation 33 follows for all $\varepsilon > 0$:*

$$P\left(\left|\frac{\sum_{k=1}^{n}\xi_k}{n} - E(\xi_1)\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0, \tag{33}$$

where the last part of the expression means that the limit of the probability is 0 as n goes to infinity.

The $\alpha = \sigma^2/n\varepsilon^2$ notation defines the significance level of our results, while ε is the margin of error. We estimated the standard deviations of the Euclidean distances and the Kendall's τ measures for the different filling structures in our simulation and used an upper bound on it. Based on this method we applied a sample size of one million elements for every (representing) graph, which results in (as an upper bound as well)

- $\alpha = 0.01$ and $\varepsilon = 0.0005$ for the computed Euclidean distances,
- and $\alpha = 0.05$ and $\varepsilon = 0.001$ for the calculated Kendall's τ measures.

As we mentioned earlier, the result of the EV weight calculation technique is similar to the LLSM, but its computational time is larger. This pattern is even stronger in the case of incompleteness (for CREV and incomplete LLSM, see for instance Csató (2013)), thus due to the large sample sizes, in our simulations we mainly focus on the LLSM weight calculation technique. The results of the CREV method were computed for smaller cases ($n \leq 5$) with a sample size of 500 000 as well, however, the ranking of filling patterns were always the same, and the indicators were almost always closer to the LLSM outcomes than the margin of error, thus we decided not to present them in much detail.

The process of the simulation for a given (n, e) pair consisted of the following steps:

1. n random weights (in general they are denoted by w_i) were generated, where $w_i \in [1, 9]$ is a uniformly distributed random real number for all $i \in 1, 2, \ldots, n$. We calculated random $n \times n$ complete and consistent PCMs, where the elements of the matrices were given by Equation 34.

$$a_{ij} = w_i / w_j \tag{34}$$

2. Then three different perturbations of the items of consistent PCMs were used to get inconsistent matrices with three well-distinguishable inconsistency levels. These levels are denoted by weak, modest and strong given by Equations 35, 36 and 37.

$$\hat{a}_{ij}^{weak} = \begin{cases} a_{ij} + \Delta_{ij} & : a_{ij} + \Delta_{ij} \geq 1 \\ \frac{1}{1 - \Delta_{ij} - (a_{ij} - 1)} & : a_{ij} + \Delta_{ij} < 1 \end{cases} \qquad \Delta_{ij} \in [-1, 1] \tag{35}$$

$$\hat{a}_{ij}^{modest} = \begin{cases} a_{ij} + \Delta_{ij} & : a_{ij} + \Delta_{ij} \geq 1 \\ \frac{1}{1 - \Delta_{ij} - (a_{ij} - 1)} & : a_{ij} + \Delta_{ij} < 1 \end{cases} \qquad \Delta_{ij} \in \left[-\frac{3}{2}, \frac{3}{2} \right] \tag{36}$$

$$\hat{a}_{ij}^{strong} = \begin{cases} a_{ij} + \Delta_{ij} & : a_{ij} + \Delta_{ij} \geq 1 \\ \frac{1}{1 - \Delta_{ij} - (a_{ij} - 1)} & : a_{ij} + \Delta_{ij} < 1 \end{cases} \qquad \Delta_{ij} \in [-2, 2] \tag{37}$$

Where $\hat{a}_{ij}^{weak}$, $\hat{a}_{ij}^{modest}$ and $\hat{a}_{ij}^{strong}$ are the elements of the perturbed PCMs, a_{ij} is the element of the consistent PCM, $a_{ij} \geq 1$ (we only perturb the elements above one and keep the reciprocity of the matrices), and Δ_{ij} is uniformly distributed in the given ranges. This perturbation method is able to produce ordinal differences as well (when $\hat{a}_{ij} < 1$). It is important to mention that we account for the contrast that can be examined above and below 1, thus our perturbed data is uniformly distributed around the original element on the scale presented by Figure 28, which also contains two examples. Our perturbation method aims to provide three different and meaningful inconsistency levels and it is, indeed, correlated with the Consistency Ratio (CR), as it is shown in Figure 29. We tested several combinations of parameters, and found that these resulted in the most relevant levels of CR.

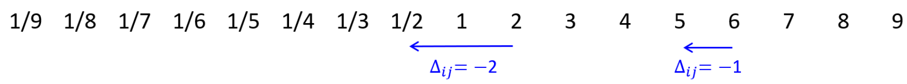


Figure 28: The ratio scale $1/9, \ldots, 9$ and the perturbation of elements according to (35)–(37).

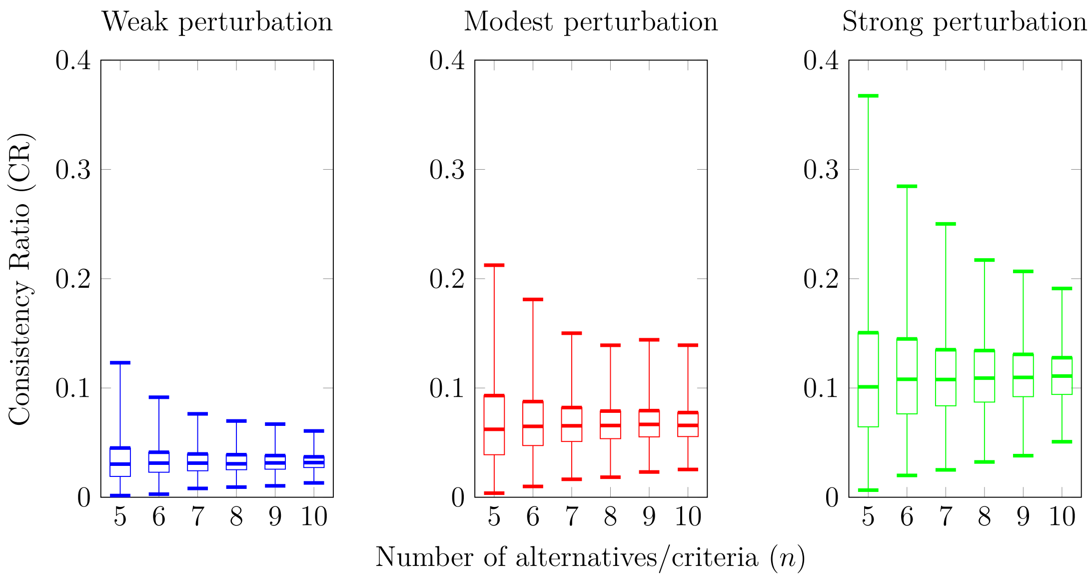


Figure 29: The relation between CR and our element-wise perturbation via Box plots. Each Box plot is based on 1000 randomly generated perturbed PCMs.

3. We deleted the respective elements of the matrices in order to get the filling structure that we were examining, and applied the LLSM (and CREV in the case of $n \leq 5$) technique(s) to obtain the weights. The certain models' Euclidean distances and Kendall's τ measures were computed with respect to the weights that were calculated from the complete inconsistent matrices. The analyzed filling in patterns included all of those that can be represented by connected non-isomorphic graphs with parameters (n, e).

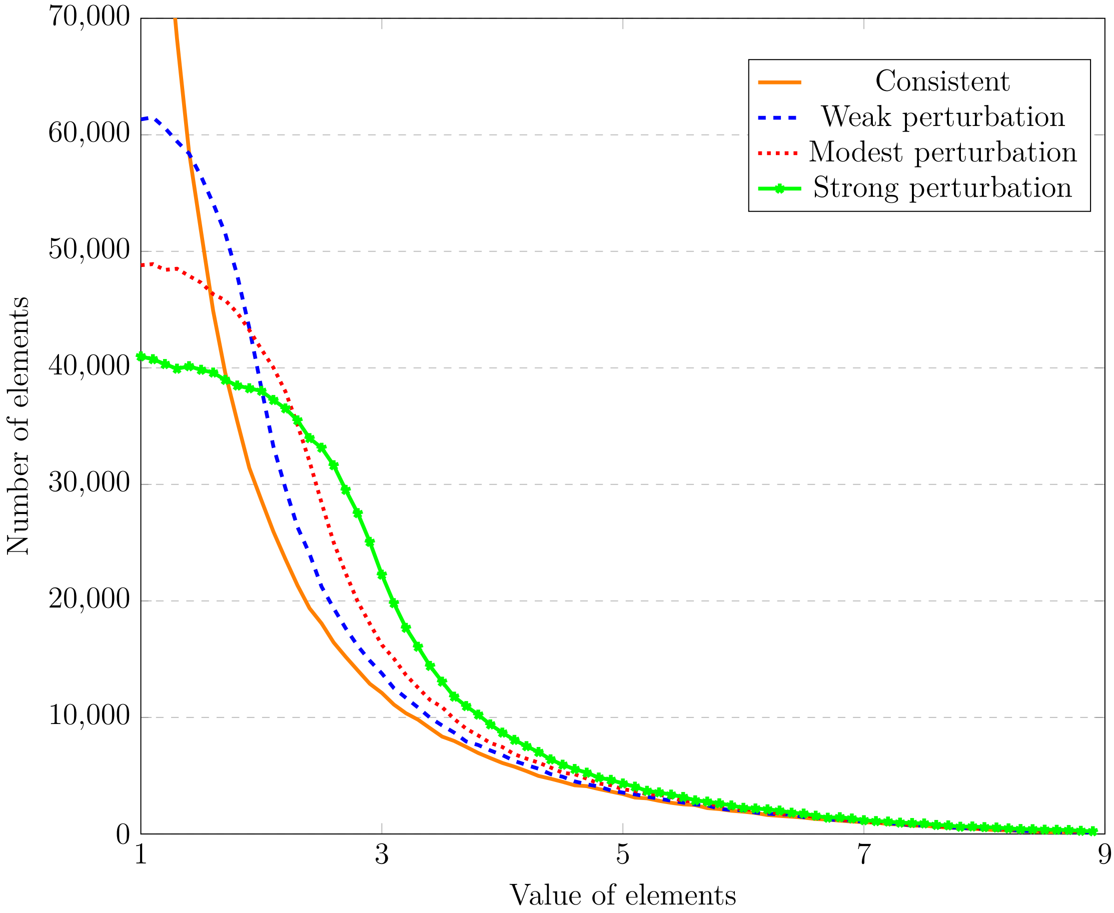


Figure 30: The histograms of the ≥ 1 elements of PCMs in case of different perturbations based on a sample of 1 million elements (with a 0.1 bin width).

4. We repeated steps 1-3 for 1 000 000 times for every level of inconsistency (thus altogether we examined 3 000 000 PCMs for a given (n, e) pair). Finally, we saved the mean of Euclidean distances and Kendall's τ measures for the different filling in patterns.

Remark 2 *The distribution of the elements of complete PCMs is independent of n. This property holds for both consistent and perturbed complete PCM cases.*

The reason behind this is that, in the simulations at first the elements of a given matrix are generated independently from n, and then they are placed into the $n \times n$ PCM. The histograms of the complete PCM elements above 1 in the different perturbation cases, based on samples containing 1 million elements each, are presented in Figure 30 (with a 0.1 bin width).

According to the histograms, a higher level of perturbation (inconsistency) leads to a higher chance to have large (extreme) matrix elements.

4.4 Results

We would like to emphasize that all of the results (and graph recommendations) presented in this section are under the following crucial assumptions.

1. We can choose the comparisons that should be made (they are not given a priori).

2. An 'optimal' graph is the one that provides the closest LLSM weight vector on average to the one calculated from the complete matrix according to the measures presented in Section 4.3.

3. There is not any prior information about the items that should be compared, thus we can handle them in a symmetric way. This also means that the 'reliableness' of every comparison is assumed to be the same.

Naturally, if one or more of the assumptions above do not hold – for instance comparing to another benchmark instead of the complete PCM (e.g., the closest complete consistent PCM) – that could lead to other outcomes.

It is important to note that the interesting cases for our research start above three alternatives (n), as in the case of $n = 2$ and $n = 3$ there is always one non-isomorphic (representing) graph for every relevant pair of (n, e) as it is shown in Figure 31.

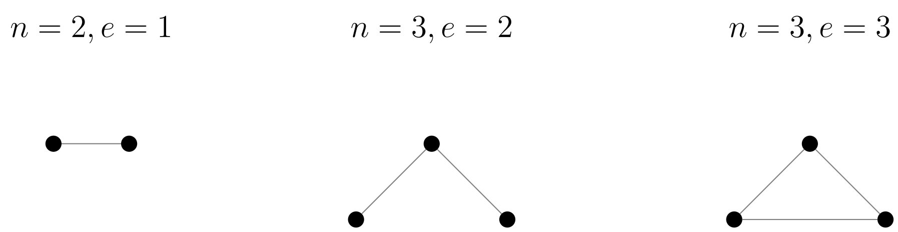


Figure 31: The connected non-isomorphic representing graphs for $n \leq 3$.

The $n = 4$ case also contains only a few possibilities, but it can be interesting in a decision problem, when there are several criteria and four alternatives, and it helps to understand the results for larger examples as well. Figure 32 presents the connected representing graphs for $n = 4$ as a GRAPH of graphs. The value of e is shown in every row of the GRAPH, in which an EDGE between two NODEs (=graphs) denotes that we can obtain one graph from the other one by adding (or deleting) exactly one edge. The GRAPH of graphs in Figure 32 is a 4-partite GRAPH with a further specific property, namely, that EDGEs go between levels k and $k + 1$ only ($k = 1, 2, 3$). Note that if all EDGEs would be oriented 'downwards' (i.e., the addition of an edge in the graph of comparisons), a partially ordered set of graphs (of comparisons) would be resulted in. We denote the graph that provided the weight vectors with the smallest average Euclidean distance and the largest average Kendall's τ respect to the vectors calculated from the complete case by green background color for every e. If two optimal graphs are connected with an EDGE, then it is a partial optimal sequence, and the respective EDGE is also denoted by green. It is important to note that the relevant values for e (the number of comparisons) are between $n-1$ (spanning trees) and $n(n-1)/2$ (complete graphs representing complete PCMs).

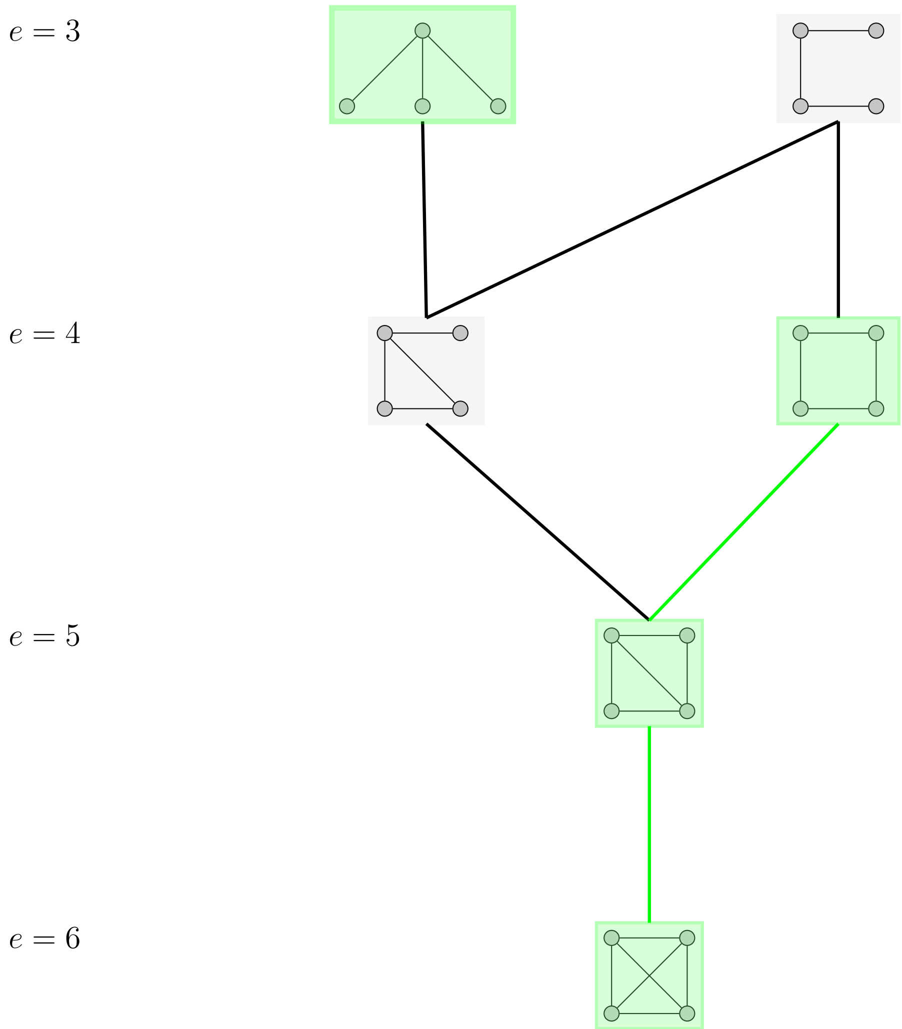


Figure 32: The GRAPH of graphs for $n = 4$, the optimal graph for a given e is highlighted by green, EDGEs between optimal graphs are colored green.

Among the spanning trees the star graph provided the smallest errors (Euclidean distances) and the largest Kendall's τ measures. This is not connected to the optimal graph with four edges, which is the 2-regular cycle. However, from this point on the optimal graphs result in an optimal filling sequence. This is not surprising, as for $e = 5$ and $e = 6$ there is only one possible non-isomorphic representing graph, but this example probably helps to understand the following cases. Tables 10 and 11 present the results provided by the graphs with $(n = 4, e = 3)$ and $(n = 4, e = 4)$ respectively, in the case of the different perturbation levels. The name of the optimal graph, and the best values in every column are highlighted with green background color.

graph	Weak		Modest		Strong	
	d_{euc}	Kendall's τ	d_{euc}	Kendall's τ	d_{euc}	Kendall's τ
Star graph	0.0918	0.7306	0.1293	0.6639	0.1620	0.6164
Line graph	0.0967	0.7194	0.1361	0.6501	0.1701	0.6020

Table 10: The average Euclidean distances and Kendall's τ measures for the graphs with $(n = 4, e = 3)$ in the case of the different perturbation levels.

graph	Weak		Modest		Strong	
	d_{euc}	Kendall's τ	d_{euc}	Kendall's τ	d_{euc}	Kendall's τ
Not regular graph	0.0650	0.8027	0.0920	0.7496	0.1156	0.7111
2-regular graph	0.0543	0.8216	0.0771	0.7705	0.0970	0.7328

Table 11: The average Euclidean distances and Kendall's τ measures for the graphs with $(n = 4, e = 4)$ in the case of the different perturbation levels.

Based on Tables 10 and 11, one can observe that for a given (n, e) pair, the same graphs provided the best results on average for the examined measures for every perturbation level. There are indeed significant differences between the examined graphs (the margin of error is 0.0005 for the Euclidean distances and 0.001 for the Kendall's τ measures). It is also easy to see that a stronger perturbation results in higher errors, while an additional edge leads to smaller distances and higher ordinal correlations. Figure 33 presents the relation between the number of comparisons (e) and the analyzed cardinal (d_{euc}) and ordinal (Kendall's τ) measures, which can help practitioners to determine the minimal sufficient number of comparisons in a given problem. Note that Figure 33 shows the results for the optimal graphs for every e, thus one optimal value is not necessarily reachable from the previous one, only in the case of partial optimal sequences.

If we know in advance that the decision maker is willing to provide exactly $e = n - 1 = 3$ comparisons, then, according to Figure 32, we recommend the star graph, i.e., filling in one (e.g. the first) row/column of the pairwise comparison matrix, namely elements a_{12}, a_{13} and a_{14} (in any order), also summarized in Table

12. It is worth noting that in this case Assumption 3 has a special importance as all items are compared to one (pivotal) item.

	1	2	3	4
1		#1'	#2'	#3'
2				
3				
4				

Table 12: Filling in sequence for $n = 4$, $e = 3$. Orders with ' are interchangeable.

If we assume that the decision maker is willing to provide more than three comparisons, the optimal filling in sequence is $\{a_{12}, a_{23}, a_{34}, a_{14}\}$ (the first four comparisons can be made in any order), followed by a_{13} and finally a_{24}, also summarized in Table 13.

	1	2	3	4
1		#1'	#5	#4'
2			#2'	#6
3				#3'
4				

Table 13: Filling in sequence for $n = 4$, $e > 3$. Orders with ' are interchangeable.

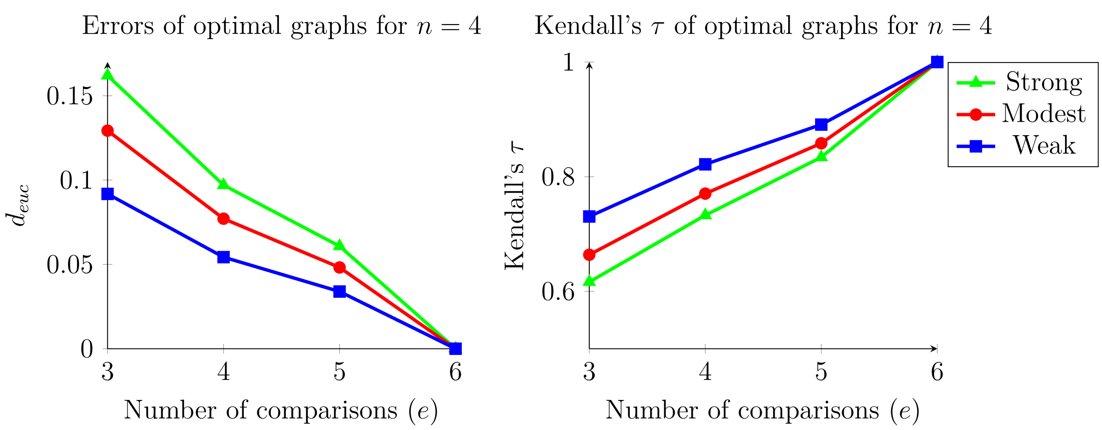


Figure 33: The relation between the number of comparisons (e), the errors (Euclidean distances) and Kendall's τ measures of optimal graphs for $n = 4$.

For larger number of alternatives (criteria, n), the possible number of connected graphs increases quickly, thus it is even more relevant to determine the optimal filling structure. In the case of $n = 5$, there are 21 connected graphs altogether. Their 7-partite GRAPH of graphs can be seen in Figure 34, using the same notations as before.

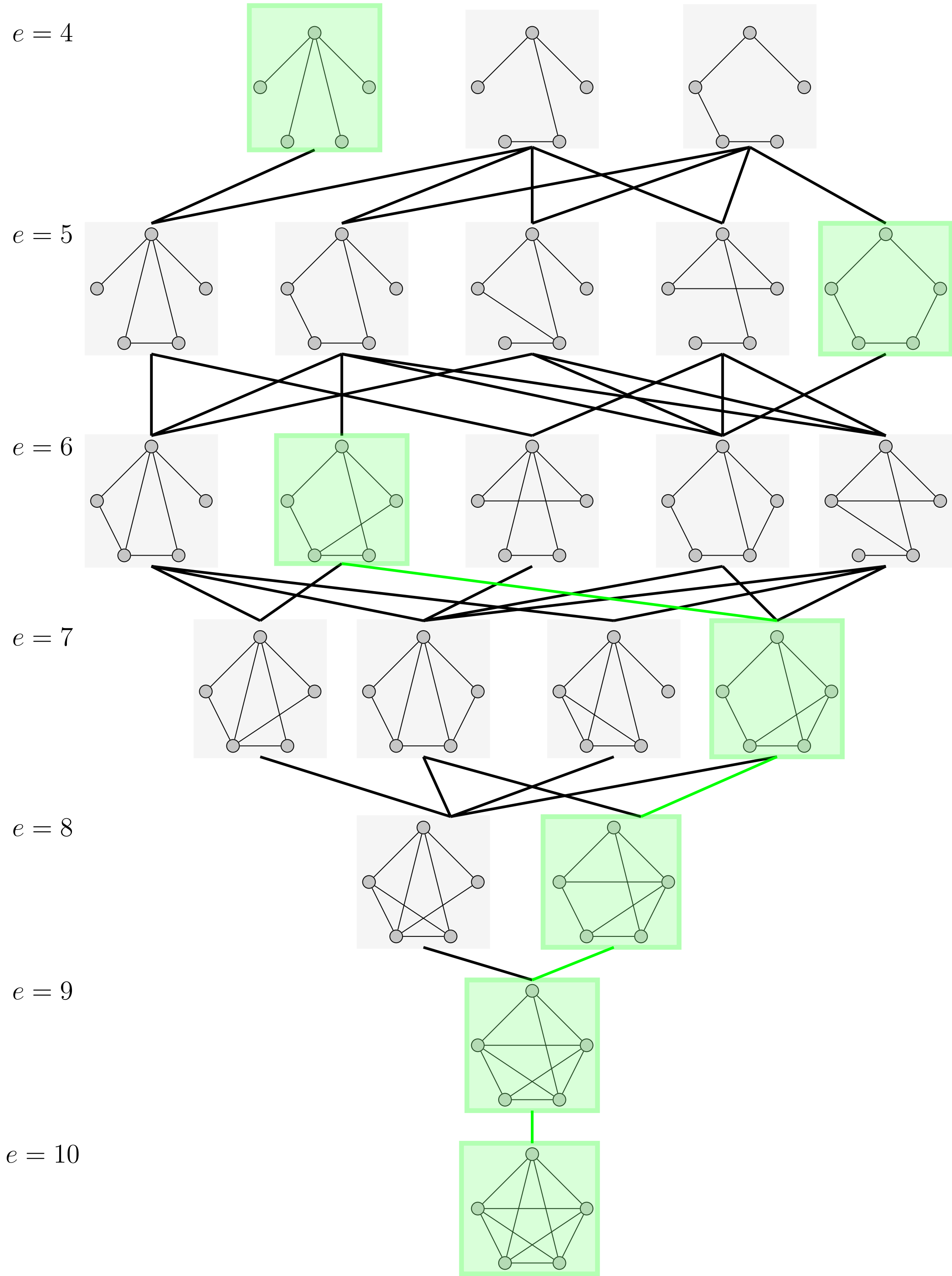


Figure 34: The GRAPH of graphs for $n = 5$, optimal graphs are highlighted by green.

One can see many similarities with the previous outcomes. The star graph resulted in the smallest Euclidean distance and the largest Kendall's τ measure among the spanning trees, once again. It is not connected to the optimal graph with $e = 5$, which is the 2-regular cycle, as before. The next optimal graph with $e = 6$ is not connected to the cycle, as well, however, from that point on there is a partial optimal sequence to the complete filling of the represented PCM. Somewhat surprisingly, the graphs providing the smallest Euclidean distances resulted in the largest Kendall's τ for every single case, except for $e = 8$. However, in that case the difference between the Kendall's τ measures for the two possible graphs is within the margin of error, thus we highlighted the graph that is better according to the Euclidean distance, which is better in both indicators if we consider the CREV weight calculation technique. It is worth mentioning that this graph is the 3-quasi-regular graph on $n = 5$.

Figure 35 shows the relation between the number of comparisons (e) and the analyzed measures for $n = 5$ in the case of optimal graphs. One optimal value is not necessarily reachable from the previous one, as before. Minimal thresholds could be determined for the number of comparisons based on this figure for certain decision problems.

If we know in advance that the decision maker is willing to provide exactly $e = n - 1 = 4$ comparisons, then, according to Figure 34, we recommend the star graph (with special attention to Assumption 3), i.e., filling in one (e.g. the first) row/column of the pairwise comparison matrix, namely elements a_{12}, a_{13}, a_{14} and a_{15}, also summarized in Table 14.

	1	2	3	4	5
1		#1'	#2'	#3'	#4'
2					
3					
4					
5					

Table 14: Filling in sequence for $n = 5$, $e = 4$. Orders with ' are interchangeable.

In the case the decision maker is willing to provide exactly $e = n = 5$ comparisons, then we should make the comparisons along an n-cycle, e.g., $\{a_{12}, a_{23}, a_{34}, a_{45}$ and $a_{15}\}$ (the five comparisons can be made in any order), also summarized in Table 15.

When we can assume that the decision maker is willing to provide more than five comparisons, the optimal filling in sequence is $\{a_{14}, a_{15}, a_{24}, a_{25}, a_{34}, a_{35}\}$ (the first six comparisons can be made in any order), followed by a_{12}, a_{13}, a_{45}, and finally a_{23}, also summarized in Table 16.

	1	2	3	4	5
1		#1'			#5'
2			#2'		
3				#3'	
4					#4'
5					

Table 15: Filling in sequence for $n = 5$, $e = 5$. Orders with ' are interchangeable.

	1	2	3	4	5
1		#7	#8	#1'	#2'
2			#10	#3'	#4'
3				#5'	#6'
4					#9
5					

Table 16: Filling in sequence for $n = 5$, $e > 5$. Orders with ' are interchangeable.

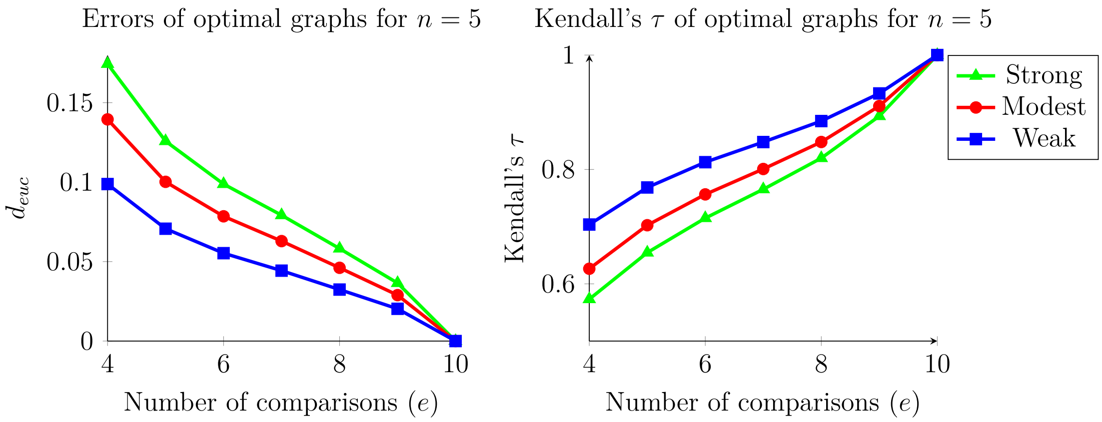


Figure 35: The relation between the number of comparisons (e), the errors (Euclidean distances) and Kendall's τ measures of optimal graphs for $n = 5$.

Finally, for $n = 6$, there are 112 possible connected (representing) graphs. Figure 36 shows the 11-partite GRAPH of graphs for this case, however, in order to keep it visible, we only denote the possible graphs with a vertex, and present the optimal cases in detail in Figure 37. For $e = 12$ and $e = 13$ the results are close to each other, and some of the differences of the Kendall's τ measures are also smaller, than the margin of error. Here the best graph according to the Euclidean distance and the Kendall's τ are different as well. However, we highlighted the graphs which were at least second according to at least one indicator by a lighter green color. These highlighted graphs for a given e practically provide the same results. As there is always a unique optimal graph according to the Euclidean distance, we denoted those with an E. We have not highlighted the EDGEs by green color on this part of the GRAPH of graphs, because of the similar results (ties). In Figure 37 for $e = 12$

and 13 the graphs that provided the best results according to the Euclidean distance are presented.

If we know in advance that the decision maker is willing to provide exactly $e = n - 1 = 5$ comparisons, then, according to Figure 36, we recommend the star graph (with special attention to Assumption 3), i.e., filling in one (e.g. the first) row/column of the pairwise comparison matrix, namely elements $a_{12}, a_{13}, a_{14}, a_{15}$ and a_{16}, also summarized in Table 17.

	1	2	3	4	5	6
1		#1'	#2'	#3'	#4'	#5'
2						
3						
4						
5						
6						

Table 17: Filling in sequence for $n = 6$, $e = 5$. Orders with ' are interchangeable.

If the decision maker is willing to provide more than five comparisons, the recommended filling in sequence is $\{a_{14}, a_{15}, a_{24}, a_{26}, a_{35}, a_{36}\}$ (the first six comparisons can be made in any order), followed by $a_{25}, a_{34}, a_{16}, a_{12}, a_{46}, a_{23}, a_{45}, a_{56}$, and finally a_{13}, also summarized in Table 18.

	1	2	3	4	5	6
1		#10	#15	#1'	#2'	#9
2			#12	#3'	#7	#4'
3				#8	#5'	#6'
4					#13	#11
5						#14
6						

Table 18: Filling in sequence for $n = 6$, $e > 5$. Orders with ' are interchangeable.

Since there is no path along all the optimal graphs, the filling in sequence above includes as many as possible. The remaining EDGEs are colored with orange in Figure 36, and we should note that the other included graphs are as close to optimal ones as possible.

One can observe many similarities with the earlier outcomes in connection with the concrete graphs, and the pattern of optimal graphs as well. Among the spanning trees, the star graph provided the best results according to both measures again. For $e = 6$ the 2-regular cycle turned out to be the optimal case, just as earlier. The optimal graphs with $e = 5, 6$ and 7 are not connected, but from that point on we can determine an optimal filling in sequence to the complete graph (if we consider all the light green cases optimal).

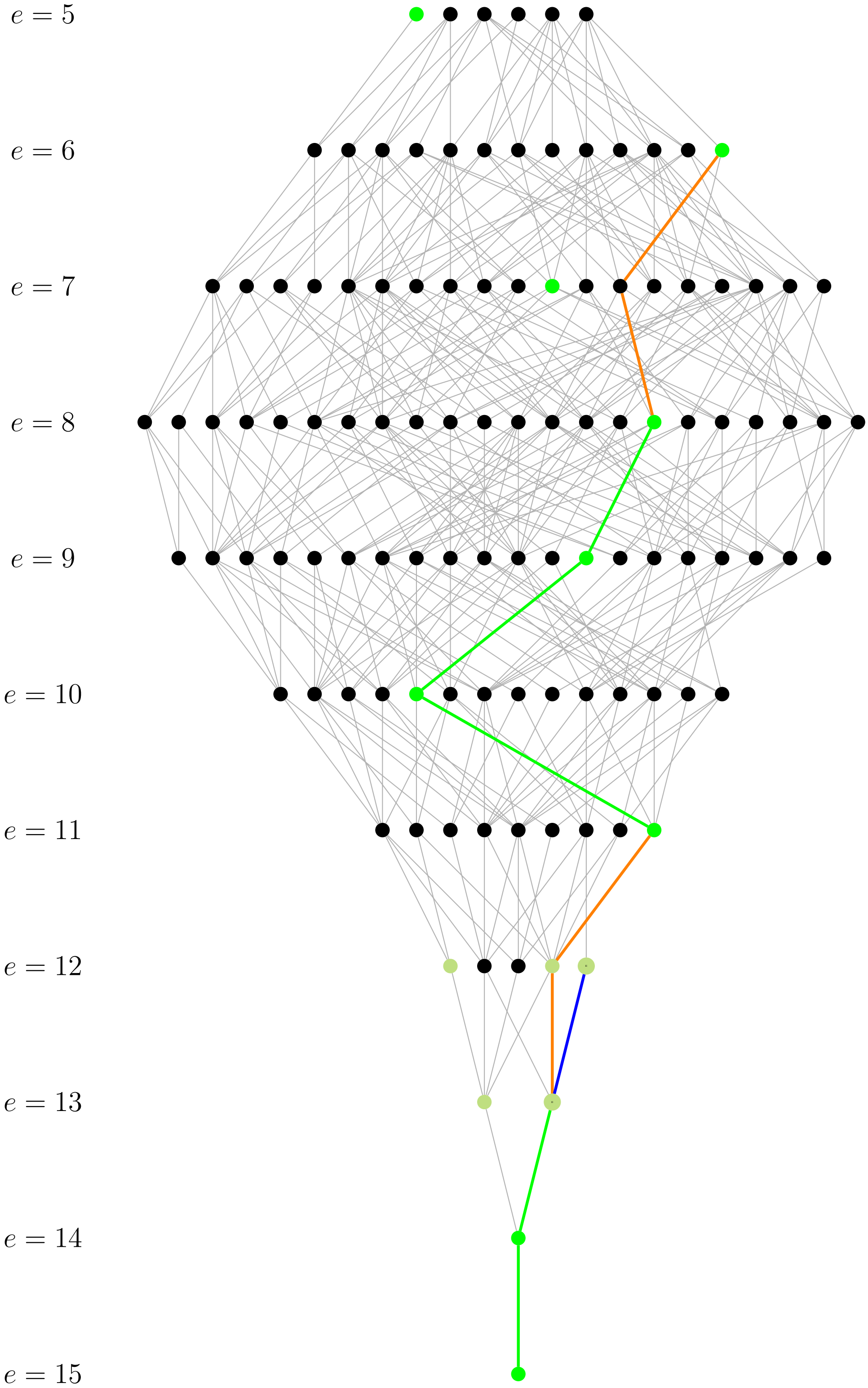


Figure 36: The GRAPH of graphs for $n = 6$, optimal graphs (=NODEs) are colored green, EDGEs between optimal graphs are colored green, too.

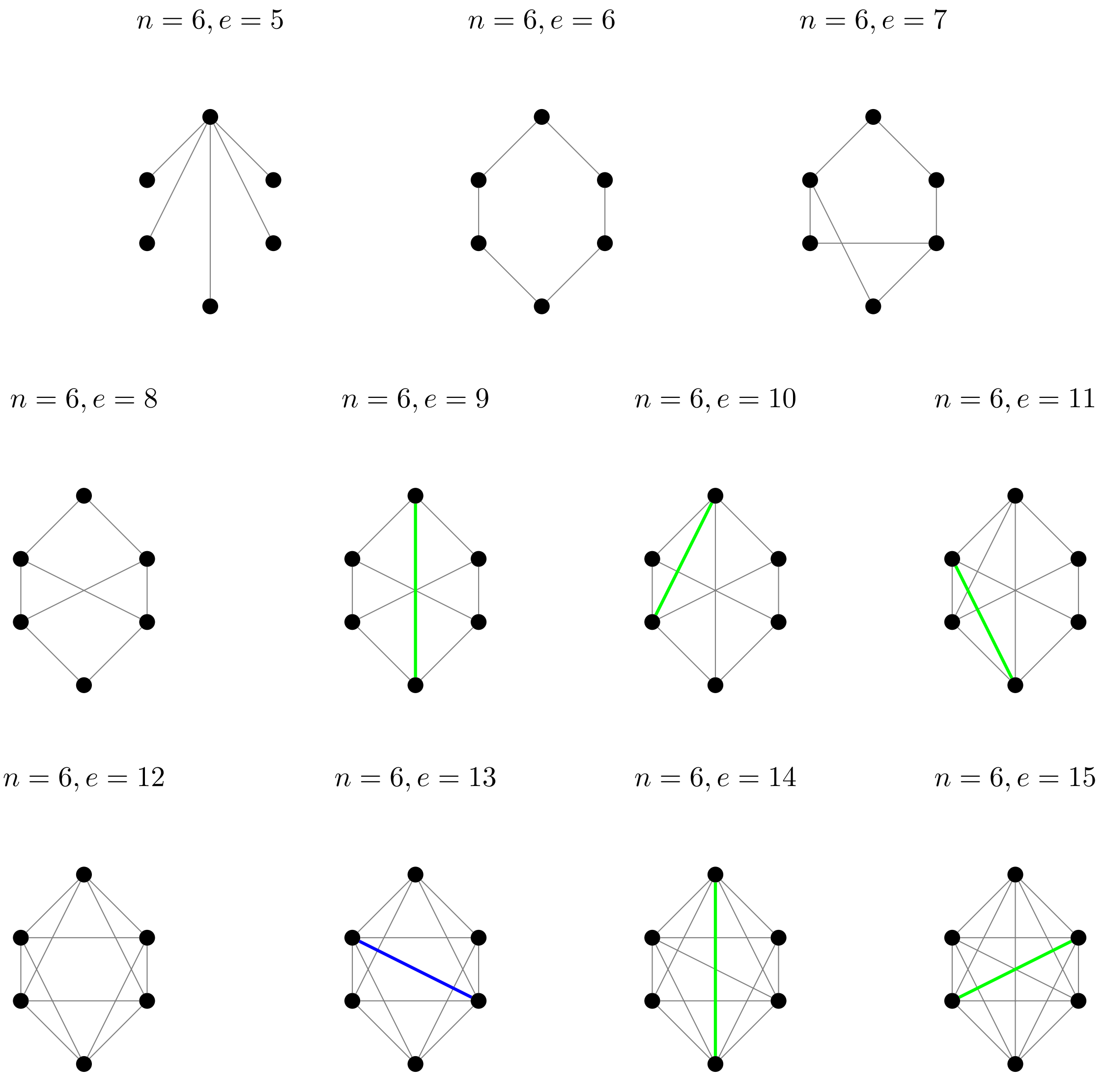


Figure 37: The optimal graphs related to the green NODEs in Figure 36. The second row shows a partial optimal filling sequence corresponding to the one in Figure 36, these graphs can be reached from each other. The additional comparisons are highlighted in every step.

Moreover, for $e = 9$ the optimal graph is the single bipartite 3-regular graph on six vertices, while for $e = 12$ the highlighted graph, which provided the best results according to the Euclidean distance and the second best according to the Kendall's τ, is the only 4-regular graph on six vertices. Based on the simulations, we can make several important remarks.

Remark 3 *The star graph provided the best results according to both measures for all examined $(n, e = n - 1)$ cases. Thus we can say that it is an optimal structure, intuitively it keeps this property for larger cases (n), as well.*

Remark 4 *For the $(n, e = k \cdot n/2)$ examples, the optimal graph is always a k-regular graph. Furthermore, k-quasi-regular graphs are optimal as well. One can say that regularity is indeed important in a more general way, as in all of the examined instances, the degree of different vertices (the number of comparisons) are as close as possible.*

Remark 5 *The optimal graphs are always bipartite graphs, or the closest ones to that.*

The analyzed indicators for optimal graphs in the case of different number of comparisons (e) can be seen in Figure 38 for $n = 6$. Again, it can serve as a guide for practitioners.

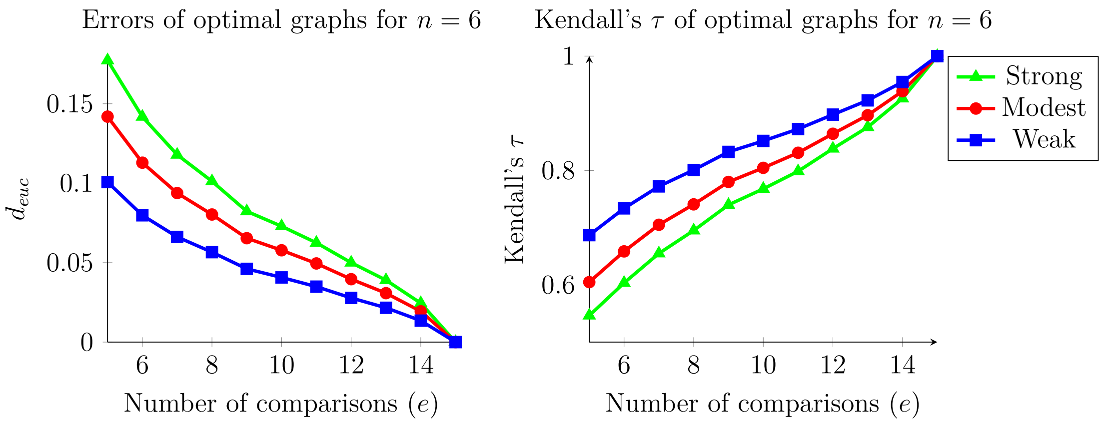


Figure 38: The relation between the number of comparisons (e), the errors (Euclidean distances) and Kendall's τ measures of optimal graphs for $n = 6$.

All of our simulation results provided optimal filling structures (representing graphs) for the examined (n, e) pairs, as well as (partial) optimal filling sequences. The outcomes show indeed similar patterns for different parameters, and can support both applications and theoretical studies.

Finally, it is worth mentioning that although a practical MCDM problem usually has several hierarchical levels, – thus many PCMs have to be filled in and the overall number of comparisons is high, – the size of the matrices usually do not exceed 6×6 (Ábele-Nagy et al., 2018). Furthermore, based on our research Gyarmati et al. (2023) found that the exact same graphs are optimal in the case of fundamentally different models as well, which are based on paired comparisons, i.e., the Bradley–Terry and Thurstone models. These findings suggest that our results are rather general and not model-specific.

4.5 Conclusion and further research

In this paper we analyzed all possible filling structures of incomplete pairwise comparison matrices when there is no prior information available for the compared items, in the case of at most six alternatives (criteria). The study heavily relied on the representing graphs of pairwise comparisons as well as on extensive numerical simulations with large samples. We compared the weight vectors (calculated by the incomplete LLSM) related to the certain filling patterns and compared them based on their Euclidean distance and Kendall's τ measure with the weights obtained from the complete case.

We found that there is a strong connection between the examined cardinal and ordinal indicators, thus we could determine the best filling structure for a given number of alternatives and comparisons, which is a significant finding itself. However, one of the main contributions of the paper is that many of these optimal graphs resulted in optimal filling in sequences as first in the literature illustrated by different paths in the examined GRAPHs of graphs.

The filling structure represented by a star graph turned out to be optimal among the graphs (filling patterns) with the same cardinality (spanning trees). Regular graphs also seem to provide optimal solutions, and regularity is a common property of the optimal cases in a more general sense.

Both theorists and practitioners can utilize our findings not just to apply the optimal filling structure in their problems, but also to use the optimal filling sequences in decision making problems where the decision maker can abandon the problem at any period of the process. Furthermore, our results on the difference between the optimal patterns and the complete case for different number of comparisons can serve as a guide to determine the minimal sufficient number of comparisons for a given problem.

A future research can investigate the certain comparisons that decrease the errors the most during the filling in process. When should we stop to ask even more questions from the decision maker? Do the last few comparisons provide significant information? How does this problem relate to the representing graph?

Empirical PCMs may differ from simulated ones, and many collections of those matrices (even with the complete filling in order) are available (Bozóki et al., 2013), thus in a future research it is important to test our findings on empirical matrices as well.

Naturally larger cases, other weight calculation methods and different distance measures can be further investigated as well. Are the findings remain true for a large number of alternatives? How much are they dependent on the used techniques and measures? What can we say when some prior information, for instance, the best or the worst alternatives, perhaps both, are known (Rezaei, 2015; Mustajoki et al., 2005; Edwards and Barron, 1994; von Winterfeldt and Edwards, 1986)?

Our results can be useful in other areas as well, for instance, in designing sport tournaments. If we would like to plan the different rounds, we should make a number of comparisons simultaneously. This leads to the general question: besides optimal direct sequences, how does the optimal graphs include each other (indirectly)?

References

Ábele-Nagy, K., Bozóki, S., and Örs Rebák (2018). Efficiency analysis of double perturbed pairwise comparison matrices. *Journal of the Operational Research Society*, 69(5):707–713. https://doi.org/10.1080/01605682.2017.1409408.

Amenta, P., Lucadamo, A., and Marcarelli, G. (2020). On the transitivity and consistency approximated thresholds of some consistency indices for pairwise comparison matrices. *Information Sciences*, 507:274–287. https://doi.org/10.1016/j.ins.2019.08.042.

Bozóki, S., Csató, L., and Temesi, J. (2016). An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research*, 248(1):211–218. https://doi.org/10.1016/j.ejor.2015.06.069.

Bozóki, S., Dezső, L., Poesz, A., and Temesi, J. (2013). Analysis of pairwise comparison matrices: an empirical research. *Annals of Operations Research*, 211(1):511–528. https://doi.org/10.1007/s10479-013-1328-1.

Bozóki, S., Fülöp, J., and Rónyai, L. (2010). On optimal completion of incomplete pairwise comparison matrices. *Mathematical and Computer Modelling*, 52(1):318–333. https://doi.org/10.1016/j.mcm.2010.02.047.

Bozóki, S. and 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(3-4):362–381. https://www.tandfonline.com/doi/abs/10.1080/03081079.2019.1585432.

Brunelli, M. (2014). *Introduction to the Analytic Hierarchy Process*. Springer. https://doi.org/10.1007/978-3-319-12502-2.

Brunelli, M. (2017). Studying a set of properties of inconsistency indices for pairwise comparisons. *Annals of Operations Research*, 248:143–161. https://doi.org/10.1007/s10479-016-2166-8.

Brunelli, M. (2018). A survey of inconsistency indices for pairwise comparisons. *International Journal of General Systems*, 47(8):751–771. https://doi.org/10.1080/03081079.2018.1523156.

Chao, X., Kou, G., Peng, Y., and Viedma, E. H. (2021). Large-scale group decision-making with non-cooperative behaviors and heterogeneous preferences: An application in financial inclusion. *European Journal of Operational Research*, 288(1):271–293. https://doi.org/10.1016/j.ejor.2020.05.047.

Crawford, G. and Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, 29(4):387–405. https://doi.org/10.1016/0022-2496(85)90002-1.

Csató, L. (2013). Ranking by pairwise comparisons for Swiss-system tournaments. *Central European Journal of Operations Research*, 21(4):783–803. https://doi.org/10.1007/s10100-012-0261-8.

Csató, L. (2017). On the ranking of a Swiss system chess team tournament. *Annals of Operations Research*, 254(1-2):17–36. https://doi.org/10.1007/s10479-017-2440-4.

Csató, L. (2021). *Tournament Design: How Operations Research Can Improve Sports Rules*. Palgrave Pivots in Sports Economics. https://doi.org/10.1007/978-3-030-59844-0.

Csató, L. and Rónyai, L. (2016). Incomplete pairwise comparison matrices and weighting methods. *Fundamenta Informaticae*, 144(3-4):309–320. https://doi.org/10.3233/FI-2016-1337.

Davidson, R. and Farquhar, P. (1976). A bibliography on the method of paired comparisons. *Biometrics*, 32(2):241–252. https://www.jstor.org/stable/2529495.

Dong, Y., Xu, Y., Li, H., and Dai, M. (2008). A comparative study of the numerical scales and the prioritization methods in AHP. *European Journal of Operational Research*, 186(1):229–242. https://doi.org/10.1016/j.ejor.2007.01.044.

Duleba, Sz., Mishina, T., and Shimazaki, Y. (2012). A dynamic analysis on public bus transport's supply quality by using AHP. *Transport*, 27:268–275. https://doi.org/10.3846/16484142.2012.719838.

Edwards, W. and Barron, F. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, 60(3):306–325. https://doi.org/10.1006/obhd.1994.1087.

Fedrizzi, M. and Giove, S. (2007). Incomplete pairwise comparison and consistency optimization. *European Journal of Operational Research*, 183(1):303–313. https://doi.org/10.1016/j.ejor.2006.09.065.

Fedrizzi, M. and Giove, S. (2013). Optimal sequencing in incomplete pairwise comparisons for large dimensional problems. *International Journal of General Systems*, 42(4):366–375. https://doi.org/10.1080/03081079.2012.755523.

Francis-Oliviero, F., Bozóki, S., Micsik, A., Kieny, M. P., and Lelièvre, J. D. (2021). Research priorities to increase vaccination coverage in Europe (EU joint action on vaccination). *Vaccine*, 39(44):6539–6544. https://doi.org/10.1016/j.vaccine.2021.09.033.

Gass, S. (1998). Tournaments, transitivity and pairwise comparison matrices. *Journal of the Operational Research Society*, 49(6):616–624. https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2600572.

Gyarmati, L., Orbán-Mihálykó, É., Mihálykó, Cs., Szádoczki, Zs., and Bozóki, S. (2023). The incomplete analytic hierarchy process and Bradley–Terry model: (In)consistency and information retrieval. *Expert Systems with Applications*, 229(Part B):120522. https://doi.org/10.1016/j.eswa.2023.120522.

Hashimoto, H. and Nikkuni, R. (2013). On Conway–Gordon type theorems for graphs in the Petersen family. *Journal of Knot Theory and Its Ramifications*, 22(9):1350048. https://doi.org/10.1142/S021821651350048X.

Horvát, S. (2020). IGraph/M. An immediately usable version of this software is accessible from its GitHub repository. https://doi.org/10.5281/zenodo.3739056.

Kułakowski, K., Szybowski, J., and Prusak, A. (2019). Towards quantification of incompleteness in the pairwise comparisons methods. *International Journal of Approximate Reasoning*, 115:221–234. https://doi.org/10.1016/j.ijar.2019.10.002.

Kułakowski, K. and Talaga, D. (2020). Inconsistency indices for incomplete pairwise comparisons matrices. *International Journal of General Systems*, 49(2):174–200. https://doi.org/10.1080/03081079.2020.1713116.

Li, Y., Kou, G., Li, G., and Peng, Y. (2022). Consensus reaching process in large-scale group decision making based on bounded confidence and social network. *European Journal of Operational Research*, 303(2):790–802. https://doi.org/10.1016/j.ejor.2022.03.040.

Lovász, L. (1977). A homology theory for spanning trees of a graph. *Acta Mathematica Academiae Scientiarum Hungaricae*, 30(3-4):241–251. https://doi.org/10.1007/bf01896190.

Lundy, M., Siraj, S., and Greco, S. (2017). The mathematical equivalence of the “spanning tree” and row geometric mean preference vectors and its implications for preference analysis. *European Journal of Operational Research*, 257(1):197–208. https://doi.org/10.1016/j.ejor.2016.07.042.

McKay, B. D. and Piperno, A. (2014). Practical graph isomorphism, II. *Journal of Symbolic Computation*, 60(0):94–112. https://doi.org/10.1016/j.jsc.2013.09.003.

Mesbahi, M. (2002). On a dynamic extension of the theory of graphs. *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, 2:1234–1239. https://doi.org/10.1109/ACC.2002.1023188.

Mustajoki, J., Hämäläinen, R., and Salo, A. (2005). Decision support by interval SMART/SWING – incorporating imprecision in the SMART and SWING methods. *Decision Sciences*, 36:317 – 339. https://doi.org/10.1111/j.1540-5414.2005.00075.x.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, 53:49–57. https://doi.org/10.1016/j.omega.2014.11.009.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281. https://doi.org/10.1016/0022-2496(77)90033-5.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill, New York.

Saw, J. G., Yang, M. C., and Mo, T. C. (1984). Chebyshev inequality with estimated mean and variance. *The American Statistician*, 38(2):130–132. https://doi.org/10.1080/00031305.1984.10483182.

Shiraishi, S. and Obata, T. (2002). On a maximization problem arising from a positive reciprocal matrix in AHP. *Bulletin of Informatics and Cybernetics*, 34(2):91–96. https://doi.org/10.5109/13511.

Shiraishi, S., Obata, T., and 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. https://doi.org/10.15807/jorsj.41.404.

Steliga, K. and Szynal, D. (2010). On Markov-type inequalities. *International Journal of Pure and Applied Mathematics*, 58(2):137–152. https://api.semanticscholar.org/CorpusID:5237038.

Szádoczki, Zs., Bozóki, S., Juhász, P., Kadenko, S. V., and Tsyganok, V. (2023). Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. *Annals of Operations Research*, 326(2):783–807. https://doi.org/10.1007/s10479-022-04819-9.

Szádoczki, Zs., Bozóki, S., and Tekile, H. A. (2020). Proposals for the set of pairwise comparisons. Proceedings of the International Symposium on the Analytic Hierarchy Process, ISAHP-2020. https://doi.org/10.13033/isahp.y2020.054.

Szádoczki, Zs., Bozóki, S., and Tekile, H. A. (2022). Filling in pattern designs for incomplete pairwise comparison matrices: (Quasi-)regular graphs with minimal diameter. *Omega*, 107:102557. https://doi.org/10.1016/j.omega.2021.102557.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286. https://doi.org/10.1037/h0070288.

Triantaphyllou, E. (2000). Multi-criteria decision making methods. In *Multi-criteria Decision Making Methods: A Comparative Study*. Applied Optimization, vol 44. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-3157-6_2.

Tsyganok, V. (2010). Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method. *Mathematical and Computer Modelling*, 52(3):538–544. https://doi.org/10.1016/j.mcm.2010.03.052.

von Winterfeldt, D. and Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge: Cambridge University Press.

Wang, K. and Takahashi, I. (1998). How to select paired comparisons in AHP of incomplete information – strongly regular graph design. *Journal of the Operations Research Society of Japan*, 41(2):311–328. https://doi.org/10.15807/jorsj.41.311.

Wolfram Research, I. (2021). Mathematica, Version 12.3. Champaign, IL, 2021. https://www.wolfram.com/mathematica.

5 Study IV. Incomplete pairwise comparison matrices: Ranking top women tennis players

Authors: József Temesi, Zsombor Szádoczki, Sándor Bozóki

Published in Journal of the Operational Research Society, 75(1): 145-157. (2024)

https://doi.org/10.1080/01605682.2023.2180447

Abstract

The method of pairwise comparisons is frequently applied for ranking purposes. This paper aims to rank top women tennis players based on their win/lose ratios. Incomplete pairwise comparison matrices (PCMs) were constructed from data obtained from the WTA (Women's Tennis Association) homepage. The database contains head-to-head results from the period between 1973 and 2022 for 28 players who had the position No. 1 in the official rankings of WTA. The weight vector was calculated from the incomplete PCM with the logarithmic least squares method and the eigenvector method. The results are not surprising: Serena Williams, Steffi Graf, and Martina Navratilova stand in the first three positions, and Martina Hingis, Kim Clijsters, and Justine Henin follow them. We also tested the frequently used probability-based Bradley-Terry method and found high rank-correlation values. Using graph representations, the results gave us a deeper insight into the properties of incomplete PCMs. Special attention was given to the nontransitive triads. A data modification was necessary to remove ties in order to apply the commonly used tests. The results indicate that ordinally nontransitive triads are not significant in the data we analysed.

Keywords: Multi-criteria, Sports, Decision analysis, Graphs

5.1 Introduction

In a wide range of individual (chess, fencing, tennis, boxing) and team sports (basketball, football, ice hockey), the title will be awarded based on pairwise match results. Various traditional systems are available for conducting these types of competitions. We do not aim to correct any of them (a theoretical approach can be found in Csató (2021)). Instead, we are interested in a historical ranking: who is the best player in the long run? Collecting results from certain databases about the wins and losses of players against each other to generate a pairwise comparison matrix seems to be a natural choice. If the pool of selected players contains pairs who have not had matches against each other, then the matrix is incomplete. These matrices play a special role in our research agenda, and applications are crucial to demonstrating our results empirically.

Several studies analysed sports results; ours focused on tennis competitions. Statistical analysis of performance data in tennis can be done for various purposes. Some articles use the data for forecasting certain results of sporting events (Kovalchik, 2016). Lisi and Zanella (2017), for instance, estimate the probability of winning with a logistic regression model. Their example is the analysis of the Grand Slam championships' results in 2013. A special approach for creating reliable forecasts applies Elo-model (Elo, 1978). Williams et al. (2021) confirm the good fitness of the model for Wimbledon 2018 results, especially for top women players. Gu and Saaty (2019) apply descriptive indicators (e.g., age, right- or left-handedness, ranking position) and performance indicators (e.g., number of aces, winning or losing service games, winning or saving break-points). Their Analytic Network Process model is based on factor analysis of the key indicators; it was tested on the results of the US Open 2015. They reported very good fitness with the real results (85%) compared to usual forecasts (70%). Ramón et al. (2012) used similar data, but they applied Data Envelopment Analysis to rank tennis players.

Another application area of sports data is team or player ranking. Langville and Meyer (2012) collected the key ideas and methods of ranking and rating with excellent historical notes and examples (not only sports applications). Their observation is that 'ranking methods . . . are largely based on matrix analysis or optimization. . . Of course, there are plenty of ranking methods from other specialties such as statistics, game theory, economics, etc.' (p.2.) They describe Keener's rating method (Keener, 1993) as a demonstrative example of using the Perron-Frobenius theory, mentioning Wei (1952), Kendall (1955) and Saaty (1987) as early and innovative users of the concept. Langville and Meyer (2006) dedicated a whole book to the PageRank method (Brin and Page, 1998). Dingle et al. (2013) published a PageRank-based tennis ranking, and Dahl (2012) introduced a parametric method based on linear algebra considering the importance of the matches. The method uses pairwise comparisons and was developed for single-elimination tournaments.

Probability-based approaches form another class of models. In the papers of Baker and McHale (2014, 2017) the paired comparisons models are formulated so, that each player being compared is associated with a strength parameter given by a function of the ratio of the strength parameters of the two players in question. The Bradley-Terry (BT) model (Bradley and Terry, 1952) assumes a logistic distribution for that function, and the Thurstone-Mosteller (TM) model (Thurstone, 1927) uses a normal distribution. Baker and McHale used Grand Slam data from more than four decades to estimate the power value of tennis players with the probabilistic dynamic model of paired comparisons. As it will be shown in Section 5.3, their final rankings for women players gave similar results to ours, while the rankings for men players show the same pattern as the results of Wang et al. (2021). Here, the authors applied a two-stage ranking method to minimize ordinal violation for pairwise comparisons to rank the male tennis players.

Orbán-Mihálykó et al. (2019) applied WTA Head-to-Head results (as we also do in this paper) to rank women tennis players using the Thurstone model to estimate parameters with the maximum likelihood method. Their ranking is similar to ours, too.

Our paper discusses the main properties of incomplete pairwise comparison matrices in Section 5.2 and describes the database used in Subsection 5.3.1. Ranking results are presented from different angles using the original PCM and its submatrices in Subsection 5.3.2. The properties of the graph representation are demonstrated next in Subsection 5.3.3. Finally, we draw some conclusions in Section 5.4.

5.2 Pairwise comparison matrices

We briefly summarise some definitions and theorems that we will use during the analysis. We introduce most of the concepts here in a more general way, not specifically for our application. Later on, we will adopt the same notations.

Let $P_1, P_2, \ldots, P_n$ denote the examined items to be compared (alternatives, criteria, voting powers, or, as in our case, players).

Definition 35 *An $n \times n$ $\boldsymbol{P} = [p_{ij}]$ matrix is called a pairwise comparison matrix (PCM), if it satisfies the following properties:*

1. $p_{ij} > 0,\ i, j = 1, 2, \ldots, n$ *(positivity)*
2. $p_{ji} = 1/p_{ij},\ \ i, j = 1, 2, \ldots, n$ *(reciprocity)*
3. *from 1 and 2 follows that* $p_{ii} = 1,\ \ i = 1, 2, \ldots, n.$

The general element of the matrix, p_{ij} shows how many times alternative P_i is better/larger/stronger than alternative P_j. From a practical point of view, the inconsistency of the matrix is crucial.

Definition 36 *A PCM is called consistent if the following holds for any three alternatives (triads):*

4. $p_{ik} = p_{ij} \cdot p_{jk}\ \ i, j, k = 1, 2, \ldots, n$ *(cardinal transitivity).*

If there exists a triad, where this equality does not hold, then the PCM is said to be inconsistent.

Moreover, a triad is called ordinally nontransitive, if the order of its alternatives determined by the appropriate matrix elements is circular. For instance, if $p_{ik} > 1$, $p_{kj} > 1$ and $p_{ij} < 1$, namely alternative P_i is better than alternative P_k, while P_k is better than P_j, however, P_j is also better than P_i. In our application and sports in general, the ordinally nontransitive triads can be interpreted well, and they occur quite often. There can be huge differences in the inconsistency of different PCMs. Measuring this problem has an extended literature (Brunelli, 2018), and there is an ongoing debate about the needed properties of the inconsistency metrics (Brunelli and Fedrizzi, 2015). In real applications, however, the CR (Consistency Ratio) inconsistency index recommended by Saaty (1977), remains the most popular. Here, the $CR < 0.1$ acceptance rule is usually used.

Definition 37 *The CR inconsistency index of an $n \times n$ $\boldsymbol{P}$ PCM is defined as follows:*

$$CR = \frac{CI}{RI},$$

where CI (Consistency Index) can be calculated as:

$$CI = \frac{\lambda_{max} - n}{n - 1},$$

where λ_{max} is the principal eigenvalue of matrix $\boldsymbol{P}$, while RI (Random Index) is the average CI calculated from a randomly generated sample of $n \times n$ PCMs.

Based on different methods, a weight vector can be calculated from a PCM, which determines the ranking (goodness, importance) of the alternatives. The two most commonly used techniques are the eigenvector method (EV) (Saaty, 1977) and the logarithmic least squares method (LLSM) (Crawford and Williams, 1985), which are defined by the following formulas:

- EV: $\boldsymbol{Pw} = \lambda_{max} \cdot \boldsymbol{w}$
- LLSM: $\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\ln p_{ij} - \ln\left(w_i/w_j\right)\right)^2 \to \min_{\boldsymbol{w}}$,

where $\boldsymbol{w}$ denotes the computed weight vector with the general element w_i ($i = 1, \ldots, n$), and λ_{max} is the principal eigenvalue of matrix $\boldsymbol{P}$, as before. If a PCM is consistent, then its elements can be written as $p_{ij} = w_i/w_j$, which means that for such a matrix every weight calculation method results in the same weight vector. It is common in the literature to estimate an inconsistent PCM with a consistent one based on different techniques (Anholcer and Fülöp, 2019). The difference between the priority vectors of the two presented weight calculation techniques also depends on the inconsistency (Kułakowski et al., 2022). Illustrative examples can be found in Brunelli (2014), as well as in Bozóki et al. (2009).

PCMs can contain some missing elements. There are several reasons for this, including: the inability to make some comparisons (as in our case), some data may have been lost, or the time constraints of the decision maker. In these cases, we are dealing with incomplete PCMs (IPCMs). The above-mentioned weight calculation techniques can be easily generalised for IPCMs, as well. The eigenvector method is based on the CR minimal completion (Shiraishi et al., 1998), while in case of the LLSM we only use the known elements of the matrix in the optimization problem (Bozóki et al., 2010). Inconsistency indices and their respective thresholds have also been generalised for the incomplete case (Ágoston and Csató, 2022).

IPCMs are easier to understand if we focus on the graph representation instead of the matrix (Gass, 1998).

Definition 38 *A $G = (V, E)$ undirected graph, where V is the vertex set and E is the edge set of the graph, is called the representing graph of IPCM $\boldsymbol{P}$, if V corresponds to the alternatives of $\boldsymbol{P}$, and an edge is in E if and only if the appropriate element in $\boldsymbol{P}$ is known.*

With the help of graph representation, many results connected to pairwise comparisons can be easily formulated.

Theorem 1 (Bozóki et al. (2010)) *The EV and LLSM techniques generalized to IPCMs have a unique solution, if and only if the representing graph of the IPCM is connected.*

A graph is called connected if there is a path between any two vertices in the graph. If there are two elements for which we cannot find a path, then we cannot determine the relation between their weights (importance) uniquely. However, it is worth investigating some of the stricter variants of connectedness for our problem.

Definition 39 *(a) A $G = (V, E)$ graph is called k-edge-connected, if it remains connected whenever fewer than k edges are removed from the graph, i.e., $G' = (V, E \setminus H)$ is connected, where $H \subseteq E$ and $|H| < k$. The edge connectivity of G is the maximal k, for which G is k-edge-connected.*

(b) A $G = (V, E)$ graph is called k-vertex-connected, if it remains connected whenever fewer than k vertices are removed from the graph, i.e., $G' = (V \setminus L, E)$ is connected, where $L \subseteq V$ and $|L| < k$. The vertex connectivity of G is the maximal k, for which G is k-vertex-connected.

It is also worth considering the confidence level of the weights between two elements for which there is only a long, indirect path that includes many comparisons. A natural measure for this problem is the diameter of the graph (Szádoczki et al., 2022).

Definition 40 *The diameter (d) of a graph $G = (V, E)$ is the longest shortest path between any two vertices of the graph:*

$$d = \max_{u,v \in V} \ell(u, v),$$

where $\ell(.,.)$ is the graph distance, namely the shortest path between two vertices (in our case the weight of every edge is 1).

Examples of the application of the graph representation can be found for instance in Gass (1998), Bozóki and Tsyganok (2019), and Szádoczki et al. (2022).

5.3 Data and results

5.3.1 Database of top women tennis players

The basics of professional tennis have not changed a lot since 1972 when the Association of Tennis Professionals (ATP) was established for protecting the interests of men players, and since 1973 when the Women's Tennis Association (WTA) was founded. The tournament system and the ranking system had their origins in the 1970s. Ranking the players is important because the seeding system is based on the ranking positions, ensuring enjoyable and spectacular competitions. The official ranking systems have special rules in order to play a relevant role in the administration of the tournaments.

ATP and WTA have databases containing the results of all official tournaments, there are search options by tournaments and by players on the homepage of both associations[1]. The H2H (Head-to-Head) statistics are also available: one can be informed about the match results of any two ranked players. The webpages report the recent ranking lists according to the official point systems[2]. These points are informative if we wish to see a kind of power ranking based on the strength of the given tournaments over a certain time frame. However, it is always debated, who the 'best' player is for a longer period, or how we can create a historical ranking. The selection of the players to be ranked is also controversial. Previously (Bozóki et al., 2016), the rankings were generated of those men players who have ever been

[1] https://www.atptour.com/, https://www.wtatennis.com

[2] https://www.atptour.com/en/rankings/singles, https://www.wtatennis.com/rankings/singles

first on the official ATP lists. We followed the same approach for the women players collecting the No. 1 players from the WTA rankings[3] from 1973 to mid-August of 2022. We have found 28 players; their names and the length of their professional careers can be seen in Figure 39.

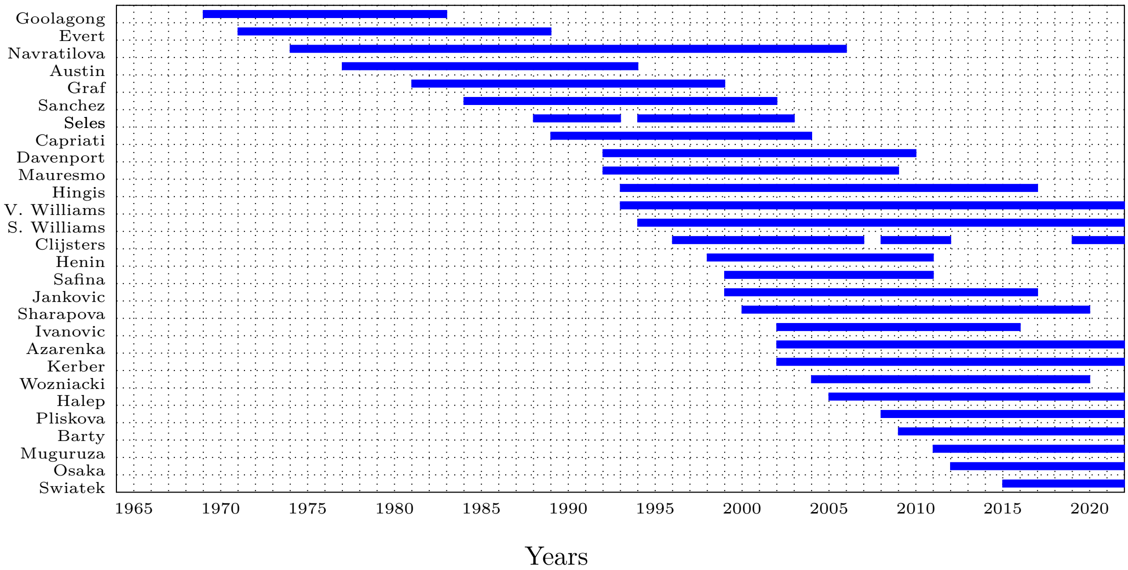


Figure 39: WTA top tennis players and the length of their professional careers.

The chart shows that Navratilova and the Williams sisters have the longest career paths (although others also have careers close to 20 years). We can also see, for instance, that Clijsters retired and resumed two times during our time. There were 11 active players at the beginning of 2022 including the Williams sisters.

We use the database to support our methodology to provide a historical ranking of the selected players. Instead of building a point system from the tournament characteristics and the advancement of a player at a given tournament, we will determine the position of a player using the match results against each other. Let us say that player P_i is 'better' than P_j if the number of her wins over P_j is greater than her losses (there is no tie in tennis), and it is measured by the win/lose (W_{ij}/L_{ij}) ratio. We can construct a matrix with the names of the players in the rows and columns, where the elements are the W/L ratios. The row player is better than the column player if the corresponding ratio is larger than 1, and they are equal if the W/L ratio is 1. If the reciprocal values will measure how much 'worse' P_i than P_j, then all of the ratios form a paired comparison matrix (PCM).

Let $P_1, P_2, \ldots, P_n$ denote the players. Our data can be described as follows:

[3]https://en.wikipedia.org/wiki/List_of_WTA_number_1_ranked_tennis_players

$$\begin{cases} z_{ij} \ \ (i,j=1,\dots,n,\ i\neq j): \text{the number of matches played between} \\ P_i \text{ and } P_j \ (z_{ij}=z_{ji}); \\ x_{ij} \ \ (i>j): \text{the number of matches between } P_i \text{ and } P_j, \\ \text{where the winner was } P_i; \\ y_{ij}=z_{ij}-x_{ij} \ \ (i>j): \text{the number of matches between } P_i \text{ and } P_j, \\ \text{where } P_i \text{ was the loser.} \end{cases} \tag{38}$$

In the incomplete $\boldsymbol{P} = [p_{ij}],\ i,j = 1,\dots,n$ pairwise comparison matrix, p_{ij} denote the W_{ij}/L_{ij} ratio between P_i and P_j:

$$\begin{cases} p_{ij} = x_{ij}/y_{ij}, \text{ if } i,j=1,\dots,n,\ i>j \text{ and } x_{ij}\neq 0, y_{ij}\neq 0; \\ p_{ji} = y_{ij}/x_{ij} = 1/p_{ij}, \text{ if } i,j=1,\dots,n,\ i<j \text{ and } x_{ij}\neq 0, z_{ij}\neq 0; \\ p_{ii}=1 \ \ (i=1,\dots,n); \end{cases} \tag{39}$$

p_{ij} and p_{ji} are missing otherwise.

Similarly to Bozóki et al. (2016), we had to make data corrections to avoid the case of 0 as a denominator in W_{ij}/L_{ij} ratios:

$$p_{ij} = x_{ij}+2, \text{ if } i>j,\ y_{ij}=0 \text{ and } z_{ij}\neq 0 \ (i,j=1,\dots,n). \tag{40}$$

The interpretation of the $p_{ij} > 0$ element is that the ith player is p_{ij} times better than the jth player.

The WTA webpage H2H section includes all results of the players. The W/L ratios are in Table 19. As can be seen from Table 19, the range of the z_{ij} elements is large: therefore the range of the values of p_{ij} is large as well. We have used the following transformation to extract the range:

$$\begin{cases} t_{ij} = p_{ij}^{z_{ij}/K}, \text{ if } z_{ij} < K, \text{ and } t_{ij} = p_{ij} \text{ otherwise}, \\ K = \max_{i,j} z_{i,j}, \end{cases} \tag{41}$$

where the transforming factor is the ratio of the number of matches played between the given two players and the maximum of the number of matches played between any two players. This can also solve the problem that the same W/L ratio based on a few matches is considered to be less reliable compared to the ratio coming from a large number of games, so we transform the ratios based on a small sample in a way that they will be closer to one. Note that if all players have the same number of matches, transformation 41 results in $t_{ij} = p_{ij}$. If the parameter K is set to a smaller value, it can be interpreted as a threshold from where we do not distinguish between the confidence of the ratios based on the number of matches. Now the initial matrix of the calculations will be matrix $\boldsymbol{T}$ with elements t_{ij}. This matrix is obviously incomplete because it is easy to find players with disjoint career intervals in Figure 39.

The graph representation of Table 19 is the network in Figure 40. The vertices of the undirected graph represent the players. The edges show that the connected players played at least one match against each other. An important property of the graph representation in Figure 40 is that it is connected. According to Theorem 1, the estimated weight vector gives a unique solution.

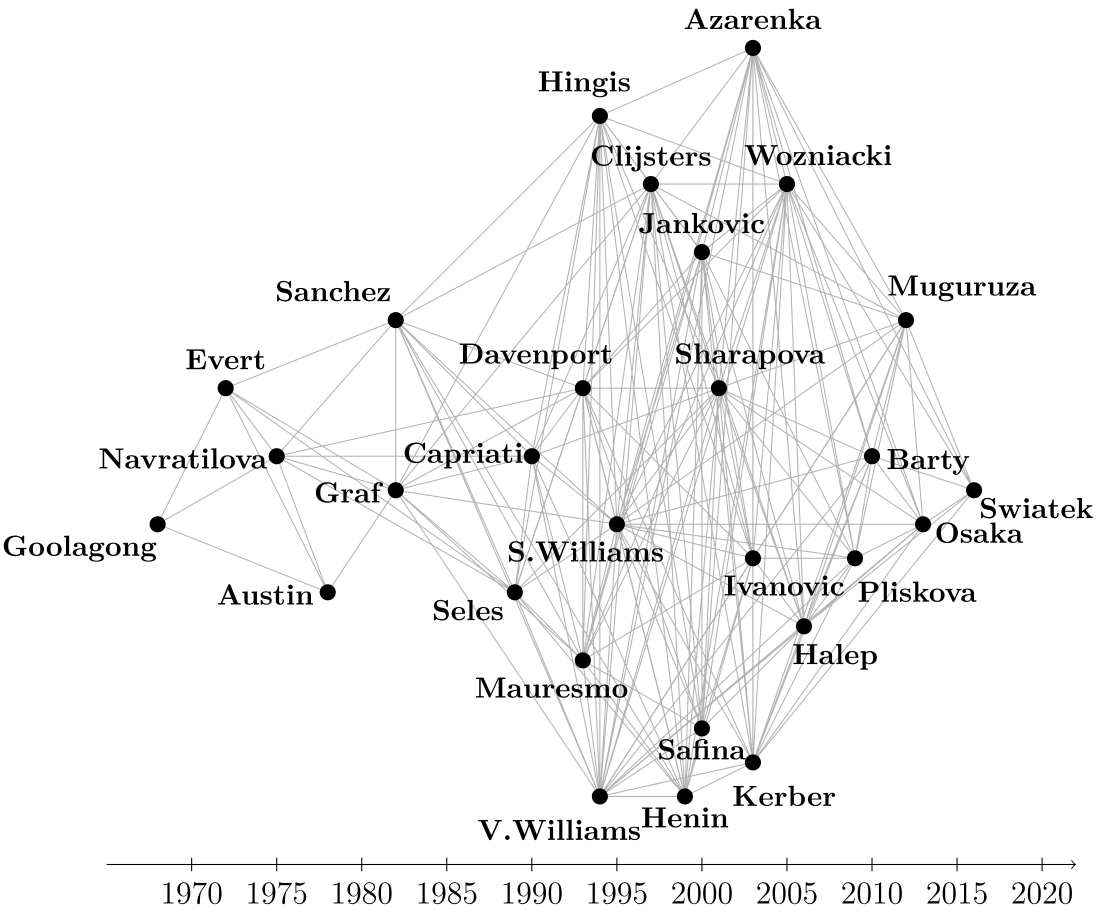


Figure 40: Graph representation of top WTA players' IPCM.

On first glance it appears that the players at the beginning and at the end of the 50-year-long period are strictly separated; however, we can find players who 'connect' them, like Sharapova or the Williams sisters. Navratilova could be one of them, but she met only 8 players of the other 27: calculating career path statisticians do not distinguish between individual and double competitions, and the latter extended her

	Goolagong	Evert	Navratilova	Austin	Graf	Sanchez	Seles	Capriati	Davenport	Mauresmo	Hingis	V. Williams	S. Williams	Clijsters	Henin	Safina	Jankovic	Sharapova	Ivanovic	Azarenka	Kerber	Wozniacki	Halep	Pliskova	Barty	Muguruza	Osaka	Swiatek
Goolagong		12/26	12/15	4/4																								
Evert	26/12		37/43	9/8	6/7	1/1	2/1																					
Navratilova	15/12	43/37		20/13	9/9	12/3	7/10	1/1	1/0																			
Austin	4/4	8/9	13/20		1/1																							
Graf		7/6	9/9	1/1		28/8	10/5	10/1	8/6	1/0	7/2	3/2	1/1	1/0														
Sanchez		1/1	3/12		8/28		3/20	6/4	7/5	1/2	2/18	3/6	4/3	0/4	1/0													
Seles		1/2	10/7		5/10	20/3		9/5	3/10	3/2	5/15	1/9	1/4	1/0	4/3			1/0										
Capriati			1/1		1/10	4/6	5/9		3/9	4/7	4/5	0/4	7/10	3/3	2/5			1/0										
Davenport			0/1		6/8	5/7	10/3	9/3		12/4	14/11	14/13	4/10	8/9	5/7	2/1	4/2	1/5	1/0		1/0	1/0						
Mauresmo					0/1	2/1	2/3	7/4	4/12		7/8	3/5	2/10	7/8	6/8	4/3	6/1	3/1	6/2	0/2		0/1						
Hingis					2/7	18/2	15/5	5/4	11/14	8/7		11/10	6/7	4/5	2/2	2/1	0/2	1/2	1/1	0/1		2/0						
V. Williams					2/3	6/3	9/1	4/0	13/14	5/3	10/11		12/19	6/7	7/2	3/1	7/7	3/5	9/3	6/3	3/6	7/1	3/4	1/2	0/2	4/2	1/1	
S. Williams					1/1	3/4	4/1	10/7	10/4	10/2	7/6	19/12		7/2	8/6	6/1	10/4	20/2	9/1	18/5	6/3	10/1	10/2	2/2	2/0	3/3	1/3	
Clijsters					0/1	4/0	0/1	3/3	9/8	8/7	5/4	7/6	2/7		13/12	8/2	8/1	5/4	6/0	4/3	0/1	3/0	1/0			0/1		
Henin						0/1	¾	5/2	7/5	8/6	2/2	2/7	6/8	12/13		5/1	10/0	7/3	5/0		2/0	1/0						
Safina									1/2	3/4	1/2	1/3	1/6	2/8	1/5		3/4	3/4	1/3	4/2		1/0	1/0					
Jankovic									2/4	1/6	2/0	7/7	4/10	1/8	0/10	4/3		1/8	3/9	4/7	2/4	5/6	1/7	1/1		2/3		
Sharapova							0/1	0/1	5/1	1/3	2/1	5/3	2/20	4/5	3/7	4/3	8/1		10/4	8/7	4/5	7/4	7/2	2/0	1/2	3/1	0/1	
Ivanovic									0/1	2/6	1/1	3/9	1/9	0/6	0/5	3/1	9/3	4/10		3/5	5/2	5/2	3/2	0/5		0/1		
Azarenka										2/0	1/0	3/6	5/18	3/4		2/4	7/4	7/8	5/3		10/1	7/4	2/3	4/4	1/1	2/3	1/3	1/2
Kerber									0/1			6/3	3/6	1/0	0/2		4/2	5/4	2/5	1/10		8/7	6/6	7/5	2/3	3/5	4/2	0/1
Wozniacki									0/1	1/0	0/2	1/7	1/10	0/3	0/1	0/1	6/5	4/7	2/5	4/7	7/8		5/2	6/4	3/4	3/3	2/1	0/1
Halep												4/3	2/10	0/1		0/1	7/1	2/7	2/3	3/2	6/6	2/5		8/4	3/1	3/4	4/1	2/2
Pliskova												2/1	2/2				1/1	0/2	5/0	4/4	5/7	4/6	4/8		2/6	9/2	3/2	
Barty												2/0	0/2					2/1		1/1	3/2	4/3	1/3	6/2		3/1	2/2	2/0
Muguruza												2/4	3/3	1/0			3/2	1/3	1/0	3/2	5/3	3/3	4/3	2/9	1/3		0/1	1/0
Osaka												1/1	3/1					1/0		3/1	2/4	1/2	1/4	2/3	2/2	1/0		1/1
Swiatek																				2/1	1/0	1/0	2/2		0/2	0/1	1/1	
	Goolagong	Evert	Navratilova	Austin	Graf	Sanchez	Seles	Capriati	Davenport	Mauresmo	Hingis	V. Williams	S. Williams	Clijsters	Henin	Safina	Jankovic	Sharapova	Ivanovic	Azarenka	Kerber	Wozniacki	Halep	Pliskova	Barty	Muguruza	Osaka	Swiatek

Table 19: W/L ratios of top WTA players.

professional career longer than the average.

5.3.2 Ranking results

The weight vector of the women players' incomplete PCM calculated with the logarithmic least squares method gives the ranking in the first column in Table 20. The result is not surprising: Serena Williams, Graf, and Navratilova top the list, followed by Hingis, Clijsters, and Henin. The new generation is represented by Barty in the 8th place. The second column of the table demonstrates that the W/L rates prove to be a good proxy of our ranking. The Spearman rank correlation value is 0.962.

	$\boldsymbol{LLSM}$	$\boldsymbol{W/L}$	$\boldsymbol{LLSM_{K=30}}$	$\boldsymbol{EV}$	$\boldsymbol{BT}$
S. Williams	1	1	1	1	1
Graf	2	2	2	2	2
Navratilova	3	6	3	3	4
Hingis	4	7	4	4	7
Clijsters	5	5	5	5	6
Henin	6	4	6	6	3
V. Williams	7	8	7	7	9
Barty	8	3	8	8	10
Davenport	9	9	10	9	8
Evert	10	10	9	10	5
Seles	11	16	11	11	12
Osaka	12	14	12	13	16
Sharapova	13	11	13	12	13
Pliskova	14	12	14	14	20
Halep	15	15	15	15	19
Swiatek	16	13	16	16	17
Azarenka	17	17	17	17	18
Muguruza	18	19	18	18	21
Mauresmo	19	18	19	20	14
Kerber	20	20	20	19	22
Wozniacki	21	22	21	21	24
Safina	22	25	23	22	25
Austin	23	21	22	23	11
Ivanovic	24	24	24	24	27
Capriati	25	26	25	25	23
Jankovic	26	27	26	26	28
Sanchez	27	28	28	27	26
Goolagong	28	23	27	28	15

Table 20: Ranking results.

Since the value of the parameter K is an outlier ($K = 80$), a ranking was generated with an average value, $K = 30$, as can be seen in column 3. There are minor changes in the positions of the players: Evert and Davenport changed positions in the 9th and 10th places; Safina and Austin in the 22nd and 23rd positions; Sanchez and Goolagong changed the last two positions. The Spearman rank correlation value is 0.998. The situation was similar to other transformation factors. The fourth column of the table contains the ranking calculated with the eigenvector method. Two rank reversals can be found: Osaka and Sharapova, and Mauresmo and Kerber. The Spearman rank correlation is 0.999.

We also applied the well-known, probability-based Bradley-Terry model (Bradley and Terry, 1952) to our data to create a ranking, and compare this approach to ours. This method assumes that there are latent random variables with logistic distribution behind the performance of the players. In the traditional model we are estimating the expected values of these random variables and based on those we can rank the objects (the larger the better). Usually, the maximum likelihood estimation method is used to determine the parameters (expected values), for which there exists a unique solution if and only if the directed graph of the comparisons is strongly connected (Ford Jr, 1957). This assumption is more stringent than the uniqueness of the PCM-based method. It basically means that even that graph should be connected, for which we delete those edges from the graph of Figure 40 where only one of the players has won all the games. This means 40 of the 192 edges in our data, however, this graph is still connected. We calculated the results of the Bradley-Terry model to compare our method to one of the most commonly used probability-based ranking methods as well. Column 5 of the table contains the ranking calculated with the maximum likelihood method of the Bradley-Terry model. As one can see, the BT-model provides a similar ranking, and the Spearman rank correlation is 0.860. The main difference is that the ranks of the earlier players (Goolagong, Evert, Austin, and Henin) are significantly better.

Our calculations with the WTA players are in line with the top ATP players in Bozóki et al. (2016). Both data systems are robust in that respect that the rankings which have been calculated from the incomplete PCMs are not sensitive to reasonable corrections, and the choice of the estimation method does not make a significant impact on the results. The rankings are similar to other orders determined by commonly used ranking models, like the Bradley-Terry method. Empirical evidence suggests that our methodology can be recommended for the given ranking exercise.

The next step of our calculations was to analyse the submatrices of the initial matrix. What happens if elements (players) were dropped or involved? How do subrankings behave? PCM1 column of Table 21 shows a ranking without the first nine players in the overall ranking. Seles, Sharapova, and Evert have the first three positions; Osaka, Kerber, and Muguruza lost several positions; the last positions did not change significantly. PCM2 is a ranking without the last nine players of the overall ranking. Serena Williams and Graf saved the first two positions, but the ranking behind them is very different from the original one. The position of Seles is very interesting: in PCM1, she is first, but in PCM2, she is last! The explanation of the changes is simple and plausible. Both the number of matches and the composition of matches changed. Some players benefitted from the modified structure (those players were missing with whom they had the poorest W/L ratios), and others became victims of the changes (their best W/L ratios disappeared). Some indirect impacts have also vanished. The most prominent example of this phenomenon is Monica Seles. The rankings are not independent of the incoming and outgoing elements – as was expected.

PCM3 is the ranking of the four most influential stars of the seventies. They follow each other under the overall ranking: it looks like their results inside of PCM3 follow the pattern outside of the block. Since everybody played against everybody here, a simple preference order can be calculated based on the winner-loser relationship as a binary relation. The order is Navratilova $\succ$ Evert $\succ$ Austin $\sim$ Goolagong. PCM4 is a ranking of 12 players from the next era. From the first six places, Hingis is the only one, who lost position, and Wozniacki is the other one, who lost position at the end of the ranking. PCM4 gives evidence that it is possible to select a relatively large number of players with their most active career in the same time period, so that their results practically determine their positions with minor changes compared to the overall ranking. PCM3 and PCM4 are complete submatrices of matrix $\boldsymbol{T}$, therefore the usual CR inconsistency indices can be calculated, too. The CR values are below 0.02 supporting our hypothesis about the robustness of the data. Furthermore, PCM4 includes 20.50% of all matches in the matrix $\boldsymbol{T}$.

Finally, in PCM5 there is a ranking of 19 players selected randomly from our pool of women players. The Williams sisters, Evert, Seles, Kerber, Azarenka, Safina, Jankovic, and Swiatek were not included. The first five players follow each other in the same order as they did in the overall ranking. The reason is likely the fact that their performance against Evert and the Williams sisters follows the same pattern not influencing the ranking based on their match results against each other. Due to the elimination of their results against the leaving players, Sharapova and Austin are in a better position. We have had the same experience with other random sets of players.

	PCM1	PCM2	PCM3	PCM4	PCM5
S. Williams	-	1	-	1	-
Graf	-	2	-	-	1
Navratilova	-	8	1	-	2
Hingis	-	9	-	6	3
Clijsters	-	10	-	2	4
Henin	-	12	-	3	5
V. Williams	-	5	-	4	-
Barty	-	3	-	-	10
Davenport	-	7	-	5	6
Evert	3	18	2	-	-
Seles	1	19	-	-	-
Osaka	12	6	-	-	15
Sharapova	2	15	-	7	8
Pliskova	6	4	-	-	13
Halep	7	11	-	-	14
Swiatek	8	13	-	-	-
Azarenka	5	16	-	-	-
Muguruza	14	14	-	-	16
Mauresmo	4	17	-	8	11
Kerber	15	-	-	-	-
Wozniacki	11	-	-	11	12
Safina	9	-	-	9	-
Austin	18	-	3	-	7
Ivanovic	10	-	-	10	18
Capriati	13	-	-	-	17
Jankovic	17	-	-	12	-
Sanchez	16	-	-	-	19
Goolagong	19	-	4	-	9

Table 21: Ranking results from various submatrices of matrix $\boldsymbol{T}$.

There are two key conclusions from these calculations. Changes in the set of players changed the rankings – as we expected. However, these changes did not blow over the original rankings entirely, the new positions could be explained with the new patterns of the modified PCM.

We also examined the rankings, when we added additional elements (tennis players) to the set of players one by one. The first subset that we analysed includes the players who were active at the beginning of 2022 (11 players). Then we stepped backward in time and included the next player, who finished her professional career the latest (in case of a tie between two players, we chose the one who started her career later). We included the elements one by one until we get the whole ranking of all 28 players. In this way, we generated 18 different rankings, which can be seen in Figure 41, as well as the changes caused by the entry of a given player (the entering players are shown at the bottom of the chart). It is important to note that there are exactly as many rank reversals due to the inclusion of a given player, as many lines cross each other between the inclusion and the former player's involvement.

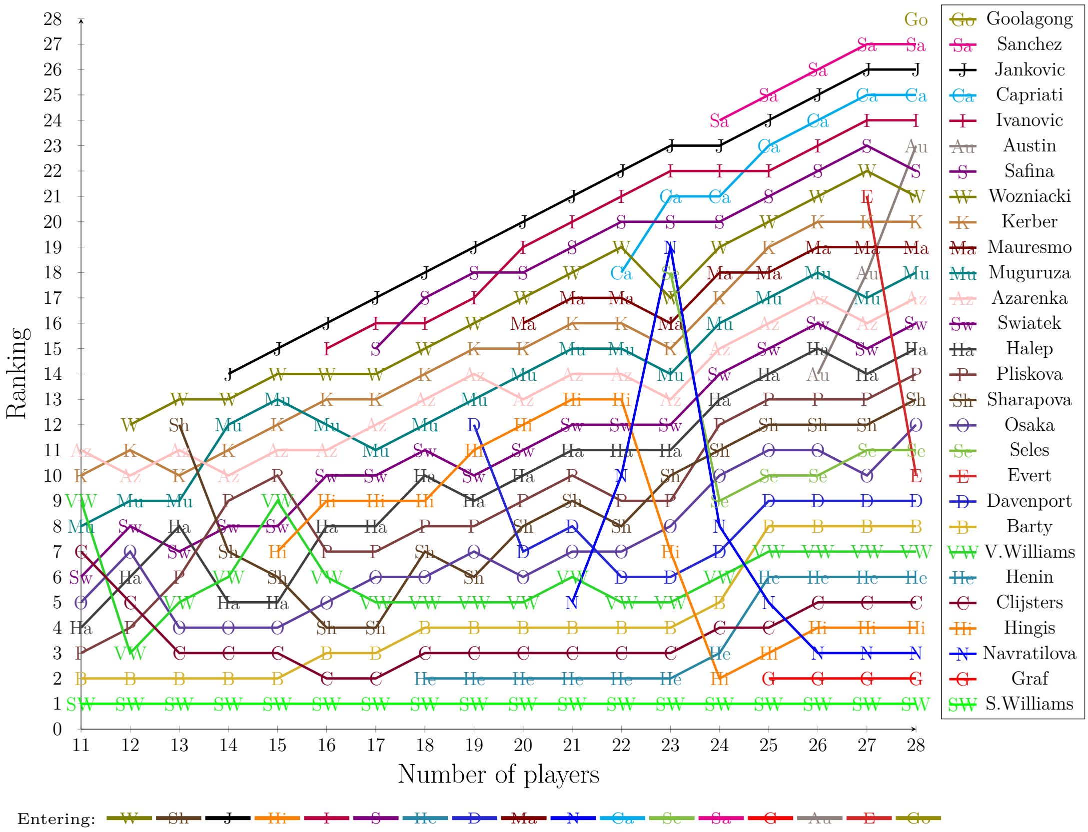


Figure 41: The differences in the rankings of women tennis players when they enter the ranking one by one.

We can see that the ranking is robust, the inclusion of a player usually does not affect the results too much, and only one or two rank reversals occur. In the rare cases when a player's rank is changed by a significant number it is since she barely played with the other players who are in the ranking so far (for instance Austin), and her comparison to the newly involved element (Evert, and then Goolagong) is more reliable. This can be seen, when the ranking of Navratilova undergoes a lot of change in the first few steps as she only played a single game with the so far included players. Of course, a player may also win (or lose) many times against a currently entered element (for instance Hingis against Seles or Sanchez and Venus Williams against Ivanovic). It is worth mentioning that the entry of a player usually has a larger impact on the players with whom she has played a lot. The beginning and the end of the ranking both seem to be robust. We can see more rank reversals in the middle, as we involve more and more elements. However, it still looks like there are clusters here, and the ranking of the players only changes within those groups.

5.3.3 Graph representations

Using graph representations gave us the possibility to have a deeper insight into the properties of incomplete pairwise comparison matrices. The representing graph of the women players can be seen in Figure 40, while its connectivity indicators are described in Table 22

	$\boldsymbol{WTA}$	$\boldsymbol{WTA_{mod}}$
Number of vertices (players)	28	23
Number of edges (players compared)	192	170
Minimal vertex degree	3	9
Maximal vertex degree	22	22
Diameter	4	2
Average shortest path	1.66	1.33
Edge connectivity	3	9
Vertex connectivity	3	8

Table 22: Properties of the representing graph for WTA players.

We can find one player (Goolagong), who had competitions with only three other players, as it is indicated by the minimal vertex degree. On the contrary, the Williams sisters had matches with 22 other players. The maximal vertex degree belongs to them. Erasing either any 2 edges or vertices we can get a connected graph. The longest shortest path (diameter) is 4, and it can be determined between players far from each other in time: Goolagong and Osaka/Muguruza/Barty/Swiatek respectively or between Evert and Swiatek.

Figure 42 illustrates the degree of vertices. That distribution is used in network theory (Albert et al., 2000) for analysing various types of systems. In our case, the average shortest path has a small value, and the degrees are distributed relatively orderly around the average. The clustering value is 0.79 (that means certain groups of players have results from the same group). These properties indicate that our network is a small-world type one (Watts and Strogatz, 1998).

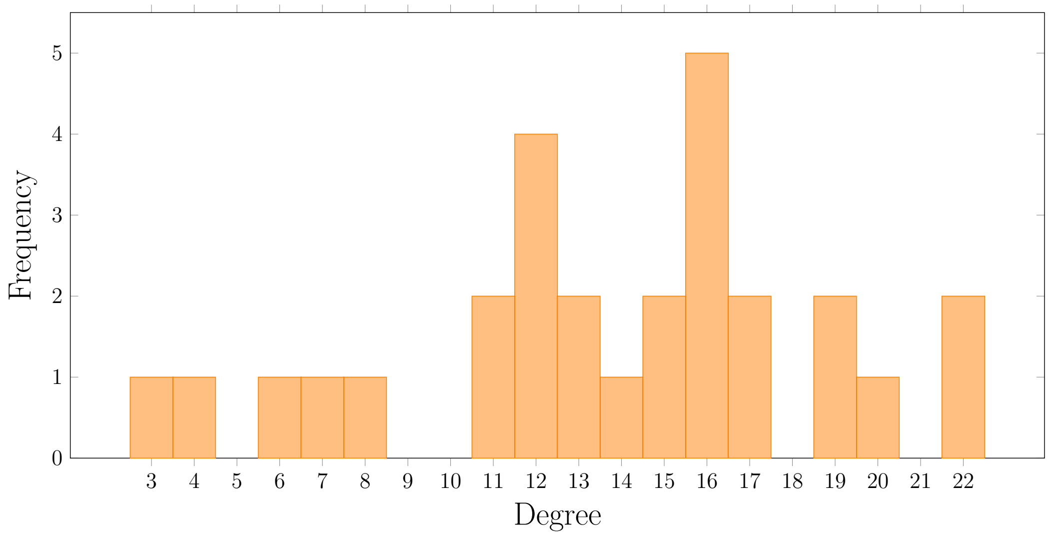


Figure 42: Distribution of the degree of vertices.

Another feature of the representing graph for WTA players is that by erasing the Williams sisters, represented by the vertices with maximal vertex degree, the connectivity properties practically do not change. However, erasing the critical triads of Evert, Navratilova, and Graf, or Evert, Navratilova, and Austin, the connectivity of the graph will be lost. On the other hand, erasing four players of the earliest period (Goolagong, Evert, Navratilova, Austin) and the most recent world number one (Swiatek), the remaining graph will have much stronger connectivity indicators, as can be seen in the second column of Table 22 ($\boldsymbol{WTA_{mod}}$). Rankings generated from these reduced graphs (submatrices) almost follow the positions in the overall ranking, suggesting that these strong relations can specify them.

It is another fact that the modified, strongly connected graph is the union of two star graphs, complemented with a few edges. The centres are the Williams sisters – meaning they played directly with all other players. Similar representing graphs can be obtained by applying the popular best-worst method (Rezaei, 2015). That structure can also be responsible for the robustness of our ranking results.

Another line of our research referred to the ordinally nontransitive triads. There are many sports competitions where $W/L > 1$ for A and B, and the same is true for B and C: A is better than B, and B is better than C. We can expect that A will be better than C; however, from the results, we get $W/L < 1$. In preference ordering that triad is called contradictory (Kwiesielewicz and van Uden, 2004). A suitable example of an ordinally nontransitive triad is Henin, Davenport, and Venus Williams in our database. We chose the positive reciprocal multiplicative PCM approach for ranking tennis players because the estimation methods are functional in the case of ordinal or cardinal nontransitivity. However, in the course of discussing the ranking results, it is important to know more about the ordinally nontransitive triads of the PCM, since their presence signals a kind of contradiction. Representing graphs are directed in the analysis of ordinal nontransitivity: an edge leads from one player to the next if the latter player lost more matches.

Kendall and Babington Smith (1940) gave the distribution of ordinally nontransitive triads in the case of a low number of elements ($n \leq 7$) and proposed a significance test. Alway (1962) extended the distribution for $8 \leq n \leq 10$; others analysed cases with larger numbers of elements. Moran (1947) proved that the distribution of the nontransitive triads goes to the normal distribution if the number of elements goes to infinity, but the convergence is slow. Knezek et al. (1998) investigated the chi-square distribution used by Kendall and Babington Smith (1940) earlier, and they found it satisfactory for more than 15 elements. Jensen and Hicks (1993) proposed a consistency coefficient and a nonparametric test for ordinal PCMs, while Iida (2009) discussed the nontransitivity tests for decision-making problems by applying them to binary PCMs without ties. It is crucial to note that all of the above-mentioned tests worked for complete PCMs without ties. In the case of ties, Kułakowski (2018) determined the maximal number of contradictory triads for any number of elements and proposed an index related to that number (without a statistical test). He extended the definition of contradictory triads to those cases when there are only one or two equalities between the elements of the triads. That kind of inconsistency analysis could not be interpreted properly in our case; we are looking for strictly inconsistent triads. That is why we did not follow the approaches of Iida or Kułakowski, and decided to hark back to the case without ties and to use the known tests.

The modified databases contain W/L set ratios for each pair. In the case of having ties even for the set ratios, the original LLSM ranking was the reference to make a precedence relation. The nontransitivity tests need complete matrices. If two players have not played with each other for any reason (no edge was found between the two vertices), then we used the same reference ranking to determine a 'winner'. A complete directed graph was created this way. Table 23 includes information about the original incomplete PCM in the first, and about its completed and tie-corrected version in the second column.

	$\boldsymbol{WTA}$	$\boldsymbol{WTA_{complete}}$
Density	192/378	378/378
Number of ties	20	0
Number of ordinally nontransitive triads	83	315
Possible maximal number of ordinally nontransitive triads	910	910

Table 23: Basic data for nontransitivity analysis.

Regarding the case before correction, we can see that the PCM has a density of around 50% (half of the elements are known), and the ratio of ties is about 10%. We have got the number of nontransitive triads from the incomplete matrices here, therefore it is not comparable with the possible maximal number of nontransitive triads obtained from the complete matrix. The second column of Table 23 informs us about the case after eliminating ties and completing the matrix. The chi-squared test value is ∼202, the corresponding p-value is practically 0: the number of ordinally nontransitive triads is not significant in our database.

5.4 Conclusion

Our results provide empirical evidence that the method of incomplete pairwise comparison matrices is appropriate for producing well-understood rankings. Our study was based on the match results of players against each other. Calculations with the whole databases and their subsets clarified that WTA data were robust enough to state that although the rankings have been changed, the differences can be explained via the analysis of the data matrices, and they are logically consistent. Our historical rankings alone may be of interest to tennis experts, but they are also relevant from a decision theory perspective.

One of the novelties of our approach was the analysis of representing graphs. We aimed to contribute to a deeper description of the properties of incomplete pairwise comparison matrices. Graph representations can open new avenues in this regard. We consider further research on ordinally nontransitive triads to be particularly important.

Our work is based on Head-to-Head statistics of the players without taking some considerations into account, which seems to limit the validity of our rankings. They are listed and commented on here with the aim of either explaining why we chose an overall and unified approach or opening new tracks with fine-tuning of the data.

Tennis fans and experts can say that: 'It is not fair to give the same weight to the matches of any player from the very early and very late periods of their professional career.' Having details about the professional career of each player it is possible to introduce correction factors. But there are at least two reasons to drop the idea. It would not be easy to determine those early and late stages, and even if we can do it, the value of the correction parameter would include a strong subjective factor in the analysis. On the other hand, we can easily find players with exceptionally good early results (e.g., Austin, Seles, Osaka), and some players retired without a declining period (e.g., Henin). That kind of time-dependent adjustment of data would bring a very controversial factor to the ranking results.

'Different surfaces need another sort of treatment.' The weighting of surfaces would introduce a subjective factor, again. Revaluation of individual results would lead to endless debates. Yes, a viable solution would be to make separate rankings for different surfaces: who is the top player on clay court, and who is the most successful on grass? Data are available, but we did not undertake that job, because it would not give extra methodological benefits.

'Match ratios can be used, but set ratios would reflect better to power differentiation.' Data are available to calculate W/L ratios from sets. We have made some calculations in the case of both men and women players. Rankings were not significantly different from the original ones, so we dropped that artificial approach.

'Ranking is restricted to the No. 1 WTA players – their performance against other players might change that ranking.' For instance, selecting the top 20 players from every year between 1973 and 2022 is possible, as data are available. We have not done the job of ranking them (or more players). It is worth mentioning that historical rankings of different player populations show very strong similarity (as is referred to in the introductory section of this paper). Another remark is that top tennis is surprisingly endogenous, the best players meet each other frequently. Even in our small sample, we can see that the ratio of 'number of matches in our database/number of matches in the entire career' is the smallest for Swiatek (∼6%), and the largest for Serena Williams (∼25%).

Acknowledgements

The authors thank the valuable comments and suggestions of the anonymous Reviewers. The comments of László Csató and András London are greatly acknowledged. This research has been supported in part by the TKP2021-NKTA-01 NRDIO grant.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Ágoston, K. Cs. and Csató, L. (2022). Inconsistency thresholds for incomplete pairwise comparison matrices. *Omega*, 108:102576. https://doi.org/10.1016/j.omega.2021.102576.

Albert, R., Jeong, H., and Barabási, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382. https://doi.org/10.1038/35019019.

Alway, G. G. (1962). The distribution of the number of circular triads in paired comparisons. *Biometrika*, 49(1-2):265–269. https://doi.org/10.1093/biomet/49.1-2.265.

Anholcer, M. and Fülöp, J. (2019). Deriving priorities from inconsistent PCM using network algorithms. *Annals of Operations Research*, 274(1):57–74. https://doi.org/10.1007/s10479-018-2888-x.

Baker, R. D. and McHale, I. G. (2014). A dynamic paired comparisons model: Who is the greatest tennis player? *European Journal of Operational Research*, 236(2):677–684. https://doi.org/10.1016/j.ejor.2013.12.028.

Baker, R. D. and McHale, I. G. (2017). An empirical Bayes model for time-varying paired comparisons ratings: Who is the greatest women's tennis player? *European Journal of Operational Research*, 258(1):328–333. https://doi.org/10.1016/j.ejor.2016.08.043.

Bozóki, S., Csató, L., and Temesi, J. (2016). An application of incomplete pairwise comparison matrices for ranking top tennis players. *European Journal of Operational Research*, 248(1):211–218. https://doi.org/10.1016/j.ejor.2015.06.069.

Bozóki, S., Fülöp, J., and Rónyai, L. (2009). Incomplete pairwise comparison matrices in multi-attribute decision making. In *2009 IEEE International Conference on Industrial Engineering and Engineering Management*, pages 2256–2260. https://doi.org/10.1109/IEEM.2009.5373064.

Bozóki, S., Fülöp, J., and Rónyai, L. (2010). On optimal completion of incomplete pairwise comparison matrices. *Mathematical and Computer Modelling*, 52(1):318–333. https://doi.org/10.1016/j.mcm.2010.02.047.

Bozóki, S. and 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(3-4):362–381. https://www.tandfonline.com/doi/abs/10.1080/03081079.2019.1585432.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4):324–345. https://doi.org/10.2307/2334029.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1):107–117. Proceedings of the Seventh International World Wide Web Conference. https://doi.org/10.1016/S0169-7552(98)00110-X.

Brunelli, M. (2014). *Introduction to the Analytic Hierarchy Process*. Springer. https://doi.org/10.1007/978-3-319-12502-2.

Brunelli, M. (2018). A survey of inconsistency indices for pairwise comparisons. *International Journal of General Systems*, 47(8):751–771. https://doi.org/10.1080/03081079.2018.1523156.

Brunelli, M. and Fedrizzi, M. (2015). Axiomatic properties of inconsistency indices for pairwise comparisons. *Journal of the Operational Research Society*, 66(1):1–15. https://doi.org/10.1057/jors.2013.135.

Crawford, G. and Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, 29(4):387–405. https://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, Palgrave Macmillan. https://doi.org/10.1007/978-3-030-59844-0.

Dahl, G. (2012). A matrix-based ranking method with application to tennis. *Linear Algebra and its Applications*, 437(1):26–36. https://doi.org/10.1016/j.laa.2012.02.002.

Dingle, N., Knottenbelt, W., and Spanias, D. (2013). On the (Page) ranking of professional tennis players. In Tribastone, M. and Gilmore, S., editors, *Computer Performance Engineering*, pages 237–247. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36781-6_17.

Elo, A. E. (1978). *The rating of chessplayers, past and present*. Arco Pub. https://www.gwern.net/docs/statistics/comparison/1978-elo-theratingofchessplayerspastandpresent.pdf.

Ford Jr, L. R. (1957). Solution of a ranking problem from binary comparisons. *The American Mathematical Monthly*, 64(8P2):28–33. https://doi.org/10.2307/2308513.

Gass, S. (1998). Tournaments, transitivity and pairwise comparison matrices. *Journal of the Operational Research Society*, 49(6):616–624. https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2600572.

Gu, W. and Saaty, T. L. (2019). Predicting the outcome of a tennis tournament: Based on both data and judgments. *Journal of Systems Science and Systems Engineering*, 28(3):317–343. https://doi.org/10.1007/s11518-018-5395-3.

Iida, Y. (2009). The number of circular triads in a pairwise comparison matrix and a consistency test in the AHP. *Journal of the Operations Research Society of Japan*, 52(2):174–185. https://doi.org/10.15807/jorsj.52.174.

Jensen, R. E. and Hicks, T. E. (1993). Ordinal data AHP analysis: A proposed coefficient of consistency and a nonparametric test. *Mathematical and Computer Modelling*, 17(4):135–150. https://doi.org/10.1016/0895-7177(93)90182-X.

Keener, J. P. (1993). The Perron–Frobenius theorem and the ranking of football teams. *SIAM Review*, 35(1):80–93. https://doi.org/10.1137/1035004.

Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, 11(1):43–62. http://www.jstor.org/stable/3001479.

Kendall, M. G. and Babington Smith, B. (1940). On the method of paired comparisons. *Biometrika*, 31(3/4):324–345. https://doi.org/10.2307/2332613.

Knezek, G., Wallace, S., and Dunn-Rankin, P. (1998). Accuracy of Kendall's chi-square approximation to circular triad distributions. *Psychometrika*, 63(1):23–34. https://doi.org/10.1007/BF02295434.

Kovalchik, S. A. (2016). Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, 12(3):127–138. https://doi.org/10.1515/jqas-2015-0059.

Kułakowski, K. (2018). Inconsistency in the ordinal pairwise comparisons method with and without ties. *European Journal of Operational Research*, 270(1):314–327. https://doi.org/10.1016/j.ejor.2018.03.024.

Kułakowski, K., Mazurek, J., and Strada, M. (2022). On the similarity between ranking vectors in the pairwise comparison method. *Journal of the Operational Research Society*, 73(9):2080–2089. https://doi.org/10.1080/01605682.2021.1947754.

Kwiesielewicz, M. and van Uden, E. (2004). Inconsistent and contradictory judgements in pairwise comparison method in the AHP. *Computers & Operations Research*, 31(5):713–719. https://doi.org/10.1016/S0305-0548(03)00022-4.

Langville, A. N. and Meyer, C. D. (2006). *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press. http://www.jstor.org/stable/j.ctt7t8z9.

Langville, A. N. and Meyer, C. D. (2012). *Who's #1?: The Science of Rating and Ranking*. Princeton University Press. http://www.jstor.org/stable/j.ctt7rwdt.

Lisi, F. and Zanella, G. (2017). Tennis betting: can statistics beat bookmakers? *Electronic Journal of Applied Statistical Analysis*, 10(3):790–808. http://siba-ese.unisalento.it/index.php/ejasa/article/view/16516.

Moran, P. A. P. (1947). On the method of paired comparisons. *Biometrika*, 34(3-4):363–365. https://pubmed.ncbi.nlm.nih.gov/18918706/.

Orbán-Mihálykó, E., Mihálykó, Cs., and 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. https://doi.org/10.1007/s10100-017-0495-6.

Ramón, N., Ruiz, J. L., and Sirvent, I. (2012). Common sets of weights as summaries of DEA profiles of weights: With an application to the ranking of professional tennis players. *Expert Systems with Applications*, 39(5):4882–4889. https://doi.org/10.1016/j.eswa.2011.10.004.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, 53:49–57. https://doi.org/10.1016/j.omega.2014.11.009.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281. https://doi.org/10.1016/0022-2496(77)90033-5.

Saaty, T. L. (1987). Rank according to Perron: A new insight. *Mathematics Magazine*, 60(4):211–213. https://doi.org/10.1080/0025570X.1987.11977304.

Shiraishi, S., Obata, T., and 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. https://doi.org/10.15807/jorsj.41.404.

Szádoczki, Zs., Bozóki, S., and Tekile, H. A. (2022). Filling in pattern designs for incomplete pairwise comparison matrices: (Quasi-)regular graphs with minimal diameter. *Omega*, 107:102557. https://doi.org/10.1016/j.omega.2021.102557.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286. https://doi.org/10.1037/h0070288.

Wang, H., Peng, Y., and Kou, G. (2021). A two-stage ranking method to minimize ordinal violation for pairwise comparisons. *Applied Soft Computing*, 106:107287. https://doi.org/10.1016/j.asoc.2021.107287.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442. https://doi.org/10.1038/30918.

Wei, T. H. (1952). *Algebraic foundations of ranking theory*. PhD thesis, University of Cambridge. https://www.repository.cam.ac.uk/handle/1810/250988.

Williams, L. V., Liu, C., Dixon, L., and Gerrard, H. (2021). How well do Elo-based ratings predict professional tennis matches? *Journal of Quantitative Analysis in Sports*, 17(2):91–105. https://doi.org/10.1515/jqas-2019-0110.