# A graph-based model to discover preference structure from choice data

**Cristóbal De La Maza (cdelamaz@andrew.cmu.edu)**
**Alex Davis (alexdavis@cmu.edu)**
**Cleotilde Gonzalez (coty@cmu.edu)**
**Inês Azevedo (iazevedo@cmu.edu)**

Carnegie Mellon University, Pittsburgh, PA 15213 USA

**Abstract**

In this paper we demonstrate how to use graph matching to uncover heterogeneity in the structure of preferences across a population of decision-makers. We propose a novel non-parametric approach to formally capture the concept of preference structure using preference graphs, thereafter clustering decision-makers based on graph embedding methods. We explore the approach with simulated choice and empirical data from the most common classes of economic and psychological models. The approach uncovers heterogeneity in preference structure across a variety of dimensions, without requiring any prior knowledge of those structures.

**Keywords:** Heuristics; Preference Structure; Graph Matching; Clustering; Transitivity

## Introduction

The study of preferences and the concept of rational choice have been relevant for cognitive science from the early conception of the fields (Simon, 1956). Theories of preferences are often concerned with the invariant axioms that describe how people decide, or the *structure* of preference. For over a century, the dominant paradigm has been a set of axioms that are necessary and sufficient for behavior to be consistent with the maximization of a well-behaved utility function, an idea dating back to the nineteenth century theorist Jeremy Bentham (Bentham, 1879). This paradigm requires decision-makers to be able to consistently rank any set of alternatives that they come across (Pareto, 1906). This well-behaved description of preference may work well in simple environments, but it is not clear how accurately represents preferences in complex, naturalistic settings. Humans are largely heterogeneous, with preferences that vary over time, and are often inconsistent (Tsetsos, Chater, & Usher, 2012). Large amounts of data are currently available that document choices people make in naturalistic settings. For example, information about purchase decisions, movie selections, and transportation patterns is widely available. This trend calls for new ways to determine insights from human preferences in the presence of large heterogeneity of naturalistic choices.

The axioms that define well-behaved preferences are both simple and quite powerful (Von Neumann & Morgenstern, 1944). The most relevant ones state that first, all alternatives must be comparable, making the preference relation *complete*. Second, preferences must be *transitive*. With these conditions it is possible to define a rank ordering of the alternatives according to the decision-maker's preferences, and there exists an ordinal utility function that corresponds to that ranking. Over the years, this dominant paradigm has not gone without challenge. Researchers in the decision sciences have found that, in many circumstances, preferences are not always well-behaved. Many descriptive theories have proliferated to explain deviations of human behavior from utility maximization. This includes ground-breaking work on bounded rationality, where decision-makers use short-cuts to deal with the limits of human cognitive capacities (Simon, 1972). For example, the cognitive burden of selecting the best alternative, considering all potential costs and benefits of each alternative, is at best psychologically implausible (Fischhoff, 2005). Instead, humans use simple rules or heuristics (Payne, Bettman, & Johnson, 1993; Gigerenzer, Todd, ABC Research Group, et al., 1999). For example, one psychologically plausible way to deal with complex choices is to simplify the task by choosing based on the attribute that is most important to the decision-maker, only examining other attributes if alternatives are sufficiently close on that attribute to be psychologically "tied". Tversky's lexicographic semiorder is such a process and can lead to intransitive behavior (Tversky, 1969).

However, there is a blind spot in choice modelling research that limits our possibilities to discover heuristic structures (Maturana & Varela, 1987). Researchers develop precise tests of their proposed models, with that testing limited to a priori defined patterns. As a result, patterns of choices are classified as either fitting a known model or not. For example, thus far structural tests of preference have been limited to specific patterns known a-priori, such as weak stochastic transitivity or the triangular condition (Regenwetter, Dana, & Davis-Stober, 2010). While this approach is promising and theory-driven, it potentially misses structures not previously considered. Some decision processes are clearly identifiable a-priori, others might not. We are looking at the choice process too closely and at the same time partially blocking our sight by using tools that are not general enough (Maturana & Varela, 1987). What is needed is an approach that can determine preference structure from choice data even when those data are inconsistent with prior models, suggesting at the same time new structures to psychological researchers or confirming old ones, and lending strength to welfare analysis or undermining it.

Next we describe our approach. The current research proposes a novel non-parametric model to formally capture the concept of preference structure using preference graphs, clusters decision-makers based on that structure, and can represent types of preferences currently not possible in existing frameworks (e.g. incomparability (Von Neumann & Morgen-

stern, 1944)). Because the approach clusters decision-makers with the same structural pattern of preferences, we provide unified method that may account for disparate preference patterns. The paper is structured as follows: we first present our graph-based model; then we test the method in simulations and in a new empirical implementations of a classic experiment in decisions between risky prospects; finally we discuss our results and present limitations of the method.

## Discovering preference structure heterogeneity

### Preference representation as graphs

Both classical utility models and newer descriptive theories imply specific patterns of choices, or *preference structures*. In this work, we exploit the idea that preference structures can be represented as *preference graphs* (Bouyssou & Vincke, 2010). For example, classical utility maximization can be represented as a *completely connected chain* (Varian, 1983; Afriat, 1972). This is, of course, not the only preference structure. For example, a lexicographic semiorder results in cyclic preferences when decision-makers change the weights they apply to attributes of alternatives (Tversky, 1969).

Graphs are a general way to represent binary relations among elements of discrete sets, including preference relations (Bouyssou & Vincke, 2010). Consider a graph $G = (V,E)$ with vertex set $V$ and edge set $E$. In a *preference graph* the vertices are interpreted as alternatives and edges as binary relations between alternatives where, for all pairs of alternatives, one and only one of the following three edges exists between them (Bouyssou & Vincke, 2010): i) if $a \succ b$, the decision maker strictly prefers $a$ over $b$, then $a \rightarrow b$ and not $b \rightarrow a$ (*strict preference* or $a \ P \ b$). If instead, $b \succ a$, the decision maker strictly prefers $b$ over $a$ ($bPa$), then $b \rightarrow a$ and not $a \rightarrow b$; ii) If $a \sim b$, the decision maker is indifferent between $a$ and $b$, then $a - b$ are connected by an undirected edge (*indifference* or $a \ I \ b$). This can also be represented as $a$ is preferred to $b$ and $b$ to $a$ or $a \leftrightarrow b$; and iii) If $a$ is incomparable with $b$, then no edge between $a$ and $b$ exists (*incomparability* or $a \ J \ b$). Figure 1 describes a graph representation of preferences.

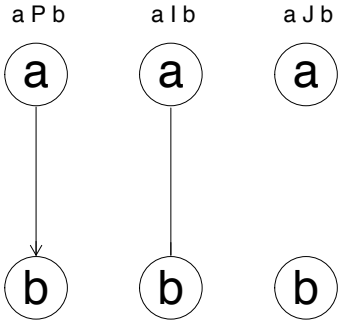


Figure 1: Preference relations in binary choice

An equivalent representation is an adjacency matrix $A = (a_{ij}) \in \{0,1\}^{nxn}$ where $a_{ij} = 1$ if $(i,j) \in E_A$ and $B = (b_{ij}) \in \{0,1\}^{nxn}$ where $b_{ij} = 1$ if $(i,j) \in E_B$, indicating preference from $i$ to $j$. Reflexive loops are usually omitted, meaning the main diagonal of the adjacency matrix has only zeros. In this work we focus on a particular type of preference graphs, namely tournaments, where every alternative is compared and only strict preference is allowed giving a complete directed graph (Bouyssou & Vincke, 2010). The number of vertices in a tournament indicates the order. In the simplest case we find transitive *tournaments* (Moon, 2015), where all relations are strict preferences and there are no cycles. Following (Bouyssou & Vincke, 2010), consider a total order giving a tournament with an adjacency matrix that will show only zeros below the diagonal. A weak order instead, will allow indifference between alternatives and hence giving a tournament with a stepped shape adjacency matrix below the diagonal. For an irrational decision maker, cycles will be observed, giving a tournament with an adjacency matrix that will show elements above and below the diagonal.

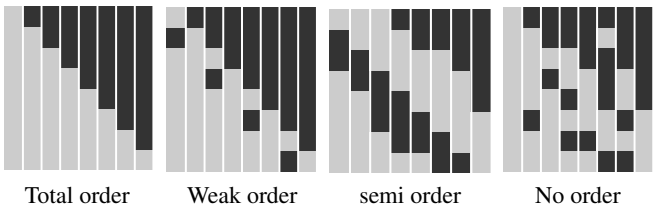


Figure 2: Tournaments with different order structure. Adjacency matrices are colored to ease interpretation with ones in black and zeros in grey.

Another representation is a format used by Moon (Moon, 2015), where graphs are drawn based on their *score vector*, which is the number of times each alternative is preferred over other alternatives. For example, with four alternatives, the maximum score is 3 (an alternative that is preferred to all others), and the minimum is zero (an alternative preferred to no others). A score vector of $s = [3,2,1,0]$ is a complete ranking of the alternatives, or a *chain*. It is drawn by sorting the score vector from highest score at the top to lowest score at the bottom, then adding down arrows from top to bottom. If arrows are omitted (to avoid clutter), this means that the upper alternative is preferred to the lower alternative. Inconsistencies are denoted by upward arrows, where an alternative with a lower score is strictly preferred to an alternative with a higher score. As shown in Figure 3, there are exactly 4 non-isomorphic structures for tournaments of four alternatives (Davis, 1954): a chain, a cycle among the top 3 alternatives, a cycle among the bottom 3 alternatives, and a single long cycle.

Notice that these structures have very different implications for decision-analysis. Given a choice between any subset of four alternatives, a decision-maker with a chain provides a ranking consistent with the global ranking over four alternatives. A decision-maker with a cycle at the top can consistently rank only the worst alternative, and likewise, the decision-maker with a cycle at the bottom can consistently

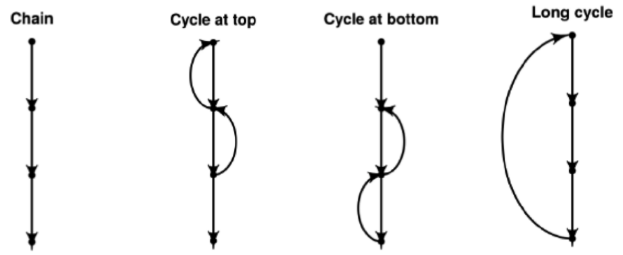


Figure 3: Tournaments on four alternatives. The score vectors are: chain $s = [3,2,1,0]$, cycle at top $s = [2,2,2,0]$, cycle at bottom $s = [3,1,1,1]$, long cycle $s = [2,2,1,1]$.

rank only the best alternative. A decision-maker with the long cycle has a consistent ranking over any subset of alternatives, but no global ranking.

## Preference graph similarity

Our primary analytical tool is a method of calculating the *distance* between graphs. Formally, a common distance metric between two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, is the minimum number of edges that need to be rearranged to make them isomorphic, known as the Hamming distance $d_H(G_1, G_2) = ||vec(G_1) - vec(G_2)||_1$ (Hamming, 1950). Decision-makers that have a small Hamming distance between their preference graphs tend to choose similar alternatives, or have similar *preference content*. For a sample of $n$ individuals, we can store the Hamming distance between all pairs of decision-makers in a symmetric $n \times n$ dissimilarity matrix $D$. From $D$, a weighted dissimilarity kernel $K$ can be constructed, with values between zero and one (Kevin, 2012). We use standard graph similarity tools to identify clusters of graphs with similar content. This approach is formally equivalent to Coombs' multidimensional unfolding (Coombs & Kao, 1960).

Preference structure cannot be obtained from these Hamming distance computations. For example, two chain preference graphs of equal size with opposing preference content will have a Hamming distance equal to the total number of unordered pairs of vertices $\binom{n}{x}$. Even though they are both chains (identical structures), the Hamming distance indicates that they are as dissimilar as possible. Thus, we need a metric that indicates that these graphs have the same structure and hence that there is a *structural distance* of zero between them. Two graphs have a structural distance of zero if they are isomorphic (Babai & Luks, 1983), meaning there is a bijection $f : V_1 \to V_2$ such that the edges of all pairs of vertices $u, v \in V_1$ in $G_1$ have the same edges for $f(u), f(v) \in V_2$ in $G_2$ (and vice versa). An *automorphism* of a graph $G$ is a graph that is isomorphic to $G$, and the *automorphism group* $Aut(G)$ is all of the graphs that are isomorphic to $G$ (Babai & Luks, 1983). We can test whether two graphs are isomorphic by checking whether any of their automorphisms are isomorphic. This is a well studied problem in computer science, called the *graph isomorphism problem* (Babai & Luks, 1983). The minimum Hamming distance between two graphs across all combinations of their automorphisms gives their *structural distance* $d_S$ (Butts & Carley, 2005): $d_S(G_1, G_2) = min(d_H(Aut(G_1), Aut(G_2)))$. If two graphs are similar (but not isomorphic), their structural distance should be small. Clusters of decision-makers with small distances between each other, indicates a common preference structure in a population of decision-makers, partially masked by noise.

## Inexact graph matching

With a few alternatives the structural distance between graphs can be quickly calculated using exhaustive search. As the number of alternatives grows, exhaustive search becomes unfeasible. In general, the problem of calculating structural distance is NP-hard (Livi & Rizzi, 2013), requiring approximation techniques for large graphs with more than 8 alternatives. To make this approximation feasible, we recast the structural distance calculation as an *inexact graph matching problem* (Livi & Rizzi, 2013), where the objective is to find the permutation matrix $\mathrm{P}^*$ over the space of permutations that makes two adjacency matrices $A$ and $B$ as similar as possible (Livi & Rizzi, 2013):

$$\mathrm{P}^* = \underset{\mathrm{P} \in \mathcal{P}}{\operatorname{argmin}} \; f(P) = \mathrm{dis}_{A \to B}(\mathrm{P}) = ||A - P^T B P|| \qquad (1)$$

where $A, B$ are the adjacency matrices for the preference graphs of two decision-makers, and $P \in \mathcal{P}$ is in the set of permutation matrices $\mathcal{P}$. If the chosen norm is the Frobenius $L2$ norm squared the problem is know as quadratic assignment (QAP) with non-deterministic polynomial time complexity (Vogelstein et al., 2011). Instead, we replaced the objective function $f(P)$ by the identity $-tr(\mathrm{A}\mathrm{P}B^T\mathrm{P}^T)$ which leads to a non-convex problem where $\nabla^2 f(P) = B \otimes A + B^T \otimes A^T$ is not positive definite, relaxing at the same time the non-convex restriction $\Pi \in P$ and replacing $P$ by its convex hull $\mathcal{D}$, where $\mathcal{D}$ is the set of doubly stochastic matrices (Vogelstein et al., 2011). We solved this problem with Frank-Wolfe algorithm (Frank & Wolfe, 1956; Vogelstein et al., 2011).

## Clustering

Once content and structural distances $d_s$ are determined for preference graphs of each pair of decision-makers, the matrix of pairwise structural (or hamming) distances between the graphs of decision-makers can be analyzed using traditional clustering techniques to classify decision-makers into groups with similar preference content and structure. Nonetheless, nothing ensures that clusters from content and structural dissimilarities will overlap. Therefore, we need to account for both structural and content dissimilarities simultaneously in the clustering stage. To do so, we first embed each dissimilarity matrix in a lower dimensional space and hereafter we bound columns of the resulting embeddings in an $n \times d$ matrix with information about content and structure for each decision-maker, with $d$ the sum of dimensions of the embeddings of both dissimilarity matrices or *embedding fusion*.

We begin by using classical multidimensional scaling to project each distance matrix onto a lower dimensional space (Torgerson, 1952), but based on its superior performance we finally used an autoencoder (Wang, Yao, & Zhao, 2016). An autoencoder is a neural network model that maps or encodes input space $x$ into a lower dimensional space $h(x)$ at its output layer and then reconstructs or decodes the original input space as $\hat{x}(h)$ (Wang et al., 2016). We pretrained the model with a Restricted Boltzmann Machine (Hinton & Salakhutdinov, 2006). Finally, to achieve a robust solution, we used k-medians algorithm to determine cluster allocation (Singh, Yadav, & Rana, 2013). We used the gap-statistic to determine the number of clusters $k$ (Tibshirani, Walther, & Hastie, 2001). To provide a more general solution, clusters are merged when necessary.

## Preference structure in simulation

We first describe the results of simulations designed to illustrate the method. In our simulation we evaluate our model's ability to separate a popular psychological model, the lexicographic semiorder (Tversky, 1969), from the more traditional expected utility maximization (Von Neumann & Morgenstern, 1944). As an example, consider choosing between pairs of gambles shown in Table 1 from Tversky's classic paper on intransitive preferences (Tversky, 1969), along with three additional gambles (f-h) added to increase graph matching difficulty. The choice task is presented in Figure 4.

Table 1: Gambles from Tversky's (Tversky, 1969) experiment 1 (a-e) plus three added for the simulation (f-h)

| Gamble | Probability | Payoff | Expected Value ($) |
|---|---|---|---|
| a | 7/24 | 5.00 | 1.46 |
| b | 8/24 | 4.75 | 1.58 |
| c | 9/24 | 4.50 | 1.69 |
| d | 10/24 | 4.25 | 1.77 |
| e | 11/24 | 4.00 | 1.83 |
| f | 12/24 | 3.75 | 1.88 |
| g | 13/24 | 3.50 | 1.894 |
| h | 14/24 | 3.25 | 1.895 |

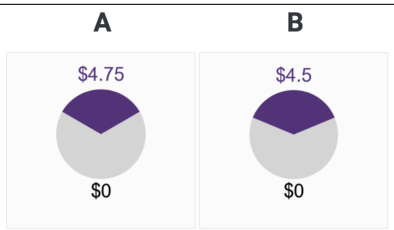


Figure 4: Choice set example alternatives b vs c

Subjects that choose based on expected value should prefer $a \succ b \succ c \succ d \succ e$, and should have a complete transitive order. Tversky hypothesized that someone following a lexicographic semiorder decision rule would first choose based on differences in gambles probabilities. If the difference in probabilities is small enough, the decision maker would switch to the next attribute and choose based on differences in payoffs. This would result in an intransitive sequence $a \succ b \succ c \succ d \succ e$ and $e \succ a$. To demonstrate that our approach can reliably cluster decision-makers into groups based on the structure of their preferences, we generated graphs for 100 decision-makers, 37 with lexicographic preferences, 33 with risk neutral expected value maximizer preferences, and 30 that would choose at random. We first mapped the simulated choices in an adjacency matrix, then computed dissimilarity matrices between adjacency matrices and finally we identified clusters of graphs with similar preference content and structure. Figure 5 summarizes our method. As exposed in Figure 5 we can separate successfully lexicographic semiorders from those who are expected value maximizers.

Here, we assumed subjects would choose deterministically. A deterministic decision rule will provide structures that are quite easy to distinguish from others because, under all circumstances, the same graph structure will emerge. Nonetheless, noise in the decision process can make preference structures harder to distinguish. For example, even though a decision rule such as expected value maximization is used, intransitive behavior is still observable if alternatives are harder to compare, confounding the later decision rule with random choice. In an extension of our simulation we observed that above a certain noise level, clustering becomes unfeasible.

## An empirical test of the model

We extended Tversky's classic experiment examining lexicographic semiorders (Tversky, 1969). Participants choose between the pairs of gambles shown in Table 1 from Tversky's classic paper on intransitive preferences (Tversky, 1969), along with the three gambles (f-h) considered in the simulation experiment and two additional gambles (i-j) where a higher probability is negatively correlated with a higher expected value (in i there is 15/24 chance of winning $3, and in j a 16/24 chance of winning $2.75). Following (Tversky, 1969), probabilities were presented as pie charts without numeric information. We presented participants with all pair combinations (45 pairs) in three repetitions with the order randomized.

We recruited 200 participants using Amazon Mechanical Turk (Mturk). Inclusion criteria were the following: age of at least 18 years, IP address in the U.S. and completion of more than 100 hits with an approval rate of 95% or higher. We provided a payment of $1 per participant and a $0.5 bonus if the participant answered an attention check correctly. The attention question was a choice set with a deterministically dominated alternative. 95% of the 200 participants was paid the bonus. Clustering by content and structure six clusters emerged: four groups with chains and two group with cycles. Structural and content heterogeneity, rather than homogeneity, is the primary takeaway. Chain graphs, are the most popular pattern. Figure 6 shows the expected adjacency matrices for the preference graphs in each cluster. Alternatives were prearranged so a lower triangular adjacency matrix indicates choices based strictly on probabilities and an upper triangular matrix indicates choices based strictly on payoffs.

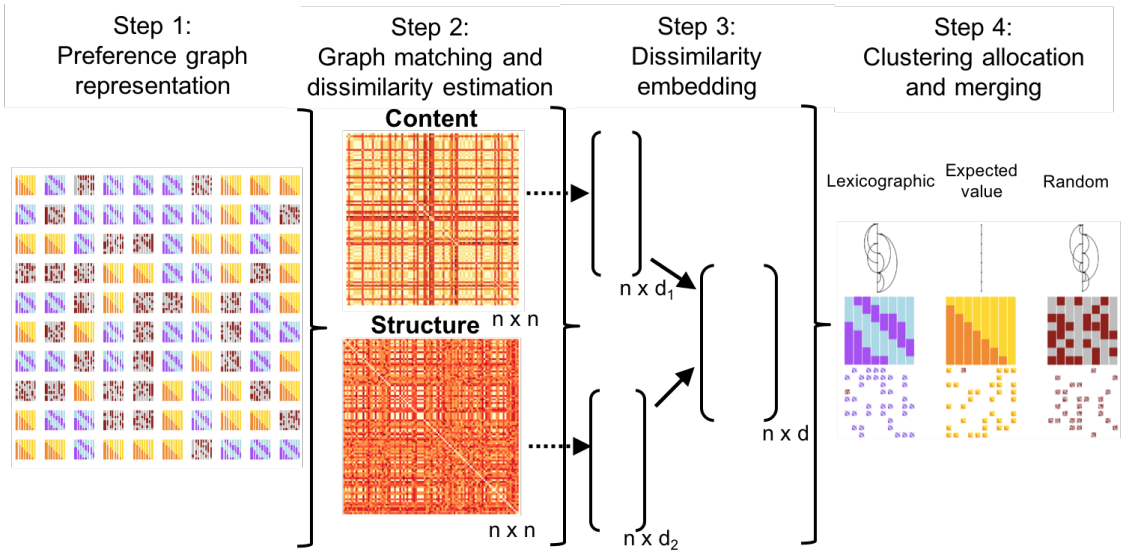


Figure 5: The schema summarizes the four steps of our method for a simulated sample of 100 decision-makers.

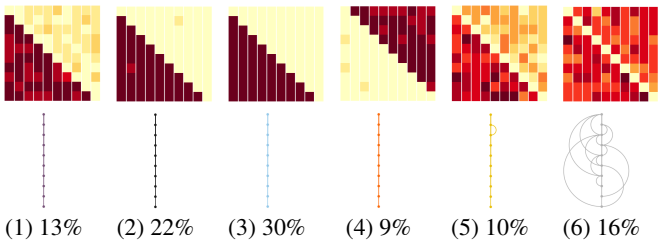


Figure 6: Weighted expected adjacency matrix in each cluster for the transitivity task. We used a color scale to easy ease interpretation with adjacency matrices colored from one in darker tones and zeros in lighter tones. Proportion in each cluster is presented in last row.

To further analyze choices, we used multinomial logit models per clusters (McFadden, 1973). For all clusters a decision rule based on a single attribute (either probabilities or payoffs) is more likely than an expected value rule. Decision-makers in clusters 1, 2 and 3 preferred the alternative with a higher probability in 87%, 96% and 100% of the choices. It is possible that this clusters respond to the same decision rule with differences in discriminant ability. We must highlight that given that probabilities are not numerically stated, recognizing the alternative with a higher probability in all problems as in cluster 3 requires a superior classification skill. We decided to merge this clusters in one group. Decision-makers in cluster 4 consistently chose the alternative with a higher payoff 93% of times, indicating a single attribute decision rule based on payoffs. Clusters 5 and 6 seem to respond to a different decision process. In cluster 6 multiple cycles are observed. The proportion of choices in cluster 6 favoring the option with the higher probability is significantly distinct from 50% ruling out random choice (p-value $\leq 0.01$). Although details of the different choice rules remain uncovered, we observed a clear tendency to choose based on probabilities (Lichtenstein & Slovic, 1971; Birnbaum & Gutierrez, 2007; Brandstätter, Gigerenzer, & Hertwig, 2006). It seems the data is more consistent with a lexicographic order (up to noise) in the sense of Fishburn (Fishburn, 1971) than a lexicographic semiorder as proposed by Tversky (Tversky, 1969).

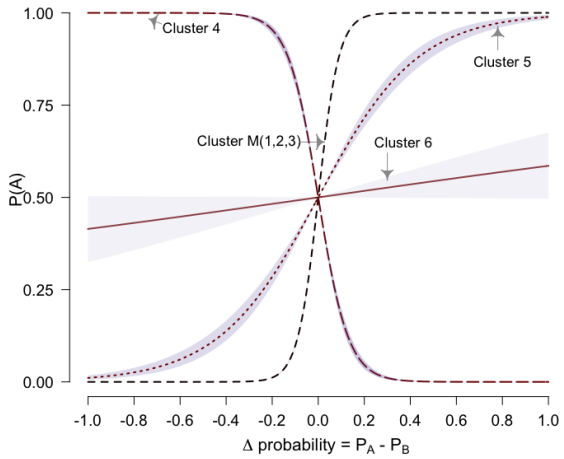


Figure 7: Logit probabilities of choosing the alternative with a higher probability of winning (A) per cluster.

## Limitations and future work

It must be noted that, although we used a majority rule to define preference across repetitions, preference strength can be reflected using weighted adjacency matrices for each individual. Among other limitations, clustering always has some arbitrariness. For example, the number of dimensions to embed the dissimilarity matrices in a lower dimensional space is defined using the elbow method. Determining the number of dimensions in the optimization process can offer a potential improvement. Future applications should also developed better ways of determining the number of clusters and herein merging similar clusters. The experimental design also provides some challenges. The number of pairwise comparisons required to complete a tournament grows exponentially with the number of alternatives, increasing the risk of observing mental fatigue. A new experimental paradigm needs to be develop in order determine apriori the minimal number of questions required to recover preference structure. Although our method is valid for any type of pairwise comparison, empirical tests should be extended to other experimental domains.

Also we could extend the method to experiments with more than two alternatives expanding choice data by rank-order explosion (Louviere et al., 2008).

## Acknowledgments

We are in debt with Nalyn Sriwattanakomen, Orsolya Kovacs and Brian Sergi. We acknowledge funding from the National Science Foundation, Decision Risk and Management Science, Award number 1530479, Center for Climate and Energy Decision Making (Grant number SES-1463492); CONICYT and CMU EPP Department.

## References

Afriat, S. (1972). Efficiency estimates of production functions. *International Economic Review*, *13*, 568–598.

Babai, L., & Luks, E. M. (1983). Canonical labeling of graphs. In *Proceedings of the fifteenth annual acm symposium on theory of computing* (pp. 171–183).

Bentham, J. (1879). *An introduction to the principles of morals and legislation*. Clarendon Press.

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semi-order. *Organizational Behavior and Human Decision Processes*, *104*(1), 96–112.

Bouyssou, D., & Vincke, P. (2010). Binary relations and preference modeling. *Decision-making Process: Concepts and Methods*, 49–84.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: making choices without trade-offs. *Psychological Review*, *113*(2), 409–432.

Butts, C. T., & Carley, K. M. (2005). Some simple algorithms for structural comparison. *Computational & Mathematical Organization Theory*, *11*(4), 291–305.

Coombs, C. H., & Kao, R. C. (1960). On a connection between factor analysis and multidimensional unfolding. *Psychometrika*, *25*(3), 219–231.

Davis, R. L. (1954). Structures of dominance relations. *Bulletin of Mathematical Biology*, *16*(2), 131–140.

Fischhoff, B. (2005). Cognitive processes in stated preference methods. *Handbook of environmental economics*, *2*, 937–968.

Fishburn, P. C. (1971). A study of lexicographic expected utility. *Management Science*, *17*(11), 672–678.

Frank, M., & Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics (NRL)*, *3*(1-2), 95–110.

Gigerenzer, G., Todd, P. M., ABC Research Group, t., et al. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell Labs Technical Journal*, *29*(2), 147–160.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, *313*(5786), 504–507.

Kevin, M. (2012). *Machine learning: a probabilistic perspective*. The MIT press.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of experimental psychology*, *89*(1), 46.

Livi, L., & Rizzi, A. (2013). The graph matching problem. *Pattern Analysis and Applications*, *16*(3), 253–283.

Louviere, J., Street, D., Burgess, L., Wasi, N., Islam, T., & Marley, A. A. (2008). Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of choice modelling*, *1*(1), 128–163.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. New Science Library/Shambhala Publications.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press: New York.

Moon, J. W. (2015). *Topics on tournaments in graph theory*. Courier Dover Publications.

Pareto, V. (1906). *Manuale di economia politica* (Vol. 13). Societa Editrice.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2010). Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in psychology*, *1*(148), 1–15.

Simon, H. A. (1956). Rational choice and the structure of the environment [Journal Article]. *Psychological Review*, *63*(2), 129-138. doi: 10.1037/h0042769

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, *1*(1), 161–176.

Singh, A., Yadav, A., & Rana, A. (2013). K-means with three different distance metrics. *International Journal of Computer Applications*, *67*(10).

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423.

Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, *17*(4), 401–419.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal [Journal Article]. *PNAS*, *109*(24), 9659-9664. doi: 10.1073/pnas.1119569109

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*(1), 31-48.

Varian, H. R. (1983). Non-parametric tests of consumer behaviour. *The review of economic studies*, *50*(1), 99–110.

Vogelstein, J. T., Conroy, J. M., Lyzinski, V., Podrazik, L. J., Kratzer, S. G., Harley, E. T., ... Priebe, C. E. (2011). Fast approximate quadratic programming for large (brain) graph matching. *arXiv preprint arXiv:1112.5507*.

Von Neumann, J., & Morgenstern, O. (1944). Theory of games and economic behavior.

Wang, Y., Yao, H., & Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing*, *184*, 232–242.