British Journal of Mathematical and Statistical Psychology (2009), 62, 663–682

The British Psychological Society

www.bpsjournals.co.uk

The metric-frequency measure of similarity for ill-structured data sets, with an application to family therapy

Célia M. D. Sales[1] and Peter P. Wakker[2]*

[1]Department of Psychology and Sociology, Universidade Autónoma de Lisboa, Rua de Santa Marta, Lisboa, Portugal

[2]Economics Department, Erasmus University, Rotterdam, The Netherlands

Similarity measures have been studied extensively in many domains, but usually with well-structured data sets. In many psychological applications, however, such data sets are not available. It often cannot even be predicted how many items will be observed, or what exactly they will entail. This paper introduces a similarity measure, called the metric-frequency (MF) measure, that can be applied to such data sets. If it is not known beforehand how many items will be observed, then the number of items actually observed in itself carries information. A typical feature of the MF is that it incorporates such information. The primary purpose of our measure is that it should be pragmatic, widely applicable, and tractable, even if data are complex. The MF generalizes Tversky's set-theoretic measure of similarity to cases where items may be present or absent and at the same time can be numerical as with Shepard's metric measure, but need not be so. As an illustration, we apply the MF to family therapy where it cannot be predicted what issues the clients will raise in therapy sessions. The MF is flexible enough to be applicable to idiographic data.

1. Introduction

The perception of similarity is basic for virtually every mental activity. Thus, quantitative measures of similarity have been developed in many disciplines, ranging from biology (Lord, Stevens, Brass, & Goble, 2003), computer science (Xiao & Zhang, 2001), economics (Gilboa & Schmeidler, 2001), linguistics (Bailey & Hahn, 2001; Navarro, 2007), medicine (Légaré, O'Connor, Graham, Wells, & Tremblay, 2006), and statistics (Shepard, 1962; Davison, 1992) to psychology, the main field of interest for our study.

For quantitative data, commonly used similarity measures consider the differences in scores on a number of predetermined metric items and aggregate those differences into

* Correspondence should be addressed to Professor Peter P. Wakker, Econometric Institute, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands (e-mail: Wakker@few.eur.nl).

DOI:10.1348/00071008X376070

an overall metric, often the Euclidean distance (Shepard, 1987; Torgerson, 1958). Tversky and Gati (1982) provided an ordinal extension. In purely qualitative settings, commonly used measures increase with the number and importance of the common features shared and similarly decrease with the distinctive features (Tversky, 1977). An intermediate approach is used in fuzzy set theory, where items are interpreted as features and their scores, between 0 and 1, reflect a vague membership. Then metric techniques similar to those mentioned above can still be used (Santini & Jain, 1999). Theories of similarity for pairwise discrimination probabilities are similarly intermediate, allowing for discrete object spaces but having continua of probabilities (Dzhafarov & Colonius, 2007).

Similarity measures have also been used to indicate the transformation distance from one subject to another (Barthélemy & Mullet, 1996; Hahn, Chater, & Richardson, 2003). This approach requires concrete cognitive structures of representations. Similarity measures that consider structural aspects have been widely studied in computer science (Resnik, 1999). Such structural aspects are, however, not assumed to be available in our study, and we will accordingly not consider structural approaches in this paper.

In many applications, the data are too unstructured and too unpredictable for traditional measures to be applicable. This occurs especially in psychology, where it often cannot be predicted how many variables will be observed and what their content may be (Lee, 2001). We introduce a measure of similarity, the metric-frequency (MF) measure, for such situations. It combines aspects from both the metric and the qualitative approach, by considering numerical differences when available but also paying special attention to the mere presence or absence of features and the information captured by corresponding frequencies. Indeed, if the number of items observed is unknown a priori, then the actual number observed contains information. Our measure will incorporate such information. The importance of combining metric and qualitative aspects was emphasized by Carroll (1976, p. 462): 'Since what is going on inside the head is likely to be complex, and is equally likely to have both discrete and continuous aspects, I believe the models we pursue must also be complex, and have both discrete and continuous components.' Navarro and Lee (2003) first proposed a model that combines both components. We present an alternative to their model that is targeted towards situations in which the number of aspects is unpredictable and thus also carries information.

The primary purpose of the MF is to be widely applicable. In particular, we wish to be able to handle situations that are ill structured and complex, as alluded to by Carroll. The measure has been designed to be easily applicable and pragmatic rather than theoretically complex. We will first illustrate the tractability of the MF through simple hypothetical examples. Then we describe an application in family therapy in more detail.

Our development of the MF did, in fact, arise from a problem in family therapy regarding the application of idiographic change measures. Idiographic change measures are especially useful for capturing the range and uniqueness of clients' conditions (Ashworth *et al.*, 2005; Elliott, Slatick, & Urman, 2001; Evans, Hughes, & Houston, 2002; Greenberg, 1986; Greenberg & Pinsof, 1986; Hill & Lambert, 2004; Levine & Luborsky 1981; Wagner & Elliott, 1999). They are tailored to each client's problems (target complaint measures: Battle *et al.*, 1966) or goals (goal attainment scaling: Kiresuk & Sherman, 1968). Items are determined by the client, or by the client in collaboration with the therapist. Each questionnaire is unique as it varies in the number of items and in their content. The use of idiographic measures in family therapy is especially useful

because each family member has a different view on the existing problems. For the therapist, it is important to know this diversity of viewpoints and to evaluate their changes during treatment (Sales *et al.*, 2006; Sales, Gonçalves, Fragoeiro, Noronha, & Elliott, 2007).

Idiographic measures result in ill-structured data sets: (a) all family members have a different questionnaire corresponding to their personal view of the existing problems; (b) there is no control over the content of the items that each family member can raise; (c) there is no limit to the number of items that are conceivable; (d) each person is free to add new items or delete previous items in subsequent questionnaire administration; (e) each item is weighted on a Likert scale. Analysing the convergence of family members' judgements on the existing problems over the course of the treatment then calls for a flexible quantitative measure of similarity. This led us to develop the MF.

The next section defines the MF. Section 3 explains the definition, and Section 4 illustrates its properties by means of hypothetical examples. Section 5 presents an application of the MF in family therapy, and Section 6 concludes.

2. The metric-frequency measure of similarity defined

As an illustrative example, we assume that two members of a family, the father and the mother, will be compared. We investigate the extent to which they have a similar perception of the problems in the family. Each item designates a problem in a family. We ask the father and the mother what the problems in their family are. The type and number of items that can be raised in such situations is so diverse that we cannot determine a prior set of all potential items in any tractable manner. Hence, the MF will not require such a prior set. In this application, we explain the measure for the example of the father and mother. It can, however, obviously be applied irrespective of what the items, scores, and subjects actually are. The precise definition of these concepts constitutes an important part of the modelling stage. We assume here that they have been defined, with further requirements specified next.

We assume that we obtain a numerical score for each item – say, concerning the seriousness of the problem raised. In a qualitative setting where only the presence or absence of an item is observable, and there is no metric structure, we use only the scores 1 (present) and 0 (absent) and then can still use the MF, as will be illustrated later.

We use a minimal score, 0 in what follows, to designate the absence of an item. For simplicity of analysis, we only consider cases where there are no scores better than absence. In other words, every explicit reference to a problem is a bad sign and not being raised (absence) is the most favourable observation that is possible for a problem. There can be many reasons why a problem was not raised even though it was serious. It is, therefore, desirable to allow as much as possible for all relevant problems to be raised. With this understood, the absence of a problem should not be treated as a missing observation but rather it is the best score possible. In the same way, a medical check-up finding no problems is the best score possible even if we can never be sure that there are no hidden undetected problems.

If there are some items that contribute positively and other items that contribute negatively to a topic of interest, then the MF should be applied separately to the positive items and to the negative items. The resulting measures can then be integrated in ways appropriate to the context considered. For example, if we ask for positive and negative points in the family, then we separately measure the similarity of the father and the

mother regarding the positive and the negative items. Theories about the integration of similarities in a two-sided scale are a topic for future research.

We also do not consider cases where both positive and negative scores can arise within one item. If such items exist, then they can be separated into one item that records any positive score and another item that records any negative score. Many studies have demonstrated that for variables that can take different signs, positive parts are usually perceived in a qualitatively different manner than negative parts (Kahneman, 2003). Thus, separate treatments of the positive and negative parts of variables are appropriate in many applications. In summary, we assume that all items contribute positively to the topic of study. Whether or not an item is present or absent will receive special treatment in our formula, which is explained in detail later.

We first display the definition of our measure and then define the symbols used. A detailed explanation is given in the following section. The *metric-frequency* (MF) *measure* is defined as

$$\frac{1}{2}\frac{\sum(1-|diff|)}{j+f+m}+\frac{1}{4}+\frac{1}{4}\sqrt{\frac{j}{N}}-\frac{1}{4}\left|\sqrt{\frac{f}{N}}-\sqrt{\frac{m}{N}}\right| \tag{1}$$

where the summation in the fraction is over all items raised by either father or mother;[1] $|diff|$ is the absolute value of the difference in 0–1 normalized scores that the two members assign to the item under consideration, with $1-|diff|$ the resulting similarity; j is the number of ('joint') items raised by both the father and the mother; f the number of items raised by the father and not by the mother; m the number of items raised by the mother and not by the father; and N an upper bound for the number of items that can be raised by one person. Software for calculating the measure can be downloaded from the second author's homepage.

3. The MF of similarity explained

The formula for the MF is explained in a number of stages and steps. Various intermediate indexes will be normalized to a 0–1 scale so as to facilitate their interpretations and comparisons. Table 1 gives an example of scores that have not yet been normalized, which is typically the format in which raw data come in. Items not raised receive a score of 0. Hence, the father will have score 0 for item H, and the mother will have that score for items E, F, and G.

3.1. Stage 1. Score similarity: The similarity based on scores

This stage explains the fraction in (1), and is as in most metric approaches to similarity measures.

Step 1.1 (Normalization of scores). We normalize the scores to a 0–1 scale by subtracting the minimum conceivable value (for the absence of an item, to be (re)scaled as 0) and then dividing by the maximum minus the minimum conceivable value. In Table 1, we

[1] Or, equivalently, over all conceivable items (if it is possible to list them), because those not raised will make a zero contribution to the summation.

Table 1. Items raised by father and mother, and their scores

Father		Mother	
Items	Scores	Items	Scores
A	7	A	5
B	6	B	6
C	1	C	2
D	1	D	1
E	3	H	1
F	2		
G	2		

Note. Here $j = 4$ (items A, B, C, D); $f = 3$ (items E, F, G); $m = 1$ (item H); $N = 20$ (explained later).

thus divide all scores by 7. The result is given in the second and third columns of Table 2. The other columns are explained later.

Table 2. Items raised by father and mother, their normalized scores, and differences of the scores of each item, $|diff|$

Items	Normalized scores of the father	Normalized scores of the mother	$\|diff\|$	Similarity $1 - \|diff\|$
A	1	5/7	2/7	5/7
B	6/7	6/7	0	7/7
C	1/7	2/7	1/7	6/7
D	1/7	1/7	0	7/7
E	3/7	0	3/7	4/7
F	2/7	0	2/7	5/7
G	2/7	0	2/7	5/7
H	0	1/7	1/7	6/7
Sum				45/7

Note. Similarity, $1 - |\text{diff}|$, is the opposite of the difference.

Step 1.2 (Similarity per item regarding the differences of scores). For each item, we calculate the absolute value of the difference of the scores of the father and the mother, $|diff|$, and the resulting similarity, $1 - |diff|$. For item C, for example, the difference is $1/7 - 2/7 = -1/7$. The absolute value is thus 1/7, with the resulting similarity $1 - 1/7 = 6/7$.

Step 1.3 (Combining the similarities of scores over all items into one measure, the score similarity). Take the sum of the similarities calculated in step 1.2:

$$\sum(1 - |diff|), \tag{2}$$

over all items raised by the father or the mother. This sum is 45/7 in Table 2. Deviating from Euclidean distances, but following Attneave (1950), we prefer not to sum squares so as not to give too much weight to outliers. The total number of similarities is $j + f + m$ ($= 4 + 3 + 1 = 8$ in Table 2). Normalize the sum obtained by dividing

by $j + f + m$, which results in

$$\frac{\sum(1 - |diff|)}{j+f+m}, \tag{3}$$

which is 0.80 for Table 2.[2] What has resulted is a 0–1 scaled similarity measure based only on the average differences of the scores of the father and the mother, as would be relevant for metric measures. It is the first component, the fraction, in (1). Visual inspection of the data suggests that a similarity of 0.80 on a 0–1 scale is plausible.

3.2. Stage 2. Frequency similarity: Similarity based on numbers of items raised

This stage analyses the special information captured by the mere presence or absence of items. We will treat the number of joint items and the number of distinct items separately, in steps 2.1 and 2.2. The results of these two steps in isolation are only intermediate steps. It is the interpretation of their combination that is relevant. This is discussed at the end of this stage.

Step 2.1 (Similarity based on the number of items raised jointly by the father and the mother). The similarity through the number of items raised by both the father and the mother is reflected by a number j/N, where N is a normalization factor that ensures that j/N, f/N, and m/N never exceed 1. N should be the same for all participants whose mutual similarity weights are calculated. Thus, it should exceed the maximum number of items raised by any single participant in the group considered. For instance, it can be the maximum number of conceivable items. Further discussion of N and its interpretation are given later. For now, let us say that, in our example, $N = 20$ has been chosen, so that $j/N = 4/20 = 0.2$.

Instead of the number j/N, we will use a transformation thereof, $\sqrt{j/N}$. The transformation is curved downwards (concave), and increases less for high values of j (and j/N) than for low values. Thus, an increase from $N = 1$ to $N = 2$ has more impact than an increase from $N = 17$ to $N = 18$, which is plausible. In our example, the transformation yields $\sqrt{0.2} = 0.45$. We have now quantified the information contained in the number of items raised by both the father and the mother.

Step 2.2 ((Dis)similarity based on the difference in the number of items raised by the father and the mother). If the father raises only a few items and the mother many, then this discrepancy in itself is a signal of low similarity. Because of the variable numbers of items raised, we cannot handle this phenomenon through scores on items not raised, but must handle it separately.

For the father the situation seems to be more serious than for the mother. The difference, taken as the absolute value, is $|f - m|$, and a normalized version is $|f/N - m/N|$. As in step 2.1, it is better to take square roots of the respective numbers, resulting in $|\sqrt{f/N} - \sqrt{m/N}|$. In our example, the result is $|\sqrt{3/20} - \sqrt{1/20}| = |0.387 - 0.224| = 0.16$. The similarity regarding this aspect is given by $1 - |\sqrt{f/N} - \sqrt{m/N}|$,

[2] In the exceptional case where both subjects raised no items, so that j, f, and m are all zero, this ratio is to be taken as 1.

which is 0.84 in our example. We have now quantified the information regarding the difference in the number of items raised by the father and by the mother.

Step 2.3 (Combining the similarities based on numbers of items into the frequency measure). This step combines the results of steps 2.1 and 2.2, by taking their average

$$\frac{\sqrt{(j/N)} + 1 - |\sqrt{(f/N)} - \sqrt{(m/N)}|}{2}, \tag{4}$$

which is $(0.45 + 0.84)/2 = 0.64$ in our example. It amounts to twice the terms in (1) except for the fraction. With four items raised by both father and mother, but the father raising two more items than the mother did, a similarity of 0.64, not far from the middle of the scale, has face validity. As in Tversky's (1977) feature-contrast model, the frequency measure developed in stage 2 is based only on the presence or absence of items. Unlike Tversky's model, our model maintains symmetry, leaving asymmetric generalizations (Johannesson, 2000; Saito, 1994) to future work.

3.3. Stage 3.The MF as overall similarity measure

The MF measure, finally, results as the half–half mid-point of the score similarity and the frequency similarity:

$$\frac{1}{2}\frac{\sum(1 - |diff|}{j + f + m} + \frac{\sqrt{(j/N)} + 1 - |\sqrt{(f/N)} - \sqrt{(m/N)}|}{2}, \tag{5}$$

which can be rewritten as (1). In the example, the MF is

$$1/2 \times 0.80 + 1/2 \times 0.64 = 0.72. \tag{6}$$

3.4. The number N

As explained in step 2.1, we want the number N to exceed the maximum number of items that were raised by any single person in the group whose mutual similarity weights are determined. This requirement can be relaxed if desired, but then either values of the measure outside the [0,1] interval should be accepted, or another normalization should be done at the end. Further, N serves as an index of the importance assigned to the scores relative to the mere presence or absence of items. The larger N is, the more the measure of stage 2 is compressed around 1/2, and the less variation it induces in the similarity measures. The latter is appropriate if it is felt that the numbers of items that are joint or different do not comprise much reliable information and, therefore, should not affect the MF much. This is typically the case if many items are conceivable. Then the presence or absence of a few items more does not provide much information, and it is appropriate that this information does not generate much variation.

The effect of N on the MF is that, as the second measure is compressed more around 1/2, the MF will be confined more to an interval only slightly bigger than the interval between 1/4 and 3/4. The information about similarity is then mostly determined by the metric part regarding the differences in scores. This influence of N is especially important if different MFs are to be aggregated, e.g. with one reflecting problems in the family and the other reflecting things going well in the family. As another example, if one measure of similarity (or rather its opposite) is used to measure changes induced by

family therapy over time, and another to measure changes induced by medicine over time, and if the number of items gives less information for family therapy than for medicine, then it is appropriate that medicine accounts for more variance in the data than family therapy.

4. Illustration of the numerical properties of the MF

We first consider a well-known example raised by James (1890). Suppose that we wish to investigate what impressions arise if we confront people with pictures. The first picture illustrates fire and the subject says that it is as luminous as can be. The second picture concerns a ball and the subject says that it is as round as can be. The third picture concerns the sun and the subject says that it is as luminous and as round as can be. Although only two items, luminosity and roundness, have been mentioned, several more could have been mentioned and there is considerable randomness in the number of items raised, which leads us to reduce the impact of frequency similarity. Hence, we take $N = 10$. (It can be seen that the choice of N happens to be immaterial for the particular observations assumed here.) Table 3 gives the item scores.

Table 3. Reactions to pictures

	Fire	Ball	Sun
Luminous	1	0	1
Round	0	1	1

Table 4 gives MF calculations. The score similarity between fire and ball is, obviously, minimal, and is 0. Their frequency similarity is the neutral value 0.5. The subject has raised the same number of items which in itself suggests a similar interest in both, thus enhancing the frequency similarity. However, no common items were raised which reduces the frequency similarity back to neutrality. Thus, the overall similarity is the average of 0 and 0.5, which is 0.25.

Table 4. Similarity calculations

	Fire and ball	Fire and sun	Ball and sun
Score similarity	0	0.5	0.5
Frequency similarity	0.5	0.5	0.5
MF	0.25	0.5	0.5

The score similarity for fire and sun is the neutral 0.5, with maximal similarity on one item and minimal on the other. The frequency similarity is also 0.5. The number of joint items raised is 1, yielding a term $\sqrt{1/9} = 1/3$ in the numerator in (4). The sun has raised an item not raised by fire, yielding a negative term $-|\sqrt{1/9} - 0|$ in the numerator, and the end result is the neutral 0.5. The overall similarity that results is 0.5. Likewise, this is the similarity between sun and ball.

As a first pragmatic approximation of similarity between the reactions of the subject, the results of the MF are plausible. They have combined the overlap of the items raised

and the intensity appearing from the number of items raised. If we take (1 − MF) as a distance measure, then ball and fire are each moderately remote from the sun with distance 0.5, but they are more remote from each other, with distance 0.75. These distances do not violate the triangle inequality, a topic central to James (1890) and other studies.

For another example, suppose that we observe the scores given in Table 5. Table 6 gives the results of the MF for all pairs of persons, where $N = 8$ was taken. This N is the minimal value that can be taken for these data, and it assigns maximal significance to the frequency similarity index. It suggests that we are confident that there is not much randomness in whether more new items may come up or not, as is plausible if we had interviewed many persons and observed so. We now discuss some of the results.

Table 5. Hypothetical example

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
A	7	7	0	0	7	0	3	3	3	3
B	7	7	0	0	7	0	3	3	3	3
C	7	7	0	0	7	0	3	3	3	3
D	7	7	0	0	7	0	3	3	3	3
E	7	7	0	0	0	7	5	4	1	0
F	7	7	0	0	0	7	5	4	1	0
G	7	7	0	0	0	7	5	4	1	0
H	7	7	0	0	0	7	5	4	1	0

Note. Ten persons $P_1, \ldots, P_{10}$ and their scores on eight items A, . . . , H, where 0 indicates the absence of an item, 1 minimal seriousness, and 7 maximal seriousness.

P_1 and P_2 gave much information, on eight items, and all of their information is identical. Therefore, it stands to reason that their similarity is maximal, which indeed it is, being 1. P_1 and P_3 are maximally different, and have minimal similarity, being 0, as is to be expected. P_3 and P_4 are also identical, but have given minimal information. One explanation may be that there are no problems at all and that there is perfect agreement between, say, the father and the mother, which would call for a maximal similarity of 1 rather than the MF value 0.75. However, an alternative explanation may be that the father and mother did not give information, and in reality may still be dissimilar. For refined measurements in sophisticated investigations, further inquiries would be desirable. The MF is, however, meant to be pragmatic and easily applicable, and to give an optimal estimate with whatever evidence we have available. In the given situation, we do not have the evidence to support maximal similarity as for P_1 and P_2. In general, the MF is more prudent and closer to neutrality the less evidence we have. The lack of evidence for P_3 and P_4 is captured through the frequency similarity, which is the neutral 0.50, yielding an overall similarity of 0.75. Thus, whereas the absence of evidence is maximally favourable as regards the seriousness of the situation (Section 2), the evidence then is too weak and has insufficient refinement to generate a maximal similarity score. Similarity is a derived concept that in itself is favourable. It cannot, however, replace the primary signals of goodness. Similarity is an additional index focusing on beliefs about the situation.

Table 6. Similarity weights of the persons in Table 5 if $N = 8$ is chosen.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
P_2	ss:1.00 fs:1.00								
	1.00								
P_3	ss:0.00 fs:0.00	ss:0.00 fs:0.00							
	0.00	**0.00**							
P_4	ss:0.00 fs:0.00	ss:0.00 fs:0.00	ss:1.00 fs:0.50						
	0.00	**0.00**	**0.75**						
P_5	ss:0.50 fs:0.50	ss:0.50 fs:0.50	ss:0.00 fs:0.15	ss:0.00 fs:0.15					
	0.50	**0.50**	**0.07**	**0.07**					
P_6	ss:0.50 fs:0.50	ss:0.50 fs:0.50	ss:0.00 fs:0.15	ss:0.00 fs:0.15	ss:0.00 fs:0.50				
	0.50	**0.50**	**0.07**	**0.07**	**0.25**				
P_7	ss:0.57 fs:1.00	ss:0.57 fs:1.00	ss:0.43 fs:0.00	ss:0.43 fs:0.00	ss:0.36 fs:0.50	ss:0.64 fs:0.50			
	0.79	**0.79**	**0.21**	**0.21**	**0.43**	**0.57**			
P_8	ss:0.50 fs:1.00	ss:0.50 fs:1.00	ss:0.50 fs:0.00	ss:0.50 fs:0.00	ss:0.43 fs:0.50	ss:0.57 fs:0.50	ss:0.93 fs:1.00		
	0.75	**0.75**	**0.25**	**0.25**	**0.46**	**0.54**	**0.96**		
P_9	ss:0.29 fs:1.00	ss:0.29 fs:1.00	ss:0.71 fs:0.00	ss:0.71 fs:0.00	ss:0.64 fs:0.50	ss:0.36 fs:0.50	ss:0.71 fs:1.00	ss:0.79 fs:1.00	
	0.64	**0.64**	**0.36**	**0.36**	**0.57**	**0.43**	**0.86**	**0.89**	
P_{10}	ss:0.21 fs:0.50	ss:0.21 fs:0.50	ss:0.57 fs:0.15	ss:0.57 fs:0.15	ss:0.43 fs:0.85	ss:0.29 fs:0.50	ss:0.64 fs:0.50	ss:0.71 fs:0.50	ss:0.93 fs:0.50
	0.36	**0.36**	**0.36**	**0.36**	**0.64**	**0.39**	**0.57**	**0.61**	**0.71**

Note. ss denotes the score similarity, fs the frequency similarity, and the MF is printed in bold face.

P_5 and P_6 may seem to be minimally similar, and their score similarity is indeed minimal. The number of items that are jointly raised by both is zero, again suggesting maximal dissimilarity. There is, however, a similarity in the sense that both raise the same number of items. This in itself suggests that the seriousness of the situation is similar in the perception of these two persons. Their frequency similarity ends up neutral, at 0.5, and the overall similarity is small, 0.25, but not minimal.

If only the score similarity is inspected, then P_7 and P_8 may seem to be equally similar as P_9 and P_{10}. The frequency similarity gives a different picture, though. P_7 and P_8 have identical numbers of items raised, and maximal frequency similarity. The frequency similarity of P_9 and P_{10} is much smaller, only the neutral 0.5, because P_{10} raised many fewer items than P_9. These results underscore the special nature of the information about the presence or absence of items. Intuitively, it stands to reason that P_9 and P_{10} are less similar than P_7 and P_8, because the difference in numbers of items constitutes a larger discrepancy.

Table 7 gives the results if $N = 20$ is chosen. This choice entails that we have less confidence in the reliability of items showing up or not, and that we feel that there can be many more items unspecified here. We would be forced to choose $N = 20$ or more if there were another person in the study who raised 20 items. This would signal to us that there are far more items and that the frequency information is not very reliable. As before, the less evidence we have, the closer MF is to the neutral 0.5. The effect of this enlarged N is, indeed, that the frequency similarities are compressed more around 0.5, implying that the overall similarities are also closer to their neutral values 0.5. Now, with $N = 20$, the maximum similarity is only 0.91, and the minimum similarity is 0.09, so that in no case do we have enough information to justify very extreme judgements.

5. Illustration of MF in the family-personal questionnaire

This section presents an application of the MF to the comparison of family members' target complaint answers before and after a family therapy treatment. Whereas traditional methods for comparing agreements in judgements between different persons have used a fixed number of predetermined variables (Légaré *et al.*, 2006), we allow for complete flexibility a priori regarding the variables that will be measured.

5.1. Participants

The original data set included 36 families ($N = 116$) under family therapy in a psychiatric day care unit (Spain), who participated in a naturalistic research project on psychotherapeutic change processes. We illustrate the use of the MF on data for one family, represented by the patient (Zarastro,[3] a 31-year-old Spanish Caucasian male, who had met criteria for paranoid schizophrenia, DSM-IV F20.02, since the age of 16), his mother and two brothers (Amadeus and José).

[3] Names and identifying variables have been modified so as to ensure patients' anonymity.

Table 7. Similarity weights of the persons in Table 6 for $N = 20$

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9
P_2	ss:1.00 fs:0.82								
	0.91								
P_3	ss:0.00 fs:0.18	ss:0.00 fs:0.18							
	0.09	**0.09**							
P_4	ss:0.00 fs:0.18	ss:0.00 fs:0.18	ss:1.00 fs:0.50						
	0.09	**0.09**	**0.75**						
P_5	ss:0.50 fs:0.50	ss:0.50 fs:0.50	ss:0.00 fs:0.28	ss:0.00 fs:0.28					
	0.50	**0.50**	**0.14**	**0.14**					
P_6	ss:0.50 fs:0.50	ss:0.50 fs:0.50	ss:0.00 fs:0.28	ss:0.00 fs:0.28	ss:0.00 fs:0.50				
	0.50	**0.50**	**0.14**	**0.14**	**0.25**				
P_7	ss:0.57 fs:0.82	ss:0.57 fs:0.82	ss:0.43 fs:0.18	ss:0.43 fs:0.18	ss:0.36 fs:0.50	ss:0.64 fs:0.50			
	0.69	**0.69**	**0.31**	**0.31**	**0.43**	**0.57**			
P_8	ss:0.50 fs:0.82	ss:0.50 fs:0.82	ss:0.50 fs:0.18	ss:0.50 fs:0.18	ss:0.43 fs:0.50	ss:0.57 fs:0.50	ss:0.93 fs:0.82		
	0.66	**0.66**	**0.34**	**0.34**	**0.46**	**0.54**	**0.87**		
P_9	ss:0.29 fs:0.82	ss:0.29 fs:0.82	ss:0.71 fs:0.18	ss:0.71 fs:0.18	ss:0.64 fs:0.50	ss:0.36 fs:0.50	ss:0.71 fs:0.82	ss:0.79 fs:0.82	
	0.55	**0.55**	**0.45**	**0.45**	**0.57**	**0.43**	**0.77**	**0.80**	
P_{10}	ss:0.21 fs:0.50	ss:0.21 fs:0.50	ss:0.57 fs:0.28	ss:0.57 fs:0.28	ss:0.43 fs:0.72	ss:0.29 fs:0.50	ss:0.64 fs:0.50	ss:0.71 fs:0.50	ss:0.93 fs:0.50
	0.36	**0.36**	**0.42**	**0.42**	**0.58**	**0.39**	**0.57**	**0.61**	**0.71**

Note. ss denotes the score similarity, fs the frequency similarity, and the MF is printed in bold face.

5.2. Measures

The Simplified Personal Questionnaire (PQ; Elliott, Shapiro, & Mack, 1999) is a modified version of the Personal Questionnaire proposed by Shapiro (1961a, 1961b) for outcome measurement of psychiatric patients. It is a target complaint measure consisting of the problems for which the patient is seeking treatment. The questionnaire is constructed in a semi-structured 45-minute interview conducted in order to help patients to state their main problems. Patients' statements were placed on individual note cards, rank-ordered and typed on to a standard form. Patients were encouraged to raise all the complaints they wanted and were instructed to 'rate each of the following problems according to how much it has bothered you during the past seven days, including today', using a 7-point scale (from 1 = not at all, to 7 = maximum possible). The PQ form was administered immediately before each session to each member of the family. Over the course of treatment, participants were allowed to change the PQ form by entering new complaints or deleting previous complaints if they wished.

5.3. Procedure

Prior to the similarity analysis, it was necessary to categorize the complaints into types of problems, because different complaints would refer to the same problem. For instance, consider: *For 16 years he did not relate to the outside world, and broke up with lifetime friends* (mother); *It's hard to maintain my few friendships* (Zarastro); *My brother Zarastro has been losing his friends* (José). These complaints report the same problem concerning the patient's impairment in social skills.

The categorization system was established using the PQ complaints of all 116 participants. In a first step, three independent judges read PQ forms guided by the question 'What are the problems this person identifies in his/her family?'. Each judge proposed a list of mutually exclusive categories that (1) were descriptive of the families' problems, and (2) captured each member's point of view. Discrepancies were discussed until consensus resulted in a categorization system with a total of 74 categories that had been raised by any of the 116 participants. Using this categorization system, the PQ complaints of Zarasto's family were independently coded by the same three judges, and discrepancies were discussed again until consensus was reached. We analyse similarity at the level of categories, taking categories as items.

We took $N = 80$ as the normalizing factor (see Section 3.4). This number safely exceeds the total number of items conceivable, and considerably exceeds the numbers of items raised by each individual participant. We could have chosen smaller values for N depending on families, but preferred this choice $N = 80$ because we felt there was considerable randomness in the number of items actually raised versus the number that could have been raised.

The similarity calculation was based on the categories of problems and their intensity, obtained pre-treatment and at the termination of treatment. If a person raised the same category more than once, then we took the highest, rather than the average, seriousness score because the multiple appearances suggest extra seriousness. For instance, prior to the treatment Amadeus raised category O twice, and we used his highest score 4.

The similarity within pairs of family members was calculated using a routine in Pascal. The program and data can be downloaded from the second author's homepage.

5.4. Results and discussion

The family reported a total of 36 complaints at the initial administration of the PQ (mother, 10 complaints; Zarastro, 10; José, 11; Amadeus, 5). Over the course of the treatment, the mother added 7 more complaints and José added 2. Thus, at discharge, there was a total of 45 complaints, corresponding to 18 problem categories (Table 8).

The results (Table 9) show that therapy increased similarity of judgement of the problems. Before the treatment, the highest agreement was between Zarastro and his mother. At discharge, Zarastro, José and the mother mostly agree with each other. Also the *increase* in agreement resulting from therapy is larger for Zarastro with his two brothers, and for José and the mother. This was enhanced by the extra items José and the mother raised during the therapy, which primarily increased the frequency similarity.

PQ similarity given by our formula provides a relational picture of the family complaints and their change resulting from therapy, which can be related to clinical information about the family. Clinical records reveal that family relationships had been characterized by long-standing generalized feelings of indifference and interpersonal avoidance, with no expression of affection. According to Zarastro, at the intake interview, his parents had 'always been indifferent'. His relationship with José was 'nowadays, bad and superficial. He blames me for doing nothing, or when I argue with someone else in the family.' Zarastro also described his relationship with Amadeus as more distant than his relation with José: 'It has been like this for as long as I can remember, we almost don't pay attention to each other.'

At the time of the intake interview, Zarastro lived confined to his home, mostly in his bedroom, avoiding any interaction with the rest of the family. He felt excluded, and he thought that the only reason his family members would approach him was to criticize his behaviour. He also blamed the family for his condition. On the other hand, out of fear of Zarastro's violent reactions, the family avoided talking with him about his condition, which in turn was felt by him as abandonment, thus increasing his isolation.

The low level of similarity between family members at the pre-treatment phase found in our results seems to reflect this family's relational pattern of emotional distance. In sum, there is isomorphism between PQ similarity patterns and the family's initial relational structure, suggesting that similarity quantifies the relational proximity within family members, in a diagnostic phase.

There is also evidence of isomorphism between PQ similarity changes and changes resulting from the therapy as reported by the family. During the course of the treatment, the mother reported that the family was more relaxed and trusting, and linked this positive change to the moments in the sessions where she could talk with Zarastro with no violent reactions. Zarastro reported that he felt that the family were giving him more help, and José noticed more expressions of affection in the family. Amadeus also reported positive changes in Zarastro and in the family mood. In sum, the treatment showed the family that there was no reason to be afraid, and showed Zarastro that the family was interested in his well-being. Family therapy has brought Zarastro closer to his family, which is captured by a general increase in the similarity of all members and Zarastro.

Nevertheless, Amadeus assumed a distant position during the sessions, rarely speaking, and failing to attend three therapy sessions. Such relational distance is also reflected in the PQ similarity, and can be observed in the similarity tree (Table 9).

Table 8. PQ complaints, categories, and pre/post-treatment (nonnormalized) scores

	PQ complaints	Category	Pre-scores	Post-scores
Mother	There were death threats to the mother	A	7	1
	I had to leave the house many times	A	7	1
	Family was very upset because of the constant threats to me	A	7	1
	For 16 years he did not relate to the outside world, and broke up with lifetime friends	B	4	4
	In the family we had continuous fights and arguments	C	4	1
	We called the police 5 times, in 2 years	A	3	1
	We went to a series of private psychiatrists	D	2	1
	He did very badly at school	E	4	2
	We had only short quiet times, but then went back to the threats, creating distress	A	6	1
	He was very strange and aggressive	A	4	1
	See him sad	R		2
	No future (for him)	E		2
	Sometimes he goes out and comes back, very tense	B		4.5
	He believes people watch him	F		2
	He has no energy, and he comments he feels tired	P		1
	He complains of tiredness. I fear he may think it is because of the pills and he may decide to stop taking them	P		4
Zarastro	I'm very shy	B	7	4
	I don't know how to keep a conversation going	B	6	4
	It's hard to maintain my few friendships	B	6	4
	It bothers me to have eye-contact with people on the street	B	5	3
	I feel that people are watching me	F	4	2
	I'm not able to show my dislike or distress	B	6	2
	I'm worried about not having a job or schooling/studies/education	E	6	3
	I have a cold relationship with my younger brother	G	3	2
	It's hard to talk and show my complaints at home	B	2	2
	I'm obsessed with the past	H	6	1
José	My brother Zarastro has a distorted perception of the relationships at home	F	7	1
	My brother Zarastro has difficulties relating to others	B	1	4
	My brother Zarastro has been losing his friends	B	1	3
	Lack of emotional communication	I	4	2
	My brother Amadeus is often impolite towards Zarastro	G	6	5
	My brother Amadeus is very cold and reserved, doesn't interact	G	5	5.5

Table 8. (*Continued*)

	PQ complaints	Category	Pre-scores	Post-scores
	My brother Amadeus has lost interest in Zarastro's problem	G	6	3.5
	My sister has severe depression	J	2	1
	It affects my job because it decreases my attention	K	5	1.5
	It affects my relationships with others	L	5	1.5
	I feel down when I have to take care of Zarastro	M	7	1.5
	My brother Zarastro doesn't relate well with people when I invite him to go out with my friends	B		4
	Zarastro does very little to improve his employment situation	E		3
Amadeus	I feel anxious sitting at the table between Zarastro and my mother	N	6	2
	My mother makes her children depend on her	O	4	2
	My mother wants her cubs surrounding her	O	3	2
	Zarastro watches TV too much	P	4	1
	We (extended family) share the same fate	Q	5	1

Note. A, patient's violent behaviour; B, patient's social skills impairment; C, conflict within the family; D, failed previous attempts to solve the problems; E, patient's educational/professional impairment; F, delirious ideas; G, emotional/affective distance between brothers; H, patient's feelings of guilt; I, generalized emotional/affective distance within the family; J, family members' diagnosed psychiatric disorders (other than the patient's); K, negative impact on family members' educational/professional life (other than the patient's); L, negative impact on family members' social life (other than the patient's); M, family members' depressed mood (other than the patient's); N, family members' anxiety (other than the patient's); O, mother–son emotional mixup; P, patient's dysthymia; Q, relational problems with the extended family; R, patient's depressed mood.

Finally, PQ similarity change was consistent with the therapist's strategies and tasks. The therapist reported a general strategy of active involvement of José in Zarastro's recovery process, in order to balance the proximity between Zarastro and the mother, who was a source of stress for the patient. José was indeed the most mature and emotionally stable person in the family, who could mediate the mother–patient relationship, and could support Zarastro's socialization endeavours. The results on PQ similarity at discharge indicate a higher proximity between Zarastro and José. Thus, PQ similarity reflects in-session activity, confirming that this measure is useful for monitoring the impact of therapeutic interventions in the family relations.

Overall, the results show that MF is useful for idiographic measure analysis in clinical contexts. PQ similarity gives the therapist a quick answer to clinically central questions, such as which members of the family share similar perceptions of the problems. Or, what impact do therapeutic interventions have in the proximity–distance relational pattern of the family members? PQ similarity patterns quantify the relational and emotional structure (proximity–distance) between family members, which is useful as a relational diagnostic measure for therapists.

Table 9. Similarity scores and graphical representation, for pre- and post-family PQ, for $N = 80$.

Pre-Treatment				
	Zarastro	Mother	José	
Mother	w:0.45 f: 0.58 **0.51**			stress formula 1 =0.0328; stress formula 2 = 0.1054; r(monotonic) squared=0.9889; r-squared (p.v.a.f.)=0.9319)
José	w:0.33 f: 0.55 **0.44**	w: 0.33 f: 0.52 **0.43**		ZARASTRO MOTHER JOSÉ
Amadeus	w:0.29 f: 0.49 **0.39**	w: 0.37 f: 0.49 **0.43**	w: 0.33 f: 0.45 **0.39**	AMADEUS
Post-Treatment				
	Zarastro	Mother	José	
Mother	w:0.81 f: 0.55 **0.68** **(0.17)**			stress formula 1 = 0.0000; stress formula 2 = 0.0000; r(monotonic) squared=1.0000;r-squared (p.v.a.f.)=0.9944)
José	w:0.81 f: 0.54 **0.68** **(0.24)**	w: 0.75 f: 0.58 **0.67** **(0.24)**		ZARASTRO JOSÉ MOTHER
Amadeus	w:0.71 f: 0.49 **0.60** **(0.21)**	w: 0.72 f: 0.50 **0.61** **(0.18)**	w: 0.70 f: 0.44 **0.57** **(0.18)**	AMADEUS

Note. w: the score-similarity; f: the frequency similarity; printed in bold: the overall similarity; within parentheses is the pre-post change, given by the difference in overall similarity of those two times. Overall similarity is graphically represented using ADDTREE/P version 1.5 – Program for Fitting Additive Trees (Corter 1982; Sattath & Tversky, 1977).

The flexibility of the MF was instrumental in the application described in this section, where it could hardly be predicted what items would be raised. The presence or absence of items played a special role, as explained in Section 3.2. The absence of an item for a particular person is in itself favourable, signalling the probable absence of the corresponding problem for this person. It may, however, decrease the similarity with a person who did signal the problem, and such a dissimilarity in belief can in itself be a drawback.

6. Conclusion

Measures of similarity are important in many domains. We have introduced the metric-frequency measure, a measure that is more widely applicable than earlier ones, being pragmatic and easily applicable to data sets with little structure. In particular, it need not be anticipated which variables will be observed, or how many variables, and they may be metric or qualitative.

References

Ashworth, M., Robinson, I.S., Godfrey, E., Parmentier, H., Shepherd, M., Christey, J., *et al.* (2005). The experiences of therapists using a new client-centered psychometric instrument, PSYCHLOPS (Psychological Outcome Profiles). *Counseling and Psychotherapy Research*, *5*, 37–42.

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology*, *63*, 516–556.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, *44*, 568–591.

Barthélemy, J.-P., & Mullet, E. (1996). Information processing in similarity judgements. *British Journal of Mathematical and Statistical Psychology*, *49*, 225–240.

Battle, C., Imber, S. D., Hoehn-Saric, R., Stone, A. R., Nash, C., & Frank, J. D. (1966). Target complaints as criteria of improvement. *American Journal of Psychotherapy*, *20*, 184–192.

Carroll, J. D. (1976). Spatial, nonspatial and hybrid models for scaling. *Psychometrika*, *41*, 439–463.

Corter, J. E. (1982). ADDTREE/P: A PASCAL program for fitting additive trees based on Sattath and Tversky's ADDTREE algorithm. *Behavior Research Methods and Instrumentation*, *14*, 353–354.

Davison, M. L. (1992). *Multidimensional scaling*. Malabar, FL: Krieger.

Dzhafarov, E. N., & Colonius, H. (2007). Dissimilarity cumulation theory and subjective metrics. *Journal of Mathematical Psychology*, *51*, 290–304.

Elliott, R., Shapiro, D. A., & Mack, C. (1999). *Simplified personal questionnaire procedure manual*. Toledo, OH: University of Toledo, Department of Psychology.

Elliott, R., Slatick, E., & Urman, M. (2001). Qualitative change process research on psychotherapy: Alternative strategies. In J. Frommer & D. L. Rennie (Eds.), *Qualitative psychotherapy research: Methods and methodology* (pp. 69–111). Lengerich: Pabst Science Publishers.

Evans, C. J., Hughes, J., & Houston, J. (2002). Significance-testing the validity of ideographic methods: A little derangement goes a long way. *British Journal of Mathematical and Statistical Psychology*, *55*, 385–390.

Gilboa, I., & Schmeidler, D. (2001). *A theory of case-based decisions*. Cambridge: Cambridge University Press.

Greenberg, L. S. (1986). Change process research. *Journal of Consulting and Clinical Psychology*, *54*, 4–9.

Greenberg, L. S., & Pinsof, W. M. (1986). *The psychotherapeutic process: A research handbook*. New York: Guilford Press.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*, 1–32.

Hill, C., & Lambert, M. J. (2004). Methodological issues in studying psychotherapy processes and outcomes. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (pp. 84–135). New York: Wiley.

James, W. (1890). *Principles of psychology*. New York: Holt.

Johannesson, M. (2000). Modelling asymmetric similarity with prominence. *British Journal of Mathematical and Statistical Psychology*, *53*, 121–139.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720.

Kiresuk, T. J., & Sherman, R. E. (1968). Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. *Community Mental Health Journal*, *4*, 443–453.

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling models for cognitive modeling. *Journal of Mathematical Psychology*, *45*, 149–166.

Légaré, F., O'Connor, A. M., Graham, I. D., Wells, G. A., & Tremblay, S. (2006). Impact of the Ottawa decision support framework on the agreement and the difference between patients' and physicians' decisional conflict. *Medical Decision Making*, *26*, 373–390.

Levine, F. J., & Luborsky, L. (1981). The core confictual relationship theme method: A demonstration of reliable clinical inferences by the method of mismatched cases. In S. Tuttmen, C. Kaye, & M. Zimmerman (Eds.), *Object and self: A developmental approach*. New York: International Universities Press.

Lord, P. W., Stevens, R. D., Brass, A., & Goble, C. A. (2003). Investigating semantic similarity measures across the gene ontology: The relationship between sequence and annotation. *Bioinformatics*, *19*, 1275–1283.

Navarro, D. J. (2007). On the interaction between exemplar-based concepts and a response scaling process. *Journal of Mathematical Psychology*, *51*, 85–98.

Navarro, D. J., & Lee, M. D. (2003). Combining dimensions and features in similarity-based representations. *Advances in neural information processing systems*, 15, 59–66.

Resnik, O. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity and natural language. *Journal of Artificial Intelligence Research*, *11*, 95–130.

Saito, T. (1994). Psychological scaling of the asymmetry observed in comparative judgement. *British Journal of Mathematical and Statistical Psychology*, *47*, 41–62.

Sales, C., Gonçalves, S., Fragoeiro, A., Noronha, S., & Elliott, R. (2007). Psychotherapists' openness to routine naturalistic idiographic research? *Mental Health and Learning Disabilities Research and Practice*, *4*, 145–161.

Sales, C., Gonçalves, S., Noronha, S., Fragoeiro, A., Rivero, C., Elliott, R., *et al.* (2006). *Research-practice integration in naturalistic clinical settings: Is it possible? Is it useful?* Paper presented at the 37th Annual Meeting of the Society for Psychotherapy Research, Edinburgh. Scotland.

Santini, S., & Jain, R. (1999). Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*, 871–883.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, *42*, 319–345.

Shapiro, M. B. (1961a). A method of measuring changes specific to the individual psychiatric patient. *British Journal of Medical Psychology*, *34*, 151–155.

Shapiro, M. B. (1961b). *The personal questionnaire. Abbreviated manual*. Unpublished manuscript.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, *27*, 125–140.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Tversky, A., & Gati, I. (1982). Similarity, separability and the triangle inequality. *Psychological Review*, *89*, 123–154.

Wagner, J., & Elliott, R. (1999). *The simplified personal questionnaire*. Unpublished manuscript, University of Toledo.

Xiao, J., & Zhang, Y. (2001). Clustering of web users using session-based similarity measures. In *Proceedings of International Conference on Computer Networks and Mobile Computing* (pp. 223–228).

Received 2 October 2007; revised version received 15 September 2008