Measuring and testing the agreement of matrices

Michael J. Brusco¹ · Douglas Steinley²

Published online: 7 December 2017
© Psychonomic Society, Inc. 2017

Abstract The problem of comparing the agreement of two $n \times n$ matrices has a variety of applications in experimental psychology. A well-known index of agreement is based on the sum of the element-wise products of the matrices. Although less familiar to many researchers, measures of agreement based on within-row and/or within-column gradients can also be useful. We provide a suite of MATLAB programs for computing agreement indices and performing matrix permutation tests of those indices. Programs for computing exact p -values are available for small matrices, whereas resampling programs for approximate p -values are provided for larger matrices.

Keywords Matrix agreement · Matrix permutation · Combinatorial data analysis

Introduction

Measurement and testing of the agreement (or concordance) of two $n \times n$ proximity matrices is commonly accomplished using the quadratic assignment paradigm (Hubert, 1987; Hubert & Schultz, 1976). When using the quadratic assignment model, an index of agreement between two matrices is

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-017-0990-7>) contains supplementary material, which is available to authorized users.

✉ Michael J. Brusco
mbrusco@business.fsu.edu

¹ Department of Business Analytics, Information Systems, and Supply Chain, College of Business, Florida State University, 821 Academic Way, Tallahassee, FL 32306-1110, USA

² University of Missouri, Columbia, USA

typically established using either the sum of products of corresponding matrix elements (Mantel, 1967), or within-row (or within-column) gradient indices that measure the internal structure or patterning of the elements (Hubert, 1978, 1987). A one-tailed significance test of these indices can be performed by generating a distribution for the index values via a permutation process. The number of permutations producing an index value equal to or more extreme than the observed index value is then divided by the total number of permutations evaluated to obtain a p -value.

Hubert and Schultz (1976, pp. 202–209) indicated that the quadratic assignment model is closely tied to a number of classic paradigms in statistical data analysis, including bivariate association, two-independent sample problems, two dependent sample problems, k -independent samples, graph intersection, and run statistics. More general applications of the quadratic assignment model in experimental psychology commonly take one of two forms. First, a researcher might have two proximity matrices measured on the same set of objects at two (or more) different points in time, such as the measure of social ties among students at different weeks of a semester. Second, two or more proximity matrices might be available for the same set of objects under different experimental conditions (e.g., different distances, different lighting, different noise levels, etc.).

As an example of the quadratic paradigm in experimental psychology, Hubert (1987) reported a study of two auditory confusion matrices based on data originally collected by Miller and Nicely (1955). Likewise, Brusco (2004) measured the concordance among 19 visual and tactual interletter confusion matrices from the perception literature. There are also some early examples of the quadratic assignment model in the areas of counseling and educational psychology. Holloway (1982) and Harris and Packard (1985) used the quadratic assignment model within the context of counsellor training and supervision, and Gliner (1981) and Medina-Diaz

(1993) report applications pertaining to cognitive structure. More recently, quadratic assignment has been successfully implemented in a variety of social network contexts, including: (i) the study of affiliation relationships in the animal kingdom (Bonnie & de Waal, 2006; Hirsch, Prange, Hauver, & Gehrt, 2013; Mitani, Merriwether, & Zhang, 2000), (ii) the investigation of corporate policy and political action (Burris, 2005; Dreiling & Darves, 2011), (iii) the interaction among students or teachers (Kwon & Lease, 2014; Molenaar, Slegers, Karsten, & Daly, 2012), and (iv) creative relationships among employees (Fliaster & Schloderer, 2010).

When considering the agreement between two matrices, there are at least two relevant issues to consider. The first of these is: How should agreement be measured? Perhaps the most common index is the sum of the products of the corresponding elements of the two matrices (Mantel, 1967). This index, known as the Mantel index, underlies the linear association (i.e., the correlation) between the matrices. Hubert (1978, 1987) recognized that a linear association index based on one-to-one products of the elements of two matrices was a legitimate but somewhat narrow perspective for establishing a measure of matrix agreement. Accordingly, as a supplement for the Mantel index, Hubert (1978, 1987) developed a variety of alternative indices for matrix agreement that measure relationships of internal structure and patterning of the two matrices.

One of the alternative indices proposed by Hubert (1987) is based on within-row gradients. This measure counts the number of consistencies and inconsistencies among triads of objects, where the first object (i) of the triad corresponds to a row and the other two objects (j and k) correspond to columns. If the element in row i and column j exceeds the element in row i and column k for both matrices, then this consistency contributes positively to matrix concordance. However, if the element in row i and column j exceeds the element in row i and column k for the first matrix, but the element in row i and column j is less than the element in row i and column k for the second matrix, then this inconsistency contributes negatively to matrix agreement.

Brusco (2004) constructed two small examples to show the potential conflict that can arise between the Mantel and within-row gradient indices. The first example showed high correlation between two matrices that exhibited little agreement with respect to within-row patterning. The second example showed low correlation between two matrices, yet perfect concordance between the within-row patterning of the matrices. Of course, there may be many other circumstances where the Mantel and within-row gradient indices will yield very comparable results. Accordingly, one of our goals in this paper is to explore these issues further using several examples.

A second issue pertains to tests of significance within the context of matrix agreement, as well as the availability of software for completing such tests (Glerean et al, 2016). In some instances, the computation of an exact p -value is

possible via the generation of the entire distribution of all possible values of the index that can be realized. This is accomplished by holding one of the two matrices fixed and obtaining the index value for all $n!$ permutations of the rows and columns for the second matrix. The computation of an exact p -value is ideal when computationally feasible. A precise limit on the value of n for which an exact p -value is feasible is difficult to establish because it depends on the hardware platform, software program, and how long the analyst is willing to wait to obtain the p -value. As a general guideline, once the value of n is approximately 12 or greater, the computational burden of enumerating all permutations tends to lessen the likelihood that an exact p -value will be sought. Therefore, in addition to a complete enumeration program, it is useful to have a program that can approximate a p -value based on a large number of sampled permutations. Accordingly, another goal of our paper is to provide MATLAB software programs for computing exact and approximate p -values for the Mantel and within-row gradient indices.

The next section of this paper provides formal descriptions of the agreement indices and the software programs. This is followed by three examples that are used to demonstrate the software programs. The paper concludes with a brief summary.

Agreement indices

Data requirements

Prior to the presentation of the specific indices for measuring agreement, we provide some preliminary definitions pertaining to the data. We begin with the definition of $S = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_Q\}$ as a set of $n \times n$ proximity matrices. The proximity matrices, $\mathbf{A}_q = [a_{ijq}]$, can take many different forms, for example: (i) confusion matrices for the same set of stimuli obtained by different authors, (ii) similarity judgments for the same set of brands measured at different times, (iii) social network ties among the same set of actors for different relations (friendship, advice-seeking, collaboration, etc.), and (iv) social network ties among the same set of actors for the same relation but at different points in time. In these applications, the main diagonal of the matrices is typically irrelevant and is, therefore, ignored in the analysis. For example, the main diagonal of a confusion matrix consists of correct recognitions of the stimuli, not confusion between different stimuli. In a similarity matrix, the judgment of the similarity of brand or product to itself is nonsensical, and self-ties are commonly ignored in the social network context. Throughout the remainder of this paper, we set $a_{iiq} = 0$ for all $1 \leq i \leq n$ and $1 \leq q \leq Q$, and assume that the main diagonal is ignored in the analysis. The zeroing out of the main diagonal is also consistent with

the assumptions of Hubert and Schultz (1976, p. 192) in their description of the quadratic assignment paradigm and derivation of the mean and variance of the Mantel statistic.

Mantel statistic

Perhaps the most common use of the quadratic assignment model for matrix agreement and permutation testing is based on the work of Mantel (1967). Mantel's index is the sum of the element-wise products of the corresponding elements of two $n \times n$ matrices. Given two $n \times n$ matrices from S , A_q , and A_r , the statistic is calculated as:

$$\Gamma_1(A_q, A_r) = \sum_{i=1}^n \sum_{j=1}^n a_{ijq} a_{ijr}. \quad (1)$$

To perform an exact test of the Mantel index at some prespecified level of significance, a reference distribution can be obtained by holding one of the matrices (A_q) fixed and recomputing the Mantel index for each of the $n!$ permutations of the second matrix (A_r). The set of all possible $n!$ permutations is denoted Ψ , and a specific permutation from Ψ is indicated by ψ . The notation $\psi(i)$ is used to refer to the object that is assigned to position i of the permutation. One of the permutations in Ψ is the identity permutation, $\psi_1 : \{\psi(1) = 1, \psi(2) = 2, \dots, \psi(n-1) = n-1, \psi(n) = n\}$. Equation (1) is the computation for ψ_1 , but this equation can now be translated to apply to any and all $\psi \in \Psi$, thus providing the statistical distribution of indices accordingly:

$$\Gamma_1(A_q, A_r(\psi)) = \sum_{i=1}^n \sum_{j=1}^n a_{ijq} a_{\psi(i)\psi(j)r} \quad \forall \psi \in \Psi. \quad (2)$$

A key aspect to note regarding Equation (2) is that the permutation, ψ , is applied to both the rows and columns of A_r . Also, as noted previously, given that the main diagonals of the matrix are assumed to be zero, the main diagonal makes no contribution to either the observed statistic or the index values for the reference distribution. Alternatively, the second summation in Equation (2) could be taken over $j \neq i$; however, when using commands in our Matlab programs, it is more economical to retain the diagonals but just assure that they are zero.

The one-tail p -value associated with such a test is the ratio of the number of permutations bearing indices as extreme or more extreme than the observed index $\Gamma_1(A_q, A_r(\psi_1))$, to the total number of permutations. Without loss of generality, we are concerned with counting the number of permutations bearing statistics greater than or equal to the observed index. Under these circumstances, the p -value is the percentage of permutations with indices at or to the right of the observed index in the distribution. The key underlying assumption associated with the computation of the p -value is that all permutations, $\psi \in \Psi$, are equally likely (i.e., an equally likely

labelling assumption for the objects). This assumption would be most improper if, for example, an algorithm had been applied to permute one of the matrices so as to optimize the observed Mantel statistic.

The computational feasibility of obtaining an exact p -value depends on n and, to some extent, on the hardware and software platforms. A Fortran or C based implementation on a fast computing machine might enable an exact p -value to be obtained for roughly $n = 13$ or 14 , whereas an implementation in MATLAB or R might be limited to $n = 11$ or $n = 12$ on a fast computing machine. For $n \geq 15$, a resampling procedure will be necessary to obtain an approximate p -value. This is accomplished by taking a random sample from Ψ to produce a reference distribution.

Two MATLAB m-files are available for the Mantel agreement index. The first program, `mantelexact.m` (see Appendix A) computes the Mantel statistic and obtains an exact p -value. This program is restricted to roughly $n = 12$. The second program, `mantelapprox.m` (see Appendix B) computes the Mantel statistic and obtains an approximate p -value based on 100,000 random samples. This program is scalable for matrices where n is well into the hundreds. For much larger matrices ($n > 1,000$), it might be necessary to reduce the number of random samples to 10,000 or 1,000 so as to obtain results in reasonable time.

Within-row gradients

The within-row gradient index originally described by Hubert (1987), and later used by Brusco (2004), is based on triads of objects. For each of the matrices in S , a three-way array is obtained as follows based on the following relationships:

$$b_{ijkq} = \begin{cases} \text{sign}(a_{ikq} - a_{ijq}) & \text{for distinct } i, j, k \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

$$\text{where: } \text{sign}(x) = \begin{cases} +1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

For a given matrix q , the value of b_{ijkq} is 1 (-1) if the element in row i and column k is greater (less) than the element in row i and column j of matrix q . A value of $b_{ijkq} = 0$ occurs when the elements in columns j and k of row i are equal. Given two $n \times n$ matrices from S , A_q , and A_r , the within-row gradient index is calculated as:

$$\Gamma_2(A_q, A_r) = \frac{\sum_{i \neq j \neq k} b_{ijkq} b_{ijk r}}{\sum_{i \neq j \neq k} |b_{ijkq} b_{ijk r}|} \quad \forall 1 \leq q < r \leq Q. \quad (4)$$

For a given pair of matrices, A_q and A_r , the numerator of Equation (4) is the number of consistent within-row triads minus the number of inconsistent within-row triads. Several

options for the denominator were offered by Hubert (1987, pp. 274–276). The particular denominator used in Equation (4) is the sum of the number of within-row consistencies plus the number of within-row inconsistencies. This denominator was also used by Hubert (1987) in his example for Miller and Nicely's (1955) auditory confusion data, as well as by Brusco (2004) in his study of visual and tactual letter recognition matrices. The selected denominator has intuitive appeal because it allows for easy interpretation. A value of $\Gamma_2(\mathbf{A}_q, \mathbf{A}_r) = 0$ indicates that the number of consistencies in the same as the number of inconsistencies. If the number of consistencies is double, triple, or quadruple the number of inconsistencies, then $\Gamma_2(\mathbf{A}_q, \mathbf{A}_r) = .333$, $\Gamma_2(\mathbf{A}_q, \mathbf{A}_r) = .5$, or $\Gamma_2(\mathbf{A}_q, \mathbf{A}_r) = .6$, respectively.

Exact and approximate significance tests of the within-row gradient index for two matrices, $\mathbf{A}_q$ and $\mathbf{A}_r$, can be performed in the same manner as those for the Mantel index. That is, Equation (4) can be translated to apply to any and all $\psi \in \Psi$ to obtain a statistical distribution for the within-row gradient indices via a re-labeling of the rows and columns of $\mathbf{A}_r$ ($\mathbf{A}_r(\psi)$), as follows:

$$\Gamma_2(\mathbf{A}_q, \mathbf{A}_r(\psi)) = \frac{\sum_{i \neq j \neq k} b_{ijkq} b_{\psi(i)\psi(j)\psi(k)r}}{\sum_{i \neq j \neq k} |b_{ijkq} b_{\psi(i)\psi(j)\psi(k)r}|} \quad \forall \psi \in \Psi. \quad (5)$$

Two MATLAB m-files are available for the within-row gradient index for concordance. The first program, `triadexact.m` (see Appendix C) computes the within-row gradient index and obtains an exact p -value. This program is restricted to roughly $n = 12$. The second program, `triadapprox.m` (see Appendix D) computes the within-row gradient index and obtains an approximate p -value based on 100,000 random samples. This program is scalable for matrices where n is well into the hundreds. For larger matrices ($n > 1,000$), it might be necessary to reduce the number of random samples to 10,000 or 1,000 to obtain results in reasonable time.

Example 1: Acoustic confusion

Data and analyses

Our first example uses $Q = 4$ acoustic confusion matrices based on data originally collected by Morgan, Chambers, and Morton (1973), and subsequently analyzed by Hubert and Golledge (1977) and Brusco (2002). The first two confusion matrices in S pertain to a *recognition* task for $n = 9$ digits (1, 2, ..., 9): matrix $\mathbf{A}_1$ corresponds to recognition of a male speaker, whereas $\mathbf{A}_2$ is for a female speaker. The latter two matrices in S are associated with *memory* tasks for the same n

$= 9$ digits: matrices $\mathbf{A}_3$ and $\mathbf{A}_4$ are associated with two different female speakers. All four matrices in S were arranged so that the rows corresponded to the presented stimulus and the columns to the response. Moreover, following the procedure used by Brusco (2004) in his analysis of the concordance among 19 visual and tactual interletter confusion matrices, all four matrices in S were normalized based on the row (stimulus) sums. These normalized matrices are displayed in Table 1. For each matrix $\mathbf{A}_q$ in S , a_{ijq} is the percentage of responses of digit j when digit i was the presented stimulus.

A variety of different tests were performed using the matrices in S . First, the concordance between all pairs of matrices in S was measured using both the Mantel index and the within-row gradient index. For each pair, an exact test was performed for both the Mantel and within-row gradient indices using the `mantelexact.m` and `triadexact.m` programs, respectively. Additionally, for each pair of matrices, approximation tests were performed using the `mantelapprox.m` and `triadapprox.m` programs to assure that they were performing adequately.

We also obtained concordance measures for the symmetry of all four matrices in S . As described by Hubert and Baker (1979), this is accomplished using $\mathbf{A}_q$ and $\mathbf{A}'_q$ as the input matrices for the MATLAB programs. Exact and approximate tests of symmetry were obtained for all four matrices using the `mantelexact.m`, `triadexact.m`, `mantelapprox.m` and `triadapprox.m` programs.

Results

The results of our analyses are displayed in Table 2. One general observation is that the significance test results associated with the `mantelapprox.m` and `triadapprox.m` programs are very consistent with those obtained by their exact test counterparts, `mantelexact.m`, `triadexact.m`, respectively. There is no instance where a different conclusion regarding significance would be reached at either the $\alpha = .05$ or $\alpha = .01$ level depending on whether the exact or approximate test was used. This benchmarking is important because, in many circumstances, there is no choice but to use the approximate tests.

With respect to the tests limited to the recognition tasks, it is noted that strong concordance between $\mathbf{A}_1$ and $\mathbf{A}_2$ was measured for both the Mantel ($\Gamma_1(\mathbf{A}_1, \mathbf{A}_2) = .8638$) and within-row gradient ($\Gamma_2(\mathbf{A}_1, \mathbf{A}_2) = .6132$) measures. For both measures, the observed index was larger than every other index value in the reference distribution, thus resulting in the smallest possible p -values of $1/9! = .00000276$ and $1/100000 = .00001$ for the exact and approximate tests respectively. The concordance measures for symmetry were generally weaker. The Mantel index for the symmetry of $\mathbf{A}_1$ is particularly weak ($\Gamma_1(\mathbf{A}_1, \mathbf{A}'_1) = .1753$) and is not significant at $\alpha = .05$ (p -value = .09985670). Contrastingly, the within-row gradient index for $\mathbf{A}_1$ is stronger

Table 1 Row (stimulus) normalized confusion matrices from Morgan et al. (1973).

		1	2	3	4	5	6	7	8	9
Matrix A₁ Recognition task (male voice)	1	.34815	.01667	.02222	.03148	.27778	.01481	.03889	.02963	.22037
	2	.04283	.21788	.29981	.06145	.02980	.15456	.09497	.05959	.03911
	3	.06310	.11472	.27342	.07075	.03824	.07075	.13576	.11472	.11855
	4	.10280	.02991	.03738	.56075	.10841	.01869	.05421	.03364	.05421
	5	.03875	.00369	.00923	.01292	.82103	.00000	.00923	.01107	.09410
	6	.03738	.02804	.05421	.03925	.02056	.64673	.09720	.05047	.02617
	7	.05882	.04364	.05882	.05693	.07780	.07211	.51992	.05693	.05503
	8	.03955	.05650	.08286	.06780	.06026	.10546	.11111	.41243	.06403
	9	.10390	.01113	.00928	.02783	.20965	.00371	.01299	.02041	.60111
Matrix A₂ Recognition task (female voice)	1	.49327	.02627	.03587	.04036	.16272	.01922	.02627	.04100	.15503
	2	.06170	.24743	.28149	.05720	.04242	.07069	.12147	.05591	.06170
	3	.07781	.14405	.30868	.08039	.05209	.06302	.10804	.08939	.07653
	4	.06274	.03073	.03457	.65493	.07618	.02433	.04097	.03713	.03841
	5	.08184	.02238	.02238	.04284	.59910	.01023	.01662	.01790	.18670
	6	.04971	.06068	.06456	.04067	.04132	.51969	.12331	.07166	.02841
	7	.07821	.05064	.05449	.06410	.05321	.07051	.51731	.06987	.04167
	8	.07374	.05886	.09120	.11449	.03816	.15653	.09961	.33182	.03558
	9	.14037	.03863	.03155	.03928	.20863	.01417	.03284	.03091	.46362
Matrix A₃ Memory task (female voice #1)	1	.89041	.00875	.01294	.01332	.01826	.00419	.01256	.01560	.02397
	2	.00152	.96385	.01065	.00457	.00228	.00457	.00799	.00228	.00228
	3	.00609	.02055	.93798	.00495	.00304	.00723	.00913	.00837	.00266
	4	.00495	.01598	.01218	.92123	.01598	.00419	.01294	.00799	.00457
	5	.00951	.00913	.00989	.02131	.88090	.00647	.02207	.01294	.02778
	6	.00152	.00419	.00723	.00266	.00381	.95358	.01712	.00837	.00152
	7	.00152	.00533	.00190	.00114	.00304	.00266	.97793	.00342	.00304
	8	.00495	.00495	.00457	.00533	.00266	.00723	.00799	.95434	.00799
	9	.01865	.01065	.00799	.00381	.02740	.00419	.00951	.00951	.90830
Matrix A₄ Memory task (female voice #2)	1	.91091	.01018	.01018	.01236	.01455	.00364	.01527	.01127	.01164
	2	.00400	.94873	.01127	.00582	.00400	.00873	.00691	.00545	.00509
	3	.00255	.01273	.94982	.00582	.00473	.00873	.00691	.00509	.00364
	4	.00909	.01127	.00836	.92000	.01964	.00836	.00873	.00473	.00982
	5	.00836	.00836	.01273	.01564	.88218	.00800	.01855	.01345	.03273
	6	.00000	.00327	.00509	.00400	.00145	.96509	.01345	.00618	.00145
	7	.00218	.00218	.00400	.00255	.00182	.00509	.97891	.00145	.00182
	8	.00582	.00618	.00473	.00509	.00655	.00836	.00873	.94545	.00909
	9	.01527	.00764	.00473	.00327	.01600	.00327	.01018	.00655	.93309

($\Gamma_2(\mathbf{A}_1, \mathbf{A}'_1) = .3415$) and is significant at $\alpha = .05$ (p -value = .00395172). This is clear evidence that the two concordance measures do not necessarily lead to the same conclusions.

Turning now to the tests limited to the memory tasks, a similar pattern is observed, yet there are some differences. Strong concordance between **A₃** and **A₄** was measured for both the Mantel ($\Gamma_1(\mathbf{A}_3, \mathbf{A}_4) = .8554$) and within-row gradient ($\Gamma_2(\mathbf{A}_3, \mathbf{A}_4) = .5745$) measures and the smallest possible p -values were observed for these measures. Once again, the concordance measures for symmetry were appreciable

weaker. However, unlike the results for the symmetry of the recognition matrices, it was the within-row gradient indices that were especially poor for the memory matrices. The Mantel indices for the symmetry of **A₃** and **A₄** were ($\Gamma_1(\mathbf{A}_3, \mathbf{A}'_3) = .3204$) and ($\Gamma_1(\mathbf{A}_4, \mathbf{A}'_4) = .2263$), respectively. Although not particularly large, both of these indices are significant at $\alpha = .05$. Contrastingly, the within-row gradient indices for the symmetry of **A₃** and **A₄** were weaker at ($\Gamma_2(\mathbf{A}_3, \mathbf{A}'_3) = .1638$) and ($\Gamma_2(\mathbf{A}_4, \mathbf{A}'_4) = .1765$), respectively, and neither of these indices is significant at $\alpha = .05$. Thus,

Table 2 Results for the acoustic confusion data – exact vs. approximate p -values

	$\Gamma_1(\mathbf{A}_q, \mathbf{A}_r)$	Exact p -value	Approx. p -value	$\Gamma_2(\mathbf{A}_q, \mathbf{A}_r)$	Exact p -value	Approx. p -value
<i>Recognition tasks</i>						
$\mathbf{A}_1 - \mathbf{A}_2$ agreement	.8638	.00000276	.00001	.6132	.00000276	.00001
$\mathbf{A}_1$ symmetry	.1753	.09985670	.09897	.3415	.00395172	.00366
$\mathbf{A}_2$ symmetry	.5899	.00067240	.00066	.4458	.00022873	.00014
<i>Memory tasks</i>						
$\mathbf{A}_3 - \mathbf{A}_4$ agreement	.8554	.00000276	.00001	.5745	.00000276	.00001
$\mathbf{A}_3$ symmetry	.3204	.00938327	.00946	.1638	.08177083	.08131
$\mathbf{A}_4$ symmetry	.2263	.03683311	.03741	.1765	.05725860	.05621
<i>Recognition vs. Memory tasks</i>						
$\mathbf{A}_1 - \mathbf{A}_3$ agreement	.3343	.01243937	.01313	.4237	.00019015	.00020
$\mathbf{A}_1 - \mathbf{A}_4$ agreement	.2629	.03598435	.03597	.5210	.00000276	.00001
$\mathbf{A}_2 - \mathbf{A}_3$ agreement	.4198	.00257937	.00256	.4310	.00012952	.00011
$\mathbf{A}_2 - \mathbf{A}_4$ agreement	.3523	.00928957	.00911	.3444	.00304233	.00285
$\mathbf{A}_1 - \mathbf{A}_3$ (transpose) agreement	.4812	.00099206	.00096	.2119	.02797068	.02881
$\mathbf{A}_1 - \mathbf{A}_4$ (transpose) agreement	.3782	.00399857	.00402	.2199	.02769786	.02780
$\mathbf{A}_2 - \mathbf{A}_3$ (transpose) agreement	.6020	.00006338	.00006	.2469	.00712357	.00688
$\mathbf{A}_2 - \mathbf{A}_4$ (transpose) agreement	.4659	.00033896	.00045	.2263	.01805381	.01892

again there is evidence that the two concordance measures do not lead to the same conclusions, but here the situation is reversed (i.e., the Mantel index concordance is stronger).

Next, we consider concordance between pairs of matrices with one member of the pair corresponding to a recognition task and the other to a memory task. The Mantel indices associated with these comparisons were fairly weak, particularly in the cases of ($\Gamma_1(\mathbf{A}_1, \mathbf{A}_3) = .3343$) and ($\Gamma_1(\mathbf{A}_1, \mathbf{A}_4) = .2629$), which were not significant at $\alpha = .01$. The within-row gradient indices were markedly stronger in most cases, and are significant for all pairs at $\alpha = .005$.

Brusco (2002) found that optimal permutations for the memory tasks were nearly the reverse of those obtained for the recognition tasks, which led to the suspicion that one set of these matrices might have been transposed. To investigate further, we examined the concordance between the recognition matrices and the transposes of the memory matrices. The Mantel indices improve substantially (and are significant at $\alpha = .005$) when the transposes of the memory matrices are used. For example, ($\Gamma_1(\mathbf{A}_2, \mathbf{A}_3) = .4198$), but ($\Gamma_1(\mathbf{A}_2, \mathbf{A}_3') = .6020$). On the other hand, the within-row gradient indices deteriorate sharply when the transposes of the memory matrices are used. For example, ($\Gamma_2(\mathbf{A}_2, \mathbf{A}_3) = .4310$), but ($\Gamma_2(\mathbf{A}_2, \mathbf{A}_3') = .2469$). Three of the four within-row gradient indices associated with the transposes are not significant at $\alpha = .01$. Accordingly, whereas the results for the Mantel index support the supposition that one set of matrices was transposed, the within-row gradient results certainly do not. Whichever the case, it is unequivocally clear that the degree of concordance measured

by the Mantel and within-row gradient indices can, in some instances, be profoundly different.

Computational sensitivity and efficiency

In this subsection, we more carefully explore the computational *sensitivity* and *efficiency* of the MATLAB programs for matrix agreement. In particular, sensitivity refers to the p -value effects associated with the choice of the number of random permutations for the `mantelapprox.m` and `triadapprox.m` programs. Efficiency pertains to the computation time required by the programs. To facilitate this examination, we began by obtaining the computation times for the `mantelexact.m` and `triadexact.m` programs for each of the test conditions in Table 2. These computation times were obtained using a 2.2 GHz Pentium 4 computer with 1 GB of RAM. In addition, we ran 10 replications of the `mantelapprox.m` and `triadapprox.m` programs (using 100,000 random samples) for each of the test conditions. We then repeated this process for 10 replications, but after reducing the number of random samples to 1,000. For each test condition, we stored the minimum, mean, and maximum p -values across the 10 replications, as well as the minimum, mean, and maximum computation times. The results of these analyses are reported in Tables 3 and 4 for the Mantel index and within-row gradient index, respectively.

The computation times for the `mantelexact.m` program when applied to the acoustic confusion data ($n = 9$) ranged from 4.300 to 4.345 s. For $n = 10$, the number of

Table 3 Results for the acoustic confusion data – computational sensitivity and efficiency results for the Mantel test. The min, mean, and max values are based on 10 replications of the `mantelapprox.m` program for either 100,000 or 1,000 random permutations. The top and bottom panels present p -value and computation time comparisons, respectively

	100,000 permutations			1,000 permutations			Exact
	min	mean	max	min	mean	max	algorithm
A ₁ – A ₂ agreement	.000010	.000012	.000030	.0010	.0010	.0010	.00000276
A ₁ symmetry	.098830	.099866	.101090	.0890	.0998	.1210	.09985670
A ₂ symmetry	.000550	.000706	.000890	.0010	.0013	.0020	.00067240
A ₃ – A ₄ agreement	.000010	.000014	.000030	.0010	.0010	.0010	.00000276
A ₃ symmetry	.009020	.009427	.009960	.0060	.0111	.0190	.00938327
A ₄ symmetry	.035820	.036729	.037570	.0240	.0403	.0540	.03683311
A ₁ – A ₃ agreement	.011580	.012337	.012840	.0090	.0155	.0210	.01243937
A ₁ – A ₄ agreement	.035300	.036045	.037070	.0320	.0377	.0460	.03598435
A ₂ – A ₃ agreement	.002240	.002523	.002790	.0020	.0040	.0060	.00257937
A ₂ – A ₄ agreement	.008770	.009213	.009570	.0090	.0123	.0170	.00928957
A ₁ – A ₃ (tr) agreement	.000890	.000977	.001100	.0010	.0022	.0040	.00099206
A ₁ – A ₄ (tr) agreement	.003660	.003923	.004140	.0020	.0050	.0080	.00399857
A ₂ – A ₃ (tr) agreement	.000050	.000077	.000140	.0010	.0010	.0010	.00006338
A ₂ – A ₄ (tr) agreement	.000220	.000333	.000430	.0010	.0018	.0030	.00033896
A ₁ – A ₂ agreement	2.909	2.918	2.927	.029	.030	.030	4.345
A ₁ symmetry	2.912	2.919	2.931	.030	.030	.030	4.325
A ₂ symmetry	2.910	2.917	2.921	.030	.030	.030	4.327
A ₃ – A ₄ agreement	2.879	2.910	2.929	.030	.030	.030	4.343
A ₃ symmetry	2.883	2.906	2.924	.030	.030	.030	4.327
A ₄ symmetry	2.907	2.918	2.937	.030	.030	.030	4.306
A ₁ – A ₃ agreement	2.911	2.915	2.922	.030	.030	.030	4.330
A ₁ – A ₄ agreement	2.912	2.917	2.927	.030	.030	.034	4.326
A ₂ – A ₃ agreement	2.910	2.917	2.925	.030	.030	.030	4.325
A ₂ – A ₄ agreement	2.910	2.924	2.952	.030	.030	.030	4.325
A ₁ – A ₃ (tr) agreement	2.910	2.919	2.940	.030	.030	.030	4.300
A ₁ – A ₄ (tr) agreement	2.910	2.920	2.929	.030	.030	.030	4.313
A ₂ – A ₃ (tr) agreement	2.910	2.920	2.934	.030	.030	.030	4.312
A ₂ – A ₄ (tr) agreement	2.910	2.917	2.932	.030	.030	.030	4.323

permutations would be 10 times greater, and thus the computation times would tend to be at least 10 times larger than those for $n = 9$. It is clear that just a few more increments of n will result in the infeasibility (or, at least, impracticality) of the `mantelexact.m` program. For example, one of our later illustrations is for a context where $n = 17$. The number of permutations for $n = 17$ ($17!$) is more than 980 million times greater than the number of permutations for $n = 9$ ($9!$). Multiplying 4.3 s by 980 million would yield a conservative estimate of roughly 133 years of computation time for $n = 17$. However, this likely understates the required time for $n = 17$ because, in addition to the change in the number of permutations, the computation necessary with each individual permutation is also greater for $n = 17$

than it is at $n = 9$ (i.e., more product terms). The computation times for the `triadexact.m` program when applied to the acoustic confusion data ($n = 9$) ranged from 30.881 to 31.099 s. The fact that these times are roughly seven times greater than those for the `mantelexact.m` program is attributable to the need for conditional evaluation of triads rather than mere element-wise products.

Tables 3 and 4 also reveal that, for a given number of random permutations, the computation times for the `mantelapprox.m` and `triadapprox.m` programs are extremely consistent, with little variability both across the test conditions and within a test condition (across the 10 replications). Moreover, as expected, the computation times for 100,000 random permutations are roughly 100 times

Table 4 Results for the acoustic confusion data – computational sensitivity and efficiency results for the within-row gradient test. The min, mean, and max values are based on 10 replications of the `triadapprox.m` program for either 100,000 or 1,000 random permutations. The top and bottom panels present p -value and computation time comparisons, respectively

	100,000 permutations			1,000 permutations			Exact
	min	mean	max	min	mean	max	algorithm
A ₁ – A ₂ agreement	.000010	.000014	.000030	.0010	.0010	.0010	.00000276
A ₁ symmetry	.003630	.003921	.004230	.0020	.0043	.0080	.00395172
A ₂ symmetry	.000160	.000231	.000330	.0010	.0011	.0020	.00022873
A ₃ – A ₄ agreement	.000010	.000014	.000020	.0010	.0010	.0010	.00000276
A ₃ symmetry	.081050	.081768	.082460	.0770	.0872	.1000	.08177083
A ₄ symmetry	.056900	.057521	.058660	.0450	.0584	.0670	.05725860
A ₁ – A ₃ agreement	.000130	.000195	.000270	.0010	.0013	.0020	.00019015
A ₁ – A ₄ agreement	.000010	.000010	.000010	.0010	.0010	.0010	.00000276
A ₂ – A ₃ agreement	.000080	.000136	.000180	.0010	.0012	.0030	.00012952
A ₂ – A ₄ agreement	.002940	.003102	.003350	.0020	.0050	.0080	.00304233
A ₁ – A ₃ (tr) agreement	.027420	.028100	.028880	.0190	.0270	.0380	.02797068
A ₁ – A ₄ (tr) agreement	.026990	.027569	.028070	.0210	.0279	.0380	.02769786
A ₂ – A ₃ (tr) agreement	.006670	.007255	.007730	.0050	.0078	.0120	.00712357
A ₂ – A ₄ (tr) agreement	.017600	.018116	.018470	.0120	.0183	.0220	.01805381
A ₁ – A ₂ agreement	13.100	13.108	13.130	.130	.131	.132	31.058
A ₁ symmetry	13.089	13.114	13.153	.130	.131	.131	30.889
A ₂ symmetry	13.108	13.127	13.144	.130	.131	.131	31.041
A ₃ – A ₄ agreement	13.104	13.117	13.130	.130	.131	.131	30.924
A ₃ symmetry	13.114	13.127	13.145	.130	.131	.131	31.049
A ₄ symmetry	13.086	13.107	13.152	.130	.131	.132	30.891
A ₁ – A ₃ agreement	13.118	13.134	13.182	.130	.131	.131	31.055
A ₁ – A ₄ agreement	13.107	13.116	13.121	.130	.131	.131	30.895
A ₂ – A ₃ agreement	13.109	13.126	13.141	.131	.131	.131	31.053
A ₂ – A ₄ agreement	13.100	13.118	13.145	.130	.131	.131	30.889
A ₁ – A ₃ (tr) agreement	13.090	13.107	13.134	.131	.131	.132	31.099
A ₁ – A ₄ (tr) agreement	13.095	13.123	13.143	.131	.131	.131	30.881
A ₂ – A ₃ (tr) agreement	13.112	13.124	13.135	.131	.131	.131	31.098
A ₂ – A ₄ (tr) agreement	13.104	13.116	13.129	.131	.131	.131	30.881

greater than the computations for 1,000 random permutations.

For the `mantelapprox.m` using 100,000 random permutations, the variability of the p -values across the 10 replicates was generally modest. There was no instance where a different conclusion regarding significance would be reached at either the $\alpha = .05$ or $\alpha = .01$ level depending on whether the minimum or maximum p -value across the 10 replications was used. However, for the A₁ symmetry test using $\alpha = .10$, the maximum p -value of .10109 would result in a failure to reject the null, whereas the minimum, mean, and exact p -values (all less than .10) p -value) would lead to rejection. When the number of random permutations was reduced to 1,000, the variability of the p -values across

the 10 replications was still fairly modest. However, there were several other circumstances where a different conclusion could be reached depending on the minimum or maximum p -value across the 10 replications was used: (i) A₃ symmetry test if $\alpha = .01$, (ii) A₄ symmetry test if $\alpha = .05$, (iii) A₁–A₃ agreement at $\alpha = .01$, and (iv) A₂–A₄ agreement at $\alpha = .01$.

For the `triadapprox.m` using 100,000 random permutations, the variability of the p -values across the 10 replicates were modest and, again, there was no instance where a different conclusion would be reached at the $\alpha = .05$ or $\alpha = .01$ level, depending on whether the minimum or maximum p -value across the 10 replications was used. When only 1,000 random permutations were used, the following test conditions would have different conclusion

depending on whether the minimum or maximum across the 10 replications was used: (1) A_4 symmetry test if $\alpha = .05$, and (2) A_2 - A_3' agreement at $\alpha = .01$.

Example 2: Visual confusion of textured materials

Data and analyses

Our second example uses $Q = 3$ visual confusion matrices based on data originally collected by Cho, Yang, and Hallett (2000), and subsequently analyzed by Brusco (2002) and Brusco and Steinley (2012). The data pertain to recognition of $n = 20$ textured materials at different distances: matrices A_1 , A_2 , and A_3 correspond to distances of 8.2, 15.5, and 22.9 m, respectively. As the confusion matrices were already row normalized, there was no reason to apply additional standardization.

The exact test programs were computationally infeasible for this example because $n = 20$ is far too large to enumerate all possible permutations. Therefore, for each pair of matrices in S , the concordance indices were computed and approximation tests performed using the `mantelapprox.m` and `triadapprox.m` programs. In addition, symmetry tests were performed for each of the three matrices using the same programs.

Results

The results of the analyses for the confusion matrices pertaining to textured materials are displayed in Table 5. It is immediately evident that the significance tests associated with the `mantelapprox.m` and `triadapprox.m` programs are all significant at the $\alpha = .00005$ level. To a large extent, this high degree of statistical significance relative to our findings in Example 1 is attributable to the larger number of stimuli ($n = 20$ in Example 2 vs. $n = 9$ in Example 1). In fact, as n increases, statistical significance is almost assured and,

Table 5 Results for the visual judgments of textured materials confusion data

	$\Gamma_1(A_q, A_r)$	Approx. p -value	$\Gamma_2(A_q, A_r)$	Approx. p -value
<i>Comparisons at different distances</i>				
$A_1 - A_2$ agreement	.7252	.00001	.7488	.00001
$A_1 - A_3$ agreement	.5561	.00001	.6156	.00001
$A_2 - A_3$ agreement	.8351	.00001	.7459	.00001
<i>Symmetry tests</i>				
A_1 symmetry	.6470	.00001	.5364	.00001
A_2 symmetry	.3587	.00003	.4390	.00001
A_3 symmetry	.3509	.00001	.5068	.00001

therefore, it is differences in the strength of the concordance indices that help to establish differences.

The concordance among the three matrices in S is generally strong, regardless of whether the Mantel or within-row gradient measure is used. The weakest agreement is between matrices A_1 and A_3 , where $(\Gamma_1(A_1, A_3) = .5561)$ and $(\Gamma_2(A_1, A_3) = .6156)$. This is not surprising given that matrices A_1 and A_3 correspond to judgments at the shortest and longest distances, respectively. When measuring concordance using the Mantel index, the agreement measured between the middle and longest distance $(\Gamma_1(A_2, A_3) = .8351)$ is a good bit stronger than the agreement between the shortest and middle distance $(\Gamma_1(A_1, A_2) = .7252)$. However, when using the within-row concordance measure, the agreements are very similar, with the agreement between the shortest and middle distance $(\Gamma_2(A_1, A_2) = .7488)$ slightly better than the agreement between the middle and longest distance $(\Gamma_2(A_2, A_3) = .7459)$.

The Mantel concordance index for the symmetry of A_1 was very strong $(\Gamma_1(A_1, A_1') = .6470)$. Although still highly significant, the Mantel indices for symmetry were appreciably weaker for the greater distances, where $(\Gamma_1(A_2, A_2') = .3587)$ and $(\Gamma_1(A_3, A_3') = .3509)$. By contrast, the within-row gradient indices for symmetry were much more comparable across the three distances: $(\Gamma_2(A_1, A_1') = .5364)$, $(\Gamma_2(A_2, A_2') = .4390)$, and $(\Gamma_2(A_3, A_3') = .5068)$.

Example 3: Sociometric rankings of fraternity members

Data and analyses

Our third example uses $Q = 3$ sociometric matrices among fraternity members, and is based on data originally collected by Nordlie (1958) and Newcomb (1961, 1968), and also published by Doreian et al. (2005, pp. 45–46). At different time points during the semester, $n = 17$ members of a (pseudo-)fraternity ranked each of the other 16 members of the fraternity with respect to who they liked the most: matrices A_1 , A_2 , and A_3 correspond to rankings collected after the first, tenth, and last weeks of the semester, respectively. For each pair of matrices in S , the concordance indices were computed and approximation tests performed using the `mantelapprox.m` and `triadapprox.m` programs. Symmetry tests were performed for each of the three matrices using the same programs.

Results

The results of the analyses for the sociometric matrices among fraternity members are displayed in Table 6. Although all of

Table 6 Results for the fraternity network affinity rankings

	$\Gamma_1(\mathbf{A}_q, \mathbf{A}_r)$	Approx. p -value	$\Gamma_2(\mathbf{A}_q, \mathbf{A}_r)$	Approx. p -value
<i>Comparisons at different times</i>				
$\mathbf{A}_1 - \mathbf{A}_2$ agreement	.3375	.00002	.2428	.00002
$\mathbf{A}_1 - \mathbf{A}_3$ agreement	.2994	.00016	.2202	.00010
$\mathbf{A}_2 - \mathbf{A}_3$ agreement	.8481	.00001	.7010	.00001
<i>Symmetry tests</i>				
$\mathbf{A}_1$ symmetry	.2830	.00001	.2368	.00001
$\mathbf{A}_2$ symmetry	.3346	.00001	.3435	.00001
$\mathbf{A}_3$ symmetry	.2720	.00001	.2943	.00001

the significance tests associated with the `mantelapprox.m` and `triadapprox.m` programs are significant at the $\alpha = .0005$ level, there is substantial disparity with respect to the magnitude of the index values. The concordance between the sociometric matrices measured at the 10th and last weeks is high, with Mantel and within-row gradient indices of ($\Gamma_1(\mathbf{A}_2, \mathbf{A}_3) = .8481$) and ($\Gamma_2(\mathbf{A}_2, \mathbf{A}_3) = .7010$), respectively. However, the concordance between other pairs of matrices is appreciably lower. For example, the concordance between the matrices measured at the first and tenth weeks is measured at ($\Gamma_1(\mathbf{A}_1, \mathbf{A}_2) = .3375$) and ($\Gamma_2(\mathbf{A}_1, \mathbf{A}_2) = .2428$) for the Mantel and within-row gradient indices, respectively. The concordance indices between the first and last week measurements are even lower. The rationale for these findings is that dramatic refinements of rankings occur early in the semester but, by the 10th week, the fraternity members are more steadfast in their rankings, which remain consistent throughout the rest of the semester.

All three of the sociometric matrices exhibited symmetry, with approximated p -values of .00001 for both the Mantel and within-row gradient indices. A degree of symmetry is to be expected, as one member's rating of the other fraternity members is apt to have some relationship with how those member's rate that particular member. It is interesting, however, that the strongest symmetry was observed for the tenth week. That is, it would be reasonable to conjecture that symmetry might strengthen over time, reaching its maximum at the final week. However, both the Mantel and within-row gradient indices decreased when moving from the tenth to the last week.

An interesting aspect of this third example is that, relative to the previous two examples, it exhibits much greater consistency between the Mantel and within-row gradient indices. This finding is likely attributable to the nature of the data. The rows of the confusion matrices in the first two examples exhibit some sharp differences with respect to the distribution of values within each row and column. However, in this third example, the

values are bounded between 1 and 16. Moreover, within each row of the matrices, the values are always a permutation of the integers 1 through 16 because of the nature of the ranking process. This reduced variability in the matrix elements likely fosters a more stable Mantel index. That is, one that is less drastically affected by the placement of a few large elements.

Summary and conclusions

The problem of comparing the agreement of matrices has an extensive history in the behavioral sciences. Applications include the analysis of recognition/confusion data, similarity judgments, and social network relations. The Mantel index is the most well-known agreement index because of its relationship to correlation. However, as noted by Hubert (1978, 1987) and Brusco (2004), the Mantel index can be profoundly affected by just a few elements in the matrices. Alternative measures of agreement based on the patterning of elements within rows or columns can also be used to study matrix agreement.

In this paper, we have shown that the Mantel and within-row gradient indices often provide similar results. However, in some instances, they yield very different assessments of agreement. Moreover, the disparity in agreement can occur in either of two ways: Mantel index agreement is strong and within-row gradient agreement is weak, or vice versa.

Exact p -values for the Mantel and within-row gradient indices can be obtained for modestly-sized matrices ($n \leq 12$, perhaps slightly larger on a fast computer), but random sampling is necessary to obtain approximate p -values for larger matrices. The number of random permutations required depends on the degree of precision necessary for the p -value, as well as the size of the matrices. For common levels of $\alpha = .05$ or $\alpha = .01$, 100,000 permutations is almost certainly sufficient, and is recommended when feasible. For larger matrices where 100,000 permutations is computationally demanding, 1,000 permutations is a reasonable starting point. If the p -value falls on the interval $[\alpha, .07]$, then a larger number of permutations should be evaluated. Finally, it should also be noted that larger matrices commonly have enough agreement to produce small p -values, even when the matrix agreement is relatively modest (see Examples 2 and 3). This occurrence is similar to the observance of statistically significant small correlations when sample sizes are large. The implication of this finding is that, when studying matrix agreement for larger matrices, it is often more relevant to compare agreement indices than conduct actual significance tests.

References

- Bonnie, K. E., & de Waal, F. B. W. (2006). Affiliation promotes the transmission of a social custom: Handclasp grooming among captive chimpanzees. *Primates*, 47, 27–34.
- Brusco, M. J. (2002). Identifying a reordering of the rows and columns of multiple proximity matrices using multiobjective programming. *Journal of Mathematical Psychology*, 46, 731–745.
- Brusco, M. J. (2004). On the concordance among empirical confusion matrices for visual and tactual letter recognition. *Perception & Psychophysics*, 66, 392–397.
- Brusco, M. J., & Steinley, D. (2012). A note on the estimation of the Pareto efficient set for multiobjective matrix permutation problems. *British Journal of Mathematical and Statistical Psychology*, 65, 145–162.
- Burris, V. (2005). Interlocking directorates and political cohesion among corporate elites. *American Journal of Sociology*, 111, 249–283.
- Cho, R. Y., Yang, V., & Hallett, P. E. (2000). Reliability and dimensionality of judgments of visually textured materials. *Perception & Psychophysics*, 62, 735–752.
- Doreian, P., Batagelj, V., & Ferligoj, A. (2005). *Generalized Blockmodeling*. Cambridge: Cambridge University Press.
- Dreiling, M., & Darves, D. (2011). Corporate unity in American trade policy: A network analysis of corporate-dyad political action. *American Journal of Sociology*, 116, 1514–1563.
- Fliaster, A., & Schloderer, F. (2010). Dyadic ties among employees: Empirical analysis of creative performance and efficiency. *Human Relations*, 63, 1513–1540.
- Glerean, E., Pan, R. K., Salmi, J., Kujala, R., Lahnakoski, J. M., Roine, U., Nummenmaa, L., Leppämäki, S., Nieminen-von Wendt, T., Tani, P., Saramäki, J., Sams, M., & Jääskeläinen, I. P. (2016). Reorganization of functionally connected brain subnetworks in high-functioning autism. *Human Brain Mapping*, 37, 1066–1079. doi:<https://doi.org/10.1002/hbm.23084>
- Gliner, G. S. (1981). A note on a statistical paradigm for the evaluation of cognitive structure in physics instruction. *Applied Psychological Measurement*, 5, 493–502.
- Harris, F. N., & Packard, T. (1985). Intensity judgments of emotion words: Implications for counselor training. *Journal of Counseling Psychology*, 32, 288–291.
- Hirsch, B. T., Prange, S., Hauver, S. A., & Gehrt, S. D. (2013). Genetic relatedness does not predict racoon social network structure. *Animal Behaviour*, 85, 463–470.
- Holloway, E. L. (1982). Interactional structure of the supervision interview. *Journal of Counseling Psychology*, 29, 309–317.
- Hubert, L. J. (1978). Generalized proximity function comparisons. *British Journal of Mathematical and Statistical Psychology*, 31, 179–192.
- Hubert, L. (1987). *Assignment methods in combinatorial data analysis*. New York: Marcel Dekker.
- Hubert, L. J., & Baker, F. B. (1979). Evaluating the symmetry of a proximity matrix. *Quality and Quantity*, 13, 77–84.
- Hubert, L. J., & Golledge, R. G. (1977). The comparison and fitting of given classification schemes. *Journal of Mathematical Psychology*, 16, 233–253.
- Hubert, L. J., & Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29, 190–241.
- Kwon, K., & Lease, A. M. (2014). Perceived influence of close friends, well-liked peers, and popular peers: Reputational or personal influence? *Journal of Social and Personal Relationships*, 31, 1116–1133.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.
- Medina-Diaz, M. (1993). Analysis of cognitive structure using the linear logistic test model and quadratic assignment. *Applied Psychological Measurement*, 17, 117–130.
- Miller, G. A., & Nicely, P. E. (1955). Analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338–352.
- Mitani, J. C., Merriwether, D. A., & Zhang, C. (2000). Male affiliation, cooperation and kinship in wild chimpanzees. *Animal Behaviour*, 59, 885–893.
- Moolenaar, N. M., Slegers, P. J. C., Karsten, S., & Daly, A. J. (2012). The social fabric of elementary schools: A network typology of social interaction among teachers. *Educational Studies*, 38, 355–371.
- Morgan, B. J. T., Chambers, S. M., & Morton, J. (1973). Acoustic confusion of digits in memory and recognition. *Perception & Psychophysics*, 14, 375–383.
- Newcomb, T. M. (1961). *The acquaintance process*. New York: Holt, Rinehart, and Winston.
- Newcomb, T. M. (1968). Interpersonal balance. In R. Abelson, E. Aronson, W. McGuire, T. Newcomb, M. Rosenberg, & O. Tannenbaum (Eds.), *Theories of cognitive consistency: A source book* (pp. 28–51). Chicago: Rand McNally.
- Nordlie, P. (1958). *A longitudinal study of interpersonal interaction in a natural group setting*. Ph.D. Thesis. Ann Arbor, University of Michigan.