A Supplemental Material

Proof of Proposition 8. Let the column sums of P be $\mathcal{C} = \{c_1, \ldots, c_v\}$ and row sums of Q be $\mathcal{R} = \{r_1, \ldots, r_u\}$. (P, Q) is a SK stable pair implies that column normalization of P equals Q, i.e. if $q_{ij} > 0$ then $q_{ij} = p_{ij}/c_j$, and further row normalization of Q equals P, i.e. if $p_{ij} > 0$ then $p_{ij} = q_{ij}/r_i$. So we have that $p_{ij} = p_{ij}/(r_i \cdot c_j) \implies r_i \cdot c_j = 1$ (Claim $(*)$). Therefore $c_j \in \mathcal{C} \implies 1/c_j \in \mathcal{R}$ and $r_i \in \mathcal{R} \implies 1/r_i \in \mathcal{C}$. In particular, let $c_{max} = \max\{c_1, \ldots, c_v\}$ and $r_{min} = \min\{r_1, \ldots, r_u\}$. We have that $c_{max} = 1/r_{min}$.

With permutation, we may assume that the columns with sum c_{max} in P are the first v_1 columns and the rows with sum r_{min} in Q are the first u_1 rows. Note that an element p_{ij} in the first v_1 columns of P is positive only if it is in the first u_1 rows. Otherwise assume that $p_{i_0 j} > 0$ for $i_0 > u_1$, then Claim $(*)$ implies that $c_j \cdot r_{i_0} = 1 \implies r_{i_0} = 1/c_j = 1/c_{max} = r_{min}$. This contradicts to $i_0 > u_1$. Similarly, we may show that an element q_{ij} in the first u_1 rows of Q is positive only if it is in the first v_1 columns. Further note that Q and P have the same pattern. So we have that for $i \leq u_1$, $p_{ij} > 0$ only if $j \leq v_1$. Therefore let B_1 be the submatrix of P formed by the first u_1 rows and first v_1 columns and P_1 (Q_1) be the submatrix of P (Q) formed by the last $u - u_1$ rows and the last $v - v_1$ columns. We just showed that $P = \text{diag}\{B_1, P_1\}$ and the column sum c_{max} of B_1 is a constant (equals u_1/v_1). (P_1, Q_1) is a *SK stable* pair with smaller dimension. Hence, inductively, the proposition holds. $\square$

Lemma A.1. *If a pair of matrices (P, Q) as in Proposition 9 exists, the pattern of any pair of limit matrices $(\mathbf{L}', \mathbf{T}')$ is intermediate between the pattern of (P, Q) and the pattern of $\mathbf{M}$, namely, $(P, Q) \prec (\mathbf{L}', \mathbf{T}') \prec \mathbf{M}$.*

Proof. Let the dimension of $\mathbf{M}$ be $u \times v$. Denote the sequence of matrices generated by *SK iteration* by $\{\mathbf{L}^n, \mathbf{T}^n\}(n > 0)$, where $\mathbf{L}^n$ are row normalized and has column sums $\{c_{jn}\}_j$, and $\mathbf{T}^n$ are column normalized and has row sums $\{r_{in}\}_i$. As explained in Pretzel (1980), there exist diagonal matrices X_n and Y_n such that $\mathbf{L}^n = X_n \mathbf{M} Y_n$ and $\mathbf{T}^n = X_n \mathbf{M} Y_{n+1}$. In particular, $X_n = \text{diag}\{x_{1n}, \ldots, x_{un}\}$ and $Y_n = \text{diag}\{y_{1n}, \ldots, y_{vn}\}$, where each x_{in} is the product of row normalizing constants (reciprocal of row sums) of row-i from step 1 to n and each y_{jn} is the product of column normalizing constants (reciprocal of column sums) of column-j from step 1 to n. Here, Y_1 is the identity matrix.

Denote the row sums of Q by $\{r_i\}_{i=1}^u$ and the column sums of P by $\{c_j\}_{j=1}^v$. Consider the following functions, we will show that they form an increasing sequence(the use of it will be clear later).

$$f_n = \prod_{i=1}^{u} x_{in}^{1+\alpha r_i} \prod_{j=1}^{v} y_{jn}^{\alpha + c_j},$$

$$g_n = \prod_{i=1}^{u} x_{in}^{1+\alpha r_i} \prod_{j=1}^{v} y_{jn+1}^{\alpha + c_j},$$

where, $\alpha \geq -\dfrac{\log s}{\log r}$, with $s = \frac{1}{\prod_j c_j^{c_j}}$ and $r = \left(\frac{v}{u}\right)^v$.

$$\frac{g_n}{f_n} = \prod_{j=1}^{v} \left(\frac{y_{jn+1}}{y_{jn}}\right)^{\alpha + c_j} = \prod_{j=1}^{v} \left(\frac{1}{c_{jn}}\right)^{\alpha + c_j} \geq \prod_{j=1}^{v} \left(\frac{1}{c_{jn}}\right)^{c_j} \prod_{j=1}^{v} \left(\frac{1}{c_{jn}}\right)^{\alpha}. \tag{3}$$

Due to Lemma 1 of Berry et al. (2007) we have, $\frac{1}{\prod_j c_{jn}^{c_j}} \geq \frac{1}{\prod_j c_j^{c_j}} = s$, i.e. the first product of the right hand side of Inequality (3) is greater or equal to s.

Moreover, by arithmetic and geometric means inequality, $\left(\prod_{j=1}^{v} c_{jn}\right)^{\frac{1}{v}} \leq \dfrac{\sum_{j=1}^{v} c_{jn}}{v} = \dfrac{u}{v}$. Hence, $\prod_{j=1}^{v} \dfrac{1}{c_{jn}} = r$. Therefore $\dfrac{g_n}{f_n} \geq s r^{\alpha} \geq 1$, where the second inequality holds because of the choice of α. Hence we have $g_n \geq f_n$. The analogous argument holds for f_{n+1}/g_n. So, we have $f_{n+1} \geq g_n \geq f_n$ (Claim $*$).

Now recall that $\mathbf{L}^n = X_n \mathbf{M} Y_n$. In particular, we have $l_{ij}^n = x_{in} m_{ij} y_{jn}$, for $m_{ij} \neq 0$. So $x_{in} y_{jn} = \frac{l_{ij}^n}{m_{ij}}$ and it is bounded above because the elements l_{ij}^n are bounded above by 1. One possible upper bound is $K = \frac{1}{\min m_{ij}}$, where min is taken over non zero elements in $\mathbf{M}$.

Moreover, let $d_{ij} = p_{ij} + \alpha\, q_{ij}$, then

$$\begin{aligned}\prod_{ij} (x_{in} y_{jn})^{d_{ij}} &= \prod_{ij} (x_{in} y_{jn})^{p_{ij} + \alpha\, q_{ij}} \\ &= \prod_i x_{in}^{\sum_j p_{ij} + \alpha\, q_{ij}} \prod_j y_{jn}^{\sum_i p_{ij} + \alpha\, q_{ij}} \\ &= \prod_i x_{in}^{1+\alpha \cdot r_i} \prod_j y_{jn}^{\alpha + c_j} = f_n.\end{aligned}$$

Furthermore, if $d_{ij} \neq 0$, then $p_{ij} \neq 0 \implies m_{ij} \neq 0$ and hence $x_{in} y_{jn} \leq K$. Therefore $f_n \leq K^d$, where $d = \sum_{ij} d_{ij}$. Together with Claim $*$, we have $(x_{in} y_{jn})^{d_{ij}} K^{(d - d_{ij})} \geq f_n \geq f_1$. So if $d_{ij} \neq 0$, then

$x_{in}y_{jn}$ is bounded away from zero. Thus it follows that l_{ij}^n is bounded away from zero for all n. Therefore $P \prec \mathbf{L}'$, where $\mathbf{L}'$ is the limit of a subsequence of $\mathbf{L}^n$. Finally, since SK iteration perseveres zero elements, $\mathbf{L}' \prec \mathbf{M}$. Together, we have $P \prec \mathbf{L}' \prec \mathbf{M}$. A similar argument holds for Q and $\mathbf{T}'$. Thus the lemma holds.

□

Remark A.2. Notice that, the choice of (P, Q) is free within the constraints (having partial pattern of $\mathbf{M}$ and being *SK stable*). In particular, such matrix pairs can be partially ordered with respect to their patterns, and (P, Q) can be selected such that they have the maximum possible pattern. Since the pattern of limit matrices must be intermediate between the pattern of (P, Q) and the pattern of $\mathbf{M}$, it follows that, all the pairs of limit matrices must have the same pattern, which must be the maximum possible.

Lemma A.3. *Any limit matrix of SK iteration on* $\mathbf{M}$ *is diagonally equivalent to* $\overline{\mathbf{M}}$.

Proof. Let $\mathbf{L}'$ be the limit of the sequence $X_n'\mathbf{M}Y_n'$ (where the $'$ signifies any sub-sequence of *SK iteration*). Then $\mathbf{L}'$ is also the limit of the sequence $X_n'\overline{\mathbf{M}}Y_n'$, since both $\overline{\mathbf{M}}$ and $\mathbf{L}'$ has the same pattern. In this case, Lemma 2 of Pretzel (1980) implies that there exist diagonal matrices X, Y such that $\mathbf{L}' = X\overline{\mathbf{M}}Y$, i.e. $\mathbf{L}'$ and $\overline{\mathbf{M}}$ are diagonally equivalent. □

Proposition 1 in Pretzel (1980) shows that:

Lemma A.4. *Let A and B be two matrices with the same row and columns sums. If there exists diagonal matrices X and Y such that $A = XBY$, then $A = B$.*

Proof of Proposition 10. We claim that: for a given $u \times v$ matrix $\mathbf{M}$, one may construct a *binary* matrix $A \prec \mathbf{M}$ such that up to permutation, A is block-wise diagonal of the form $\text{diag}(B_1, \ldots, B_k)$ where each B_i is a row or column vector of ones (i.e. $B_i = (1, \ldots, 1)$ or $B_i = (1, \ldots, 1)^T$). Let P, Q be the row and column normalizations of A respectively. It is straightforward to check that (P, Q) is *SK stable* and $P \prec \mathbf{M}$. Therefore, we only need to prove the claim.

We will prove the claim inductively on the dimension of $\mathbf{M}$. Let $n = \max\{u, v\}$. When $n = 1$, $\mathbf{M}$ is an 1×1 matrix and the claim holds. Now assume that the claim holds when $n \leq k - 1$, we will show that for a $u \times v$ matrix $\mathbf{M}$ with $\max\{u, v\} = k$, the claim still holds. Without loss, we may assume that $v = k$. There are now two cases.

Case 1 When $u < k$, let $\mathbf{M}'$ be the sub-matrix formed by the first $k-1$ columns of $\mathbf{M}$. Then $\mathbf{M}'$ is a $u \times (k-1)$ matrix. $\mathbf{M}$ has no zero columns implies that $\mathbf{M}'$ has no zero columns. (1) If $\mathbf{M}'$ contains no zero rows, according to the inductive assumption, there exists a binary matrix $A' \prec \mathbf{M}'$ having the desired form. Note that the last column of $\mathbf{M}$ contains a non-zero element m_{tk}. Let $\mathbf{v} = (v_1, \ldots, v_u)^T$ be the column vector with $v_i = 0$ if $i \neq t$ and $v_t = 1$. The desire $A \prec \mathbf{M}$ is then constructed using A' and $\mathbf{v}$ as following. The t-th row of A' must have a non-zero element a'_{ts} (as A' has no zero row). Denote the block contains a'_{ts} by B'. If B' is a row vector with all ones, then A is obtained by augmenting A' by $\mathbf{v}$, i.e. $A = [A', \mathbf{v}]$. Otherwise, we may replace $a'_{s,t}$ in A' by zero and denote the resulting matrix by A''. A is obtained by augmenting A'' by $\mathbf{v}$, i.e. $A = [A'', \mathbf{v}]$.

(2) If $\mathbf{M}'$ contains zero rows, let $\mathbf{M}^*$ be the matrix obtained from $\mathbf{M}'$ by omitting rows with indices in S_{zero} where S_{zero} is the index set of zero rows. Then there exists $A^* \prec \mathbf{M}^*$ according to the inductive assumption. Let $A' \prec \mathbf{M}'$ be the matrix obtained from A^* by inserting back the zero rows (at indices S_{zero}). Note that $\mathbf{M}$ contains no zero rows implies that $m_{ik} > 0$ for any $i \in S_{zero}$. Let $\mathbf{v} = (v_1, \ldots, v_u)^T$ be the column vector where $v_i = 1$ if $i \in S_{zero}$ and $v_i = 0$ otherwise. A is obtained by augmenting A' by $\mathbf{v}$, i.e. $A = [A', \mathbf{v}]$.

Case 2 When $u = k$, let $\mathbf{M}'$ be the sub-matrix formed by the first $k - 1$ rows $\mathbf{M}$. Then depending whether M' contains zero column, one may construct A as in case 1. In all circumstances, it is easy to check that the defined A has the desired format by construction. Hence claim also holds for any matrix $\mathbf{M}$ (or its transpose) of the form $u \times k$. □

Proof of Corollary 14. Let $(\mathbf{L}, \mathbf{T})$ and $(\bar{\mathbf{L}}, \bar{\mathbf{T}})$ be the limit of SK iteration on $\mathbf{M}$ and $\overline{\mathbf{M}}$ respectively. It is enough to show that $\mathbf{L} = \bar{\mathbf{L}}$. Lemma A.3 implies that both $\bar{\mathbf{L}}$ and $\mathbf{L}$ are diagonally equivalent to $\overline{\mathbf{M}}$. Therefore, $\mathbf{L}$ is diagonally equivalent to $\bar{\mathbf{L}}$. Further since both $\mathbf{L}$ and $\bar{\mathbf{L}}$ have the same pattern as $\overline{\mathbf{M}}$, Proposition 8 shows that they have the same row and column sums. Hence, Lemma A.4 implies that $\mathbf{L} = \bar{\mathbf{L}}$. □

Proof of Proposition 19. Since $\mathbf{M}$ and $\overline{\mathbf{M}}$ have exactly the same positive diagonals, we may assume that $\mathbf{M}$ has total support. Suppose that $\mathbf{M} \in \Phi^{-1}(\mathbf{L})$, i.e. $\Phi(\mathbf{M}) = \mathbf{L}$. Since $\mathbf{M}$ has total support, Sinkhorn and Knopp (1967b) implies that there exists diagonal matrices $X = \text{diag}(x_1, \ldots, x_n)$ and $Y = \text{diag}(y_1, \ldots, y_n)$ such that $\mathbf{M} = X\mathbf{L}Y$. In particular, $m_{ij} = x_i \times l_{ij} \times y_j$ holds, for any element m_{ij}. Let $D_1^{\mathbf{M}} = \{m_{i,\sigma(i)}\}$, $D_2^{\mathbf{M}} = \{m_{i,\sigma'(i)}\}$ be two positive diagonals of $\mathbf{M}$ and $D_1^{\mathbf{L}} = \{l_{i,\sigma(i)}\}$, $D_2^{\mathbf{L}} = \{l_{i,\sigma'(i)}\}$ be the corresponding positive diagonals in $\mathbf{L}$. Then:

$$\begin{aligned} CR(D_1^{\mathbf{M}}, D_2^{\mathbf{M}}) &= \frac{\Pi_{i=1}^n m_{i,\sigma(i)}}{\Pi_{i=1}^n m_{i,\sigma'(i)}} = \frac{\Pi_{i=1}^n x_i \times l_{i,\sigma(i)} \times y_{\sigma(i)}}{\Pi_{i=1}^n x_i \times l_{i,\sigma'(i)} \times y_{\sigma'(i)}} \\ &= \frac{\Pi_{i=1}^n x_i \times \Pi_{i=1}^n l_{i,\sigma(i)} \times \Pi_{i=1}^n y_{\sigma(i)}}{\Pi_{i=1}^n x_i \times \Pi_{i=1}^n l_{i,\sigma'(i)} \times \Pi_{i=1}^n y_{\sigma'(i)}} \\ &= \frac{\Pi_{i=1}^n l_{i,\sigma(i)}}{\Pi_{i=1}^n l_{i,\sigma'(i)}} = CR(D_1^{\mathbf{L}}, D_2^{\mathbf{L}}) \end{aligned} \tag{4}$$

We have established the 'if' direction. Now, for the 'only if' direction, suppose that $\mathbf{M} \overset{cr}{\sim} \mathbf{L}$. Let $\mathbf{M}^* = \Phi(\mathbf{M})$. Then $\mathbf{M}^* \in \mathcal{B}$ and $\mathbf{M} \in \Phi^{-1}(\mathbf{M}^*)$. According to Equation (4), $\mathbf{M}^* \overset{cr}{\sim} \mathbf{M}$ and so $\mathbf{M}^* \overset{cr}{\sim} \mathbf{L}$. Let $k = d_1^{\mathbf{M}^*}/d_1^{\mathbf{L}} > 0$, where $D_1^{\mathbf{M}^*}$ and $D_1^{\mathbf{L}}$ are positive diagonals determined by the same $\sigma \in S_n$. $\mathbf{M}^* \overset{cr}{\sim} \mathbf{L}$ implies that $\Pi_{i=1}^n m^*_{i,\alpha(i)} = k \times \Pi_{i=1}^n l_{i,\alpha(i)}$ for any $\alpha \in S_n$. Note that *distinct* doubly stochastic matrices do not have proportional corresponding diagonal products (For a proof see Sinkhorn and Knopp (1969)). Hence, $\mathbf{L} = \mathbf{M}^* = \Phi(\mathbf{M})$. $\square$

Theorem A.5 (Birkhoff-von Neumann theorem)**.** *(Dufossé and Uçar (2016)) For any $n \times n$ doubly stochastic matrix A, there exist $\theta_i \geq 0$ with $\sum_{i=1}^k \theta_i = 1$ and permutation matrices $\{P_1, \ldots, P_k\}$ such that $A = \sum_{i=1}^k \theta_i P_i$. This representation is also called Birkhoff-von Neumann (BvN) decomposition of A.*

Proof of Proposition 26. $\mathbf{M}^1$ has total support implies that there exist two diagonal matrices $X = \text{diag}\{x_1, \ldots, x_n\}$ and $Y = \text{diag}\{y_1, \ldots, y_n\}$ such that $\mathbf{M}^1 = X\mathbf{L}^1 Y$. Let $\mathbf{M}^2 = X\mathbf{L}^2 Y$ and $C = \max_{ij}\{x_i y_j\}$. Then $d(\mathbf{L}^1, \mathbf{L}^2) \leq \epsilon \implies |l^1_{ij} - l^2_{ij}| \leq \epsilon \implies |x_i \cdot l^1_{ij} \cdot y_j - x_i \cdot l^2_{ij} \cdot y_j| \leq C\epsilon \implies |m^1_{ij} - m^2_{ij}| \leq C\epsilon$. Thus, $d(\mathbf{M}^1, \mathbf{M}^2) \leq C\epsilon$. $\square$

Construction of Homeomorphic Φ.

As mentioned above, for any $\mathbf{M} \in \overline{\mathcal{A}}$, there exist two diagonal matrices X and Y such that $\mathbf{M} = X\Phi(\mathbf{M})Y$. Note the choice of X and Y is unique only up to a scalar. This can be made deterministic by requiring the last positive element of Y to be 1, i.e. $y_n = 1$. In this way Φ can be viewed as a map : $\overline{\mathcal{A}} \to \mathcal{R}_+^{2n-1} \times \mathcal{B}$ where $\mathbf{M} \mapsto [(x_1, \ldots, x_n, y_1, \ldots, y_{n-1}), \Phi(\mathbf{M})]$. Tverberg (1976) showed that:

Proposition A.6. *$\Phi : \overline{\mathcal{A}} \to \mathcal{R}_+^{2n-1} \times \mathcal{B}$ is continuous, invertible and the inverse is also continuous. Thus Φ is homeomorpic.*

Role of zeros: Results derived in Section 3-4 do not depend on zero elements. The lower bounds of CI derived in Section 5 are closely related to the amount of zeros and their locations. However, together with the sensitivity analysis, these bounds can still be used as an approximation for lower bounds of CI for matrices with very small elements (instead of exact zeros). In some models, zero elements could appear. For instance, in linguistic applications, if an utterance is not consistent with a referent, its corresponding element would be zero (Golland et al., 2010). More generally, it is an interesting question as to whether more models should assign zero probability to some possible outcomes. As noted, most probabilistic models do not currently assign zero probability to any outcomes. However, if one, for example, wants models that are explainable via examples our results show that assigning zero probability to some outcomes is a desirable feature. Beyond explainability, as far as we know, there are no principled reason for not assigning zero probability to some possible outcomes. Finally, having zeros reduces the number of positive diagonals, which is a special case of the more general problem of establishing bounds based on cross-ratio (Corollary 20). This is considerably more challenging and a direction for future work.

Other connections. Sinkhorn iteration finds its way in many other applications in variety of fields. To name a few: transportation planning to predict flow in a traffic network (Fienberg et al., 1970), contingency table analysis which has many uses in biology, economics etc. (Fienberg et al., 1970), decreasing condition numbers which is of importance in numerical analysis (Osborne, 1960). Moreover, there are many algorithms implemented as generalizations of Sinkhorn matrix balancing to solve problems such as Edmonds problem (Gurvits, 2003), Sudoku Solvers (Moon et al., 2009) and web page ranking algorithms (Knight, 2008). More applications and a comprehensive discussion can be found in (Idel, 2016) and references therein.