Correlation of Gene Function Annotation Lists through Enhanced Spearman and Kendall Measures

Davide Chicco(1)(2), Eleonora Ciceri(1), Marco Masseroli(1)

(1) Dipartimento di Elettronica Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo Da Vinci 32, 20133 Milan, Italy
(2) Institute for Genomics and Bioinformatics, Information and Computer Science Building 2, University of California Irvine, 92697-3445 Irvine, CA, USA

Email: davide.chicco@gmail.com, eleonora.ciceri@mail.polimi.it, masseroli@elet.polimi.it

Keywords: Spearman coefficient, Kendall distance, top-K queries, Gene Ontology.

Abstract. Gene function annotations are paramount in bioinformatics, and computational methods able to predict them provide a fundamental contribution. Several machine learning algorithms for this purpose are available today, although their relevant parameters might strongly influence the output list of predicted annotations. Here, we propose a method to evaluate this issue by introducing two list correlation measures, based on the Spearman rank correlation coefficient and Kendall rank distance respectively, which are able to state the level of similarity between ordered annotation lists. We show the application of these measures to Gene Ontology annotation datasets, which leads to unveil interesting patterns between predicted annotation lists and express some statements about the prediction algorithms used.

1 Scientific background

In bioinformatics, a controlled gene function annotation is the association of a gene with a controlled term that represents a functional feature; such term can be part of a terminology, or structured in an ontology such as the Gene Ontology (GO) [1]. Thus, the annotation states that the gene has the functional feature represented by the annotation term. For instance, <VMA9, transmembrane transport> is the annotation of the gene VMA9 to the biological function transmembrane transport. Despite their biological significance, some issues concern available annotations (e.g. erroneous or missing annotations) [2]. Thus, computational methods and software tools able to produce ranked lists of reliably predicted annotations are an excellent contribution to the field.

In the past, we designed and developed some algorithms in this field. We started from a state-of-the-art algorithm based on truncated Singular Value Decomposition (tSVD) [3] and developed some variants [4]. Then, in [5] we designed an algorithm to choose the best truncation level for the tSVD and in [6] we designed and tested some topic modeling techniques. All these methods involve key parameters that influence the output. To understand how the resulting annotation lists change accordingly to variations of these key parameters, a similarity measure that compares different output lists is required. To accomplish this, currently the most useful and consistent measures are the Spearman rank correlation coefficient [7] and Kendall rank distance [8].

Here, we depart from a recent work by Ciceri et al. [9] to develop new weighted correlation metrics that compare ordered annotation lists. The remainder of this paper is organized as follows. Section 2 explains the prediction of gene function annotations and introduces the Spearman and Kendall measure variants that we developed for the comparison of ordered annotation lists. The section 3 shows some significant test results of the proposed measure variants and discusses them, while Section 4 concludes.
2 Material and methods

Let \( A = [a_{ij}] \) be a \( m \times n \) matrix, where each row \( i \) corresponds to a gene and each column \( j \) corresponds to a functional feature term of a terminology or ontology, with \( a_{ij} = 1 \) if the gene \( i \) is annotated to the feature term \( j \), or \( a_{ij} = 0 \) otherwise. Let \( \theta \) be a fixed threshold value and suppose that a prediction algorithm elaborates the matrix \( A \) to produce an output matrix \( \tilde{A} \), with the same dimensions of \( A \), where each value \( \tilde{a}_{ij} \) represents the likelihood of the annotation of the gene \( i \) to the feature \( j \). Thus, a high \( \tilde{a}_{ij} \) value indicates that the probability of the gene \( i \) to be associated with the feature \( j \) is high. An annotation list, ordered according to the \( \tilde{a}_{ij} \) values, is finally defined and each annotation \( \langle \text{gene}_i, \text{feature}_j, \tilde{a}_{ij} \rangle \) is classified in one of the following categories:

- **Annotation Confirmed (AC):** \( a_{ij} = 1 \land \tilde{a}_{ij} > \theta \) (similar to True Positive - TP)
- **Annotation Predicted (AP):** \( a_{ij} = 0 \land \tilde{a}_{ij} > \theta \) (similar to False Positive - FP)
- **Non-Annotation Confirmed (NAC):** \( a_{ij} = 0 \land \tilde{a}_{ij} \leq \theta \) (similar to True Negative - TN)
- **Annotation to be Reviewed (AR):** \( a_{ij} = 1 \land \tilde{a}_{ij} \leq \theta \) (similar to False Negative - FN)

According to these categories, two annotation sublists are defined: an APlist, i.e. Annotation Predicted list, and a NAClist, i.e. Non-Annotation Confirmed list, which contain those annotations from the original list that were classified in the AP or NAC class, respectively. Furthermore, these four categories are used to build the Receiver Operating Characteristic (ROC) curve, which is a graphical plot depicting the performance of a binary classifier system for different discrimination threshold values. Similarly to its original definition, which uses TPrate and FPrate, our ROC curve depicts the trade-off between the ACrate and APrate, where:

\[
ACrate = \frac{AC}{AC + AR} \quad APrate = \frac{AP}{AP + NAC}
\]

for all possible values of \( \theta \). Notice that, in statistical terms, \( ACrate = Sensitivity \) and \( APrate = 1 - Specificity \).

Each prediction algorithm has parameters that, when changed, lead to different output annotation lists. For instance, tSVD may produce quite different APlists when its truncation level \( k \) varies. Thus, to understand how the selected parameter values influence the output lists, it is important to define similarity metrics that compare annotation lists resulting from different algorithm parameterizations. Two similarity measures for ordered predicted annotation list comparison are here introduced as follows.

### 2.1 Extended Spearman rank correlation coefficient

The **Spearman rank correlation coefficient** \( (\rho) \) measures the statistical dependence of two variables \( X \) and \( Y \). The measure expresses either positive correlation, i.e. \( Y \) increases (decreases) when \( X \) increases (decreases), or negative correlation, i.e. \( Y \) decreases (increases) when \( X \) increases (decreases). A similar definition can be applied to pairs of lists \( l_a \) and \( l_b \) of ranked elements. Suppose \( l_a \) and \( l_b \) have the same length (number of elements) \( n \) and are composed of the same elements. Given an element \( i \), let \( x_i \) denote its position in \( l_a \), \( y_i \) its position in \( l_b \) and \( d_i = |x_i - y_i| \) the distance of its positions in \( l_a \) and \( l_b \). The final normalized Spearman \( \rho \) value is computed as: 

\[
\rho = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)} \]

A maximum positive correlation \( \rho = +1 \) occurs when \( l_a \) and \( l_b \) are identical (i.e. with the same elements in the same order), while the maximum negative correlation \( \rho = -1 \) occurs when \( l_a \) and \( l_b \) contain the same elements but in inverse order. The minimum correlation \( \rho = 0 \) (i.e. the maximum diversity in the element order) arises when the element order in \( l_a \) and \( l_b \) strongly diverges.

However, when two lists \( l_a \) and \( l_b \) have different length and/or contain different elements, they are not properly handled by the classical Spearman rank correlation coefficient. To manage these cases, based on Cicieri et al. \[9\], we introduce a new **Weighted Spearman rank correlation coefficient** \( (\rho_w) \), which features penalty distance weights \( w_{si} \) for each element \( i \) absent from one of the two lists.
Let \( q = |l_a \cup l_b| \). Thus, the penalty weight \( w_{st} \) for an element \( i \) in list \( l_a \) and/or \( l_b \) is computed as follows, by penalizing elements present in only one of the two lists:

\[
w_{st} = \begin{cases} 
1 - \frac{1}{|x_i - y_i| + 1} & \{ i \in l_a \wedge i \in l_b \} \\
1 & \text{otherwise}
\end{cases}
\]  

(2)

The Weighted Spearman \( \rho_w \) value is then computed as: \( \rho_w = (\sum_{i=1}^{q} w_{st}) / q \). High correlation is found when \( \rho_w \approx 0 \) (i.e. very few penalties are assigned), while low correlation is found when \( \rho_w \approx 1 \) (i.e. many penalties are assigned). If the two lists are equal, \( \rho_w = 0 \), otherwise, if they have no common elements, i.e. \( q = |l_a| + |l_b| \), \( \rho_w = 1 \).

Yet, the Weighted Spearman rank correlation coefficient has a flaw: all elements \( \{ i : i \notin l_a \lor i \notin l_b \} \) are weighted equally. Let suppose to have other two lists \( l_{a2} \) and \( l_{b2} \), and that, if an element \( i \) is not present in \( l_a (l_b) \), it could be present, or not, in \( l_{a2} (l_{b2}) \). For example, in our context (e.g. \( l_a = A\text{P}l\text{i}st_1, l_b = A\text{P}l\text{i}st_2, l_{a2} = N\text{A}C\text{l}ist_1 \) and \( l_{b2} = N\text{A}C\text{l}ist_2 \)), if an annotation is not present in an \( A\text{P}l\text{i}st_1 \) (\( A\text{P}l\text{i}st_2 \)), it could be present, or not, in the related \( N\text{A}C\text{l}ist_1 \) (\( N\text{A}C\text{l}ist_2 \)). If both cases, e.g. \( \{ i' : i' \notin l_a, i' \notin l_{a2} \} \) and \( \{ i'' : i'' \notin l_a, i'' \notin l_{a2} \} \), the penalty weight would be equally maximum (i.e. \( w_{st} = w_{st''} = 1 \)). Thus, we defined a new penalty weight \( v_i \), where \( i' \) gets a lower penalty than \( i'' \). To do so, we first defined the penalized position of each element \( i \) in \( l_{a2} (l_{b2}) \), with respect to the length of the related list \( l_a (l_b) \), as: \( \hat{z}_i = z_i + n \cdot 2 \), where \( z_i \) denotes the position of \( i \) in \( l_{a2} (l_{b2}) \), \( n \) is the length of the related \( l_a (l_b) \) and 2 is a penalty factor for \( i \) not to be in \( l_a (l_b) \) but in \( l_{a2} (l_{b2}) \). Then, we expressed the new penalty weight as follows:

\[
v_{si} = \begin{cases} 
1 - \frac{1}{|x_i - y_i| + 1} & \{ i \in l_a, i \in l_b \} \\
1 - \frac{1}{|x_i - \hat{z}_i| + 1} & \{ i \notin l_a, i \notin l_{a2}, i \notin l_b \} \lor \{ i \notin l_a, i \notin l_{a2}, i \notin l_{b2} \} \\
1 & \{ i \notin l_a, i \notin l_{a2}, i \notin l_b, i \notin l_{b2} \}
\end{cases}
\]  

(3)

where \( x_i \) is the \( i \) element position in \( l_a \), \( y_i \) is its position in \( l_b \) and \( \hat{z}_i \) is its penalized position in \( l_{a2} (l_{b2}) \). A new Extended Spearman rank correlation coefficient (\( \rho_e \)) is then computed as: \( \rho_e = (\sum_{i=1}^{q} v_{si}) / q \). As for the Weighted Spearman rank correlation coefficient, high \( \rho_e \) values lead to low correlation, while \( \rho_e \approx 0 \) suggests high correlation.

### 2.2 Extended Kendall rank distance

The Kendall rank distance \( (\tau) \) [8] is related to the number of pairwise disagreements between two lists \( l_a \) and \( l_b \) of ranked elements, i.e. the number of bubble-sort swaps needed to sort the two lists in the same order. Obviously, when the two lists are identical, \( \tau = 0 \). On the other hand, if the two lists are equal but in reverse order, then \( \tau = 1 \). Let \( l_a \) and \( l_b \) be two lists of length \( n \). Given an element \( i \), let \( x_i \) be its position in \( l_a \) and \( y_i \) its position in \( l_b \). Thus, the set \( K \) of required swaps between list elements is computed as:

\[
K(l_a, l_b) = \{ (i, j) : (x_i < y_i \land x_j > y_j) \lor (x_i > y_i \land x_j < y_j) \}
\]  

(4)

Then, the normalized Kendall rank distance \( \tau \) is given by: \( \tau = |K| / (n \cdot (n - 1) / 2) \). Notice that the Kendall rank distance \( \tau \) does not express negative correlation between lists. Moreover, while the Spearman rank correlation coefficient \( \rho \) is focused on the distance between the ranks of each element in the two lists, the Kendall rank distance \( \tau \) considers just the number of swaps in the element rank.

Similarly to what concerns the Spearman rank correlation coefficient, a flaw of the classical normalized Kendall rank distance is its usability only when the two lists have the same length and the same elements. To avoid this limitation, we introduce weights to consider, but with a penalty, the elements that are present in one list but absent from the other [9]. The weight \( w_{ki} \) for each swapped element \( i \) in a list is defined as:

\[
w_{ki} = \begin{cases} 
\frac{1}{\log(x_i + 2) - \log(x_i + 3)} & \{ i \in l_a \land i \in l_b \} \\
0.5 & \text{otherwise}
\end{cases}
\]  

(5)
Again, similarly to what concerns the Weighted Spearman rank correlation coefficient, not included in an additional related list available (i.e. $l$ values, whereas a penalty weight $v$ is used to define $\tau$ positions more important. The Weighted Kendall rank distance $\tau_w$ is thus computed as:

$$\tau_w = \frac{\sum_{(i,j) \in K(l_a,l_b)} w_{ki} w_{kj}}{|K|} \quad (6)$$

As for the classical Kendall rank distance $\tau$, low correlations correspond to high $\tau_w$ values, whereas $\tau_w = 0$ if the two lists are identical (since no swap occurs).

Again, similarly to what concerns the Weighted Spearman rank correlation coefficient, all elements $\{i : i \notin l_a \land i \notin l_b\}$ are weighted equally, independently if they are, or are not, included in an additional related list available (i.e. $l_{a2}$ or $l_{b2}$). Thus, we also define a penalty weight $v_{ki}$ for the swapped elements as:

$$v_{ki} = \begin{cases} 
\frac{1}{\log(x_i+2)} - \frac{1}{\log(x_i+3)} & \{i \in l_a, i \in l_b\} \\
0.5 - h & \{i \notin l_a, i \notin l_{a2}, i \notin l_{b2}\} \lor \{i \in l_a, i \notin l_{b2}, i \in l_{b2}\} \\
0.5 & \{i \in l_a, i \notin l_{a2}, i \in l_{b2}\} \lor \{i \in l_{a2}, i \notin l_{b2}, i \notin l_{b2}\}
\end{cases} \quad (7)$$

where, when an element $i$ missed from $l_a$ ($l_b$) is present in a related $l_{a2}$ ($l_{b2}$) list (e.g. $l_a = A\text{Plist}_1$, $l_b = A\text{Plist}_2$, $l_{a2} = N\text{AClist}_1$ and $l_{b2} = N\text{AClist}_2$), $i$ gets a lower penalty weight, i.e. $v_{ki} = 0.5 - h$. In particular, the penalty reduction $h$ can be defined as: $h = 0.5 - z_i/(m \cdot 2)$, where $z_i$ is the position of $i$ in $l_{a2}$ ($l_{b2}$) and $m$ is the length of $l_{a2}$ ($l_{b2}$). Alternatively, if the elements in $l_{a2}$ ($l_{b2}$) have an associated likelihood value (e.g. the $\tilde{a}_{ij}$ likelihood value in a $N\text{AClist}$), $h$ can be set equal to such value. (Notice that the annotations $\tilde{a}_{ij}$ have likelihood value in the interval $[\theta, 1]$ in an $A\text{Plist}$, while in the interval $[0, \theta]$ in a $N\text{AClist}$, in both cases ranked from maximum to minimum value; thus, the likelihood value $h$ used to define $v_{ki}$ decreases along the rank). We chose the weight function in Equation $[7]$ so that is decreases when the element gets a lower rank. Thus, we introduce the Extended Kendall rank distance $\tau_e$, where an element $i$ may be absent from $l_a$ ($l_b$) but present in its related list $l_{a2}$ ($l_{b2}$), as:

$$\tau_e = \sum_{(i,j) \in K(l_a,l_b)} v_{ki} v_{kj} \quad (8)$$

As for the Weighted Kendall rank distance, $\tau_e$ is high when $l_a$ and $l_b$ are very different, whereas $\tau_e \approx 0$ when the two lists $l_a$ and $l_b$ are very similar.

3 Results

We evaluated the extended $\rho_e$ and $\tau_e$ measures on annotation lists generated using the tSVD prediction method, while varying the truncation level $k$. The tSVD was applied on the following datasets (downloaded from Genomic and Proteomic Data Warehouse [10]): 1. *Homo sapiens* gene annotations to GO Cellular Component (CC) terms (#genes: 7,868, #terms: 684, #annotations: 14,381); 2. *Bos taurus* (cattle) gene annotations to GO Molecular Function (MF) terms (#genes: 543, #terms: 422, #annotations: 934); 3. *Gallus gallus* MF terms (#genes: 309, #terms: 225, #annotations: 509).

3.1 tSVD truncation trends

Table[1] shows the AP and NAC amounts of *Homo sapiens* gene CC annotations generated by the tSVD for different truncation levels $k$. The listed $k$ values are the ones progressively evaluated by our best truncation level algorithm [5], which chose $k = 402$ as the best value; with this $k$ value, the tSVD algorithm produces List0. Table[1] shows that small changes in the truncation level $k$ may lead to quite different AP and NAC amounts and lists. Yet, by sorting such lists on the basis of their truncation level $k$ and comparing them through the defined $\rho_e$ and $\tau_e$ measures, interesting results appear (Table[2]).
Proceedings of CIBB 2014

Table 1: AP and NAC amounts of Homo sapiens, Gallus gallus and Bos taurus gene GO annotations generated by tSVD when varying the truncation level \( k \), and their ROC AUC percentage. The likelihood threshold \( \theta \) was set to \( \theta = 0.49 \) for Homo sapiens annotations and \( \theta = 0.50 \) for the other annotations.

<table>
<thead>
<tr>
<th>k</th>
<th>AP</th>
<th>NAC</th>
<th>AUC%</th>
<th>k</th>
<th>AP</th>
<th>NAC</th>
<th>AUC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>List0 378</td>
<td>8</td>
<td>4,458,751</td>
<td>83.49</td>
<td>List5 233</td>
<td>48</td>
<td>4,458,711</td>
<td>51.82</td>
</tr>
<tr>
<td>List1 402</td>
<td>2</td>
<td>4,458,757</td>
<td>53.64</td>
<td>List6 175</td>
<td>78</td>
<td>4,458,681</td>
<td>48.80</td>
</tr>
<tr>
<td>List2 390</td>
<td>7</td>
<td>4,458,752</td>
<td>53.58</td>
<td>List7 117</td>
<td>86</td>
<td>4,458,673</td>
<td>45.11</td>
</tr>
<tr>
<td>List3 291</td>
<td>19</td>
<td>4,458,740</td>
<td>53.14</td>
<td>List8 59</td>
<td>95</td>
<td>4,458,664</td>
<td>38.82</td>
</tr>
<tr>
<td>List4 349</td>
<td>8</td>
<td>4,458,751</td>
<td>52.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

While in this case the Extended Spearman rank correlation coefficient \( \rho_e \) gives no hints about specific trends, the Extended Kendall rank distance \( \tau_e \) shows that the higher the difference between tSVD truncation levels is, the higher the Extended Kendall rank distance \( \tau_e \) between the generated lists is, and the less similar the lists are. Except for the comparison \((List1, List2)\), all other list comparisons show \( \tau_e \) value trends that increase when the distance between tSVD truncation level values increase. For example, List8 shows almost the maximum \( \tau_e \) dissimilarity values from all the other lists compared.

Table 2: Extended Spearman rank correlation coefficient \( \rho_e \) values (gray cells) and Extended Kendall (E. Kendall) rank distance \( \tau_e \) values (white cells) from the comparison of AP lists obtained for the Homo sapiens gene CC annotation dataset using tSVD with different truncation level \( k \) values. MaxCorrelation = 0 and MinCorrelation = 1 for both \( \rho_e \) and \( \tau_e \).

<table>
<thead>
<tr>
<th></th>
<th>List1</th>
<th>List2</th>
<th>List3</th>
<th>List4</th>
<th>List5</th>
<th>List6</th>
<th>List7</th>
<th>List8</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>0.129</td>
<td>0.621</td>
<td>0.558</td>
<td>0.894</td>
<td>0.962</td>
<td>0.995</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>K</td>
<td>0.07</td>
<td>0.224</td>
<td>0.245</td>
<td>0.819</td>
<td>0.512</td>
<td>0.495</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>e</td>
<td>0.512</td>
<td>0.491</td>
<td>0.724</td>
<td>0.570</td>
<td>0.535</td>
<td>0.995</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>n</td>
<td>0.535</td>
<td>0.324</td>
<td>0.656</td>
<td>0.841</td>
<td>0.708</td>
<td>0.495</td>
<td>0.998</td>
<td>0.997</td>
</tr>
<tr>
<td>d</td>
<td>0.708</td>
<td>0.656</td>
<td>0.872</td>
<td>0.841</td>
<td>0.838</td>
<td>0.995</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>a</td>
<td>0.838</td>
<td>0.872</td>
<td>0.841</td>
<td>0.302</td>
<td>0.306</td>
<td>0.283</td>
<td>0.998</td>
<td>0.996</td>
</tr>
<tr>
<td>l</td>
<td>0.535</td>
<td>0.324</td>
<td>0.656</td>
<td>0.841</td>
<td>0.838</td>
<td>0.995</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>l</td>
<td>0.838</td>
<td>0.872</td>
<td>0.841</td>
<td>0.302</td>
<td>0.306</td>
<td>0.283</td>
<td>0.998</td>
<td>0.996</td>
</tr>
</tbody>
</table>

3.2 ROC AUC trends

Our tSVD method uses the Receiver Operating Curve (ROC) and maximizes the area under the curve (AUC) to select the best prediction parameter values. Yet, while the tSVD truncation level \( k \) varies, these AUC percentages may have an oscillatory trend and the AP lists vary. Thus, it is relevant to evaluate the correlation between ROC AUC percentages and the generated AP list similarity; this can be performed by comparing the AUC percentage values with the values of the Extended Spearman rank correlation coefficient \( \rho_e \) and Extended Kendall rank distance \( \tau_e \) between the generated AP lists. By doing so, we found interesting trends. As general examples, in Table 1 and Table 3 we show these values for the Gallus gallus and Bos taurus gene MF annotations. As one may notice, in these cases the Extended Spearman \( \rho_e \) metric shows that AP lists
generated by predictions with similar AUC percentages (e.g. List0, List1 and List2), have similar low $\rho_e$ values; thus, these lists mostly contain the same annotations with similar ranking. This proves that our method to select the best prediction parameter values, based on the optimization of the ROC AUC percentages, is quite robust.

### Table 3: Extended Spearman rank correlation coefficient $\rho_e$ values (gray cells) and Extended Kendall (Ext. K.) rank distance $\tau_e$ values (white cells) from the comparison of AP lists obtained for the *Gallus gallus* and *BosTaurus* gene MF annotation datasets using tSVD with different truncation level $k$ values. Low $\rho_e$ and $\tau_e$ values showing high correlation for lists with similar AUC percentages are in bold.

<table>
<thead>
<tr>
<th></th>
<th>List0</th>
<th>List1</th>
<th>List2</th>
<th>List3</th>
<th>List4</th>
<th>List0</th>
<th>List1</th>
<th>List2</th>
<th>List3</th>
<th>List4</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>0.370</td>
<td>0.200</td>
<td>0.620</td>
<td>0.850</td>
<td>0.192</td>
<td>0.166</td>
<td>0.364</td>
<td>0.794</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>0.790</td>
<td>0.330</td>
<td>0.780</td>
<td>0.940</td>
<td>0.687</td>
<td>0.279</td>
<td>0.425</td>
<td>0.888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>0.640</td>
<td>0.790</td>
<td>0.180</td>
<td>0.740</td>
<td>0.812</td>
<td>0.822</td>
<td>0.128</td>
<td>0.772</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>0.980</td>
<td>0.960</td>
<td>0.950</td>
<td>0.780</td>
<td>0.980</td>
<td>0.980</td>
<td>0.950</td>
<td>0.773</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4 Conclusion

By focusing on either element position difference (Spearman) or number of elements with different positions (Kendall), our defined Extended Spearman rank correlation coefficient $\rho_e$ and Extended Kendall rank distance $\tau_e$ metrics resulted effective in computing the similarity level between two lists of elements, although the compared lists include (partially) different elements. Results show that, when used for the evaluation of gene function annotation lists, the proposed metrics reveal similarity levels between predicted annotation lists which correlate with prediction key parameter variations.

**Acknowledgments**

This work is partially funded by CUbRIK project (www.CubrikProject.eu).

**References**


