

## Supplementary methods

We want to compute a substitutability score such as:

1. Two items are highly substitutable if they are consumed in similar contexts.
2. Two items are less substitutable if they are consumed together (association).
3. Substitutability is a symmetrical relationship.

Let us denote, for an item  $x$ , the context set  $C_x$  as the set of dietary contexts in which  $x$  is a consumed item. If the number of items in  $C_x$  (denoted as  $|C_x|$ ) is high, then  $x$  will be substitutable in many dietary contexts.

For two items  $x$  and  $y$ , the condition '1' is described by the intersection of  $C_x$  and  $C_y$ , (noted as  $C_x \cap C_y$ , *i.e.* the dietary contexts in which  $x$  and  $y$  appears). If  $|C_x \cap C_y|$  (*i.e.* the number of dietary contexts in which  $x$  and  $y$  appears) is high, then  $x$  and  $y$  are consumed in similar contexts. As  $|C_x \cap C_y|$  will inevitably be higher if  $x$  or  $y$  are frequently consumed items,  $|C_x \cap C_y|$  needs to be corrected by  $|C_x \cup C_y|$  (*i.e.* the number of dietary contexts in which  $x$  or  $y$  appears) to provide a first estimation of substitutability between  $x$  and  $y$ .

We denote  $A_{x:y}$  the set of contexts of  $x$  where  $y$  appears. The number of items in  $A_{x:y}$  denotes how  $y$  is associated to  $x$ . Condition '2' also implies that the substitutability score of  $x$  by  $y$  must be corrected by  $|A_{x:y}| + |A_{y:x}|$ . We thus propose the substitutability score inspired by the Jaccard index (24):

$$f(x, y) = \frac{|C_x \cap C_y|}{|C_x \cup C_y| + |A_{x:y}| + |A_{y:x}|}$$

This score equals 1 when  $x$  and  $y$  appear in the same contexts and  $|A_{x:y}| = |A_{y:x}| = |\emptyset|$ . If  $x$  and  $y$  are never consumed in the same context then the score equals 0. The higher  $|A_{x:y}| + |A_{y:x}|$  is, the higher the association of  $x$  and  $y$  is and the lower the score is. It is important to note that, in general,  $|A_{x:y}| \neq |A_{y:x}|$ . For instance let's consider *bread* and *jam*, overall, *bread* is rarely associated with *jam* as a consequence  $|A_{bread:jam}|$  is quite small, whereas since *jam* is often associated with *bread*,  $|A_{jam:bread}|$  is high.

Here is a simple example of calculation of a substitutability score; let's consider the 6 following meals:

Meal 1: (*pasta, tomato sauce, chicken*)

Meal 2: (*pasta, tomato sauce, meatballs*)

Meal 3: (*pasta, pesto, meatballs*)

Meal 4: (*pasta, pesto, chicken*)

Meal 5: (*rice, chicken*)

Meal 6: (*chips, meatballs*)

We then obtain the following sets:

$$C_{chicken} = \langle \text{pasta, tomato sauce} \rangle; \langle \text{pasta, pesto} \rangle; \langle \text{rice} \rangle$$

$$C_{meatballs} = \langle \text{pasta, tomato sauce} \rangle; \langle \text{pasta, pesto} \rangle; \langle \text{chips} \rangle$$

$$|C_{chicken} \cap C_{meatballs}| = 2$$

$$|C_{chicken} \cup C_{meatballs}| = 4$$

$$|A_{chicken:meatballs}| = |\emptyset| = 0$$

$$|A_{meatballs:chicken}| = |\emptyset| = 0$$

$$f(chicken, meatballs) = \frac{|C_{chicken} \cap C_{meatballs}|}{|C_{chicken} \cup C_{meatballs}| + |A_{chicken:meatballs}| + |A_{meatballs:chicken}|} = 0.5$$