Let us see how the value of the relation of the neighbourhoods Δ_r and Δ_s alters on account of their common duplicates. The following definition is introduced to make precise how much the relation increases because of one of their common "complete" duplicates in X, i.e., such a connected piece in X which is not longer than the relating neighbourhood, containing one copy of each name from Δ_r and Δ_s (taking the multiplicities into account). If we represent Δ_r and Δ_s as two chronicles having a common inverse image from Y, then the complete duplicate is the complete chronicle combining the names of these two.

Definition 8. We call the number

$$E_0\big(\Delta_r(k), \Delta_s(k)\big) = \frac{c}{(2k+1)^2} \sum_{i=r-k}^{r+k} \sum_{j=s-k}^{s+k} \frac{1-\delta_{ij}}{c(a_i, a_j)},$$

where $\delta_{ij} = 1$ if $i = j$, and $= 0$ otherwise, and $c(a_i, a_j)$ was defined in (4), c is the multiplier from (5), *the proper relations unit for two defining neighbourhoods.*

Let X contain duplicates. We call two determining neighbourhoods *independent* if they have no common duplicates and are non-intersecting in X. We call the remaining neighbourhood pairs *dependent.* We assume for simplicity that there are few duplicates, so that the relation between two independent neighbourhoods is similar to the correct list.

Consider the three followings cases, viz.:

(1) The neighbourhoods Δ_r and Δ_s are independent. Then the mean value of their relation equals $c \cdot \alpha$.

(2) The neighbourhoods Δ_r and Δ_s coincide, with Δ_r having no duplicates. The mean value of the relation in this case equals $c\alpha + E_0(\Delta_r, \Delta_s)$, and the neighbourhood is its complete duplicate.

(3) Two non-intersecting neighbourhoods Δ_r and Δ_s in X possess j common complete duplicates. The mean value of their relation is then equal to $c\alpha + j \cdot E_0(\Delta_r, \Delta_s)$.

We have to separate cases (1) and (3), for which we shall try to find the optimal radius of the determining neighbourhood (radius k). Note that, by increasing k we decrease the scattering with respect to α (the variance of the relation L_0), which increases the precision of the separation. However, for too large k, the duplicate completion degree lessens, thus leading to actually decreasing the factor j in (3). The value of k must not exceed the typical length of the elementary chronicle Z_i (see Item 10). The optimal value is chosen from experience.

Remark. Since the system $\{\Delta_{k+1}, \Delta_{k+2}, \ldots, \Delta_{N-k-1}\}$ is that of "current" neighbourhoods in the list X, less pure neighbourhood duplicates than the "precise" one are neighbouring to it. To distinguish the most complete duplicates, we will only retain local maxima in the relation matrix $a_{rs} = L_0(\Delta_r, \Delta_s)$ and consider the relation $L_0(\Delta_r, \Delta_s)$ only in the case when $L_0(\Delta_r, \Delta_s) \geqslant L_0(\Delta_r, \Delta_{s-1}) - \varepsilon$, and $L_0(\Delta_r, \Delta_s) \geqslant L_0(\Delta_r, \Delta_{s+1}) - \varepsilon$, or else it is replaced by zero. This remark does not concern the construction of the frequency histograms (see below). The value of ε was chosen to be equal to the length of the interval to be divided in constructing the frequency histogram (see Item 13).