directly on account of the common names, since the mechanism leading to such an increase does exist. However, this is not so. For demonstration, we introduce another two relation measures. Consider a neighbourhood pair $\Delta_r(k)$ and $\Delta_s(k)$, and let

$$\Delta_r \supset \Delta_r' = \{\text{set of entries of } \Delta_r \text{ with different names}\},$$

$$\Delta_s \supset \Delta_s' = \{\text{set of entries of } \Delta_s \text{ with different names}\},$$

$$\Delta_r' \supset \Delta_{r,s}'' = \{\text{set of entries of } \Delta_r' \text{ whose names are known to be coincident with those from } \Delta_s'\}.$$

Thus the neighbourhoods Δ_r' and Δ_s' contain one representative of each name; besides, Δ_s' and $\Delta_r' \setminus \Delta_{r,s}''$ contain no common names. Denote the length (number of terms) of a neighbourhood by $|.|$. By definition, we put

$$L_1(\Delta_r, \Delta_s) = \frac{c}{|\Delta_r'|\,|\Delta_s'|} \sum_{a \in \Delta_r', b \in \Delta_s'} l(a, b), \tag{6}$$

$$L_2(\Delta_r, \Delta_s) = \frac{c}{|\Delta_{r,s}''|\,|\Delta_s'|} \sum_{a \in \Delta_{r,s}'', b \in \Delta_s'} l(a, b); \tag{7}$$

it is easy to verify that $L_2(\Delta_r, \Delta_s) = L_2(\Delta_s, \Delta_r)$.

The quantity $L_2(\Delta_r, \Delta_s)$ in no way relates to the common names in Δ_r and Δ_s; they are not involved in its definition. However, the frequency histogram for $L_2(\Delta_r, \Delta_s)$ for the lists P and N and fixed value of $0(\Delta_r, \Delta_s)$ show that the dependence of L_2 on $0(\Delta_r, \Delta_s)$ is the same as that of L_0 on $0(\Delta_r, \Delta_s)$. The same is valid for $L_1(\Delta_r, \Delta_s)$ which signifies that a certain common factor leading to their statistical dependence is at the foundation of two outwardly unrelated values $L_2(\Delta_r, \Delta_s)$ and $0(\Delta_r, \Delta_s)$. It is clear that the availability of common duplicates is a factor of this kind. Hence, the discovered dependence supports the hypothesis regarding the existence of duplicates in P and N.

The relation matrices for P and N constructed by means of L_0, L_1 or L_2, respectively, turned out to lead to the same conclusion, i.e. distinguishing the same duplicate systems. Therefore, we shall sometimes write simply $L(\Delta_r, \Delta_s)$ meaning one of the three relations L_0, L_1, or L_2.

Note the difference between the relation matrices constructed by means of $L(\Delta_r, \Delta_s)$, and that derived from the common names for P and N: the former yield a more complete and 'purer' picture. In particular, if the value of $0(\Delta_r, \Delta_s)$ is large, then as a rule, $L(\Delta_r, \Delta_s)$ is large; however, the converse is not valid.

The thresholds separating large relations (which should lead to the conclusion regarding the dependence of neighbourhoods) from small ones (the conclusion being that the neighbourhoods are independent) were chosen in accordance with the magnitude of $0(\Delta_r, \Delta_s)$ as follows: the relation frequency histogram was constructed from the matrix

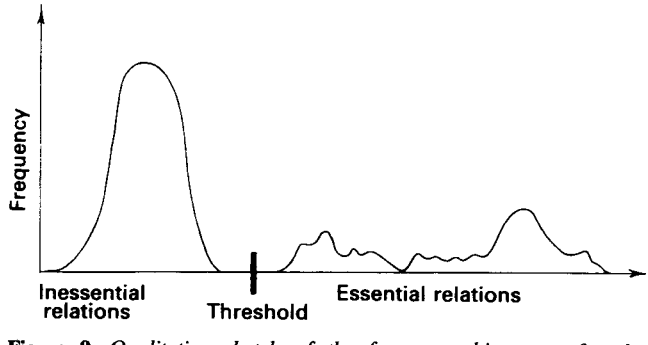

Figure 9. *Qualitative sketch of the frequency histogram for the neighbourhood pair relation in the matrix (number of common names $0(\Delta_r, \Delta_s)$ being fixed).*