$$sim(s_q, s_t, D) = \frac{\sum\limits_{f \in F} \{e^{-{p_f}^2/2\sigma^2} | f \subseteq s_q \wedge f \subseteq s_t\}}{\sum\limits_{f \in F} \{e^{-{p_f}^2/2\sigma^2} | f \subseteq s_q \vee f \subseteq s_t\}} \cdot \frac{n_{s_q}^{frequent}}{n_{s_q}} \cdot \frac{n_{s_t}^{frequent}}{n_{s_t}} \tag{1}$$

with
$p_f | f \subseteq s_q \wedge f \subseteq s_t$ ... significance of fragment f that occurs in s_q *and* s_t
$p_f | f \subseteq s_q \vee f \subseteq s_t$ ... significance of fragment f that occurs in s_q *or* s_t
σ ... standard deviation of the gaussian distribution (0.3)
F ... set of significant features
n_{s_q} ... number of fragments in the query structure
$n_{s_q}^{frequent}$ ... number of query structure fragments that occur frequently enough for statistical evaluation
n_{s_t} ... number of fragments in the neighbors
$n_{s_t}^{frequent}$ ... number of neighbors fragments that occur frequently enough for statistical evaluation
Minimum frequencies for statistical significance are derived from the χ^2 definition (with Yates correction) under the assumption that the fragment occurs only in asingle class.