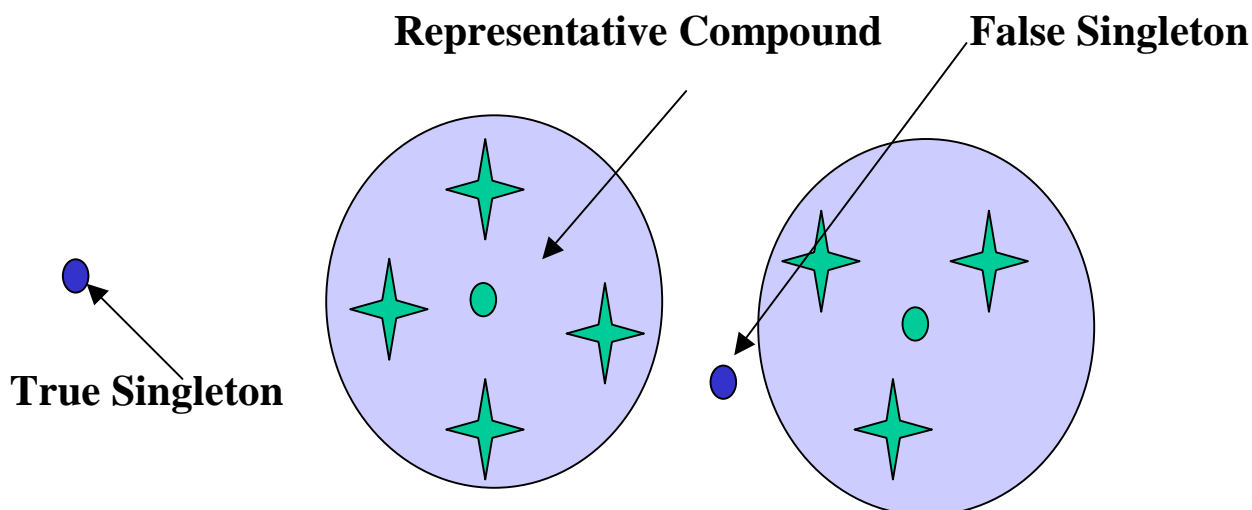


Grouping Module



Clustering compound collections has numerous applications, such as compound acquisition, lead optimization in High Throughput Screening, etc. The **Grouping Module** includes exclusion region grouping algorithms^{1,2}. Taking binary strings as input (e.g., MDL keys or Daylight fingerprints), similarities are generated via user-defined measures (e.g., Tanimoto or Tversky). Grouping with such discrete representations often generates ambiguous results³ however. Thus the module contains a tool that determines ambiguity indices, enabling the user to find similarity thresholds that produce the least ambiguity. True singletons and false singletons are identified. The latter are also merged into the group that contains their respective nearest neighbor. Both disjoint and non-disjoint (overlapping), either symmetric or asymmetric, versions are included. Each of the four versions report "ties" statistics³. MXN similarity searching is also supported. ASCII and Excel readable cluster and similarity reports are generated. All programs are available as standalone programs or C++ library functions.

1. JCICS, 1995, 35, 59-67. 2. JCICS, 1999, 39, 747-750. 3. JCICS, 2001, 41, 134-146.

System Architecture: Linux Red Hat 7.x, Irix 6.5, Windows 2000

Software Requirements: Fingerprint Module (or other binary fingerprint input, or dissimilarity matrices as input).

Mesa Analytics & Computing, LLC
505-983-3449 info@mesaac.com