Extraction and Semi-metric Analysis of Social and Biological Networks

3. Distance Functions

A valued function such that a nonnegative, symmetric, real-valued function known as a distance function. Most distance graphs obtained violate the triangle inequality expected of Euclidean distances. This type of distance function is known as a semi-metric. We show that the semi-metric behavior of these distance graphs, can be used for identifying specific implicit associations in the graph, and thus useful to identify trends in communities associated with the sets of documents from where associations were extracted.

4. Semi-metric behavior

\[ d(k_1, k_2) \leq d(k_1, k_3) + d(k_3, k_2) \]

5. Recovering Missing Knowledge

5.1. Co-Occurrence Proximity

Given a binary relation \( R \) between sets of keywords \( K \) and documents \( D \) we extract a co-occurrence proximity measure. \( KDP(k_i, k_j) \) is the probability that both keywords \( k_i \) and \( k_j \) co-occur in the same document \( d \in D \).

\[ KDP(k_i, k_j) = \frac{N(k_i, k_j)}{N_D \cdot (N_i - 1) \cdot (N_j - 1)} \]

Where

\( N(k_i, k_j) \) is the number of times \( k_i \) and \( k_j \) co-occur.

\( N_D \) is the number of documents.

\( N_i \) and \( N_j \) are the counts of documents containing \( k_i \) and \( k_j \) respectively.

\( PDP(k_i, k_j) \) is the probability that both keywords \( k_i \) and \( k_j \) co-occur in the same document.

\[ PDP(k_i, k_j) = \frac{N(k_i, k_j)}{N_D \cdot N_i \cdot N_j} \]

5.2. Power of Proximity Measures

Expert Knowledge: Immunology Testcase

Cytokines: Problem Terms

Receptor Molecules: Choice Terms

Example: Terrorist Networks

Random Deletions (Full and Partial)

Distance from People (terrorists) Document Proximity

PDP against Random Graphs