I. Introduction of network theory and Barabasi’s work

A. Scale-free networks of hubs plus nodes with fewer connections

B. Connection frequencies fit a power law distribution

C. Also define degree, degree centrality, and other graph theory concepts that should be included

II. What semantics can contribute to network theory

A. By looking at a network as a semantic structure, we can analyze the semantic relationships between the different components.

B. Semantic predications representing a coherent body of text can be arranged to form a network with a scale-free node distribution. (describe a SemRep-produced semantic predication).

III. Semantic analysis may begin to explain why scale-free networks look this way

A. The basic quantitative properties of nodes in a scale-free network describe it, but do not explain why it formed in such a way

i. For example, they don’t describe *why* a particular node is the most connected

IV. In this paper, we want to exploit network semantic properties in order to move toward an explanation of network characteristics

1. The first step is to analyze the quantitative semantic characteristics constituting a hub (a central concept, and all edges and nodes of 1 path-length). A single connection between the central concept and a node also represents a predication.
2. Semantically speaking, predications are restricted. Each edge is an arc (has direction). Arguments and predicates are restricted by logic in how they combine to form a reasonable assertion. Arguments are further defined in an ontology as a semantic class.
3. This analysis could reinforce the hypothesis that semantic characteristics can explain why networks form as they do.

V. Methods – We use graphs of semantic predications extracted from text from the biomedical domain. We start with all predications, then limit to different central concepts.

A. For four central concepts of different classes…

1. We choose four concepts with different semantic types, and that are distinct within the MeSH hierarchy

2. For each concept, we pull every semantic predication from the SemMed database that includes it as a subject or object (using EDAT), within EDAT values 2002 – 2012 (or all predications from these years, using the EDAT timestamps). We aggregate predications by month.

3. For each instance (or each month), we count (a) frequency of the central concept, (b) number of unique opposing arguments, (c) number of unique predicates, and (d) number of unique semantic types.

4. We use k-means clustering, with 4 centroids (because we have four concepts) to see if instances are unique enough to be clustered by class (i.e., concept). High clustering accuracy indicates that these quantitative values, when combine and aggregated by month, are distinctive to their given concept, when compared with those of other concepts. They are like a fingerprint, of sorts.

B. For central concepts within the same class…

1. For each of the four original concepts, we choose four more concepts that are descendants of the original, according to the MeSH hierarchy, and that share the same semantic type with the original concept.

2. We repeat the steps above, and analyze the clustering accuracy. High accuracy is indicative of distinctive values, as described above. Low accuracy indicates that the four sibling concepts are not as distinctive in terms of the quantitative values we gathered.

C. For the above experiments that yielded high accuracy, we use the data to train a multi-perceptron and test it with new data from another year span

1. To test if a given “fingerprint” is sufficient to identify data with another timestamp

2. Further tests idea that quantitative semantic values contribute to network formation