Adam Olenderski
27 March 2006
CS 790R
Complex Systems

Discussion Notes

- Language can be described in terms of a graph of word interactions
  - Nodes in the graph are the words in the language
  - Edges of the graph are the relationships between individual words, as described by their close (within a distance of 2 words or less) cooccurrence.
- This graph has the properties of a small-world network, namely:
  - High clustering coefficient
  - Low average path length
- This graph also displays the characteristics of a scale-free network, in that the probability P(k) of having a node with degree k scales as $P(k) = k^{-\gamma}$.
- This representation can capture verb-adverb, adjective-noun, verb-adjective, etc. correlations because they are manifest in relatively short distances, but does not capture longer-distance correlations
- The small-world/scale-free structure of the network is given as the reason for humans' ability to produce coherent sentences quickly as well as being able to richly describe a given concept, because any given word can be reached with, on average, less than three intermediate words.
- Rationalization for the author's claim that this network is representative of the way humans store language is shown in the forms of agrammatism and paragrammatism, syndromes in which humans lose the ability to use function words or substitute other words for function words, respectively.
- The paper does not go into any history of language or give a concrete rationalization for how or why words, when added to a lexicon, attach preferentially to already highly-connected words (as per the scale-free model)
- The networks that are created are static in that once a link is created between two words, it remains forever, and that links are binary. Either there is a link between two words or there isn't, and there is no facility given for strengthening or weakening those links according to the frequency with which they cooccur.
- There was an argument made that a network based on purely syntactic information cannot be fully descriptive, in that it ignores semantic relationships (such as synonyms or antonyms), which are also important in sentence production.