**Introduction**

The basic elements of language enable us to communicate in several mediums and contexts. Research in formal semantics illuminates how arguments logically fit together with a binding predicate to communicate a logical assertion (ref?). Consider the following example, taken from PubMed, an electronic medium that provides information from a biomedical context:

Nonsteroidal anti-inflammatory drugs (NSAIDs) are widely used for the treatment of pain in rheumatic disorders and chronic pain syndromes. (PMID 19715384)

In this example Nonsteroidal anti-inflammatory drugs and pain are the arguments of the communication, and treatment is the predicate value that reveals the relationship uniting them. Nonsteroidal anti-inflammatory drugs and pain also belong to classes of larger, more general meaning. For example, Nonsteroidal anti-inflammatory drugs also belongs to the general class drugs and pain belongs to the general class disease symptoms.

**Objectives**

This research has two objectives. First, to determine if quantitative properties associated with a concept such as Nonsteroidal anti-inflammatory drugs are substantially unique to distinguish it from other concepts. The second objective is to determine the feasibility of creating a classifier to distinguish concepts, using their associated quantitative properties (given a positive outcome for the first objective).

**Background**

SemRep (ref), a natural language processing application, extracts semantic predications (i.e., arguments bound by a predicate) from biomedical text, using the Unified Medical Language System (UMLS) (ref) as a knowledge base. For the above text, it extracts this semantic predication:

|C0003211|Anti-Inflammatory Agents, Non-Steroidal|phsu|phsu|||TREATS|C0030193|Pain|sosy|sosy||

Using MetaMap technology, SemRep maps Nonsteroidal anti-inflammatory drugs and pain to their respective UMLS Metathesaurus preferred arguments, Anti-Inflammatory Agents, Non-Steroidal and Pain. SemRep also identifies treatment as the predicate value, and maps it to TREATS within the UMLS Semantic Network. The output also includes the semantic types of the arguments, which serve as more general classes within the UMLS Semantic Network. The semantic type of Anti-Inflammatory Agents, Non-Steroidal is Pharmacologic Substance and the semantic type of pain is Sign or Symptom. To someone sufficiently familiar with this domain, these arguments logically fit together with this predicate. One could not say “rheumatic disorders treat nonsteriodal anti-inflammatory drugs” or “pain causes nonsteriodal anti-inflammatory drugs” and successfully convey logical statements, because the foundational knowledge within the biomedical domain indicates that such communication is irrational.

The propositional components of communication can be visualized in a graph. For example, the above semantic predication can be viewed in a network, where the arguments are nodes, and the predicate is the connecting edge:

Research in network theory has provided a framework for analyzing several types of data, including those illustrating communication instances as noted in published text. Since Barabasi’s landmark work describing the behavior of networks (ref) several others have added to this active research field (for example, see refs # - #). As Barabasi, Onnela et al. (2007) found that networks of mobile phone calls (serving as proxies for communication instances) exhibited behavior expected within a scale-free topology, such as power law distributions in basic node connections. Our preliminary research in semantic predicate networks yielded similar findings.

Because communication is constrained by what semantic components logically fit together, it is reasonable to conclude that each component interacts with a limited set of other components. For example, the concept *pain* has limited sets of predicates and arguments with which it interacts; the arguments interacting with *pain* also belong to finite sets of classes. The quantitative properties of these sets, as derived from text, may indicate their associated concepts. These communication components can also be seen as agents (parts?) within a network. Findings in this field may contribute to ongoing efforts in network theory research. (in the discussion we will note similar findings to Barabasi, and the potential for extra bits of info like link type/predicate could contribute to network theory).

Additional comments from Tom:

This needs to be done at the conceptual level (not with all the details). You need more explicit motivation, and a clearer, more thought out description of how the quantitative analysis you are proposing contributes to the semantic analysis you are proposing as an enhancement to current quantitative analysis in network research. I would say the line of reasoning should be something like the following: 1) start with what Barabasi has accomplish, 2) note the deficiencies of purely quantitative analysis in explaining the nature of communication. 3) discuss how semantic analysis can contribute. 4) explain how the current proposal contributes to effective semantic analysis for the purpose of contributing to Barabasi’s accomplishments. You don’t need description of SemRep and MetaMap in what you send to Trevor.

Liz’s comments: This paper, as I envisioned it, would be a purely quantitative analysis. It would be a tacit introduction to the second paper. The second paper (looking at slope and predicate properties) will likely include a qualitative analysis of the predicate links.