Measuring Word Meaning

Scott Grimm, Ph.D.
Department of Linguistics

Research on language during the last half century has developed increasingly insightful models of language knowledge, acquisition, processing, and production. A core component of this effort involves modeling nominal reference: how language users employ nouns to describe things in the world as well as to establish and track reference to them. A reasonable degree of success has been achieved so far; however, standard methodology in theoretical linguistics makes generalizations based on data sets, which, for a variety of reasons, are typically restricted in size and scope.

This talk gives an overview of ongoing work taking place in the Quantitative Semantics Lab developing data science methods and resources to investigate meaning from a more global perspective, ultimately providing a window onto how language is structured to make reference to objects, concepts, or events.

The hypothesis guiding the project is that measuring the grammatical behavior of lexical items yields insight into their semantic structure. The broadest goal of this project is to deliver a large-scale classification of nouns. In effect, this effort is comparable to foundational work in botany measuring plant features, such as sepal width of different flowers, for identifying taxonomic organization. I will discuss how results from this work bear on various questions that have been raised in natural language semantics, philosophy of language, and cognitive science, such as the distinction between nouns for which the referents can be counted (“three dogs”) and those for which it is impossible (“three sands”), as well as questions relating to the nature of nouns designating abstract entities (“justice”) or events (“war”).

On-Going Research:
Executable Models for Pathway Analysis

Rohith Palli
Department of Biochemistry and Biophysics

Pathway analysis is widely used to gain mechanistic insights from high throughput omics data. However, most methods do not utilize signal integration, leading to enrichment of convergent pathways when downstream genes are dysregulated. Novel methods that incorporate network topology and signal flow could rank the pathways based on dysregulation in key regulatory genes. Discrete state network modeling can facilitate this advanced implementation due to simplicity in parameterization. Here, we model cellular heterogeneity using discrete state dynamics and apply these models to measure pathway activities in cross-sectional data. We introduce a new algorithm, Boolean Omics Network Invariant-Time Analysis (BONITA), for network signal propagation, rule inference, and pathway analysis. Our signal propagation approach models heterogeneity in omics data as arising from intercellular heterogeneity rather than intracellular stochasticity and propagates binary signals repeatedly across networks. Context-specific rules inferred by genetic algorithm followed by local search are used to determine the impact of each node in a pathway and score the probability of pathway dysregulation. We show that BONITA pathway analysis has higher sensitivity at lower levels of signal dysregulations and performs as well as other methods at higher levels of signal dysregulations using simulated data. Additionally, application of BONITA pathway analysis to previously published and validated RNA-sequencing studies identifies additional relevant pathways in in vitro human cell line experiments. Thus, we demonstrate the power of this approach in prioritizing not only pathways but also specific genes and their interactions as compared to the existing methods. BONITA is available at: https://github.com/thakarlab/BONITA.