

# Incremental Distant Supervision for Extraction of Regulation Relations from Biology Text

David Sidi

School of Information  
The University of Arizona

IT Summit, Autumn 2016



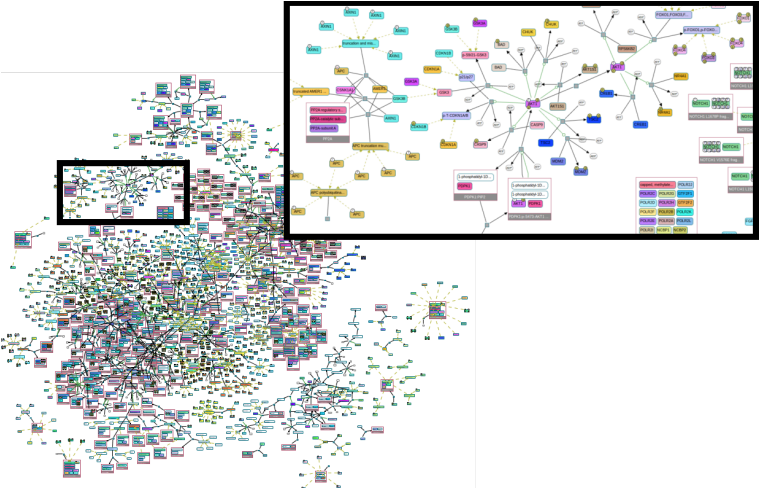
# Acknowledgements



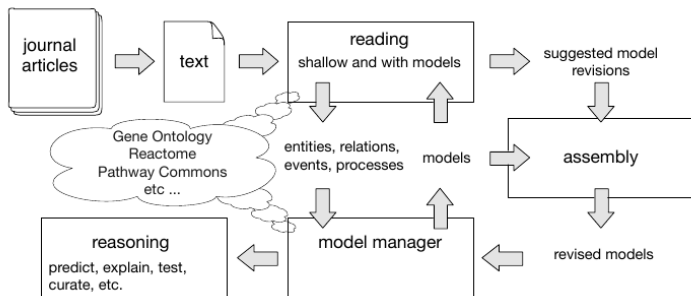
also, the members of clulab

- Mihai Surdeanu
- Gustave Hahn-Powell
- Marco Valenzuela-Escárcega
- Dane Bell
- Enrique Noriega
- Rebecca Sharp
- Tom Hicks

# BIDS is part of an effort to build “Big Mechanisms”

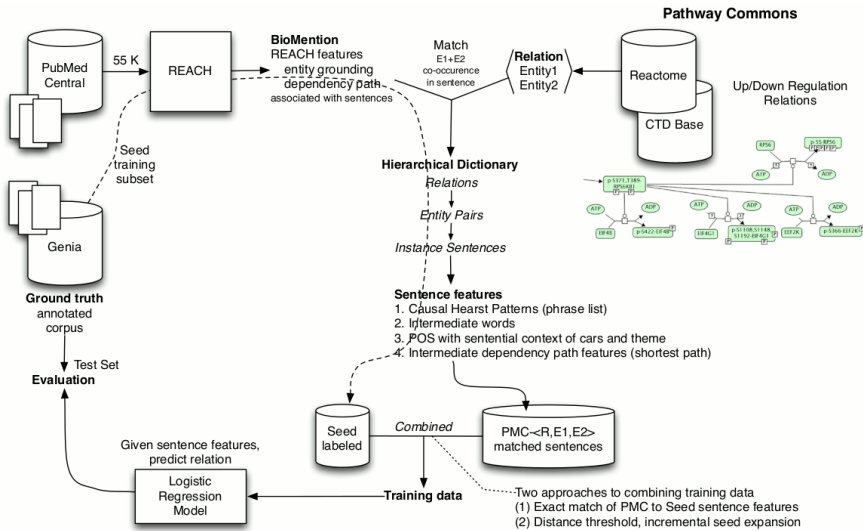


# BIDS is a way to read with a model



(Cohen, 2015)

# BIDS is Distant Supervision that builds on what it knows



The signal used to expand training data is noisy; BIDS uses high-precision, low-recall features

### TNF-alpha positively regulates NF-Kappa B

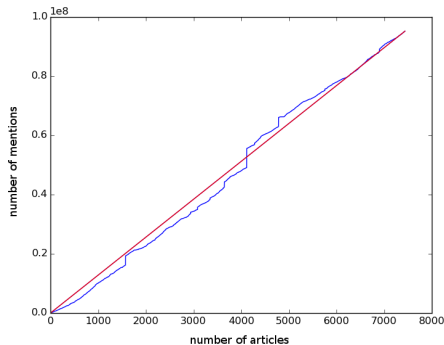
“Secreted TNF-alpha mediated the enhancement of nuclear factor kappa B (NF-kappa B) and activator protein-1 (AP-1) binding activity.”

### TNF-alpha and NF-Kappa B co-occur, but positive relation is not expressed

“Both TNF-alpha and NF-kappa B are also involved in the stimulation of the pathway that leads to the expression of major histocompatibility complex (MHC) class I molecules and, hence, antigen presentation to the T cells.”

# Lots of data is needed for high-precision, low-recall features

- $\sim 2.5$ M regulation relations in PathwayCommons DBs, 24.4K regulation relations annotated in GENIA-Event
- $\sim 2.56$ M entity mentions extracted from provisional 200K articles. These are searched for each regulation relation.
- Total number of articles in journals of interest to us is now up to over 1M



# Thank you

**To be notified when BIDS has its first release, please email  
dsidi@email.arizona.edu.**