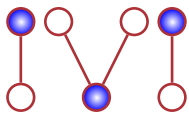
FROM TEXTUAL EVIDENCE TO PROBABILISTIC MODELS

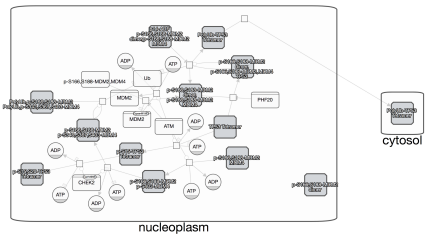
Adarsh Pyarelal

IVILab Colloquium, August 21, 2018

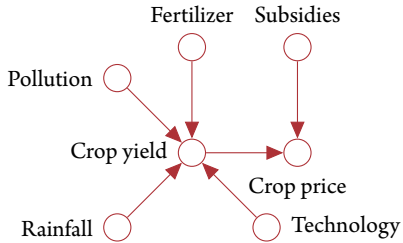


Domain: Complex systems that can be represented by graphs

Biological Pathways



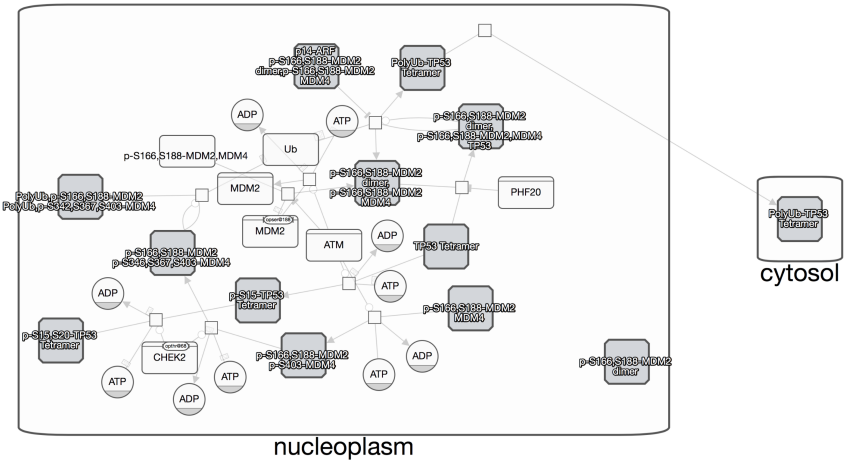
Agricultural Pathways



Part I

Big Mechanism

Example biological pathway: Stabilization of p53



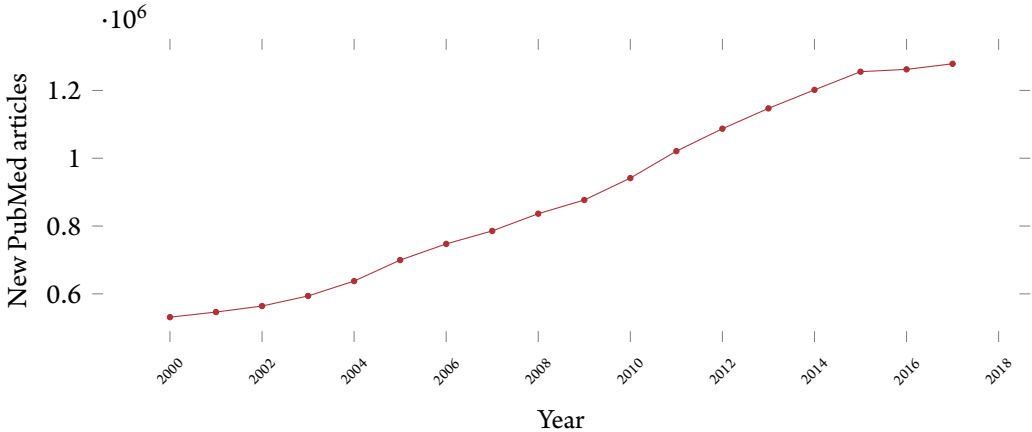
Why study biological pathways?

Understanding their structure can help us design new, targeted drugs to treat diseases with complicated mechanisms, such as cancer.

The problem of fragmented knowledge

Biomedical journal articles typically focus on fragments of pathways, rather than complete pathways.

Humans can't keep up with the biomedical literature!



Solution:

Use machines to extract fragments and assemble pathways.

Our research focus

Approaches to Assembly

Deterministic

Output: A single pathway

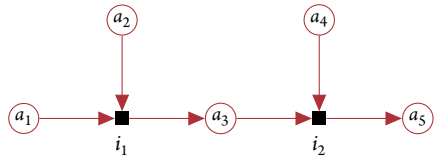
Probabilistic (Our approach)

Output: A probability distribution over pathways.

Hypergraph representation

$$M = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{array} \begin{array}{c} \begin{array}{cc} i_1 & i_2 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \end{array}$$

Incidence Matrix



Hypergraph

Part II

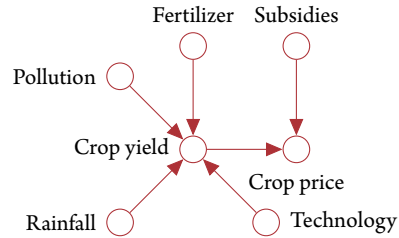
World Modelers

Definition of food insecurity

'Limited or uncertain availability of nutritionally adequate and safe foods or limited or uncertain ability to to acquire acceptable foods in socially acceptable ways'

- Anderson, S. A., *Core indicators of nutritional state for diffcult-to-sample populations,* The Journal of Nutrition (USA) (1990)

Food insecurity is a complex function of many factors, which can be modeled by a *causal influence network*, like the one on the right.



Building on Big Mechanism

The causal influence network will be assembled from textual evidence, similar to biological pathways.

Making predictions

In addition to understanding the structure of the causal influence network, we would ideally like to use it to make predictions about the factors in the future.

To do this, we use *dynamic Bayes networks*.

Model Uncertainty

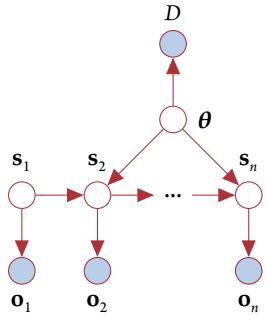
Consider the phrase: *A small increase in X causes a large decrease in Y.*

1. Can we extrapolate from this to talk about the effect of a small *decrease* in X?
2. What does the writer mean by 'small' and 'large'? Does their interpretation match that of other writers?
3. What if we have other evidence sentences about X and Y that differ in
   - Magnitude:
     - Small increase in X causes large increase in Y
     - Tiny increase in X causes huge increase in Y
   - Polarity
     - A small increase in X causes large decrease in Y
     - A small increase in X causes a large increase in Y
4. What if we have no evidence sentences, are there reasonable assumptions?

Idea:

Consider the *model itself* to be a random variable.

Linear Dynamical System with random transition model



Simple causal linear dynamical system

How far can we get with just the following sentence?

→ *A small increase in X causes a large decrease in Y.*

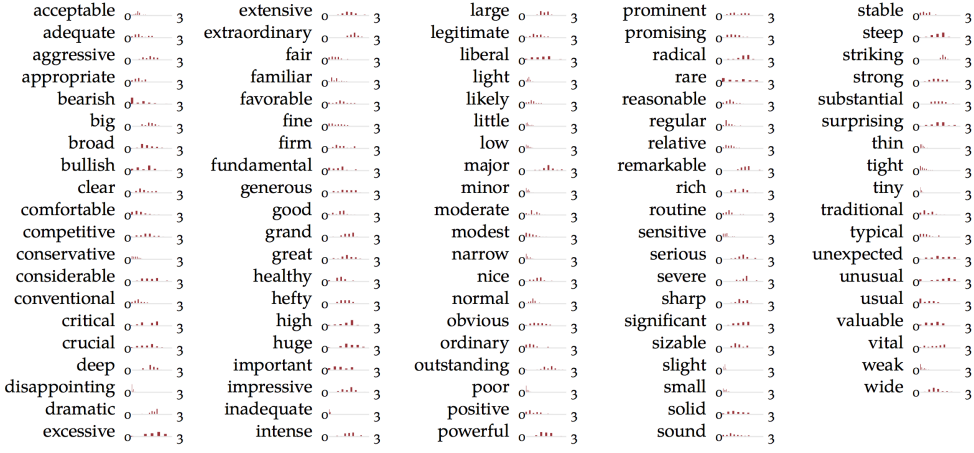
The sentence

A small increase in X causes a large decrease in Y

is telling us about the shape of $\frac{\partial Y}{\partial X}$.

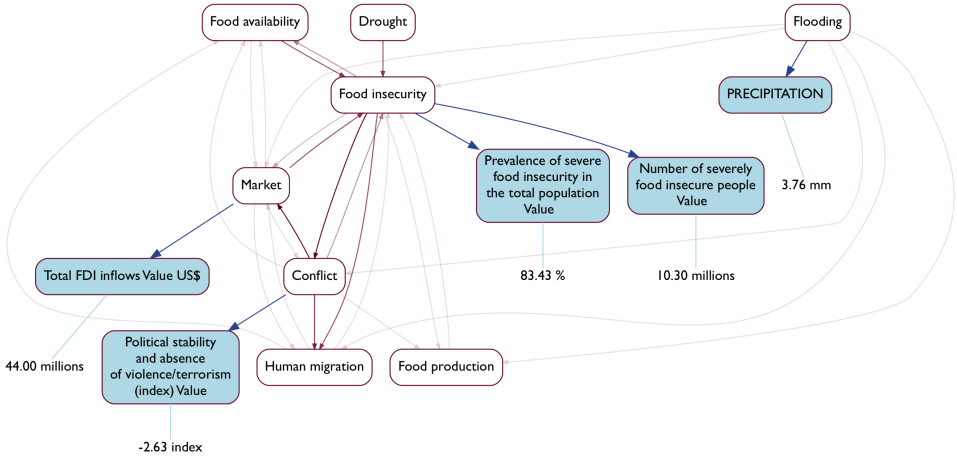
$$
\begin{aligned}
Y_{n+1} &\cong Y_n + \left.\frac{dY}{dt}\right|_n \Delta t \\
&= Y_n + \left(\left.\frac{\partial Y}{\partial t}\right|_n + \left.\frac{\partial Y}{\partial X}\frac{\partial X}{\partial t}\right|_n \right) \Delta t \\
&= Y_n + \left(\dot{Y}_n + \left.\frac{\partial Y}{\partial X}\right|_n \dot{X}_n \right) \Delta t
\end{aligned}
$$

Quantifying gradable adjectives through crowdsourcing



Units: Standard deviations from the mean: $\bar{r} = \frac{r-\mu}{\sigma}$

Causal Analysis Graph for South Sudan (2014)



Undergraduate Opportunities

Big Mechanism

- *Task*: Analyzing REACH extractions from > 1 million papers
  Skills: Working with graph databases (Neo4j), data analysis.
- *Task*: Implementing a hypergraph sampler in C++
  Skills: C++, Python

World Modelers

- *Task:* Creating a web-based survey to understand how people describe models in natural language.
  Skills: Javascript (including d3.js)
- *Task*: Implementing unit and regression tests for Delphi.
  Skills: Python, unit and regression testing basics (we use `pytest`).