A brief introduction to graphical models and deep methods in computational network biology

Clayton W. Seitz

February 27, 2022

Outline

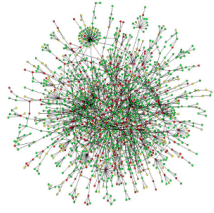
Computational network biology

Emerging research field that encompasses theory and applications of network models to study complex interactions of cells, DNA, RNA, proteins, and metabolites

Say we have a set of variables $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ which might have some statistical dependence. $\boldsymbol{x}$ might be RNA or protein expression data, for example

- Often we are handed a batch of empirical samples $\boldsymbol{X} = \{\boldsymbol{x_1}, .., \boldsymbol{x_k}\}$
- We want to learn about the generating distribution $P(\boldsymbol{x}, t)$

Joint effort between physics, computer science, and biology



A gene interaction network

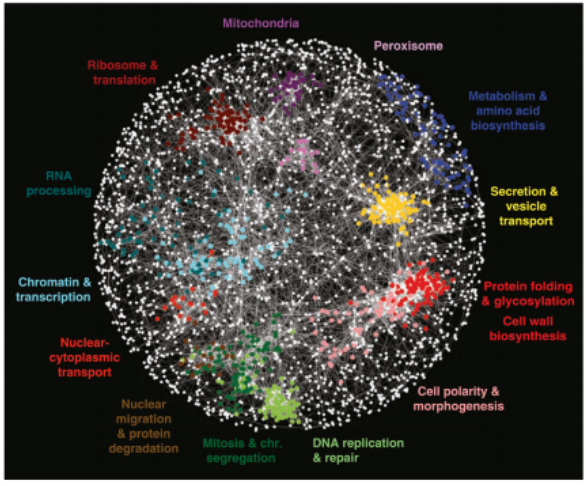


Figure 1: **Landscape of genetic interactions** in cells. Edges between genes denote Pearson correlation coefficients ($\rho > 0.2$) calculated from the complete genetic interaction matrix.

A protein interaction network

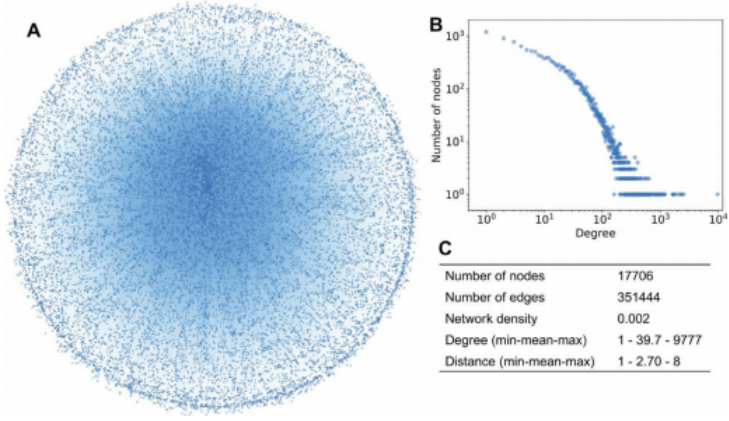


Figure 2: **Human protein interactome** of 17,706 proteins and 351,444 interactions (A) Overall complex network of human interactome. (B) Degree (connectivity) distribution of proteins by following a power-law tail. (C) Several selected network topological characteristics of the interactome.

A cellular interaction network (model neurons)

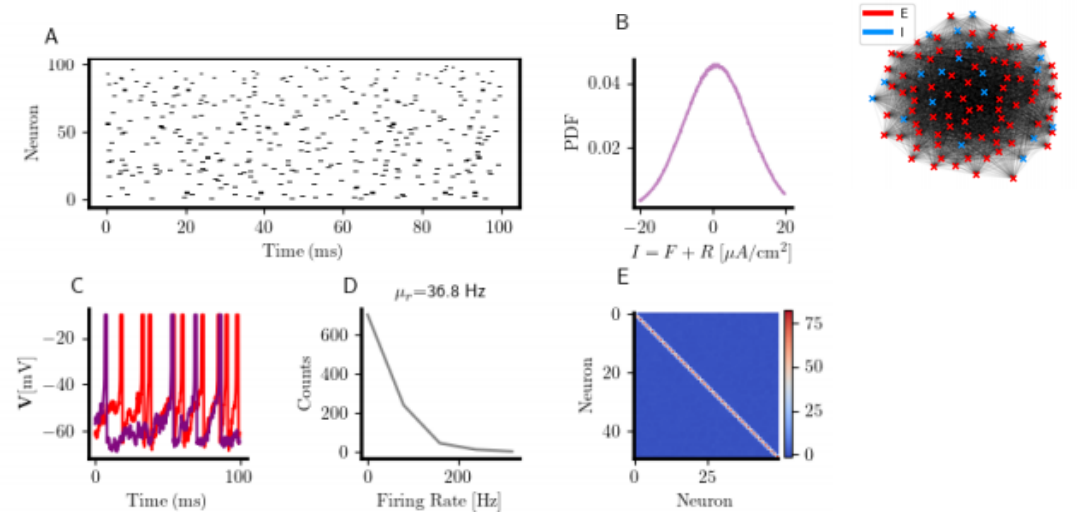


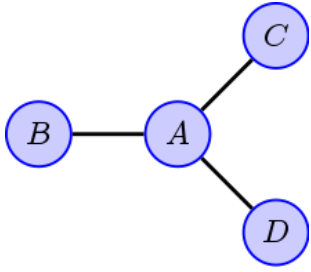
Figure 3: **Asychronous spiking of model neurons** (A) Steady-state raster plot of $N = 100$ uncoupled EIF neurons undergoing stimulation with GWN with $\mu = 2\mu A/\mathrm{cm}^2$ and $\sigma = 9\ \mu A/\mathrm{cm}^2$

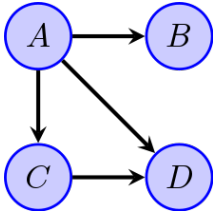
Probabilistic graphical models (PGMs)

Probabilistic graphical models are a class of machine learning algorithms that represent statistical dependencies of probability distributions as graphs

Two main types used in machine learning:
Bayesian Networks (BNs), Markov Random Fields (MRFs), but there are others

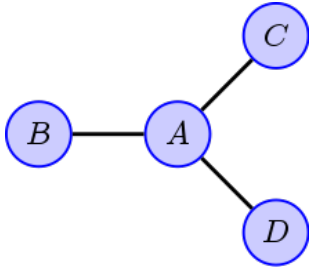
Major advantage is that they are structured models
They do not scale as easily as deep networks





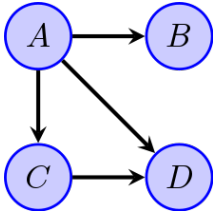
Probabilistic graphical models (PGMs)

Say we have a joint probability over gene expression $P(\boldsymbol{X})$
A PGM describes how $P(\boldsymbol{X})$ factors



Markov Random Fields (MRFs) e.g., Ising model

$$P(\boldsymbol{X};\Theta) = \frac{1}{Z}\prod_{i=1}^{N} P(\boldsymbol{X_i}, \mathcal{C}(X_i);\Theta_i)$$



Bayesian Network (BNs) - include causality

$$P(\boldsymbol{X}|\mathcal{G},\Theta) = \prod_{i=1}^{N} P(\boldsymbol{X_i}|\mathcal{C}(X_i),\Theta_i)$$

BNs as well as hybrid models have been used to examine gene expression

An example graphical model

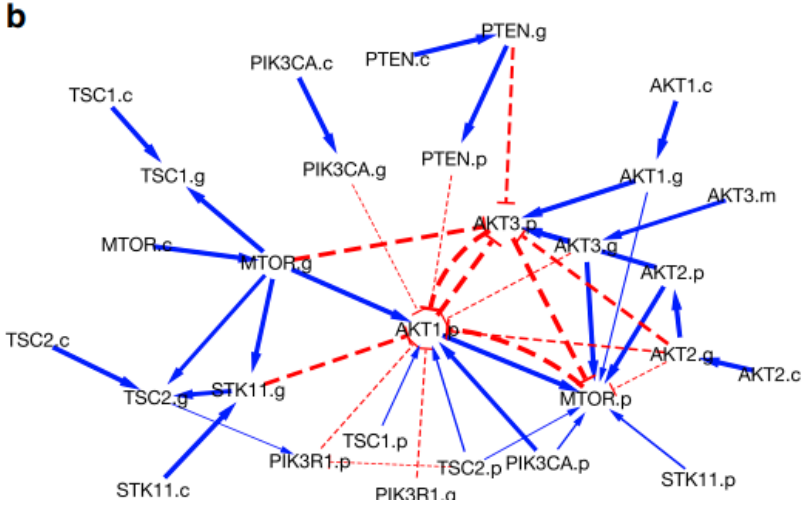


Figure 4: **PI3K pathway graph** discovery using graphical modeling (Ni et al. Bioinformatics 2018). c - transcript count, g - gene, p - protein, m - methylation

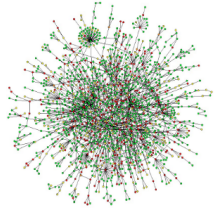
A tradeoff between mechanistic understanding and scale

Fine structure of molecular interactions sometimes can be resolved for low dimensionality

Computational complexity often scales exponentially with an increase in variables, density of interactions

In high-dimensional biological networks we often turn to classic dimensionality reduction or deep methods

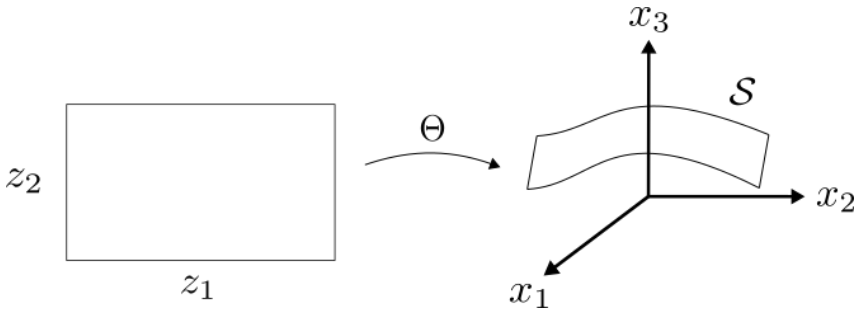
Introducing latent variables into the model can reduce computational complexity



Latent variables

Modeling all possible conditional dependencies quickly becomes intractable, lots of parameters

Introducing latent variables ***z*** can reduce the number of needed parameters



Variational autoencoders (VAEs)

The VAE architecture has been very succesful when applied to RNA-seq datasets see (Lopez Nature Methods 2020)

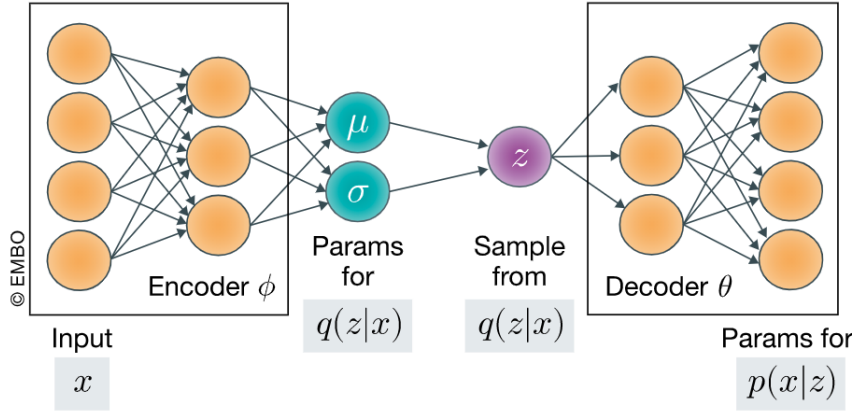
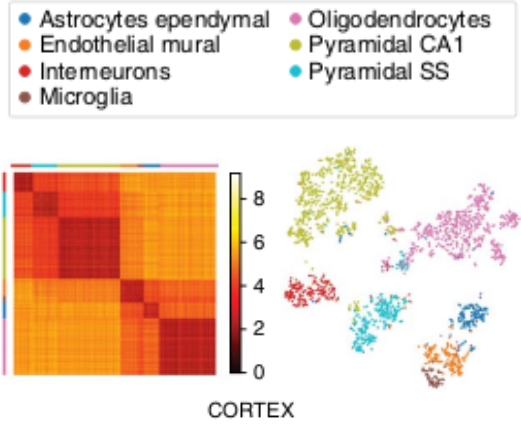


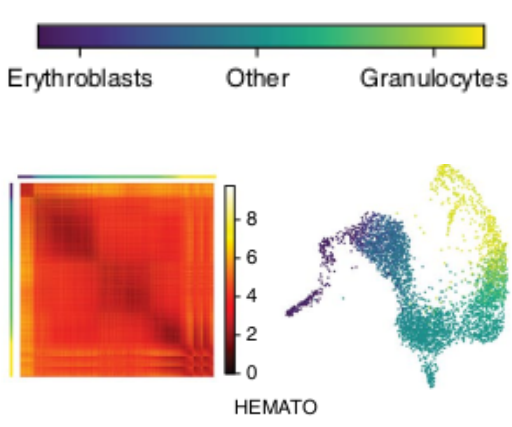
Figure 5: **Variational autoencoder architecture** Lopez 2020 EMBO

Using the VAE for cell phenotyping

Can do hypothesis testing (Bayes factor) but does not explicitly capture visible-visible causal relationships (captures latent-visible)

558 genes/3005 cells for CORTEX, 7,397 genes/4016 cells for HEMATO (Lopez Nature Methods 2020)





References I