Network-based interpretation and integration

Sushmita Roy
sroy@biostat.wisc.edu

Computational Network Biology
Biostatistics & Medical Informatics 826
https://compnetbiocourse.discovery.wisc.edu

Nov 29th 2018
Perturbations in networks

• Understanding genetic perturbations are important in biology
• Genetic perturbations are useful to identify the function of genes
  – What happens if knock gene A down?
    • Measure some morphological phenotype like growth rate or cell size
    • Measure global expression signatures
• Perturbations can be artificial or natural
  – Artificial perturbations
    • Deletion strains
  – Natural perturbations
    • Single nucleotide polymorphisms
    • Natural genetic variation
• Perturbations in a network can affect
  – Nodes or edges
  – Edge perturbations
    • Mutations on binding sites
Types of algorithms used to examine perturbations in networks

• Graph diffusion followed by subnetwork finding methods
  – HOTNET
• Probabilistic graphical model-based methods
  – Factor graphs
  – Nested Effect Models (NEMs)
• Information flow-based methods (also widely used for integrating different types of data)
  – Min cost max flow
  – Prize collecting steiner tree
Probabilistic graphical models for interpreting network perturbations

• “Inference of Patient-Specific Pathway Activities from Multi-Dimensional Cancer Genomics Data Using PARADIGM. Bioinformatics” https://academic.oup.com/bioinformatics/article/26/12/i237/282591


Factor graphs

- A type of graphical model
- A bi-partite graph with variable nodes and factor nodes
- Edges connect variables to potentials that the variables are arguments of
- Represents a global function as product of smaller local functions
- Perhaps the most general graphical model
  - Bayesian networks and Markov networks have factor graph representations
Fig. 1. A factor graph for the product $f_A(x_1)f_B(x_2)f_C(x_1, x_2, x_3) \cdot f_D(x_3, x_4)f_E(x_3, x_5)$.

From Kschischang, Frey, Loeliger 2001
Probabilistic graphical models for interpreting network perturbations

• “Inference of Patient-Specific Pathway Activities from Multi-Dimensional Cancer Genomics Data Using PARADIGM. Bioinformatics” https://academic.oup.com/bioinformatics/article/26/12/i237/282591


Nested Effect Models

Fig. 1. An introduction to Nested Effects Models. Plot (a) shows a toy dataset consisting of phenotypic profiles for eight perturbed genes \(A, \ldots, H\). Each profile is binary with *black* coding for an observed effect and *white* for an effect not observed. The eight profiles are hierarchically clustered, showing that they fall into four pairs of genes with almost identical phenotypic profiles: \((A, B)\), \((C, D)\), \((E, F)\) and \((G, H)\), as shown in plot (b). An important feature of the data missed by clustering is the subset structure visible between the profiles in the data set: the effects observed when perturbing genes \(A\) or \(B\) are a superset to the effects observed for all other genes. The effects of perturbing \(G\) or \(H\) are a subset to all other genes’ effects. The pairs \((C, D)\) and \((E, F)\) have different but overlapping effect sets. The directed acyclic graph (DAG) shown in plot (c) represents these subset relations, which are shown in plot (d). Compared to the clustering result in plot (b) the NEM additionally elucidates relationships between the clusters and thus describes the dominant features of the data set better.

Markowetz et al, 2007
Key properties of Factor Graph-NEMs (FG-NEMs)

• NEMs assume the genes that are perturbed interact in a binary manner
• But many interactions have sign
  – inhibitory or stimulating action
• FG-NEMs capture a broader set of interactions among the perturbed genes
• Formulation based on a Factor Graph
  – Provide an efficient search over the space of NEMs
Notation

• **S**-genes: Set of genes that have been deleted individually
• **E**-genes: Set of effector genes that are measured
• \( \Theta \): The attachment of an effector gene to the S-gene network
• \( \Phi \): The interaction matrix of S-genes
• \( X \): The phenotypic profile, each column gives the difference in expression in a knockout compared to wild type
  – Rows: E-genes
  – Columns: S-genes
• \( Y \): Hidden effect matrix, each entry is \{-1, 0, +1\} which specifies whether an S-gene affects the E-gene
An example of 4 S-genes and 13 E-gens

A-\rightarrow C is reflected in the scatter plot. When \( X_A \) is up, \( X_C \) is up. When \( X_A \) is down, \( X_C \) is down or no change.

B-\rightarrow D is also reflected in the scatter plot. \( X_D \) is a subset of opposite changes from \( X_B \).
S-gene interaction modes and their expression signatures

- Depending on the connections of the S-genes to one another and to the E-genes, a disruption in an S-gene will cause E-genes to either increase or decrease in expression relative to wild-type. For example, E-gene $E_{7}$ decreases under $D_B$ relative to wild-type because the wild-type activation by $B$ is absent in the deletion. On the other hand, the expression of $E_{10}$ also decreases under $D_B$ relative to wild-type but as a result of a different mechanism. In wild-type, $E_{10}$ is expressed at a baseline level because its repressor, the product of gene $D$, is inhibited by $B$'s product. However, in the $B$ deletion, $D$ is derepressed, leading to inhibition of $E_{10}$. This toy example illustrates that the disambiguation of inhibition and activation, both for S-gene interactions and E-gene attachments, make it possible to account for an expanded set of mechanisms leading to the observed expression changes.

The E-gene expression changes are available in a data matrix $X$ where each column gives the difference in expression of each E-gene under the deletion of a single S-gene relative to wild-type. $X$ may also contain replicates in the form of repeated S-gene knock-downs. The entry $X_{eAr}$ represents $e$'s expression change under the $r$th replicate of $D_A$. Furthermore, we assume that an unknown expression ''state'' for each E-gene under each knock-down, determines its set of expression changes observed across the $\{X_{eAr}\}$ replicates in the microarray data. The matrix, $Y$, records a hidden state for each E-gene under each knock-down, where entry $Y_{eA}$ is the state of E-gene $e$ under $D_A$. We allow the states to be ternary-valued $\{+1, 2, 0\}$ representing whether $e$ is up-regulated, down-regulated, or unchanged, respectively.

Figure 1. Predicting Pair-wise Interaction Using Quantitative Nested Effects.

(A) Hypothetical example with four S-genes, $A$, $B$, $C$, and $D$. The graph contains one inhibitory link, $B \times D$ (left). A heatmap of E-gene expression under knockdown of each S-gene shows both inhibitory and stimulatory effects (middle). Scatter plots of the $C$, $A$, $B$, and $D$ knock-outs show that expression fits in the shaded preferred regions of each interaction (right). The inhibitory link explains some of the ''observed'' data: expression changes under $D_B$ (bright red or bright green entries in the heatmap) occur in a subset of the E-genes for which the opposite changes occur in $D_B$. (B) Data from a known inhibitory interaction. Expression levels of effect genes under the DIG1/DIG2 knock-out (y-axis) plotted against their levels under the STE2 knock-out (x-axis) as detected in [17]. Expression changes significant at $\alpha = 0.05$ indicated in gray lines. DIG1/DIG2 is known to inhibit STE12. (C) Interaction modes. Observed E-gene expression changes are compared to five possible types of interactions between two S-genes, $A$ and $B$ (i–v). The top row illustrates the expected nested effects relationship for each type of interaction mode: circles represent sets of E-genes with expression changes consistent with either activation (blue circles) or inhibition (yellow circles). Scatter-plots for each interaction mode show the hypothetical expression changes under $D_A$ (x-axis) and $D_B$ (y-axis) for all E-genes (circles). E-gene levels are either consistent (filled) or inconsistent (open) with the mode. Shaded regions demark expression levels consistent with each interaction model. The example shows expression changes that most closely match the inhibition mode (indicated by the greatest number of closed circles).
Probabilistic model for NEMs

- Goal is to find a network, $\Phi$ and $\Theta$ that best fit the observed data ($X$)
- This is an inference problem
- Use a Maximum a posterior (MAP) approach

$$J(X) = \max_{\phi, \theta} P(\phi, \theta | X)$$

$$J(X) = \max_{\phi, \theta} \sum_{Y} P(\phi, \theta, Y | X)$$

- $Y$ encodes the “true” expression state of effector genes ($X$).
- $X$ is a noisy measurement of $Y$. $Y$ is the quantity we need to sum over
Probabilistic model continued

\[
J(X) = \max_{\Phi, \Theta} \left\{ P(\Phi) \sum_{Y} \prod_{e \in E} P(Y_e | \Phi, \theta_e) P(X_e | Y_e) \right\} \quad \text{Independence over all E-genes}
\]

\[
= \max_{\Phi, \Theta} \left\{ P(\Phi) \prod_{e \in E} \sum_{Y} P(Y_e | \Phi, \theta_e) P(X_e | Y_e) \right\} \quad \text{Re-arranging the terms}
\]

\[
= \max_{\Phi, \Theta} \left\{ P(\Phi) \prod_{e \in E} L_e \right\}
\]
Digging inside the $L_e$ term

- Note: $Y_e = \{Y_{eA}, Y_{eB}, Y_{eC}, \ldots, Y_{eN}\}$, where $N$ is the total number of $S$-genes.
- Define $L'_e$, proportional to $L_e$ using a set of pairwise potentials

$$
L'_e = \sum_{A, B \in S} \prod_{Y_{eA}, Y_{eB}} P(Y_{eA}, Y_{eB} | \phi_{AB}, \theta_{eAB}) P(X_{eA} | Y_{eA}) P(X_{eB} | Y_{eB})
$$

- $\phi_{A, B}$ The S-gene interaction
- $\theta_{eAB}$ Attachment of gene $e$ with respect to $A$ or $B$
Digging inside the $L_e$ term

- $\phi_{A,B}$ The S-gene interaction
- $\theta_{eAB}$ Attachment of gene e with respect to A or B

Now the joint can be written in a more tractable way

$$J(X) = \max_{\Phi} \left\{ P(\Phi) \prod_{e \in E, A,B \in S} \max_{\theta_{eAB}} \right\}$$

$$\sum_{Y_{eA}, Y_{eB}} P(Y_{eA}, Y_{eB}|\phi_{AB}, \theta_{eAB}) P(X_{eA}|Y_{eA}) P(X_{eA}|Y_{eA})$$

Each of these conditional distributions will correspond to a factor
Defining the factors

\[
P(Y_{eA}, Y_{eB} | \phi_{AB}, \theta_{eAB}) P(X_{eA} | Y_{eA}) P(X_{eA} | Y_{eA})
\]

Four variable factor, over discrete variables
\(Y_{eA}\): binary variables

\(\phi_{AB}\) Four values for each possible type of interaction: inhibitory, activating, equivalent, no interaction

\(\theta_{eAB}\) Interaction of \(e\) with \(A\) or \(B\): inhibited or activated by \(A\) or \(B\) or no action

This factor has value=1 if the E-gene \(e\) is attached to either \(A\) or \(B\) and \(e\)'s state is consistent with the interaction mode between \(A\) and \(B\).
The prior over S-gene graph

- The prior $P(\Phi)$ can incorporate prior knowledge of interactions among genes in pathways.
- At its simplest, it should encode a transitivity relationship to force all pairwise interactions to be consistent among all triples.

\[
P(\Phi) \propto \left( \prod_{A,B,C \in S} \tau_{ABC}(\phi_{AB},\phi_{BC},\phi_{AC}) \right) \left( \prod_{A,B \in S} \rho_{AB}(\phi_{AB}) \right)
\]

Transitivity constraint for triples

Physical network constraints

Example transitivity: If $A \rightarrow B$, $B \rightarrow C$, Then, $A \rightarrow C$
Factor graph representation of NEMs

\[ P(Y_{eA}, Y_{eB} | \phi_{A,B}, \theta_{eAB}) \]

\[ P(X_{eA} | Y_{eA}) \]
Inference on the factor graph

- Find most likely configurations for $\phi_{A,B}$
- Use a message passing algorithm called the Max-Product algorithm (standard for factor graphs)
- Message passing happens in two steps
  - Messages are passed from observations $X_{eA}$ to the $\phi_{A,B}$
  - Messages are passed between the interaction and transitivity factors until convergence
Does FG-NEM capture activating and inhibitory relationships?

FG-NEM: capture inhibitory and activating relationships
uFG-NEM: capture only unsigned interactions
FG-NEM AVT: FG-NEM run on absolute value data
Solid lines: structure recovery
Dashed lines: sign recovery
Pathway expansion

- Attach new E-genes to S-gene network
- An attached gene $e$ to S-gene $s$ asserts that $e$ is directly downstream of $s$
- All E-genes attached to the S-gene network are called frontier genes
- An E-gene’s connectivity is examined based on the Log-likelihood Attachment Ratio

$$LAR(e) = \log \left( \frac{\max_{i \neq 0} P(X_e | \Phi, \theta_e = i)}{P(X_e | \Phi, \theta_e = 0)} \right)$$

One of the S genes
FG-NEM based pathway expansion in yeast

Figure 4. Yeast knock-out compendium predictions.

(A) Precision/recall comparison. Each method's ability to expand a pathway was compared. Thick lines indicate mean precision and shaded regions represent standard error of mean calculated over the networks with the five highest AUCs from any of the tested methods.

(B) Network expansion comparison. Networks were predicted for a non-redundant set of GO categories containing four or more S-genes in the Hughes et al. (2000) compendium and used to predict held-out genes from the same category (see Methods). The area under the curve (AUC) for each pathway was calculated for each method. AUC ratios (y-axis) were calculated for each method relative to the lowest AUC.

(C) Compatibility of physical evidence and predicted S-gene interactions. Each point is the margin of compatibility (MOC, see Methods) of a predicted genetic interaction to high-throughput physical interaction data when physical interaction evidence was used (y-axis) and when it was not used (x-axis). Coloring indicates two-dimensional density estimation of points. Inset shows detail of the highest density region.

Prediction methods that are significantly better than the lowest performing method, excluding random, at the 0.05 level (*) and 0.01 level (**) were determined by a proportions test on the top 30 predictions from each method.

(D) Predicted S-gene networks for the ion homeostasis pathway. Shown are predicted networks from the FG-NEM method (Signed) and the uFG-NEM method (Unsigned). Arrows indicate activating interactions and t's indicate inhibiting interactions. The absence of a link between a pair of S-genes indicates the most likely mode for the pair was the non-interaction mode. Equivalence interactions are indicated with double lines and S-genes connected by equivalence are grouped into dashed ovals.

doi:10.1371/journal.pcbi.1000274.g004

Template matching: rank E genes based on similarity in expression to an “idealized template”
FG-NEM infers a more accurate network than the unsigned version in yeast

- FG-NEM and uFG-NEM networks inferred in the ion-homeostasis pathway
- FG-NEM inferred more genes associated with ion homeostasis compared to uFG-NEM
FG-NEM application to colon cancer

Cancer invasion network identification. We applied FG-NEMs to recover a network for the second-tier genes. We included E-genes that demonstrate a robust and significant effect under at least two of the knock-downs included in the Irby et al. (2005) study. We selected genes whose log$_2$ ratios differ by less than 0.5 in replicate arrays and had an absolute log$_2$ expression change at least equal to the mean absolute level of the activated distribution (1.75) in at least two arrays. Using these criteria, we identified 185 E-genes to use for model inference. Figure 5A shows the expression data of these E-genes plotted in order of their predicted attachment points as identified by FG-NEMs. For the most part, E-gene expression changes moved in the same direction following knock-down across the panel of five S-genes, indicating the presence of mostly stimulatory links among the S-genes (Figure 5A). This is in contrast to Figure 1A, where expression changes of a single E-gene move in the opposite direction following knock-down of S-genes connected by an inhibitory link.

The absence of inhibitory links among S-genes is expected since, according to the selection criteria, all of the S-genes were found previously to act in the same direction (invasion promotion). The method does find many inhibitory links to E-genes, which dramatically increases the fit of the model on the data points. These predicted attachment signs provide information about how an E-gene's involvement in the invasion process can be tested in follow-up experiments. The model predicts that invasion can be suppressed by knocking down genes connected by stimulatory attachments or by over-expressing genes connected by inhibitory attachments.

FG-NEM recovered the network shown in Figure 5B. KRT20 and RPL32 are predicted to be equivalent. Also, the model predicts TFDP1 and DHX32 are downstream of KRT20 and RPL32. The equivalent interaction of KRT20 and RPL32 received significantly high likelihoods ($P < 0.001$) as well as a strong excitatory downstream connection to TFDP1 ($P < 0.001$).

Figure 5. Invasive colon cancer network predictions.

(A) Expression changes of selected E-genes following targeted S-gene knock-downs in HT29 colon cancer cells. Gene expression was measured in HT29 cells treated with a shRNA specifically targeting an S-gene (column of the matrix) relative to cells treated with a scrambled control shRNA (Irby et al., 2005). Colors indicate putatively inhibited E-genes (rows of the matrix) with up-regulated levels relative to control (red), activated E-genes with down-regulated levels relative to control (green), and unaffected E-genes with expression levels not significantly different from control (black). Biological replicates were available for KRT20, TFDP1, and GLS knock-downs. Genes were sorted by their attachment point and then by their LAR scores. (B) Cancer invasion network predicted by FG-NEM. For each pair of S-genes, the most likely interaction mode is shown. The same conventions used for illustrating interactions predicted for the yeast networks were used here. Some interactions were found to be significant at the 0.05 level (*) or 0.01 level (**) using a permutation test (see Methods). KRT20 and RPL32 were predicted to be equivalent and are therefore grouped together in a dashed oval. (C) Matrigel invasion assay in HT29 colon cancer cells. Genes predicted to be significantly attached to the network, CAPN12 and expressed sequence tag AA099748, resulted in a loss of the invasiveness phenotype when knocked-down by RNA interference. Genes not significantly attached to the network, MYO1G, BMPR1A, and COLEC12, did not result in significant loss of the invasive phenotype. A scrambled non-sense sequence also served as a negative control and did not result in a loss of HT29 cell invasiveness. Gene knock-downs in HT29 cells were validated by quantitative real time RT-PCR where mRNA levels of targeted genes were decreased by 70–80% compared to scrambled control shRNA-treated cells (data not shown). Data shown are the mean ± S.E. of five independent experiments performed in quadruplicate. *Significantly different from scrambled control shRNA-treated cells ($P < 0.05$) by ANOVA and post hoc Tukey test. doi:10.1371/journal.pcbi.1000274.g005

Most interactions in S-gene network are activating

Novel expanded genes that have significant effect on the invasive phenotype
Summary

• FG-NEMs: A general approach to infer an ordering of genes from knock-down phenotypes

• Strengths
  – FG-NEMs could be used in an iterative computational-experimental framework
  – Handles signed interactions between S-genes

• Weaknesses
  – Computational complexity of the inference procedure might be high
    • Required independence among E-genes
    • Model pairs of S-genes at a time
Overall conclusion

- Networks are powerful models for interpreting sequence variants or genetic perturbations as such.
- We have seen two classes of methods:
  - Extract a weighted graph based on the influence of a mutation on one node to another.
  - Probabilistic approaches.
- A systematic comparison of these two classes of methods has not been done so far.
Data integration strategies
Biological data is of many different types

- Genomics
  - Epigenomics
  - ~25,000 genes

- Transcriptomics
  - ~10^9 RNA transcripts

- Proteomics
  - ~10^8 protein

- Metabolomics
  - ~10^4 metabolites

- Phenomics
  - Exposomics
  - ~10^8 compounds

- Metagenomics
  - ~10^14 microorganisms

12 tumor types:
- Leukemia (LAML)
- Lung adenocarcinoma (LUAD)
- Lung squamous (LUSC)
- Kidney (KIRC)
- Bladder (BLCA)
- Endometrial (UCEC)
- Glioblastoma (GBM)
- Head and neck (HNSC)
- Breast (BRCA)
- Ovarian (OV)
- Colon (COAD)
- Rectum (READ)

Image credit: TCGA, Gligorevic et al., Proteomics 2015
We are getting better at collecting lots of different types of biological datasets.
Need for systematic approaches for data integration

- The approach to integrate different data types depends upon the end goal and the types of data available
- Three considerations
  - Number of samples per data type
  - Supervised or unsupervised
  - Types of measurements
    - Gene sets versus quantitative profiles
Network-based approaches for integrating data

• Network-inference based
  – Learning mixed graphical models where different variable types (different probability distribution families) represent different omic data types

• Diffusion based
  – Similarity Network Fusion (Wang et al., Nature Methods 2014)
  – MASHUP (Cho et al., Cell Systems 2016)
  – GeneMania (Mostafavi et al, Genome Biology 2008)

• Information flow based methods
  – Especially suited if we have a small number of samples
    • Max flow
    • Steiner tree
Similarity Network Fusion

• Given N different types of measurements for different individuals

• Do
  – Construct a similarity matrix of individuals for each data type
  – Integrate the networks using a single similarity matrix using an iterative algorithm
  – Cluster the network into a groups of individuals
Similarity network fusion with two data types

Similarity network fusion (Nodes are patients, edges represent similarities).
Problem definition of information flow

• Given
  – two node sets and a weighted directed network with edge weights corresponding to the flow between two nodes

• Do
  – Find the subnetwork that maximizes the flow between the two node sets
Information flow between sink to source nodes

Source nodes

Sink nodes

**Skeleton Graph:**
Potential gene network derived from a given confidence level

**Selected Signaling Pathway:**
Selected by different set of algorithms to predict the most significant and relevant paths
Information flow-based methods

• Used for integrating different types of data, as well as for examining perturbations and their effect

• Integration of different types of “omics” data
  – Min cost max flow (ResponseNet; Yeger-Lotem et al 2008)
  – Prize-collecting Steiner tree variants (Huang & Fraenkel 2009, OmicsIntegrator)
Notation

• A flow network is defined as directed graph $G=(V,E)$, with capacities for each edge
  – $V$: vertex set
  – $E$: edge set
• $s$: source node
• $t$: sink node
• $c(u,v)>0$: Capacity of edge $(u,v)$
Flow in a graph $G$

- A flow in $G$ is defined by a function $f$ that has the following properties for each edge $(u,v)$:

\[
f(u, v) \leq c(u, v) \quad \text{Capacity constraint}
\]

\[
\sum_{v \in V, v \neq s,t} f(u, v) = 0 \quad \text{Conservation of flow}
\]

- The value of a flow is defined as

\[
|f| = \sum_{v \in V} f(s, v)
\]
An example flow network

Flow network $G$

A flow of 19 on $G$

Only positive flows are shown
Max-flow problem

• Given
  – A flow network $G$, source $s$ and sink $t$

• Do
  – find a flow $f$ with maximum value

• How
  – Ford-Fulkerson algorithm
Variation: Min cost max flow

• Often the question is not to maximize flow, but to find the most efficient/least expensive way of doing this.
• In addition to the flow, there is also a cost associated with each edge.
  – For example, the cost might be inversely proportional to the edge confidence.
• So we would try to maximize the overall flow at the smallest cost.
Min cost max flow

• Define cost of each edge as $a(u,v)$
• Overall cost: $\sum_{(u,v) \in E} a(u,v)f(u,v)$
• Minimize cost while maximize flow as follows:

$$\sum_{(u,v) \in E} a(u,v)f(u,v) - \gamma \sum_{v \in V} f(s,v)$$

• This idea was used in ResponseNet tool
Alternate problem definition of information flow

• Given
  – A node set
  – A weighted network

• Do
  – Find the minimal graph connecting the nodes, where minimal is defined by the graph with the lowest total weight

• We will use a Steiner tree approach to address this problem
Steiner tree

• Let’s start by defining a Steiner tree
• Given
  – edge-weighted graph $G=\{V, E, w\}$
  – A subset $S$ of $V$
• A Steiner tree is a minimal length tree connecting $S$, including potentially intermediate nodes
• This problem is NP-complete
Steiner tree examples

$S=\{A, B, C\}$

$S=\{A, B, C, D\}$
Prize-collecting Steiner tree objective function

- $p(i)$: Define prize of node $i$ as
- $y(i)$: include a node $i$
- $a(i,j)$: Define cost of edge $(i,j)$
- $x(i,j)$: include an edge
- Constrain that subnetwork must be a tree
- PCST objective

$$\max \sum_i p(i)y(i) - \lambda \sum_{(i,j)} a(i,j)x(i,j)$$

Trade-off between cost and prize

- Solve using variety of optimization techniques
  - E.g. integer linear programming-based method (Ljubic et al, 2006)
Prize-collecting Steiner trees (PCST) connect signaling proteins to gene regulation

- Top: functional screen hits, bottom: mRNA response
- Predicts relevant nodes, paths, transcription factors
- Cannot directly predict transcriptome effect from perturbations; edges are not oriented

Image from Tuncbag et al, 2012
PCST to phosphoproteomic and transcriptomic data to find genes relevant to glioblastoma multiforme

Huang et al, 2013, Figure 2
Types of approaches

• Network-based approaches
  – Network inference
  – Similarity network fusion
  – Information flow based methods

• Matrix factorization based approaches
  – Also known as clustering/dimensionality reduction based approaches
  – Multi-omics factor analysis
  – Non-negative matrix tri-factorization
Multi-omics Factor Analysis (MOFA)

(1) Allows for missing partially overlapping datasets, (2) Based on a probabilistic model, (3) Learns sparse factors

Ricard Argelaguet et al, MSB 2019
Using MOFA for Chronic Lymphocytic Leukaemia
Non-negative Matrix Tri Factorization for predicting gene drug interactions

Web and software resources

• GeneMANIA (Network integration and diffusion-based subnetworks)
  – http://www.genemania.org
• HOTNET (Diffusion-based subnetworks)
• ResponseNet (flow network)
  – http://netbio.med.ad.bgu.ac.il/respnet/
• OmicsIntegrator (PCST)
Concluding remarks

• We have seen a suite of problems, algorithms and applications in a real setting
• These ranged from network inference, dynamic network inference, network modules, network alignment and network-based interpretation
• We saw less of
  – Integration of different types of networks
  – Experimental design for better learning of networks
• If you remain interested in these topics or would like to learn more, feel free to reach out to me.