Application of Transfer Entropy to Assess Causality in Metabolomic Networks

Markus Ferdinand Dablander

UCL CoMPLEX MRes Project Report 3

Project Supervisors:

Dr. Tomaso Aste
University College London

Dr. Fotios Drenos
University College London
University of Bristol

June 5, 2017

Abstract

The human metabolome forms a complex, ever-changing network of biochemical interactions. Being able to understand, predict and manipulate these interactions is of high interest for medical and biological researchers and much effort has been made to shed light on the hidden causal pathways governing the dynamics of metabolomic change.

In this project, we explore the potential of *transfer entropy* to discover time-directed causal interactions between pairs of metabolic measures. Transfer entropy is an asymmetric, entropy-based statistical technique to assess predictive causality between pairs of time series and has been used successfully in fields such as economics and neuroscience. We apply transfer entropy to measures of $m = 67$ metabolites obtained from $n = 1452$ participants from the Avon Longitudinal Study of Parents and Children (ALSPAC) at the approximate ages of 7, 15 and 17 years.

Our method was able to confirm several known pathways. The results of our analysis were used to construct a statistically validated network with 25 directed edges indicating the information flow between different metabolites. This network can be used to stimulate the creation of new hypotheses about unknown interactions. We emphasize the distinction between predictive and interventionist causality and discuss problems as well as promising future applications of transfer entropy and network analysis in the context of metabomolics.

Contents

Word Count: 5133.

1 Introduction

The metabolome of a single organism or part of it refers to the complete set of small-molecule metabolites that can be found within its tissue. It usually consists of a wide range of different biological compounds and includes metabolic intermediates as well as hormones and other signalling molecules. The metabolome changes dynamically and can be seen as a chemical fingerprint that contains information about the current state of the biological system at any given point in time. Metabolites are the phenotypic result of gene expression and environmental factors and provide a "snapshot" view of the metabolic state. The change of the metabolic profile of an organism is partly governed by stochastic fluctuations and partly by causal links between different metabolites.

The complicated and rich information content hidden in the metabolome can potentially be used to derive meaningful biomarkers for medical purposes and to clarify pathophysiological mechanisms leading to disease. Therefore, health researchers have a keen interest in understanding the causal relations between different metabolites; a clear understanding of these causal pathways is the most important step towards the ability to predict and eventually manipulate the dynamics of the metabolome to produce desirable outcomes.

Metabolites, however, form a complex ever-changing network of compounds that are often strongly correlated. This creates the need for technical methods that can be used to distinguish causal from non-causal associations in order to uncover the directed information flow within this network. Much of the scientific field of metabolomics is devoted to developing quantitative techniques to adress this problem.

Steuer et. al. [2003] [1] provided a short overview over metabolomic network analysis and discussed the difficulties involved in the interpretation of large numbers of undirected linear correlation coefficients between metabolites.

Numata et. al [2008] [2] noted that strong correlations are not necessarily only found between metabolomic compounds that are neighbours in the underlying reaction network. They applied computer simulations to explore the origins of correlations between non-neighbouring compounds and identified a series of cases in which such phenomena can occur.

Drenos [2017] [3] has provided a review on the integration of metabolomic and genomic information and on the use of genetics to biologically interpret metabolic associations.

Camacho et. al. [2005] [4] used *mutual information* to investigate the dependency structure among metabolite concentrations from *Arabidopsis thaliana*. Mutual information is an entropy-based measure of dependence between two random variables and is able to detect both linear and nonlinear dependence. Unfortunately, the mutual information between two random variables is nondirectional, which makes causal interpretations difficult.

In this project, we employed *transfer entropy* to analyze short metabolomic time series data from $n = 1452$ children. Transfer entropy is a powerful information-theoretic, time-directed measure of nonlinear dependence and can be interpreted as a quantification of predictive causality.

We depicted the results of our analysis in a metabolomic network that can be interpreted biologically and potentially used to create hypotheses about causal biochemical pathways. Our method is presented in detail and we discuss the advantages as well as possible difficulties associated with the data and the application of transfer entropy as a measure of causality.

2 Data

Our data was taken from the Avon Longitudinal Study of Parents and Children (ALSPAC), which is a prospective study based on a European birth cohort. The study started with the monitoring of 14541 unborn children of pregnant mothers in the Bristol area and researchers have since collected information about these children in different developmental phases of their lives. For a subset of these children, a high-throughput serum nuclear magnetic resonance (NMR) metabolomics platform was used to quantify 230 metabolic measures at different age levels. These measures represent high-resolution pictures of the metabolic states of the children at the times of measurement. The measured variables are known to be present in several metabolic pathways. They include lipoprotein lipids and subclasses, fatty acids and fatty acid compositions, glycolysis precursors and amino acids. The study website [1] provides detailed information on the data as well as a searchable dictionary that describes all available measurements.

From the ALSPAC data set, we extracted measures from $n = 1452$ children and analysed values for $m = 67$ metabolites that were acquired at $l = 3$ points in time. For each child, the measurements were taken approximately at the ages of $t_1 = 90.00$ months ($\sigma_1 = 3.60$), $t_2 = 184.82$ months ($\sigma_2 = 3.06$) and $t_3 = 212.60$ months ($\sigma_3 = 4.13$). The measurements available for the metabolite with number $k \in \{0, ..., m-1\}$ can be seen as n independent realizations

$$(x_{k,1,1}, x_{k,2,1}, x_{k,3,1}), ..., (x_{k,1,n}, x_{k,2,n}, x_{k,3,n}) \tag{1}$$

of a 3-dimensional vector of random variables

$$X_k := (X_{k,1}, X_{k,2}, X_{k,3}). \tag{2}$$

Here $X_{k,j}$ describes the level of metabolite k at age t_j. The vector X_k forms a time series of experimental measurements of length $l = 3$, whereby each component represents one age level. A complete list of the analyzed metabolites is given in appendix A.

Each of the $n = 1452$ participants can be associated with a total number of $3 * 67 = 201$ measurements. In 260 participants, at least one measurement of one metabolite at one time point was missing. In 252 of these participants less than 10 and in 8 participants 10 to 27 out of 201 measurements were missing. For each particular child, all non-missing values were used as linear regressors to replace possible missing values. The coefficients for all linear regression models were constructed on the basis of the data subset corresponding to the 1192 children for which the complete set of 201 measurements was available.

Finally, the data for each metabolite with number $k \in \{0, ..., m-1\}$ at time $j \in \{1, 2, 3\}$

$$(x_{k,j,1}, ..., x_{k,j,n}) \tag{3}$$

was transformed to normality via a rank-based inverse normal transformation [5]:

[1] www.bris.ac.uk/alspac

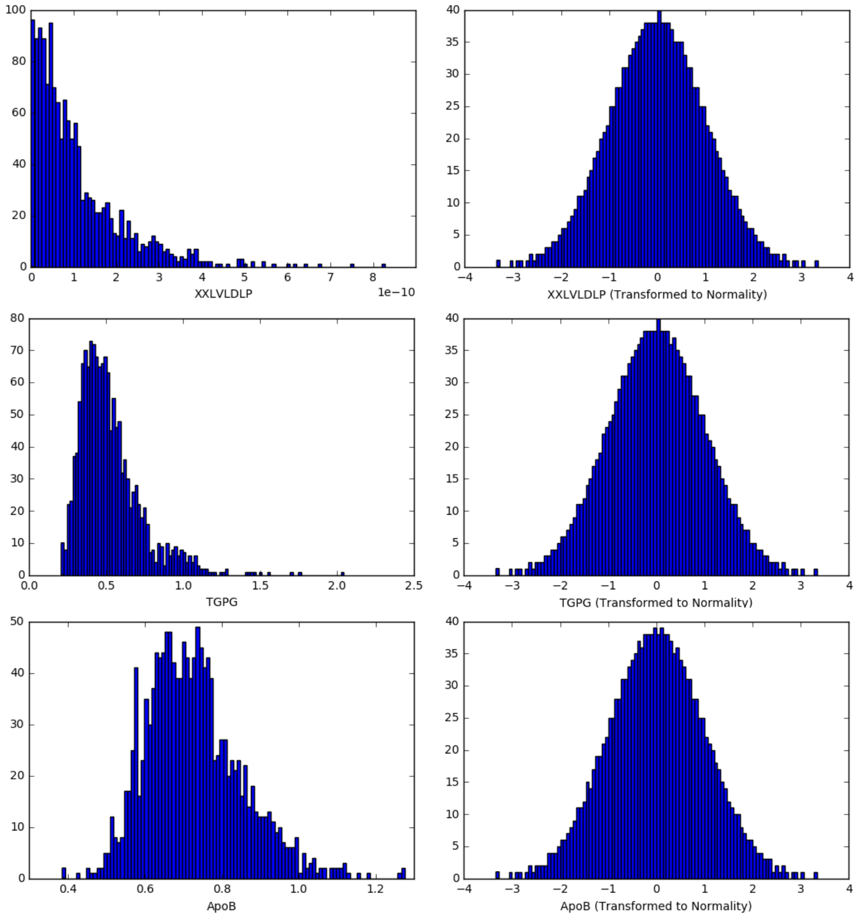

Figure 1: Measurements of XXLVLDLP, TGPG and ApoB at age t_1, before and after rank based inverse normal transformation.

$$x_{k,j,i} \;\mapsto\; \Phi^{-1}\Big(\frac{r_i - 3/8}{n + 1/4}\Big). \tag{4}$$

Here Φ is the standard normal cumulative distribution function and r_i describes the rank of $x_{k,j,i}$ when (3) is ordered according to size with the smallest value having rank 1. If $x_{k,j,i} = x_{k,j,i+1}$, then both values are assigned the same rank. A selection of histograms of metabolic measurements at time t_1 before and after transformation to normality is depicted in figure 1.

In the rest of our report, the quantities $x_{k,j,i}$ and random variables $X_{k,j}$ always refer to their associated values after replacement of missing values and transformation to normality.

3 Method

3.1 Causality

The aim of our project is using the above data to investigate the causal relationships between the measured metabolites. Therefore it is necessary to shortly discuss the term *causality.* It is not by any means clear how to precisely define the concept of causality in a general context. Characterizing ”true” causality is considered a complex philosophical problem and it is not possible to give an overview over all existing theories surrounding the debate in this report.

Though, one of the few facts that seems to be widely agreed on is that causality has a temporal direction: the cause must happen *before* the effect. Moreover, in the context of science the term causality almost always refers to one of two related, but distinct concepts: *predictive* causality and *interventionist* causality. To explain the intuitive difference between the two, let us look at two distinct, hypothetical events A and B with different possible outcomes.

- If A happens before B and information obtained from observing the outcome of A in the present can be used to improve the accuracy of predictions about the outcome of B in the future, then this is referred to as *predictive causality.*

- If A happens before B and manipulating the outcome of A by intervention influences the future outcome of B, then we speak of *interventionist causality.*

An example for predictive causality is the relationship between an oak tree loosing or keeping its leaves in the present (A) and cold or warm temperatures three months in the future (B). If the tree looses its leaves, we can interpret this as a sign that winter is probably on its way. This, however, does not represent interventionist causality, since somehow preventing the tree from loosing its leaves in autumn will not change the fact that temperatures are going to fall.

On the other hand, an example of interventionist causality is the relationship between the event of an oak tree loosing or keeping its leaves today (A) and the amount of dead leaves lying on the ground around the tree tomorrow (B). Manipulating the amount of leaves a tree is loosing today (for example by cutting them off) is likely to have a direct influence on the amount of dead leaves lying on the ground tomorrow.

Both predictive and interventionist causality may depend on the context; A might cause B in context C_1 but not in context C_2. Moreover, interventionist causality is stronger than predictive causality. If manipulation of A influences the outcome of B, then the outcome of A must contain information that can be used to improve predictions of the outcome of B.

One of the ultimate goals of metabolomics is to identify a network of *interventionist* causal relations that covers as much of the metabolome as possible. Precise, quantitative knowledge about how the manipulation of the concentration of one metabolite in the presence influences the concentration of another metabolite in the future can be expected to be an immensely powerful tool for medical research and drug design.

There are different ways to test statements about causality. Of course, a classical method to assess causality is the conduction of a controlled experiment that minimizes the effects of possible confunding variables. On the theoretical side, Lizier and Prokopenko [2010] [6] have provided an especially clear discussion on the differences between the predictive and the interventionist approach and on quantitative tools associated with both ideas. Furthermore, Bradford Hill [1965] [7] proposed a set of theoretical criteria to assess the likelihood of interventionist causality in situations where experiments are impossible or impractical. But before most methods can be applied, researchers first have to identify connections between observations whose probabilities of being causal seem to be especially high.

The powerful tools of mathematics and statistics can help to find such connections. In the following section, we will describe *transfer entropy*, a measure of a special kind of predictive causality. Since inverventionist causality implies predictive causality, predictive relations detected via transfer entropy can be seen as promising candidates for more interesting interventionist relations. This can stimulate the creation of causal hypotheses and motivate the conduction of empirical research.

3.2 Mutual Information and Transfer Entropy

Transfer entropy is a measure of information transfer between time series that was introduced by Schreiber [2000] [8] and has since been used in a variety of fields. It is based on the concept of the Shannon entropy [1948] [9] and some of its generalizations, namely *mutual information* (MI) and *conditional mutual information* (CMI).

Let

$$X, Y : \Omega \to \mathbb{R} \tag{5}$$

be two random variables. If X and Y are both discrete, then their MI is defined by

$$I(X;Y) := \sum_{y \in Y(\Omega)} \sum_{x \in X(\Omega)} \Pr(X = x, Y = y) \log_2 \left(\frac{\Pr(X = x, Y = y)}{\Pr(X = x)\ \Pr(Y = y)} \right). \tag{6}$$

If X and Y are continuous with densities f_X and f_Y and joint density $f_{X,Y}$, their MI is defined via:

$$I(X;Y) := \int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x,y) \log_2 \left(\frac{f_{X,Y}(x,y)}{f_X(x)\, f_Y(y)} \right) dx\, dy. \tag{7}$$

MI has several desirable properties that make it a suitable measure for stochastic independence. The two most important ones of these properties are:

- $I(X;Y) \geq 0$ and

- $I(X;Y) = 0 \iff X$ and Y are independent.

In particular, MI can detect linear as well as nonlinear relationships between random variables, which makes it a substantially more powerful tool for the investigation of dependency structures than the classical Pearson correlation coefficient.

Let us now consider a third discrete or continuous random variable

$$Z : \Omega \to \mathbb{R} \tag{8}$$

and a value $z \in Z(\Omega)$. In the continuous case, we assume the existence of a density f_Z for Z. We define the expression

$$I_Z(X;Y|z) \tag{9}$$

by taking $I(X;Y)$ and exchanging all probabilities Pr (resp. all densities f) by the conditional probabilities $\Pr(\ .\ |\ Z = z)$ (resp. by the conditional densities $f(\ .\ |\ Z = z)$). This can be used to speficy [2] a random variable

$$I_Z(X;Y|Z(\ )) : \Omega \to \mathbb{R}, \qquad I_Z(X;Y|Z(\ )) \mapsto I_Z(X;Y|Z(\omega)). \tag{10}$$

We can now define the conditional mutual information (CMI) of X and Y conditioned on Z as the expected value of (10):

$$I(X;Y|Z) := E(I_Z(X;Y|Z(\ ))). \tag{11}$$

The CMI of X and Y conditioned on Z can be interpreted as the average dependence between X and Y when the outcome of Z is known. Just like MI, CMI is always nonnegative.

We now have everything we need to define transfer entropy. Let

$$X := (X_1, X_2, ...) \tag{12}$$

and

$$Y := (Y_1, Y_2, ...) \tag{13}$$

be two stochastic processes in discrete time. Then the transfer entropy between X and Y at time $t \in \mathbb{N}_{\geq 1}$ is defined via:

$$T_{X \to Y, t \to t+1} := I(X_t; Y_{t+1}|Y_t). \tag{14}$$

$T_{X \to Y, t \to t+1}$ can be interpreted as the amount of uncertainty about the future value Y_{t+1} resolved by knowing the present value X_t, over and above the degree of uncertainty about the future value Y_{t+1} already resolved by its own present value Y_t. If $T_{X \to Y, t \to t+1}$ is positive, then we know that predictions about Y_{t+1} that are based on both X_t and Y_t are better than predictions of Y_{t+1} that are

[2] In the continuous case, this random variable is defined almost everywhere, since the set $\{w \in \Omega \mid f_Z(Z(w)) = 0\}$ has measure 0 and the conditional densities are defined everywhere but on this set.

solely based on Y_t. Furthermore, definition (14) can easily be generalized by considering longer time lags $(X_t, Y_t) ..., (X_{t-k}, Y_{t-k})$ instead of just X_t and Y_t when trying to predict Y_{t+1}.

Transfer entropy is a non-parametric, time-directed measure of information flow between stochastic processes and provides a quantification of the concept of predictive causality discussed in the previous section. It can be seen as a more general, non-parametric version of the well-known *Granger causality* [10] and is frequently applied in the field of time series forecasting.

Example. To illustrate the ideas from above, let us imagine two discrete-time stochastic processes $X = (X_1, X_2, ...)$ and $Y = (Y_1, Y_2, ...)$ that only take discrete values in the set $\{-1, 1\}$. In the beginning, we define

$$\Pr(X_1 = 1) = \Pr(X_1 = -1) = \Pr(Y_1 = 1) = \Pr(Y_1 = -1) = 1/2. \tag{15}$$

For all $t \in \mathbb{N}_{\geq 1}$, we define

$$X_{t+1} := X_t \tag{16}$$

and

$$\Pr(Y_{t+1} = X_t) = 2/3. \tag{17}$$

This means that both processes start with a uniform distribution on $\{-1, 1\}$. Then X continues to flip deterministically at each time step, completely autonomous from any outside influence. On the other hand, Y proceeds by copying the value of X with a chance of $2/3$ and producing the opposite value of X with a chance of $1/3$. If we compute the transfer entropy between X and Y for $t = 1$, we find that

$$T_{X \to Y, 1 \to 2} = I(X_1; Y_2 | Y_1) = I(X_1; Y_2) = \frac{2 \log_2(4/3) + \log_2(2/3)}{3} > 0 \tag{18}$$

while

$$T_{Y \to X, 1 \to 2} = I(Y_1; X_2 | X_1) = 0. \tag{19}$$

More generally, one can show that both $T_{X \to Y, t \to t+1} > 0$ and $T_{Y \to X, t \to t+1} = 0$ hold for all $t \in \mathbb{N}_{\geq 1}$.

This is in perfect alignment with our intuition of predictive causality. If we know the current value of X, then knowing the current value of Y produces no advantage in predicting the next value of X. On the other hand, knowing the current value of both X and Y is definitely better than only knowing the current value of Y when trying to predict the next value of Y. A detailed mathematical treatment of this example as well as an introduction to transfer entropy and its applications to a variety[3] of complex systems can be found in the book by Bossomaier et. al [2016] [11].

[3] Cellular automata, spin models, oscillators, random boolean networks,...

3.3 Network Construction

Let us now consider two distinct metabolic time series X_k and X_r of the form described in (2) with $k, r \in \{0, ..., m-1\}$. We are interested in determining the predictive value of X_k for X_r. The transfer entropy at $t = 1$

$$T_{X_k \to X_r, 1 \to 2} = I(X_{k,1}, X_{r,2} \mid X_{r,1}) \tag{20}$$

measures the information in $X_{k,1}$ and $X_{r,2}$ that cannot be found in $X_{r,1}$. We remember that, if (5) is positive, this means that predictions of $X_{r,2}$ that are based both on $X_{k,1}$ and $X_{r,1}$ are better than predictions of $X_{r,2}$ that are merely based on $X_{r,1}$. In an analogous way we can interpret the transfer entropy for $t = 2$

$$T_{X_k \to X_r, 2 \to 3} = I(X_{k,2}, X_{r,3} \mid X_{r,2}). \tag{21}$$

To summarize both quantities in order to obtain a global measure of predictive causality, we combine (20) and (21) via:

$$T_{X_k \to X_r} := \frac{T_{X_k \to X_r, 1 \to 2} + T_{X_k \to X_r, 2 \to 3}}{2}.^{4} \tag{22}$$

We want to use this measure to construct a causal network for our metabolites. Unfortunately, using a finite amount of samples to numerically calculate information-theoretic quantities associated with continuous random variables is a non-trivial problem. In the discrete case, the computation of mutual information and related measures is more straightforward, which is why many MI algorithms are based on discretization of continuous data via binning. Uniform binning, however, can lead to a substantial bias and loss of information and several suggestions have been made to find optimal binning methods [12].

In our case, the metabolite concentrations are continuous and in order to keep the maximum amount of information, we avoided discretization of the variables. An estimate $T_{n, X_k \to X_r}$ for $T_{X_k \to X_r}$ was computed using `Python` along with the `Non-Parametric Entropy Estimation Toolbox (NPEET)` [5]. The algorithms in the NPEE toolbox are based on the mutual information estimator proposed by Kraskov et. al [2004] [13]. Unlike conventional estimators that are based on binning, the Kraskov MI estimator is based on entropy estimates from k-nearest neighbour distances. It therefore avoids the theoretical problems associated with discretization.

To construct a causal network, we interpreted each of the $m = 67$ metabolites in our analysis with a vertex in a graph and used formula (22) to compute transfer entropy weights for all possible $67 * 66 = 4422$ directed edges in this network. In the next section, we will discuss the significance analysis of those edgeweights.

[4] Of course, this definition ignores the asymmetry in the time gaps corresponding to $T_{X_k \to X_r, 1 \to 2}$ and $T_{X_k \to X_r, 2 \to 3}$. In fact, the gap $t_2 - t_1$ is more than three times the size of $t_3 - t_2$. A possible way to account for this would be to experiment with different weights for both measures.

[5] https://www.isi.edu/ gregv/npeet.html

3.4 Significance Analysis

Of course, many of the weights in the complete, directed graph constructed in the previous section are positive by mere coincidence while the "true" underlying value of $T_{X_k \to X_r}$ is in fact 0. This produces the need for a proper significance analysis to statistically validate all edges and find candidates for those edges whose weights are truly positive.

Although the Kraskov algorithm offers several numerical advantages in estimating transfer entropy, a rigorous significance analysis of the results for a large number of statistical tests with different underlying null distributions proves to be difficult without discretization of the data. Barnet and Bossomaier [2012] [14] showed that conventional discrete estimators of transfer entropy are asymptotically χ^2-distributed if the associated transfer entropy value is 0; a result that can be readily exploited to perform significance analysis. Unfortunately, we could not find comparable results about the null distributions of transfer entropy estimators that are based on k-nearest-neighbour-algorithms instead of binning. In particular, the relationship between the null distributions of these estimators and the distributions of the time series values they take as input is unclear.

This implies that the optimal way to perform significance analyis in our case would be to derive an empirical estimate $\hat{F}_{k,r,n}$ for the cumulative null distribution function $F_{k,r,n}$ *for each pair* of metabolites (k, r). Here, $F_{k,r,n}$ describes the distribution of our n-sample estimate $T_{n,X_k \to X_r}$ under the null hypothesis that $T_{X_k \to X_r} = 0$. We remember that we are computing our n-sample estimate for $T_{X_k \to X_r}$ on the basis of the data for X_k

$$(x_{k,1,1}, x_{k,2,1}, x_{k,3,1}), ..., (x_{k,1,n}, x_{k,2,n}, x_{k,3,n}) \tag{23}$$

and the data for X_r

$$(x_{r,1,1}, x_{r,2,1}, x_{r,3,1}), ..., (x_{r,1,n}, x_{r,2,n}, x_{r,3,n}). \tag{24}$$

If we randomly redefined the order of the vectors in (23) and (24), computed $T_{n,X_k \to X_r}$ for this shuffled data set and repeated this process a sufficiently large number of times, then we could derive an empirical estimate $\hat{F}_{k,r,n}$ for $F_{k,r,n}$. This estimate could then be used to assign a p-value to the edge (k, r) in our network.

Unfortunately, this method is precise but numerically very expensive and it would have taken too long to compute all empirical null distributions $\hat{F}_{k,r,n}$ for all distinct pairs (k, r) considering the time and computational resources available for this project. We therefore simplified the significance analysis by assuming an identical distribution for all 3-dimensional time series random vectors $X_1, ..., X_m$. Since all $(x_{k,j,1}, ..., x_{k,j,n})$ were transformed to a standard normal distribution, all components of the vectors $X_1, ..., X_m$ are already Gaussian with mean 0 and variance 1. The claim that all $X_1, ..., X_m$ follow the same distribution, hoewever, is stronger since it implies that not only the marginal distributions of all components of all $X_1, ..., X_m$ are the same but also the inner dependency relations among the components of each X_k.

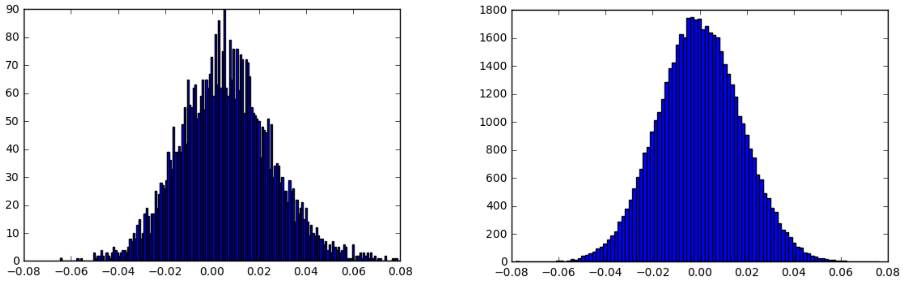

Figure 2: Histogram of all 4422 edgeweights computed with $T_{n,X_k \to X_r}$ from our original data (left) and simulated null distribution $\hat{F}_n$ under the assumption that $T_{X_k \to X_r} = 0$ (right). Taking a look at the interval $[0.05, 0.08]$ confirms that the left distribution has higher probabilities for larger values of $T_{n,X_k \to X_r}$, which is to be expected.

This assumption makes it possible to assume that $F_{k,r,n} := F_n$ does not depend on k and r. In particular, this means that it is sufficient to only estimate a single empirical null distribution $\hat{F}_n$ in order to assign p-values to our edgeweights.

We estimated F_n by randomly drawing two distinct sequences of metabolite measurements

$$(x_{k,1,1}, x_{k,2,1}, x_{k,3,1}), ..., (x_{k,1,n}, x_{k,2,n}, x_{k,3,n}) \tag{25}$$

and

$$(x_{r,1,1}, x_{r,2,1}, x_{r,3,1}), ..., (x_{r,1,n}, x_{r,2,n}, x_{r,3,n}). \tag{26}$$

from our data set, shuffling the vectors in these sequences to destroy possible interdependence and then computing $T_{n,X_k \to X_r}$ for these shuffled sequences. This step was repeated 50.000 times to make $\hat{F}_n$ sufficiently smooth. The resulting null distribution is depicted in figure 2 and compared to the distribution obtained from computing $T_{n,X_k \to X_r}$ for our original non-shuffled data set. Figure 2 strongly suggests that $T_{n,X_k \to X_r}$ follows a Gaussian distribution under the null hypothesis. We can also observe that unlike $T_{X_k \to X_r}$ our estimate $T_{n,X_k \to X_r}$ can take negative values.

After using $\hat{F}_n$ to compute p-values for all edgeweights, we used the Benjamini-Hochberg-Yekutieli procedure to control for false discoveries. The *Benjamini-Hochberg-Yekutieli procedure* offers a simple way to control the expected amount of false discoveries (type 1 errors) in a large series of statistical tests. To illustrate the underlying algorithm, let us consider a set of null hypotheses

$$H_1, ..., H_m \tag{27}$$

with corresponding p-values

$$p_1, ..., p_m. \tag{28}$$

W.l.o.g. we assume that these p-values are ordered according to size:

$$p_1 < ... < p_m. \tag{29}$$

We now define a threshold function $t : \mathbb{N}_{\geq 1} \times \mathbb{N}_{\geq 1} \times [0, 1] \to \mathbb{R}$ via

$$t(k, m, \alpha) = \frac{k\alpha}{m \sum_{i=1}^{m} \frac{1}{i}}. \tag{30}$$

and a threshold index

$$k_{max}(m, \alpha) := \max\{k \in \{1, ..., m\} \mid p_k \leq t(k, m, \alpha)\}. \tag{31}$$

If we reject the first $H_1, ..., H_{k_{max}(m,\alpha)}$ of our m null hypotheses, the expected fraction of falsely rejected hypotheses is smaller or equal to $\alpha \in [0, 1]$. A detailed theoretical justification of this technique can be found in the article by Benjamini and Yekutieli [2001] [15].

`Python` was used to implement the Benjamini-Hochberg-Yekutieli procedure and apply it to the transer entropy estimates in order to select edges in our network at a false discovery rate of $\alpha = 0.3$. In the last section, we will take a look at this network.

4 Metabolomic Network and Discussion

4.1 Graph

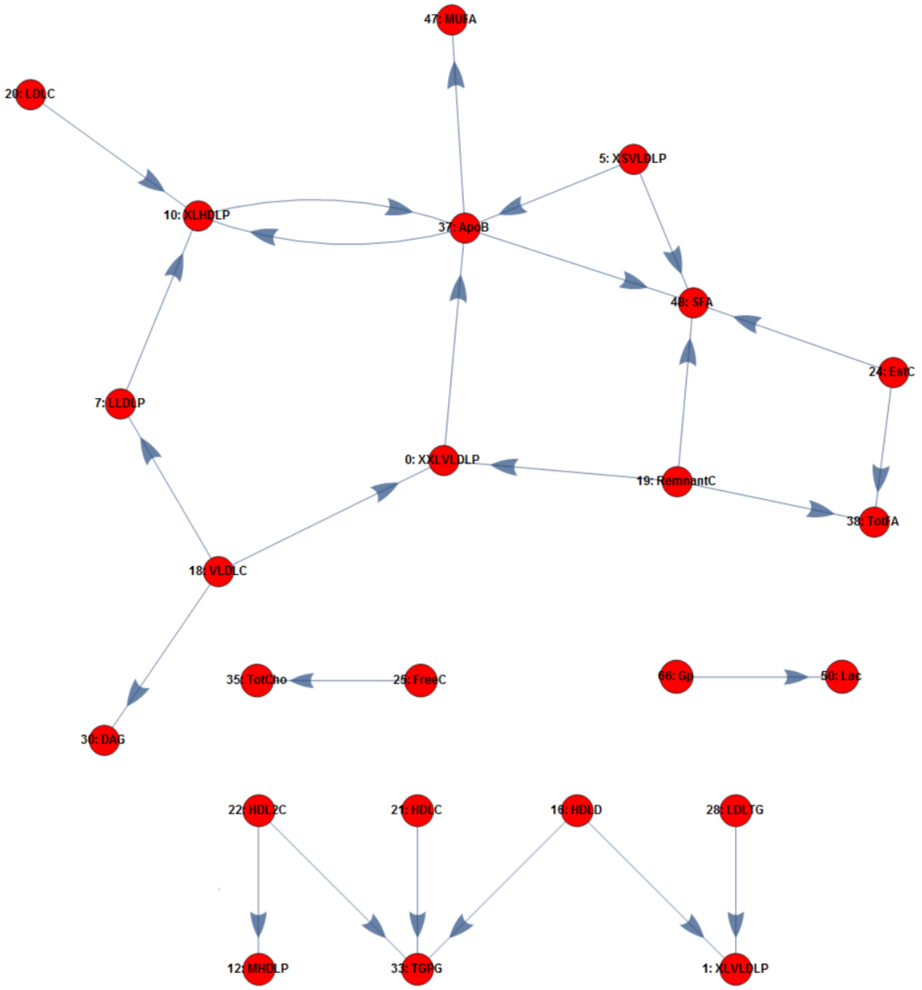

Figure 3: Metabolomic Network.

The metabolomic network shown in figure 3 consists of 24 metabolites that are connected by 25 directed edges. The plot was made with `Wolfram Mathematica` and the precise edgeweights as well as their associated *p*-values can be found in Appendix *B*. There is an edge from metabolite *A* to metabolite *B*, if the transfer entropy from *A* to *B* is significantly bigger than 0, which implies that *A* can be used to predict *B*. The network consists of 4 components. The vertex ApoB has 6 links associated with it (degree 6) and seems to hold the most central position in the network, followed by SFA with 4 links (degree 4).

Our method was able to confirm several known metabolic pathways. The association between esterified cholesterol (24) and both total fatty acids (38) and saturated fatty acids (48) are known (but probably small). Similarly, there are known links between XXLVLDLP (0) and ApoB (37) as well as between XSVLDLP (5) and ApoB (37).

Let us now for example take a look at the loop between XLHDLP (10) and ApoB (37). One imaginable explanation for this structure is that this loop is caused by negative time-lagged linear correlation - if one measure goes up, then the other measure goes down after a while and vice versa. This kind of time-lagged linear coupling leads to oscillating behaviour.

Morever, the metabolite VLDLC (18) has 3 outgoing links but 0 incoming links. One could guess that might be a sign that $VLDLC$ holds an especially high place in the causal hierarchy of the metabolic network, which means that it tends to influence other measures while having a high degree of independence.

Of course the above ideas are highly speculative. The predictive causal relationships depicted in figure 3 are promising *candidates* for interventionist causal relationships. The purpose of the metabolomic network constructed in this project is to form the basis for the formulation of hypotheses about such interventionist connections. It can be used as a tool by biological researchers to stimulate and guide the direction of future research.

4.2 Conclusion and Further Thoughts

The biggest challenge to the application of transfer entropy to the data available to us is the unusual length and distribution of the time lags between the measurements. The metabolome is an ever-changing system and it seems likely that transfer-entropy-based measures loose most of their exploratory powers if the time spans between measurements are several years.

Moreover, there are no exact mathematical results available about how the rank-based inverse normal transformation applied in section 2 affects the transfer entropy values. Since the normal transformation preserves at least part of the interdependence structure between different metabolic time series, it seems intuitive that the changes in transfer entropy should be small. Still, the exact change of transfer entropy under normal transformations remains to be investigated rigorously and provides an opportunity for further mathematical research. [6]

[6] It is known that mutual information values are invariant under homeomorphisms. It might be possible to extend weaker forms of this result to CMI under certain normal transformations.

In spite of these obstacles, transfer entropy has proven to be successful in uncovering several known and seemingly new metabolic relationships. This is a clear encouragement for further applications of information-based techniques in the field of metabolomics.

A promising future research project could be to apply our measure to a data set with much smaller and uniformly distributed time lags in the scale of days or hours instead of years. This data structure could help to exploit the full powers of transfer entropy. It seems very probable that this would lead to a much richer network with more highly significant edges at a smaller false discovery rate α.

An even further step could be to compare the constructed metabolomic network with an equivalent random network. By using computer simulations, one could construct a large number of random graphs in order to derive empirical null distributions for associated strutural descriptors (clustering coefficient, average path length, entropy of degree distribution, small-worldness,...). By using these null distributions to detect significant deviations from random structures in the metabolomic network, one could get insights into its idiosyncratic topological characteristics. These stuctural characteristics could provide insights into the global organization of the network and help us to understand its working mechanisms on a biological level.

5 Appendixes

5.1 A: List of Analyzed Metabolites

- (0) XXLVLDLP: Concentration of chylomicrons and extremely large VLDL particles (mol/l)
- (1) XLVLDLP: Concentration of very large VLDL particles (mol/l)
- (2) LVLDLP: Concentration of large VLDL particles (mol/l)
- (3) MVLDLP: Concentration of medium VLDL particles (mol/l)
- (4) SVLDLP: Concentration of small VLDL particles (mol/l)
- (5) XSVLDLP: Concentration of very small VLDL particles (mol/l)
- (6) IDLP: Concentration of IDL particles (mol/l)
- (7) LLDLP: Concentration of large LDL particles (mol/l)
- (8) MLDLP: Concentration of medium LDL particles (mol/l)
- (9) SLDLP: Concentration of small LDL particles (mol/l)
- (10) XLHDLP: Concentration of very large HDL particles (mol/l)
- (11) LHDLP: Concentration of large HDL particles (mol/l)
- (12) MHDLP: Concentration of medium HDL particles (mol/l)
- (13) SHDLP: Concentration of small HDL particles (mol/l)
- (14) VLDLD: Mean diameter for VLDL particles (nm)
- (15) LDLD: Mean diameter for LDL particles (nm)
- (16) HDLD: Mean diameter for HDL particles (nm)
- (17) SerumC: Serum total cholesterol (mmol/l)
- (18) VLDLC: Total cholesterol in VLDL (mmol/l)
- (19) RemnantC: Remnant cholesterol (non-HDL, non-LDL -cholesterol) (mmol/l)
- (20) LDLC: Total cholesterol in LDL (mmol/l)
- (21) HDLC: Total cholesterol in HDL (mmol/l)
- (22) HDL2C: Total cholesterol in HDL2 (mmol/l)
- (23) HDL3C: Total cholesterol in HDL3 (mmol/l)

- (24) EstC: Esterified cholesterol (mmol/l)
- (25) FreeC: Free cholesterol (mmol/l)
- (26) SerumTG: Serum total triglycerides (mmol/l)
- (27) VLDLTG: Triglycerides in VLDL (mmol/l)
- (28) LDLTG: Triglycerides in LDL (mmol/l)
- (29) HDLTG: Triglycerides in HDL (mmol/l)
- (30) DAG: Diacylglycerol (mmol/l)
- (31) DAGTG: Ratio of diacylglycerol to triglycerides
- (32) TotPG: Total phosphoglycerides (mmol/l)
- (33) TGPG: Ratio of triglycerides to phosphoglycerides
- (34) PC: Phosphatidylcholine and other cholines (mmol/l)
- (35) TotCho: Total cholines (mmol/l)
- (36) ApoA1: Apolipoprotein A-I (g/l)
- (37) ApoB: Apolipoprotein B (g/l)
- (38) TotFA: Total fatty acids (mmol/l)
- (39) FALen: Estimated description of fatty acid chain length, not actual carbon number
- (40) UnsatDeg: Estimated degree of unsaturation (fatty acids)
- (41) DHA: 22:6, docosahexaenoic acid (mmol/l)
- (42) LA: 18:2, linoleic acid (mmol/l)
- (43) CLA: Conjugated linoleic acid (mmol/l)
- (44) FAw3: Omega-3 fatty acids (mmol/l)
- (45) FAw6: Omega-6 fatty acids (mmol/l)
- (46) PUFA: Polyunsaturated fatty acids (mmol/l)
- (47) MUFA: Monounsaturated fatty acids; 16:1, 18:1 (mmol/l)
- (48) SFA: Saturated fatty acids (mmol/l)
- (49) Glc: Glucose (mmol/l)
- (50) Lac: Lactate (mmol/l)

- (51) Pyr: Pyruvate (mmol/l)
- (52) Cit: Citrate (mmol/l)
- (53) Ala: Alanine (mmol/l)
- (54) Gln: Glutamine (mmol/l)
- (55) His: Histidine (mmol/l)
- (56) Ile: Isoleucine (mmol/l)
- (57) Leu: Leucine (mmol/l)
- (58) Val: Valine (mmol/l)
- (59) Phe: Phenylalanine (mmol/l)
- (60) Tyr: Tyrosine (mmol/l)
- (61) Ace: Acetate (mmol/l)
- (62) AcAce: Acetoacetate (mmol/l)
- (63) bOHBut: 3-hydroxybutyrate (mmol/l)
- (64) Crea: Creatinine (mmol/l)
- (65) Alb: Albumin (signal area)
- (66) Gp: Glycoprotein acetyls, mainly a1-acid glycoprotein (mmol/l)

5.2 B: Significant Transfer Entropy Values

- (0) XXLVLDLP —> (37) ApoB: 0.0647875442735 (p = 0.00016)
- (5) XSVLDLP —> (37) ApoB: 0.068632551462 (p = 0.0001)
- (5) XSVLDLP —> (48) SFA: 0.0771489669164 (p = 0.0)
- (7) LLDLP —> (28) LDLTG: 0.0691846747587 (p = 8.00000000001e-05)
- (10) XLHDLP —> (1) XLVLDLP: 0.073493405603 (p = 4e-05)
- (16) HDLD —> (1) XLVLDLP: 0.0673378661871 (p = 0.00012)
- (16) HDLD —> (33) TGPG: 0.065771643814 (p = 0.00016)
- (18) VLDLC —> (0) XXLVLDLP: 0.0740876145966 (p = 4e-05)
- (18) VLDLC —> (7) LLDLP: 0.0655098965212 (p = 0.00016)
- (18) VLDLC —> (30) DAG: 0.0786233872699 (p = 0.0)
- (19) RemnantC —> (0) XXLVLDLP: 0.0775333150789 (p = 0.0)
- (19) RemnantC —> (38) TotFA: 0.0711103186347 (p = 4e-05)
- (19) RemnantC —> (48) SFA: 0.0674880674173 (p = 0.00012)
- (20) LDLC —> (28) LDLTG: 0.065132488548 (p = 0.00016)
- (21) HDLC —> (33) TGPG: 0.0646811919685 (p = 0.00016)
- (22) HDL2C —> (12) MHDLP: 0.0703209292886 (p = 4e-05)
- (22) HDL2C —> (33) TGPG: 0.0884026728494 (p = 0.0)
- (24) EstC —> (38) TotFA: 0.0690249898816 (p = 8.00000000001e-05)
- (24) EstC —> (48) SFA: 0.0822164704488 (p = 0.0)
- (25) FreeC —> (35) TotCho: 0.066124668097 (p = 0.00016)
- (28) LDLTG —> (37) ApoB: 0.0675151704785 (p = 0.00012)
- (37) ApoB —> (28) LDLTG: 0.0660390397262 (p = 0.00016)
- (37) ApoB —> (47) MUFA: 0.0644154638847 (p = 0.00018)
- (37) ApoB —> (48) SFA: 0.0895360178805 (p = 0.0)
- (66) Gp —> (50) Lac: 0.0671283342601 (p = 0.00012)

References

[1] Ralf Steuer, Jürgen Kurths, Oliver Fiehn, and Wolfram Weckwerth. Interpreting correlations in metabolomic networks, 2003.

[2] Jorge Numata and Oliver Ebenhöh. Measuring correlations in metabolomic networks with mutual information. *Genome Informatics*, 20:112–122, 2008.

[3] Fotios Drenos. Mechanistic insights from combining genomics with metabolomics. *Current opinion in lipidology*, 28(2):99–103, 2017.

[4] Diogo Camacho, Alberto De La Fuente, and Pedro Mendes. The origin of correlations in metabolomics data. *Metabolomics*, 1(1):53–63, 2005.

[5] T Mark Beasley, Stephen Erickson, and David B Allison. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behavior genetics*, 39(5):580, 2009.

[6] Joseph T Lizier and Mikhail Prokopenko. Differentiating information transfer and causal effect. *The European Physical Journal B-Condensed Matter and Complex Systems*, 73(4):605–615, 2010.

[7] Austin Bradford Hill. The environment and disease: association or causation?, 1965.

[8] Thomas Schreiber. Measuring information transfer. *Physical review letters*, 85(2):461, 2000.

[9] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[10] Lionel Barnett, Adam B Barrett, and Anil K Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical review letters*, 103(23):238701, 2009.

[11] Terry Bossomaier, Lionel Barnett, Michael Harré, and Joseph T Lizier. *An introduction to transfer entropy: information flow in complex systems*. Springer, 2016.

[12] Georges A Darbellay and Igor Vajda. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory*, 45(4):1315–1321, 1999.

[13] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[14] Lionel Barnett and Terry Bossomaier. Transfer entropy as a log-likelihood ratio. *Physical review letters*, 109(13):138105, 2012.

[15] Yoav Benjamini and Daniel Yekutieli. *Annals of statistics*, pages 1165–1188, 2001.

List of Figures