Higher order analysis of gene correlations by tensor decomposition

Farzane Yahyanejad
Department of Computer Science
University of Illinois at Chicago
fyahya2@uic.edu

March 8, 2019

Abstract

This study advances our understanding of inter- and intra-pathways higher order signaling in the cellular system and it leads to new discovery of multiple intracellular structures in signal transduction pathways in *yeast Saccharomyces*. We present a new tensor decomposition algorithm in reconstructing the pathways based on higher correlations among genes that compose a cellular system. The higher order gene correlation (HOGC) analysis has the power to elucidate gene's higher interaction dependencies which has been barely understood. Recent studies i.e. [24] have experimentally revealed that multiple signaling proteins, yet sometimes infinite, may assemble to meaningful structure to transmit a receptor activation information. In this paper we reveal 3-order genomic correlations among significant component of the cellular system. This is the first time such a systematic and computational model provided for analysis of higher order correlations among genes. We use new fast algorithm to formulate a genes $\times$ genes $\times$ genes decorrelated rank-1 sub-tensors (complexes) which can be associated with functionally independent pathways. Then we model higher order tensor decomposition $\mathcal{T} \in R^{d^{\otimes 4}}$ which is constructed by K tensors of genes $\times$ genes $\times$ genes. Each new tensor is constructed by an orthogonal projection of data signal onto a designated basis signal to keep common sub-tensors in both signals. Our model for decomposing tensor order-4 approximates series of tensors as linear components of decorrelated rank-1 sub-tensors over tensor of order-3 and rank-3 triplings among sub-tensors. The linear components represent intra-pathway in cell signaling and triplings implicate inter-pathways higher order signaling. Through structural studies of inter- and intra- higher order signaling pathways, we uncover different scenario that involves triple formation of signaling proteins into higher order signaling machines for transmission of receptor activation information to cellular responses.

Keywords— Tensor decomposition, Signal processing, Higher order correlation, Signal transduction, Yeast *Saccharomyces*.

1 Introduction and motivation

The biological functioning and life of a cellular system is controlled by signaling and energy transfer interactions among its numerous constituents such as proteins, RNAs, DNAs and other small molecules, and usually involve a *cascade* of biochemical reactions or other physical interactions among these constituents. Consequently, cellular systems generate genomic signals, such as mRNA expression and DNA-bound proteins' occupancy levels, that can be measured experimentally using various methods such as DNA microarrays. Often, biologists model such a cellular system by presenting the interaction data in the form of a network diagram (some type of graph), *optionally* with some mathematical formulation of its dynamics. Usual mathematical formulations of the dynamics typically assume that each node u in the network has an associated "state" variable (representing, for example, concentration of the corresponding protein) that is

a function of the "time" variable t , and describes how the *value* of this variable at a node ("state" of the node) depends on the state of the nodes interacting with it. A major drawback of using such graph-theoretic tools on a single network diagram lies in ignoring the time or ignoring the *higher-order* correlations of the interactions which may lead to *inaccurate* or *incomplete* analysis. For example, a network diagram only encodes pairwise correlations of node state variables, and thus cannot represent a joint k -way correlation among k state variables for any $k > 2$. If precise equations of time evolutions of state variables are given then we could of course completely ignore the network diagrams and work with the given equations, but then we lose the advantage of employing graph-theoretic tools and instead fall back on analysis techniques which are often *hard* to employ effectively because of difficulties of estimating precise equations and the non-trivial *non-linear* natures of these equations. The main topic of this paper is to study higher order (*beyond pairwise*) genomic correlations among significant components (*e.g.*, genes) of the cellular systems based on the measured signaling data. This is the first time such a mathematical model is given to uncover structural assemblies of higher order correlations among genes. Higher-order correlations among genes conditioned by the biological and experimental settings in which they are observed are *not* yet very well understood because of the difficulties to detect them genetic mapping studies, and not too many examples of such correlations have been described in the literature. However, evidences from model organisms suggest that these higher-order correlations among genes contribute *frequently* to genetic studies [9, 12, 15, 22, 26]. Informally, the main goal of this paper is to use powerful *tensor analysis methods* to provide the foundations of *systematic* and *computationally efficient* approaches to distinguish significant higher order correlations among the elements of biological systems.

For genes $\times$ genes networks, several previous researchers have proposed and analyzed models that separate genome-scale signals (*e.g.*, mRNA expression) into mathematically defined patterns that correlate with the independent biological and experimental processes and cellular states that compose the signals [1–3]. In this paper, we take this approach one step further to detect three-way correlations between the components of the biological system using tensor decompositions. These higher relations among the activities of genes may be conditioned by the biological and experimental settings in which they are observed. We call them higher order signaling pathways. For example, the mRNA expression patterns of the yeast *Saccharomyces* genes *CIS3*, *SWI4* and *HTA1* have higher order correlations during cell-cycle progression. A single genomic signal matrix of correlations cannot describe differences in such higher level relations. More specifically, in this paper we describe the usage of higher order tensor decomposition via fast spectral algorithms (based on power iterations) to reconstruct pathway-dependent relations among the genes of a cellular system from higher-order correlations represented as symmetric tensors from measured genomic signals. Our method computes a genes $\times$ genes $\times$ genes tensor from signaling data, and decomposes the tensor to a *series* of genes $\times$ genes $\times$ genes decorrelated rank-1 sub-tensors. The rank-3 triplings among these sub-tensors are associated with higher-order inter-pathway correlations of the signaling pathways. To accurately approximate the higher-order correlations of the genes, we use a series of only those sub-tensors that are *common* to both signals by using the Moore-Penrose pseudo-inverse projection; this projects the signaling data onto a designated "basis" signal and simulate observations of only those pathways that are manifest under both sets of the biological and experimental conditions in which the data and basis signals are measured. Sub-tensors and tripling of sufficiently high significance represent independent pathways or higher-order correlations among them common to all or exclusive to a subset of the signals, and discretized sub-tensors and their triplings may highlight unknown differentials (*i.e.*, pathway-dependent relations) between genes. We illustrate our new fast decomposition algorithm and the new approach for finding higher-order correlations by analyzing the mRNA expression data of all common genes with manifested stages in the cell cycle of yeast (*S. cerevisiae*) [6] and DNA-binding data of the yeast transcription factors that are involved in cell-cycle, development, and biosynthesis programs [16] in pheromone-response classifications.

2 Method

Following widely used conventions, we will denote tensors, matrices and column vectors by calligraphic uppercase letters (e.g., $\mathcal{X}$), boldface lowercase letters (e.g., $\mathbf{x}$) and non-bold lowercase letters with overhead arrows (e.g., $\vec{x}$), respectively. Individual elements will be denoted by the corresponding (non-bold) lowercase letter with appropriate indices, e.g., $x_{i,j}$ for matrix $\mathbf{x}$. We assume that the reader is familiar with the basic concepts and definitions associated with tensor analysis such as in [13, 14, 18].

2.1 Mathematical Models

Let the *symmetric* tensor $\mathcal{T}^{(1)}$ of size $N \times N \times N$, where N is the number of genes, tabulate the non-directional tensor of three-way correlations among the genes of a cellular system. We say tensor $\mathcal{T}^{(1)}$ is *symmetric* if the entries $(\mathcal{T}^{(1)}_{i,j,k})$ are invariant under permuting the indices. The tensor $\mathcal{T}^{(1)}$ is computed from a genome-scale signal, referred to as the data signal, of mRNA expression levels (or similar other data). Assume that these expression levels come from measurements conducted in a set of M_1 samples of the cellular system and are subsequently tabulated as a $N \times M_1$ matrix $\mathbf{a}^{(1)}$, such that each entry of the tensor is:

$$\mathcal{T}^{(1)}_{i,j,k} = \sum_{m=1}^{M_1} \mathbf{a}^{(1)}_{i,m} \mathbf{a}^{(1)}_{j,m} \mathbf{a}^{(1)}_{k,m}$$

To find significant gene components of the correlation tensor, we decompose the tensor to M_1 series of rank-1 sub-tensors of size $N \times N \times N$ by a fast spectral algorithm (to be described later):

$$\mathcal{T}^{(1)} = \sum_{m=1}^{M_1} \lambda_{1,m} \overrightarrow{\alpha_{1,m}}^{\otimes 3} \quad (1)$$

where $\lambda_{1,m}$ is a scalar, $\otimes$ denotes the tensor product operation, $\overrightarrow{\alpha_{1,m}}^{\otimes 3}$ is a short-hand notation for $\overrightarrow{\alpha_{1,m}} \otimes \overrightarrow{\alpha_{1,m}} \otimes \overrightarrow{\alpha_{1,m}}$ and the m^{th} sub-tensor $\overrightarrow{\alpha_{1,m}}^{\otimes 3}$ is the 3rd order tensor product of the m^{th} gene component $\overrightarrow{\alpha_{1,m}}$. The M_1 gene-values $\lambda_{1,m}$'s and M_1 gene-components $\overrightarrow{\alpha_{1,m}}$'s define the M_1 non-negative higher-order gene level of gene expressions such that the expression of the m^{th} gene-value in the m^{th} gene-component is the m^{th} gene expression of subspace of $\overrightarrow{\alpha_{1,m}}^{\otimes 3}$. The significance of the m^{th} sub-tensor $(\alpha_{1,m} \otimes \alpha_{1,m} \otimes \alpha_{1,m})$ is indicated by the m^{th} "fraction of gene-value" $s_{1,m} = \lambda_{1,m} / (\sum_{m=1}^{M_1} \lambda_{1,m})$ which is the higher expression correlation captured by m^{th} sub-tensor relative to that captured by all sub-tensors. Each sub-tensor is decorrelated of all other sub-tensors since by our algorithm we find orthogonal components under a small amount of error.

3 Results

Tensor order 3 decomposition

In this part we give an algorithm that finds gene-components and gene-values in (1) with presence of very small spectral norm error. Our algorithm is inspired by orthogonal tensor decompositions proposed by Anandkumar et al. [4] and Ma et al. [11]. In general, we want to recover the gene-values $\lambda_{k,m}$ and gene-components $\overrightarrow{\alpha_{k,m}}$ for M_κ samples of cellular system by finding a certain orthogonal tensor decomposition of every symmetric tensor $\mathcal{T}^{(\kappa)} \in \mathbb{R}^{N^{\otimes 3}}$ which is a 3rd order tensor over $\mathbb{R}^N$ such that $\mathcal{T}^{(\kappa)} = \sum_{i,j,k} T_{i,j,k}^{(\kappa)} \otimes e_i \otimes e_j \otimes e_k$ where $e_1, e_2, \dots, e_N$ is the standard basis of $\mathbb{R}^N$.

Decomposing general tensors is a delicate issue, tensors may not even have unique decomposition under a mild non-degeneracy condition. An orthogonal decomposition of a symmetric tensor $\mathcal{T}^{(\kappa)}$ is a series

Algorithm 1 Fast spectral algorithm for 3^{rd} order tensor decomposition

- 1: **Input** $\hat{T} \in \mathbb{R}^{N \times N \times N}$
- 2: Draw $U^{(0)}$ uniformly at random from $\mathbb{R}^{N \times N}$
- 3: **for** $m = 1$ to $\log N$ **do**
- 4:

$$U^{(m)} = \frac{(Id \otimes Id \otimes ((Id \otimes Id \otimes U^{(m-1)})\hat{T})^T)\hat{T}}{\|(Id \otimes Id \otimes (Id \otimes Id \otimes U^{(m-1)})\hat{T})^T\hat{T}\|} \quad (2)$$

- 5: Initialize $\vec{\alpha}^{(0)}$ the top eigenvector of $U^{(\log N)}$
- 6: **for** $m = 1$ to $\log N$ **do**
- 7:

$$\vec{\alpha}^{(m)} = \frac{(Id \otimes \vec{\alpha}^{(m-1)T} \otimes \vec{\alpha}^{(m-1)T})\hat{T}}{\|(Id \otimes \vec{\alpha}^{(m-1)T} \otimes \vec{\alpha}^{(m-1)T})\hat{T}\|} \quad (3)$$

- 8: $\lambda = (\vec{\alpha}^{(\log N)T} \otimes \vec{\alpha}^{(\log N)T} \otimes \vec{\alpha}^{(\log N)T})\hat{T}$
 - 9: **return** the estimated eigenvector/eigenvalue pair $(\vec{\alpha}, \lambda)$
-

of orthonormal unit vectors $\{\vec{\alpha}_{\kappa,1}, \vec{\alpha}_{\kappa,2}, \dots, \vec{\alpha}_{\kappa,M_\kappa}\}$ together with corresponding position scalars $\lambda_{\kappa,m} \geq 0$ which in this paper we call them gene-value. For our algorithm we consider the case we have an orthogonally decomposable symmetric tensor $\hat{T}^{(k)} \in \mathbb{R}^{N \times N \times N}$ in tensor $\mathcal{T}^{(\kappa)} = \hat{T}^{(\kappa)} + \mathcal{E}_\kappa$, where $\mathcal{E}_\kappa$ is the perturbation in $\mathbb{R}^{N \times N \times N}$ with small spectral norm. For a bipartition P_1, P_2 of the index set $\mathcal{T}^{(\kappa)}$, the spectral norm of the matrix unfolding $\|\mathcal{T}^{(\kappa)}\|_{P_1, P_2}$ with rows indexed by the indices in P_1 and columns indexed by indices in P_2 ,

$$\|\mathcal{T}^{(\kappa)}\|_{P_1, P_2} = \max_{\substack{x \in (\mathbb{R}^N)^{\otimes |P_1|}, y \in (\mathbb{R}^N)^{\otimes |P_2|} \\ \|x\| \leq 1, \|y\| \leq 1}} \sum_{i,j,k} \mathcal{T}_{i,j,k}^{(\kappa)} x_{i_{P_1}} y_{i_{P_2}}$$

Since the tensor is *symmetric*, all possible spectral norms $\|\mathcal{T}_{\{1,2\}\{3\}}^{(\kappa)}\|$, $\|\mathcal{T}_{\{1\}\{2,3\}}^{(\kappa)}\|$, and $\|\mathcal{T}_{\{2,3\}\{1\}}^{(\kappa)}\|$ are the same. Also, since the order of our tensor is odd without loss of generality we can add the requirement that $\lambda_{\kappa,m}$ is positive.

We remark that in our multilinear operation on tensors, we use the form $\mathcal{T}^{(\kappa)} \mapsto (U_i \otimes U_j \otimes U_k)\mathcal{T}^{(\kappa)}$, where U_i, U_j, U_k are matrices with N columns and,

$$(U_i \otimes U_j \otimes U_k)\mathcal{T}^{(\kappa)} = \sum_{i,j,k} \mathcal{T}_{i,j,k}^{(\kappa)} (U_i e_i) \otimes (U_j e_j) \otimes (U_k e_k)$$

If some of the matrices of U_m are identity matrix, then the corresponding operation is called tensor contraction. For example, for 3^{rd} -order tensor $\mathcal{T}^{(\kappa)}$ and matrix $U \in \mathbb{R}^{N \times N}$, we call $(Id \otimes Id \otimes U)\mathcal{T}^{(\kappa)}$ the contraction of the third mode of $\mathcal{T}^{(\kappa)}$ with matrix U and the result of this contraction is a vector. For vector $\vec{v} \in \mathbb{R}^N$, we call $(Id \otimes Id \otimes \vec{v})\mathcal{T}^{(\kappa)}$ the contraction of the third mode of $\mathcal{T}^{(\kappa)}$ with vector $\vec{v}$ and the result of this contraction would be matrix.

The key step of the algorithm to find the single component of the tensor $\mathcal{T}^{(k)}$ is repeating two contractions of the tensor under power iteration in (2). We first generate matrix U , a uniform $N \times N$ dimensional matrix and contract tensor $\mathcal{T}^{(\kappa)}$ with U to get a vector. Then we contract the tensor $\mathcal{T}^{(\kappa)}$ by contracted $\mathcal{T}^{(\kappa)}$ one more time and we get a matrix. Let us assume the result of the contraction of the tensor $\mathcal{T}^{(\kappa)}$ with U is vector $\vec{v} = (v_1, \dots, v_N)$,

$$\vec{v} := (Id \otimes Id \otimes U)\mathcal{T}^{(\kappa)} \longleftrightarrow v_n = \sum_{n=1}^N (Id \otimes Id \otimes \vec{u}_n^T)\mathcal{T}_n$$

where $\vec{u}_n$ is n^{th} column of the matrix U and $\mathcal{T}_n^{(\kappa)}$ is n^{th} slice of tensor $\mathcal{T}^{(\kappa)}$. Now we contract tensor $\mathcal{T}^{(\kappa)}$ by vector $\vec{v}$ one more time by the following contraction and find matrix U which we normalize and use it in our power iteration update in (2),

$$U := (Id \otimes Id \otimes \vec{v}^T) \mathcal{T}^{(\kappa)} \longleftrightarrow U = \sum_{n=1}^N v_n \mathcal{T}_n$$

Since we have two contractions at each step and updated result would be a matrix instead of a vector in (2), Algorithm 1 establishes the convergence for extracting a single component of the orthogonal decomposition faster than the convergence of tensor power method for orthogonal decomposition in [4]. After $\log N$ power iteration we get converged matrix U . Then we run $\log N$ power iteration on top eigenvector of matrix U in (3) and the final output would be $\vec{\alpha}$, the single component of the tensor. The algorithm succeeds with probability at least $1/\log n$ over the randomness of the algorithm when the ratio between the largest and second largest eigenvalue of contracted tensor is at least $1 + 1/\log n$. We can find full analysis of similar method is in [11]. To find the corresponding $\lambda_{\kappa,m}$ we contract tensor $\mathcal{T}^{(\kappa)}$ in all 3 ways by the component. By running Algorithm 1 M_κ times we can extract an approximated decomposition of $\mathcal{T}^{(\kappa)}$.

Computation of subsequent gene-component can be computed with deflation $\hat{\mathcal{T}}$, i.e., $\hat{\mathcal{T}} = \mathcal{T}^{(\kappa)} - \sum_{m=1}^{M_\kappa} \lambda_{\kappa,m} \vec{\alpha}_{\kappa,m} \otimes 3$.

Tensor order 4 decomposition

Let the fourth-order tensor $\{\check{\mathcal{T}}^{(\kappa)}\} \in \mathbb{R}^{N \otimes 4}$ of size $K \times N$ -genes $\times N$ -genes $\times N$ -genes tabulates a series of K genome scale order-3 tensors $\mathcal{T}^{(\kappa)}$,

$$\mathcal{T}_{i,j,k}^{(\kappa)} = \sum_{m=1}^{M_\kappa} \mathbf{a}_{i,m}^{(\kappa)} \mathbf{a}_{j,m}^{(\kappa)} \mathbf{a}_{k,m}^{(\kappa)},$$

such that each tensor is constructed by a new genomic signal (i.e. DNA-bound protein's occupancy levels) which is created by Moore-Penrose pseudo-inverse projection of genome scale signal $\mathbf{a}^{(1)}$ on designated basis signal $\mathbf{p}^{(\kappa)}$ of size N -genes $\times M_\kappa$ -arrays, of, e.g., proteins' DNA binding levels measured in a set of M_κ samples of the cellular system,

$$\mathbf{a}^{(\kappa)} = \mathbf{p}^{(\kappa)} \mathbf{p}^{(\kappa)\dagger} \mathbf{a}^{(1)} \quad (4)$$

By this computation we want to keep only those sub-tensors which are common to both signals $\mathbf{a}^{(1)}$ and new projected signal $\mathbf{a}^{(\kappa)}$. Now we define and compute a higher order tensor decomposition of the tensor of tensors $\{\check{\mathcal{T}}^{(\kappa)}\}$,

$$\check{\mathcal{T}} := \sum_{\kappa=1}^K \mathcal{T}^{(\kappa)} = \sum_{m=1}^M \lambda_m \vec{\alpha}_m \otimes 3$$

The result of above decomposition are gene-components $\vec{\alpha}_m$ and gene-values λ_m which are decorrelated in overall tensor $\check{\mathcal{T}}$ as our algorithm find the orthogonal components with good approximation. These decorrelated gene-components in overall tensor might not be decorrelated in any one of individual tensors, since $\lambda_{\kappa,m,l,n} \neq 0$. Here we propose a new model based on higher order tensor decomposition above which formulates each individual tensor order-3 $\mathcal{T}^{(\kappa)}$ as a series of $M := \sum_{\kappa=1}^K M_\kappa$ rank-1 symmetric decorrelated sub-tensors and the series of $\binom{M}{3}$ rank-3 tripling among these sub-tensors such that,

$$\begin{aligned} \mathcal{T}^{(\kappa)} = & \sum_{m=1}^M \lambda_{\kappa,m} \vec{\alpha}_m^{\otimes 3} + \sum_{m=1}^M \sum_{l=m+1}^M \sum_{n=l+1}^M \lambda_{\kappa,m,l,n} (\vec{\alpha}_m \otimes \vec{\alpha}_l \otimes \vec{\alpha}_n + \vec{\alpha}_m \otimes \vec{\alpha}_n \otimes \vec{\alpha}_l \\ & + \vec{\alpha}_l \otimes \vec{\alpha}_m \otimes \vec{\alpha}_n + \vec{\alpha}_l \otimes \vec{\alpha}_n \otimes \vec{\alpha}_m + \vec{\alpha}_n \otimes \vec{\alpha}_m \otimes \vec{\alpha}_l + \vec{\alpha}_n \otimes \vec{\alpha}_l \otimes \vec{\alpha}_m) \end{aligned} \quad (5)$$

for all $\kappa = 1, 2, \dots, K$. Tripling of components $\vec{\alpha}_l, \vec{\alpha}_m, \vec{\alpha}_n$ is sum of tensor product of all combinations of these sub-tensors for all $m \neq l \neq n$. By this definition we can find the significance of HOGC m^{th} sub-tensor in the κ^{th} tensor. The significance is indicated by the m^{th} fraction of gene-component of the κ^{th} tensor $s_{\kappa,m} = \lambda_{\kappa,m} / (\sum_{k=1}^K \sum_{m=1}^M \lambda_{\kappa,m})$, i.e., the correlation of components captured by the m^{th} sub-tensor in the κ^{th} tensor relative to that captured by all sub-tensors. Similarly, the amplitude of the fraction $s_{\kappa,m,l,n} = \lambda_{\kappa,m,l,n} / (\sum_{k=1}^K \sum_{m=1}^M \lambda_{\kappa,m})$ indicates the significance of the tripling among the m^{th}, l^{th} and n^{th} sub-tensors in the κ^{th} tensor.

3.1 Biological interpretation of Yeast intracellular signaling pathways in higher order assemblies

To determine HOGC inter- and intra- pathways in Yeast cellular system, we compute $K = 4$ HOGC tensors by signal matrices $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \mathbf{a}^{(3)}$, and $\mathbf{a}^{(4)}$. The genomic signal matrices $\mathbf{a}^{(2)}, \mathbf{a}^{(3)}$, and $\mathbf{a}^{(4)}$ are driven by projection $\mathbf{p}^{(1)}, \mathbf{p}^{(2)}$ and $\mathbf{p}^{(3)}$ onto basis signals $\mathbf{a}^{(1)}$ (4), which tabulates the relative binding DNA-bound protein occupancy levels of the yeast transcription factors that are involved in cell-cycle, development, and biosynthesis programs respectively under experimental conditions of Spellman et al. and Roberts et al. [16, 20, 21]. We compute all tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \mathcal{T}^{(3)}$, and $\mathcal{T}^{(4)}$ on common genes among $\mathbf{a}^{(1)}, \mathbf{p}^{(1)}, \mathbf{p}^{(2)}$ and $\mathbf{p}^{(3)}$ with manifested stages for either pheromone or cell cycle classification which tabulates relative mRNA expression levels of 27 yeast genes with valid data in the number of presented samples of a cell cycle time course of a culture synchronized by the mating pheromone α factor. Before computing $\mathcal{T}^{(1)}$ from the data signal $\mathbf{a}^{(1)}$, we use IterativeSVD to fill missing data in $\mathbf{a}^{(1)}$ and to approximately center the expression pattern of each gene in its invariant level (mean). IterativeSVD is matrix completion by iteratively low-rank svd decomposition which is similar to SVDimpute from [Missing value estimation methods for DNA microarrays] [23, 27]. Tensors $\mathcal{T}^{(2)}, \mathcal{T}^{(3)}$, and $\mathcal{T}^{(4)}$ are computed by $\mathbf{a}^{(2)}, \mathbf{a}^{(3)}$, and $\mathbf{a}^{(4)}$ respectively.

For both pheromone and cycle classifications we associate sub-tensors and the tripling among them with most likely expression correlations based on the P -values calculated by hypergeometric probability distribution of the Q triples of annotations among T which is total number of triples,

$$P(t; q, Q, T) = \frac{\binom{Q}{q} \binom{T-Q}{t-q}}{\binom{T}{t}}$$

By combinatorics and assuming a hypergeometric probability distribution [19], we can find the probability of a specific annotation in Q triples of genes' annotations from a population of N genes that contains exactly T triples with that annotation, where in each annotation is one of the possible combination of triple among cell cycle and pheromone response states. The association of all sub-tensors of projected tensors to annotations from two different classifications of cell-cycle and pheromone-response are brought in Table 1.

The Cell cycle classification includes five cellular states (G_1, S, G_2 , and M) [8]. For pheromone response classification we have either state Up or Down [10]. The stages might be None which means we don't know about the state of that gene regarding the classification. We select $t = 120$ higher correlations among $T = 2925$ triples of $N = 27$ genes at the intersection of basis signals to find the association of most likely cellular states with each sub-tensor and tripling among them for both classifications.

Figure 1: Discretized significant sub-tensors of the tensor $\mathcal{T}^{(1)}$ in the subset of 120 higher correlations largest in amplitude among all genes with their cell-cycle classifications, M/G_1 , G_1 , S , S/G_2 , and G_2/M and pheromone-response classifications, up-regulated and down-regulated. Most of the participated genes in higher order correlation among 27 genes in the intersection of DNA binding data are not reported to be regulated by pheromone but new structures of cell-cycle dependent among proteins can be observed. U and D are the gene code according to pheromone response classifications up-regulated and down-regulated respectively

Significant independent sub-tensors associated with intra-pathways in higher order signaling machines In this part we analyze genes are pathway dependent in higher order signaling machines. Genes are involved in independent intra-pathway higher order signaling are assembled to the structure regarding of their cell-cycle phases. The algorithm also drives the impact of each independent higher order signaling pathway in overall multiple intracellular signaling. The most three significant sub-tensors which our algorithm uncovers on $\mathcal{T}^{(1)}$, capture $\approx 67\%$, 5% , and 7% respectively of the higher expression correlation of $\mathcal{T}^{(1)}$. There are independent pathways associated to the HOGC sub-tensors which are manifest in the data signal $\mathbf{a}^{(1)}$. The result follows by the P values for the distribution of the 10 genes. The genes microarray-

classified as pheromone regulated and 25 genes which traditionally or microarray classified as cell cycle regulated genes with the subset of $t=120$ triples of genes with highest levels of expression correlation among all $T = 2925$ triples. We visualize these three discretized sub-tensors of the $\mathcal{T}^{(1)}$ that constitute 120 higher correlations in each sub-tensors with large amplitude among all high correlations Figure1. The result uncovers genes *MKC7*, *YNK1*, *YEH1* and *GAS2* are not pathway dependent in higher order signaling as they have not appeared in any HOGC sub-tensors.

Figure1a shows that the higher correlations among the genes in the first and most significant sub-tensor are pheromone-response independent, however we can see that all higher correlations, have at least one gene with cell-cycle stage S or G_1 . For example, *CIS3*, *SWI4*, and *YHR1* are not correlated since none of them peaks at either cell cycle S or G_1 , but we have HOGC among *CIS3*, *SWI4*, and *HTA1* where *HTA1* encodes cell cycle S .

In the second sub-tensor, The higher relations among the genes depend only on cell cycle classification phases M/G_1 and G_1 Figure1b. This sub-tensor discovers there is no higher correlation among genes with other classification phases, and therefore it filters out all decorrelated genes *IST2*, *VAC17*, *YDR0*, *HTB1*, *HTA1*, *CIS3*, *ECM17*, *CWP2*, *YLR3*, *DIC1*, *SML1*, and *PLB3* of $\mathcal{T}^{(1)}$ ¹

The third HOGC sub-tensor omits those genes from cell cycle classification of S and G_2/M since they are decorrelated from each other in higher order relations in this sub-tensor Figure1c. More precisely the higher relation of genes in this sub-tensor depends only on genes with cell cycle progression G_1 , M/G_1 , and S/GG_2 . For example, *CIS3*, *SWI4*, and *HTA1* are not correlated since *HTA1* encodes cell cycle S . Besides, genes presented in HOGC during mating are cycle dependent. Genes that are down regulated in response to pheromone are involved in the third higher order signaling intra-pathway if their cell-cycle stages are either G_1 or S/G_2 . Also, We can see gene *DSE1* with down regulated response to pheromone during mating is involved in all higher order correlations in signaling pathways. For all tensors, most likely parallel associations of the first three significant independent HOGC of signaling pathways and number of participated genes in each one are presented in Table 1.

¹*YDR0*, *YFL0*, *YHR1*, *YLR3*, *YLR4*, and *YLR46* are short form of YORFs *YDR089W*, *YFL064C*, *YHR126C*, *YLR345W*, *YLR4262W*, and *YLR4264W* since the genes name in our data were unknown.

Table 1: Most likely parallel associations of the most three significant independent higher order of signaling pathways and number of participated genes in each pathway of tensor $\mathcal{T}^{(1)}$ created by data signal $\mathbf{a}^{(1)}$, $\mathcal{T}^{(2)}$ created by $\mathbf{a}^{(2)}$ the projection of $\mathbf{a}^{(1)}$ onto the cell cycle, $\mathcal{T}^{(3)}$ created by $\mathbf{a}^{(3)}$ the projection of $\mathbf{a}^{(1)}$ onto the development, and $\mathcal{T}^{(4)}$ created by $\mathbf{a}^{(4)}$ the projection of $\mathbf{a}^{(1)}$ onto the biosynthesis basis signals, according to the traditional and microarray classifications of cell cycle- and pheromone- regulated yeast genes.

Tensor	sub-tensor	N(genes)	Cell Cycle classification		Pheromone response classification	
			association	P-value	association	P-value
Cell expression	1	25	$G_1 G_1 S$	5.5×10^{-19}	Up Down None	2.6×10^{-14}
	2	13	$G_1 G_1 G_1$	2.7×10^{-22}	None None None	2.5×10^{-4}
	3	14	$G_1 G_1 S/G_2$	5.1×10^{-9}	None None Up	1.9×10^{-7}
Cell Cycle	1	27	$G_1 G_1 S$	4.3×10^{-9}	Up Up None	6.5×10^{-5}
	2	14	$G_1 G_1 G_1$	7.5×10^{-23}	None None None	5×10^{-9}
	3	17	$G_1 S S/G_2$	2.9×10^{-8}	Down Down None	4.7×10^{-7}
Development	1	17	$G_1 G_1 S$	9.3×10^{-5}	Down Down None	3.4×10^{-5}
	2	16	$G_1 G_1 M/G_1$	3.8×10^{-14}	None None None	7.3×10^{-21}
	3	17	$G_1 G_1 S$	9.3×10^{-5}	Down Down None	3.4×10^{-5}
Biosynthesis	1	23	$G_1 G_1 S$	1.2×10^{-23}	Down Down None	4×10^{-8}
	2	19	$G_1 G_1 G_1$	2.8×10^{-36}	None None None	8.6×10^{-6}
	3	17	$G_1 S S/G_2$	9.7×10^{-6}	Down Down Down	3.8×10^{-23}

Figure 2: Discretized significant sub-tensors of the Tensor $\{\tilde{\mathcal{T}}\} \equiv \mathcal{T}^{(1)} + \mathcal{T}^{(2)} + \mathcal{T}^{(3)} + \mathcal{T}^{(4)}$ in the subset of 120 higher correlations largest in amplitude.

Table 2: Most likely parallel associations of the sub-tensors of tensor $\{\tilde{\mathcal{T}}\}$ constructed by four tensors $\mathcal{T}^{(1)}$, $\mathcal{T}^{(2)}$, $\mathcal{T}^{(3)}$, and $\mathcal{T}^{(4)}$ and their tripling according to the traditional and microarray classifications of cell cycle- and pheromone- regulated yeast genes.

sub-tensors or Triples between sub-tensors	Cell Cycle classification		Pheromone response classification	
	association	P-value	association	P-value
1	$G_1 G_1 S$	8.5×10^{-21}	Down Down Up	7.4×10^{-7}
2	$G_1 G_1 G_1$	1.2×10^{-17}	None None None	1.2×10^{-4}
3	$S/G_2 S/G_2 S$	1.6×10^{-18}	Down Down None	4.8×10^{-8}
4	$S/G_2 S/G_2 S$	1.6×10^{-18}	Down Down None	4.8×10^{-8}
1 $\leftrightarrow$ 2 $\leftrightarrow$ 3	$G_1 G_1 S$	5.6×10^{-32}	None None None	2.6×10^{-15}
1 $\leftrightarrow$ 2 $\leftrightarrow$ 4	$G_1 G_1 M/G_1$	9.6×10^{-7}	None None Down	5.1×10^{-3}
1 $\leftrightarrow$ 3 $\leftrightarrow$ 4	$S/G_2 G_1 S$	1.4×10^{-13}	None None Down	5.1×10^{-3}
2 $\leftrightarrow$ 3 $\leftrightarrow$ 4	$G_1 G_1 S$	4.4×10^{-11}	None None Down	1.6×10^{-10}

sub-tensors and their tripling are associated with intra- and inter-pathways' higher order signaling

Our model which is presented in (5) for higher order correlations among tensors $\mathcal{T}^{(1)}$, $\mathcal{T}^{(2)}$, $\mathcal{T}^{(3)}$, and $\mathcal{T}^{(4)}$ uncovers 4 significant sub-tensors and tripling among these sub-tensors which capture higher order correlations in each individual tensor Figure3. The significant decorrelated sub-tensors that are manifested in overall tensor capture $\approx 60\%$, 4% , and 6% , and $\ll 1\%$ of the higher correlations among genes respectively. The associations of aforementioned sub-tensors and triplings among them is computed for 27 common genes (genes with manifested stage either in pheromone classification or cell-cycle) in $\mathbf{a}^{(1)}$, $\mathbf{a}^{(2)}$, $\mathbf{a}^{(3)}$, and $\mathbf{a}^{(4)}$ Table2.

The 3D visualization of four significant sub-tensors of higher order decomposition of individual tensors Figure2 reveals new hidden subsets of higher order correlation of intra-pathway genes. The sub-tensors are associated with the higher order signaling in independent pathways that are manifest in the overall tensor of higher correlations as well as the individual tensors Figure2. The higher order intra-pathway relation of genes in the first sub-tensor of overall decomposition brought in Figure2a.

Figure2b uncovers that only genes with the cell cycle classifications of G_1 , M/G_1 , and S belong to this independent pathway in higher correlation. Genes with down regulated in response to pheromone that are involved in this higher order signaling intra-pathway depend on either cell-cycle phase G_1 or S .

In the third sub-tensor, The higher relations among the genes depend only on cell cycle classification of G_1 , S/G_2 and S Figure2c. Genes that are down regulated in response to pheromone are involved in the third higher order signaling intra-pathway if their cell-cycle stages are either S or S/G_2 .

The fourth sub-tensor keep those genes from cell cycle classification of G_1 , S/G_2 and M/G_1 Figure2d. This sub-tensor is also is pheromone dependent in the way that it highlights no higher correlation among involved genes are Up-regulated in pheromone-response classification. Genes that are down regulated in response to pheromone are involved in the third higher order signaling intra-pathway if their cell-cycle stages are either G_1 or S/G_2 .

Figure3c, Figure3d shows the the contribution of the first sub-tensor to the higher expression correlation of development tensor $\mathcal{T}^{(3)}$ and biosynthesis tensor $\mathcal{T}^{(4)}$ which are constructed based on projected signals $\mathbf{a}^{(3)}$ and $\mathbf{a}^{(4)}$ is negligible, but it contributes to the higher expression correlation of tensor $\mathcal{T}^{(1)}$ and cell cycle tensor $\mathcal{T}^{(2)}$ Figure3a, 3b. The second sub-tensor, which is associated with signal transduction pathway that only involves genes with cell cycle stage G_1 , M/G_1 , and S in higher order correlations contributes

Figure 3: Significance of the sub-tensors of higher order tensor $\{\mathcal{T}\}$ in each individual tensor $\mathcal{T}^{(1)}$, $\mathcal{T}^{(2)}$, $\mathcal{T}^{(3)}$, and $\mathcal{T}^{(4)}$ and the contribution of tripling among them.

Figure 4: Discretized significant Tripling among the significant sub-tensors of Tensor $\{\tilde{T}\}$ in the subset of 120 higher correlations largest in amplitude.

to the expression correlation of the higher correlations of tensor $\mathcal{T}^{(1)}$ and biosynthesis tensor $\mathcal{T}^{(4)}$ mostly Figure3a, Figure3d, and third sub-tensor to the higher expression correlation of tensor $\mathcal{T}^{(1)}$ and development tensor $\mathcal{T}^{(3)}$ Figure3a, Figure3c. The most contribution of fourth sub-tensor which is associated with signal transduction pathway in higher order formations of genes with cell cycle G_1 , S/G_2 , and M/G_1 is in cell cycle tensor $\mathcal{T}^{(2)}$ and development tensor $\mathcal{T}^{(3)}$ Figure3b, Figure3c, while it's contribution to tensor $\mathcal{T}^{(1)}$ and biosynthesis tensor $\mathcal{T}^{(4)}$ is negligible Figure3a, Figure3d.

Trough structural studies of inter- pathways signaling sub-tensors, we began to see different scenarios regarding formation of higher order signaling machines. For example, we identified this scenario that histones and putative proteins are involved in all higher order signaling of inter-pathways. The participated Histones in these HOGC inter-pathways' higher order signaling are *HTA1* and *HTB1* and putative proteins are *YLF0* and *YLR4* (*YLF0* and *YLR4* are short form of YORFs *YFL064C* and *YLR462W* since their gene's name are unknown). The discretized sub-tensors and triplings highlight unknown higher or-

Figure 5: Higher order assemblies of cell cycle classification of genes in triplings among the significant HOGC sub-tensors of signal transduction Tensor $\{\tilde{T}\} \in \mathbb{R}^{\otimes 4}$ in cellular system in the subset of 120 higher correlations largest in amplitude. Each layer shows the stage of one of the involved genes in triple correlation for each tripling. For example, (a) shows that in first tripling among $\mathcal{T}^{(1)}$, $\mathcal{T}^{(2)}$, and $\mathcal{T}^{(3)}$ the first gene is with cell-cycle stage G_1 , the second gene has cell-cycle stage S or S/G_2 , and the third one can be any of them.

der relations among genes that are in agreement with current understanding of the cellular system of yeast Figure4, Figure5. The tripling is associated with the higher order signaling among these independent pathways that are manifest in individual tensors.

The tripling among first, second and third sub-tensors is associated with G_1 and one of the S/G_2 or S and it has contribution in all individual tensors except $\mathcal{T}^{(4)}$ Figure3. To obtain comprehensive view of the relationship between higher order gene organization and induction of developmentally regulated gene-expression profile, our comparative analysis of intra-pathway of first, second and third reveals distinct set of cell cycle classification of (G_1, S) or $(G_1, S/G_2)$ in every triple of genes in this higher order level Figure5a. We might conclude that in higher correlation among tensor of expression correlations of $\mathcal{T}^{(1)}$, and development projected tensor of $\mathcal{T}^{(3)}$ gene with type G_1 from cycle classification is an activation platform for correlation of other genes in this tripling. The most participated genes in this tripling is *YFL0*, *YLR*, and *YLR6* Figure4b.

The tripling among first, second and fourth sub-tensors reveals G_1 as a permanent cell cycle association and it contributes in all individual tensors Figure3 Figure5b. The most participated genes in this tripling is *YFL0*, *YLR*, and *YLR6* Figure4a.

In all triples among first, third and fourth sub-tensors we have S and this tripling contributes in $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(4)}$ while its contribution in individual tensors $\mathcal{T}^{(2)}$ and $\mathcal{T}^{(3)}$ is negligible Figure3g Figure3h. Figure4c shows the most significant genes in this tripling are *HTA1*, and *HTB1*. Therefore, this tripling highlights higher order correlations among the histones and other genes that are involved in triple of first, third and fourth sub-tensors.

The tripling among second, third and fourth sub-tensors which has contribution in $\mathcal{T}^{(2)}$, $\mathcal{T}^{(3)}$, and $\mathcal{T}^{(4)}$ reveals distinct subsets of $G_1, S, G_1, S/G_2, M/G_1, S/G_2$ and $M/G_1, M/G_1$ Figure5d. Figure4d also shows the most significant genes in this tripling are *YFL0*, *YLR*, *HTA1*, and *HTB1*.

4 Discussion

Many questions are raised by higher order correlation analysis of genes, which need to be explored both theoretically and experimentally. Elucidating the biophysical principle governing the higher order signaling analysis among genes, regulated fashion may reveal the structural basis of clustering and uncover paradigm in cell singling. Higher-order assemblies may be an important aspect of many biological processes because they enable formation of precisely organized molecular machines from constituents present in inactive states at low concentrations to promote biochemical reactions in cells. We have defined tensors based on genome scale signal of, e.g., mRNA expression and proteins' DNA-binding relations among genes and described an algorithm for efficient new tensor decomposition based on power iteration to separate genome scale nondirectional tensors into mathematically defined decorrelated components. We have demonstrated its usefulness on a three dimensional and four dimensional higher correlation data to uncover HOGC sub-tensors and HOGC tripling relations among them with clear biological and statistical significance. This approach tends to focus mainly on identifying higher order signaling in intra- pathways in cellular system and higher order signaling among these transduction signaling pathways (inter-pathway signaling) and also multiple intracellular structures of cell cycle genes. The proposed tensor decomposition was a simply approach to extract the orthogonal decomposition with high accuracy in analyses of higher order gene correlations in genome scale yeast data. These components and tripling among them include reconstruction of higher order signaling of inter- and intra-pathways from nondirectional tensor of higher correlations. The result uncovers new higher order coordinated differential relations mostly among cell-cycle and also some pheromone regulated genes in higher order pathway-dependent signalings of these genes. There are several interesting ways in which this model can be extended or changed. The model can be extended to higher-dimensional data or this factorization can be benefited multiple-tissue gene expression data sets.

References

- [1] O. Alter, P. O. Brown and D. Botstein, *Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms*, Proceedings of National Academy of Sciences, 100(6), 3351-3356, 2003.
- [2] O. Alter and G. H. Golub, *Integrative analysis of genome-scale data by using pseudoinverse projection predicts novel correlation between DNA replication and RNA transcription*, Proceedings of National Academy of Sciences, 101(47) 16577-16582, 2004.
- [3] O. Alter and G. H. Golub, *Reconstructing the pathways of a cellular system from genome-scale signals by using matrix and tensor computations*, Proceedings of National Academy of Sciences, 102(49), 17559-17564, 2005.
- [4] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade and M. Telgarsky, *Tensor Decompositions for Learning Latent Variable Models*, Journal of Machine Learning Research, 15(1), 2773-2832, 2014.
- [5] A. Bhaskara, M. Charikar, A. Moitra and A. Vijayaraghavan, *Smoothed analysis of tensor decompositions*, Proceedings of the forty-sixth annual ACM symposium on Theory of Computing, 594-603, 2014.
- [6] M. P. S. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares and D. Hausler, *Knowledge-based analysis of microarray gene expression data by using support vector machines*, Proceedings of National Academy of Sciences, 97(1) 262-267, 2000.
- [7] M. Civelek and A. J. Lusis, *Systems genetics approaches to understand complex traits*, Nature Reviews Genetics, 15, 34-48, 2014.

- [8] GM Cooper, *The Eukaryotic Cell Cycle*, The cell: a molecular approach, Washington, D.C: ASM Press, 2000.
- [9] S. Dutta, J.-P. Eckmann, A. Libchaber and T. Tlusty, *Green function of correlated genes and the mechanical evolution of protein*, Proceedings of the National Academy of Sciences, 2018.
- [10] S. Fields, *Pheromone response in yeast*, Trends Biochem Sci, 270-3, 1990.
- [11] S. B. Hopkins, T. Schramm, J. Shi and D. Steurer, *Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors*, Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, 178-191, 2016.
- [12] V. Hore, A. Viñuela, A. Buil, J. Knight, M. McCarthy, K. Small and J. Marchini, *Tensor decomposition for multiple-tissue gene expression experiments*, Nature Genetics, 48(9), 1094-1110, 2016.
- [13] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Review, 51(3), 455-500, 2009.
- [14] P. M. Kroonenberg, *Applied Multiway Data Analysis*, John Wiley & Sons, 2008.
- [15] X. Li, Y. Ye, Q. Wu and M. K. Ng, *MultiFacTV: Finding modules from higher-order gene expression profiles with time dimension*, IEEE International Conference Bioinformatics and Biomedicine, 1-6, 2012.
- [16] R. Martone, G. Euskirchen, P. Bertone, S. Hartman, T. E. Royce, N. M. Luscombe, J. L. Rinn, F. K. Nelson, P. Miller, M. Gerstein, S. Weissman and M. Snyder, *Distribution of NF-B-binding sites across human chromosome 22*, Proceedings of National Academy of Sciences, 100(21), 12247-12252, 2003.
- [17] A. T. McKenzie, I. Katsyv, W.-M. Song, M. Wang and B. Zhang, *DGCA: A comprehensive R package for Differential Gene Correlation Analysis*, BMC Systems Biology, 10, 106, 2016.
- [18] M. Mørup, *Applications of tensor (multiway array) factorizations and decompositions in data mining*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 24-40, John Wiley & Sons, Inc., 2011.
- [19] I. Rivals, L. Personnaz, L. Taing, and M. Potier, "Enrichment or depletion of a GO category within a class of genes: which test?", Bioinformatics. 23 (4): 401-407, 2007.
- [20] CJ. Roberts , B. Nelson, MJ. Marton ,R. Stoughton, MR. Meyer, HA. Bennett , YD. He ,H. Dai ,WL. Walker , TR. Hughes, M. Tyers, C. Boone , SH. Friend, *Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles*, Science. 2000 Feb 4;287(5454):873-80.
- [21] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, B. Futcher, *Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization*, Mol. Biol. Cell 9, 3273-3297, 1998.
- [22] M. B. Taylor and I. M. Ehrenreich, *Higher-order genetic interactions and their contribution to complex traits*, Trends in Genetics, 31(1), 34-40, 2015.
- [23] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, RB. Altman, *Missing value estimation methods for DNA microarrays.*, Bioinformatics 520-5, 2001.
- [24] Hao. Wu, *Higher-order Assemblies in a New Paradigm of signal transduction*, Cell 153, 2013.

- [25] M. Wu, J. Huang and S. Ma, *Identifying gene-gene interactions using penalized tensor regression*, Statistics in Medicine, 37(4), 598-610, 2017.
- [26] X. Zhang, F. Pan and W. Wang, *Finding High-Order Correlations in High-Dimensional Biological Data*, Link Mining: Models, Algorithms, and Applications, P. S. Yu *et al.* (eds.), 505-534, Springer, New York, NY, 2010.
- [27] <https://pypi.org/project/fancyimpute>.

A Appendix

Figure 6: Common higher correlated genes among all sub-tensors of tensor $\tilde{\mathcal{T}}$. These higher order correlation of genes are not pathway dependent in higher order signaling machine. Color-coded for genes inside the box is according to their cell-cycle classifications, M/G_1 (green), G_1 (blue), S (red), S/G_2 (orange), and G_2/M (pink), and separately according to their pheromone-response classifications, we highlight the up-regulated purple and down-regulated gray. The graph nodes are depicted in columns and rows. Each column in the graph shows HOGC sub-tensors of an individual tensor. For example $\mathcal{T}_1^{(2)}$, $\mathcal{T}_2^{(2)}$, and $\mathcal{T}_3^{(2)}$ are most significant HOGC sub-tensors of tensor $\mathcal{T}^{(2)}$ which is Cell Cycle tensor. Each row represents most significant HOGC sub-tensors of all individual tensors at same significant level. For example $\mathcal{T}_1^{(1)}$, $\mathcal{T}_1^{(2)}$, $\mathcal{T}_1^{(3)}$ and $\mathcal{T}_1^{(4)}$ are first and most significant sub-tensors in individual tensor $\mathcal{T}_1^{(1)}$, $\mathcal{T}_1^{(2)}$, $\mathcal{T}_1^{(3)}$ and $\mathcal{T}_1^{(4)}$ respectively in the overall tensor of cell expression, cell cycle, development and biosynthesis tensor respectively. Each path shows the common genes with higher order correlations among these sub-tensors. Each HOGC path is illustrated by different color. The left side box shows the path color for each HOGC. For example purple path passes through $\mathcal{T}_1^{(1)}$, $\mathcal{T}_3^{(2)}$, $\mathcal{T}_1^{(3)}$ and $\mathcal{T}_3^{(4)}$ shows correlation among genes *HTB1*, *HTA1*, and *DSE1* which are common in first sub-tensors of $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(3)}$ and third significant sub-tensors of $\mathcal{T}^{(2)}$ and $\mathcal{T}^{(4)}$.

Figure 7: Expression patterns across gene-components in all individual tensors of $\mathcal{T}^{(1)}$, $\mathcal{T}^{(2)}$, $\mathcal{T}^{(3)}$ and $\mathcal{T}^{(4)}$. In each subfigure blue line, green line, and orange one represent expression level of genes in first, second, and third most significant HOGC sub-tensor respectively in that individual tensor.

Figure 8: Line-joined display of spectral error of decomposition after each iteration to find a gene-components in tensors $\mathcal{T}^{(1)}$ (blue), $\mathcal{T}^{(2)}$ (orange), $\mathcal{T}^{(3)}$ (green), and $\mathcal{T}^{(4)}$ (red) and $\{\tilde{\mathcal{T}}\} \equiv \mathcal{T}^{(1)} + \mathcal{T}^{(2)} + \mathcal{T}^{(3)} + \mathcal{T}^{(4)}$ (purple). The error is the spectral norm of the deflated tensor at each step.

$\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$		
	MPS3	SML1 YLR4 HTA1
SWI4	HTB1	YLR4 DSE1
	MKC7	DSE1 DIC1
MKC7	DSE1	DIC1
(a)		
$\mathcal{T}^{(2)}$ and $\mathcal{T}^{(4)}$		
	YLR4	TSL1 DSE1 DIC1 VAC17 YFL0
HTA1	DIC1	YFL0 TSL1 YFL0
	YFL0	VAC17 SWI4
	DSE1	TSL1
YFL0	DIC1	VAC17 YLR4 DSE1
	YLR4	VAC17
YLR4	DSE1	DIC1
(c)		
$\mathcal{T}^{(1)}$ and $\mathcal{T}^{(3)}$		
	HTB1	MKC7 GAS2 DIC1 YHR1 CWP2 HTA1 ECM17
(b)		
$1 \leftrightarrow 3 \leftrightarrow 4$ and $2 \leftrightarrow 3 \leftrightarrow 4$		
	DIC1	YLR46 IST2 HTA1 VAC17 YFL0
HTB1	YLR4	YLR46 IST2 HTA1 VAC17 YFL0
	DSE1	YLR46 HTA1 VAC17 YFL0
	HTA1	YLR46
(d)		
$1 \leftrightarrow 2 \leftrightarrow 3$ and $1 \leftrightarrow 2 \leftrightarrow 4$		
DSE1	SWI4	YLR46 PRY3
	MPS3	YLR46
(e)		

Table 3: Boolean AND intersections of discretized significant sub-tensors in $\{\tilde{\mathcal{T}}^{(\kappa)}\} \in \mathbb{R}^{N^{\otimes 4}}$ of the series of 4 individual tensors $\{\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \mathcal{T}^{(3)}, \mathcal{T}^{(4)}\}$ and their tripling in the subset of 120 higher correlations in largest amplitude among all traditionally-classified cell cycle genes of highlighted series of inter-pathway relations of the cellular system of yeast. Color-coded according to their cell-cycle classifications, M/G_1 (green), G_1 (blue), S (red), S/G_2 (orange), and G_2/M (pink), and separately according to their pheromone-response classifications, we highlight the up-regulated purple and down-regulated gray. Yet, the annotation of *DIC1* in cell-cycle is not reported. Our model finds the total number of 15 extracted higher-correlations of *DIC1* with other genes among 120 higher correlations in sub-tensors. In 14 of these HOGC, one of the genes is at least from G_1 stage, and the other one shows HOGC of *HTB1*, *HTA1*, and *DIC1* where *HTB1* and *HTA1* encode S . Therefore, our analyses predicts that *DIC1* encodes G_1 under experimental conditions of Spellman et al. and Roberts et. al. [20, 21].

Figure 9: Line-joined display of gene-values for tensors $\mathcal{T}^{(1)}$ (blue), $\mathcal{T}^{(2)}$ (orange), $\mathcal{T}^{(3)}$ (green), and $\mathcal{T}^{(4)}$ (red). The m^{th} gene-value of corresponding gene-component $\vec{\alpha}_m$ in tensor $\mathcal{T}^{(\kappa)}$ is $\sqrt[3]{\lambda_{\kappa,m}}$. Note that for better graphical representation of it we scaled down the gene-values to third root of it which doesn't effect on the interpretation. Each tensor is decomposed to the number of designated signal's samples rank one sub-tensors. We find these sub-tensors and corresponding gene-value iteratively. These gene-values for each HOGC sub-tensor is computed in line 8 in Algorithm 1.

(h) $\mathcal{T}_1^{(1)}$ and $\mathcal{T}_1^{(3)}$

(i) $\mathcal{T}_1^{(1)}$ and $\mathcal{T}_2^{(3)}$

(j) $\mathcal{T}_1^{(1)}$ and $\mathcal{T}_3^{(3)}$

(k) $\mathcal{T}_2^{(1)}$ and $\mathcal{T}_2^{(3)}$

(l) $\mathcal{T}_3^{(1)}$ and $\mathcal{T}_1^{(3)}$

(m) $\mathcal{T}_3^{(1)}$ and $\mathcal{T}_2^{(3)}$

(n) $\mathcal{T}_3^{(1)}$ and $\mathcal{T}_3^{(3)}$

(o) $\mathcal{T}_1^{(1)}$ and $\mathcal{T}_1^{(4)}$

(p) $\mathcal{T}_1^{(1)}$ and $\mathcal{T}_3^{(4)}$

(q) $\mathcal{T}_2^{(1)}$ and $\mathcal{T}_1^{(4)}$

(r) $\mathcal{T}_2^{(1)}$ and $\mathcal{T}_2^{(4)}$

(s) $\mathcal{T}_3^{(1)}$ and $\mathcal{T}_1^{(4)}$

Figure 6: Intersection of significant HOGC sub-tensors of individual tensors in the subset of 120 relations among largest amplitude among all traditionally classified cell cycle genes of higher correlations inter-pathway dependent relations of the cellular system of yeast. Color-coded according to their cell-cycle classifications, M/G_1 (green), G_1 (blue), S (red), S/G_2 (orange), and G_2/M (pink), and separately according to their pheromone-response classifications, we highlight the up-regulated purple and down-regulated gray. (g), (n), and (u) represents intersection of the third significant HOGC sub-tensors where two genes of three genes are with S/G_2 and one gene is with G_1 cell-cycle classification in all higher order signaling. The intersection uncovers higher order correlations among cell wall genes. For example $CIS3$ or $CWP2$. Higher order correlations in third significant individual tensors would follow the same pathway-dependence. For example in all HOGC such as both $SML1$, $PRY3$, and $CIS3$ and $SML1$, $MCD1$, and $CIS3$, we have at least one gene with down-regulated response to pheromone with cell-cycle stage S/G_2 . Therefore, this analysis suggests that $CWP2$ in HOGC $SML1$, $YFL0$, and $CWP2$ which is not reported to be regulated by pheromone is down-regulated in response to pheromone. (d), (k), and (r) intersection of all second significant HOGC sub-tensors in individual tensors of $\mathcal{T}$ highlights higher correlation among genes with G_1 and M/G_1 .