Correlation Patterns in Gene Expressions along the Cell Cycle of Yeast

Jelena Živković , Marija Mitrović and Bosiljka Tadić

Abstract Currently available genome-wide expression measurements and the respective data bases represent the reliable entry information for gene interaction research. In order to unravel the collective behavior of genes, methods and approaches pertinent to complex dynamical systems are necessary. Using the network theory, we study correlation patterns in the time series of gene expressions of Yeast measured along the cell cycle. We select a subset of genes by their leading participation in the scale-invariant features of the expression data. Applying standard filtering of the correlation matrix reveals inhomogeneous mesoscopic structure of the related graph with several well defined modules of genes. The findings are corroborated by the spectral analysis of the correlation matrix and the eigenvector localization on the graph. The topologically distinct groups of genes which are co-expressed within a given phase of the cell cycle belong to different functional categories but often share the same localization, i.e., nucleus, cytoplasm, or mitochondria, inside the cell.

1 Introduction

Mapping of a complex dynamical system on a mathematical graph, i.e., by identifying its nodes and their connections (edges), provides a ground for quantitative study of complexity by methods of the formal graph theory and statistical physics of structured networks [1, 2]. In this picture, the dynamic processes of the system

Jelena Živković
SPM Group, IMM, Faculty of Science, Radboud University, Nijmegen, The Netherlands, e-mail: jelena.zivkovic@gmail.com

Marija Mitrović
Scientific Computing Laboratory, Institute of Physics, Belgrade, Serbia e-mail: mitrovic@scl.rs

Bosiljka Tadić
Department of theoretical physics, Jožef Stefan Institute, Ljubljana, Slovenia e-mail: bosiljka.tadic@ijs.si

can be viewed as a set of *time series* related to the fluctuations of the activity of each node on the graph. Such time series are often measured in the real systems and massive empirical data are currently available, for instance, the time series in the stock-price fluctuations in the stock market [3] and temporal fluctuations of the activity of routers in the Internet traffic [4]. Analysis of the time series and models reveals the occurrence of the long-range correlations and collective behavior related to the structure of the underlying dynamical system [4, 5, 6].

Studies of the correlations in time-series were attempted using various methods in order to unravel the underlying functional structure of the complex system, which results in these time series. Owing to the nonlinearity of the dynamics and the complexity (emergence) of the *collective dynamical effects* which do not occur at the level of isolated units, this problem represents a major challenge in contemporary science of complex systems. Often the mathematically elaborate methods theoretical models are necessary to interpolate between the observed data and true interactions in the system. On the other hand, massive data of genome-wide expressions measurements [7] are available, which contain hidden information about gene interactions. Assuming a model of gene dynamics, analysis of gene expression data yielded rather limited information about *pairwise* gene interactions (gene network) [8, 9]. Using formal analysis we recently studied the traffic time series on known modular networks [10]. We have shown that, unlike individual links, the network modularity (structure at the mesoscopic scale) can be revealed quite accurately from the correlations in the time series. The accuracy increases when the network is fully partitionable into modules. (See also Refs. [11] and [12] for different approaches to the problem of resolution of the network community structure.)

In this work we study correlations in the empirical time series of gene expressions measured [13] for each gene in the genome of Yeast *Saccharomyces cerevisiae (S.c.)*. Compared to the typical time series, say in the stock prices or Internet traffic, the time series of the gene expressions differ in the following: the expressions are sparsely measured (every 10 minutes) resulting in fewer number of points (the statistical importance of the data points is discussed in [13]); more importantly, the time evolution is naturally related to the *cell cycle*, therefore, when a cycle is completed (approximately 2 hours), the gene activity returns to the beginning of a new cycle. Our focus here is on the gene clustering that can be detected through the analysis of these time series. Our analysis is based entirely on the expression data without further reference to the transcription regulation between genes with already known transcription factors [7, 2]. We also restrict the analysis to a subset of 603 genes whose activity is most prominent in one of the phases of the cell-cycle (the "cell-cycle type"). Identification of the cell-cycle related genes and their mutual correlations is an important and still open problem (see [14] and references therein).

In Section 2 we present the statistical analysis of the genome-wide expression data and select the genes with leading contributions in the cell-cycle phases. We further analyze the correlations of the selected subset of genes in Section 3, where we determine the groups of genes using filtered correlation matrix and its eigenvalue spectrum. Section 4 is devoted to the study of topology of the network constructed on the basis of correlations, and Section 5 makes a brief summary.

2 Scale invariance in the expressions of CC genes

We consider empirical data for the time-course expression of the whole genome of yeast *S.c.*, measured in Ref. [13] at 17 equidistant time points along two full cell cycles. The statistical analysis of the data [15] (see also [2]) revealed scale-invariance in the ranking of the gene expressions and the broad distributions with power-law tails, which are characteristic for collective behavior, e.g., in self-organized dynamical systems. In particular, the average expressions of genes during the cell cycle obey a power-law ranking, also known as Zipf's statistics [21, 22], which indicates different contributions of genes to the dynamics during the cell cycle. Here we use the scale-invariance of the expression data to select the genes which contribute at most to the observed scale-invariance.

For this purpose we analyse the *differential expression* which is defined as $\Delta X_i(t) = h_i(t) - h_i(t-1)$ for each gene $i = 1, \cdots, N = 6406$ ORF (genes). The ranking distributions of all measured entries ($N \times 17$) is given in Fig. 1a, top curve. Similar ranking statistics for all N genes but for a single time at $t = 1, 8, 16$ are also given by other three curves. Broad distribution with the Zipf's law suggests *non-random correlations* for genes with the expressions above a threshold (marked by the horizontal line, which actually coincides with often used discrimination level $2 \times h_0$, where h_0 is the average expression in the whole system and for all instances of time). The occurrence of a smaller slope for genes with the largest expressions (upper part of the ranking curve) suggests that a community structure might be present with several genes playing a central role in each community. The scale-invariance of the expressions is also seen in the distribution for the differential expressions in Fig. 1b, again for the three time points and for all data. This distribution exhibits a power-law tail after a characteristic scale $\sim \Delta_0$, and can be fitted by the q-exponential form

$$P(\Delta X) = B_q \left[1 - (1-q)\frac{\Delta X}{\Delta_0} \right]^{\frac{1}{1-q}} \; ; \; q \neq 1 \quad , \tag{1}$$

where $q \simeq 1.36$ represents the non-extensivity parameter [20].

For the purpose of this work we select the genes with the expressions above the threshold $< h_i >> 2h_0$, as explained above and marked in Fig. 1a. The selected set consists of $N_s = 1216$ genes. Among these the first 612 in the ranking order are the genes which are expressed throughout the entire cell cycle, while the following 604 genes are mostly expressed within one or two of the cell cycle phases (G1,S,G2,M). In the following we will focus on this latter group of genes (cell-cycle genes). For the illustration, in Fig. 2 are shown the expressions for several cell-cycle genes within the first cycle.

For each pair of genes we compute the correlation coefficients of the temporal differential gene expressions $\Delta X_i(t)$, which is given by

$$C_{ij}(t-t') = \frac{\sum_t (\Delta X_i(t) - \langle \Delta X_i \rangle)(\Delta X_j(t-t') - \langle \Delta X_j \rangle)}{\sigma_i \sigma_j} \quad . \tag{2}$$

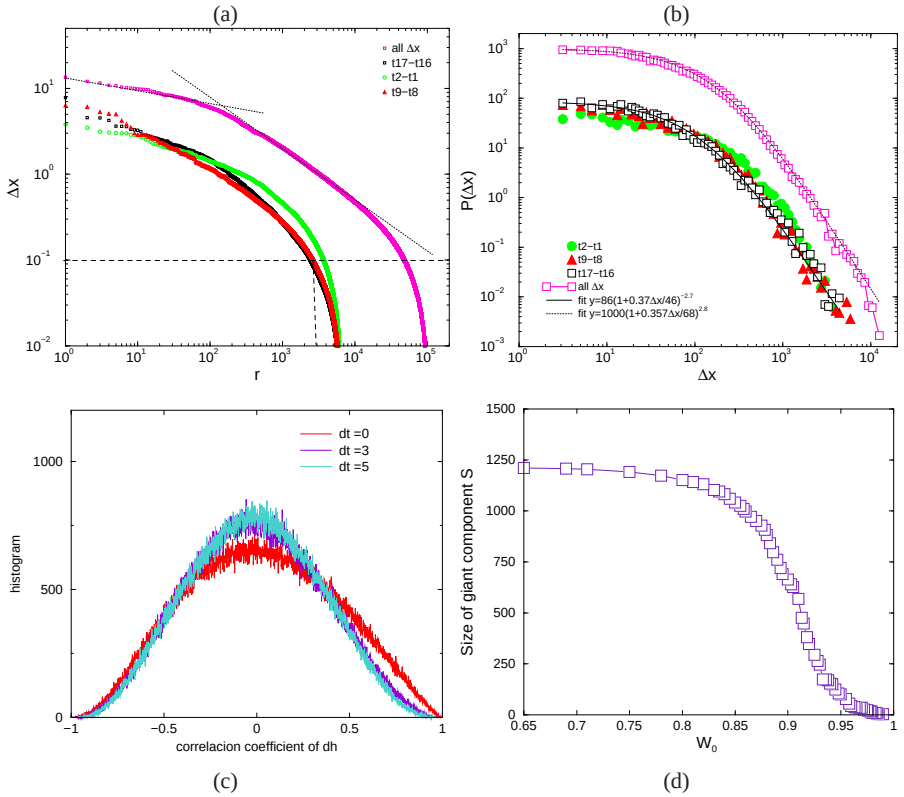

Fig. 1 (a) and (b) Statistical analysis of the differential expressions data for all genes in the genome of yeast *S.c.* shown in ranking order (a) and histogram (b). In both panels, top curve includes data of all measurements, while lower three curves are for all genes at a given instant of time in the cell cycle. The discrimination dashed line in (a) indicates how the first $N_s = 1215$ genes in the ranking order are selected. (c) For the selected genes: the distribution of correlation coefficients C_{ij} for co-expression ($dt = 0$) and for two time-delayed correlations. (d) Size of the giant component as a function of the threshold correlation W_0 in the selected set of N_s genes.

where σ_i and σ_j are the standard deviations of the respective time series of the genes i and j. The distribution of the correlation coefficients for the selected set of N_s genes is given in Fig. 1c, where $dt \equiv t - t'$ indicates the time-lagged correlations. As the Fig. 1c shows, in this set of genes the strongest deviations from normal distribution are seen in the equal-time correlations $dt = 0$, which we will consider in the following. In order to extract the relevant correlations, e.g., in the tail of the distributions in Fig. 1c, from the random correlations around zero, one considers only the correlations above a threshold value $|C_{ij}| > W_0$. However, for the analysis of the community structure, the threshold needs to be low enough such that all nodes of the network based on C_{ij} as the connectivity matrix belong to a single connected component. Large values of the threshold lead to fragmentation of such network. On

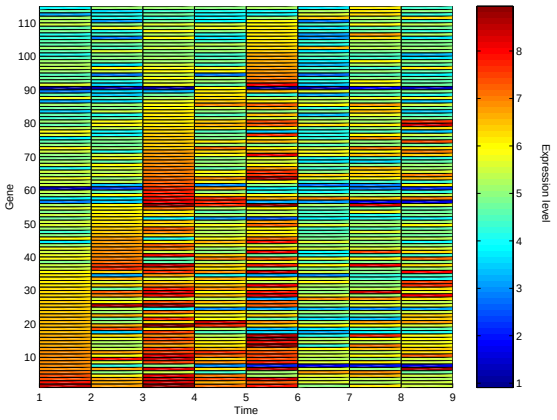

Fig. 2 Three-dimensional plot of the gene expressions within the first cycle for several leading Cell-Cycle genes ordered according the time of their maximum expression.

the other hand, we may observe how the giant cluster is formed by slowly reducing the threshold starting from $W_0 = 1$ (in Ref. [15] formation of the giant cluster for the entire genome of Yeast was studied). In the case of selected N_s genes which we consider here, the largest cluster S_{max} as function of the threshold W_0 forms slowly and all N_s genes are connected only at $W_0 \geq 0.6$, as shown in Fig. 1d. Therefore, we will consider the critical threshold $W_0 = 0.6$ for the study of the correlation matrix of N_s genes and their subset—CC genes, in the next session.

3 Detecting modules in the correlation matrix

As mentioned above, the network based on the correlation matrix **C** with the elements $C_{ij} > W_0 = 0.6$ is connected, however, it has very large number of links. As mentioned in the Introduction, in order to find a meaningful structure in the correlations, one needs to apply a filtering procedure in order to reduce the number of 'spurious' links. One of the methods which we use here is based on affinity transformation [17]. For the filtering procedure[17], the matrix elements C_{ij} of the correlation matrix are first mapped to the positive interval $[0,1]$. Then each element C_{ij} is multiplied with a factor M_{ij} which is constructed from the elements of the rows i and j in the correlation matrix: Excluding the diagonal elements C_{ii} and C_{jj}, the remaining matrix elements are first reordered to form the $n \equiv (N-1)$-dimensional vectors $\{C_{ij}, C_{i1}, ..., C_{iN}\}$ and $\{C_{ji}, C_{j1}, ..., C_{jN}\}$. Then with the components of these vectors the Pearson's coefficient M_{ij} is computed according to the general expression in Eq. (2). The matrix elements C^M_{ij} of the *filtered correlation matrix* $\mathbf{C^M}$ are given by the respective products

$$C^M_{ij} = M_{ij} C_{ij} \ . \tag{3}$$

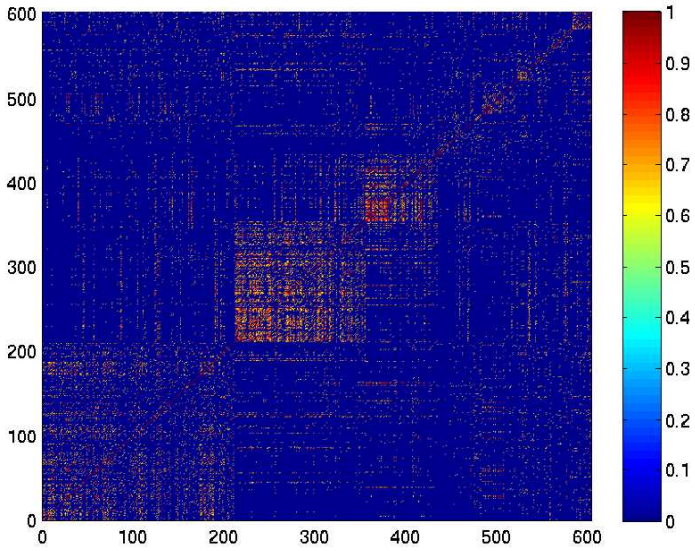

Fig. 3 Filtered correlation matrix showing several modules (diagonal blocks) of CC genes.

In this way, the correlation between the nodes i and j is enhanced if the corresponding meta-correlation element M_{ij} is large, i.e., the nodes i and j are connected to the rest of the system in a similar way (which will be also manifested in their time series), and reduced otherwise. Therefore, after the filtering procedure we expect to find enhanced correlations between 'similar' nodes, and reduced in the case when nodes play different role in the system. Applying then the same threshold as above, we find groups of nodes with enhanced correlations inside the group and generally reduced correlations (number of links above the threshold) outside the group. In Fig. 3 we show the filtered correlation matrix of the 603 CC genes. The filtered matrix clearly exhibits the modular structure with several modules, blocks along the diagonal, of different sizes. It is also clear that some of the modules are not entirely homogeneous, having a strongly expressed gene inside the module leads to enhanced correlations between the other genes in the module and with other modules. As shown in ref. [10] for the model network, the internal inhomogeneity of the modules leads to extra correlations between the modules, which can not be filtered out by standard methods. However, the size and the association of the nodes with the modules is correctly matched. We expect the same is true in the case of gene correlations. The structure of the modules can be visualized and analyzed in detail using the network representation, with the adjacency matrix defined by the elements of the filtered correlation matrix $\mathbf{C^M}$. In the following we first analyze the spectrum and subsequently the structure of the network.

3.1 Spectral analysis of the gene correlation matrix

The modular structure of the network can be also visualized via the spectral analysis of its connectivity matrix and other matrices, e.g., Laplacian, related to the structure [16]. Here we solve the complete eigenvalue problem of the filtered correlation matrix shown in Fig. 3. The matrix with weighted links (excluding the self correlations) above the threshold 0.6 is considered. The results for the eigenvalues λ_i are shown in Fig. 4, ranked according to their values. In the eigenvalue spectra of structured networks [23] the largest eigenvalue is separated from the rest of the spectrum and the components of its eigenvector are related to the eigenvector-centrality measure [1, 16]. In the presence of the structural modules (or communities), additional eigenvalues appear between the largest eigenvalue and the main part of the spectrum. The number of such eigenvalues is directly related to the number of topologically distinct modules. In the upper part of the plot in Fig. 4 six such eigenvalues occur, corresponding to six (sufficiently large and distinct) modules in the correlation matrix.

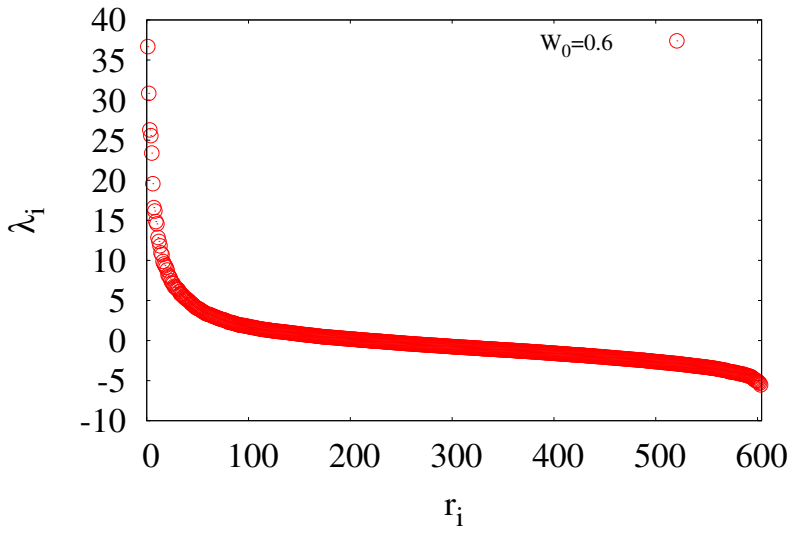

Fig. 4 Eigenvalues of the filtered correlation matrix of CC genes ranked according to value.

Further remarkable property of the eigenvalue problem of a modular network is illustrated by the localization of the eigenvectors of the large eigenvalues of its connectivity matrix (or similarly, of the small non-zero eigenvalues of the Laplacian matrix [16]) on the network modules. Formally, the term *localization* of the eigenvector $V_i \equiv V(\lambda_i)$ belonging to the eigenvalue λ_i denotes that it has a *nonzero*

component V_i^κ, corresponding to the node with the index κ. In modular networks the nodes with indexes κ corresponding to positive/negative values of the eigenvector components belong to different modules. Moreover, the extended branched view of the scatter plot indicates a structured network. In our case with the filtered correlation matrix we show the eigenvectors associated with the three large eigenvalues just below the λ_{max}. The 3-dimensional scatter plot of the components these eigenvectors $(V_1^\kappa, V_2^\kappa, V_3^\kappa)$ is shown in Fig. 5. Each point in this plot correspond to the same index κ of the three eigenvectors, and thus represents one node on the network. The separate branches in the scatter plot correspond to different modules on the network (diagonal blocks in its adjacency matrix). In our case each node represents a specific gene. Therefore, genes belonging to different branches can be identified by their names and other known biological properties [19]. For the illustration, the genes at the far ends in the scatter plot are node with the index (cf. matrix in Fig. 3) 471 corresponding to gene DIM1 or ORF "YPL266W"(most right in Fig. 5), index 130 corresponding to the gene CLB5 or "YPR120C" (at the front tip of the middle branch), index 367 or "YOR153W" and gene PDR5 (lower left tip), and index 221 representing the gene TY2B "YBL101W-b" (top of the middle branch in the scatter plot on Fig. 5). Identity of more genes grouped in the vicinity of these four are listed in the separate sections in the Table 1. Their biological properties, taken from the MIPS database [19], indicate that the genes in each group span the whole spectrum of different biological functions, however, they mostly share the same physical localization in the cell (nucleus, cytoplasm, mitochondria, cell periphery, etc.).

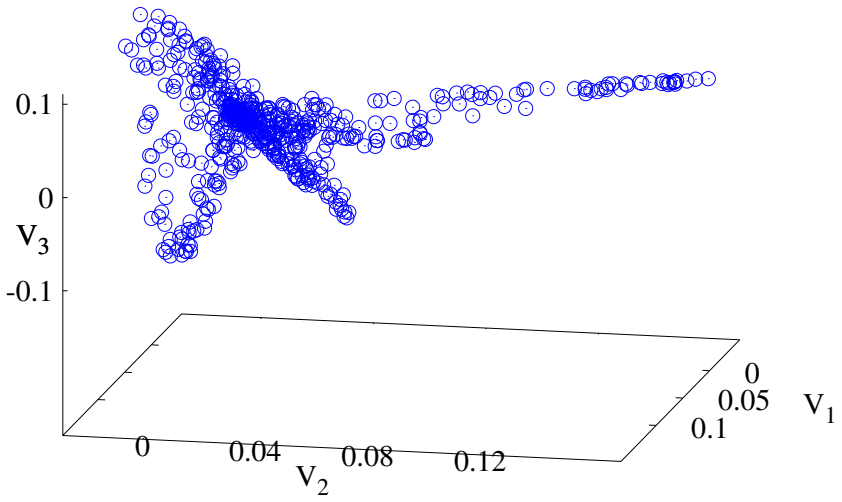

Fig. 5 Scatter plot: Components of the eigenvectors associated to three large eigenvalues $\lambda_3 = 25.568807$, $\lambda_2 = 26.276988$, and $\lambda_1 = 30.840979$. Most distant points on the plot correspond to the genes DIM1, CLB5, and TY2B, located in nucleus, cytoplasm, and mitochondria, respectively.

4 Structure of correlation network and group-identity of genes

The structure of the modules of co-expressed genes can be studied systematically by analysis of topology of the network whose adjacency matrix is based on the filtered correlation matrix. In Fig. 6a we show such network with nodes representing genes and edges indicating the expression correlations above the threshold (shown are only the links $C_{ij} > 0.8$). Colors (red, green, yellow, blue) on the nodes represent one of the phases (G1, S, G2, M) of the cell cycle where the gene has its peak expression. The presence of modules are seen on the network, however, the number

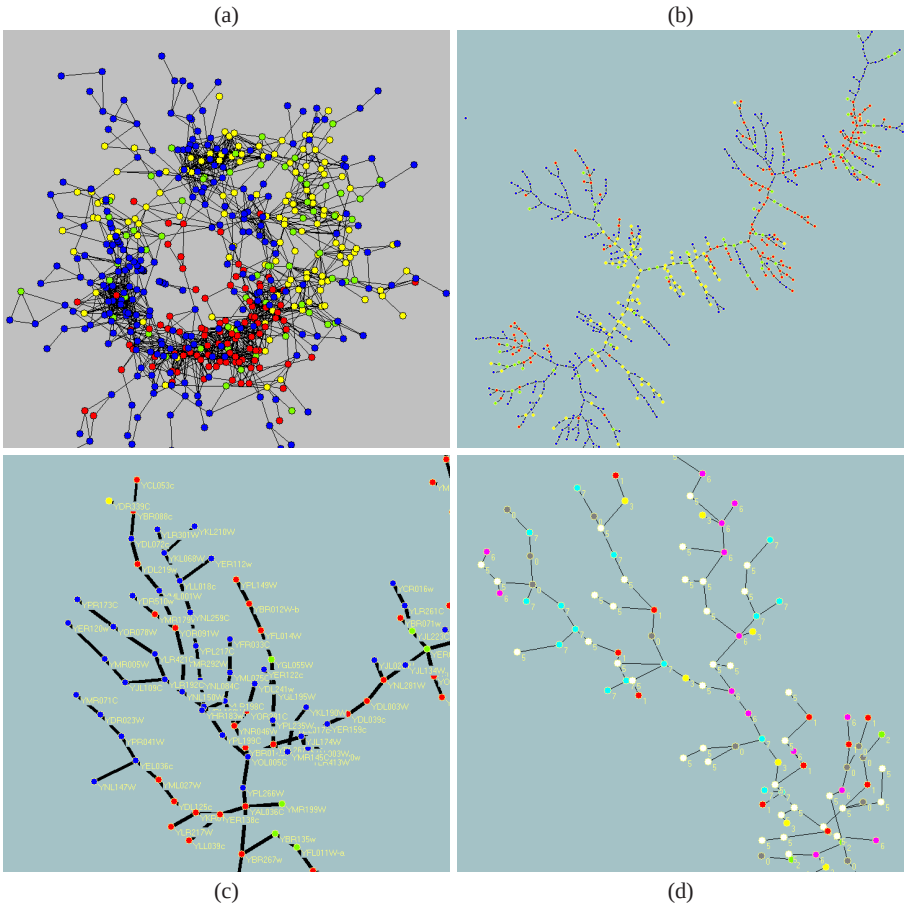

Fig. 6 (a) Network based on the filtered correlation matrix of CC genes (shown are links $C_{ij} > 0.8$)). (b) Maximum correlation spanning tree of the network. (c) Zoomed branch of the tree with genes marked by their ORF labels in the original data. Colors on nodes in (a-c) indicate the cell-cycle phases (G1:red, S:green, G2:yellow, M:blue) in which the gene has maximum measured expression. (d) Biological functions (grouped in seven categories) for genes within same branch.

of links inside these modules is too large. Therefore, we use the *spanning tree* as another way to visualize the structure of these weighted networks. We construct the spanning tree of the gene correlation network (shown in Fig. 6b), where each gene is connected to the rest of the tree by its *strongest* link (maximum correlation spanning tree). Separate branches on the tree correspond to seven modules in the correlation matrix (cf. Fig. 3). The predominant color in parts of the tree (branches) indicates gene groupings according to the phase of their peak activity. Detailed identification of the genes within a branch or sub-branch, an example is displayed in Fig. 6c with the names of genes, shows that genes within the same phase of the cell cycle have variety of biological functions. An example of the tree branch is shown in Fig 6d, with gene functions indicated by colors and numbers from 0-7 (for seven groups of known gene functions: metabolism, energy, cell cycle/cell rescue, transcription, protein synthesis, cellular transport, cell type differentiation and development, and unknown). The genes along the same branch of the tree have additional similarity in their localization in the cell. For instance, according to the MIPS database [19], the majority of the genes belonging to the branch in Fig. 6c are localized either in the nucleus or in the nucleus and cytoplasm.

5 Conclusions

We have studied correlations in the activity of genes of yeaast *S.c.* based on the empirical data [13] in the form of time series, in which the fluctuation in the expression of each gene are measured along the cell cycle. We employed three methods suitable for the analysis of *collective dynamical behavior* in self-organized systems:

- *Scale-invariance* in various statistical measures of the differential expressions suggests uneven role of different genes in the cell-cycle dynamics. With the ranking distribution (Zipf's law) we selected the genes with leading contributions to the scaling behavior, what implies prominent correlations within the selected set.
- *Spectral analysis* of the appropriately filtered correlation matrix is suitable for detecting modules of strongly correlated genes. We find several such modules of different size in the subset of the cell-cycle genes. The eigenvectors associated with the largest eigenvalues of the correlation matrix show a pattern of localization of the non-zero components on gene indexes belonging to different modules.
- *Network topology* based on the filtered correlation matrix (above a threshold) gives a systematic survey of the content of different groups (modules). Identifying the genes in different modules suggest that they share similarity in the phase of the cell cycle with their peak activity and also in the physical localization inside the cell, checked against the biological information in MIPS database.

We have shown that the genes co-expressed along the cell-cycle show certain patterns of correlated activity which is unraveled by the formal analysis within the network theory without further reference to bio-chemical interactions (transcription regulations) or a specific mathematical model for the gene dynamics.

Table 1 Characteristics of genes associated with four tips of the branches in the scatter plot in Fig. 5. Data according to MIPS database [19]. Abbreviations: UP (UNCLASSIFIED PROTEINS); BIOG (BIOGENESIS), M (METABOLISM), CC (CELL CYCLE), CT (CELLULAR TRANSPORT), E (ENERGY), CTD (CELL TYPE DIFFERENTIATION), TR (TRANSCRIPTION), ENV (Interaction with Environment), PF (regulation of PROTEIN FUNCTION), PM (PROTEIN FATE , PROTEIN MODIFICATION), PS (PROTEIN SYNTHESIS), DEV (DEVELOPMENT), CF (CELL FATE), CR (CELL RESCUE); Localization: CP (cell periphery), CYT (cytoplasm), NUC (nucleus), VAC (vacuole), Ec (extracellular), MIT (mitohondria), ER (endoplas.reticulum).

ORF reading	Gene Name	Functional Category	Disruption	Localization
YPL266W	DIM1	TRANSCRIPTION	letal	NUC
YOL005c	RPB11	DNA binding, TR	lethal	NUC
YPL267W	ACM1	UP	viable	NUC
YKR013W	PRY2	UP	viable	NUC, VAC
YNL281w	HCH1	CELL RESCUE	viable	CYT,NUC
YDL039c	PRM7	DEVELOPMENT	viable	?
YDL003W	MCD1	CC & DNA processing	lethal	CYT, NUC
YAL036c	RBG1	GTP binding	viable	CYT
YJL174W	KRE9	CTD, M, BIOG, E	lethal	Extracellular
YDL241W	-	UP	viable	CYT,NUC
YPL235W	RVB2	M, CC & DNA processing, TR	lethal	?
YBL101W-b	TY2B	Viral& PP	-	(cyt)
YLR257W	-	UP	viable	CYT
YFL035c-b	MOB2	CC & DNA, PM, M, PF	lethal	CP, CYT
YHR135C	YCK1	ENV, CC & DNA, PM, M, CTD	viable	PM, CYT, ER, NUC
YPL226W	NEW1	M, PB, PS	viable	CYT,NUC,MIT
YML085C	TUB1	DEV, CC & DNA processing, M	lethal	CYT, cytoskeleton
YKL001C	MET14	METABOLISM	viable	CYT
YDR245W	MNN10	CTD, M, BIOG, PM	viabl	ER, golgi
YPL032C	SVL3	CTD, CF, BIOG	viable	CP, CYT
YGR118W	RPS23A	PROTEIN SYNTHESIS	viable	CYT
YPL028W	ERG10	METABOLISM	lethal	CYT, NUC
YDR346c	-	unclassified	viable	CYT, NUC
YDL124w	-	PB, M	viable	CP, CYT, NUC
YBL064c	PRX1	CR, DEFENSE & VIRULENCE	viable	CYT, NUC, MIT
YJL217w	-	UP	viable	CYT
YBR194w	SOY1	UP	viable	CYT, NUC
YER001w	MNN1	M, PM	viable	golgi, VAC
YNR065c	YSN1	UP	viable	?
YDL166c	FAP7	CR, TRANSCRIPTION	letal	CYT, ER, NUC
YEL003w	GIM4	BIOG, PM, PB	viable	CYT
YDR098c	GRX3	CR, PM, CT	viable	CYT, NUC
YAR002c-a	ERP1	CT, PM	viable	golgi, MIT, VAC
YKL142W	MRP8-	PROTEIN SYNTHESIS	viable	CYT, MIT
YPR120C	CLB5	CC& DNA processing	viable	MIT
YKL190w	CNB1	TR, PB	viable	CYT
YPL135w	SU1	M, ENV	viable	MIT
YCR070w	CPR4	PROTEIN FATE MODIFICATION	viable	ER, VAC
YLR017w	MEU1	METABOLISM	viable	CYT, NUC
YML010W	SPT5	TR, CC&DNA processing	viable	NUC, MIT
YER074w	RPS24A	PROTEIN SYNTHESIS	viable	CYT, MIT
YBR048w	RPS11B	PROTEIN SYNTHESIS	viable	CYT
YOR153w	PDR5	ENV, CT, PB, CR	viable	CP, CYT, MIT

Acknowledgements Research supported by the program P1-0044 (Slovenia) and national project OI141035 (Serbia), and international projects BI-RS/08-09-047 and MRTN-CT-2004-005728. We used the computer system of the Department of theoretical physics, Jožef Stefan Institute, Ljubljana, and on the AEGIS e-Infrastructure, supported in part by EU FP6 and FP7 projects CX-CMCS, EGEE-III and SEE-GRID-SCI, at the Scientific Computing Lab, Institute of Physics, Belgrade.

References

1. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, *Physics Reports*, 424:175–308 (2006)
2. S. Tavazoie, J.D. Hughes, M.J. Campbell, R.J. Cho, and G.M. Church, Nature Genet. **22**, 281 (1999); F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, Proc. Natl. Acad. Sci. **101**, 4781 (2004); D. Balcan and A. Erzan, Eur. Phys. J. B **38**, 253 (2004)
3. R. N. Mantegna, Hierarchical structure in financial markets, *European Physical Journal B*, 11:193–197 (1999)
4. M. Takayasu, H. Takayasu and T. Sato, Critical behavior and 1/f noise in information traffic, *Physica A* 233:824-834 (1996)
5. B. Tadić, G.J. Rodgers and S. Thurner, Transport on Complex Networks: Flow, Jamming & Optimization, Int. J. Bifurcation and Chaos, **17**, Issue 7, 2363-2385 (2007)
6. M. Abel, K. Ahnert, J. Kurths, and S. Mandelj, Additive nonparametric reconstruction of dynamical systems from time series, *Phys. Rev. E*, 71(1):015203 (2005)
7. D. J. Lockhart, and E. A. Winzeler, Genomics, gene expression and DNA arrays, *Nature*, 405:827-836 (2000)
8. T. S. Gardner and J. J. Faith, Reverse-engineering transcription control networks, *Physics of Life Reviews*, 2:65–88 (2005)
9. D. Stokić, R. Hanel and S. Thurner, A fast and efficient gene-networks reconstruction method-from multiple overexpression experiments, ArXiv e-prints: (2008)
10. B. Tadić and .M. Mitrović, Jamming and correlation patterns in traffic of information on sparse modular networks, ArXiv e-prints: (2009)
11. S. Fortunato and M. Barthelemy, Resolution limit in community detection, PNAS **104**, 36-41 (2007)
12. L. Danon, A. Diaz-Guilera, A. Arenas, Effect of size heterogeneity on community identification in complex networks, J. Stat. Mechanics: Theory & Experiment, P11010 (2006)
13. R.J.Cho, Campbell, et. al., A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle, Molecular Cell **2** 65-73 (1998); http://arep.med.harvard.edu/cgi-bin/ExpressDByeast.
14. C. Caretta-Cartozo, P. De Los Rios, F. Piazza, P. Lio, Bottleneck Genes and Community Structure in the Cell-Cycle Network of S.pombe, PLoS Comput. Biol. 3(6): e103 (2007)
15. J. Živkovic, B. Tadić, N. Wick, S. Thurner, Statistical Indicators of Collective Behavior and Functional Clusters in Gene Expression Network of Yeast, European Physical Journal B **50**, 255 (2006)
16. M. Mitrović and B. Tadić, Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneity, *arxiv.org:0809.4850,* (2008)
17. A. Madi, Y. Friedman, D. Roth, T. Regev, S. Bransburg-Zabary, and E.B. Jacob, Genome holography: Deciphering function-form motifs from gene expression data, *PLoS ONE*, 3(7):e2708, (2008)
18. Z. Eisler and J. Kertesz, Random walks on complex networks with inhomogeneous impact, *Physical Review E*, 71(5):057104 (2005)
19. http:// mips.gsf.de
20. C. Tsallis, J. Stat. Phys **52**, 479 (1988)
21. G.K. Zipf, Psycho-Biology of Languages (Houghton-Mifflin, 1935; MIT Press, 1965)
22. C. Furusawa and K. Kaneko, Phys. Rev. Lett. **90**, 088102 (2003)
23. I.J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek, Spectra of real-world graphs: Beyond the semicircle law, *Phys. Rev. E*, 64(2):026704 (2001)