QIGTD: identifying critical genes in the evolution of lung adenocarcinoma with tensor decomposition

Bolin Chen[1,2]*, Jinlei Zhang[1], Ci Shao[1], Jun Bian[3], Ruiming Kang[4] and Xuequn Shang[1,2]

*Correspondence:
blchen@nwpu.edu.cn

[1] School of Computer Science, Northwestern Polytechnical University, Xi'an 710012, China
[2] Key Laboratory of Big Data Storage and Management, Northwestern Polytechnical University, Ministry of Industry and Information Technology, Xi'an 710012, China
[3] Department of General Surgery, Xi'an Children's Hosptial, Xi'an Jiaotong University Affiliated Children's Hosptial, Xi'an 710003, China
[4] Rewise (Hangzhou) Information Technology Co., LTD, Hangzhou 310000, China

Abstract

Background: Identifying critical genes is important for understanding the pathogenesis of complex diseases. Traditional studies typically comparing the change of biomecules between normal and disease samples or detecting important vertices from a single static biomolecular network, which often overlook the dynamic changes that occur between different disease stages. However, investigating temporal changes in biomolecular networks and identifying critical genes is critical for understanding the occurrence and development of diseases.

Methods: A novel method called Quantifying Importance of Genes with Tensor Decomposition (QIGTD) was proposed in this study. It first constructs a time series network by integrating both the intra and inter temporal network information, which preserving connections between networks at adjacent stages according to the local similarities. A tensor is employed to describe the connections of this time series network, and a 3-order tensor decomposition method was proposed to capture both the topological information of each network snapshot and the time series characteristics of the whole network. QIGTD is also a learning-free and efficient method that can be applied to datasets with a small number of samples.

Results: The effectiveness of QIGTD was evaluated using lung adenocarcinoma (LUAD) datasets and three state-of-the-art methods: T-degree, T-closeness, and T-betweenness were employed as benchmark methods. Numerical experimental results demonstrate that QIGTD outperforms these methods in terms of the indices of both precision and mAP. Notably, out of the top 50 genes, 29 have been verified to be highly related to LUAD according to the DisGeNET Database, and 36 are significantly enriched in LUAD related Gene Ontology (GO) terms, including nuclear division, mitotic nuclear division, chromosome segregation, organelle fission, and mitotic sister chromatid segregation.

Conclusion: In conclusion, QIGTD effectively captures the temporal changes in gene networks and identifies critical genes. It provides a valuable tool for studying temporal dynamics in biological networks and can aid in understanding the underlying mechanisms of diseases such as LUAD.

Keywords: Temporal network, Tensor decomposition, Critical genes, LUAD, Evolution

Chen *et al. BioData Mining*   (2024) 17:30

Page 2 of 14

Introduction

Complex networks are common in real life and the research has shifted from discovering the macro laws of structure and dynamics to uncovering the role of macro elements as nodes in real systems [1–5]. In the past several decades, designing effective centrality methods to identify critical nodes in complex networks has become a significant topic. Many methods have been designed to measure the importance of nodes in static networks, such as degree [6], closeness [7] and betweenness [8, 9]. Those centrality measures have been used in predicting essential proteins [10], identifying influential nodes from social networks [11, 12], finding critical links [13]. With the development and evolution of organisms, the structure of biological networks changes dynamically over time [14–17], by increasing or decreasing the number of nodes or edges [18, 19]. The research on dynamic biological networks and the identification of critical nodes are helpful to better understand biological processes [20].

At present, several methods have paying attention to identify critical genes in biological networks [21–23]. Liu et al. [24] identified potential critical genes related to the pathogenesis and prognosis of gastric cancer by protein-protein interaction (PPI) network and Cox proportional hazards. Li et al. [25] identified critical miRNAs, genes and transcription factors of lung adenocarcinoma by analyzing Gene Ontology terms, pathways, and PPI networks. Liu et al. [26] used the robust rank aggregation method, reconstructed the PPI network and performed modules analysis to identify critical genes. However, most of those methods are based on static network and ignoring the stage heterogeneity of complex diseases. He et al. [20] investigated miRNAs in serum exosome-like microvesicles to identify stage-common and stage-specific miRNAs, but ignored the connections between stages. Kim et al. [27] defined the temporal version of degree, closeness and betweenness on temporal networks, which reduced a dynamic network to a static one with directed flows. Nevertheless, those methods simply calculated the degree, closness and betweenness centrality of nodes in different time snapshots and obtained a mean value. The information of nodes changing with time would be lost. In our previous studies, we have proven that the studies of cancer stages is important for understanding the evolution of cancers [28, 29].

In this study, a lightweight and effective method that quantify the importance of genes with tensor decomposition (QIGTD) was proposed to identify the critical genes along with the progression of lung adenocarcinoma (LUAD). To start with, a time-series network was constructed to represent the molecular connections of individual pathological stages of LUAD, and a third-order tensor was employed to capture topological information of both intra-stage and inter-stage. The intra-stage topological connections were obtained from gene co-expression relationships, while the inter-stage topological connections were calculated by combining both local similarities and a pre-defined parameters. Then, a tensor decomposition method was proposed to identify critical gene from the temporal network, which considers not only the intra-stage topological information, but also the inter-stage temporal characteristics. It is also a learning free method, which can work well with a small amount of samples. The precision and mAP are presented to evaluate the performance of QIGTD, and the other three state-of-the-art methods: temporal versions of degree, betweenness and closeness [30, 31] were employed as benchmark methods. The overall framework of the proposed method was show in Fig. 1.
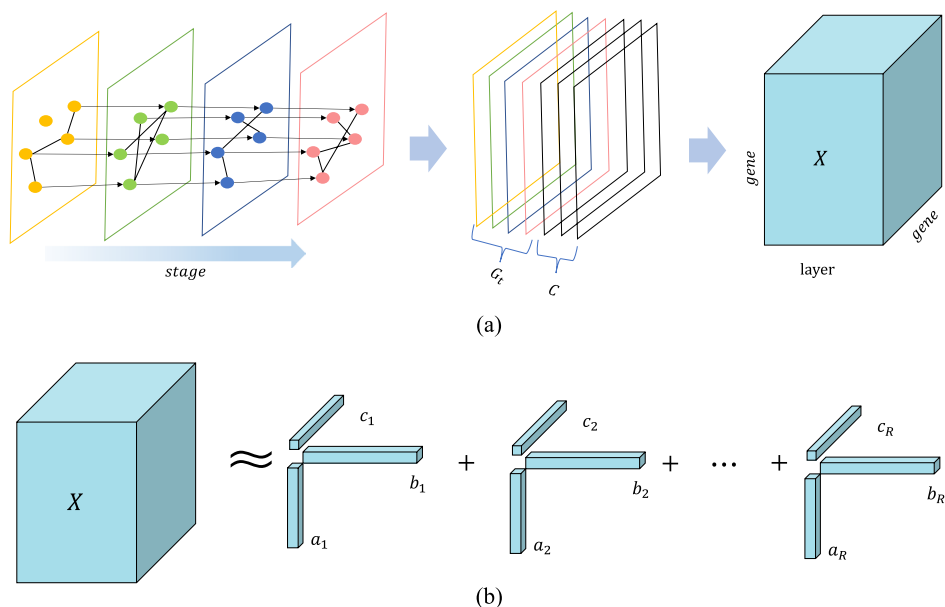
(a)

(b)

Fig. 1 The module of forming the tensor and decomposing tensor. **a** The way of constructing temporal network into tensor. The edges, consist of inter-stage edges and intra-stage edges are both taken into consideration. The yellow one is the network of Stage I, the green one is Stage II, the blue one is Stage III and the red one is Stage IV. The black represents inter-stage network. **b** The decomposition of tensor

Materials and methods

Data collection and processing

The critical genes are identified in the Stage I - Stage IV temporal networks of LUAD. The LUAD related gene expression dataset were downloaded from Xena (https://xenabrowser.net/), where there are 206 samples in Stage I, 93 samples in Stage II, 59 samples in Stage III, and 20 samples in Stage IV.

The networks of the four stages are constructed separately according to the *PCC* and the obtained *p*-value. In this study, the selection criteria were *p*-value< 0.01 and $|PCC| > 0.8$ according to the characteristics of biological networks. As a result, there are 17,830 edges in Stage I, 21,951 in Stage II, 11,170 in Stage III and 611 edges in Stage IV.

The known critical genes can be obtained from DisGeNET (https://www.disgenet.org/), where there are 3,899 genes appearing in the temporal network, and 566 of them have been verified to be associated with LUAD.

The temporal network construction

The network representing each stage of LUAD was constructed with pearson correlation coefficient (*PCC*) calculated with gene expression. Besides the connections within stage, there was also a fixed set of genes connected between different stages of LUAD.

Currently, there are two typical ways to construct connections between networks of adjacent stages. One is to use a fixed constant to represent the interlayer relationship, and the value of the parameter can indicate the strength of the interlayer relationship. The other method is to use similarity metrics to measure inter layer relationships. It was stated that the features in temporal networks are studied by converting time into

Chen *et al. BioData Mining*   (2024) 17:30

Page 4 of 14

a snapshot sequence of the network, so the similarity measurement method between nodes in static networks can be extended to the node relationships between adjacent layers.

In this study, a novel method for measuring inter layer node similarity in temporal networks was proposed by combining the calculation of node local similarity index with fixed parameters, which is

$$TLS_i^{(t,t+1)} = C + \frac{\sum_j w_{ij}^t + \sum_j w_{ij}^{t+1}}{2N} + \frac{|SN_i^{(t,t+1)}|}{N} \tag{1}$$

where C is a constant parameter that indicate the strength of the interlayer relationship, N is the number of vertices in the network, if $\sum_j w_{ij}^t = 1$, then the vertex i and vertex j has connection in the network of G_t, while if $\sum_j w_{ij}^t = 0$, then the vertex i and vertex j does not have connection. $SN_i^{(t,t+1)}$ represents the number of common vertices in two adjacent network G_t and G_{t+1}.

The first part of Eq. 1 is a constant parameter, which can be setup according to the experimental requirement. If a relatively small parameter was used, then it enhances the importance of vertices with high inter layer similarity in temporal networks, while selecting larger fixed parameters strengthens the importance of isolated vertices. In this study, the value of C was set to 0.5 based on the characteristics of the biological network. The second part of Eq. 1 is network local similarity, which represents the proportion of local neighbors of adjacent snapshot nodes in the entire network at different times. The third part of Eq. 1 represents the proportion of shared neighbors of adjacent snapshot vertices in the entire network. Hence, the overall value of *TLS* can characterize the degree of vertices in a temporal network and the inter layer relationship of node adjacency in different time snapshots. The larger the *TLS* value, the higher the probability of the node continuously appearing on two snapshot layers, and the more stable the node adjacency relationship.

The tensor description of the temporal network

The temporal network was represented as $X = \{G_t, C\}$. The G_t is the network of different stages of LUAD and C is the set of interconnections between different networks. The elements in C were concerned as 'cross network'. The temporal network could be represented in tensor as follows. Let $X \in \mathbb{R}^{I \times J \times K}$. The elements can be defined according to Formula 2.

$$X_{ijk} = \begin{cases} w_{ijk} & \text{if } X_{ijk} \in G_t \\ c_{ijk} & \text{if } X_{ijk} \in C \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $0 \le i < I, 0 \le j < J, 0 \le k < K$, the w_{ijk} is the element in G_t and c_{ijk} is the element in C.

The process to form the tensor from temporal network is shown in Fig. 1a, where different colors represent different stages. The edges between the different stages compose the cross network. There are four kinds of networks in G_t and three kinds of cross networks in C.

Chen *et al. BioData Mining*　　(2024) 17:30

Page 5 of 14

The tensor could be transformed into matrix by unfolding or flattening. In this study, we expanded the n-order tensor X along mode-n into a matrix X_n. The mode-1 corresponds to the 1-order of tensor, mode-2 corresponds to the 2-order of tensor and mode-3 corresponds to the 3-order of tensor. After the matricization of the tensor, the Kronecker, Khatri-Rao, and Hadamard products can be calculated respectively as follows.

$$A \odot B = [a_1 \otimes b_1 a_1 \otimes b_1 a_2 \otimes b_3 \cdots a_N \otimes b_N] \tag{3}$$

$$A \otimes B = \begin{vmatrix} a_{11}B & \cdots & a_{1N}B \\ \vdots & \ddots & \vdots \\ a_{N1}B & \cdots & a_{NN}B \end{vmatrix} \tag{4}$$

$$A * B = \begin{vmatrix} a_{11}B_{11} & \cdots & a_{1N}B_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1}B_{N1} & \cdots & a_{NN}B_{NN} \end{vmatrix} \tag{5}$$

The Canonical Polyadic (CP) decomposition of tensor

A tensor can be expressed as the sum of finite rank tensors. In this study, the 3-order could be decomposed as follows.

$$X \approx [[A, B, C]] = \sum_{r=1}^{R} a_r \circ b_r \circ c_r \tag{6}$$

where

$$X \in \mathbb{R}^{I \times J \times K}$$
$$A = (a_1, a_2, a_3, ..., a_R) \in \mathbb{R}^{I \times R}$$
$$B = (b_1, b_2, b_3, ..., b_R) \in \mathbb{R}^{J \times R}$$
$$C = (c_1, c_2, c_3, ..., c_R) \in \mathbb{R}^{K \times R}$$

The symbol "∘" is the outer product, the vector $a_r \in \mathbb{R}^I$ is column r of factor matrix $A \in \mathbb{R}^{I \times R}$, the vector $b_r \in \mathbb{R}^J$ is column r of factor matrix $B \in \mathbb{R}^{J \times R}$, and the vector $c_r \in \mathbb{R}^K$ is column r of factor matrix $C \in \mathbb{R}^{K \times R}$.

The outer product of these vectors is a rank one tensor, so the R rank-one tensors was used to approximate the original data, which is shown in Fig. 1b. By utilizing the factor matrix, the 3-order tensor can be decomposed as follows.

$$min_A \sum_{i,j,k} \left(x_{ijk} - \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr} \right)^2 = min_A \left\| X_{(1)} - A(C \odot B)\top \right\|_F^2$$
$$min_B \sum_{i,j,k} \left(x_{ijk} - \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr} \right)^2 = min_B \left\| X_{(2)} - B(C \odot A)\top \right\|_F^2 \tag{7}$$
$$min_C \sum_{i,j,k} \left(x_{ijk} - \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr} \right)^2 = min_B \left\| X_{(2)} - B(C \odot A)\top \right\|_F^2$$

The formulas can be approximately as

$$X_{(1)} \approx [A(C \odot B)\top]$$
$$X_{(2)} \approx [B(C \odot A)\top]$$
$$X_{(3)} \approx [C(B \odot A)\top] \tag{8}$$

Consequently, the $A_{[n]}$ could be calculated with back propagation and gradient descent. Since the goal is to make the tensor $\hat{X}$ estimated by A, B, C as close as possible to the original tensor X, the loss function is set as follows.

$$Loss_1 = \frac{1}{2}[X_{(1)} - A(C \odot B)\top]$$
$$Loss_2 = \frac{1}{2}[X_{(2)} - B(C \odot A)\top]$$
$$Loss_3 = \frac{1}{2}[X_{(3)} - C(B \odot A)\top] \tag{9}$$

The partial derivative of A, B and C could be quantified and the parameters could be updated by the following formulas.

$$A = A - \alpha * \frac{\partial Loss_1}{\partial A}$$
$$B = B - \alpha * \frac{\partial Loss_2}{\partial B}$$
$$C = C - \alpha * \frac{\partial Loss_3}{\partial C} \tag{10}$$

where α is the learning rate. The vertex centrality is now can be calculated as

$$s_i = \frac{1}{T} \sum_{t=1}^{T} ((a_1)_i(c_1)_t + (b_1)_i(c_1)_t) \tag{11}$$

In this study, where $I = J$ represents the number of genes, and $K = 4$ represents the number of layers in the network. The importance score of every gene could be determined with either $I = A \odot C$ or $I = B \odot C$. Additionally, if R is set to 1, so $I = A$ or $I = B$.

Results
The evaluation indices
The performance of QIGTD is evaluated by the precision, mean average precision (mAP) and fold enrichment.

The precision show the true positive ratio by giving a list of predictions, which is

$$precision = \frac{TP}{TP + FP} \tag{12}$$

The primary objective revolves around the task of ranking, where precision alone may not insufficiently reflect the algorithm's performance. The mAP does not only consider the accuracy of identifying the critical genes, but also considers the differences in genes order. More robustly, the mAP is utilized to reflect the model's performance, which can be defined as Formula 13.

$$AP(q_i) = \frac{\sum_{i \in i_1, i_2, \ldots, i_M} P_i \times L_i}{M} \tag{13}$$

where $P_i = \frac{1}{\sum_{j=1}^{i} j}$, and the L_i is the label of the i-*th* gene. In this study, the label comprises 0 and 1. Since there is only one query in the problem, so the mAP is equal to the AP.

Chen *et al. BioData Mining* (2024) 17:30

Page 7 of 14

Comparing with the benchmark methods

The important score of every gene in temporal network were calculated with three benchmark methods: T-degree, T-closeness, T-betweenness and the proposed QIGTD method.

The T-degree is defined as

$$T\text{-}deg(v) = \frac{\sum_{t=1}^{T} D_t(v)}{T} \tag{14}$$

where $D_t(v)$ represents the degree of vertex v at the tth network snapshot, and T is the total number of network snapshot.

The T-closeness is defined as

$$T\text{-}clo(v) = \sum_{1 \leq t \leq T} \sum_{u \in V \setminus v} \frac{1}{\Delta_{t,T(u,v)}} \tag{15}$$

where $\Delta_{t,T(u,v)}$ represents the shortest path length between vertex u and vertex v.

The T-betweenness is defined as

$$T\text{-}bet(v) = \sum_{1 \leq t \leq T} \sum_{s \neq v \neq d \in V} \frac{\sigma(s,d,v)}{\sigma(s,v)} \tag{16}$$

where $\sigma(s,d,v)$ represents the number of shortest path between vertex s and vertex v that through vertex d, while $\sigma(s,v)$ represents the number of all shortest path between vertex s and vertex v.

The precision of different methods are summarized in Table 1. The QIGTD consistently performs better than the other three state-of-the-art methods from the top 10 to top 500 predictions. In top 10, the precision of QIGTD is 0.50, while the best in the other three methods is T-betweenness with the precision of only 0.20. This situation also hold for predictions from the top 50 to top 500. The top three important genes calculated by QIGTD are highly related to LUAD.

The results of mAP@M, presented in Table 2, indicate that QIGTD outperforms the other three methods. QIGTD exhibits superior performance in accurately identifying the LUAD related genes without learning.

The fold enrichment is carried out to measure the performance of the model, which indicates how precisely the method can locate disease-related genes. QIGTD

Table 1 The precision of QIGTD and the three SOTA methods

Rank	T-deg.	T-clo.	T-bet.	QIGTD
Top 10	0.10	0.00	0.20	**0.50**
Top 50	0.22	0.08	0.16	**0.58**
Top 100	0.28	0.11	0.19	**0.52**
Top 150	0.27	0.15	0.17	**0.46**
Top 500	0.21	0.20	0.20	**0.23**

The rank is calculated with the four methods. And the precision is obtained by comparing with DisGeNET. For example, the precision of QIGTD is 0.50 in top 10 and it means there are 50% genes appear in DisGeNET among top 10 genes. The precision of QIGTD consistently surpasses that of the other three methods

Chen *et al. BioData Mining* (2024) 17:30

Page 8 of 14

Table 2 The mAP of the four methods

Rank	T-deg.	T-clo.	T-bet.	QIGTD
mAP@5	0.000	0.000	0.250	**0.458**
mAP@10	0.017	0.000	0.120	**0.214**
mAP@20	0.008	0.003	0.063	**0.125**
mAP@50	0.009	0.003	0.028	**0.061**
mAP@100	0.007	0.002	0.016	**0.034**

The mAP of QIGTD demonstrates higher performance compared to the other methods
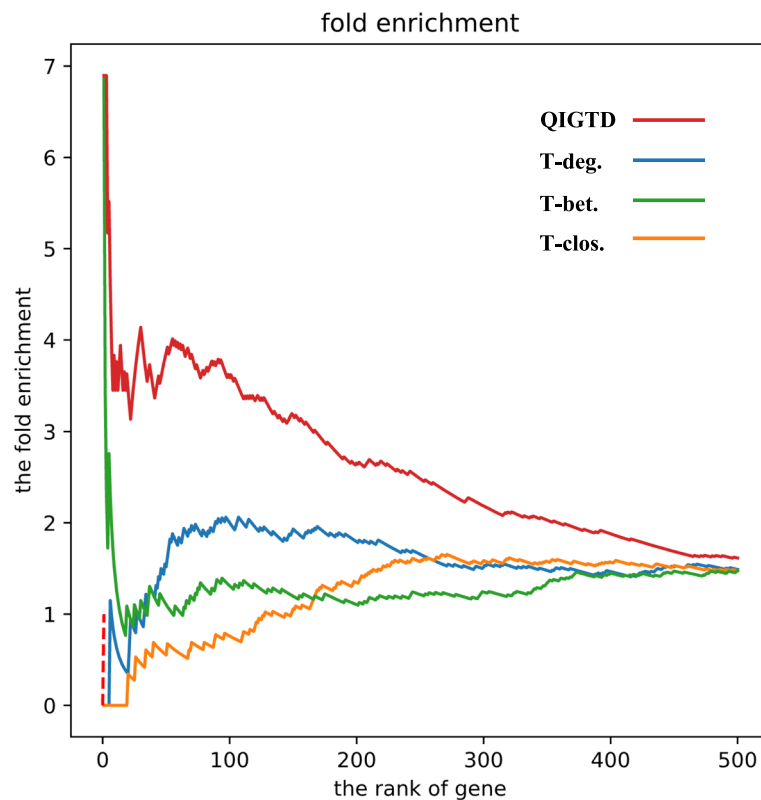
Fig. 2 The curve of fold enrichment in the top 500 genes. The x axis is the rank of genes in every method. The y axis is the score of fold enrichment. The fold enrichment could be calculated with precision and the correlation rate between all genes and LUAD

consistently exhibits significantly higher values compared to the other three methods as shown in Fig. 2.

Biological evidences of the predictions

Table 3 illustrated top 10 genes identified by the four methods as well as their rank. Among the top 10 genes identified by QIGTD, 5 genes are associated with LUAD in Dis-GenNET, indicating a higher level of association with the disease compared to the other methods.

Additionally, the rest 5 genes have the potential to become biomarkers of LUAD. The NCAPH was verified to be negatively associated with Mcl-1 in non-small cell

lung cancer [32]. Nguyen et al. [33] found that CDCA5 (cell division cycle associated 5) upregulated in the majority of lung cancers. The study of Wei et al. [34] found that the knockdown of HJURP inhibits non-small cell lung cancer cell proliferation, migration, and invasion. Coincidentally, there are many researchers found that BUB1 may hopefully become a novel marker and therapeutic target for LUAD [35–37]. The BUB1B was also identified to be a significant biomarker for a poor prognosis and poor clinicopathological outcomes in patients with LUAD [38].

The differentially expressed analysis is also performed on the top 10 genes, which is shown in Fig. 3. The blue box is the gene expression in control and the rest 4 boxes are that in four stages. The top 10 genes are obviously differentially expressed in stages compared to control.

Moreover, in top 50 genes identified by QIGTD, 29 genes are verified have a strong association with the disease, which is demonstrated in Table 4.

The sub-networks of top 10 genes are extracted as Fig. 4a. The figure shows the sub-graphs of top 10 genes of Stage I to Stage IV respectively. The thickness of the edge in the figure indicates the weight of the edge. The thick edges gradually decrease from Stage I to Stage IV, and some edges also disappear at Stage IV, thus the subgraphs of top 10 genes exhibit the evolution of LUAD.

The sub-networks of top 50 genes are extracted as Fig. 4b. The red nodes are top 10 genes, the purple are top 20 genes and the green are top 50. The thickness of the edges is not obviously as there are too many edges in the networks, but the sub-networks gradually become sparse with the stages, which shows a signal of the evolution of LUAD.

Among top 50 genes, 36 genes are enriched to 5 GO terms in Fig. 5. The GO terms are nuclear division, mitotic nuclear division, chromosome segregation, organelle fission and mitotic sister chromatid segregation, all of which have been verified to be associated with LUAD [39–41]. The different color of the ribbon represents the different GO terms. The numbers of ribbon means the number of GO terms that genes enrich. For example, the DLGAP5 has five ribbons, which means it enriches all 5 GO terms. The SPC25 has a green ribbon, which means it only enriches the chromosome segregation.

Table 3 The top 10 genes identified by the four methods and verified by DisGeNET

T-deg.	Dis.	T-clo.	Dis.	T-bet.	Dis.	QIGTD	Dis.
SASH3	0	CD4	0	ZEB2	1	TPX2	**1**
NCKAP1L	0	IKZF1	0	GIMAP8	0	NCAPG	**1**
CD53	0	ARHGEF6	0	CD53	0	KIF23	**1**
LCP2	0	PLEK	0	OLFML1	0	NCAPH	0
BTK	0	ACAP1	0	TCF4	1	KIF2C	**1**
PTPRC	1	C17orf87	0	WAS	0	CDCA5	0
IL10RA	0	MPEG1	0	STARD8	0	HJURP	0
CD4	0	FGD2	0	IL16	0	BUB1	0
PLEK	0	SELPLG	0	FLI1	0	KIF4A	**1**
ARHGAP25	0	MS4A4A	0	PLXDC2	0	BUB1B	0

There is the rank of top 10 genes calculated with the four methods. In the gene columns, genes are arranged in descending order based on the scores obtained in each method. In the disease columns, 1 represents the gene has been verified to be associated with LUAD in DisGeNET. The top 3 ranked with QIGTD have all verified in dataset. And 5 in top 10 genes have verified to be highly related to LUAD according to DisGeNET
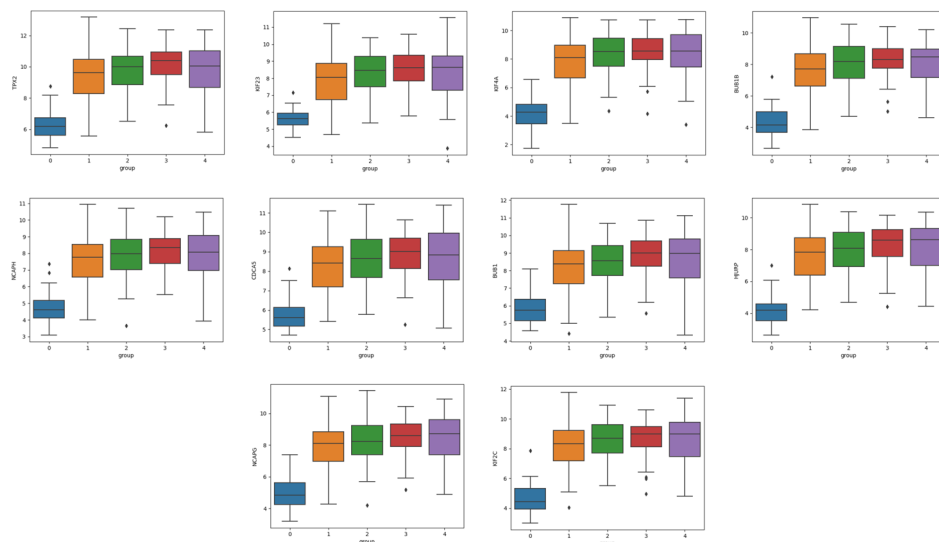
Fig. 3 The boxplot of the expression of top 10 genes. The different color in the plots is different stages. The blue box represents the expression in control. It shows that the genes identified with QIGTD are differentially expressed genes

Table 4 The top 50 genes identified by QIGTD and verified in DisGeNET

Gene	Score	Disease	Gene	Score	Disease
TPX2	0.13189	1	PLK1	0.11190	1
NCAPG	0.12907	1	CKAP2L	0.11183	1
KIF23	0.12859	1	TOP2A	0.10984	1
NCAPH	0.12847	0	NEK2	0.10982	1
KIF2C	0.12699	1	SPAG5	0.10966	1
CDCA5	0.12596	0	NDC80	0.10933	0
HJURP	0.12580	0	KIFC1	0.10894	0
BUB1	0.12566	0	GTSE1	0.10841	0
KIF4A	0.12427	1	ZWINT	0.10797	0
BUB1B	0.12413	0	CDC6	0.10767	0
CENPA	0.12363	1	KIF18B	0.10750	1
SGOL1	0.12327	0	RRM2	0.10728	1
MCM10	0.12241	1	SKA3	0.10700	0
TTK	0.12233	1	NUF2	0.10657	0
EXO1	0.12208	0	RAD54L	0.10550	0
CCNB2	0.12115	0	SKA1	0.10485	0
CCNA2	0.12071	1	KIF15	0.10482	1
KIF11	0.11880	0	SPC25	0.10462	1
DEPDC1	0.11584	1	KIF14	0.10388	1
DLGAP5	0.11541	0	FAM72B	0.10358	0
NUSAP1	0.11408	0	CDK1	0.10251	1
CDCA8	0.11397	0	FOXM1	0.10137	1
PRC1	0.11379	1	ESPL1	0.10133	1
CEP55	0.11346	1	ASPM	0.10064	1
CDC20	0.11205	1	PLK4	0.10004	1

The scores in the table represent the significance of each gene under the QIGTD method, where higher scores indicate a higher ranking. Among the top 50 genes, 29 genes are associated with LUAD
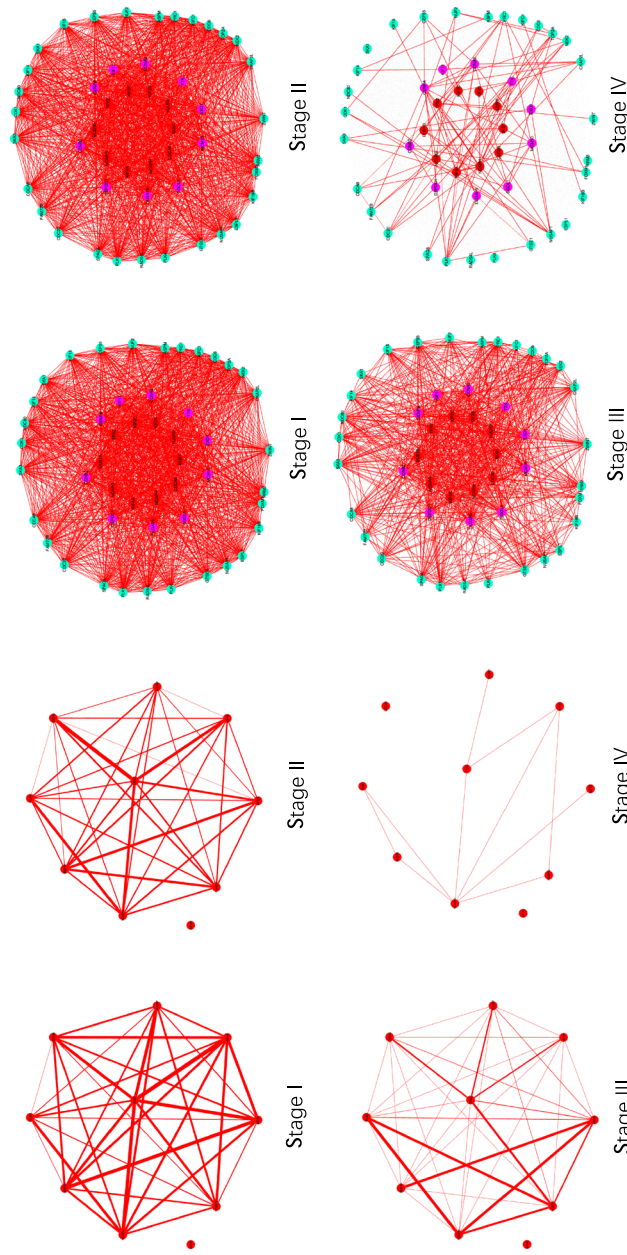
Fig. 4 The sub networks of Top 10 genes and Top 50 genes

Chen *et al. BioData Mining*   (2024) 17:30

Page 12 of 14

Discussion

The investigation of critical genes in temporal networks has become increasingly prevalent. Most of previous studies concentrate on the the structure of the network itself, but ignore the connections and changes between network at adjacent stages. Inspired by tensor decomposition, QIGTD is proposed in this research. Both the connections of genes inter and intra are taken into consideration.

The experimental results show that QIGTD outperforms the other three SOTA methods, especially in identifying the most critical genes. In the result, 5 genes in top 10 identified by QIGTD have been verified to be critical. At the same time, the other five genes may also be critical according to recent researches. The top 10 genes also differentially expression in stages compared to control. Furthermore, 29 genes are highly related to LUAD in top 50. The GO terms show indicate the top 50 genes ranked by QIGTD is associated with LUAD. The sub network of top 10 to top 50 undergoes changes across stages, which means the genes identified are potential to be biomarkers of the evolution of LUAD.
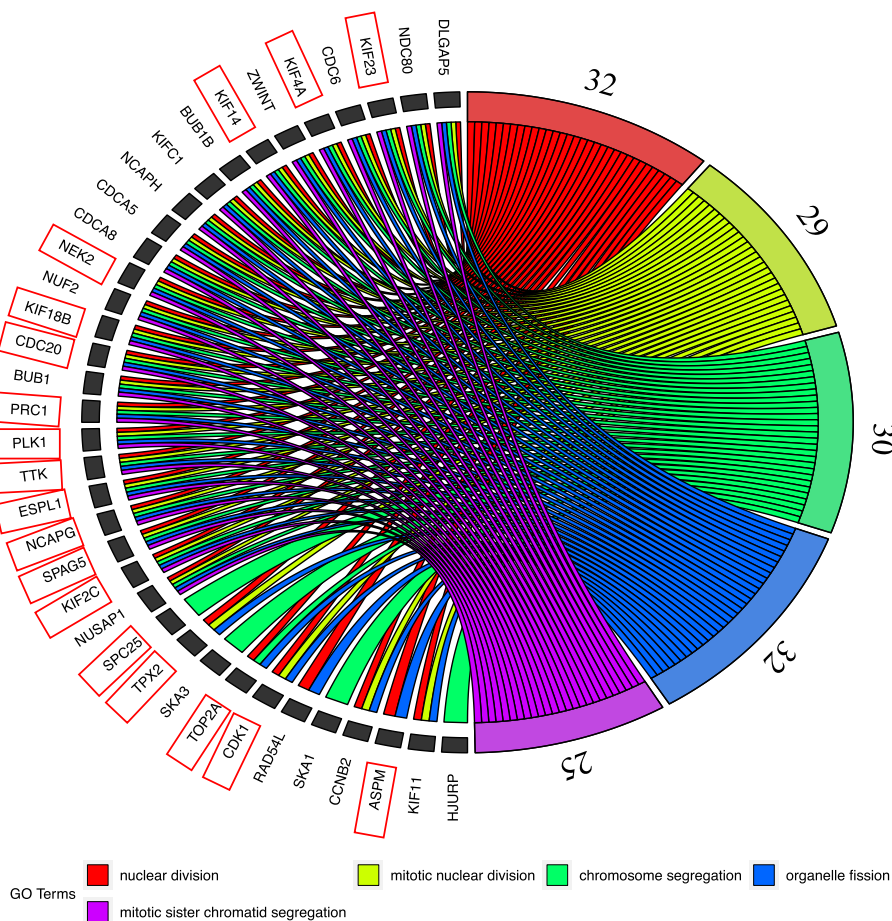

Fig. 5 The GO enrichment of top 50 genes. 5 Go terms chosen from the result of the enrichment are exhibited. The ribbons in different colors represents different GO terms. The number of the ribbons in gene is the number of GO terms it enriches. The boxed genes are LUAD-related genes. There are 32 genes enriched on nuclear division, 29 enriched on mitotic nuclear division, 30 on chromosome segregation, 32 on organelle fission and 25 enriched on mitotic sister chromatid segregation

Chen *et al. BioData Mining*   (2024) 17:30

Page 13 of 14

Additionally, QIGTD is a learning free and effective method, which does not require too many samples. The QIGTD has a low computational complexity and can be utilized in large-scale networks, which also could be easily embedded into the research of other complex problems.

Availability of data and materials
The publicly dataset of LUAD could be downloaded from Xena: https://xenabrowser.net/. The gene-disease association information could be collected via DisGeNET: https://www.disgenet.org/.

Declarations

Competing interests
The authors declare no competing interests.

References
1. Boccaletti S, Bianconi G, Criado R, Del Genio CI, Gómez-Gardenes J, Romance M, et al. The structure and dynamics of multilayer networks. Phys Rep. 2014;544(1):1–122.
2. Jin S, Li Y, Pan R, Zou X. Characterizing and controlling the inflammatory network during influenza A virus infection. Sci Rep. 2014;4(1):1–14.
3. Li Y, Jin S, Lei L, Pan Z, Zou X. Deciphering deterioration mechanisms of complex diseases based on the construction of dynamic networks and systems analysis. Sci Rep. 2015;5(1):1–11.
4. Morone F, Makse HA. Influence maximization in complex networks through optimal percolation. Nature. 2015;524(7563):65–8.
5. Lü L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T. Vital nodes identification in complex networks. Phys Rep. 2016;650:1–63.
6. Bonacich P. Factoring and weighting approaches to status scores and clique identification. J Math Sociol. 1972;2(1):113–20.
7. Freeman LC. Centrality in social networks conceptual clarification. Soc Networks. 1978;1(3):215–39.
8. Zhang J, Luo Y. Degree centrality, betweenness centrality, and closeness centrality in social network. In: Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017), vol. 132. Atlantis press. 2017. pp. 300–303.
9. Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977;40:35–41.
10. Tang X, Wang J, Zhong J, Pan Y. Predicting essential proteins based on weighted degree centrality. IEEE/ACM Trans Comput Biol Bioinforma. 2013;11(2):407–18.
11. Srinivas A, Velusamy RL, Identification of influential nodes from social networks based on Enhanced Degree Central-ity Measure. In: 2015 IEEE international advance computing conference (IACC). IEEE; 2015. pp. 1179–84.
12. Okamoto K, Chen W, Li XY. Ranking of closeness centrality for large-scale social networks. In: International workshop on frontiers in algorithmics. Springer; 2008. pp. 186–195.
13. Veremyev A, Prokopyev OA, Pasiliao EL. Finding critical links for closeness centrality. INFORMS J Comput. 2019;31(2):367–89.
14. Hintze A, Adami C. Evolution of complex modular biological networks. PLoS Comput Biol. 2008;4(2):e23.
15. Tenazinha N, Vinga S. A survey on methods for modeling and analyzing integrated biological networks. IEEE/ACM Trans Comput Biol Bioinforma. 2010;8(4):943–58.
16. Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, et al. Using graph theory to analyze biological networks. BioData Min. 2011;4(1):1–27.
17. Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007;8(6):450–61.
18. Kramer MA, Kolaczyk ED, Kirsch HE. Emergent network topology at seizure onset in humans. Epilepsy Res. 2008;79(2–3):173–86.
19. Gao W, Gilmore JH, Giovanello KS, Smith JK, Shen D, Zhu H, et al. Temporal and spatial evolution of brain network topology during the first two years of life. PLoS ONE. 2011;6(9):e25278.

20. Yang Q, He S, Huang L, Shao C, Nie T, Xia L, et al. Serum Exosomal miRNAs as Biomarkers of Early Diagnosis and Progression in Parkinson's Disease. Transl Neurodegener. 2021;10:25.
21. Liu X, Hong Z, Liu J, Lin Y, Rodríguez-Patón A, Zou Q, et al. Computational methods for identifying the critical nodes in biological networks. Brief Bioinform. 2020;21(2):486–97.
22. Abedi M, Gheisari Y. Nodes with high centrality in protein interaction networks are responsible for driving signaling pathways in diabetic nephropathy. PeerJ. 2015;3:e1284.
23. Rezaei J, Zare Mirakabad F, Marashi SA, MirHassani SA. The assessment of essential genes in the stability of PPI networks using critical node detection problem. AUT J Math Comput. 2022;3(1):59–76.
24. Liu X, Wu J, Zhang D, Bing Z, Tian J, Ni M, et al. Identification of potential key genes associated with the pathogenesis and prognosis of gastric cancer based on integrated bioinformatics analysis. Front Genet. 2018;9:265.
25. Li J, Li Z, Zhao S, Song Y, Si L, Wang X. Identification key genes, key miRNAs and key transcription factors of lung adenocarcinoma. J Thorac Dis. 2020;12(5):1917.
26. Liu L, He C, Zhou Q, Wang G, Lv Z, Liu J. Identification of key genes and pathways of thyroid cancer by integrated bioinformatics analysis. J Cell Physiol. 2019;234(12):23647–57.
27. Kim H, Anderson R. Temporal node centrality in complex networks. Phys Rev E. 2012;85(2):026107.
28. Chen B, Wang Y, Zhang J, Han Y, Benhammouda H, Bian J, et al. Specific feature recognition on group specific networks (SFR-GSN): a biomarker identification model for cancer stages. Front Genet. 2024;15:1407072.
29. Chen B, Chakrobortty N, Saha AK, Shang X. Identifying colon cancer stage related genes and their cellular pathways. Front Gen. 2023;14:1120185.
30. Tang J, Musolesi M, Mascolo C, Latora V, Nicosia V. Analysing information flows and key mediators through temporal centrality metrics. In: Proceedings of the 3rd Workshop on Social Network Systems. New York, Paris: Association for Computing Machinery; 2010. pp. 1–6. https://doi.org/10.1145/1852658.1852661.
31. Tsalouchidou I, Baeza-Yates R, Bonchi F, Liao K, Sellis T. Temporal betweenness centrality in dynamic graphs. Int J Data Sci Anal. 2020;9(3):257–72.
32. Xiong YC, Wang J, Cheng Y, Zhang XY, Ye XQ. Overexpression of MYBL2 promotes proliferation and migration of non-small-cell lung cancer via upregulating NCAPH. Mol Cell Biochem. 2020;468(1):185–93.
33. Nguyen MH, Koinuma J, Ueda K, Ito T, Tsuchiya E, Nakamura Y, et al. Phosphorylation and activation of cell division cycle associated 5 by mitogen-activated protein kinase play a crucial role in human lung carcinogenesis. Cancer Res. 2010;70(13):5337–47.
34. Wei Y, Ouyang G, Yao W, Zhu Y, Li X, Huang L, et al. Knockdown of HJURP inhibits non-small cell lung cancer cell proliferation, migration, and invasion by repressing Wnt/β-catenin signaling. Eur Rev Med Pharmacol Sci. 2019;23(9):3847–56.
35. Wang L, Li S, Wang Y, Tang Z, Liu C, Jiao W, et al. Identification of differentially expressed protein-coding genes in lung adenocarcinomas. Exp Ther Med. 2020;19(2):1103–11.
36. Jeganathan K, Malureanu L, Baker DJ, Abraham SC, Van Deursen JM. Bub1 mediates cell death in response to chromosome missegregation and acts to suppress spontaneous tumorigenesis. J Cell Biol. 2007;179(2):255–67.
37. Zhou X, Yuan Y, Kuang H, Tang B, Zhang H, Zhang M. BUB1B (BUB1 Mitotic Checkpoint Serine/Threonine Kinase B) Promotes Lung Adenocarcinoma by Interacting with Zinc Finger Protein ZNF143 and Regulating Glycolysis. Bioengineered. 2022;13(2):2471–85.
38. Chen J, Liao Y, Fan X. Prognostic and clinicopathological value of BUB1B expression in patients with lung adenocarcinoma: a meta-analysis. Expert Rev Anticancer Ther. 2021;21(7):795–803.
39. Sun ZY, Wang W, Gao H, Chen QF. Potential therapeutic targets of the nuclear division cycle 80 (NDC80) complexes genes in lung adenocarcinoma. J Cancer. 2020;11(10):2921.
40. Rao CV, Yamada HY, Yao Y, Dai W. Enhanced genomic instabilities caused by deregulated microtubule dynamics and chromosome segregation: a perspective from genetic studies in mice. Carcinogenesis. 2009;30(9):1469–74.
41. Chiang YY, Chen SL, Hsiao YT, Huang CH, Lin TY, Chiang IP, et al. Nuclear expression of dynamin-related protein 1 in lung adenocarcinomas. Mod Pathol. 2009;22(9):1139–50.

Publisher's Note