A Structural Analysis of the Hypoxia Response Network

jianjie li
Army medical university  https://orcid.org/0000-0002-6061-4464

Yuqi Gao (✉ gaoy66@yahoo.com )
https://orcid.org/0000-0001-9614-062X

Research

Keywords: Hypoxia response network, Complex networks, Network topology, Bow-tie structure, Structural complexity

Posted Date: June 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-27433/v1

License: ☑️ Read Full License

Version of Record: A version of this preprint was published at PeerJ on April 6th, 2021. See the published version at https://doi.org/10.7717/peerj.10985.
Abstract

**Background:** The hypoxia-inducible factor-1 (HIF-1) signaling pathway is an important topic in high-altitude medicine. Network analysis is a novel method for integrating information on different aspects and levels of biological networks. However, this method has not been used in research on the HIF-1 signaling pathway network. To introduce this method into HIF-1-related research fields and verify its feasibility and effectiveness, we used a network analytical method to explore the structural attributes of the HIF-1 signaling pathway network.

**Methods:** First, we analyzed the overall network of the HIF-1 signaling pathway using information retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG). We performed topology analysis, centrality analysis, and subgroup analysis of the network. Then, we analyzed the core network based on the overall network analysis. We analyzed the properties of the topology, the bow-tie structure, and the structural complexity of the core network.

**Results:** We obtained topological structure diagrams and quantitative indicators of the overall and core networks of the HIF-1 signaling pathway. For the structure diagrams, we generated topology diagrams of the network and the bow-tie structure of the core network. As quantitative indicators, we identified topology, centrality, subgroups, the bow-tie structure, and structural complexity. The topology indicators were the number of nodes, the number of lines, the network diameter, and the network density. The centrality indicators were the degree, closeness, and betweenness. The cohesive subgroup indicator was the components of the network. The bow-tie structure indicators included the core, input, and tendril-like structures. The structural complexity indicators included a power-law fitting model and its scale parameter.

**Conclusions:** The core network could be extracted based on the subgroup analysis of the overall network of the HIF-1 signaling pathway. The critical elements of the network could be identified in the centrality analysis. The results of the study show the feasibility and effectiveness of the network analytical method used to explore the network properties of the HIF-1 signaling pathway and provide support for further research.

**Background**

Hypoxia affects the work abilities and health of soldiers at high altitudes [1, 2]. Research on the hypoxia response network (HRN) is a necessary field of high-altitude medicine that aims to promote the health and work abilities of soldiers at high altitudes. The hypoxia-inducible factor-1 (HIF-1) signaling pathway is one of the crucial components of the HRN [3, 4]. Most studies on the HRN have been carried out based on classical biomedical methods, and these studies have made remarkable achievements. With the development of systems biology, researchers have tried to explore the HRN from a systematic view to compensate for the shortcomings of traditional research methods. Kohn et al. [5] proposed a theoretical HRN model based on ordinary differential equation behavior. They analyzed the structural composition of core subsystem responsible for HRN conversion. The network
has been decomposed into multiple primary paths using the extreme path analysis method [6]. It can be well matched with the consensus of existing research, showing that the path switching or branching effect may be the cause of the intense response to the oxygen concentration. Heiner et al. [7] subsequently constructed the general Petri net structural model of the HRN and analyzed the relative modules and properties of the network algorithmically. Additionally, network analysis methods and tools have been used to analyze biological networks [8], providing a novel perspective for studying traditional biological problems. Ding et al. [9] studied the structural and functional properties of the giant strong component of the \textit{B. thuringiensis} metabolic network. Zhang et al. [10] performed a protein-protein interaction network analysis of insecticide resistance molecular mechanisms. Ma et al. [11] studied the connectivity of the metabolic networks of 65 biological species, which did not include the core network of HIF-1 signaling pathways. They found that these biological metabolic networks are similar to the Internet in their macrostructure, which also presents a bow-tie structure. Network analysis methods have brought about new knowledge of the HRN. They are useful methods that focus on the properties of the connections of things and investigate these connections as a whole [12]. However, this method has not been used to analyze the HRN.

This study describes an analytical method for the hypoxia network involving the analysis of the quantitative network indicators of the HIF-1 signaling pathway from the perspective of complex networks. It aims to explore the structural properties of this network and verify the feasibility of the network analytical method. We analyzed the structural attributes of this network after constructing it from the corresponding biomedical database and then checked the complex properties of the network.

**Methods**

**Datasets**

This study focused on the structure of the hypoxia-inducible factor-1 (HIF-1) [13, 14] signaling pathway of the HRN and retrieved related information from the entries in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. KEGG provides data resources related to the high-level functions and operations of cells, organs, ecosystems, and other life systems based on molecular-level information, especially information used for parsing molecular datasets generated by gene sequencing and other high-throughput experimental techniques. There are two essential elements of KEGG discussed in this paper: the KEGG PATHWAY and KEGG ORTHOLOGY databases. The KEGG PATHWAY database is a collection of knowledge-based hand-drawn maps of pathways based on molecular interaction networks covering metabolic, genetic information processing, environmental information processing, cellular processes, biological systems, human diseases, and drug discovery. KEGG PATHWAY provides a concrete explanation of the meaning of the markers of the HIF-1 signaling pathway. The KEGG ORTHOLOGY (KO) database is a set of ortholog assemblages manually defined to denote the nodes (boxes) in the KEGG PATHWAY maps. The distinctive identifier that determines each KO entry is referred to as the K number ('K' followed by a five-digit number).
We analyzed the data in the KEGG Markup Language (KGML) file. These data contain information that corresponds to the HIF-1 signaling pathway in KEGG PATHWAY. The KGML, which enables the automated mapping of KEGG pathways, is an extensible markup language (XML) representation of the KEGG pathway. It is conducive to the computer-aided analysis and model building of gene/protein and chemical networks. There are two types of graphic elements in the KGML-based metabolic pathways. One is a rectangle, which represents an enzyme, connected by a "relationship." The other is a circle, which represents a compound, connected by a "reaction." In non-metabolic pathways, there are only rectangular elements, which indicates that there are only proteins joined by "relationships" in these pathways. The HIF-1 signaling pathway addressed in this study belongs to the metabolic pathways.

Overall Analytical Methods

The methods of this study are based on network analysis. The network of the HIF-1 signaling pathway is a metabolic network. It is a directed unprivileged network; i.e., the direction of a line between two nodes needs to be considered rather than the weight in the network. The corresponding analytical methods applied in this study are as follows: Process the text in the KGML file of the HIF-1 signaling pathway. Refine the network topology information and transform it into a network file that is identifiable for network analysis. Analyze the network indicators and draw the network topology maps. Analyze and test the structural complexity of the network. The generation of the network file is based on regular expression using the text-editing software Notepad++ and functions using the spreadsheet software Excel. The process of network analysis is based on network analysis models using Pajek. These methods and tools are comprehensively utilized to realize the qualitative and quantitative integration analysis of the network structural properties of the HIF-1 signaling pathway.

Topology

The network topology indicators include the node number, the line number, the line value, the network density, and the network diameter. The node is an entity that forms the network. The lines between the nodes represent the connections between network entities. The line value is the weight of a line, which indicates the strength of the relationship between network entities. The diameter of the network is the longest path of the shortest distance between pairs of nodes in the network, which is the number of maximum steps required to connect any pair of nodes in the network. The network density is equal to the numeric ratio of the actual connections to the possible connections. It indicates the degree of closeness of the relationships between the nodes in the network.

Centrality

Centrality is an essential concept of network analysis. A highly centralized network supports the convenient transmission of information. The central node has a critical influence on the transmission of information.
information in the network. The metrics for centrality used in this paper include the centrality of the degree, the closeness, and the betweenness.

**Degree centrality** The degree centrality of a node is defined as the number of connections of a node. It provides the most intuitive conceptual form of centrality indicator. It can be classified as the overall, input, or output degree centrality. These values correspond to the total number of connections, the number of input connections, and the number of output connections of the node, respectively. The regular formula of the degree centrality is as follows [15]:

\[ DC_i = Din_i + Dout_i \]

where \( DC_i \) represents the overall degree, \( Din_i \) represents the input degree, and \( Dout_i \) represents the output degree.

**Closeness centrality** A node's closeness centrality is the value obtained by dividing the number of all other nodes by the sum of the geodesic distance between the node and all other nodes. The geodesic distance is the number of connections included in the shortest path between the two nodes. The farther away a node is from other nodes, the lower the closeness centrality of the node is, and vice versa. Similar to the degree centrality, the closeness centrality of the nodes involves the overall closeness centrality, the input closeness centrality, and the output closeness centrality. The closeness centrality of a node reflects the nearness of a node to other nodes. The closer a node is to other nodes, the easier it is for information to reach the node, and the higher its closeness centrality is. The regular formula of the closeness centrality is as follows [15]:

\[ CC_i = \frac{N - 1}{N} \frac{1}{\sum_{j=1}^{N} d_{ij}} \]

where \( i \neq j \), is the number of nodes, and \( d_{ij} \) is the shortest pathway between nodes and .

**Betweenness centrality** A node's betweenness centrality is the ratio of the shortest pathways passing through this node to all of the shortest pathways between any two nodes in the network. The degree centrality and closeness centrality are based on the reachability of a node in the network. In view of betweenness centrality, if this indicator of a node is high, its importance as an intermediary node in the network is higher, and its bridging ability is active.

The normal formula for the betweenness centrality is as follows [15]:

\[ \text{Loading [MathJax]/jax/output/CommonHTML/jax.js} \]
\[ BC_i = \frac{2 \times \sum_{j \leq k} g_{jk}(i) / g_{jk}}{(N - 1)(N - 2)} \]

where \( i \neq j \neq k \), \( g_{jk} \) represents the number of the shortest pathways between nodes and \( g_{jk}(i) \) represents the number of the shortest pathways containing \( i \), and \( g_{jk} \) represents the number of nodes.

### Cohesive Subgroups

Some entities in the network relate to each other so tightly that they form a small local group known as a cohesive subgroup. The number of cohesive subgroups in the network as well as the local and global associations between subgroups are analyzed via cohesive subgroup analysis. This analysis includes component analysis, K-core analysis, and island analysis based on different analytical indicators and perspectives [16]. The component analysis is mainly applied in this study. The component is an essential indicator of the cohesive subgroup analysis, which refers to the largest connected subnetwork in the network; i.e., there is a way to reach other nodes between any nodes in the subnetwork. Biometabolism networks often contain components that are not connected, and the most significant component is often the one that needs attention.

### Bow-tie Structure

The bow-tie property is an essential component of these properties. Border [17] et al. revealed that there is a bow-tie structure in the topology structure of the Internet at the macro level.

Ma et al. [11] divided the macrostructure of the biological metabolic network into four parts: the giant strong component (GSC), substrate subset (S), product subset (P), and isolated subset (IS). These components correspond to the strongly connected core, input, output, and disconnected components of the Internet bow-tie structure, respectively, while the exact opposite situation is not observed. Any two metabolites within GSC can be generated from each other by a series of reactions. Any metabolite in S can be converted into the corresponding metabolite in GSC, but not vice versa. Any metabolite in P can be transformed via a series of the corresponding metabolites in GSC, while the exact opposite situation is not observed. The metabolites in IS cannot be converted from those in GSC, nor can they be converted into the corresponding metabolites in GSC.

### Structural Complexity

Qian et al. [18] provided a strict definition of complex networks. A network with some or all of the following characteristics is known as a complex network. These characteristics are self-organization, self-

Loading [MathJax]/jax/output/CommonHTML/jax.js
sophisticated attributes of the network. The scale-free property refers to the invariance of the network's scale, meaning that the degree of the network nodes obeys the law of a power exponential distribution (i.e., the power-law distribution). It is also known as the Pareto distribution rate or the Zipf rate. This distribution law describes the distribution characteristics of the degree of network nodes. A small number of nodes in the network tend to present a large number of connections, but the number of connections of most nodes is very limited. This paper uses the Pareto distribution rate:

\[ P_r[X \geq x] \sim x^{-a} \]

where \( P_r \) is the probability corresponding to the actual measurement scale of the network, and \( a \) is the scale parameter of the power rate distribution.

The degree distribution law of the network nodes needs to be fitted to the power exponential distribution law to verify this characteristic of the core network of the HIF-1 signaling pathway. That is, the fitting value, \(^\wedge a\), of the scale parameter, \( a \), is obtained. Then, the characteristic is judged by the fitting effect.

**The Power-law Fitting Of The Degree Distribution**

The general method of power-law fitting is to use the least-squares linear fitting based on double-logarithmic coordinates and to use the \( R^2 \) test to measure the fitting effect. However, Goldstein et al. believe that the scale result obtained by this method shows a significant error in relation to the corresponding value of the actual measurement network scale [19]. Therefore, Clauset [20] and Barabási [21] et al. proposed a maximum likelihood estimation method for power rate fitting and tested the fitting results using KS statistics. At present, no research shows that all the data corresponding to the actual measurement scale of the studied network obey the power-law distribution, but there is a critical value. The data corresponding to these scales obey the power-law distribution only when the value corresponding to the actual measurement scale, \( x_{\text{min}} \), is higher than \( x_{\text{min}} \). This study used Clauset’s universal method, applicable to both discrete and continuous data, to estimate \( x_{\text{min}} \). For actual network-scale data, the following formula is used to estimate the scale parameter, \(^\wedge a\), of the power-law distribution:

\[
^\wedge a \approx 1 + n \left[ \sum_{i=1}^{n} \ln \frac{x_i}{x_{\text{min}} - 0.5} \right]^{-1}
\]

where \( x_{\text{min}} \) can be accurately obtained by calculating the maximum difference, \( \Delta \), between the actual measurement-scale data and the corresponding data of the fitted model.
\[ L = \max_{x \geq \min} |S(x) - P(x)| \]

where \( S(x) \) is the corresponding data of the actual measurement, \( P(x) \) is the corresponding data of the fitted power-law distribution model, and \( x_{\min} \), which minimizes \( L \), is the optimal value.

**Similarity Test**

This study used the K-S (Kolmogorov-Smirnov) method to test the "distance" between the actual measurement data and the fitted power-law distribution model. The model constructed from actual measurement data is denoted as \( S(x) \). This model produces sets of data. There is a set of data whose "distance", \( m/n \), from the corresponding fitted model, \( P(x) \), is greater than the "distance", \( m/n \), between the actual measured data and the fitted model, \( S(x) \). The number of such datasets is \( m/n \) is represented as and is known as the value. If the value is large (close to 1), it can be considered that the statistical fluctuations alone cause the difference between the actual measurement data and the fitted model. If the value is small, there is room for adjustment in the rationality of the fitted model. If \( p \leq 0.1 \), it can be considered that the actual measured data do not obey the power-law distribution [22].

**Results**

Figure 2 shows the overall topology of the HIF-1 signaling pathway network based on the KGML file from KEGG. Quantitative analysis of the network indicators is performed, and the results are as follows: The numbers of nodes and lines are 85 and 61, respectively. The network has a ring and ringless density of 0.00844291 and 0.00854342, respectively. The average node degree (i.e., the average number of connections) is 1.43529412. The network diameter is 7, and the average distance between the nodes is 2.26761. There are 59 orthologous enzymes, 16 compounds, eight pathways, and two unknown groups in this metabolic network that have no references in KEGG. The numbers of activation reactions, expression reactions, and inhibition reactions are 29, 26, and 6, respectively.

The core network of the HIF-1 signaling pathway is identified based on cohesive subgroup analysis. The network is extracted for the convenience of further study, as shown in Figure 3. The quantitative indicators of the network are as follows: the number of nodes is 34, and the number of lines is 34. The network has a ring density of 0.02941176 and a ringless density of 0.03030303. The average node degree is 2. The network diameter is 7, and the average distance between the nodes is 2.57018. There are 24 orthologous enzymes and 10 compounds in this metabolic network. The number of unknown groups that have no references in KEGG is 1; the number of activation reactions is 27; the number of expression reactions is 2; and the number of inhibition reactions is 5.
Centrality Analysis

**Degree centrality** The degree centrality of the overall network of the HIF-1 signaling pathway is analyzed and includes the overall centrality, the input centrality, and the output centrality. The top 10 nodes are listed in Table 1.

| Rank | Overall Degree Centrality | Input Degree Centrality | Output Degree Centrality |
|------|---------------------------|--------------------------|--------------------------|
|      | Node Number | Value | Node Number | Value | Node Number | Value |
| 1    | 85          | 24    | 6           | 6     | 85          | 24    |
| 2    | 6           | 7     | 10          | 4     | 24          | 2     |
| 3    | 10          | 4     | 61          | 2     | 23          | 2     |
| 4    | 20          | 4     | 20          | 2     | 20          | 2     |
| 5    | 61          | 3     | 74          | 2     | 78          | 2     |
| 6    | 24          | 3     | 60          | 1     | 17          | 2     |
| 7    | 23          | 3     | 80          | 1     | 63          | 1     |
| 8    | 78          | 3     | 29          | 1     | 62          | 1     |
| 9    | 17          | 3     | 78          | 1     | 61          | 1     |
| 10   | 60          | 2     | 58          | 1     | 60          | 1     |

Four nodes appeared in the top 10 list in terms of the three-degree centrality indicators, numbered 20, 60, 61 and 78. They were K07203, K10160, K(04735, 02580), and K04371 in KEGG, respectively.

**Closeness centrality** The analysis of the closeness centrality of the overall network of the HIF-1 signaling pathway is performed to calculate the total closeness centrality, the input closeness centrality, and the output closeness centrality. Table 2 lists the top 10 nodes in terms of the closeness centrality. There are two nodes with both the overall and the input closeness centrality ranked in the top 10, numbered 29 and 82. They are K12077 and K02161 in KEGG, respectively. The number of the node with the overall and output closeness centrality in the top 10 is 85, which is undefined in KEGG. The node with the input closeness centrality and output closeness centrality ranked in the top 10 is numbered 20, which is K07203 in KEGG.
Table 2
Top 10 nodes of the overall network closeness degree centrality of the HIF-1 signaling pathway

| Rank | Overall Closeness Centrality | Input Closeness Centrality | Output Closeness Centrality |
|------|-------------------------------|----------------------------|----------------------------|
|      | Node Number | Value | Node Number | Value | Node Number | Value |
| 1    | 85          | 0.2941 | 10          | 0.1008 | 85          | 0.2941 |
| 2    | 29          | 0.1502 | 6           | 0.0647 | 17          | 0.0462 |
| 3    | 58          | 0.1502 | 74          | 0.051  | 24          | 0.0449 |
| 4    | 57          | 0.1502 | 20          | 0.0424 | 12          | 0.0398 |
| 5    | 82          | 0.1502 | 61          | 0.0392 | 23          | 0.0392 |
| 6    | 56          | 0.1502 | 76          | 0.037  | 25          | 0.0366 |
| 7    | 55          | 0.1502 | 77          | 0.0338 | 20          | 0.0353 |
| 8    | 54          | 0.1502 | 82          | 0.0235 | 78          | 0.0353 |
| 9    | 53          | 0.1502 | 60          | 0.0235 | 27          | 0.035  |
| 10   | 52          | 0.1502 | 29          | 0.0235 | 22          | 0.0324 |

**Betweenness centrality** Table 3 shows the betweenness analysis results of the overall network of the HIF-1 signaling pathway.

Table 3
Top 10 nodes of the overall network betweenness degree centrality of the HIF-1 signaling pathway

| Rank | Node Number | Value |
|------|-------------|-------|
| 1    | 20          | 0.0036 |
| 2    | 24          | 0.0034 |
| 3    | 25          | 0.003  |
| 4    | 23          | 0.0026 |
| 5    | 27          | 0.0014 |
| 6    | 6           | 0.0014 |
| 7    | 19          | 0.0014 |
| 8    | 76          | 0.0014 |
| 9    | 17          | 0.0014 |
| 10   | 78          | 0.0013 |
Two nodes appeared in the top 10 list in terms of the betweenness centrality and degree centrality, numbered 20 and 78, which are K07203 and K04371 in KEGG, respectively.

**Cohesive Subgroup Analysis**

First, the distribution of the subset of the overall network is determined. Then, the core network based on the distribution is determined. The minimum component size is 1; i.e., the smallest subnet can be an isolated node without connections. Table 4 shows the analysis results.
Table 4
The maximum connected subnet distribution of the overall network of the HIF-1 signaling pathway

| Subnet Number | Number of nodes included | Cumulative number of nodes | Representative network node number |
|---------------|--------------------------|----------------------------|-----------------------------------|
|               | n                        | %                          | n                                | %                                |
| 1             | 1                        | 1.1765                     | 1                                | 1.1765                           | 1                                |
| 2             | 1                        | 1.1765                     | 2                                | 2.3529                           | 2                                |
| 3             | 1                        | 1.1765                     | 3                                | 3.5294                           | 3                                |
| 4             | 1                        | 1.1765                     | 4                                | 4.7059                           | 4                                |
| 5             | 25                       | 29.4118                    | 29                               | 34.1176                          | 5                                |
| 6             | 34                       | 40                         | 63                               | 74.1176                          | 6                                |
| 7             | 1                        | 1.1765                     | 64                               | 75.2941                          | 7                                |
| 8             | 1                        | 1.1765                     | 65                               | 76.4706                          | 8                                |
| 9             | 3                        | 3.5294                     | 68                               | 80                               | 9                                |
| 10            | 1                        | 1.1765                     | 69                               | 81.1765                          | 13                               |
| 11            | 1                        | 1.1765                     | 70                               | 82.3529                          | 14                               |
| 12            | 1                        | 1.1765                     | 71                               | 83.5294                          | 15                               |
| 13            | 1                        | 1.1765                     | 72                               | 84.7059                          | 16                               |
| 14            | 1                        | 1.1765                     | 73                               | 85.8824                          | 21                               |
| 15            | 1                        | 1.1765                     | 74                               | 87.0588                          | 21                               |
| 16            | 1                        | 1.1765                     | 75                               | 88.2353                          | 30                               |
| 17            | 1                        | 1.1765                     | 76                               | 89.4118                          | 31                               |
| 18            | 1                        | 1.1765                     | 77                               | 90.5882                          | 32                               |
| 19            | 1                        | 1.1765                     | 78                               | 91.7647                          | 33                               |
| 20            | 1                        | 1.1765                     | 79                               | 92.9412                          | 34                               |
| 21            | 2                        | 2.3529                     | 81                               | 95.2941                          | 35                               |
| 22            | 1                        | 1.1765                     | 82                               | 96.4706                          | 37                               |
| 23            | 1                        | 1.1765                     | 83                               | 97.6471                          | 59                               |
| 24            | 1                        | 1.1765                     | 84                               | 98.8235                          | 65                               |
| 25            | 1                        | 1.1765                     | 85                               | 100                              | 81                               |
The data in the table show that the overall network of the HIF-1 signaling pathway includes 25 subnets. The minimum size of the subnet is 1; i.e., there is only one node in the subnet. The maximum size is 34; i.e., 34 nodes are in the subnet.

We colored different subnets differently for the convenience of analysis in this paper. The number of the subnet to which the node belongs is identified. The subnet distribution map of the overall network of the HIF-1 signaling pathway is obtained based on Pajek, as shown in Fig. 2. Figure 2 depicts the subnet distribution of the overall network of the HIF-1 signaling pathway graphically. For example, "[6.00] 20" means that node 20 belongs to the number 6 subnet.

**Bow-tie Structure Analysis**

The Bow-tie analysis module of Pajek software was used to analyze the macrostructure of the core network of the HIF-1 signaling pathway. Table 5 shows the distribution of the structure.

| Structure category                     | Frequency of occurrence | Cumulative frequency | Representative reactants |
|----------------------------------------|-------------------------|----------------------|--------------------------|
|                                        | n  | %    | n   | %    |                                |
| The core section of strong connection  | 1  | 2.9412 | 1   | 2.9412 | ko:K08268                      |
| The input section                      | 20 | 58.8235 | 21  | 61.7647 | ko:K09592                      |
| The tendril-like section               | 13 | 38.2353 | 34  | 100   | ko:K04526 ko:K04357 ko:K05459 |

We colored the different structures in the core network topology of the HIF-1 signaling pathway, as shown in Fig. 3.

**Structural Complexity Analysis**

We analyzed the distribution of network node degrees. That is, the scale of the core network of the HIF-1 signaling pathway is measured. The degree distribution of the core network of the HIF-1 signaling pathway is shown in Table 6.
| Degree | Frequency of occurrence | Cumulative frequency | Representative reactants |
|--------|-------------------------|----------------------|--------------------------|
|        | n | %       | n | %       |                                |
| 1      | 14 | 41.1765 | 14 | 41.1765 | ko:K04526 ko:K04357 ko:K05459 |
| 2      | 12 | 35.2941 | 26 | 76.4706 | ko:K00922 ko:K02649            |
| 3      | 5  | 14.7059 | 31 | 91.1765 | ko:K04527 ko:K04361 ko:K05087 ko:K05083 |
| 4      | 2  | 5.8824  | 33 | 97.0588 | ko:K08268                      |
| 7      | 1  | 2.9412  | 34 | 100     | ko:K09592                      |

Based on the above methods, the power-law fitting results of the core network scale of the HIF-1 signaling pathway in this study are shown in Fig. 4.

Based on Clauset's method, the power-law fitting model for the core network of the HIF-1 signaling pathway can be calculated as follows:

\[ P_r[X \geq x] \sim x^{-3.28} \]

The scale parameter of the power-law distribution of the core network of the HIF-1 signaling pathway is 3.28, and \( x_{\text{min}} = 2 \). That is, the nodes with a degree of 2 or more in the network obey the Pareto distribution rate. The similarity detection index of this fitted model is much larger than 0.1 and closer to 1. This indicates that the fitting effect is good.

**Discussion**

The overall network centrality analysis of the HIF-1 signaling pathway suggests that K07203 appears in the top 10 list in terms of the betweenness centrality, input closeness centrality, and output closeness centrality. It corresponds to the mTOR enzyme, which is known as an essential element in the field of HRN research. These results suggest that the mTOR enzyme has a crucial influence on HIF-1 signaling, which conforms to the known findings, indicating the effectiveness of the network analytical method.

With regard to the component analysis of the overall network of the HIF-1 signaling pathway, the largest connected subnet of the overall network of the HIF-1 signaling pathway is the number 6 subnet; i.e., the number 6 subnet is the core network of the HIF-1 signaling pathway, which is consistent with the findings of current biomedical research.
With respect to the bow-tie structure analysis, the work scope of Ma et al. [11] does not involve the macrostructure analysis of the core network of the HIF-1 signaling pathway. Therefore, this study can be regarded as a supplement to their work and as a verification of their conclusions. The core network of the HIF-1 signaling pathway has a bow-tie structure. It includes macrostructures such as a strongly connected core, an input, and a tendril-like structure. It is worth noting that there are no output components in the structure; that is, HIF-1α has no metabolites. The reason is that the corresponding orthologous enzymes are not listed in the KGML file of KEGG. In KEGG’s hand-painted HIF-1 signaling pathway diagram, HIF-1α has two output pathways: aerobic degradation and hypoxic DNA expression. However, there are no orthologous enzyme products. Therefore, the diagram is consistent with the analysis results obtained by this study. According to the analysis, HIF-1α is the strongly connected core of the bow-tie structure in the core network of the signaling pathway. The bow-tie structure of the core network of the HIF-1 signaling pathway has no formal output components. However, the actual judgment based on a priori knowledge of the degradation pathway of HIF-1α under normoxic conditions suggests that the bow-tie structure of the network contains a potential output structure. Moreover, this structure is HIF-1α itself. In this sense, the bow-tie structure of the core network of the HIF-1 signaling pathway is relatively complete. Additionally, this study showed that the core network of the HIF-1 signaling pathway has a tendril-like structure. This finding further verifies the similarity of the biological metabolic network and the Internet in terms of macroscopic structure, based on the work of Ma et al.

The results of the structural complexity analysis show that the core network of the HIF-1 signaling pathway has scale-free characteristics. Therefore, the HIF-1 signaling pathway network is a relatively typical complex network. It can be studied using complexity-based research methods.

We plan to analyze the network of the HIF-1 signaling pathway from a network perspective since the network analytical method is systematic and holistic. It is efficient in analyzing complex biological networks, and the HIF-1 signaling pathway is an essential component of the HRNs. However, the present study has a few limitations. For example, to verify the effectiveness of the network analytical method, we only analyzed the structural information of the network of the HIF-1 signaling pathway, which is insufficient to explore the unknown critical elements and the relationships of the elements of the pathway. Therefore, further studies need to be carried out to investigate the complex interactions of the elements of the HRNs at different levels using a network analytical method so that new relationships and critical elements of the networks can be identified.

Conclusions

In this study, we investigated the HRN from the perspective of complex networks. In the analysis of the overall network, we identified the topological properties of the overall network of the HIF-1 signaling pathway and the biomedical meaning of the network indicators based on the construction of the network using the data retrieved from the KEGG database. The centrality indicators of the overall network, the remarkable nodes, and the biomedical meaning of these nodes could be identified through centrality analysis. Based on the network indicators, cohesive subgroups of the overall network were obtained; the
different subnets were indicated; and the core HRN was identified. In the analysis of the core network, we refined the core network of the HIF-1 signaling pathway via topology analysis of the network information based on the overall network analysis. We obtained a topological structure diagram and the quantitative indicators of this network. Then, we determined the bow-tie structure existing in this network through the analysis of the bow-tie structure. This structure is composed of three types of components: the core structure, the input structure, and the tendril-like structure. We could verified and further added to the existing research conclusions by analyzing the node distribution of each part. Additionally, a power-law fitting model of this network was constructed through the verification of its complex attributes. The results showed that the power-law fitting model of the core network of the HIF-1 signaling pathway presented a good fitting effect. This proved that the core network of the HIF-1 signaling pathway has scale-free characteristics and is a relatively typical complex network. Although this network analysis is preliminary work, it enriches and improves upon the existing research conclusions. It provides a basis for in-depth research on HRNs at different levels using network analytical methods. In particular, these results can be applied in parallel with traditional studies of the HRN, which will promote and complement each other.

**Abbreviations**

HRN: Hypoxia response network  
HIF-1: Hypoxia-inducible factor-1  
KEGG: Kyoto Encyclopedia of Genes and Genomes  
KO: KEGG Orthology  
KGML: KEGG Markup Language  
XML: Extensible Markup Language

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Availability of data and materials**

The datasets during and/or analyzed during the current study are available from https://www.kegg.jp/kegg-bin/download?entry=hsa04066&format=kgml.

**Consent for publication**

Loading [MathJax]/jax/output/CommonHTML/jax.js
Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Youth Development Program of the Army Medical University under Grant No. 2018XQN08.

Authors’ contributions

JJL wrote the article, and YQG provided the concept of the study. Both researchers investigated the network and screened and analyzed the data. All authors have read and approved the final manuscript.

Acknowledgements

The data were sourced from the Kyoto Encyclopedia of Genes and Genomes [KGML Data file]. Retrieved from http://www.genome.jp

References

1. AG Truesdell, RL Wilson: Training for medical support of mountain operations. Mil Med 2006, 171:463-7.
2. M Qiang, Z Zhi-qing, C Xue-wei, C Bo, C Zhao-li, L Pei-bing, W Lei: Site assessment and analysis of military operation ability of plateau training troops. Military Medical Sciences 2014, 38:668-671.
3. AL Harris: Hypoxia - A key regulatory factor in tumour growth. Nat Rev Cancer 2002, 2:38-47.
4. M Hockel, P Vaipel: Tumor hypoxia: Definitions and current clinical, biologic, and molecular aspects. Jnci-J Natl Cancer I 2001, 93:266-276.
5. KW Kohn, J Riss, O Aprlikova, JN Weinstein, Y Pommier, JC Barrett: Properties of switch-like bioregulatory networks studied by simulation of the hypoxia response control system. Mol Biol Cell 2004, 15:3042-52.
6. B Palsson: Systems biology properties of reconstructed networks. Cambridge; New York: Cambridge University Press; 2006.
7. M Heiner, K Sriram: Structural analysis to determine the core of hypoxia response network. Plos One 2010, 5:e8600.
8. A Barabási, ZN Oltvai: Network biology: understanding the cell’s functional organization. Nat Rev Genet 2004, 5:101-113.
9. DW Ding, YR Ding, LN Li, YJ Cai, WB Xu: Structural and Functional Analysis of Giant Strong Component of Bacillus thuringiensis Metabolic Network. Braz J Microbiol 2009, 40:411-6.
10. G Zhang, W Zhang: Protein–protein interaction network analysis of insecticide resistance molecular mechanism in Drosophila melanogaster. Arch Insect Biochem 2018, 100:e21523.
11. HW Ma, AP Zeng: The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics 2003, 19:1423-1430.

12. S BOCCALETTI, V LATORA, Y MORENO, M CHAVEZ, D HWANG: Complex networks: Structure and dynamics. Physics Reports 2006, 424:175-308.

13. CW Pugh, PJ Ratcliffe: Regulation of angiogenesis by hypoxia: role of the HIF system. Nat Med 2003, 9:677-684.

14. GL Semenza: Regulation of mammalian O-2 homeostasis by hypoxia-inducible factor 1. Annu Rev Cell Dev Bi 1999, 15:551-578.

15. S Wasserman, K Faust: Social Network Analysis Methods and Applications. New York and Cambridge: Cambridge University Press; 1994.

16. W De Nooy, A Mrvar, V Batagelj: Exploratory Social Network Analysis with Pajek, Third Edition edn. Cambridge and New York: Cambridge University Press; 2018.

17. A Broder, R Kumar, F Maghoul, P Raghavan, S Rajagopalan, R Stata, A Tomkins, J Wiener: Graph structure in the Web. Comput Netw 2000, 33:309-320.

18. Q Xuesen, Y Jingyuan, D Ruwei: A new discipline of science — The study of open complex giant system and its methodology. J Syst Eng Electron 1993, 4:2-12.

19. ML Goldstein, SA Morris, GG Yen: Problems with fitting to the power-law distribution. The European Physical Journal B-Condensed Matter and Complex Systems 2004, 41:255–258.

20. A Clauset, CR Shalizi, MEJ Newman: Power-Law Distributions in Empirical Data. Siam Rev 2009, 51:661-703.

21. A Barabási, R Albert, H Jeong: Mean-field theory for scale-free random networks. Physica A: Statistical Mechanics and its Applications 1999, 272:173 - 187.

22. W Ping: Study on Power-law Distribution Law of Website Resources Citation. Journal of Intelligence 2014, 33:105-110.

Figures
Figure 1

The bow-tie structure of the network. This structure is composed of the input section, the core section of strong connection, the output section, the tendril section, the pipeline section, and the component with no connection.

Figure 2

The subnet distribution map of the overall network of the HIF-1 signaling pathway. In the figure, the color of the node and the number in square brackets loading [MathJax]/jax/output/CommonHTML/jax.js
identify the subnet to which the node belongs. The square nodes, round nodes, and elliptical nodes
represent the orthologous enzymes, compounds, and pathways involved in metabolism, respectively. An
arrow represents a metabolic association relationship. The shaft of the arrow connects the metabolic
reactant, and the arrowhead connects the metabolites. The solid black line indicates the "activated"
metabolic behavior, while the solid blue line indicates the "expression" metabolic behavior. Moreover, the
red dashed line indicates "suppressive" metabolic behavior.

![Diagram](Image)

**Figure 3**

The core network bow-tie structure of the HIF-1 signaling pathway. According to the analysis results, the
bow-tie structure of the core network of the HIF-1 signaling pathway includes three types of components.
The first is a strongly connected core, corresponding to the yellow node, ko: K08268, in the figure, and the
corresponding reactant in the KEGG database is HIF-1α. The second is the input components, including
20 green nodes, such as ko: K09592, corresponding to 20 reactants, such as PHD (EGLN, HPH), in the
KEGG database. The third is a tendril-like structure including 13 pink nodes, such as ko: K04526, ko:
K04357, and ko: K05459, which correspond to 13 reactants, such as GF (INS, EGF, IGF1), in the KEGG
database.
Figure 4

Power-law fitting of the core network scale of the HIF-1 signaling pathway. In the figure, the abscissa corresponds to the network scale. The ordinate is the distribution rate of the network scale. The black square points are the actual measured power-law distribution points. The dotted line is the slope of the fitted straight line.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement4.pdf