

NETWORK BIOLOGY: UNDERSTANDING THE CELL'S FUNCTIONAL ORGANIZATION

Albert-László Barabási* & Zoltán N. Oltvai‡

Was sind die grundlegenden architektonischen und evolutionären Prinzipien zellulärer Netzwerke?

Die grundlegenden architektonischen und evolutionären Prinzipien zellulärer Netzwerke lassen sich mithilfe der sich rasch entwickelnden Theorie komplexer Netzwerke quantifizieren und verstehen. Es hat sich herausgestellt, dass die architektonischen Merkmale molekularer Interaktionsnetzwerke in Zellen weitgehend mit denen anderer komplexer Systeme (wie dem Internet oder sozialen Netzwerken) übereinstimmen, was auf ähnliche steuernde Gesetze hindeutet.

Die zelluläre Funktion entsteht aus komplexen Interaktionen zwischen zahlreichen zellulären Bestandteilen wie Proteinen, DNA, RNA und kleinen Molekülen, die zusammen ein komplexes zelluläres Interaktionsnetzwerk bilden – ein „Netzwerk von Netzwerken“.

Grundlegende Architektonische Prinzipien

Die Architektur zellulärer Netzwerke wird durch spezifische topologische Merkmale bestimmt:

1. Skalenfreie Topologie (Scale-Free Topology)

Die wichtigste architektonische Erkenntnis ist, dass die meisten Netzwerke innerhalb der Zelle eine skalenfreie Topologie aufweisen.

- Gradverteilung ($P(k)$): Im Gegensatz zu Zufallsnetzwerken, bei denen die meisten Knoten ungefähr den gleichen Grad (Anzahl der Verbindungen) haben, folgen skalenfreie Netzwerke einer Potenzgesetz-Verteilung ($P(k) \sim k^{-\gamma}$).
- Hubs (Knotenpunkte): Diese ungleichmäßige Verteilung bedeutet, dass die Netzwerke hochgradig ungleichmäßig sind. Die meisten Knoten haben nur wenige Verbindungen, aber eine kleine Anzahl von Knoten weist eine sehr große Anzahl von Verbindungen auf und wird als Hubs bezeichnet. Diese Hubs halten das Netzwerk zusammen.
- Vorkommen: Die skalenfreie Topologie wurde zuerst in Stoffwechselnetzwerken von 43 verschiedenen Organismen nachgewiesen, wobei die meisten Substrate nur an ein oder zwei Reaktionen beteiligt sind, während einige wenige (wie Pyruvat oder Coenzym A) als metabolische Hubs fungieren. Auch Protein-Protein-Interaktionsnetzwerke und genetische Regulationsnetzwerke zeigen diese Merkmale.

2. Hierarchische Modularität (Hierarchical Modularity)

Zelluläre Funktionen werden wahrscheinlich stark modular ausgeführt.

- Module: Module (oder Cluster) sind Gruppen von physikalisch oder funktionell verbundenen Molekülen (Knoten), die zusammenarbeiten, um eine relativ abgrenzbare Funktion zu erreichen.
- Hohe Clusterbildung: Biologische Netzwerke weisen im Allgemeinen eine hohe mittlere Clusterbildung ($\langle C \rangle$) auf, die signifikant höher ist als bei einem Zufallsnetzwerk.
- Hierarchie: Die Skalenfreiheit und die Modularität koexistieren, was auf eine hierarchische Netzwerktopologie hindeutet. Das wichtigste messbare Merkmal der hierarchischen Modularität ist, dass der Clustering-Koeffizient vom Grad des Knotens abhängt und $C(k) \sim k^{-1}$ folgt.
- Funktion der Hubs in der Hierarchie: Knoten mit wenigen Verbindungen haben einen hohen Clustering-Koeffizienten und gehören zu hochgradig vernetzten kleinen Modulen. Die stark vernetzten Hubs hingegen haben einen niedrigen Clustering-Koeffizienten; ihre Rolle besteht darin, verschiedene, ansonsten nicht kommunizierende Module zu verbinden.

3. Motive (Motifs)

Motive sind elementare Einheiten zellulärer Netzwerke.

- Definition: Motive sind Untergraphen (spezifische lokale Interaktionsmuster), die im Vergleich zu einer randomisierten Version desselben Netzwerks überrepräsentiert sind.
- Beispiele: Gerichtete Dreiecksmotive, bekannt als Feed-Forward Loops, treten sowohl in transkriptionsregulatorischen als auch in neuronalen Netzwerken auf.
- Motiv-Cluster: Motive aggregieren typischerweise zu größeren Motiv-Clustern.

4. Kleine-Welt-Eigenschaft und Disassortativität

- Kleine-Welt-Effekt: Ein gemeinsames Merkmal aller komplexen Netzwerke ist, dass jeder beliebige Knoten über einen Pfad von nur wenigen Verbindungen mit jedem anderen Knoten verbunden werden kann.
- Ultra-Small-World: Skalenfreie Netzwerke sind ultra-klein ($\langle l \rangle \sim \log \log N$), was bedeutet, dass ihre Pfadlänge viel kürzer ist, als es der Kleine-Welt-Effekt vorhersagt. Im Metabolismus können die meisten Paare von Metaboliten durch Pfade von nur drei bis vier Reaktionen verbunden werden.
- Disassortativität: Zelluläre Netzwerke sind disassortativ; das heißt, Hubs (stark vernetzte Knoten) neigen dazu, Verbindungen zu Knoten mit nur wenigen Interaktionen einzugehen und vermeiden es, sich direkt miteinander zu verbinden.

5. Robuste Topologie

Die skalenfreie Architektur hat direkte Auswirkungen auf die Robustheit des Systems.

- Robustheit gegen zufällige Fehler: Skalenfreie Netzwerke sind erstaunlich widerstandsfähig gegen zufällige Komponentenausfälle, da zufällige Fehler hauptsächlich

die zahlreichen Knoten mit geringem Grad betreffen, deren Fehlen die Netzwerk-Integrität nicht stört.

- Anfälligkeit für Angriffe: Diese Abhängigkeit von Hubs führt jedoch zu einer Anfälligkeit für Angriffe (attack vulnerability) – die Entfernung nur weniger Schlüssel-Hubs kann das System in kleine, isolierte Cluster zerteilen.
- Funktionelle Bedeutung von Hubs: Die Wichtigkeit eines Proteins für die Viabilität oder das Wachstum der Zelle hängt stark von seinem Grad an Konnektivität (Hub-Status) ab.

Grundlegende Evolutionäre Prinzipien

Die skalenfreie Architektur ist das Ergebnis grundlegender evolutionärer Prozesse.

1. Wachstum (Growth)

Die meisten Netzwerke entstehen durch einen Wachstumsprozess, bei dem über einen längeren Zeitraum hinweg neue Knoten in das System eintreten.

2. Präferenzielle Anbindung (Preferential Attachment)

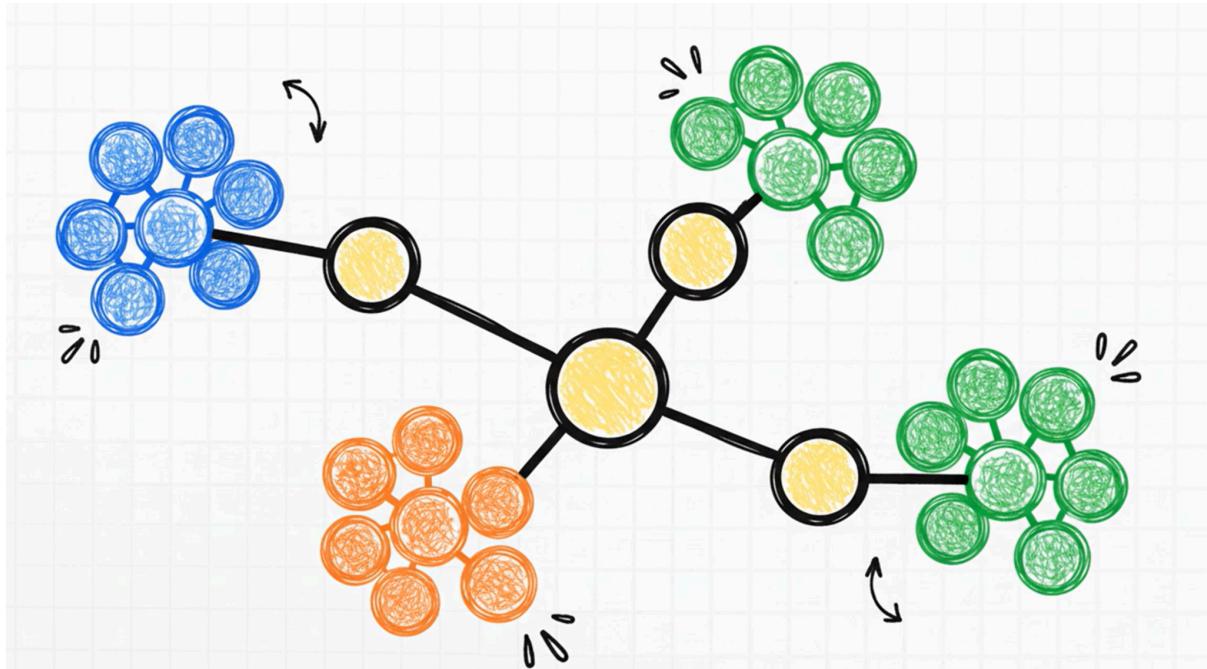
Neue Knoten neigen dazu, sich mit Knoten zu verbinden, die bereits viele Verbindungen besitzen. Dieser Mechanismus führt zu einem „Rich-gets-richer“-Phänomen, bei dem bereits stark vernetzte Knoten schneller neue Verbindungen gewinnen und sich so in Hubs verwandeln.

3. Gen-Duplikation als Ursprung der Präferenziellen Anbindung

In Proteinnetzwerken haben Wachstum und präferenzielle Anbindung ihren Ursprung wahrscheinlich in der Gen-Duplikation.

- Mechanismus: Dupliizierte Gene produzieren identische Proteine, die mit denselben Partnern interagieren. Dadurch gewinnt jedes Protein, das mit dem duplizierten Protein in Kontakt steht, eine zusätzliche Verbindung.
- Vorteil für Hubs: Hochgradig vernetzte Proteine haben einen natürlichen Vorteil: Sie haben eine höhere Wahrscheinlichkeit, mit einem duplizierten Protein verbunden zu sein als schwächer vernetzte Proteine, und gewinnen daher schneller neue Links. Dies stellt eine subtile Form der präferenziellen Anbindung dar.
- Beweise: Die Inspektion metabolischer Hubs zeigt, dass evolutionär ältere Substrate (wie Coenzym A, NAD, GTP, und Elemente alter Stoffwechselwege wie die Glykolyse) zu den am stärksten verbundenen Substraten gehören. Darüber hinaus haben evolutionär ältere Proteine im Durchschnitt mehr Verbindungen zu anderen Proteinen als jüngere Gegenstücke, was die präferenzielle Anbindung empirisch belegt.

Die Hierarchie der Module ist ebenfalls evolutionär geprägt. Die Akkumulation lokaler Veränderungen, die die kleinen, hochgradig integrierten Module betreffen, wirkt sich langsam auf die größeren, weniger integrierten Module aus. Dies deutet darauf hin, dass Evolution und natürliche Selektion existierende Module wiederverwenden, um die Überlebenswahrscheinlichkeit und die Komplexität des Organismus zu erhöhen.



Protein-protein interaction networks (PPI) and complex diseases

Nahid Safari-Alighiarloo¹, Mohammad Taghizadeh², Mostafa Rezaei-Tavirani

Generally, traditional analyses exploit a univariate approach to study gene expression and identify genes with meaningful individual differential expression in the phenotype of interest. However, the underlying mechanisms of complex diseases, which arise from the interplay among multiple genetic and environmental factors, cannot be explicated by such univariate approaches.

There are three main progressions in graph theory at 20th century, random graph theory, small world networks and scale free networks.

Small-world networks: shortcuts lead to shorter than expected path length, while maintaining high clustering of neighbors

Scale-free networks were first formally introduced by Albert and Barabasi (3). Their main feature is that the degree distribution (i.e. the number of connections or edges a node possesses) follows the power-law rule in which the vast majority of nodes have a low degree, while a smaller than expected numbers of nodes, known as hubs, have a very high degree of connectivity. It was later suggested that protein–protein interaction networks obey such power-law distribution (18).

There are four higher-level topological indices including average degree (K), clustering coefficient(C), average path length (L), and diameter (D). It is possible to calculate four topological distributions such as degree distribution $P(k)$, degree distribution of cluster coefficients $C(k)$, shortest path distribution $SP(i)$, and topological coefficient distribution $TC(k)$, which take more attentions (29–31) and are comprehensively used in cellular networks, such as PPI networks

Table 1. Basic definitions in PPI network

Term	Definition
Node or (Vertices)	Each protein in network
Edge or (link)	Physical or functional interactions between proteins
Hub	Each “high-degree” node of a network
Modules	Group of sub networks in which each sub network includes a high number of inside-sub network links and a low number of between- sub network links
Degree (k)	The number of adjacent links
Average degree ($\langle k \rangle$)	The mean of all degree values of nodes in a network.
Clustering Coefficient ($\langle C \rangle$)	The proportion of links between the nodes within the i -neighborhood divided by the number of links that could possibly exist between them
Shortest Path Length	The steps (number of links) needed to connect every pair of nodes through their shortest path.
Diameter	The minimum number of links that separate the two most distant nodes in a network.
Betweenness centrality	Measures how often nodes occur on the shortest paths between other nodes
Heterogeneity of a network	the coefficient of variation of the degree distribution