Long loops of information flow in genetic networks highlight an inherent directionality

Roi Itzhack, Lea Tsaban and Yoram Louzoun*

Department of Mathematics; Bar-Ilan University; Ramat Gan, Israel; Gonda Brain Research Center; Bar-Ilan University; Ramat Gan, Israel

Keywords: genetic regulatory networks, circles, graphs, directionality, flow

Introduction

Biological networks, such as those describing gene regulation, signal transduction and neural synapses, are the representations of large-scale dynamic systems. Organizing principles of biological networks may be deduced from the connection between networks structure and the flow of information in the system they represent. Indeed, the combination of a large number of interactions into networks was used to reveal a macroscopic organization in multiple domains. We here propose that the cycle length distribution in these networks highlights a large scale information flow. Specifically, one can treat networks as an assembly of randomly connected small functional modules represented by short cycles, or as large scale regulatory feedback loops that integrate the network into a collective response. We show here that the latter is probably correct.

Network development and growth have attracted much interest in the last two decades. Multiple network growth models have been proposed to explain the large scale structure of networks.1-28 These include, among many others, the duplication-mutation models that suggest that network growth occurs through the duplication of existing nodes and mutations of edges.8 Other models include random static network models, where edges are randomly connected25 and small-world network models based on an interpolation between regular ring lattices and randomly connected graphs.7

However, these proposed models do not explain the large scale organization of real world networks. Specifically, they do not address the emergence of large scale structures beyond local connectivity, and the global flow of information along the network. The concept of global information can relate to multiple aspects including, among others, a possible hierarchy between nodes, where some nodes are more important than others. A second approach is to study the flow of information in the network and to analyze the feedback loops therein. A third approach could be the detection of a core and periphery as is performed in the K-core analysis.29 Order in the sense of hierarchy was recently studied by Muchnick et al. that suggested a methodology to extract a hierarchical model from a directed network.30 This methodology highlights the presence of “importance” levels in the network from nodes in high hierarchical positions to lower ones. Other methods include closeness centrality,31 defined as the average distance (geodesic path length) to all reachable vertices. In this measure, it is assumed that important vertices reach other vertices through short paths. The PageRank algorithm32 is a form of eigenvector centrality33 where the score of each vertex depends on

*Correspondence to: Yoram Louzoun; Email: ylouzoun@gmail.com
Submitted: 10/01/12; Revised: 11/08/12; Accepted: 02/07/13
http://dx.doi.org/10.4161/sysb.24471
the PageRank score of the vertices pointing to it. The PageRank score vector is a stationary probability distribution expressing the chance of reaching a vertex by a random walk. Finally, in a spectral network drawing algorithm, the score vector is a unique minimizer of an energy that is supposed to order vertices in a hierarchy, and define a general directionality in the network.

Here, we study the general organization of networks through the information flow in the network and attempt to detect a non-random large scale information flow and the structure of this information flow. Existing hierarchy based methods fail to detect the large scale regulatory structure of the network, which can be mediated by long feedback loops. We here propose a novel method to show that genetic regulatory networks contain a large scale flow of information among distant parts of the network. This flow is determined by very long feedback loops. We propose a test to detect such a flow, and show that such a flow exists in multiple genetic networks.

In order to approximate the number of feedback loops, we formally define a “loop” in a graph as a cycle (close walk) composed of back and forth minimal paths between a pair of nodes that contains no shorter cycle, and measure the number of such loops. Note that a loop in graph theory is sometime defined as a self edge. We do not use this notation here. A precise definition of that is given in the method section. We find that the number of such loops are highly sensitive to the flipping of even a small fraction of the edges directions. Such extreme behavior is not observed in any of the other features of genetic networks or in loops of random networks. In the following sections, we describe a novel algorithm to measure the loop distribution in networks, and show that the long loops emerging in intra-cellular networks are composed of essential genes.

Results

An important question in the analysis of genetic regulatory networks is whether they should be treated as a randomly connected set of interactions forming a “meaningless” large scale structure, where the paths do not have a clear direction, or as a large scale well organized structure with a clear, yet hidden directionality and long-term flow. In order to test for such a flow, we compared the effect of randomizing the direction of edges (i.e., replacing an edge from a to b by an edge chosen randomly to be either a to b or b to a), and tested the effect of this randomization. If a large scale flow exists, then the direction of each edge is important. If there is no large scale structure, there will be no significant difference in the statistical properties of the network between the real network and a randomized one. We here test whether clear directional information flow and large scale structures are indeed observed in genetic regulatory networks.

To directly test the effect of the network directionality, we need a baseline that is not affected by other elements. Often, fully randomized networks are used for a comparison (Erdos Renyi networks25). These networks have a Poisson degree distribution, and short average (geodesic) distances between nodes.55 In other cases, randomized networks that maintain the degree and/or clustering coefficient were used as a baseline comparison. In the present case, we are only interested in the directionality of the edges; we have therefore used a normalization based on edge inversion that does not affect the undirected network induced by ignoring the direction of edges.

We have tested the effect of randomizing the direction of the edges in three genetic regulatory networks: E. coli, yeast, and the mouse. Inverting the edges had no effect on the degree distribution and the clustering coefficient, as is indeed expected from a randomization that does not change the undirected network induced by the directed one (data not shown). The in-degree and out-degree were also very weakly affected by randomly switching the direction of edges.

A more appropriate method to detect flow is probably related to the presence or absence of feedback loops. Such loops must be presented as loops in the graphs. Loops have been indeed proposed to be related to information flow in biological networks,36,37 We have thus examined the length distribution of loops, as shall be further discussed. As mentioned above, we define here a loop as a cycle in a graph composed of a minimal path from one node to the other and the minimal path back to the original node.

We found that the loop length distribution is indeed highly sensitive to very small random changes in the direction of edges (Fig. 1). Regulatory genetic networks have typically more long loops than expected in random direction networks, and less short loops than random direction networks. However, as soon as the direction of a small fraction of the edges is changed, the loop distribution is drastically changed. Switching the direction of 2.5% of the edges decreases the number of long loops (above 10 nodes) by more than an order of magnitude, and increases the number of short loops (3–5 nodes) by more than an order of magnitude (Fig. 1).

When the fraction of swapped edges was increased from 0% to 50%, a clear effect was observed after changing the direction of only 2.5% of the edges. However, both the increase in short loop frequency and the decrease in long loop frequency do not saturate until the edge directions are fully random (Fig. 1).

Note that in the yeast network, the distribution is closer to the random distribution than in the E. coli or the Mouse. A simple reason for that could be the presence of erroneous edges in the yeast network. Indeed adding random edges to the E. coli networks makes it more similar to the random network and to the Yeast network (Fig. S1).

The decrease in the long loop frequency may in principle be an artifact of the increase in the frequency of short loops. Since adding short loops can create shortcuts within long loops. If such shortcuts exist, we would not count the long loops. In order to check that this is not the case, we performed two validations. First, we flipped 20% of the edges, and obtained a large number of short loops and much fewer long loops. Then, we removed the majority (90%) of 3-loops (loops of length 3), and computed the loop distribution. Even when practically all 3-loops were removed, there was practically no increase in the number of long loops, and no loop with length more than 10 emerged, in contrast with the original network (Fig. S2). Another test performed was simply to remove 3-loops from the yeast network, where there were initially few long loops, and to check whether long loops
sense the opposite of short directed 3-loops (A points to B, which points to C, which points back to A) (Fig. 3). Indeed, the numbers of short FFL are much higher in genetic networks than in their random direction counterpart, and as edges are flipped to a random direction, their number decreases to the expected base-line. Actually, a clear parallel between the increase in directed 3-loops and the decrease in FFL can be observed in all studied networks (Fig. 4), where their sum add to a constant, but as the edges are flipped, directionality is lost, and the number of directed 3-loops increases, while the number of FFL decreases. This raises the interesting possibility that the absence of short loops in genetic regulatory networks is derived from an inherent directionality of these networks. If an order exists between genes, and edges are mainly between genes high in the order to genes low in the order, one would expect directed loops to be very rare, and feed forward loops to be very frequent.

To test this explanation, we developed a simple toy network model, where a random Erdos-Renyi network was built and each node was assigned a random value. We then switched all edges to be from high to low values (i.e., if an edge existed between A and B, and B had a higher score than A, we replaced it by an edge between B and A). This leads to a directed acyclic graph (DAG) that can have FFL, but no directed 3-loops.

When the direction of edges in the toy model is randomly switched, the number of FFL decreases and the number of directed 3-loops increases, precisely as in the other graphs.

Another possibility could have been that the change in the loop length distribution is a simple mirror of the distribution of the path length. It is possible that the long loops only represent longer (geodesic) paths in the real network compared with randomized networks. Note that a loop is computed through a combination of two paths (from i to j and back to i). Thus, increasing the path length leads to an increase in the loop length. However, if the effect of flipping edges on a loop was mediated by the path length, then the observed non-random long loop distribution would follow the paths length distribution. However, this is not the case in the networks studied here. When the flipping ratio is increased, the increase in the number of node pairs with a short distance is much slower than the increase in the number of short loops. Similarly, the decline in the number of long loops is much faster than the decline in the number of long paths (Fig. 2). Thus the path length distribution is not the reason for the extreme sensitivity of the loop length distribution to the direction of a small number of edges.

The low number of short loops is paralleled by a high number of feed-forward loops (FFL). These loops are composed of three nodes, where A points to B and C, and B points to C (Fig. 3). Note that FFL are not formally loops using the definition above, since they do not close a directed cycle in a graph. We here use the term FFL to follow the standard terminology. FFL are in some sense the opposite of short directed 3-loops (A points to B, which points to C, which points back to A) (Fig. 3). Indeed, the numbers of short FFL are much higher in genetic networks than in their random direction counterpart, and as edges are flipped to a random direction, their number decreases to the expected baseline. Actually, a clear parallel between the increase in directed 3-loops and the decrease in FFL can be observed in all studied networks (Fig. 4), where their sum add to a constant, but as the edges are flipped, directionality is lost, and the number of directed 3-loops increases, while the number of FFL decreases.

This raises the interesting possibility that the absence of short loops in genetic regulatory networks is derived from an inherent directionality of these networks. If an order exists between genes, and edges are mainly between genes high in the order to genes low in the order, one would expect directed loops to be very rare, and feed forward loops to be very frequent.

To test this explanation, we developed a simple toy network model, where a random Erdos-Renyi network was built and each node was assigned a random value. We then switched all edges to be from high to low values (i.e., if an edge existed between A and B, and B had a higher score than A, we replaced it by an edge between B and A). This leads to a directed acyclic graph (DAG) that can have FFL, but no directed 3-loops.

When the direction of edges in the toy model is randomly switched, the number of FFL decreases and the number of directed 3-loops increases, precisely as in the other graphs.
and removed all directed 3-loops. This does indeed very slightly increase the number of 6 and 7 loops, but it has no effect of all the other attributes of the network. Thus, such a mechanism does not seem plausible (Fig. S4).

Assuming a directionality is the source of the absence of short loops, we have tested whether this directionality could be the source of the very long loops observed in most genetic regulatory networks. An inherent directionality could in theory increase the number of long feedback loops, since we only analyze loops that do not contain a shorter loop within. Thus, by reducing the number of short loops, the number of long loops may be in principle increased. However, the toy model does not produce the long loops observed in real networks (Fig. 5). Moreover, the increase in the number of short loops and the decrease in the FFL occur quite slowly and require the flipping of a large number of edges. The decrease in the number of long feedback loops occurs in a very drastic way following the inversion of a very small fraction of the edges, which cannot be explained by the general directionality of the networks.

Thus, two different mechanisms seem to be operating in genetic networks: a local order mechanism preventing short feedback loops and enhancing FFL and a more global mechanism of long range flow in the network. While the order in the network is non-specific, the selection for the presence of very large loops must occur in specific genes. We have thus checked whether the genes participating in a large number of large loops are essential to the organism survival. We used the full mapping of the Yeast...
genetic interactions as mapped by Tong et al., which provided a score for synthetic lethal interactions per gene. Indeed, the fraction of genes with mainly long loops (at least half the loops they participate in have a length of at least 6) that is essential (18%) is much larger than the fraction of essential genes in the total population (2%) (chi square test p < 1.e-3). Thus, the genes that participate mainly in long loops are indeed an essential part of the genetic regulation circuitry of the organism. One could assume that the essentiality of these genes is induced by their high-degree or a high betweenness centrality. However, while the total number of loops in which each node participates is highly correlated with both the degree and the betweenness centrality, the fraction of long loops has a negative (yet non-significant) correlation with the degree as well as with the centrality. Thus, the observed correlation between essentiality and long loops is not an artifact of the degree.

Discussion

Genetic networks are used to represent the aggregated information of multiple genetic regulation mechanisms. Beyond their obvious physical parameters, such as the number of genes and genetic interactions, these networks have mathematical properties that can reveal larger organizational principles. Such measures include among others the degree distribution, clustering coefficient, centrality and the path length distributions. Some of these measures highlight the local organization of the network, while others are related to macroscopic properties of the network. However, these measures fail to detect the long range information flow in such networks. We have used here a novel method to understand this large scale flow.

We have checked the effect of the direction of edges on genetic networks properties. We found an extreme sensitivity of the loop numbers in the network to the edge direction, where we have defined a loop as a cycle containing a given pair of nodes, with no shorter cycle containing these nodes. The main difference between the original networks and networks with inverted edges is the low number of short directed loops and high number of long loops in the real networks. Inverting even 2–3% of the network edges can reduce the number of long loops by an order of magnitude and lead to a similar increase in the short loop number.

Two possible mechanisms can explain the absence of short loops: a genetic process limiting useless, very short information loops, or an inherent order in the network. Indeed, ordering the nodes according to an arbitrary order and having most of the edges follow this order does lead to a limited number of short loops. The same can happen if we simply choose three-node loops and invert one of the edges. Note that the same mechanism explains the frequency of feed-forward loops. Thus, the absence of short loops and the presence of FFL may simply be two aspects

![Figure 4. 3-loop direction comparison. Decline in the feed forward loop frequency (full dark line) compared with the rise of the directed 3-loop frequency (dot dashed line) as a function of the flipping ratios for three genetic networks and a toy model. For all networks the total number of 3-loops (triangles in graph theory terms) is maintained. Thus, the total number of 3-loops is not affected by the flipping ratio (central line). The expected number of FFL and directed 3-loops in the absence of a predefined order is simply the value obtained for random flipping (flipping ratio of 0.5).](image)
An important aspect not discussed in the current analysis is the coherence of these long loops. There may be very significant differences between positive and negative feedback loops. While very long positive feedback loops may be inefficient, very long negative feedback loops may be crucial to ensure the appropriate function of the cell. We plan to further analyze this relation in multiple networks.

**Materials and Methods**

**Definitions**

Through this analysis, we use multiple concepts from graph theory. Following are some definitions of these concepts:

- **Distance** is the geodesic distance between a pair of nodes. The geodesic distance between a pair of nodes is the length of the shortest path between them.

- **Degree**. The degree of a node is the number of neighbors it has in the network.

- A loop between nodes $i$ and $j$ is a cycle in the graph that contains $i$ and $j$, with no shorter cycles containing both $i$ and $j$. Note that the graph may contain shorter cycles containing either node $i$ or node $j$. In some times, in graph theory, a loop is a self-edge. This is not the notation used here.

- **Feed forward loop (FFL)** is a combination of three nodes ($A$, $B$, and $C$) with the following edges ($A \to B$, $B \to C$, and $A \to C$).

- A directed 3-loop is a combination of three nodes ($A$, $B$, and $C$) with the following edges ($A \to B$, $B \to C$, and $C \to A$).

- **Betweenness centrality** of a node $i$ is the number of geodesic paths passing through it.

**Degree distribution**

The degree $k(i)$ of a node is defined as the number of neighbors of the node $i$. It can be divided into the in-degree (degree of

---

**Figure 5.** Toy network. Loop length distribution in the toy model. One can clearly see that while there are some long loops in a toy model with an almost perfect order (low the flipping ratio), their number is much lower than in realistic networks.
incoming edges), out-degree (degree of outgoing edges) and total degree. \( P(k) \) is the probability that a randomly selected node has a degree of \( k \).

**Geodesic distance distribution**

The shortest path or geodesic distance \( d(i,j) \) in a directed graph is defined as the minimal number of edges from node \( i \) to \( j \). Note that the shortest path \( d(i,j) \) from \( i \) to \( j \) is usually not equal to the shortest path \( d(j,i) \) between \( j \) and \( i \). In fact, both are longer or equal to the undirected shortest path, since the undirected path often violates the correct edges direction. The average shortest path is defined as the average of the distance matrix that represents the length of the shortest paths between all connected nodes.

**Clustering coefficient**

The clustering coefficient \( C.C(i) \) of a node \( i \) is the fraction of its neighbors that are connected one to each other. Given the set of \( k(i) \) neighbors of a node \( i \), one could in principle have \( k(i) \) \((k(i)-1)/2\) undirected edges between them. The number of connected neighbors divided by \( k(i)(k(i)-1)/2 \) is the C.C. For a large, completely random graph the average C.C is \( < k > /N \), while for a clique (group of nodes which are completely connected to each other) the C.C is 1.

**Loop enumeration**

The loop enumeration software is available in an online tool at the following: http://peptibase.cs.biu.ac.il/graph_analysis/

A directed loop is a closed directed path containing no repetitions (i.e., none of the nodes in the loop appears twice). We define a minimal loop as a loop composed of a minimal path from one node to the other and the minimal path back to the original node.

We have developed a rapid algorithm to enumerate all such loops. The naïve algorithm would find the minimal geodesic path between all node pairs in a directed graph, \( G(V,E) \), using a BFS (Breath First Search) algorithm, and check for each source node \( v_{\text{source}} \) whether nodes at the end of the path have a direct edge to the source node \( v_{\text{source}} \). This algorithm would however count each loop multiple times (Fig. S1). The only way to count each loop once would be to maintain a list of all loops, and check for those only present once in the list. This naïve algorithm is extremely costly, and becomes impractical even for networks of a few hundred nodes.

We developed a practical algorithm that counts each node only once. If a loop of size \( r \) is counted \( j \) times in the naïve algorithm, the loop counter of size \( r \) increases only by \( 1/j \), each time this loop is measured. \( j \) is the number of nodes in the loop, with a directed distance of \( r \) to the node pointing to them in the loop. In other words, if in a loop of length 8 (for example), the distance from a node \( v_i \) to its precedent is less than 8, it implies that there is a shortcut in this loop, and one will not count the loops when starting from \( v_i \) (Fig. S5). A loop of size \( r \) would be counted exactly \( r \) times only if the direct distance from each node to its precedent is \( r-1 \). We thus do not use a data structure but only count for each loop how many nodes have a distance of \( r-1 \) to their precedent.

The total cost of the algorithm is \( O(V^2E^2L) \) where \( L \) denotes the average geodesic path length. The total memory cost is \( O(V^4E^2) \) (in the worst case), where \( V \) denotes the number of all the shortest paths between pairs of nodes, and \( L \) denotes the average geodesic path length.

**Translation of networks to a graph**

The genetic networks studied here were obtained from multiple sources. The yeast network was obtained from the Biological General Repository for Interaction Datasets (BioGRID; http://bidata.mshri.on.ca/grid). We considered only genetic interactions, such as Dosage Growth Defect, Dosage Lethality, Phenotypic Enhancement etc., each TF was considered as a node and each genetic interaction was considered as an edge with equal weight. The E. coli genetic network was obtained from the regulodb database. The mouse gene regulatory network was created according to the Ahn et al. analysis. We used an upper threshold of 0.4, and obtained a network with 20,143 genes and 128,551 connections.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Supplemental Materials**

Supplemental materials may be found here: www.landesbioscience.com/journals/systemsbiomedicine/article/24471

**References**

1. Amaral LA, Scala A, Barthelemy M, Stanley HE. Classes of small-world networks. Proc Natl Acad Sci U S A 2000; 97:11149-52; PMID:11005838; http://dx.doi.org/10.1073/pnas.200327197
2. Barabási AL, Oltvai ZN. Network biology: understanding the cell’s functional organization. Nat Rev Genet 2004; 5:101-13; PMID:14735121; http://dx.doi.org/10.1038/nrg1272
3. Blank A, Solomon S. Power laws in cities population, financial markets and internet sites (scaling the cell’s functional organization. Nat Rev Genet 2004; 5:101-13; PMID:14735121; http://dx.doi.org/10.1038/nrg1272
4. Blank A, Solomon S. Power laws in cities population, financial markets and internet sites (scaling the cell’s functional organization. Nat Rev Genet 2004; 5:101-13; PMID:14735121; http://dx.doi.org/10.1038/nrg1272
5. Michalis F, Ferros F, Christos F. On power-law relationships of the Internet topology. SIGCOMM Comput Commun Rev 1999; 29:251-62; http://dx.doi.org/10.1145/316194.316229
6. Noh JD. Exact scaling properties of a hierarchical network model. Phys Rev E Stat Nonlin Soft Matter Phys 2003; 67:045103; PMID:12786419; http://dx.doi.org/10.1103/PhysRevE.67.045103
7. Ravasz E, Barabási AL. Hierarchical organization in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 2003; 67:026112; PMID:12636753; http://dx.doi.org/10.1103/PhysRevE.67.026112
8. Vázquez A, Flammini A, Maritan A, Vespignani A. Modeling of Protein Interaction Networks. Complexus 2003; 1:38-44; http://dx.doi.org/10.1159/000067642
9. Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. Nature 1998; 393:440-2; http://dx.doi.org/10.1038/30918
10. Willinger W, Govindan R, Jamin S, Paxson V, Shenker S. Scaling phenomena in the Internet: critically examining criticality. Proc Natl Acad Sci U S A 2002; 99(Suppl 1):2573-80; PMID:11875212; http://dx.doi.org/10.1073/pnas.012583099
11. Adams LA, Huberman BA. Power-Law Distribution of the World Wide Web. Science 2000; 287:2115; http://dx.doi.org/10.1126/science.287.5461.2115a
12. Agranovich A, Louzoun Y, Shnerb N, Moolan S. Catalyst-induced growth with limited catalyst lifespan and competition. J Theor Biol 2006; 241:307-20; PMID:16412475; http://dx.doi.org/10.1016/j.jtbi.2005.11.031
13. Agrawal H. Extreme self-organization in networks constructed from gene expression data. Phys Rev Lett 2002; 89:268702; PMID:12484863; http://dx.doi.org/10.1103/PhysRevLett.89.268702
14. Amaral LA, Scala A, Barthelemy M, Stanley HE. Classes of small-world networks. Proc Natl Acad Sci U S A 2000; 97:11149-52; PMID:11005838; http://dx.doi.org/10.1073/pnas.200327197
15. Barabási AL, Albert R. Emergence of scaling in random networks. Science 1999; 286:509-12; PMID:10521342; http://dx.doi.org/10.1126/science.286.5459.509
16. Barabási AL, Bonabeau E. Scale-free networks. Sci Am 2003; 288:60-9; PMID:12701331; http://dx.doi.org/10.1038/scientificamerican0903-60
17. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic networks. Nature 2000; 407:651-4; PMID:11054217; http://dx.doi.org/10.1038/35066277
18. Barabási AL, Bonabeau E. Scale-free networks. Science 2001; 63:062101; PMID:11415146; http://dx.doi.org/10.1126/science.107374
19. Bollobás B, Riordan O. The Diameter of a Scale-Free Random Graph. Combinatorica 2004; 24:5-34; http://dx.doi.org/10.1007/s00493-004-0002-2
20. Bornholdt S, Ebel H. World Wide Web scaling exponent from Simon's 1955 model. Phys Rev E Stat Nonlin Soft Matter Phys 2001; 63:056125; PMID:11456732; http://dx.doi.org/10.1103/PhysRevE.63.056125
21. Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, et al. The Graph structure in the Web. Computer Networks: The International Journal on Computer and Telecommunications Networking 2000; 33:309-20; http://dx.doi.org/10.1016/S1389-1286(00)00083-9
22. Dorogovtsev SN, Mendes JFF. Evolution of networks. Adv Phys 2002; 51:1079-187; http://dx.doi.org/10.1080/00018730110112519
23. Dorogovtsev SN, Mendes JF, Samukhin AN. Size-dependent degree distribution of a scale-free growing network. Phys Rev E Stat Nonlin Soft Matter Phys 2001; 63:062101; PMID:11415146; http://dx.doi.org/10.1103/PhysRevE.63.062101
24. Dorogovtsev SN, Mendes JFF. Evolution of networks. Adv Phys 2002; 51:1079-187; http://dx.doi.org/10.1080/00018730110112519
25. Erdos P, Renyi A. On the evolution of random graph. Institute of Mathematics Hungarian Academy of Sciences, 1959
26. Fermi I Cancho R, Solé RV. The small world of human language. Proc Biol Sci 2001; 268:2261-5; PMID:11674874; http://dx.doi.org/10.1098/rspb.2001.1800
27. Newman MEJ. Scientific collaboration networks. I. Network construction and fundamental results. Phys Rev E Stat Nonlin Soft Matter Phys 2001; 64:016131; PMID:11461355; http://dx.doi.org/10.1103/PhysRevE.64.016131
28. Newman MEJ. The structure of scientific collaboration networks. Proc Natl Acad Sci U S A 2001; 98:404-9; PMID:11149952; http://dx.doi.org/10.1073/pnas.98.2.404
29. Newman MEJ. The mathematics of networks. The New Palgrave Encyclopedia of Economics 2. 2008
30. Newman MEJ. Scientific collaboration networks. I. Network construction and fundamental results. Phys Rev E Stat Nonlin Soft Matter Phys 2001; 64:016131; PMID:11461355; http://dx.doi.org/10.1103/PhysRevE.64.016131
31. Freeman LC. Centrality in social networks conceptual clarification. Soc Networks 1979; 1:215-39; http://dx.doi.org/10.1016/0378-8733(78)90021-7
32. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the web. 1999
33. Newman MEJ. The structure and function of complex networks. SIAM 2003; 61:567; http://dx.doi.org/10.1209/epl/i2003-10006-9
34. Carmel L, Harel D, Koren Y. Combining hierarchy and energy for drawing directed graphs. IEEE Trans Vis Comput Graph 2004; 10:46-57; PMID:15382697; http://dx.doi.org/10.1109/TVCG.2004.1260757
35. Newman MEJ. The Structure and Function of Complex Networks. SIAM 2003; 45:167-256; http://dx.doi.org/10.1137/0005614450342480
36. Inzhak R, Louzoun Y. Random distance-dependent attachment as a model for neural network generation in the Caenorhabditis elegans. Bioinformatics 2010; 26:647-52; PMID:20081220; http://dx.doi.org/10.1093/bioinformatics/btp015
37. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. Science 2002; 297:3551-5; PMID:12208230; http://dx.doi.org/10.1126/science.107374
38. Mangan S, Alon U. Structure and function of the feedforward loop network motif. Proc Natl Acad Sci U S A 2003; 100:11980-5; PMID:14530388; http://dx.doi.org/10.1073/pnas.2133841100
39. Mangan S, Alon U. Structure and function of the feedforward loop network motif. Proc Natl Acad Sci U S A 2003; 100:11980-5; PMID:14530388; http://dx.doi.org/10.1073/pnas.2133841100
40. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. Science 2004; 303:808-13; PMID:14764870; http://dx.doi.org/10.1126/science.1091317
41. Tuite WT. Graph Theory. Cambridge University Press; 2004
42. Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavala A, Muñiz-Rascado L, Solano-Lira H, et al. RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). Nucleic Acids Res 2011; 39(Database issue):D98-105; PMID:21051347; http://dx.doi.org/10.1093/nar/gkq1110
43. Ahn S, Wang RT, Park CC, Lin A, Leahy RM, Lange K, et al. Directed mammalian gene regulatory networks using expression and comparative genomic hybridization microarray data from radiation hybrids. PLoS Comput Biol 2009; 5:e1000407; PMID:19521529; http://dx.doi.org/10.1371/journal.pcbi.1000407