Information networks: concept, classification and application

Nazif Salihović, Adisa Hasković Džubur

University of Sarajevo, Zmaja od Bosne 8, Sarajevo 71000, Bosnia and Herzegovina

Abstract

An information network is a structure used for transmitting various forms and types of information. In basic structure, it consists of branches that connect certain nodes. Many scientists and researchers have dealt with the problem of defining an information network, depending on the functional organization and data transmission, and also with the classification of information networks. Information networks have a very wide application in almost all scientific disciplines. A large number of researches are carried out on the application of information networks (e.g. bisociative, deep information network, heterogeneous information network, and space information network) in the field of medicine for easier detection of diseases, drug development, etc., and other needs to support real-time communication, massive data transmission, and data processing. In accordance with the above mentioned, the aim of this paper is to offer different approaches in defining and classifying general forms of information networks and to notice their wide application in different research disciplines.

Keywords: *information network, bisociative information networks, heterogeneous information networks*

1 Introduction

A mathematical definition of an information network was constructed for the purpose of developing a theory used to answer questions about the transmission of information. Information network includes: users, information resources, information centers, and the overall structure of information transfer [1]. The concept of the information network caused different reactions in the period of faster development of technologies.

According to Reynard Swank, the basic characteristics of an information network are [2]:

1. Information resources
2. Users
3. Intellectual organization of documents and data
4. Resource delivery methods
5. Formal organization
6. Bidirectional communications network

In general, two approaches were used to define the concept of information networks. The first approach defines a network in terms of its functional organization (used by Swank). The second approach relies on the structure of information transfer and deals less with functions. Some authors followed both approaches.

In [3] twelve critical components essential to the proper development of network objectives were described. Several network configurations were performed, considering the structures by which information can be transmitted.

In this paper, we introduce the concept of information network, its significance, and its application in various scientific disciplines. The remaining of this paper is organized as follows. In the sction 2, we consider the basic concept of the information network. Various definitions and basic characteristics of information networks are given. Different network classifications are also mentioned. The formula for calculating network flexibility is presented, as well as the adjacency matrix. Section 3 introduces the wideness of the application of the information network. Specifically, we consider three examples: bisociative information networks, deep information networks, heterogeneous information networks, and we will point out possible advantages of their application. The final Section 4 contains concluding remarks.

2   Classification of information networks

Duggan in [3] describes twelve critical components essential to the proper development of network objectives.
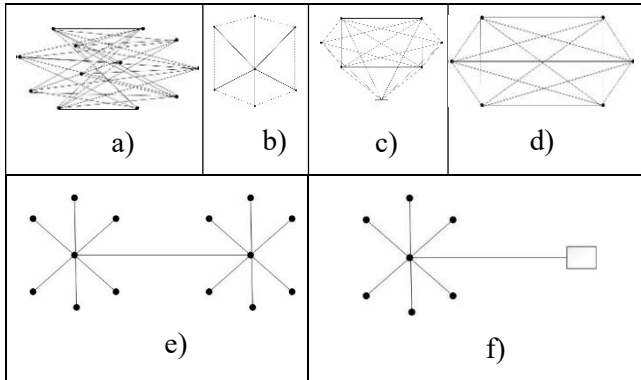
Figure 1. Network classification according to (a,b,c,d,e,f) [3]

There were listed several network configurations, through which she recognizes the different structures by which information can be transmitted. She proposes six structural forms shown in Figure 1 and Table 1 where number of links (C) that each structure requires were presented Network classification according to [3] is shown as follows:

a)   interface of two directed networks
b)   directed network
c)   undirected network with a specialized centre
d)   undirected network
e)   interface of two directed networks
f)   directed network with a specialized center

Table 1. The number of links in the function of nodes number for specific network type given in Figure 1

Network type	The number of links in function of nodes number
Interface of two undirected networks	$C = N\left(\frac{N-1}{2}\right) = 12\left(\frac{12-1}{2}\right) = 66$
Directed network	$C = N - 1 = 6$
Undirected network with a specialized centre	$C = N\left(\frac{N-1}{2}\right) = 21$
Undirected network	$C = N\left(\frac{N-1}{2}\right) = 15$
Interface of two directed networks	$C = 13$
Directed network with a specialized centre	$C = N - 1 = 7$

Ruth Davis also used both approaches in her description of the „National Biomedical Communication Network ". She represents the four types of network organization shown in Figure 2. The centralized and decentralized structure corresponds to the directed and undirected form described by Duggan. Also, the composite centralized structure is similar to Duggan's representation of the interface of two directed networks. The fourth structure described by Davis is very significant. She shows that the complexity of the control system increases with the hierarchical structure, while at the same time greater flexibility of network and intercommunication is achieved [4].

Network classification according to Ruth Davis [4]:

a)   fully decentralized network
b)   fully centralized network
c)   composite centralized network
d)   a third level hierarchical network
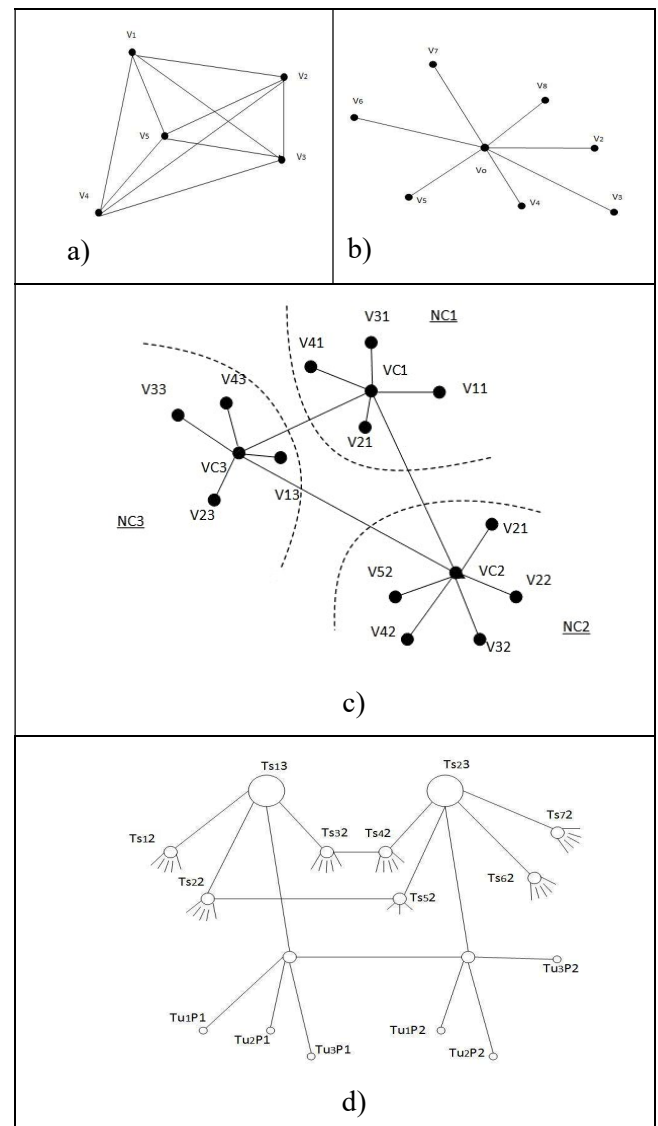
Figure 2. Network classification according to Duggan (a,b,c,d) [4]

According to Richard Nance, we can define information network as a set: $N = \{U, I, C, A, f, f'\}$ where [1]:

- U is a set of users of information resources $i \in I$
- I is a set of information resources accessed by users $u \in U$
- C is a set of information centers
- A represents a set of all branches connecting the nodes, each branch has one or both labels: m denotes the transmission of a message between two nodes, and b is the transfer of a document
- Functions f and f' constitute information transfer structure

An information center is defined as an entity where requests for information can be submitted by users, as well as an entity in which information resources are located. There are two types of information transfer within an information network [1]:

1. Message transmission (accomplished using function f) – these are the information required to access information resources.
2. Document transmission (accomplished using function f') – used to supply members of the set of users (U) with information resources (1)

Therefore, the first transmission refers to unmarked information, i.e. no final destination is defined for the transfer. While the last transmission refers to marked information, the i.e. transmission of information $i \in I$ to the known destination. Function f is used to gain access (structure of the message transmission), and f' for delivery or response (structure of the document transfer).

2.1 Structural Classification of Information Networks

If we assume $N = |C| > 1$, [1] [5]

1. The information network N with N centres is cyclic if and only if:

$$id(v_1) = od(v_1) = 1, \forall v_1 \in C \qquad (1)$$

2. The information network N with N centres is decentralized if and only if:

$$id(v_1) = od(v_1) = N - 1, \forall v_1 \in C \qquad (2)$$

Notation *id* and *od* refer to the number of input and output branches of the node respectively.

3. The information network N is strictly hierarchical if the graph is obtained by substituting all two-cycle arcs in G by a non-directional branch.

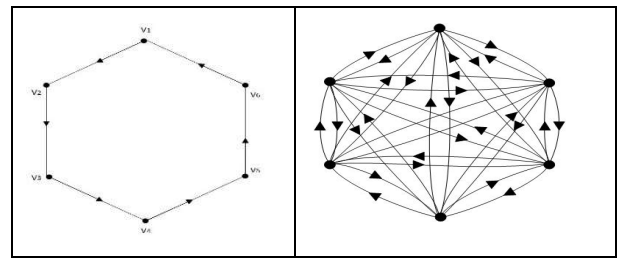Graphs of the cyclic and decentralized networks are shown in Figure 3: (a) cyclic network, (b) decentralized network.

Figure 3. Cyclic and decentralized network [1]

For a network with N nodes, minimum (q^0) and maximum (q^1) number of branches are given:

$$q^0 = N \text{ and } q^1 = N(N - 1) \qquad (3)$$

Each node must have two branches. However, according to the definition of an information network, access to each node must be provided, and also each node, must be able to be accessed by each node. So, we have:

$$id(v_1) \geq 1 \text{ } i \text{ } od(v_1) \geq 1 \qquad (4)$$

Minimum in this case is:

$$id(v_1) = od(v_1) = 1 \qquad (5)$$

This can be met only if G is a directed cyclic graph with N branches. In order to get the maximum number of branches, each node should have the number of output branches of $N - 1$. Now we can define:

$$q(G) = \sum_{i=1}^{N} \frac{[id(v_1)+od(v_1)]}{2} = \sum_{i=1}^{N} id(v_1) = \sum_{i=1}^{N} od(v_1) \qquad (6)$$

Where q represents the number of branches in the message transmission structure for the information network with N nodes. For a cyclic network $q = q^c = q^0 = N$. In a cyclic network structure, for each node $v \in G$, from which the message is sent, there is one and only one node to which the message can be directed. This is the structure with the most restrictions. The decentralized structure has fewer constraints. We denote a cyclic network as 0-flexible and decentralized as 1-flexible [1]. Now we introduce the variable $Z(N, q)$ to calculate the degree of network flexibility with q branches and N nodes.

$$Z(N, q) = \frac{q - q^0}{q^1 - q^0} = \frac{q - N}{N(N-2)} \qquad (7)$$

Since $Z(N, q^0) = 0$ for a cyclic network with N nodes, and $Z(N, q^1) = 1$ for a decentralized network, any network can be called z-flexible, depending on the value of the variable.
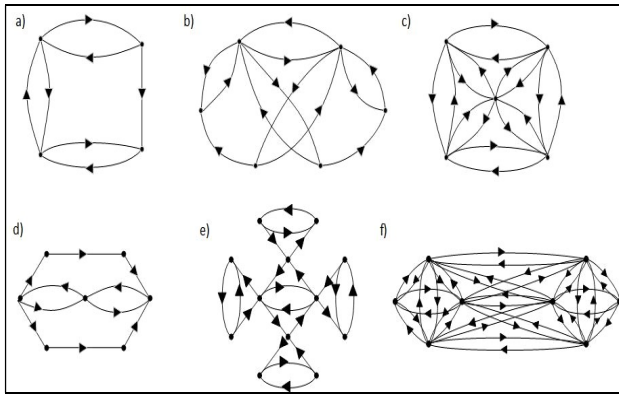
Figure 4. Examples of different network structures [1]

Figure 4 shows several network structures for which a degree of flexibility has been determined:

- is 0.38-flexible (4 nodes)
- is 0.25-flexible (6 nodes)
- is 0.73-flexible (5 nodes)
- is 0.9-flexible (7 nodes)
- is 0.10-flexible (12 nodes)
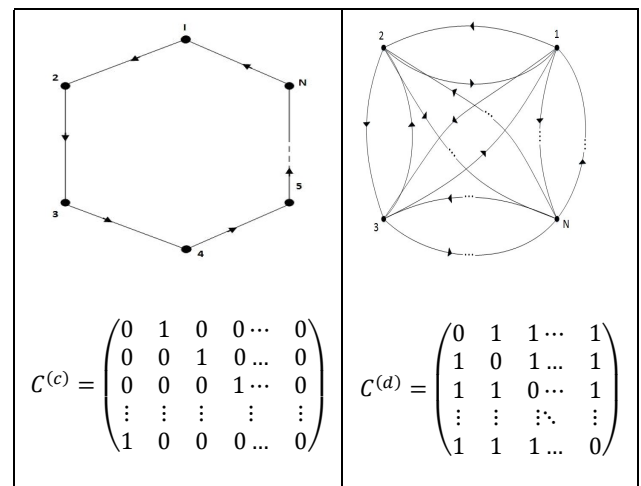- is 0.62-flexible (8 nodes)

2.2   Using Graphic-theoretical concepts

Diagrammatic representation of graphs proved to be very useful in establishing basic definitions, but when we explore individual structures, the matrix representation is also very important. Let's define C as a matrix of size NxN with values c_{ij}, where c_{ij}:

$$c_{ij} = \begin{cases} 1, if \; branch \; (v_i, v_j) \in G \\ 0, any \; other \; case \end{cases} \qquad (8)$$

The matrix C is called the adjacency matrix of a particular graph G. Now we add the labels c and d to the matrix C to emphasize the cyclic and decentralized structure. Figure 5 shows a graph of the cyclic structure and decentralized structure.

Figure 6 shows a strictly hierarchical and block structure. Next to this, the general form of the adjacency matrix for this structure is shown

.

$$C^{(c)} = \begin{pmatrix} 0 & 1 & 0 & 0 \cdots & 0 \\ 0 & 0 & 1 & 0 \ldots & 0 \\ 0 & 0 & 0 & 1 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \ldots & 0 \end{pmatrix} \quad C^{(d)} = \begin{pmatrix} 0 & 1 & 1 \cdots & 1 \\ 1 & 0 & 1 \ldots & 1 \\ 1 & 1 & 0 \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 \ldots & 0 \end{pmatrix}$$

Figure 5. The cyclic and decentralized structure [1]

Note that no node sends messages to itself, so the main diagonal of the matrix is filled with zeros. For a strictly hierarchical network shown in Figure 6. a), the adjacency matrix is symmetric, and treating only the upper triangle, we can obtain a block structure shown in Figure 6 b).
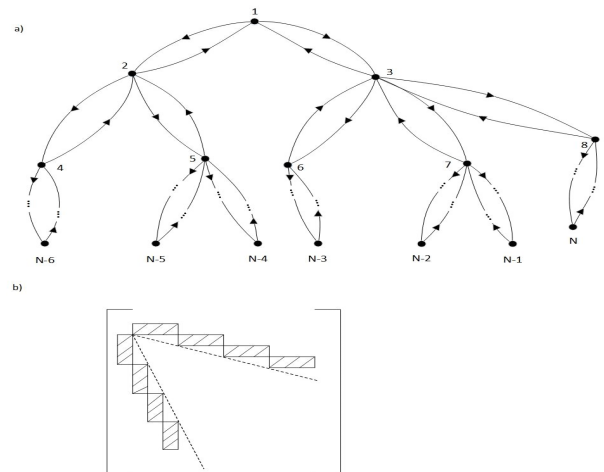
Figure 6. Strictly hierarchical and block structure of the adjacency matrix [1]

3 Application of information networks

Here we consider four types of information networks:

1. Bisociative information network
2. Deep information network
3. Heterogeneous information network
4. Space information network

3.1 Bisociative information network

Bisociation can be defined as a set of concepts that bridge two partially connected domains [6]. Bisociative knowledge discovery is a very challenging task motivated by the trend of specialization in research and development, which usually results in deep and isolated pieces of knowledge. Arthur Koestler considered bisociation to be the foundation of human creativity in science and art. He also thought that bisocial thinking occurs when a problem is perceived simultaneously in two or more different domains [7]. Figure 7 presents an example of bisociative network.

Nodes in Bisonets represent arbitrary information units such as: gene, protein, specific molecules, document, ideas, events etc. Nodes of the same type are grouped into vertical partitions (dotted lines). Depending on the view, vertical partitions can represent links or information units. Let's consider the network of movies. Movies can be described by a link between the actors. On the other hand, actors can be described by the link between the films in which they act. Links between the nodes are represented by branches. The certainty of a connection is represented by the 'weight' of the branch. A higher weight means higher certainty [8].
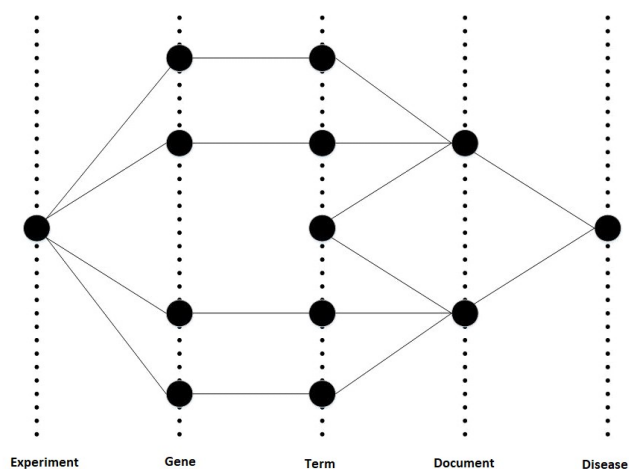

Figure 7. Example of bisociative network [8]

3.1.1 Bridging concepts

The bridging concepts connect subgraphs from different domains. The first approach to discover bridging concepts is the discovery of concept graphs in integrated data. Concept graphs can be used to identify existing and missing concepts in networks by searching related subgraphs. Once a concept graph has been discovered, domains and nodes can be analyzed in order to find concepts that connect information units from different domains [8].
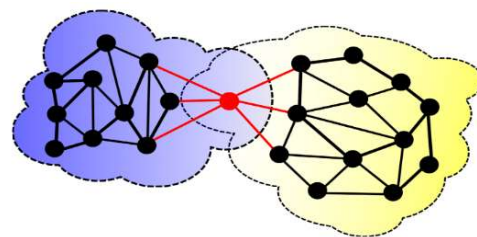

Figure 8. Example of bridging concept [8]

Other than network-based representation (shown in Figure 8), we can also use other representations such as textual [6].

3.1.2 Bridging graphs

The bridging graphs are subgraphs that connect concepts from different domains. Two different domains are connected mainly by a subset of concepts that are interconnected [6].

Bridging graphs can lead to surprising information coming from different domains, as they link unrelated domains. Bridging graph can also lead to linking two unrelated concepts from the same domain via a link through that or some unrelated domain [8].

The first step in the direction of detecting bridging graphs is the formalization and detection of cross-domain subgraphs[9]. Detected subgraphs can be further ranked according to their potential curiosity. Therefore, this curiosity is measured by considering domain connectivity, the rarity of links between the domains, and the distribution of neighbors of bridging nodes [8]. Figure 9 shows examples of bridging graphs.
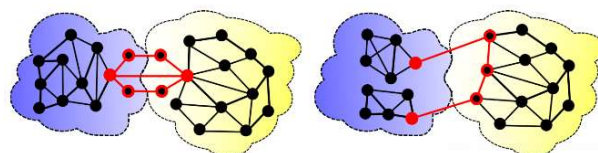

Figure 9. Examples of bridging graphs [8]

3.1.3   *Bridging by graph similarity*

The most complex type of bisociation does not rely on a simple type of connection between the two domains, but models such connection on a higher level. In both domains, two subsets of concepts that share structural similarities can be identified [6]. This is the most abstract pattern of bisociation that can lead to discoveries by linking domains that have no connection other than a similar interaction of the bridging concepts and their neighbors [8]. Here, we can combine spatial and structural similarities to detect bisociations based on the structural similarity of unrelated subgraphs [10]. Figure 10 shows an example of bridging by graph similarity.
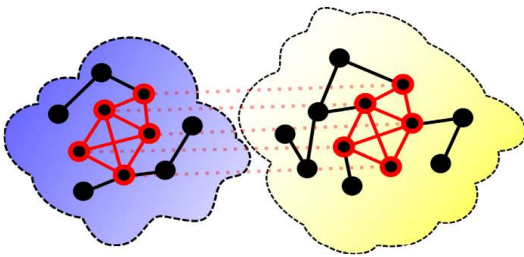

Figure 10. Example of bridging by graph similarities [8]

3.2   Deep information network

Deep information networks are based on information node, which uses input samples X_{in} to generate output samples X_{out} according to conditional probabilities $P(X_{out} = j | X_{in} = i)$ obtained by minimizing mutual information $I(X_{in}; X_{out})$ with the constraint of a given mutual information $I(X_{in}; Y)$ between the input and destination. Outputs from two or more nodes are combined without the loss of information to generate samples that need to be passed to the next information node. The last node outputs the class of each input data [11].

Each information node has an input vector X_{in}, whose elements get values from the cardinality set N_{in}, and output vector X_{out}, whose elements get value from the cardinality set N_{out}. The target class vector y_{train} is also available as an input, with elements from the set $[0, X_{class} - 1]$, all vectors have the same size N_{train}. If we set $N_{out} < N_{in}$, the information node performs compression [12]. Figure 11 shows a schematic representation of the information node, here we can see the input and output vectors with cardinality N_{in} and N_{out} and target vector y available during the training phase. As an example of the application of a deep information networks we will use an experiment in kidney disease. The objective of the experiment is to correctly classify patients affected by chronic kidney disease. The data set has a total of 24 medical characteristics.
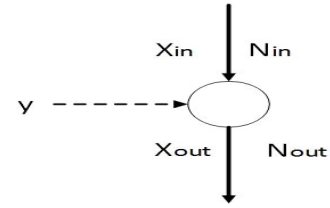

Figure 11. Schematic representation of the information node [12]

In the proposed information network layer zero has as many information nodes as there are features (24). These nodes are trained parallel. Then the outputs of layer zero, two at the same time, are mixed, and they become the input of layer one. In layer one we have 12 information nodes. These 12 information nodes are merged into six nodes in layer two, then these six nodes are merged into three in layer three. The three obtained nodes are combined into the final node, whose cardinality of the output is equal to two (which corresponds to the outputs: sick or healthy). The 24 input features are uniformly quantized with different values of $N_{in}{}^{(0)}$, depending on the feature. On the other hand, the value of N_{out} is three for all nodes except the last one [12].

Figure 12 shows how the mutual information $I(X, Y)$ develops in a deep information network and how it can be estimated to increase from the first to the last layer. Specifically, $I(X_{out}(2k, i), Y)$ and $I(X_{out}(2k + 1, i), Y)$, represented in the form of circles, are connected via segments on $I(X_{in}(k, i + 1), Y)$, represented in the form of triangles, to show that the lower limit goes from i=0 (blue markers) to i=3 (black markers). We can also conclude that $I(X_{in}(k, 1), Y)$ (blue triangle) is greater than $I(X_{out}(k, 1), Y)$ (green circle), which occurs due to compression by the information node which reduces cardinality of $X_{in}(k, 1)$ to three values. This is generally true when $I(X_{in}(k, i), Y)$ is compared with $I(X_{out}(k, i), Y)$ for all values of k and i [12].
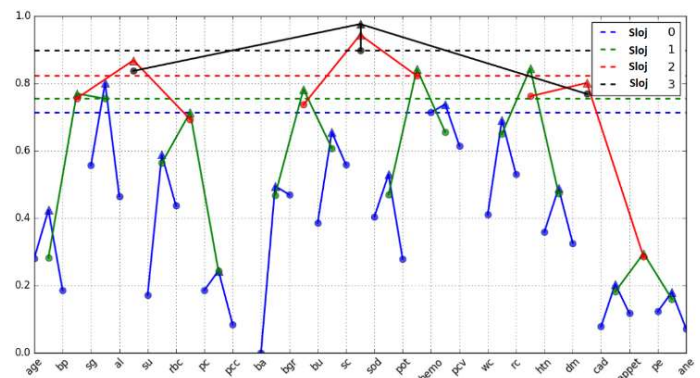

Figure 12. Mutual information values in the network [12]

Circles correspond to inputs, triangles to outputs, each dotted line represents the maximum value of mutual information at the input of a given layer. The proposed deep information network shows good results in terms of accuracy and represents a new modular structure, flexible and useful in various applications [11].

### 3.3   Heterogeneous information network

A heterogeneous information networks (HIN) can effectively integrate complex and interconnected data into a single framework, which is successfully applied in solving many biological problems. In doing so, nodes and branches in HIN can represent components and relevant interactions in the biological system [13]. HIN-based methods have two main advantages in analyzing biological data. First, SNR can be improved. Second, biological data from different sources can be combined and linked based on HIN [14].

HIN-based methods can be divided into five categories [14]:

- Network motive detection. A small subnet is composed of a group of nodes that represent interactions with each other. Network motifs are subnets that have a relationship with biological function

- Module detection. A subnet that has high-level nodes is called a module. In modules, local regions are tightly connected, and nodes with a similar or related function are strongly joined. Therefore, module detection approaches are always based on network clustering algorithms.

- Diseases and biomolecules. A biomolecule can be divided into macromolecules and small molecules. Discovering the links between disease and macromolecule is still hot research in the field of biology and bioinformatics. In addition, it has been shown that small molecules can also regulate biological processes.

- Drug development. Network analysis and appropriate devices are primarily used for computer-aided drug design, which can be applied to test pharmacology hypotheses as well as drug discovery.

- Interactions within the cell. The whole set of forms of molecular interactions can be represented by HIN.

Definition: A directional graph $G = (V, E)$ can define an information network that has an object type mapping function: $\phi: V \rightarrow A$ and link - type mapping function $\psi: E \rightarrow R$. Each node $v \in V$ has a unique object type, and each branch $e \in E$ has a unique relation type. Thus, assuming that the same relation type includes two links, these two links have the same type of start and end object. Compared to a traditional network, an information network can distinguish object and relation types [15].

Homogeneous networks have one object type (nodes) and one relation type (branches), while heterogeneous ones have several types of mentioned elements. The main procedure for the analysis of HIN-based biological data is shown in Figure 13. First, raw data from different data sets are analyzed as an adjacency matrix of several types of biomedical objects. Then knowledge from different sources can be integrated into HIN. Finally, computational methods are primarily used to analyze biological data in HIN [14].
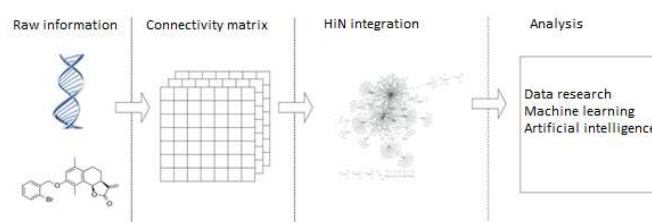

Figure 13. Biological data analysis based on HIN [14]

Disruption and decomposition of different types of biomolecules would lead to a serious disease that is often fatal. Therefore, the discovery of disease-related biomolecules has attracted much interest from biologists. So far, multiple HIN-based methodologies have been proposed in to effectively predict which biomolecules cause disease. The main processes of HIN-based computational methods for determining biomolecule-related disease priorities are shown in Figure 14.

Diagnosis and treatment of cancer can be improved by detecting disease-related genes. To date, many HIN-based methods have been developed to identify disease-related genes. Well - known associations of disease and genes can contribute to the discovery of candidate genes that can cause disease [14].

Bi-layer HIN, especially one consisting of a disease network, gene network is the most popular network model. In bi-layer HIN, detecting disease-causing genes is actually predicting the links between disease and gene nodes.
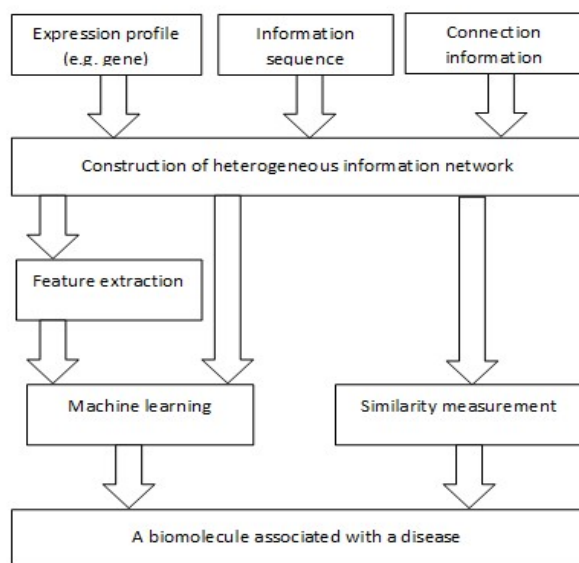
Figure 14. The process of predicting disease-related biomolecules [14]

3.4   Space information network

Space information networks are integrated networks based on various space platforms, including GEO, M/LEO satellites, and aircrafts on high altitude platforms (HAPS) to support real-time communication, massive data transmission, and data processing. In the last ten years, many internet giants have proposed various space information network development projects with the objective to provide internet access anywhere in the world. Compared to a terrestrial network, a space information network has a wider scope of application. Due to their unique characteristics, they are expected to play a key role in the application of communications, IoT, etc. [16].

Space information networks are composed of multiple satellites in the orbit. Due to the high dynamics of satellite movement, each satellite changes its position over time, so the communication links between the satellites are interrupted, leading to various changes in the network topology. A stable topology is not only a foundation of information exchange and resource sharing but also a prerequisite for network management. Therefore, the development of a satellite network topology algorithm is key [17].

In recent years, there have been many proposed topological algorithms for the space information network. C. Pan used an improved simulated annealing algorithm to solve a multi-objective optimization topology model with average and maximum delay in the network. This algorithm minimized communication delay, but only considered delay and did not consider network viability. Zhang and Xuan proposed an ant

colony algorithm to reduce network link congestion. Although the overall efficiency of the algorithm is great, the problem is slow convergence so the local optimum is easily reached [17].

Zhang and others have mainly dealt with architecture and key technologies for multiple-user transmission in space information networks. They represent some of the basic advantages of using multiple antennas, multiple access techniques, and cooperative transmission [18].

Space information network can also be described as $G = (V, E)$ in graph theory, where $V = \{1, 2, \dots, N\}$ represents a set of satellite nodes, and $E = \{e_{ij} \big| i, j \in V, \ i \neq j\}$ represent links between satellites. If the communication link is not blocked by the Earth and the distance between the two satellites is less than the maximum communication distance then we can say that the two satellites are visible [17].

Studies in the field of IoT are also very important. Namely, Bacco and others explored IoT applications and services within space information networks. Horizontal solutions were analyzed in order to allow interoperability between different protocol stacks and services. These solutions act as relay elements, which can be implemented vertically in different network segments [16].

4   Conclusion

The development of technologies, many of which are based on information networks and their basic concepts, has enabled numerous studies that discover new advantages and possibilities of the application of information networks in various scientific disciplines.

There is particularly great potential in the application of information networks in medicine. An experiment in kidney disease has shown how flexible, accurate, and useful a deep information network can be in determining whether a patient is healthy or ill.

The heterogeneous information network has shown great potential in biological system analysis, the identification of disease-related genes, discovery of disease-related biomolecules, and the development of drugs. We can conclude that the wider application of information networks in the future is unquestionable, but much more studies, experiments, and resources will be needed to discover and take full advantage of all the benefits and opportunities provided by these networks.

References

[1] R. E. Nance, R. R. Korfhage, and U. N. Bhat, "Information networks: Definitions and message transfer models," *J. Am. Soc. Inf. Sci.*, vol. 23, no. 4, pp. 237–247, 1972.

[2] C. Raynard, "Interlibrary Cooperation, Interlibrary Communications, and Information Networks-- Explanation."

[3] DUGGAN M, "Library Network Analysis and Planning," *J Libr. Autom.*, vol. 2, no. 3, pp. 157–175, 1969.

[4] R. M. Davis, "The national biomedical communications network as a developing structure.," *Bull. Med. Libr. Assoc.*, vol. 59, no. 1, pp. 1–20, 1971.

[5] U. N. Bhat, R. E. Nance, and R. R. Korfhage, "Information networks: A probabilistic model for hierarchical message transfer," *Inf. Sci. (Ny).*, vol. 9, no. 2, pp. 169–184, 1975.

[6] M. R. Berthold, "Towards bisociative knowledge discovery," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7250, pp. 1–10, 2012.

[7] N. Lavrač, M. Martinc, S. Pollak, M. Pompe Novak, and B. Cestnik, "Bisociative Literature-Based Discovery: Lessons Learned and New Word Embedding Approach," *New Gener. Comput.*, vol. 38, no. 4, pp. 773–800, 2020.

[8] T. Kötter and M. R. Berthold, "From information networks to bisociative information networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7250, no. May, pp. 33–50, 2012.

[9] U. Nagel, K. Thiel, T. Kötter, D. Piątek, and M. R. Berthold, "Towards discovery of subgraph bisociations," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7250, pp. 263–284, 2012.

[10] K. Thiel and M. R. Berthold, "Node similarities from spreading activation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7250, pp. 246–262, 2012.

[11] G. Franzese and M. Visintin, "Probabilistic ensemble of deep information networks," *Entropy*, vol. 22, no. 1, p. 100, 2020.

[12] G. Franzese and M. Visintin, "Deep Information Networks," no. 1, pp. 1–10, 2018.

[13] K. Tsuyuzaki and Nikaido I., "Biological Systems as Heterogeneous Information Networks: A Mini-review and Perspective," *HeteroNAM'18*, vol. 9, no. 3, pp. 209–224, 2018.

[14] P. Ding, W. Ouyang, J. Luo, and C. K. Kwoh, "Heterogeneous information network and its application to human health and disease," *Brief. Bioinform.*, vol. 21, no. 4, pp. 1327–1346, 2019.

[15] S. Y. and Han J., "Mining Heterogeneous Information Networks: A Structural Analysis Approach," *SIGKDD Explor.*, vol. 55, no. 9, pp. 1125–1129, 2012.

[16] L. Bai, T. De Cola, Q. Yu, and W. Zhang, "Space information networks," *IEEE Wirel. Commun.*, vol. 26, no. 2, pp. 8–9, 2019.

[17] S. Wan, P. Yang, M. Zhuo, and S. Hu, "An improved space information networks topology algorithm," *2021 IEEE Int. Conf. Inf. Commun. Softw. Eng. ICICSE 2021*, pp. 227–232, 2021.

[18] X. Zhang, L. Zhu, T. Li, Y. Xia, and W. Zhuang, "Multiple-user transmission in space information networks: Architecture and key techniques," *IEEE Wirel. Commun.*, vol. 26, no. 2, pp. 17–23, 2019.