The Usage of Graphs to Represent Big-Data in an Expert System

Sri. Eashwar K B[1], Sri. Y. B. Venkatakrishnan[2], Sivaranjani R[3]

[1]Assistant Professor, Department of CSE, SASTRA Deemed University, Srinivasa Ramanujan Center, Kumbakonam,
[2]Associate Professor, Department of Mathematics, School of Humanities and Sciences, SASTRA Deemed University, Thirumalaisamuthiram, Thanjavur,
[3]Post Graduate Student, SASTRA Deemed University, Thirumalaisamuthiram, Thanjavur.

Abstract: *This world of humans is changing from its current state to world of machines. Even humans are depending on machines to take complex decisions. So it is inevitable to construct such machines for complex real-world problems, for example bacterial infections prediction, automated traffic in smart cities, etc,. Expert systems are in the process of construction from the time AI (Artificial Intelligence) was emerged. There exists lot of stages in the process of building expert systems. One such a phase is Knowledge Base (KB) representation. Now-a-days voluminous of data are processed by Decision Making Systems. To handle the KB, it needs a proper structure to represent them, access them and to correlate them. One such a feasible structure to model big data is graphs. To select appropriate structures among the graphs for the specific set data is the problem to be resolved. The proposed idea is to discuss the characteristics and physical properties of the graphs for suitable type of data, to facilitate the storage of those data. Especially, when more and more voluminous of data are used in today's daily routines there arises a need to have such a system with the support of graphs to handle them. The idea behind choosing graph structures to represent the KB is because of its dynamic nature.*

Keywords: **Expert Systems, Knowledge Base, Artificial Intelligence, Graphs.**

1. INTRODUCTION

Since the data handling is getting tougher in these days, because of the heterogeneous nature of the data. So the need for the proper data structure is essential to store these types of dynamic-natured data elements. The latest trend to handle such a complex structured data is AI (Artificial Intelligence). Though, there exist several data bases or structures, they aren't suit to hold the heterogeneous type of data. So there is a need arise to manage these type of data. When a research was invoked on this issue, several models were considered. Several parameters were regarded and analysed. From that, they have identified that graph-based models are suitable to handle such a kind of evolving data. These models are having mathematical background. That is, there exists a way to define them, to reach them, to access the elements of the graphs, to add meaning to the connectivity, to define the order of direction, in which the elements gonna be accessed. So there is a need to have a proper procedure to execute the above mentioned tasks regarding the graph structures [4]. A simple graph may be expressed as follows, according to the application domain:

"A graph is a structural representation of any real-world entities, which are connected either via a directed or undirected paths."

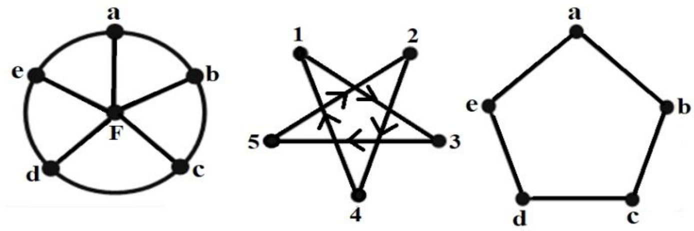


Figure 1. Various Graph Structures

Each and every structure of a graph models has its own meaning and way to access the elements (nodes) of the given graph. For example, the graph structures that are shown in *Figure 1*, show the unique forms of graphs such as wheel, star, and

The applications of graph theory or graph models in the field of AI and Expert Systems is remarkable and to be deeply concerned in the perception of research.

Now, the focus is turning towards the expert systems in various domains. Especially in the field of Bio-Chemistry[1], Medicine, Maintaining road traffic in an automated way (smart cities)-*probably giving suggestions to the drivers to pick the path which is congestion-free,* to assess the reason for the delay of assembled parts in a manufacturing plant, etc,. Wherever a critical decision has to be made like an experienced master of that domain, there an expert system can play a vital role.

2. Expert System

As from the word itself, the expert system resolves difficult decision-making problems through the collected facts and a special type of methods that allows taking conclusions quickly and rationally, called heuristics. These expert systems usually have the capability to inference with the deep knowledge about that domain. They are unconventional systems that work in a non-procedural way for solving situations like in game theory (example Chess), financial planning, assembling components of a machine, monitoring systems (for example flight monitors), where previously all actions rely on human expertise [5].

2.1. Expert System and human interaction

It has certain primary components and interfaces with human individuals who plays vital role in the development of the system. The entire arrangement is shown in *Figure 2*.

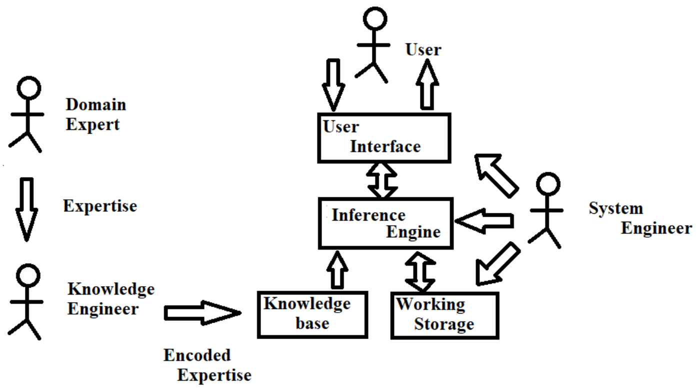


Figure 2. Components of an Expert System

2.1.1. Components and interfaces

- **Knowledge base:** Entire representation of the facts of that domain expertise.
- **Working Storage:** The data which are in need to solve the issues of the given problem.
- **Inference engine:** This is the core area of the system, which actually receives recommendations from the knowledge base and issue-specific data in Working Storage.
- **User Interface:** It establishes the connectivity between user and the system.

2.1.2. Roles being played by the individuals with the system

- **Domain Expert:** The individuals who are expert in a particular area.
- **Knowledge Engineer:** The individual who transforms the expert's knowledge in a particular acceptable form, which is going to be used by the expert system.
- **User:** The individual or a group or a firm, (who or) which is in need of some solution to resolve a complex problem [5].

2.1.3. Examples to Expert System

The first expert system, DENDRAL (1970s, Stanford university), which interpreted the output of the mass spectrometer as accurately as chemical experts.

MYCIN was an one of the pioneer expert system that used the idea of artificial intelligence to identify bacteria causing severe infections, and to recommend medicines, with the dosage adjusted for patient's physical properties like body weight, temperature, age, etc,. It used “certainty factors” to predict the diseases accurately [5].

CADUCEUS was a medical expert system finished in the mid-1980s, which could diagnose 1000 different diseases. It started in 1970s-but could end up late because of the time taken to build its vast knowledge base [5].

PROSPECTOR is an expert system constructed for decision-making problems in mining of minerals. It helps the geologists in evaluating the region of interest of an exploration site or region for occurrences of deposits of particular types of minerals [5]. The knowledge base of the PROSPECTOR used semantic network similar to the graph models (but not exactly) to express three kinds of relationships among the nodes such as,

1. Logical Relations.

2. Plausible Relations.

3. Contextual Relations [5].

Nutrition Advisor Expert System (NAES)
It is an expert system provides expert decisions and guidance into the dynamic domain of management of nutrition. It performed deep nutritional analysis on the food items then find out the deficiencies and suggested the food items recommended by U.S. Government [7].

3. Importance of Graphs in Expert Systems

As from the Figure 2.0., it is well-known that the importance of designing the knowledge-base for an expert system is inevitable. It is in fact that the reliance on data is growing like anything in our recent emerging trends in technology. Also the size of the data sets is bulging. Then added to the complexity on data, it's quite-often changing behaviour increases the urge to push the research community to search for new way of representing such a voluminous and dynamic-natured data (Big-Data) sets [1].

This necessary pushed the attention on mathematical models. The behavioural attributes of the current data are suitable to the nature of graphs, so they are, at present, in the focus.

Initially, graphs were used to model the geographical location related problems. Then its usage in modeling and visualizing the data set has been widened [1].

4. Application of Graphs in Chemistry

In 1857, Cayley founded that the graphs can be used to represent every chemical molecule, as every atom of that molecule is replaced with a vertex and atomic bonds between leaf level atoms, can be represented as edges. For example Uracil ($\mathbf{C_4H_4N_2O_2}$) can be represented as follows:

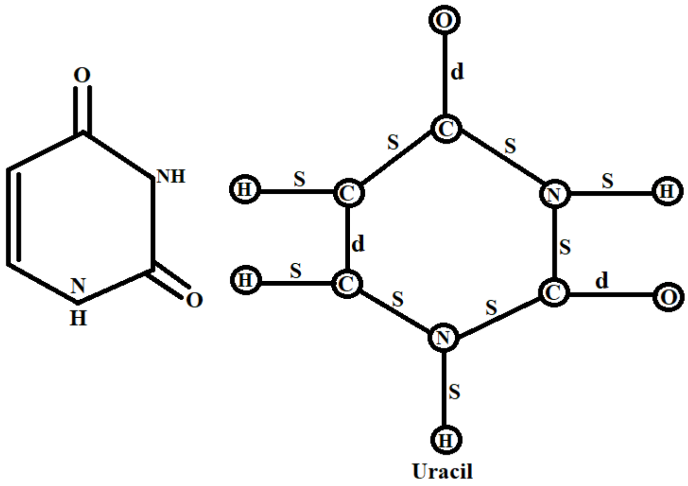


Figure 3. Graph Representation of Uracil

The valence of the respective atom is given by the degree of each vertex of such a graph.

Degree of vertex: *"Let G be a graph and v є V. The number of edges incident at v in G is called the degree of the vertex v in G and is denoted by* d_G *(v); or simply d (v)."*

Isomers: *"Molecules with different structural properties, but having same chemical formula are called isomers."*

According to graph theory, the isomers can be defined as follows: "The two graphs, which are not isomorphic, have the degree sequence as the same." The two molecular graphs, actually that are represented as trees (Figure. 4.0.), used to represent two isomers of the molecule C_3H_7OH (propanol). The graph that is shown in Figure. 5.0., represents aminoacetone C_3H_7 NO. The pair of multiple edges between Carbon and Oxygen atoms is used to represent the multiple bonding between them [6]:

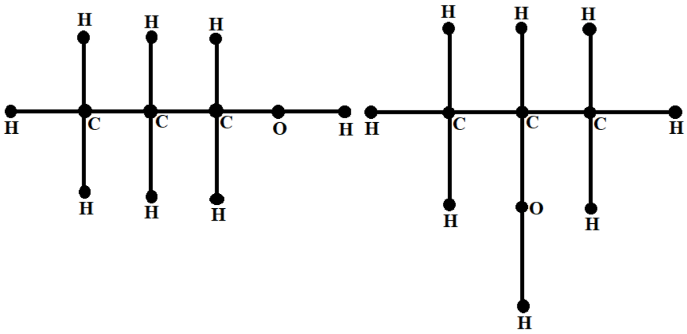


Figure 4. Applications of Graphs in Chemistry (Two Isomers of C_3H_7OH**)**

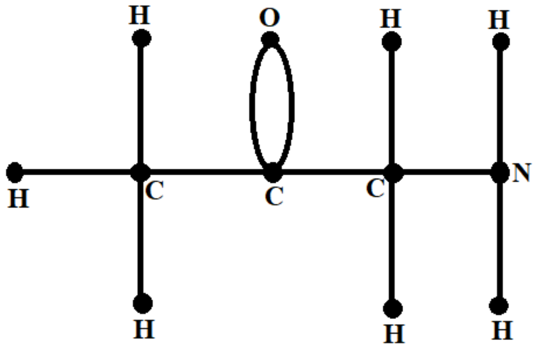


Figure 5. Applications of Graphs in Chemistry (Aminoacetone of C_3H_7NO)

5. Application of Graphs to Social Psychology

In 1960, at University of Michigan, Department of Group Dynamics, the trio Cartwright, Harary, and Norman, introduced the application of graph theory in Social Psychology.

Group dynamics is the domain about a specific communal group that shares relationships. A commonly used graph representation to express these relationships is *signed graphs*. It is a graph G, with + or – signs embedded with each of its edges. A specific edge *e* of G is positive, if the sign denoted on it is + (plus); otherwise the edge *e* is negative, denoted by a – (minus) on it.

Always a positive sign represents that the same social attributes are shared between two persons '*a*' and '*b*', i.e., they are "related".

A negative sign implies a alternative synonym. The communal oddness may be "same political ideology," "companionship," "likes particular cultural rituals," and so on. A cluster of individuals sharing these relational attributes creates a social system. A balanced social system has two individuals sharing a positive relationship. On the other hand, if we classify the group into two subgroups, then there exist a positive relation between the two individuals of the same subgroup and a negative relation between the two individuals of the different subgroups. For example, consider *a* and *b* belong to same subgroup, and *c* belongs to a different subgroup. If so, *a* and *b* must have positive relationship provided they have negative relationship with *c* [2].

A graph G, which has the two subset of a vertex set A as A_i; i=1,2, can be defined as a signed and balanced one. This arrangement is used to represent a balanced social system. In which, one subset may be empty (any edge in $G[A_i]$ is positive), the other, which is negative (for any edge between A_1 and A_2) (Figure 6.0).

For example, the signed graph G1 of Figure.7.0 (a) is balanced (take A1={a1,a2} and A2={a3,a4,a5}). However, the signed graph G2 of Figue.7.0. (b) is not balanced.

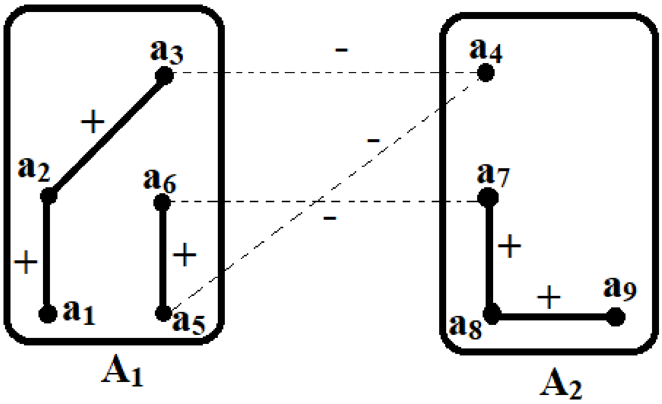
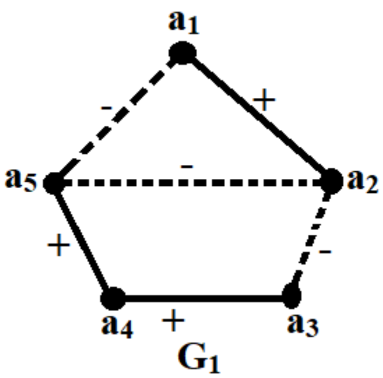


Figure 6. A Negative $a_1 - a_9$ path



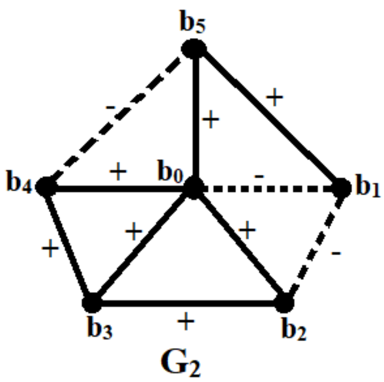


Figure 7.a. Balanced Graph **Figure 7.b. Unbalanced Graph**

A signed graph G, which consists of a path or cycle, is considered as positive when it has even count of negative signs; otherwise, it is framed as negative [6].

6. Samples of Knowledge Graphs Representing Rules and Facts

Usually, in any expert system, the knowledge base consists of production rules such as the following example, which is used to find the given character is, a vowel or not:

If (c = ‘a’ or c = ‘e’ or c = ‘i’ or c = ‘o’ or c = ‘u’)
Display “The entered character in ‘c’ is a vowel”
Else
Display “The entered character in ‘c’ is not a vowel”.

This logic can also be represented using a graph model as in the Figure 8.0.

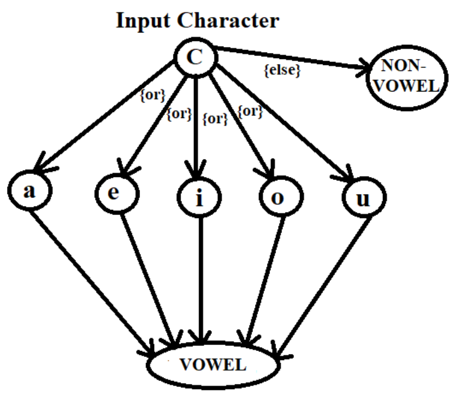


Figure 8. Knowledge Graph representing the logic

Similarly a directed graph can be used to represent a fact, “If there exist dense clouds and less wind then surely there is a possibility for the rain” as in Figure 9.0:

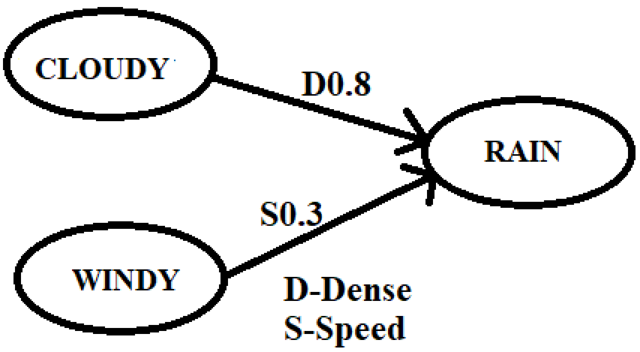


Figure 9. A Weight-edged Directed Graph representing the fact

Another example to use the graphs as the defined production-rules of a compiler is shown in Figure 10.0.

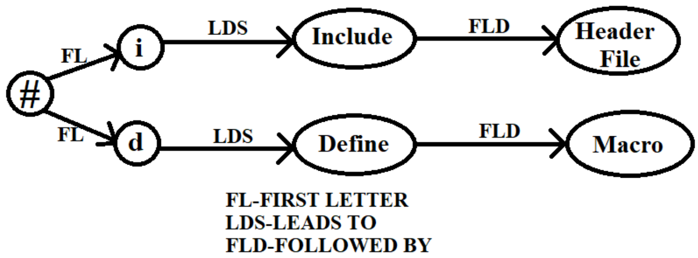


Figure 10. Knowledge Graph used in Parsing

Moreover, the main process of searching the appropriate information-graph from the KB, for the given query graph, can be done through the comparison between given query graph and the stored ones in the KB using graph isomorphism (Figure 11.0).

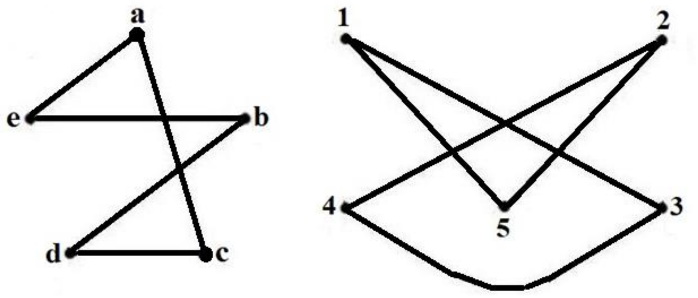


Figure 11. Graph Isomorphism

7. The processing of Graphs in Knowledge base

The information sent by the domain experts is given as *Input.* This information will be processed by an *information processor*, converts the information into a graph mode. When a user supplies a query as in a normal form, will be converted into a graph by the *graph processor* and send it to the *search engine*, which accepts graphs as input and search the *knowledge base* for the appropriate graph to retrieve. The retrieved graph will be interpreted as a user-understandable format [3] and send it to the user (Figure. 12.0.).

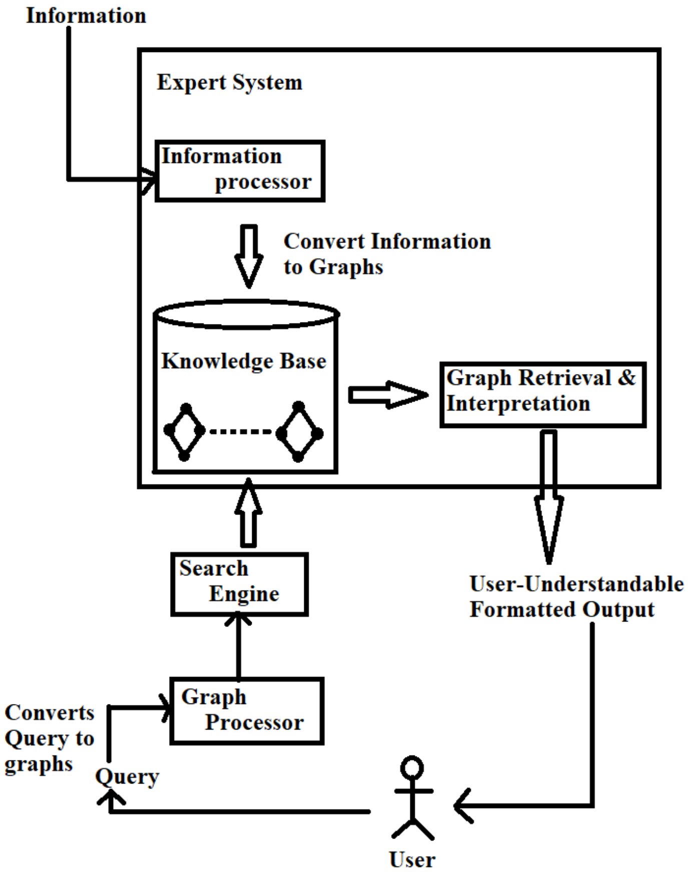


Figure 12. Processing of Graphs in an Expert System

8. Conclusion

From the given examples of the applications of graphs as a representation of data, in the domain of chemical compounds and social psychology, shows that the information can be processed efficiently. The applications of Expert system in these fields with the graph models as knowledge base representation, can exhibit efficiency in data (facts) retrieval and processing of graphs. Thus the nature of the real-world data can be successfully insulated into the graphs. There exist many efficient tools to process graphs into information and to reverse the process. There exist lot of real-world firms like Google, Facebook, and Amazon which are on the run to dig out more on the efficient use of graphs to represent large voluminous data. In future, more dynamic natured data representations will be in need. So there arises a need for more graph models to represent different data types.

9. References:

9.1. Journal

[1] *D. Pal, P. Rao, V. Slavov, and A. Katib, "Fast processing of graph queries on a large database of small and medium-sized data graphs," Journal of Computer and System Sciences, vol. 82, no. 6, pp. 1112–1143,* ***2016****.*

[2] *D. M. Walker, A. Zaitouny, and D. C. Correa, "On using the modularity of recurrence network communities to detect change-point behaviour," Expert Systems with Applications, vol. 176, p. 114837,* ***2021****.*

[3] *S. Yang, Z. Yang, X. Chen, J. Zhao, and Y. Ma, "Distributed aggregation-based attributed graph summarization for summary-based approximate attributed graph queries," Expert Systems with Applications, vol. 176, p. 114921,* ***2021****.*

[4] *Leszek Kotulski, Adam Sedziwy, Barbara Strug, " Conditional Synchronization in Multi- Agent Graph-based Knowledge System", Procedia Computer Science, vol 51, pages: 1043-1051,* ***2015.***

9.2. Books

[5] *Elaine Rich, Kevin Knight, Shivashankar B Nair, "Artifical Intellegence", 3rd edition, The McGraw-Hill Companies.*

[6] *R.Balakrishnan,K.Ranganathan,A,"Textbook of Graph Theory",2nd edition, Universitext, Springer,****2012****.*

9.2. Internet Resources

[7] *www.Science.gov.*