Algorithms for Graph Similarity

Tushar Agarwal

1783910056

Agarwal.tushar2905@gmail.com

March 16,2019

Abstract

This day's usage of computers is increasing in human life. Everything is becoming computer oriented. While this advancement is at its peak the most of the largely used applications one way or the other use graph theory, like search engines are largely based on graphs. We deal with independent but related problems, those of graph similarity. For the problem of graph similarity, we develop and test a new framework, for solving the problem using belief propagation and related ideas. We make substantial progress compared to the existing methods for problems. Proper understanding of various graphs present in graph theory is required to achieve understanding in real world applications. In this paper we demonstrate various graphs with their definitions, basic understanding and finally their importance and applications in real world. In our research we have identified different graphs.

Introduction

Graphs are important because graph is a way of expressing information in pictorial form. A graph shows information that equivalent to many words. A graph can give information that might not be possible to express in words. IN a letter to C. Huygens of 1679, G.W. Leibniz expressed his dissatisfaction with the standard coordinate geometry treatment of geometric figures and maintained that "we need yet another kind of analysis, geometric or linear, which deals directly with position, as algebra deals with magnitude". Graphs arise very naturally in many situations - examples vary from the web graph of documents, to a social network graph of friends, to road-map graphs of cities. Over the last two decades, the field of graph mining has grown rapidly, not only because the number and the size of graphs has been growing exponentially (with billions of nodes and edges), but also because we want to extract much more complicated information from our graphs (not just evaluate static properties, but infer structure and make accurate predictions). This leads to challenges on several fronts - proposing meaningful metrics to capture different notions of structure, designing algorithms that can calculate these metrics, and finally finding approximations or heuristics to scale with graph size if the original algorithms are too slow.

Graph Similarity

Our setting for graph similarity is as follows. We have two graphs on the same set of N nodes, but with possibly different sets of edges (weighted or unweighted). We assume that we know the correspondence between the nodes of the two graphs Graph similarity involves determining the degree of similarity between these two graphs (a number between 0 and 1). Intuitively, since we know the node correspondences, the same node in both graphs would be similar if its neighbors are similar (and its connectivity, in terms of edge weights, to its neighbors). Again, its neighbors are similar if their neighborhoods are similar, and so on. Or we can say that If Traversing of the graph is same about any vertex so, we can say that graph are similar about that vertex. Given two graphs, it is often really hard to tell if they ARE isomorphic, but usually easier to see if they ARE NOT isomorphic.
Here is our first idea to help tell if two graphs are isomorphic.
Suppose we have two graphs. In the first graph there are v1 vertices and e1 edges. In the second graph there are v2 vertices and e2 edges. Then in order for the two graphs to be isomorphic we must have:

v1 = v2

e1 = e2

In words, isomorphic graphs must have the same number of vertices and edges. It is important to note that just having v1 = v2 and e1 = e2 is
NOT a guarantee that two graphs will be isomorphic.

Suppose we have two graphs where each graph has the same number of vertices, v1 = v2 = n. Write the degrees of each vertex (with repeats) in ascending order for Graph 1. This gives a list of numbers that we can represent generally as d1, d2, d3, . . ., dn.

If the two graphs are isomorphic then when listing the degrees of Graph 2 in ascending order, we get the exact same list as above.
In short, ISOMORPHIC GRAPHS HAVE THE SAME DEGREE LISTS. More useful though, IF THE DEGREE LISTS ARE DIFFERENT, THE TWO GRAPHS ARE NOT ISOMORPHIC.

Basic Theory

Before we can understand application of graphs we need to know some definitions that are part of graphs theory. Authors of this paper has identified these definitions and has represented it in very easy to understand manner.

1. *Graph:* A graph – usually denoted G (V, E) or G = (V, E) – consists of set of vertices V together with a set of edges E. The number of vertices in a graph is usually denoted n while the number of edges is usually denoted m [1].
Example: The graph given in figure 1 has vertex set V= {1,2,3,4,5,6} and edge set= {(1,2), (1,3), (2,3), (3,4), (3,5), (4,5), (5,6)}.
2. *Connected graph*: A graph G= (V, E) is said to be connected graph if there exists a path between every pair of vertices in graph G.
3. *Degree of a vertex:* Number of edges that are incident to the vertex is called the degree of the vertex.
4. *Regular graph:* In a graph if all vertices have same degree (incident edges) k than it is called a regular graph.
5. *Complete graph:* A simple graph G= (V, E) with n mutually adjacent vertices is called a complete graph G and it is denoted by Kn. or A simple graph G= (V, E) in which every vertex in mutually adjacent to all other vertices is called a complete graph G.
6. *Cycle graph:* A simple graph G= (V, E) with n vertices ($n \geq 3$), n edges is called a cycle graph.
7. *Directed graph:* A directed graph in which. each edge is represented by an ordered pair of two vertices, e.g. (Vi, Vj) denotes an edge from Vi to Vj (from first vertex to second vertex).

Applications

Graphs are used to model many problems of the real word in the various fields. Graphs are extremely power full and yet flexible tool to model. Graph theory includes many methodologies by which this modeled problem can be solved. Authors of the paper have identified such problems, some of which are mentioned in this paper.

1.Computer networks are extremely popular in today’s life. In computer networks nodes are connected to each other via links. This final network of nodes forms a graph. In computer network graph is used to form a network of nodes and enable efficient packet routing in the network. This includes finding the shortest paths between the nodes, analyze the current network traffic and find fasted root between the nodes, finding cost efficient route between the nodes. Standard algorithms such as Dijkstra’s algorithm, Bellman-Ford algorithm are used to in the various ways with graph to find the solutions.
2. Structure of a websites containing many pages can be represented using a directed graph. Each page can be considered as a vertex. A link between exists if there is a link between two pages. This way it can be identified that which page is accessible form which page.
3.In electronic chip design each component is considered as a vertex of the graph. The machine that creates connection between this components a printed circuit board takes input
in the form of a graph where edges denote that there is a connection between the pair of components. The head that creates this connection on the board then find the optimal to moves across the chip to get the desired resultant circuit.
4. The new semantic search engine, which is known as **Facebook Graph Search** introduce by Facebook in March 2013. In general all search engine gives result in list of link, but Facebook
Graph Search give the answer to user in nature language rather than a list of links [10]. In Facebook Graph Search engine graph Search

feature combines external data into a search engine providing user-specific search results and the big data acquired from its over one billion users [10]. In Facebook Graph Search engine search algorithm is same, as Google search engine algorithm so searching will very faster in Facebook site.

5. For probabilistic decoding of LDPC and turbo codes in belief network Factor graph is used. The bipartite graph can also be used in **Query Log Analysis,** which is used for improve search engine capability [4]. Query Log Analysis would maintain the query with each respective website so searching becomes easy in search engine, the bipartite graph between search engine and URLs (Uniform Resource Locator). In Query Log Analysis method edges connected the query with its appropriate URL and capture some semantic relation between the query and the URLs [4]. Figure shows the example of the Query Log Analysis method in which left partition represents the query and the right partition represents the respective URL. This model is based on the raw click frequency (CF). Raw click frequency is to weight the query and URL on click graph.

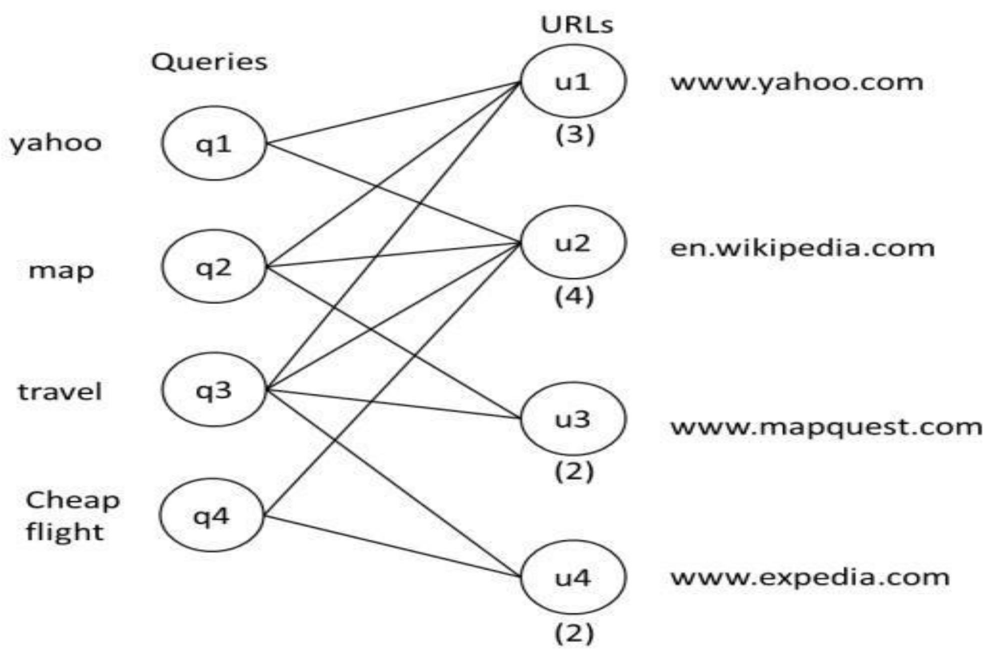

Figure: Example of Query Log Analysis

6. Chemistry: Graphs are used to model molecule structures for computer processing. Here atoms can be considered as vertices of a graph the bonds that connects them are represented as edges between them.
7.Biology: A Graph Theory is a very vast subject; it is also extensively used for the analysis in biological networks. In biology analysis the number of components of the system and their interactions is distinguish as network and they are normally represented as graphs where lots of nodes are connected with thousands of vertices [6].

Conclusion

In this paper authors have provided basic definitions that are crucial part of graph theory. These definitions are very easy to understand and provide clear idea of different types of graphs. All the necessary terminologies of graph theory are covered by these definitions. We tackle related problems in data mining: graph similarity. They are motivated by similar applications and objectives to study and analyze graphs that occur naturally as biological networks, social networks, web graphs and many others. Later various applications of graph theory have been identified and divided as per their fields. This paper explains where different graphs of graph theory are used in these real-world applications. Hence this paper gives clear idea of use of terminologies of the graph theory in real world applications, covering both basic knowledge and brief of where these terminologies are applied. One can easily understand these terminologies and get idea how they are used in real world. Here, we approach the problem with relevant ideas from belief propagation. We use the Linearized Belief Propagation (Fa-BP) algorithm with a similarity metric that is a normalized version of the Euclidean distance. This produces extremely intuitive results and is effective in both the weighted and unweighted, connected and disconnected graph settings.