Contingency tables by bipartite graphs and intersection graphs

S.AbdElrahman(3) El-Zohny H(1),M.M. Khalil(2)

[1,2,3]*Mathematics Department, Faculty of scienceAlAzhar UniversityCairo- Egypt*

Abstract: *In this paper we will convert weighted bipartite graph to intersection graph. Also we introduce the variation of the algorithm of bipartite graph and intersection graph.*
Keywords: *Graph, Contingency table, bipartite graph, intersection graph, clique, algorithms.*

I. Introduction

Definitions:-
Abstract graph:
 A graph G= (V,E) consists of a set of vertices V and a set of edges E. Elements of V are distinct. Elements of E are 2- sets of the form {x, y}, where x and y are both vertices in V.

Contingency table:
 A simple 2 way Contingency table: is a common method of analyzing the association between two categorical(qualitative) variables.
A p-way contingency table when (p>2).

Bipartite graph:
A graph G= (V, E) is called bipartite where V=X∪Y, with X∩Y=Ø, and every edge of G of the form {a ,b} with a ϵX ,b ϵY.

Intersection graph:
 A graph G= (V, E) is called intersection graph where the set of vertices is formed by a family of sets. Given two vertices a and b there is an edge {a, b} in E iffa∩b≠Ø.
Both Bipartite graph and intersection graph has no loops or multiple edges.
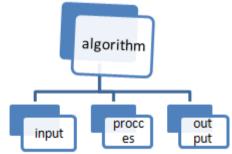
Multivariate data analysis:
Data with more than one variable

Clique:
defined to be a sub graph in which each pair of vertices joined by edge.

Algorithm:
In mathematics, computer science, and related subjects, an algorithm is a finite sequence of steps expressed for solving a problem. Algorithm can be expressed as

II. Main results

 We use graph theory to read the data given in contingency table , and cluster it .by this way the data converted to weighted bipartite graph and weighted intersectiongraph.

First:- For bipartite graph
Example (1)

Consider the simple contingency table (1) for 50 metal to detect specularity according to their size

	small	large
Specular	4	13
Non specular	15	18

Table (1)

We can convert this table to weighted bipartite graph, where the weight on the edges corresponding to the frequencies of the cell in the table.
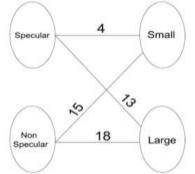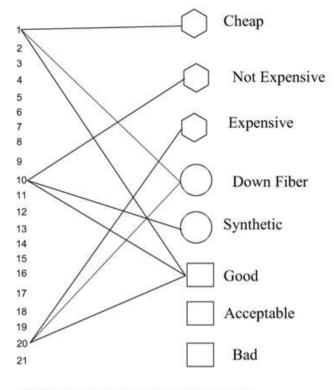

Figure (1) Weighted Bipartite graph

Example (2)
Consider the data base for different animals shown in the table (2):

		Cheap	Not Expensive	Expensive	Down Fibers	Synthetic	Good	Acceptable	bad
	Sleeping Bag		Price			Fiber	Quality		
1	One kilo Bag	1	0	0	0	1	1	0	0
2	Sund	1	0	0	0	1	0	0	1
3	Kompack Basic	1	0	0	0	1	1	0	0
4	FinmarkTour	1	0	0	0	1	0	0	1
5	InterlightLyx	1	0	0	0	1	0	0	1
6	Kompact	0	1	0	0	1	0	1	0
7	Touch the Cloud	0	1	0	0	1	0	1	0
8	Cats Meow	0	1	0	0	1	1	0	0
9	Igloo Super	0	1	0	0	1	0	0	1
10	Donna	0	1	0	0	1	0	1	0
11	Tyin	0	1	0	0	1	0	1	0
12	Travellers Dream	0	1	0	1	0	1	0	0
13	Yeti Light	0	1	0	1	0	1	0	0
14	Climber	0	1	0	1	0	0	1	0
15	Viking	0	1	0	1	0	1	0	0
16	Eiger	0	0	1	1	0	0	1	0
17	Climber Light	0	1	0	1	0	1	0	0
18	Cobra	0	0	1	1	0	1	0	0
19	Cobra Comfort	0	1	0	1	0	0	1	0
20	Foxfire	0	0	1	1	0	1	0	0
21	Mont Blanc	0	0	1	1	0	1	0	0

Bipartite graph due to the data from conengency table

Figure (2)

Second:- For intersection graph:

The difference between intersection graphs and bipartite graphs is the vertices where in the case of intersection graphs the vertices are sets.

We can convert weighted bipartite graph in example (1) into weighted intersection graph by put specular (1-yes,0-no),and size(1-small,0-large).as shown in the figure (3)
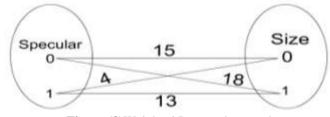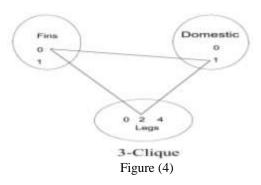

Figure (3) Weighted Intersection graph

By using intersection graphs the observations can be read as clique [if we have 3variables we get 3clique,and if we have 4 variable we get 4 clique].

Example :suppose we observe3 characteristic of chicken {Fins(0,1),domestic(0,1),legs(0,2,4)}.and represent them in the traditional tabular form

Tabular form

Chicken	0	1	2
	Fins	Domestic	Legs

3-Clique

Figure (4)

Example (3)

Consider sample fromzoo data set that consist of 17 characteristic of 101 animals. For simplicity we selected a subset of these sample consist of 15 animals and 4 variables. The animals are

1) Antelope   2)bear   3)carp4)catfish   5) chicken
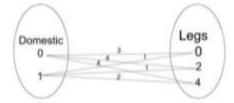
6) clam7)crow8)duck9)goat10)sealion

11) worm12)elephant13)cat   14)monkey15)bat

 and the variables are {wild(0,1),Fins(0,1),domestic(0,1),legs(0,2,4)}.

Simple 2- way contingency table:

We have two variables (Domestic, Legs) to convert contingency table into intersection graph

	Legs			Total
	0	2	4	
Domestic 0	3	4	4	11
Domestic 1	1	1	2	4
Total	4	5	6	15

Weighted intersection graph

An algorithm to convert 2-way contingency tables into intersection graphs:-

[1] Start

[2] read the data sample $X_{i,j}$,$W_{i,j}$(where i rows, and j columns , $W_{i,j}$ is the frequency between i,j)

[3] For (i=0,i<=1,i++)

[4] For (j=0,j<=4,j+2)

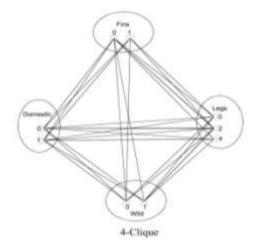[5] If∃ relation between i,j draw an edge between i,j

[6] Print$W_{i,j}$ on the edge

[7] Else go to line 3

[8] End

Consider the table
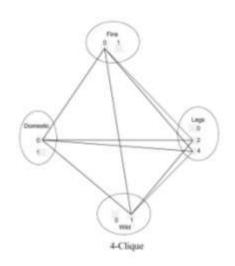
		Fins	Domestic	Legs	Wild
1	Antelope	0	0	4	1
2	Bear	0	0	4	1
3	Carp	1	1	0	0
4	Catfish	1	0	0	0
5	Chicken	0	1	2	1
6	Clam	0	0	0	0
7	Crow	0	0	2	1
8	Duck	0	0	2	1
9	Goat	0	1	4	1
10	Sea lion	1	0	2	0
11	Worm	0	0	0	1
12	Elephant	0	0	4	1
13	Cat	0	1	4	1
14	Monkey	0	0	4	1
15	Bat	0	0	2	1

The data in the table converted to intersection graph where the vertices of the graph are the variables.

4-Clique

Conditional algorithm for clusters intersection graphs:

[1] Start

[2] Read the data base Xi,j,k,l with Wi,j,k,l

[3] Draw graph with nodes i,j,k,l

[4] For (i=0,i<=1,i++)

[5] For (j=0,j<=1,j++)

[6] For (k=0,k<=4,k+2)

[7] For (l=0,l<=1,l++)

[8] If the relation between I,j,k,l occurs twice or more draw the relations

[9] Else

[10] Go to line 4

[11] End

4-Clique

References

[1].   Cook ,D.J.,Holder,L.,B.,2006.Mining graph data. New York :john.

[2].   Intersection graph –based representation of contingency tables,Valter Martins Vairinhos.

[3].   Anderson,T.W.(1984).An introduction to multivariate statistical Analysis,John Wiley.

[4].   Intersectiongraph :an introduction ,Madhumangal Pal 2013.

[5].   SatuElisa,:graph clustering, Computer Science Review, (2007),27-64.