

# Estructures de grafs amb equivalències d'arestes aplicades a l'anàlisi de dades relacionals

Laura Rodríguez Navas

laura.rodriguez.navas@upc.edu

Universitat Politècnica de Catalunya

## Resum

La teoria de les 2-estructures [1] proporciona una infraestructura matemàtica per a la descomposició de grafs. Permet representar els grafs en una única estructura algebraica, una 2-estructura. En aquest treball d'investigació es contribueix en l'estudi de les 2-estructures. Per un costat, es dissenya i s'implementa un conjunt d'algoritmes per a la creació i la visualització de les 2-estructures i per un altre costat, s'estableixen teoremes i/o resultats per a enfortir l'anàlisi de dades relacionals mitjançant l'ús d'aquestes estructures.

## Introducció

En els últims anys s'ha produït un increment significatiu en la capacitat d'emmagatzemar i compartir dades. Segons IBM (International Business Machines), el 90 % de les dades del món s'han generat en els últims anys. La importància estratègica de les dades massives no consisteix en la quantitat sinó en les aplicacions potencials que ofereixen. Alguns exemples. Les dades en blocs i xarxes socials s'aprofiten per a dissenyar estratègies polítiques. Les dades poden ajudar a gestionar ciutats o recursos naturals, a estudiar el canvi climàtic o a promoure el desenvolupament de països. També en la caracterització de malalties complexes a escalar molecular combinada amb l'historial mèdic i del tractament amb proves diagnòstiques o d'imatge afavoreix oportunitats sense precedents per a personalitzar la medicina. I altres.

Davant de l'allau de dades o tsunami de dades, en aquest document s'hi reflecteixen alguns punts teòrics que en milloren l'anàlisi.

## Dades Relacionals

Les dades relacionals amb les que es treballa inicialment es poden incloure en un dels tres tipus de fitxers:

- ARFF (Attribute-Relation File Format)

Arxius de text en format ASCII que descriuen una llista d'instàncies que comparteixen un conjunt d'atributs.

- TXT

Arxius de text estàndard que contenen text sense format.

- DB

Arxius de bases de dades genèriques.

Els fitxers es transformen en taules SQLite [2]. Les taules d'aquest tipus de les bases de dades relacionals s'utilitzen en aquest treball ja que alhora d'iniciar una línia d'investigació és molt important la senzillesa i la generalització que aporten.

## Recorreguts en una taula SQLite

Una vegada s'han transformat les dades d'un fitxer a taula SQLite, es recorren les files de la taula per extreure'n les dades. Amb les dades es crea un graf dirigit complet<sup>1</sup>. Cada element no repetit de la taula representarà un node del graf. I cada node estarà enllaçat amb la resta de nodes. Una arista discontinua representa un enllaç.

Després es realitza un segon recorregut que identifica les parelles de nodes que apareixen a la taula.

	A	B	C	D
1	a	b	d	e
2	a	f	d	e
3	c	c	c	c

Figura 1. Exemple de taula SQLite.

Sent  $i$ , un índex que recorre les columnes de la figura 1 i  $j$ , un índex que recorre les files de la figura 1, llavors per a cada parella de valors  $(i, j)$ ,  $i \neq j$ , existirà un enllaç entre els elements de la parella. Una arista contínua representa un enllaç. El graf resultant s'anomena graf de Gaifman [3] o graf primer.

**Exemple 1.** A la figura 2 es pot veure el graf de Gaifman resultant de la figura 1.

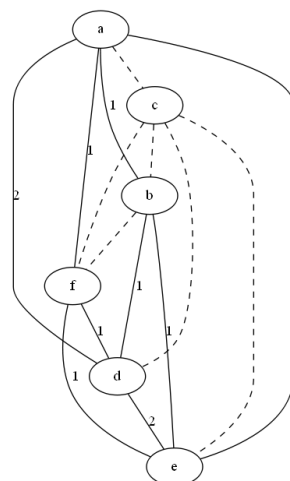


Figura 2. Graf de Gaifman.

<sup>1</sup> un graf és dirigit i complet quan entre totes les parelles de nodes del graf existeix una arista.

## Grafs

En aquest treball, a partir dels grafs de Gaifman, es desenvolupen quatre tipus de grafs:

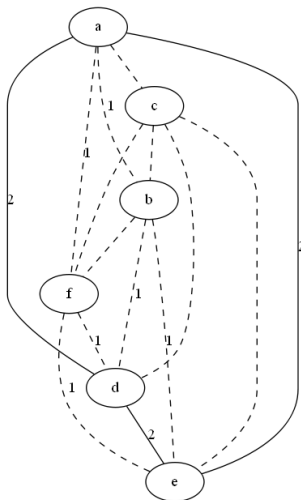
### - Graf pla

Graf de Gaifman que conté dues classes d'equivalències<sup>1</sup>. La primera classe d'equivalència representa les parelles de nodes que apareixen a la taula (arestes contínues). La segona classe d'equivalència representa les parelles de nodes que no apareixen a la taula (arestes discontinúes). El graf de la figura 2 representa el graf pla de la taula de la figura 1.

### - Graf pla amb llindar

Graf pla que també conté dues classes d'equivalències. Una classe d'equivalència representa les arestes amb el nombre d'equivalències<sup>2</sup> per sota el llindar<sup>3</sup> (arestes discontinúes). L'altra classe d'equivalència representa les arestes amb el nombre d'equivalències és superior o igual a el llindar (arestes contínues).

**Exemple 2.** Considerant la taula de la figura 1, la figura 3 mostra el graf pla resultant amb llindar igual a dos. Les arestes etiquetades amb el valor igual o superior al llindar són les arestes contínues. I les arestes etiquetades amb el valor inferior al llindar són les arestes discontinúes.



**Figura 3.** Graf pla amb llindar.

### - Graf lineal

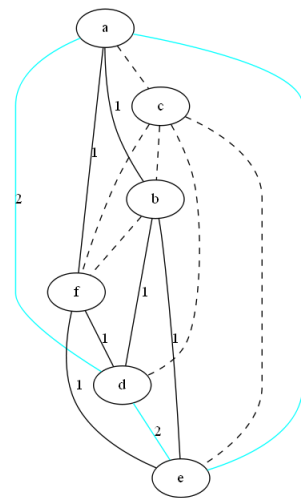
Graf pla que conté més de dues classes d'equivalències. Segons el nombre d'equivalències de cada aresta, la classe d'equivalències que la representarà segueix la nomenclatura de la taula següent:

Nombre d'equivalències	Classes d'equivalències
0	Negre/Discontínua
1	Negre/Contínua

2	Turquesa
3	Verd
4	Magenta
5	Taronja
6	Blau
7	Vermell
8	Groc
9	Marró
$\geq 10$	Gris

**Taula 1.** Classes d'equivalències.

**Exemple 2.** Considerant la taula de la figura 1, la figura 4 mostra el graf lineal resultant.



**Figura 4.** Graf lineal.

### - Graf exponencial

Graf lineal que agrupa les arestes en diferents classes d'equivalències. Cada classe d'equivalència comprèn un interval exponencial, que s'inicia en  $2^0$  i acaba en  $2^n$ . L'interval exponencial indica l'agrupació de nombres d'equivalències que representarà cada classe d'equivalència seguint la nomenclatura de la taula següent:

Nombre d'equivalències	Classes d'equivalències
(0, 1)	Negre/Discontínua o contínua
(2, 3)	Turquesa
(4, 7)	Verd
(8, 15)	Magenta
(16, 31)	Taronja
(32, 63)	Blau
(64, 127)	Vermell

(128, 255)	Groc
(256, 511)	Marró
$\geq 512$	Gris

**Taula 2.** Taula grups classes d'equivalències.

<sup>1</sup> Les classes d'equivalències representen les arestes acolorides d'un graf. Dues arestes són del mateix color si pertanyen a la mateixa classe d'equivalència.

<sup>2</sup> Nombre d'aparicions d'una parella de valors en una taula.

<sup>3</sup> El lllindar és un nombre enter introduït per l'usuari.

## Clans

Las subsiguientes segunda, tercera y cuarta páginas del documento se escribirán en un formato uniforme de 2 columnas de acuerdo con los márgenes indicados en la sección 2 de este documento. Por supuesto, no olvide mantener Times New Roman 10 para el texto con espaciado simple entre líneas.

Una recomendación final. **No necesita numerar las páginas** ya que esto se hará cuando se compongan las Actas definitivas.

## Estructura

Una de las partes más complejas a la hora de editar un artículo es sin lugar a dudas las figuras. Por esta razón no consideramos muy eficiente dar excesivas reglas acerca de su formato. Solamente citar varios consejos generales. Siempre que sea posible, se deberán colocar las figuras lo más cerca posible del lugar del texto donde se les haga referencia por primera vez. Por otra parte se procurará adecuar el tamaño de la figura para que ocupe la anchura de una columna, siempre y cuando la claridad de la misma no se vea comprometida, en cuyo caso se podrán utilizar las dos columnas, hasta un máximo de 17.5 cm.

Además de la libertad dada para el posicionamiento y edición de las figuras, nos gustaría proponerle una recomendación muy simple para la uniformidad, que concierne a los **pies de las figuras**. Deben ser escritos justo debajo de la figura, preferiblemente centrados con referencia al formato de la figura y escritos en fuente **Times New Roman 9 itálica** para diferenciarlos del texto principal. Las figuras se numeran secuencialmente y su número debe aparecer en negrita, exactamente como está indicado en nuestra figura 1. El número de orden así como la abreviación fig. deberán ir en negrita.

Distorsión armónica	2 Arm.	3 Arm.	4 Arm.
Señal A	-51 dB	-53 dB	-54 dB
Señal B	-76 dB	-65 dB	-44 dB

**Tabla 1.** Ilustración de la edición de una tabla

## Resultats

Tiene ahora todas las pautas básicas que nos gustaría que usara al preparar el documento definitivo para presentar al JSWEB 2006. Por favor, tenga en cuenta de nuevo los **márgenes** a respetar en referencia a un **DIN-A4**, la longitud máxima de 8 páginas, las **áreas** dedicadas a cada una de las diferentes partes de su documento así como los **tipos y tamaños de las fuentes** que deben usarse para diferenciarlas. Como habrá adivinado hasta ahora, esta plantilla de muestra ha sido escrita en total conformidad con tales pautas.

## Treball futur

Aquí se pondrían los agradecimientos relacionados con el proyecto u organización que ha financiado la investigación.

## Referències

Para las referencias se utilizará una fuente Times New Roman 9. Deberán figurar en el mismo orden en el cual han aparecido en el texto.

- [1] A. Ehrenfeucht i G. Rozenberg. "The Theory of 2-Structures: A Framework for Decomposition and Transformation of Graphs". World Scientific, 1999.
- [2] SQLite. <https://www.sqlite.org/>
- [3] Hodkinson, I. Otto, M. "Finite conformal hypergraph covers and Gaifman cliques in finite structures". The Bulletin of Symbolic Logic, 2003.