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 [R] 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:

* ARFF1 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.

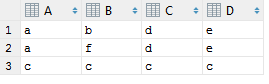
Els fitxers es transformen en taules SQLite [R]. Les taules d’aquest tipus de les bases de dades relacionals s’utilitzen en aquest treball ja que alhora d’iniciar una línia d’investigació és molt important la senzillesa i la generalització que aporten.

**Recorreguts en una taula SQLite**

Una vegada s’han transformat les dades d’un fitxer a taula SQLite, es recorrent les files de la taula per extreure’n les dades. Amb les dades es crea un graf complet2. Cada element no repetit de la taula representarà un node del graf. I cada node estarà enllaçat amb la resta de nodes. Una aresta discontínua representa un enllaç.

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

j

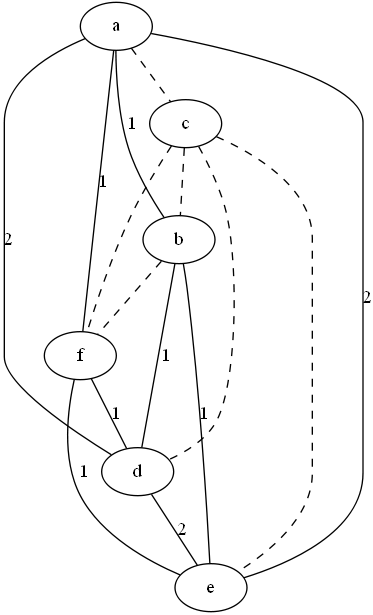


i

**Figura 1**. Exemple de taula SQlite.

Sent i, un índex que recórrer les columnes de la figura 1 i j, un índex que recórrer les files de la figura 1, llavors per a cada parella de valors (i, j), i ≠ j, existirà un enllaç entre els elements de la parella. Una aresta contínua representa un enllaç. El graf resultant s’anomena graf de Gaifman [R] 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.

Grafs

Además del Título y los nombres y datos de los Autores la primera página contiene el Resumen del documento. Éste es un elemento bastante importante que da una primera visión rápida de lo que trata su documento y las nuevas contribuciones que propone. Por esa razón debe ser escrito con extremo cuidado para contener toda la información necesaria y relevante de la forma más clara y concisa posible. Debido a su importancia en la estructura global del documento, el Resumen aparece en la parte superior de la primera columna y no debe ocupar más de 15 líneas de texto. Esperamos que haga un esfuerzo para adherirse a esta regla.

Después del Resumen, vemos el desarrollo de varias secciones en las que está organizado el documento. Primero, la Introducción, donde normalmente exponemos el motivo principal y el trabajo de fondo en relación con su propio trabajo así como las contribuciones que son propuestas y la forma de presentarlas. Es normalmente aquí donde encontramos el cuerpo principal de las referencias al trabajo presentado. Éstas son, por supuesto, extremadamente importantes y deben ser seleccionadas cuidadosamente entre las publicaciones del mundo científico relacionadas con su trabajo. El modo sugerido de llamar a una referencia en el cuerpo principal del texto es simplemente insertar entre corchetes un número de secuencia cada vez que se indica una nueva referencia. Por ejemplo, [1], [2], y así. La identidad completa de estas referencias, incluyendo autores, título, medios de publicación y datos específicos de la misma (números de páginas, fecha de publicación) pueden encontrarse al final del documento. Es muy fácil (y a menudo ocurre) confundir números de páginas e incluso fechas de publicación, lo que puede causar una frustración significativa cuando un lector interesado intente consultar la referencia. Por favor, sea tan cuidadoso como pueda para asegurar que la información concerniente a todas las referencias sea completa y fiable. Ésta es obviamente una parte muy importante y relevante de su propio trabajo y no debe ser pasada por alto.

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.



Fig. 1. Ejemplo de pie de figura

6. Acerca de las tablas

Las pautas para editar y situar las tablas son muy simples. Sólo hay que escribirlas como texto y situarlas como figuras. Se puede usar negrita y diferentes tamaños (desde Times New Roman 12 hasta Times New Roman 10) para diferenciar la información relevante que contienen, como está indicado debajo.

Respecto a la anchura de las tablas, las pautas a seguir serán las mismas que en el caso de las figuras.

|  |  |  |  |
| --- | --- | --- | --- |
| 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] M. Autin, M. Biey, M. Hasler, “Order of discrete time nonlinear systems determined from input-output signals”, *Proc. IEEE Int. Symp. Circ. Syst.*, San Diego, 1992, pp. 296-299

[2] L. Ljung, *System Identification-Theory for the User*, Prentice Hall, 1987

[3] S.K. Rao, T. Kailath, “Orthogonal digital filters for VLSI implementation”, *IEEE Trans. Circuits Syst.*, CAS-31, 1984, pp. 933-945