



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

TEORÍA DE GRÁFICAS Y SU APLICACIÓN
PARA RESOLVER PROBLEMAS
PRÁCTICOS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICA

PRESENTA:

MITZI DAYANA MONTIEL GRACIDA

DIRECTORA DE TESIS:

DRA. HORTENSIA GALEANA SÁNCHEZ



Ciudad Universitaria, CD. MX, 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno

Montiel

Gracida

Mitzi Dayana

Universidad Nacional Autónoma de México

Facultad de Ciencias

Matemáticas

312309893

2. Datos de la tutora

Dra.

Galeana

Sánchez

Hortensia

3. Datos del sinodal 1

Dr.

Goldfeder

Ortiz

Ilán Abraham

4. Datos del sinodal 2

Dra.

Sánchez

López

María del Rocío

5. Datos del sinodal 3

Dr.

Benítez

Bobadilla

Germán

6. Datos del sinodal 4

Dra.

Rojas

Monroy

María del Rocío

7. Datos del trabajo escrito

Teoría de gráficas y su aplicación para resolver problemas prácticos

85 p

2021

*A MI ABUELO ARNULFO
MONTIEL Q.E.P.D*

Agradecimientos

Quiero agradecer a la Dra. Hortensia Galeana Sánchez, por sus consejos, constancia, conocimientos que fueron esenciales para culminar este trabajo, por ayudarme a aterrizar mis ideas cuando divagaba y principalmente darme su orientación a pesar de la distancia, sin su apoyo no hubiera sido posible este trabajo. Ha sido un honor haber sido su tesista.

A todos mis Sinodales Ilán Goldfeder, Rocío Sánchez, Germán Benítez y Rocío Rojas por sus valiosas aportaciones, correcciones, ideas, para mejorar mi trabajo, por compartirme sus conocimientos de manera profesional, por su tiempo, dedicación al leer mi trabajo y por ayudarme a enriquecerlo.

Principalmente a mis abuelos, que son parte importante en mi vida, por impulsarme día a día, por su apoyo incondicional. Hoy cuando concluyo mis estudios universitarios les dedico este logro, una meta más conquistada y vamos por más.

A mi familia, en especial a mi mamá por creer en mí, por ser ese motor que impulsa mis sueños y esperanzas, por apoyarme de manera incondicional, estar a mi lado en los días y noches más difíciles durante mis horas de estudio.

A mi animalito confiar en mí, por estar conmigo incondicionalmente, por todos los momentos que hemos pasado juntas y por no dejarme caer. También a las personas que se tomaron su tiempo para escucharme, estar al pendiente de mi tesis y emocionarse tanto como yo.

A la UNAM, especialmente a la Facultad de Ciencias por darme una segunda casa y darme una formación y educación de calidad.

Investigación realizada gracias al Consejo Nacional de Ciencia y Tecnología (CONACyT), proyecto FORDECYT-PRONACES/39570/2020, y al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la Universidad Nacional Autónoma de México (UNAM), proyecto IN102320 Operaciones en gráficas y digráficas III.

Índice general

Introducción	6
1. Antecedentes	8
2. Un primer acercamiento a las gráficas	12
2.1. <u>Conceptos básicos</u>	12
2.2. <u>Caminos, paseos, trayectorias y ciclos</u>	15
2.3. <u>Gráficas isomorfas, subgráficas, conexidad y un poco más</u>	17
2.4. <u>Algunas familias de gráficas</u>	21
3. Aplicación de gráficas en la teoría de redes	27
3.1. <u>Las gráficas ayudaron a los británicos a ser menos olorosos</u>	29
3.1.1. <u>Midiendo distancias</u>	30
3.2. <u>Árboles de expansión mínima</u>	32
3.2.1. <u>Red de alcantarillado en la Ciudad de México</u>	32
3.2.2. <u>Algoritmo de Kruskal</u>	34
3.3. <u>La teoría de gráficas como medio para una mejor organización</u>	39
3.3.1. <u>Definiciones y reglas para construir la red <i>PERT</i></u>	40
3.3.2. <u>Vamos a organizar una fiesta</u>	42
4. Las gráficas y su relación con la sociedad	51
4.1. <u>¿Mis amigos tienen más amigos que yo?</u>	51
4.2. <u>Gráficas versus epidemias</u>	56
5. Gráficas y datos	66
5.1. <u>Bases de datos orientadas a gráficas</u>	67
5.2. <u>Neo4j</u>	70
5.2.1. <u>¿Cuál es el secreto de <i>Neo4j</i>?</u>	72
5.2.2. <u><i>Cypher</i></u>	73
5.3. <u>Recomiéndame una película</u>	75
Conclusiones	82
Bibliografía	83

Introducción

*“No hay rama de la matemática,
por abstracta que sea, que no pueda
aplicarse algún día a los fenómenos
del mundo real”.*

Nikolai Lobachevski

La ciencia es sin duda una parte muy importante de nuestro día a día que muchas veces no nos damos cuenta y estamos aplicando ciencia, desde niños empezamos experimentando, cuestionando, observando, hasta tener una respuesta, y de esta manera se adquiere un nuevo conocimiento. Gracias a la ciencia se ha desarrollado múltiples disciplinas, como lo son las matemáticas que tenemos información desde los primeros registros de hace más de cinco mil años, y surgen como necesidad del hombre de comprender fenómenos que veían, como el día y la noche, los cambios de clima, conforme se les presentaban problemas, buscaban la forma de solucionarlos. En las matemáticas hay diversas ramas como el cálculo, la estadística, álgebra, geometría, etc. Una muy interesante, tal vez aún no tan conocida y que se ha utilizado en diversas áreas, es la **Teoría de gráficas**. Muchas personas confunden las gráficas, con las que se usan en la estadística para analizar datos, o en cálculo para graficar funciones, y es entendible, ya que el mismo nombre se usa para otras áreas. Pero algo que tienen en común y que es con la primera idea intuitiva con la que hay que quedarse es, una gráfica es una visualización de objetos cuantitativos que tienen alguna relación.

Mi principal interés personal para elegir este tema, comenzó cuando tomé el seminario de matemáticas aplicadas, impartido por la Dra. Hortensia. Lo que llamó mi atención fue ver que iba orientado a aplicaciones de la teoría de gráficas, la cual ya había tomado diversos cursos, todos muy teóricos, me agradó ver que había más de 50 posibles aplicaciones de gráficas, el seminario fue mi motivación para investigar y escribir el presente trabajo, me gustaría transmitir a otras personas mi gusto por la aplicación, claro para toda aplicación hay que tener una buena base teórica.

Esta tesis proporciona una gran oportunidad para contribuir al conocimiento de la Teoría de gráficas, los objetivos que me he planteado son: desarrollar un mejor entendimiento acerca de la teoría de gráficas, mediante diversas aplicaciones prácticas y de alguna forma despertar el interés y la curiosidad de las personas que están en el campo de la ciencia. Así como hacer ver que podemos aplicar las matemáticas en situaciones que se presenten en nuestro día a día.

La estructura del presente trabajo se ha dividido en 5 capítulos. En el primer capítulo llamado “*Antecedentes*”, muestra la historia y el origen de la Teoría de gráficas, es importante saber la historia que hubo tras esta rama, para poder apreciar el presente y tener una perspectiva a futuro.

En el capítulo 2 llamado “*Un primer acercamiento a las gráficas*”, se dan las bases teóricas necesarias, para poder entender mejor las aplicaciones planteadas más adelante, cabe señalar que hay muchas más definiciones y teoría, para lo cual si es interés del lector, en la bibliografía encontrará referencias, para estudiar por cuenta propia.

En el capítulo 3 llamado “*Aplicación de gráficas en la Teoría de Redes*”, trata de como las gráficas junto con las redes, a consecuencia de un suceso histórico que se presentó, empezó a tomar fuerza e importancia en la aplicación a partir de algoritmos, los cuales propuse para crear una red de alcantarillado en la Ciudad de México y otra aplicación para organizar una fiesta.

En el capítulo 4 llamado “*Las gráficas y su relación con la sociedad*”, trata de como se usan para poder hacer modelos, y entender la manera en que interactuamos con otras personas, esto nos ha llevado a pensar ¿Mis amigos tienen más amigos que yo? lo cual es un fenómeno muy común y que tiene una explicación, además de como se puede aplicar para frenar una epidemia.

En el capítulo 5 llamado “*Gráficas y datos*”, se trata sobre la relación que hay entre la parte de computación y las gráficas, a partir de las bases de datos que de hecho los datos son algo que hoy en día están creciendo a gran escala, han cobrado gran importancia, el ver como funcionan, así como la manera en que se organizan, interpretan y analizan, para ofrecernos servicios que ocupamos día a día, ya sea en redes sociales, plataformas de video y música, compras en línea, etc. Un ejemplo propuesto es hacer un sistema de recomendación de películas. Por último, se mencionan las conclusiones a las que se llegó, con toda esta investigación y algunas referencias bibliográficas consultadas, para más información si es de su interés.

Capítulo 1

Antecedentes

Bien dicen que mirar al pasado, nos ayuda a entender el presente y a la inversa, observar nuestro presente y analizarlo, nos hace ver el alcance y la importancia, que tuvieron determinadas ideas, problemas, situaciones y teorías, formadas tiempo atrás. La Teoría de gráficas no es la excepción, así que se empezará por ver sus orígenes.

La historia dice que en la ciudad de Königsberg en Prusia (ahora Kaliningrado, Rusia), situada junto al río Pregel, en la costa del Mar Báltico. Este río atravesaba la ciudad formando dos islas, a su paso por la ciudad, y para no perder comunicación entre los pobladores la ciudad contaba con siete puentes. Los habitantes eran personas muy tranquilas y unidas, gozaban a menudo de plantear retos, o desafíos para entretenerse, entre los cuales surgió: ***¿Es posible recorrer todas las zonas de la ciudad, atravesando todos los puentes, una y solo una vez cada uno de ellos y regresar al mismo punto de inicio?***

En el año de 1736, gracias a un grupo de jóvenes de la misma ciudad, se enteraron de que había un gran matemático, que había llegado, su nombre era *Leonhard Euler*, el mismo relata lo siguiente:

“Me pusieron un problema sobre una isla en la ciudad de Königsberg, que se encuentra rodeada por un río al que cruzan 7 puentes: me preguntaron si alguien podría dar un paseo que cruzara por todos los puentes pasando por cada uno solo una vez. Fui informado que hasta ahora nadie había mostrado que esto fuera posible, ni demostrado que no lo fuera. La pregunta es banal, pero me pareció digna de atención porque ni la geometría, ni el álgebra, ni aun el arte de contar fueron suficientes para resolverla. En vista de esto, se me ocurrió preguntarme si pertenecería a la geometría de la posición tan buscada alguna vez por Leibniz. Así que después de alguna deliberación, obtuve una regla, simple pero firme, con cuya ayuda uno puede

decidir inmediatamente para todos los ejemplos de este tipo, con cualquier número de puentes arreglados de cualquier modo, si tal paseo es posible o no.” (Leonard Euler, 1736) [27].

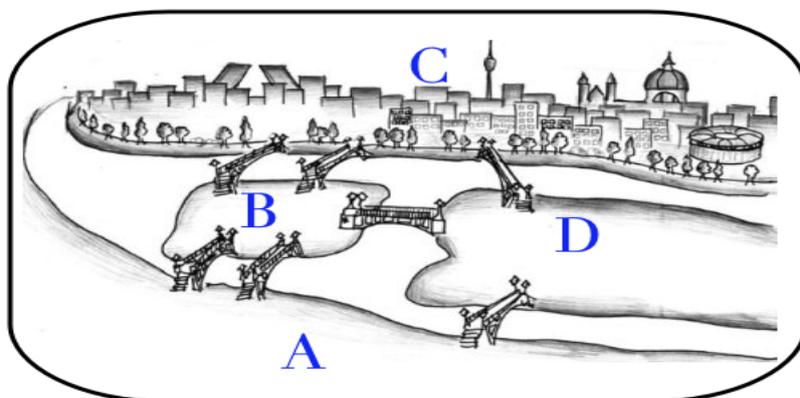


Figura 1.1: La ciudad de Königsberg y sus puentes.

Lo que hizo Euler, para empezar a resolverlo, fue pasar el problema a un esquema más sencillo de visualizar, justo como el que aparece en la figura 1.2. Donde los distintos territorios en los que los puentes dividieron la ciudad, se convirtieron en puntos, es decir, en *vértices*; y los puentes se convirtieron en líneas, lo que llamamos *aristas*. Euler demostró que no era posible, ya que el número de líneas que inciden en cada punto no es par (condición necesaria para entrar y salir de cada punto, y para regresar al punto de partida, por caminos distintos en todo momento). Los estudios de Euler, fueron el detonante de la Teoría de Gráficas, convirtiendo un reto simple de entretenimiento en una disciplina científica.

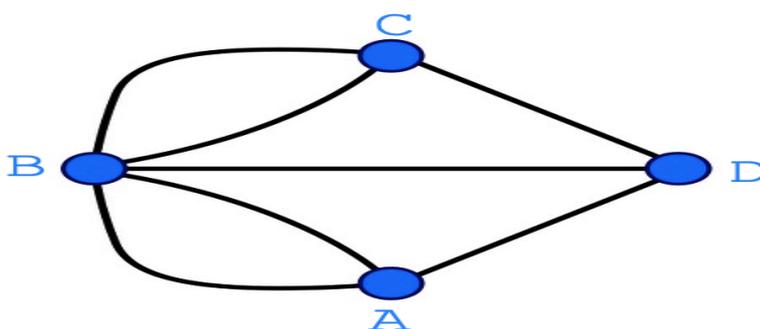


Figura 1.2: Representación de la ciudad de Königsberg y sus puentes.

Otro problema muy conocido y de gran importancia para la Teoría de Gráficas, fue el problema de los cuatro colores. Se remonta al año de 1852 cuando *Francis Guthrie*, matemático de la época, mientras intentaba colorear el mapa de los condados de Inglaterra, notó que cuatro colores eran suficientes, para que se cumpliera la condición de que dos regiones con frontera común, tuvieran colores distintos (si dos regiones se tocan en un único punto se entiende que no tienen frontera común). Francis le preguntó a su hermano *Frederick Guthrie*, si era cierto que cualquier mapa se puede colorear usando cuatro colores, a su vez Frederick se lo planteó a *Augustus de Morgan* (profesor suyo en un curso de matemáticas en aquel momento), que aunque no supo responderle, se encargó de difundir el asunto entre otros matemáticos [35]. En 1878, *Arthur Cayley* lo presenta formalmente a la *London Mathematical Society* y así el problema queda abierto, de la siguiente forma:

“Todo mapa plano puede colorearse con máximo cuatro colores, con la condición de que regiones con frontera común tengan colores distintos.” [23]



Figura 1.3: Mapa de México coloreado con 4 colores.

El intercambio de ideas entre dos ramas diferentes de la ciencia es a menudo muy beneficioso para ambas. Las ideas e investigaciones en el campo de la química fueron inspiración para parte de la terminología que ahora es estándar en teoría de gráficas, incluida la palabra “gráfica”. Todo comenzó con el avance de la teoría química del siglo XIX, en 1850 se estaba volviendo evidente que los compuestos químicos se forman a partir de ciertos elementos, que además se combinan en ciertas proporciones fijas, dando como resultado fórmulas químicas. Los investigadores de la época empezaron a notar, que podían representar átomos y enlaces de manera visual, mediante círculos y líneas. Esto fue de mucha ayuda, para no quedarse solamente con una imagen mental, dándole el nombre de: notación gráfica.

Aunque, estas representaciones de compuestos químicos son puramente estáticas, fueron de gran ayuda para poder tener una nueva forma de estudiar la química, y así aportar nuevas ideas. *James Joseph Sylvester* concibió la noción, de que existe una relación entre química y álgebra, ejemplificado por el uso de notación gráfica. Publicó una nota sobre sus ideas en febrero de 1878, en la revista científica *Nature*. En esta nota, la palabra “gráfica” se usa por primera vez; se deriva claramente de la “notación gráfica” química. Entonces, el crédito por el uso de este término debe atribuirse a Sylvester [2].

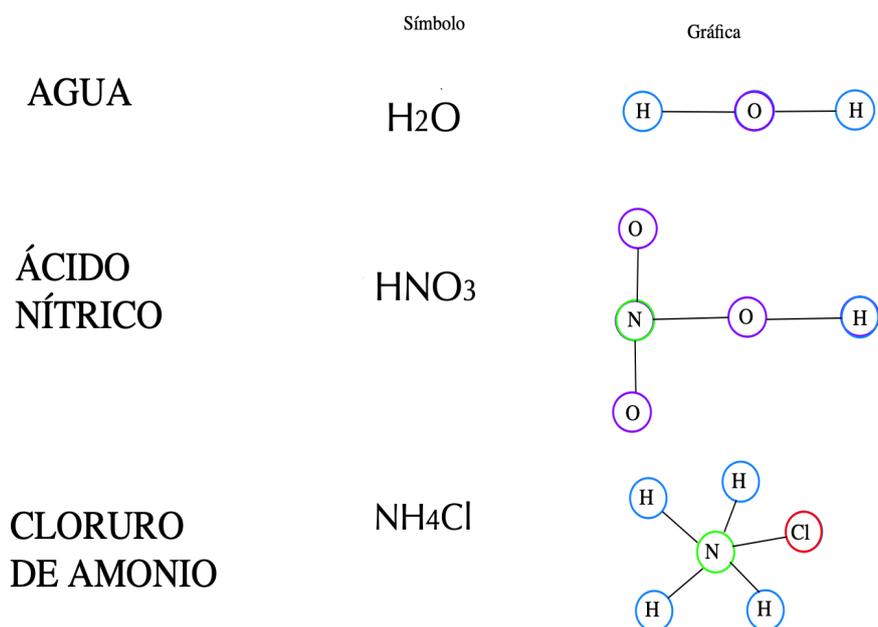


Figura 1.4: Representación gráfica de algunos compuestos químicos.

Capítulo 2

Un primer acercamiento a las gráficas

2.1. Conceptos básicos

Se define una **gráfica** G como una pareja ordenada $(V(G), A(G))$, donde $V(G)$ es un conjunto de elementos finito y no vacío, donde los elementos reciben el nombre de *vértices* (puntos) y $A(G)$ es el conjunto de pares no ordenados de distintos elementos de $V(G)$, también llamados *aristas* (línea que une dos puntos cualesquiera). Puesto que la cardinalidad de los vértices y aristas es finita, se está hablando de una gráfica finita, por lo tanto se dice que una gráfica G es de **orden** p y **tamaño** q si $|V(G)| = p$ y $|A(G)| = q$.

La definición de gráfica se puede ampliar si se permiten **lazos**, que son aristas cuyos extremos están en el mismo vértice, es decir son de la forma vv . Si $a = xy$ es una arista, los vértices x, y se les llama **extremos**. También si se permiten **aristas múltiples o paralelas**, que son aristas que tienen los mismos extremos. En este caso se hablaría de una **multigráfica** [15]. En caso contrario será una **gráfica simple**, en la que a lo más una arista une dos vértices cualesquiera. Por el momento lo que se expondrá a continuación es para gráficas simples a menos que se especifique lo contrario.



Figura 2.1: Ejemplo de una multigráfica y una gráfica simple.

Sean $\mu, \nu \in V(G)$, se dice que μ es **adyacente** a ν si $a = \mu\nu \in A(G)$, es decir, son adyacentes si una arista los une.

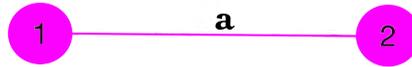


Figura 2.2: Gráfica donde los vértices 1 y 2 son adyacentes por la arista a.

Por otro lado, sean $a_1, a_2 \in A(G)$ se dice que a_1 es **adyacente** a a_2 si $a_1 = \eta\mu$ y $a_2 = \mu\nu$ con $\eta, \mu, \nu \in V(G)$, es decir que dos aristas son adyacentes si tienen un vértice en común.

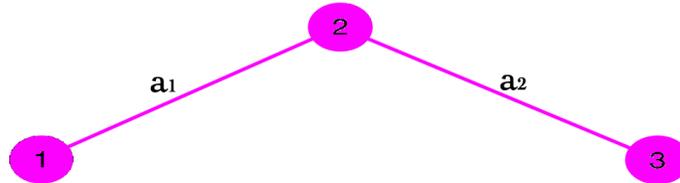


Figura 2.3: Gráfica donde las aristas a_1, a_2 son adyacentes.

El **grado** de un vértice v_i en una gráfica G es el número de aristas que inciden en el vértice v_i , y se denota por $\delta(v_i)$. En la figura 2.4, se tiene que $\delta(v_1) = 4$, $\delta(v_2) = 1$, $\delta(v_3) = 1$, $\delta(v_4) = 1$ y $\delta(v_5) = 1$.

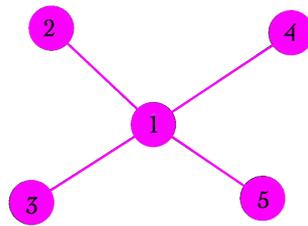


Figura 2.4: Ejemplo de gráfica.

En el caso de un vértice con grado 0, se dice que es un **vértice aislado**, y los vértices de grado 1 se les llama **hojas** o finales. Por definición, todo vértice puede

ser adyacente a lo más a todos los vértices menos a él mismo, Por lo tanto, de la definición de gráfica tenemos la siguiente desigualdad para toda gráfica G .

$$0 \leq \delta(v) \leq p - 1 \quad \forall v \in V(G)$$

Considerando la gráfica de la figura 2.5 ya que que $p = 7$ y por la desigualdad mencionada anteriormente, se tiene que todos los vértices tendrán su grado acotado entre 0 y 6. Entonces se tiene que $\delta(v_1) = 2$, $\delta(v_2) = 3$, $\delta(v_3) = 3$, $\delta(v_4) = 3$, $\delta(v_5) = 2$, $\delta(v_6) = 1$ y $\delta(v_7) = 0$.

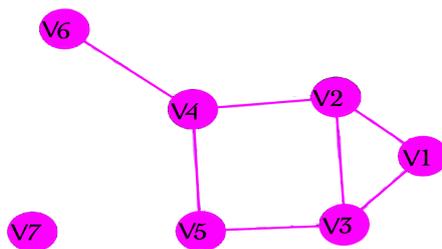


Figura 2.5: Ejemplo de gráfica con hoja y un vértice aislado.

Se define el **grado mínimo** de G como $\delta(G) = \min\{ \delta(v) \mid v \in V(G) \}$ y el **grado máximo** como $\Delta(G) = \max\{ \delta(v) \mid v \in V(G) \}$. Por lo tanto, reescribiendo la desigualdad antes mencionada, se puede afirmar que la siguiente desigualdad siempre se cumple:

$$0 \leq \delta(G) \leq \delta(v) \leq \Delta(G) \leq p - 1 \quad \forall v \in V(G)$$

Si $\delta(G) = \Delta(G) = r$, se dice que la gráfica es **r -regular**. En la figura 2.6 $\delta(G_1) = 2$, $\Delta(G_1) = 4$ y G_2 es 3-regular.

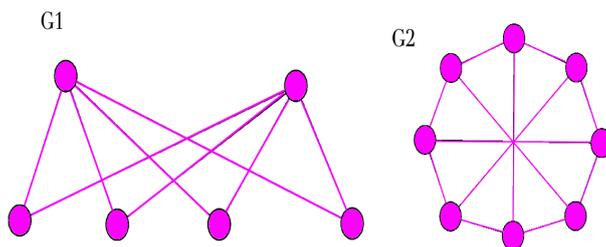


Figura 2.6: $\delta(G_1) = 2$, $\Delta(G_1) = 4$ y $\delta(G_2) = \Delta(G_2) = 3$.

2.2. Caminos, paseos, trayectorias y ciclos

En el estudio de las gráficas, es normal empezar a pensar en las posibles maneras de recorrerlas, las siguientes definiciones ayudarán a comprender mejor estos recorridos. Un **camino** C es una sucesión de vértices $C = (v_0, v_1, \dots, v_n)$, con $v_{i-1} v_i \in A(G)$ para todo $i \in \{1, 2, \dots, n\}$, pero de igual manera se puede definir como una sucesión alternada de vértices y aristas de la siguiente forma: $C = (v_0, a_1, v_1, a_2, v_2, a_3, \dots, v_{n-1}, a_n, v_n)$, tal que $a_i = v_{i-1} v_i$ con $i \in \{1, 2, \dots, n\}$. Cabe mencionar que en el camino se pueden repetir vértices y aristas, al vértice $u = v_0$ del camino C le llamamos vértice inicial y al vértice $v = v_n$ le llamamos vértice final; así, se dice que C es un **uv-camino**. Por otra parte C es un **camino cerrado** si $v_0 = v_n$, quiere decir que empieza y termina en el mismo vértice, por lo tanto es correcto afirmar que si un camino no es cerrado se le llama camino abierto.

A continuación, en la figura 2.7 se tiene dos veces a la gráfica G , donde la sucesión $C = (A, a_1, B, a_5, E, a_4, D, a_6, C)$ el cual es un AC -camino y por otra parte, se tiene la sucesión $C1 = (B, a_2, C, a_5, E, a_4, D, a_6, B)$ que es un camino cerrado.

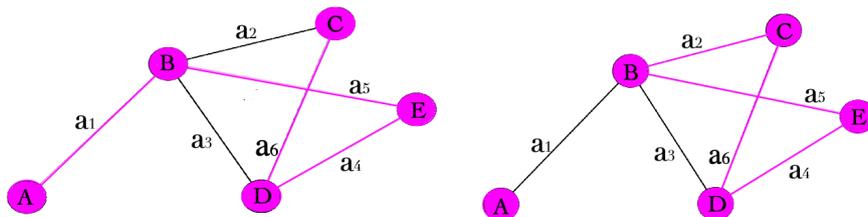


Figura 2.7: Gráficas con un camino simple y un camino cerrado.

Si un camino no repite aristas se llama **paseo**, un camino que no repite vértices se llama **trayectoria**, de manera análoga tenemos la notación de $v_0 v_n$ -paseos, paseos cerrados y las $v_0 v_n$ -trayectorias. Un **ciclo** es un camino cerrado que no repite vértices salvo el primero y el último, con al menos 3 vértices.

La **longitud de un camino** $C = (v_0, a_1, v_1, a_2, v_2, a_3, \dots, v_{n-1}, a_n, v_n)$, denotada por $l(C)$ se define como el número n . Si $l(C) = 0$, se dice que el **camino** es **trivial**. Observe que en un ciclo el número de vértices es igual al número de aristas ya que no repite vértices, entonces un ciclo con l vértices se denota por \mathcal{C}_l , tendrá longitud al menos 3.

El **cuello** de la gráfica G se define como el mínimo de las longitudes de los ciclos de G , denotado por $g(G)$. Si la gráfica no contiene ciclos, se dice que es **acíclica** y que su cuello es ∞ , en el caso de una multigráfica se podría encontrar un cuello de longitud 1, que corresponde a un lazo.

A continuación, se muestran algunos ejemplos a partir de la figura 2.8

- V_1V_8 camino de longitud 5.
 - $C = (V_1, V_4, V_7, V_5, V_6, V_8)$
- Un paseo que es una trayectoria.
 - $P = (V_6, V_2, V_3, V_8, V_7, V_5, V_1, V_4)$
- Trayectoria de máxima longitud.
 - $T = (V_1, V_4, V_3, V_2, V_6, V_5, V_7, V_8), l(T) = 7$
- Paseo cerrado de longitud 8.
 - $P_1 = (V_2, V_3, V_4, V_1, V_5, V_7, V_8, V_6, V_2)$
- Cuello de la gráfica G .
 - $g(G) = 4$
- Un ciclo en G .
 - $C_4 = (V_1, V_2, V_3, V_4, V_1)$

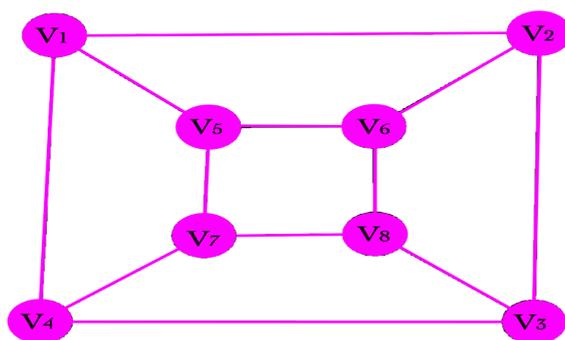


Figura 2.8: Gráfica usada para resolver los ejemplos planteados.

2.3. Gráficas isomorfas, subgráficas, conexidad y un poco más

Siguiendo con el estudio y acercamiento hacia las gráficas, una pregunta interesante es: ¿qué características deben cumplir dos gráficas para que se consideren iguales en la teoría de gráficas?

Por un lado, se tiene que pensar que un vértice de la gráfica 1 tendrá su semejante en la gráfica 2 y así con todos los demás vértices, por lo tanto se necesita una función biyectiva entre los conjuntos de vértices de las gráficas. Asimismo, nos gustaría conservar las aristas, en el sentido de que si dos vértices son adyacentes en una de las gráficas sus imágenes en la otra gráfica también deben ser adyacentes.

Entonces dadas dos gráficas G_1 y G_2 decimos que G_1 es **isomorfa** a G_2 si existe una función *biyectiva* $\phi : V(G_1) \rightarrow V(G_2)$ que cumpla las siguientes propiedades:

- Si $uv \in A(G_1)$, entonces $\phi(u)\phi(v) \in A(G_2)$.
- Si $uv \notin A(G_1)$, entonces $\phi(u)\phi(v) \notin A(G_2)$.

Estas propiedades son equivalentes a decir que $uv \in A(G_1) \Leftrightarrow \phi(u)\phi(v) \in A(G_2)$.

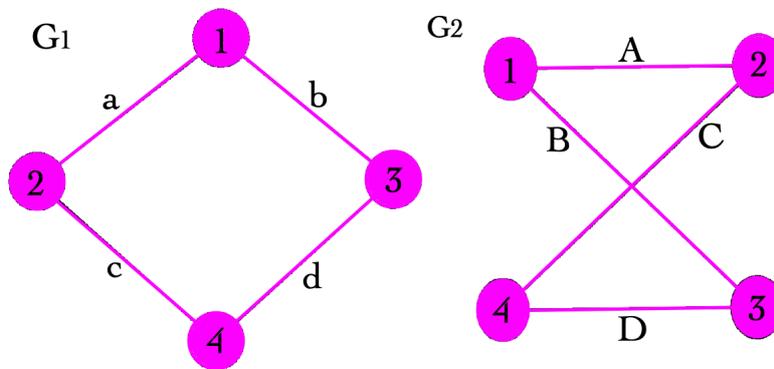


Figura 2.9: Ejemplo donde G_1 es una gráfica isomorfa a G_2 .

Para probar que dos gráficas G y H son isomorfas basta con dar una función biyectiva y comprobar que efectivamente cumple las condiciones anteriores.

En la figura 2.10, al aplicar ϕ de $\{1,2,3,4,5\}$ a $\{a,b,c,d,e\}$ definida por $\phi(1)=a$, $\phi(2)=b$, $\phi(3)=c$, $\phi(4)=d$, $\phi(5)=e$, puesto que al corresponder uno a uno los vértices de G y H , cumpliendo la propiedad de adyacencia, se concluye que son gráficas isomorfas.

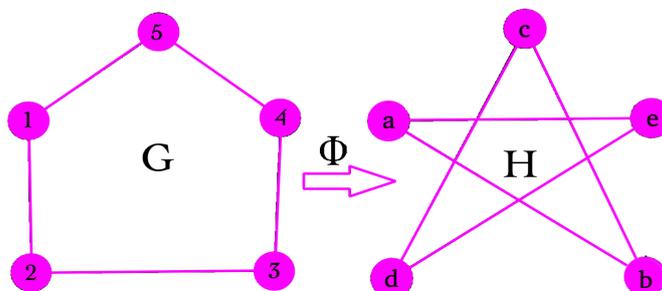


Figura 2.10: La gráfica G isomorfa a la gráfica H bajo la función ϕ .

Las gráficas isomorfas cumplen las siguientes propiedades:

- G y H tienen el mismo orden.
- G y H tienen el mismo tamaño.
- G posee un ciclo si y solo si H posee un ciclo.

Una gráfica S es **subgráfica** de G si el conjunto de los vértices de S es un subconjunto de los vértices de G y el conjunto de las aristas de S es subconjunto de las aristas de G , es decir $V(S) \subseteq V(G)$ y $A(S) \subseteq A(G)$. Todas las aristas y vértices de G pueden no estar presentes en S , pero si un vértice o arista esta presente en S forzosamente debe estar en G . Claramente toda gráfica G es subgráfica de sí misma.

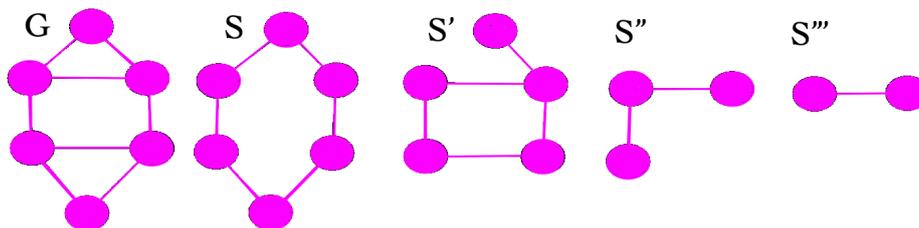


Figura 2.11: Ejemplo de algunas posibles subgráficas que se pueden generar a partir de la gráfica G .

Sea G una gráfica y S un subconjunto de los vértices de G , la **subgráfica inducida** por vértices de S en G , denotada como $G[S]$, tiene por conjunto de vértices a S y dos vértices en $G[S]$ son adyacentes si y solo si son adyacentes en G . Si H es subgráfica de G , decimos que H es *inducida* si $H = G[V(H)]$ [13].

Para evitar confusiones, también decimos que $G[S]$ es una subgráfica inducida por aristas, si S es su conjunto de aristas y conserva todos los vértices originales, es decir, su conjunto de vértices es $V(G)$.

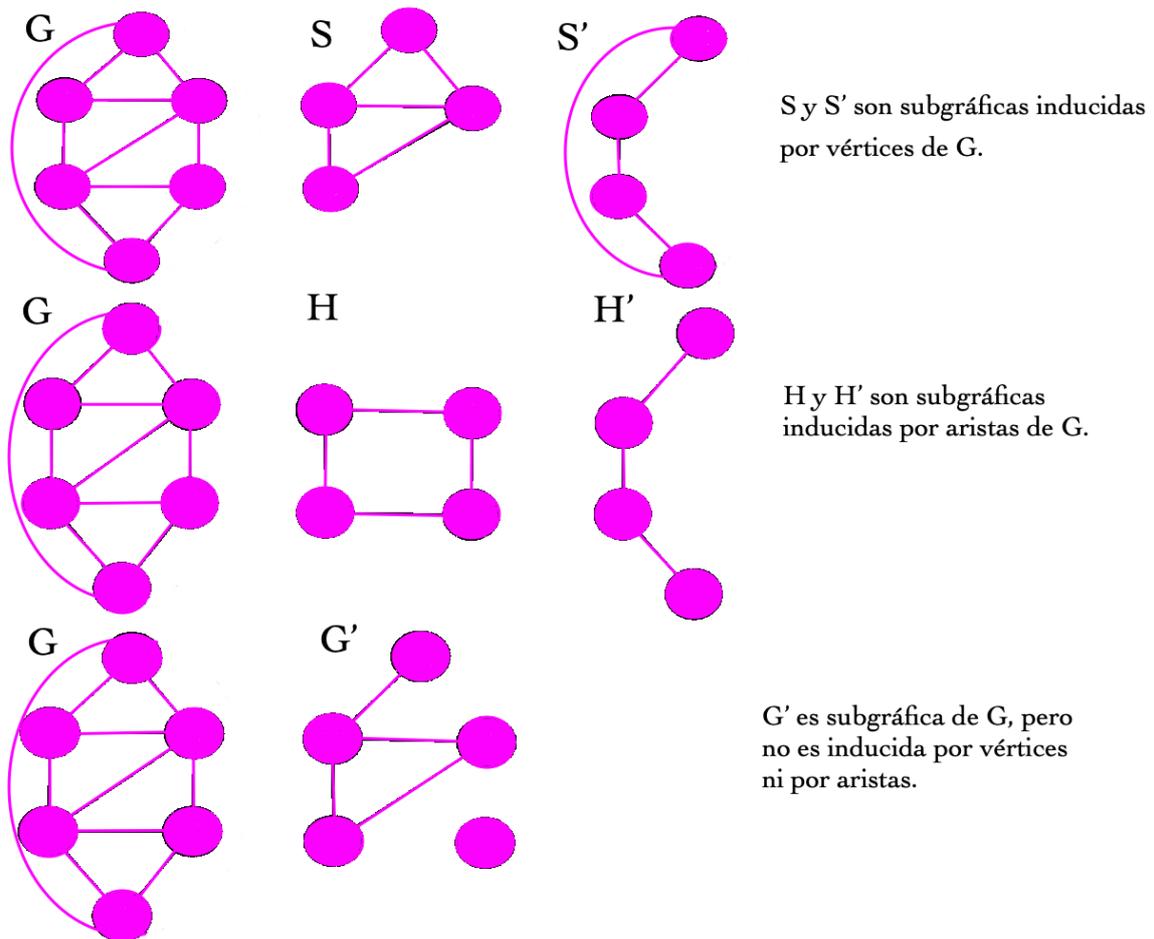


Figura 2.12: Ejemplos de algunas gráficas inducidas por vértices y por aristas a partir de G .

Se dice que si H es subgráfica de G , diremos que H es una **subgráfica generadora** de G si $V(H) = V(G)$.

Siguiendo con las definiciones básicas, se dice que una gráfica G es **conexa** si para cualquier par de vértices u y v de G , existe un uv -camino, de otra manera decimos que G es **inconexa** o que está **desconectada**, intuitivamente, una gráfica conexa está formada por un solo pedazo. Si una gráfica tiene más de un solo pedazo, estos pedazos de la gráfica se les llama **componentes conexas**, una componente conexa de G es una subgráfica de G máxima por contención con la propiedad de ser conexa.

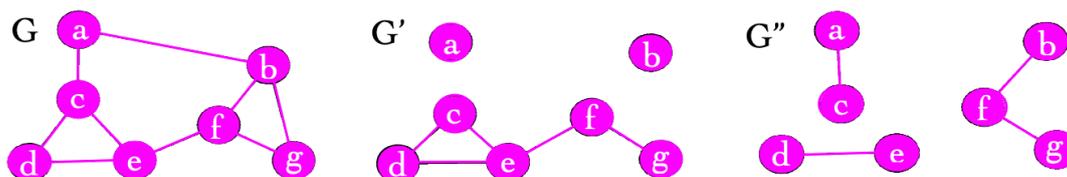


Figura 2.13: Ejemplos de gráficas conexas e inconexas y sus componentes.

En la figura 2.13 se tiene una gráfica G , la cual es conexa, ya que siempre hay un camino entre cualesquiera dos vértices. La gráfica G' es inconexa ya que los vértices a y b están aislados, por tanto no hay forma de llegar a ellos por un camino. Por último la gráfica G'' es una gráfica inconexa, ya que no hay un cd -camino, ce -camino o ef -camino, además G'' tiene tres componentes conexas, la primera componente es la subgráfica inducida por el conjunto de vértices $\{a, c\}$, la segunda componente es la subgráfica inducida por el conjunto de vértices $\{d, e\}$ y la tercer componente es la subgráfica inducida por el conjunto de vértices $\{b, f, g\}$.

2.4. Algunas familias de gráficas

A continuación se hablará de algunas familias de gráficas, las cuales se usan con frecuencia en ejemplos y se utilizarán en aplicaciones que más adelante veremos. Una **gráfica completa** es aquella en la que todo par de vértices son adyacentes. Si G es una gráfica completa de orden n , usualmente se denota por K_n donde su orden es exactamente $\frac{n(n-1)}{2}$ aristas.

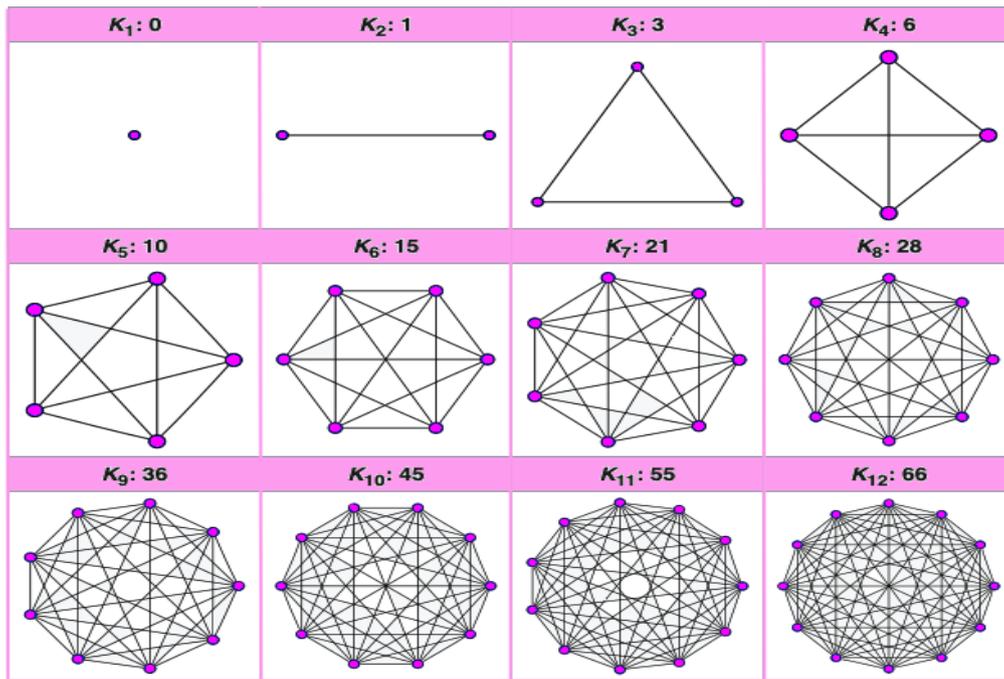


Figura 2.14: Gráficas completas de $n=1,2,3,\dots,12$ imagen cortesía de [40].

Otra familia importante son las **gráficas bipartitas**, sea G una gráfica, si el conjunto de vértices $V(G)$ puede ser partido en dos conjuntos disjuntos, no vacíos, V_1 y V_2 , llamados subconjuntos de bipartición de G , tal que toda arista de G tiene un extremo en V_1 y otro en V_2 . Por tanto, una gráfica bipartita se puede expresar como $G = (V_1 \cup V_2, A(G))$. Es importante señalar que la definición no pide forzosamente que existan aristas, solo si llega a tener, sus extremos estén en conjuntos diferentes.

En la figura 2.15 se tiene un ejemplo de una gráfica G donde $V(G) = \{M, A, T, E, M, \acute{A}, T, I, C, A, S\} \cup \{A, S, T, R, O, N, O, M, \acute{I}, A\}$, notemos cada uno de los vértices de V_1 es una letra que juntos forman una palabra, de igual forma pasa en V_2 , además las adyacencias se dan entre los caracteres que son iguales en conjuntos diferentes, como no hay adyacencias entre elementos del mismo conjunto la gráfica es bipartita.

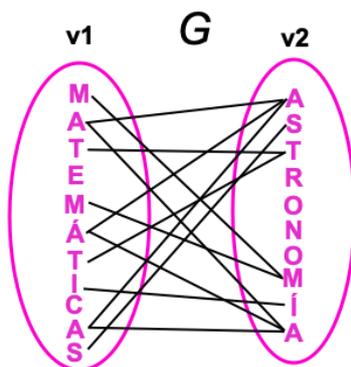


Figura 2.15: Ejemplo de gráfica bipartita con palabras.

Una forma para saber si una gráfica G es bipartita, es colorear a los vértices, con la condición de que solo se pueden usar dos colores y que dos vértices no pueden tener el mismo color si están unidos por una arista. Un ejemplo está en la figura 2.16, donde la gráfica G_1 no es bipartita, puesto que un vértice al ser adyacente con otros dos vértices de distinto color rompería la condición mencionada antes, en cambio la gráfica G_2 cumple las condiciones de coloración y adyacencia por tanto es bipartita.

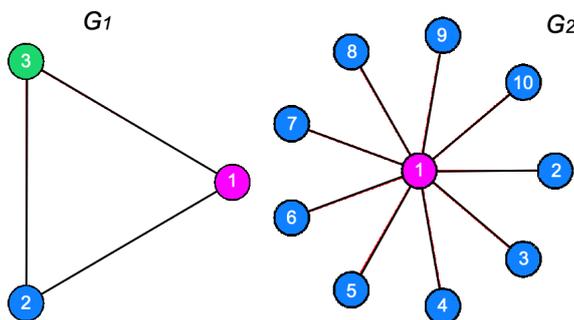


Figura 2.16: Gráfica completa K_3 y gráfica estrella de 10 vértices.

Una gráfica G es **bipartita completa** si todo vértice de V_1 es adyacente a todo vértice de V_2 . Si $|V_1| = n$ y $|V_2| = m$ se denota a G como $K_{n,m}$, además se cumple que $|V(K_{n,m})| = n + m$ y $|A(K_{n,m})| = nm$ [15].

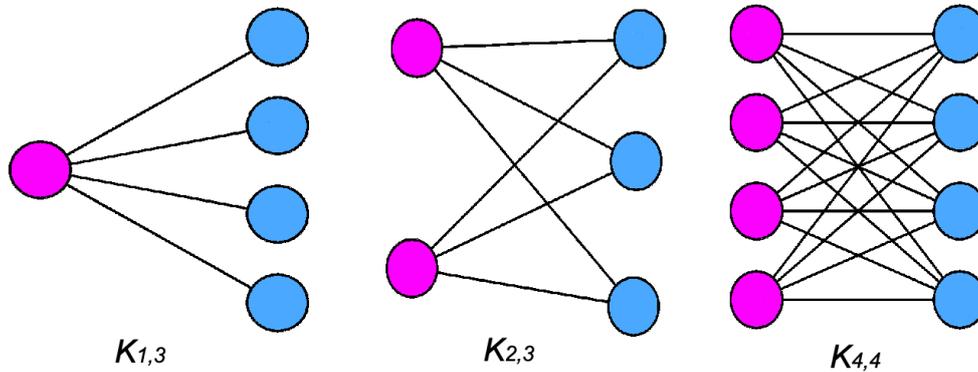


Figura 2.17: Gráficas bipartitas completas.

En la figura 2.17 se muestran algunos ejemplos, donde es fácil comprobar que la suma de las cardinalidades de los dos conjuntos de la partición, da el total de vértices de la gráfica, y que la multiplicación de vértices de los conjuntos da el total de aristas de dicha gráfica.

Otra familia de gráficas muy importante y con una variedad de aplicaciones son los árboles. Una gráfica G es **árbol** si es conexa y acíclica, es decir, no tiene ciclos.

¿Cómo saber que gráficas son árboles?, por ejemplo, las gráficas completas de uno y dos vértices K_1 y K_2 , son árboles. Pero K_3 no es un árbol porque posee un ciclo. Entonces toda gráfica completa de orden mayor o igual a tres, no será árbol. Se podría incluso pensar que si G tiene muchas aristas (con respecto al número de vértices que posee), entonces es poco probable que G sea un árbol [14].

Su importancia reside en que los árboles son gráficas que conectan todos los vértices con el menor número posible de aristas. Los árboles tienen las siguientes propiedades:

- G es conexo y acíclico.
- Al añadir alguna arista se forma un único ciclo.
- G es conexo y al quitar una arista se vuelve inconexo.

- Dos vértices cualesquiera de G están conectados por una única trayectoria.

Las demostraciones de estas propiedades las puede consultar en [21]. Un **bosque** es una gráfica sin ciclos que puede o no, ser conexa, en caso de no ser conexa sus componentes conexas son árboles.

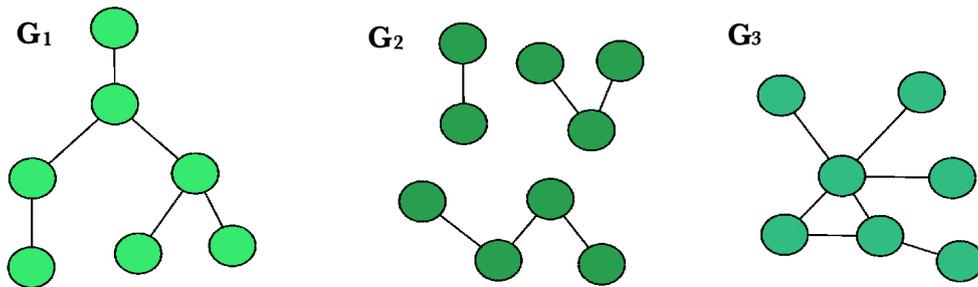


Figura 2.18: G_1 es un árbol, G_2 es un bosque y G_3 que no es árbol ni bosque.

Sea G una gráfica simple, un **árbol generador** de G es una subgráfica conexa que contiene todos los vértices de G y es un árbol.

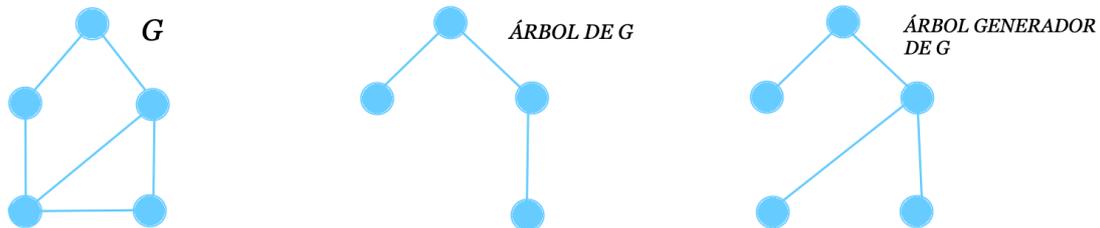


Figura 2.19: Gráfica G con sus respectiva gráfica árbol y árbol generador.

Hasta el momento se han mencionado gráficas simples, pero ¿qué pasaría si estas gráficas en lugar de aristas simples, tuvieran una dirección? Se define como **digráfica** a la pareja ordenada $D = (V(D), F(D))$ donde $V(D)$ es un conjunto finito, no vacío, de objetos, llamados vértices y $F(D)$ es un conjunto de pares ordenados de vértices de D llamados flechas. Sea $a = (u, v)$ una flecha, se dirá que u y v son los extremos de a , también se dice que u es vértice inicial y v es vértice final. Dos flechas son adyacentes si comparten alguno de sus extremos.

Como en el caso de gráficas simples y multigráficas, de igual forma se tiene *digráficas simples y multidigráficas*.

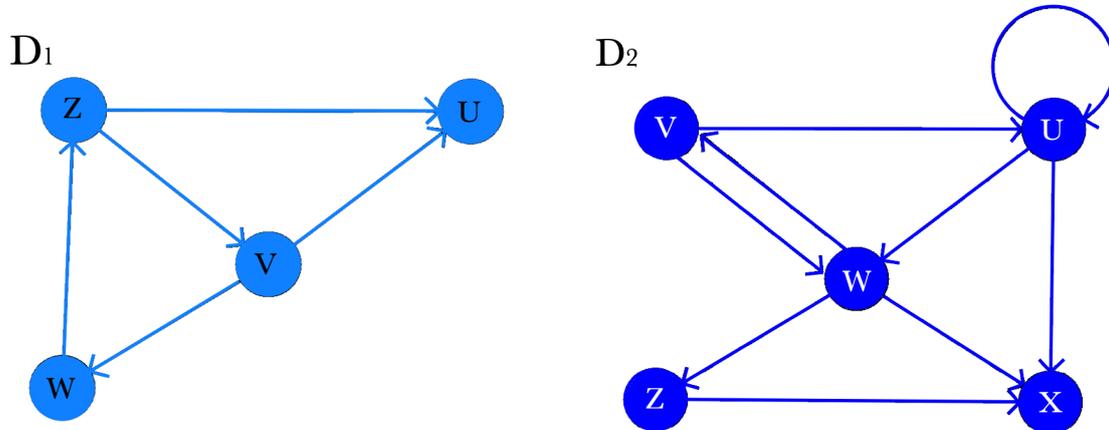


Figura 2.20: Ejemplos donde D_1 es una digráfica y D_2 una multidigráfica.

Al igual que las gráficas simples, las digráficas también tienen sus formas para recorrerlas. Un **camino dirigido** W en una digráfica D es una sucesión de vértices $(u_0, u_1, u_2, \dots, u_n)$ tal que $(u_i, u_{i+1}) \in F(D)$ con $i \in \{0, 1, \dots, n-1\}$. Un **paseo dirigido** P en una digráfica D es un camino dirigido en el que no se repiten flechas. Una **trayectoria dirigida** T en una digráfica D es un camino dirigido el cual no repite vértices.

Recordando que en gráficas se tienen caminos cerrados, paseos cerrados y ciclos, de igual forma se tienen en digráficas, por lo tanto se tienen las siguientes definiciones: Un **camino dirigido cerrado** en una digráfica D es un camino dirigido tal que el primer y el último vértice son iguales. Un **paseo dirigido cerrado** en una digráfica D es un paseo dirigido tal que el primer y último vértice son iguales. Un paseo dirigido cerrado también recibe el nombre de *circuito*. Por último, un **ciclo dirigido** C en una digráfica D es un camino cerrado que no repite vértices, salvo el primero y el último [36].

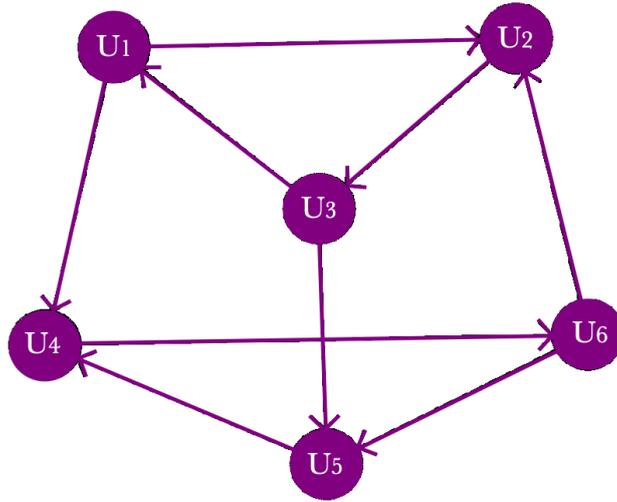


Figura 2.21: Ejemplo de una digráfica D con $V(D) = \{u_1, u_2, u_3, u_4, u_5, u_6\}$ y $F(D) = \{(u_1, u_2), (u_1, u_4), (u_2, u_3), (u_3, u_1), (u_3, u_5), (u_4, u_6), (u_5, u_4), (u_6, u_2), (u_6, u_5)\}$.

En la figura 2.21 tenemos que $W = (u_1, u_4, u_6, u_2, u_3, u_5)$ es un *camino dirigido*, $P = (u_2, u_3, u_1, u_4, u_6)$ es un *paseo dirigido*, $T = (u_3, u_5, u_4, u_6, u_2)$ es una *trayectoria dirigida* y $C = (u_6, u_2, u_3, u_1, u_4, u_6)$ un *ciclo dirigido*.

Más definiciones y conceptos vistos en este capítulo, puede consultarse en [3], [6], [16], [14].

Capítulo 3

Aplicación de gráficas en la teoría de redes

La primera aplicación que se verá es en la teoría de redes, área que se dedica al estudio de gráficas, utilizándolas para representar y modelar situaciones, principalmente de naturaleza combinatoria; hace referencia a la capacidad de organizar elementos que poseen ciertas propiedades y están involucrados en un problema determinado, así como las posibilidades y dimensiones en las que pueden interactuar, es decir, relaciona rutas, cortes, árboles y otros objetos para obtener las soluciones mediante el uso de algoritmos especiales creados para tal fin.

Formalmente se define una **red** como $R = (N, A, f)$, donde N corresponde a los nodos, también llamados vértices o elementos totales de la red, y se denotan por los símbolos $v_1, v_2, v_3, v_4, \dots, v_N$, si un nodo v_i está conectado con otro nodo v_j esta conexión se representa por una pareja ordenada $(v_i, v_j) \in A$ que son las aristas o flechas según sea el caso, se define como $A(R)$ al conjunto de aristas de la red R y $f: A(R) \rightarrow \mathbb{N}$ como una función que va de las aristas de la red a los naturales, a lo que se llamará flujo.

Los problemas de redes pueden clasificarse esencialmente en cinco áreas: rutas más cortas, flujo máximo, árbol de expansión mínima, flujo a costo mínimo, planeación y control de proyectos [17], dependiendo del área de aplicación se utilizará una gráfica con ciertas características.

Entonces se define como una **red dirigida** donde sólo existen flechas es decir, el desplazamiento de un nodo a otro solo es posible en un sentido, como se muestra en la figura 3.1.

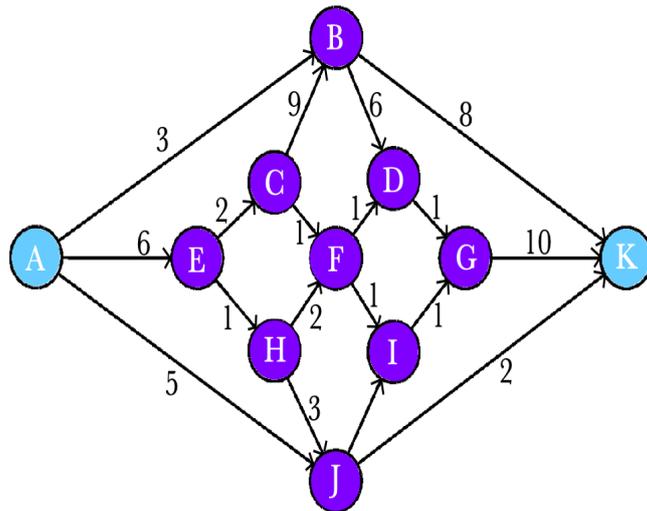


Figura 3.1: Ejemplo de una red dirigida $R = (11, 18)$, donde también se muestra la regla de correspondencia de f .

Por otra parte se define como **red no dirigida** donde existen aristas, es decir, cuando el desplazamiento entre dos nodos es posible en ambos sentidos.

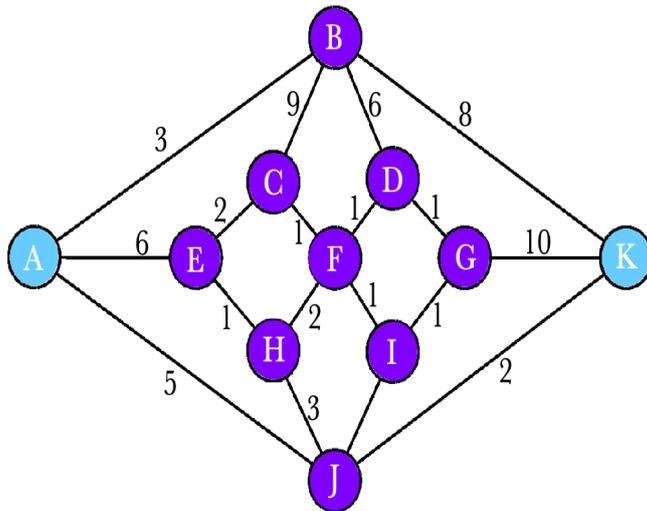


Figura 3.2: Ejemplo de una red no dirigida $R = (11, 18)$, donde también se muestra la regla de correspondencia de f .

3.1. Las gráficas ayudaron a los británicos a ser menos olorosos

Una mañana de julio a mediados del siglo XIX, la gente de Londres se despertó con un hedor repugnante. No podían salir de sus casas sin resultar enfermos. Los más acomodados se ataban los pañuelos con perfume y caminaban cubriéndose permanentemente la cara. Muchos de los pobres abandonaron la ciudad para buscar trabajo en el campo porque simplemente no podían soportarlo. Fue sin duda el incidente más oloroso de la historia británica, mejor conocido como *La gran peste de 1858* [22].

Esto se debía a que el río Támesis estaba lleno de varios siglos de desechos humanos que llegaban a través del sistema de drenaje bastante deteriorado y viejo. Y dadas las altas temperaturas poco usuales del verano y la cantidad de lodo acumulado, formaban junto con los desechos un hedor que alcanzaba todo Londres.



Figura 3.3: La muerte ronda sobre el Támesis, reclamando la vida de los que no pagan por la limpieza del río, durante el Gran Hedor [31].

La corporación de la ciudad de Londres se dio cuenta de que ya era suficiente e invitó a enviar propuestas para el diseño de un nuevo plan de alcantarillado para la ciudad y fue así que el 2 de agosto de 1858 se aprobó una ley dándole a la Junta Metropolitana de Obras la autoridad y dinero para embarcarse en el proyecto de ingeniería civil más grande del siglo el año siguiente, con *Joseph Bazalgette* a cargo [4].

Pero hubo dos preguntas importantes desde el inicio para hacer el proyecto viable y sustentable: *¿Cómo minimizar la ruta de alcantarillado entre dos*

puntos cualquiera de la red?, y ¿cuáles son los puntos de conexión más importantes?

La hazaña de ingeniería de Bazalgette fue un ejemplo de algunos de los primeros usos del campo de la teoría de gráficas, e ilustra la importancia de dos conceptos que observamos todo el tiempo hoy en relación con las redes: *la distancia entre vértices y la importancia de los vértices.*



Figura 3.4: Michael Faraday da su tarjeta al Padre Támesis [34].

3.1.1. Midiendo distancias

La distancia en una gráfica es un concepto sencillo y muy útil en la vida práctica, especialmente a la hora de modelar redes para la solución de problemas. En redes lo que se busca es minimizar la distancia dado dos vértices, si no hubiera forma de llegar de un vértice a otro se dice que la distancia es infinita. Para recorrer estas redes se utilizarán los conceptos definidos en el capítulo anterior para gráficas y digráficas, dependiendo de si la red tiene aristas o flechas.

En la figura 3.5 hay una red, donde se tienen algunos caminos distintos para llegar del nodo A a G , por ejemplo: $C_1 = (A, B, C, G)$, $C_2 = (A, E, F, G)$, $C_3 = (A, D, G)$. Como se dijo anteriormente lo que se busca es minimizar la distancia, en este caso la ruta óptima es C_3 , ya que pasa por la menor cantidad de aristas.

Pero en ocasiones la ruta de menor distancia no es viable, a la hora de modelar situaciones de la vida cotidiana puede que se necesite disminuir costos y la ruta más

corta sea la más cara, o que forzosamente se necesite pasar por todos los vértices que llegan a representar ciudades.

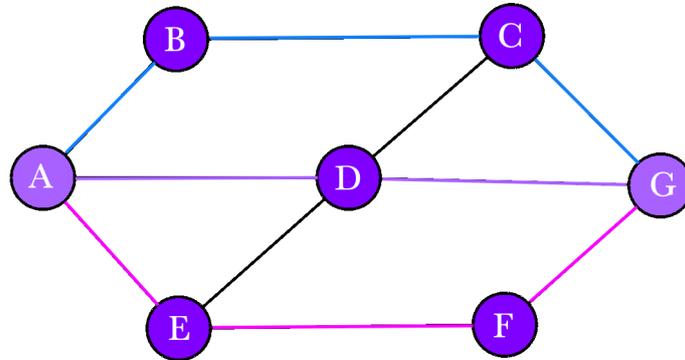


Figura 3.5: Ejemplo de una red no dirigida, donde se tienen algunos posibles caminos para llegar del vértice A al vértice G .

Las redes como se menciona anteriormente pueden ser dirigidas, esto quiere decir que se consideran flechas en lugar de aristas. Entonces una red es dirigida si todas sus aristas tienen una orientación. Por lo tanto, se define como **nodo fuente** de la red a aquel nodo en el cual sus aristas se encuentran orientadas hacia afuera, y **nodo destino** de la red es aquel en el cual sus aristas se encuentran orientados hacia él como se muestra en la figura 3.6.

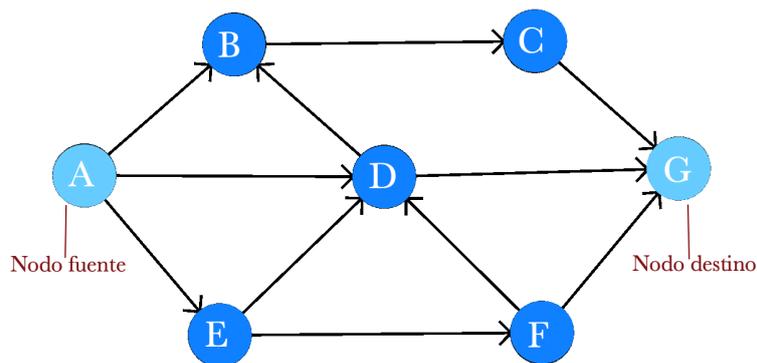


Figura 3.6: Ejemplo de red dirigida donde se aprecia su nodo fuente y su nodo destino.

3.2. Árboles de expansión mínima

Antes de empezar a plantear un problema, recordar que en el capítulo 2 se dio la definición de árbol, al pasarlo a un problema de redes, se debe conocer los costos o distancias entre diferentes vértices de la red, cabe aclarar que para esta sección se usarán redes no dirigidas y el objetivo es encontrar un árbol que comunique todos los vértices de la red, pero cuya distancia o costo total es mínimo. Dada una gráfica conexa y no dirigida $G = (V, E)$ se define como un **árbol de expansión** o como se conoce en inglés *spanning tree* de G , a un árbol que se extiende; es decir, incluye todos los vertices de G , con un subconjunto de aristas de G .

Este tipo de problemas abarca lo que son redes telefónicas, eléctricas, aéreas, marítimas, carreteras, calles, etc. donde los nodos pueden representar ciudades, aeropuertos, terminales de autobuses, plantas de luz, etc. y las aristas pueden ser las carreteras, líneas de tensión eléctricas, líneas telefónicas, trayectorias de aviones o barcos. Para solucionar este tipo de problemas se utilizan algoritmos para obtener el **árbol de expansión mínima** que es un árbol de expansión que tiene un costo total mínimo, o mejor dicho es un modelo de optimización de redes que consiste en enlazar todos los nodos de la red de forma directa o indirecta, con el objetivo de que la longitud total de las aristas sea mínima (entiéndase por longitud de la arista a una cantidad variable según el contexto operacional de minimización y que bien puede representar una distancia o unidad de medida).

3.2.1. Red de alcantarillado en la Ciudad de México

Recordando el problema de la gran peste que sufrió Londres en 1858, y como el ingeniero Joseph Bazalgette diseñó una red de alcantarillado para solucionar la situación, se plantea la siguiente situación:

El sistema de alcantarillado de la Ciudad de México está formado por diferentes elementos, entre los que figuran presas y lagunas de regulación, colectores, estaciones de bombeo, canales y túneles profundos, entre otros. Un mal mantenimiento de las infraestructuras hidráulicas, exceso de lluvias, edificación o actuaciones urbanísticas en lugares donde podía circular agua, entre otras situaciones hacen que el sistema de drenaje disminuya su capacidad de desalojo, principalmente son afectadas las tuberías, debido a las grandes sobrepresiones para transportar el agua, además que estas tuberías llegan a estar obstruidas por basura, lo que provoca severas inundaciones [1]. Para solucionar esto se propone colocar una estación de bombeo en cada una de las dieciséis delegaciones, para drenar de forma óptima el agua de tal forma que queden conectadas entre sí, pero se tiene que cuidar los costos.

En la figura 3.7 se muestran los distintos puntos, de color verde que es donde se pretende colocar las estaciones de bombeo, se puede observar las líneas de color rojo que son los caminos posibles que se pueden lograr entre cada estación de bombeo, cada una de estas aristas tendrá un costo, el cual va a ir variando dependiendo de la distancia que haya entre dos estaciones de bombeo, también influye el tipo de suelo, ya que este puede facilitar o complicar la conexión. Antes de resolver esto se hablará de un algoritmo que nos facilitará la solución del problema.

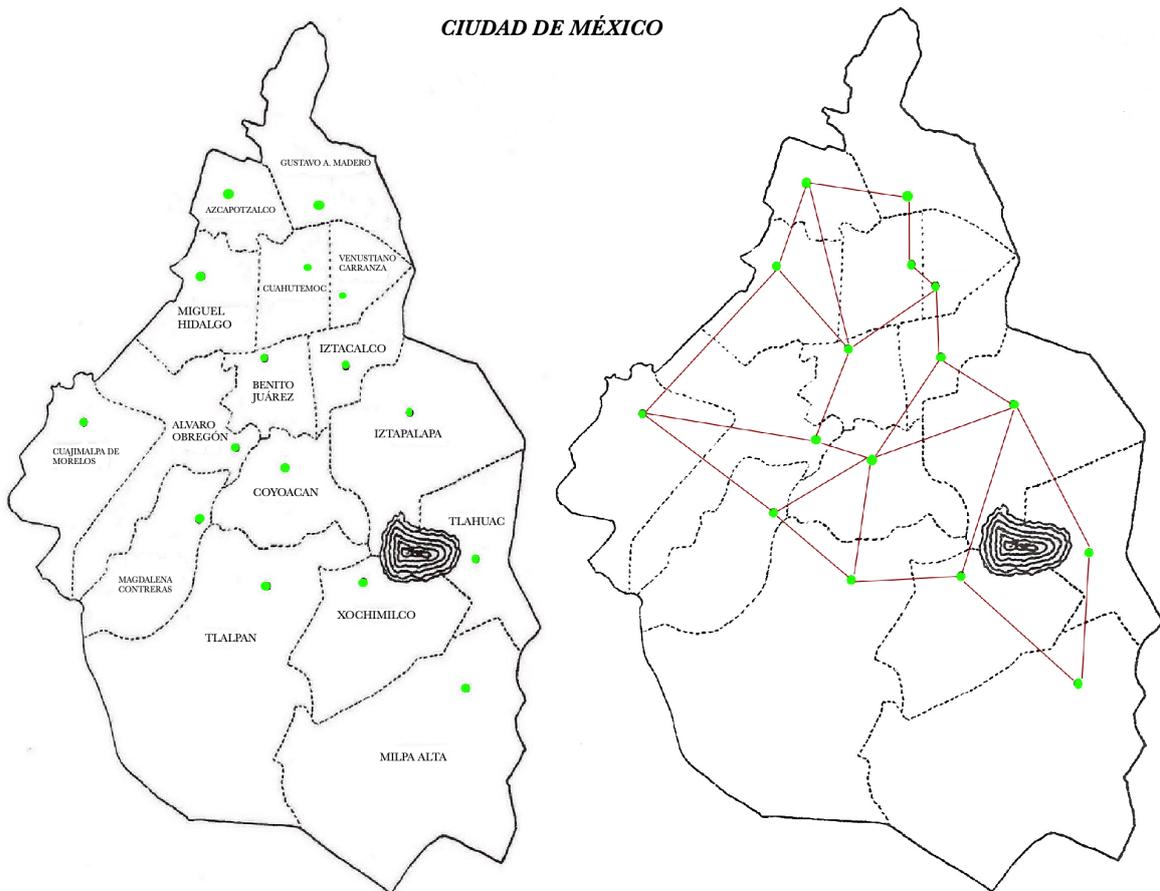


Figura 3.7: Mapa de las alcaldías de la Ciudad de México junto al mapa donde están las posibles rutas para la red de estaciones de bombeo.

3.2.2. Algoritmo de Kruskal

En 1956 el matemático norteamericano *Joseph Kruskal* descubrió un algoritmo que su aplicación garantiza encontrar un árbol generador de peso mínimo en cualquier gráfica conexa con pesos en sus aristas. Un árbol con peso es una gráfica donde cada arista tiene asociado un número, la suma de todos los números de todas las aristas del árbol, se llama peso del árbol. La idea principal de Kruskal consiste en ser tacaño, y escoger siempre la arista de menor peso, cuidando que la arista escogida en cada paso no forme un ciclo. Este algoritmo de la teoría de gráficas busca un subconjunto de aristas que incluyan todos los vértices formando un árbol y donde todas las aristas elegidas cumplan con que sus pesos sean mínimos.

Propósito:

En una red $R = (N, A, f)$ conexa con $N = n$ como se definió al inicio del capítulo, se pide obtener el árbol generador de peso mínimo, que se definirá como $T = (N', A')$, donde T es la gráfica generada por el algoritmo de Kruskal, por construcción T es acíclica y tiene $n - 1$ aristas y $N' \subset N$ y $A' \subseteq A$.

Pasos del algoritmo:

1. Ordenar las aristas de menor a mayor con respecto a su peso.
2. Se usará un contador J para saber el número de iteraciones y un contador K para saber el número de aristas agregadas. Empezamos con $J = K = 0$.
3. Empezamos con la iteración esto significa que $J = J + 1$.
4. Se escoge la arista de menor peso que se encuentran en R , en caso de tener dos o más del mismo peso, se escoge una aleatoriamente. Preguntamos: ¿La arista $a_j \in A(R)$ si se agrega forma un ciclo en T ?
Si la respuesta es No, agregar a T , esto implica que $a_j \in A'(T)$. Por lo tanto, $K = K + 1$ y eliminar la arista de la lista.
Si la respuesta es Sí, ir a 5.
5. Si $K = n - 1$ terminar, si $K < n - 1$ ir a 3.

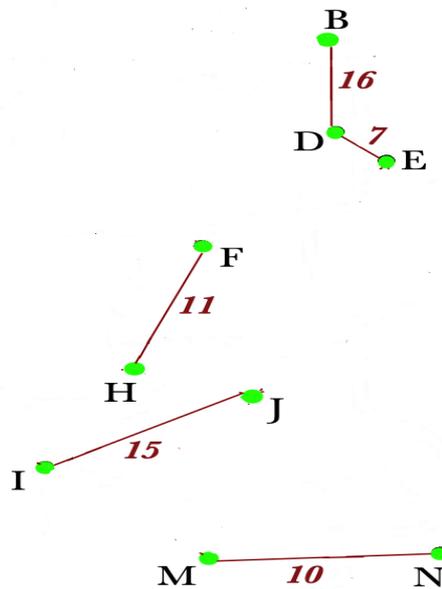
Retomando el problema de la red de alcantarillado para la Ciudad de México, en la figura 3.8 se muestran los costos de construir la conexión entre las diversas estaciones de bombeo. Los costos son en miles de pesos, para facilitar los cálculos se quitarán los últimos tres ceros. Para resolver el problema de lograr conectar todas las estaciones a un costo mínimo se utilizará el algoritmo de Kruskal.

Considerar que la red $R = (N, A, f)$ tiene $N = \{16\}$.

Paso 2. Ahora como se muestra en la siguiente tabla, empezamos el contador en 0 al igual que el número de aristas del árbol T .

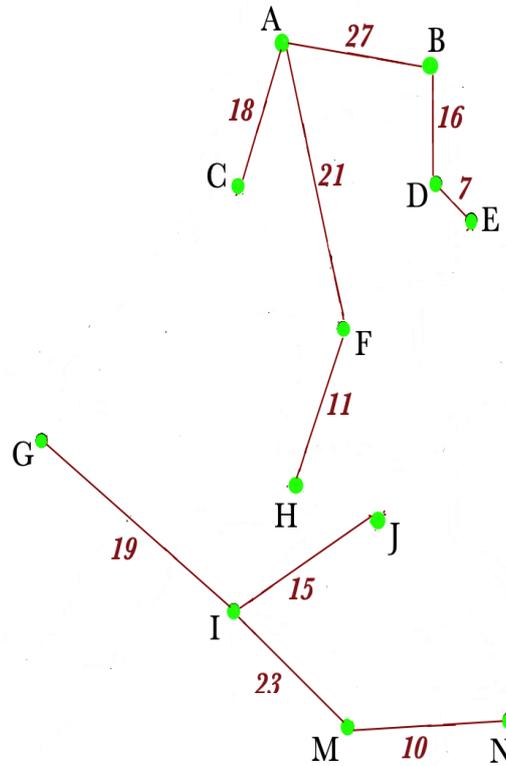
$J = \text{contador de it.}$	arista agregada	$K = n^{\circ} \text{ de aristas del árbol}$
0	—	0
1	a_1	1
2	a_2	2
3	a_3	3
4	a_4	4
5	a_5	5

Aquí hacemos una pausa para ver como se va viendo el árbol T , después de las primeras cinco iteraciones.



$J = \text{contador de it.}$	arista agregada	$K = n^{\circ} \text{ de aristas del árbol}$
6	a_6	6
7	a_7	7
8	a_8	8
9	a_9	9
10	a_{10}	10
11	—	10

Si se agrega la arista a_{11} se formaría un ciclo, puesto que $K < n-1$, no se agrega arista y seguimos iterando.



$J = \text{contador de it.}$	arista agregada	$K = n^{\circ}$ de aristas del árbol
12	—	10
13	a_{13}	11
14	a_{14}	12
15	a_{15}	13
16	—	13
17	—	13
18	—	13
19	a_{19}	14
20	a_{20}	15

Se termina las iteraciones puesto que $n = 16$ y el algoritmo indica parar cuando $K = n - 1$.

Ya que se tiene el árbol de peso mínimo lo que sigue es hacer la suma de los pesos que hay en las aristas del árbol. Al realizar la suma da un *Peso total igual a 361*, como las cantidades se les quitaron tres ceros para hacer la suma, al final se volverán a agregar y recalcular los valores, se concluye que el costo mínimo para lograr conectar las estaciones de bombeo será de \$3,610,000 pesos.

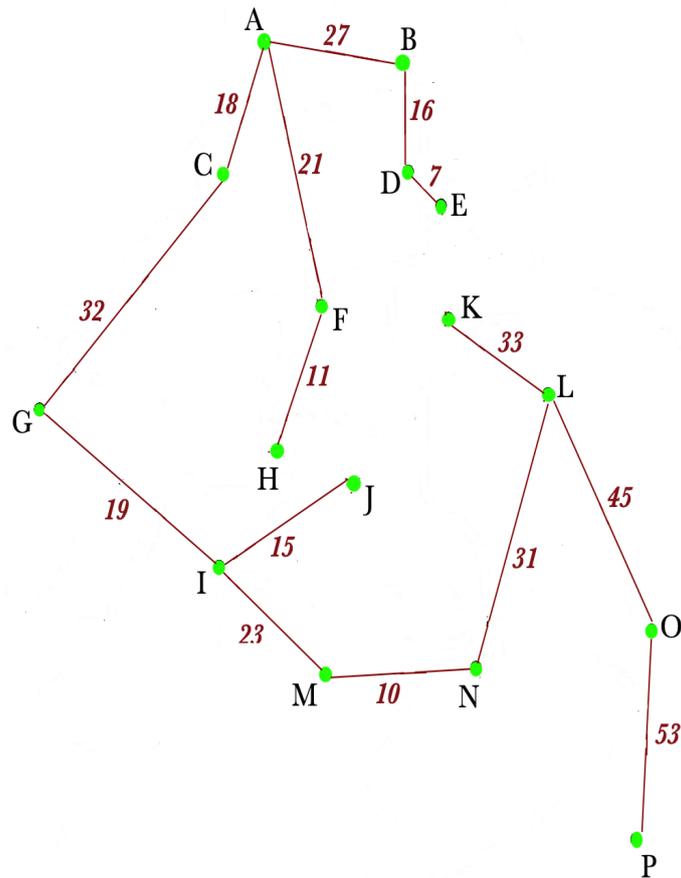


Figura 3.9: Árbol generador de peso mínimo para resolver el problema de alcantarillado en la Ciudad de México.

3.3. La teoría de gráficas como medio para una mejor organización

Desde la aparición del hombre en la tierra, se ha visto con la necesidad de organizarse, los primeros hombres se dividían en clanes de 25 a 50 miembros [39], los cuales tenían diversas actividades como cazar, hacer la comida, recolectar frutos, construir chozas, elaborar herramientas, etc. Su forma de dividir las tareas eran según el sexo, edad, estatus y parentesco con el fin de terminar de forma más rápida y tener mayores oportunidades de sobrevivencia [18]. Un ejemplo claro de organización fueron los egipcios con la construcción de las pirámides, aunque a la fecha es un misterio su construcción [28] a pesar de sus diversas teorías, si algo es seguro es que no pudieron construirse sin planes previos y una organización compleja de recursos.

Adelantándose un poco más en la historia, pasando al contexto de la Guerra Fría, la Armada de los Estados Unidos en 1957 creó el método *PERT* (*Project Evaluation and Review Techniques*), es un algoritmo basado en la teoría de gráficas, diseñado para facilitar la planificación de proyectos. La armada utilizó esta técnica para la planificación del proyecto Polaris un misil balístico basado en submarinos, construido con armas nucleares, ellos tenían planeado que la duración de Polaris estaría listo en 5 años, pero gracias al método *PERT* se adelantó dos años la fecha de terminación de su construcción [20]. Toda una ventaja si se habla de un contexto bélico.



Figura 3.10: William Francis Raborn (1905-1990), militar estadounidense, a cargo del proyecto Polaris.

También el proyecto Apolo, que permitió que el hombre pusiera el pie en la luna, fue programado mediante *PERT*. El método *PERT* es una herramienta especialmente valiosa para gestionar proyectos complejos, a largo plazo y en los que interactúan muchos actores. Como toda herramienta necesitaba perfeccionarse, ya que se encontró ciertas deficiencias, por ejemplo el tiempo en el que se ejecuta una determinada tarea está dado en estimaciones probabilistas, por lo tanto no se tenía una fecha de término exacta, sino aproximada.

Así que nació el método *CPM* (*Critical Path Method*), fue desarrollado en 1957 en Estados Unidos, por un centro de investigación de operaciones para las firmas Dupont y Remington Rand, cuyo objetivo era controlar y optimizar los costos de operación de las actividades de un proyecto. En este método, cada una de las tareas tenía una duración exacta, conocida de antemano.

Ambos métodos aportan desde su perspectiva la forma de determinar la ruta crítica en los proyectos y es conveniente emplearlos en conjunto; ya que permiten tener control en tiempos de ejecución y optimizar los costos, con la finalidad de obtener el proyecto en el menor tiempo y al menor costo. Es por ello que este método se emplea para la planeación y control de diversas actividades, tales como actividades de construcción, reparación de embarcaciones, investigación de mercados, estudios económicos regionales, auditorías, planeación de carreras universitarias, ampliaciones de fábrica, planeación de itinerarios para cobranzas, planes de venta, censos de población, etc.

3.3.1. Definiciones y reglas para construir la red *PERT*

1. Cada actividad se debe representar por una sola flecha.
2. Los vértices en la digráfica marcan cuando inicia el evento y cuando finaliza, de cada una de las distintas actividades.



3. En cada flecha debe ir el tiempo estimado que se tiene para realizar cada actividad.

4. Dos o más actividades diferentes no pueden identificarse por los mismos nodos iniciales y finales. Véase la figura 3.11.

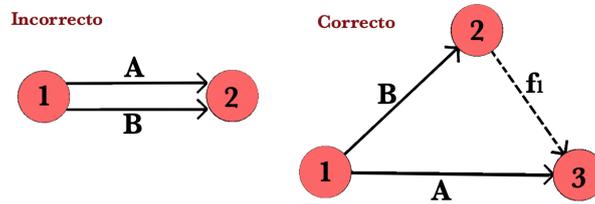


Figura 3.11: Actividades A y B que inician y finalizan al mismo tiempo.

5. Cada actividad debe estar identificada por dos vértices distintos, en el caso de existir actividades que se inicien al mismo tiempo, o bien que el inicio de una actividad dependa de la finalización de dos o más actividades distintas, se debe agregar una actividad ficticia, representadas por flechas punteadas. Véase figura 3.12.

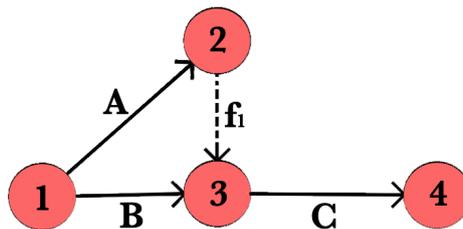


Figura 3.12: Ejemplo donde la actividad C debe comenzar cuando finalicen las actividades A y B .

6. Una *actividad ficticia* no consume recursos, ni tiempo por tanto su valor es cero, pero si indica una relación de precedencia con alguna actividad.
7. Cada nodo debe tener las siguientes notaciones:
- E_i con $i \in \mathbb{N}$, que representa el tiempo más temprano de realización de un evento.
 - L_i con $i \in \mathbb{N}$, que representa el tiempo más tardío de realización del evento.

- H_i con $i \in \mathbb{N}$, es el tiempo de holgura, esto representa que tanto tiempo puede tardarse en realizar una actividad, sin afectar todo el proyecto.
8. Ninguna actividad puede comenzar hasta que se hayan terminado las que le preceden.
 9. El número de identificación de cada nodo debe ir de orden ascendente de izquierda a derecha.
 10. Los nodos en los que al calcular la holgura sea 0, serán parte de la ruta crítica.
 11. Toda red *PERT/CPM* tiene un vértice inicial y un vértice final.

3.3.2. Vamos a organizar una fiesta

Algo muy común en la sociedad es ir a fiestas, pero normalmente preferimos asistir que organizarlas, ya que implican distintas tareas por hacer, se plantea la siguiente situación: Ana y Juan están próximos a casarse, tienen una lista las actividades que consideran son más difíciles de hacer, con el tiempo que estiman concluirán cada una de ellas, con ayuda del método *PERT/CPM* se buscará lograr una mejor organización y lograr tener todo listo en el menor tiempo posible antes de la fecha de la boda.

Clave	Actividad	Predecesora	Tiempo optimista (semanas)	Tiempo probable (semanas)	Tiempo pesimista (semanas)
A	Elegir el salón	<i>N/A</i>	2	4	12
B	Mandar a hacer invitaciones	<i>A</i>	3	4	5
C	Decidir decoración del salón	<i>A</i>	1	2	3
D	Escoger el menú de la comida	<i>A</i>	5	6	7
E	Arreglos florales	<i>A</i>	4	5	6
F	Escoger la música	<i>C, D</i>	2	3	4
G	Diseño y sabor del pastel	<i>B, F</i>	3	4	5
H	Elegir vinos y licores	<i>E, G</i>	1	2	3

Cuadro 3.1: Lista de actividades a realizar.

Paso 1: Calcular el tiempo estimado.

Debe estimarse la cantidad de esfuerzo de trabajo requerido y la cantidad de recursos para completar la actividad. Lo cual permite determinar la cantidad de periodos de trabajo (*duración de la actividad*) necesarios para completar la actividad. En la técnica *PERT*, el tiempo de la actividad se considera como una variable

aleatoria según una distribución de probabilidad *beta*. Una variable aleatoria intuitivamente, puede definirse como cualquier característica medible que toma diferentes valores con probabilidades determinadas. Toda variable aleatoria posee una distribución de probabilidad que describe su comportamiento.

Tomando como ejemplo la primera actividad que es escoger el salón, se tiene como: $T_o = 2$ (Tiempo optimista)
 $T_m = 4$ (Tiempo probable)
 $T_p = 12$ (Tiempo pesimista)

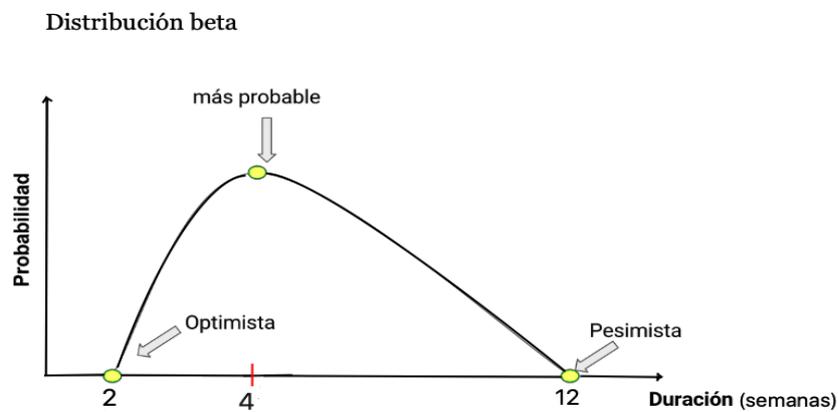


Figura 3.13: Comportamiento de la distribución beta.

El tiempo esperado de la actividad debe estar en algún punto de la curva como aparece en la figura 3.13. Para calcularlo puede hacerse por medio de la fórmula de aproximación [30]:

$$T_e = \frac{T_o + 4T_m + T_p}{6}$$

Sustituyendo los valores se tiene:

$$T_e = \frac{2 + 4(4) + 12}{6} = 5$$

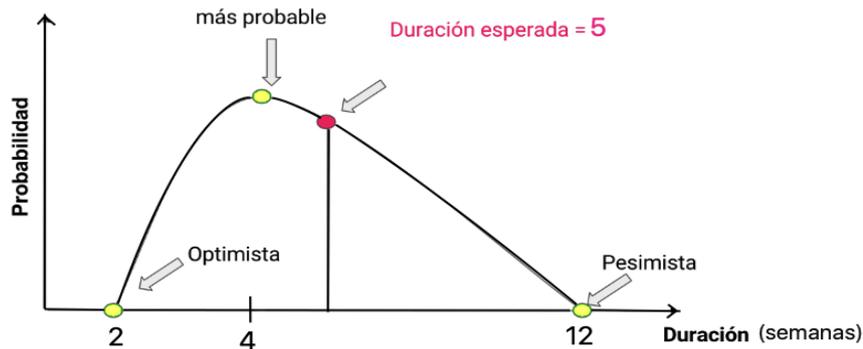


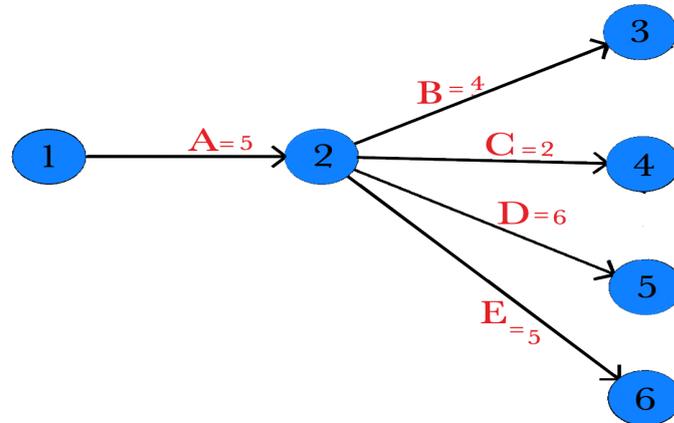
Figura 3.14: Punto de la curva donde se encuentra el tiempo estimado de la actividad A.

Al hacer los respectivos cálculos del tiempo estimado se obtienen nuevos valores, los cuales modifican el cuadro 3.1, entonces queda una nueva tabla donde ahora se tendrá la columna tiempo estimado, cuyos valores ya se han calculado previamente.

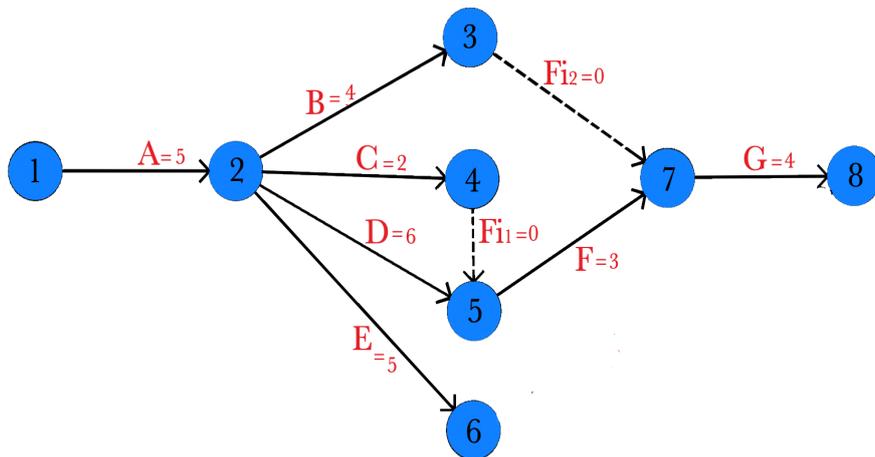
Clave	Actividad	Predecesora	Tiempo estimado (semanas)
A	Elegir el salón	<i>N/A</i>	5
B	Mandar hacer invitaciones	<i>A</i>	4
C	Decidir decoración del salón	<i>A</i>	2
D	Escoger el menú de la comida	<i>A</i>	6
E	Arreglos florales	<i>A</i>	5
F	Escoger la música	<i>C, D</i>	3
G	Diseño y sabor del pastel	<i>B, F</i>	4
H	Elegir vinos y licores	<i>E, G</i>	2

Cuadro 3.2: Lista de actividades con tiempo estimado calculado.

Paso 2: Ahora sigue elaborar la red *PERT* con base en las reglas mencionadas anteriormente. Lo primero será considerar la actividad inicial, aquella que no necesita de otra para empezar, en este caso es la actividad *A*. Posteriormente considerar las actividades que pueden empezar una vez finalizada *A*, las cuales son *B, C, D, E*, tendrán mismo vértice inicial, pero deben tener distinto vértice final, ya que si no se estaría violando la regla número 3, mencionada anteriormente, quedando como se ve en la figura 3.15.

Figura 3.15: Comienzo de la red *PERT*.

Posteriormente se tiene que la actividad F se empezará una vez concluida la actividad C y D , puesto que vienen del mismo vértice inicial marcado con 2 y no pueden tener mismo vértice final, se agregará una actividad ficticia F_{i_1} , y así podrá dar inicio F . Para la actividad G , que da inicio después de terminar B y F , puesto que vienen de distintos nodos iniciales, hay que agregar una actividad ficticia F_{i_2} , tal como se muestra en la figura 3.16.

Figura 3.16: red *PERT* con actividades ficticias.

De igual manera se agregará una actividad ficticia del vértice 6 al 8, para dar inicio a la actividad H , la cual empieza una vez terminada la actividad G , dando como resultado la red que se muestra en la figura 3.17.

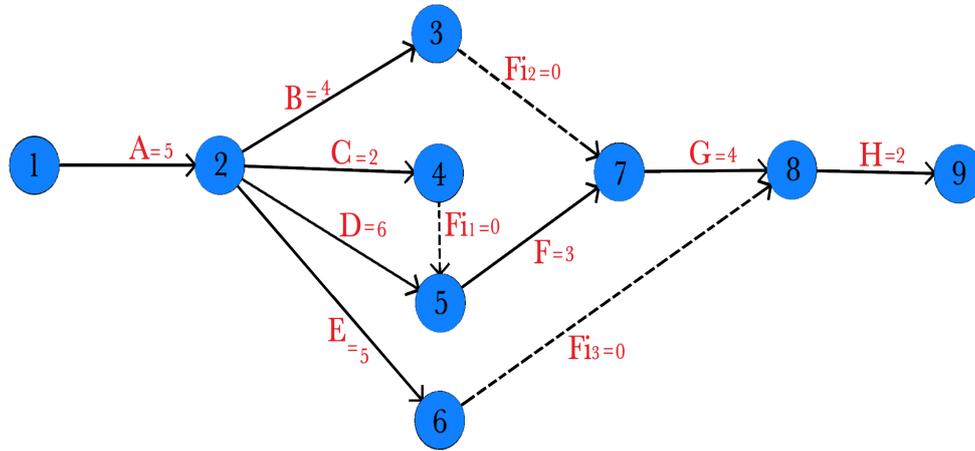


Figura 3.17: Red *PERT*.

Paso 3: Ahora se procede a calcular el indicador de ocurrencia más temprana (E_i), el cual hace referencia a lo más temprano que puede iniciar una actividad. Para hacer el respectivo cálculo se debe considerar:

- E_1 del nodo inicial es 0.
- E_i (del nodo i -ésimo) es igual a E_{i-1} del nodo anterior más la duración de la actividad (tiempo estimado) que finaliza en el nodo i .
- La red se recorre de izquierda a derecha.
- Si en un nodo finaliza más de una actividad, se toma el tiempo de la actividad con mayor valor.

Por ejemplo para calcular E_5 , puesto que en el nodo finaliza más de una actividad, se tienen dos casos:

$$E_5 = E_4 + F_{i_1} = 7 + 0 = 7$$

$$E_5 = E_2 + D = 5 + 6 = 11$$

Por lo tanto el nodo cinco quedará etiquetado con el de mayor valor que es 11.

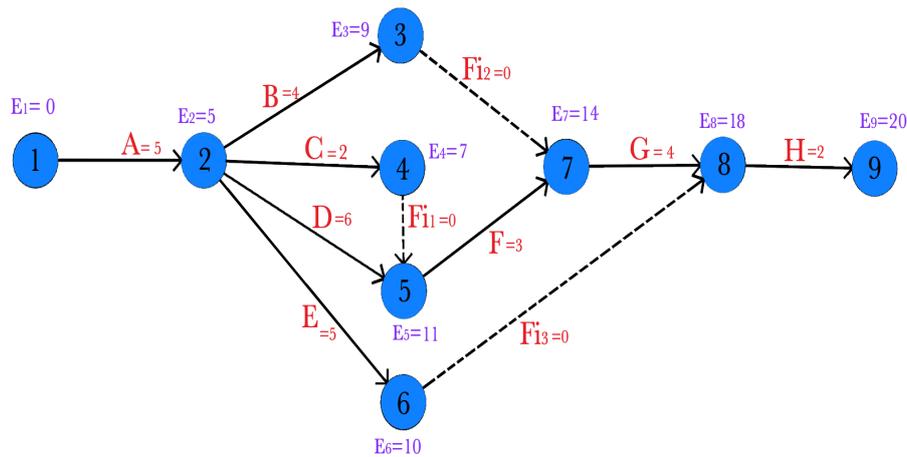


Figura 3.18: Red *PERT* con la ocurrencia temprana calculada.

Paso 4: Ahora se calculará la ocurrencia tardía (L_i), que representa lo más tarde que se puede empezar una actividad sin retrasar el proyecto. Para realizar el cálculo se debe considerar lo siguiente:

- La red se recorre de derecha a izquierda.
- L_i del primer vértice de derecha a izquierda, es igual a E_i de ese mismo vértice.
- L_i del nodo n es igual a L_i del nodo *sucesor* menos la duración de la actividad que se inicia.
- Si en un vértice finaliza más de una actividad, se toma el tiempo de la actividad con menor valor.

En la red mostrada en la figura 3.19, como el primer vértice de derecha a izquierda es el 9, se tiene que $L_9 = E_9 = 20$, cumpliendo la segunda condición. Ahora para calcular el vértice 2, se tienen cuatro posibles casos:

$$L_2 = L_3 - 4 = 14 - 4 = 10$$

$$L_2 = L_4 - 2 = 11 - 2 = 9$$

$$L_2 = L_5 - 6 = 11 - 6 = 5$$

$$L_2 = L_6 - 5 = 18 - 5 = 13$$

Se escoge la de menor tiempo, por lo tanto $L_2 = 5$.

Paso 5: Ahora se va a calcular el **tiempo de holgura** (H_i) u holguras, esta variable corresponde al tiempo máximo que se puede retrasar una actividad sin afectar el resto del proyecto. Para calcular la holgura se considera lo siguiente:

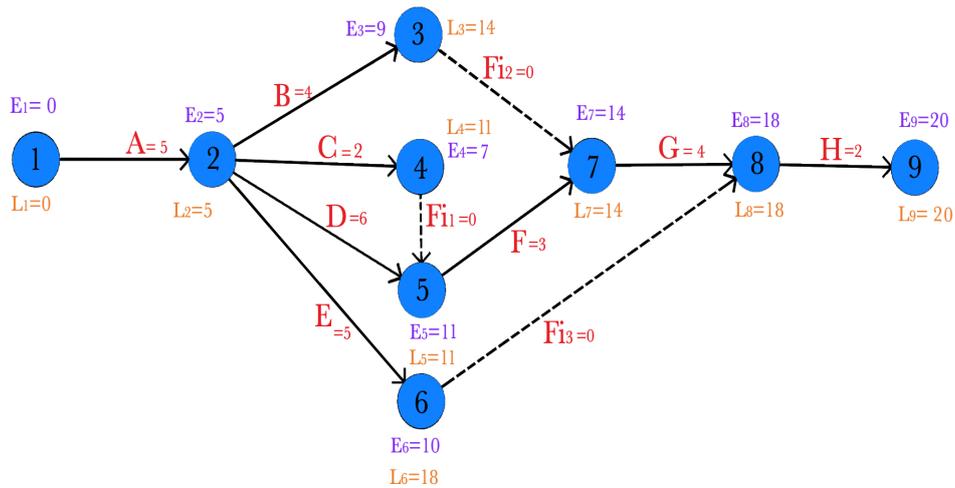


Figura 3.19: Red *PERT* con la ocurrencia tardía calculada.

- H_i de un vértice i es igual a la ocurrencia tardía L_i menos la ocurrencia temprana E_i
- Si la holgura $H_i = 0$, se dice que la **actividad es crítica**, quiere decir que dicha actividad no puede tener atraso, de lo contrario afectaría el tiempo estimado para terminar todo el proyecto.

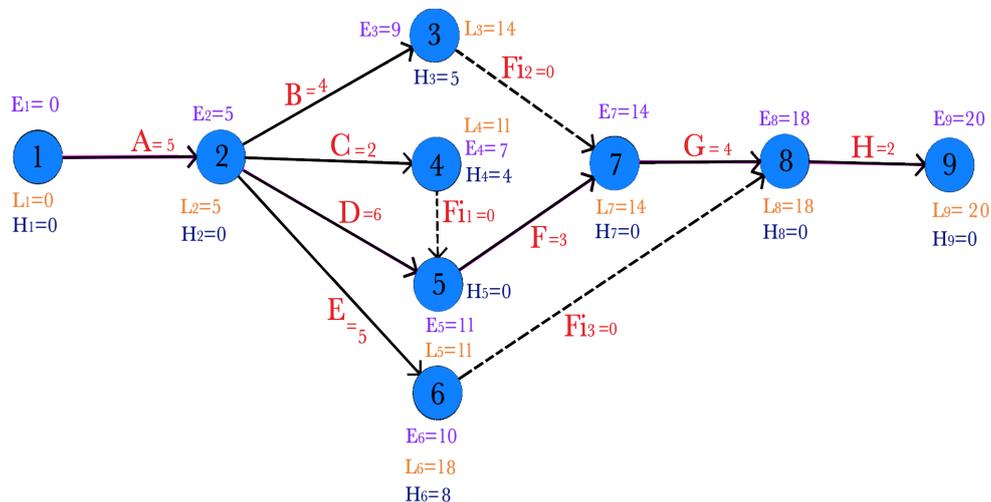


Figura 3.20: Red *PERT* con las holguras calculadas.

Paso 6: Calcular la ruta crítica. Se debe considerar lo siguiente:

- La **ruta crítica** se encuentra constituida por las actividades críticas, es decir donde las holguras son cero.
- Las actividades críticas por definición constituyen la ruta más larga que abarca el proyecto, es decir que la suma de las actividades de una ruta crítica determinará la duración estimada del proyecto.
- Puede darse el caso en el que se encuentren más de una ruta crítica.

En la figura 3.21, se observa de color rosa la ruta crítica de la red. Esta ruta se encuentra compuesta por las actividades A, D, F, G, H . Al hacer la suma de las actividades da un total de 20, esto significa que la duración para tener listos los preparativos de la boda será de 20 semanas.

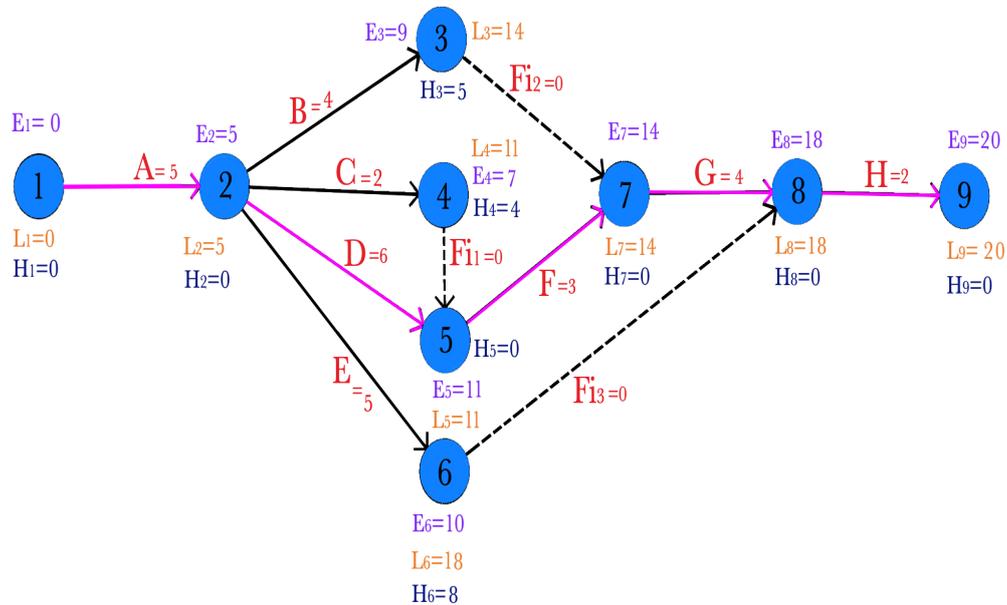


Figura 3.21: Red *PERT* con su ruta crítica.

Paso 7: Establecer un cronograma, el cual ayuda visualmente a tener un control y orden para llevar a cabo cada actividad.

Para establecer un cronograma deberá considerarse varios factores, los más importantes son: relación de precedencia, duración las actividades, destacar las actividades críticas y las holguras. Como herramienta de ayuda, se utilizará el diagrama de Gantt, permite seguir el curso de cada actividad, a partir de una representación gráfica sobre dos ejes. En el vertical se ponen las actividades y en el horizontal el tiempo.

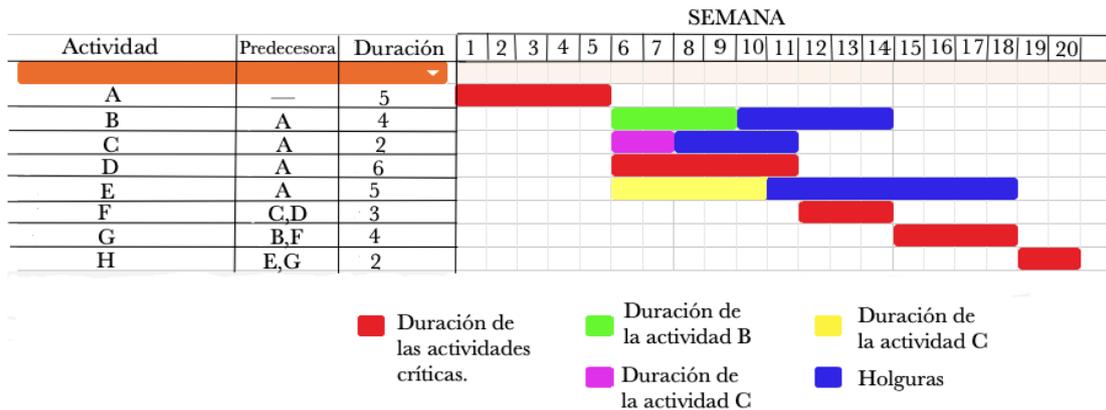


Figura 3.22: Diagrama de Gantt.

Capítulo 4

Las gráficas y su relación con la sociedad

4.1. ¿Mis amigos tienen más amigos que yo?

Nuestra sociedad actual se encuentra sumergida en el mundo de las redes sociales, lo cual ha sido una gran herramienta, pero a su vez se ha convertido en el mayor distractor. Las redes sociales más conocidas como Facebook, Instagram, Twitter, Snapchat, etc. Guardan gran cantidad de información sobre nosotros, pero algo muy importante que destaca, es la cantidad de “amigos” que se tienen agregados. Hay muchas variables a considerar cuando se agregan amigos a una red social, ya sea porque se conocen de algún lugar, comparten algunos intereses, trabajan juntos o simplemente quieren hacer crecer su lista de amigos. Las redes sociales han tenido gran impacto que nos pueden hacer sentir popular o poco sociable.

En algún momento revisando ciertos perfiles de amigos, me surge la pregunta: *¿Mis amigos tienen más amigos que yo?*, intuitivamente se podría pensar que en promedio podrían tener el mismo número de amigos que nosotros. Sin embargo, en 1991 el sociólogo Scott L. Feld tras varios análisis de redes sociales, y una serie de experimentos, llegó a la conclusión sorprendente de que **en promedio, nuestros amigos tienen más amigos que nosotros mismos** [11]. Lo que en realidad se conoce como *paradoja de la amistad*.

A partir de aquí, al decir redes sociales no se está hablando de plataformas como Facebook o Twitter (aunque tampoco las excluiríamos como medio), sino al conjunto de personas con las que se tiene algún vínculo: amigos, parejas, familiares, etc., y los vínculos que a su vez esas personas tienen. Para explicar mejor esto, se adaptará el ejemplo de Scott que utilizó en [11], donde se toma una muestra de 12 escuelas, se

entrevistó a los alumnos pidiéndoles que nombraran a sus amigos y los que coincidieron en amistad se tomaron en cuenta, en este caso, se utilizarán personas famosas que se desarrollan en diferentes ámbitos. Una red social está representada por una gráfica $G = (V, A)$ donde el conjunto de vértices, V , corresponde a la gente de la red social y el conjunto de aristas, A , a la relación de amistad entre pares de personas.

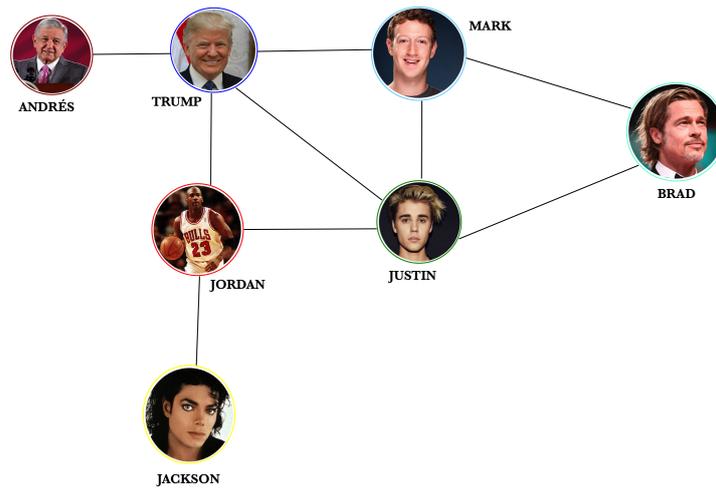


Figura 4.1: Ejemplo de gráfica social.

		NÚMERO DE AMIGOS DE SUS AMIGOS						
		ANDRÉS	TRUMP	MARK	BRAD	JUSTIN	JORDAN	JACKSON
P E R S O N A J E S	ANDRÉS		4	—	—	—	—	—
	TRUMP	1		3	—	4	3	—
	MARK	—	4		2	4	—	—
	BRAD	—	—	3		4	—	—
	JUSTIN	—	4	3	2		3	—
	JORDAN	—	4	—	—	4		1
	JACKSON	—	—	—	—	—	3	

Figura 4.2: Tabla con la información de la gráfica social.

La tabla que aparece en la figura 4.2 se puede leer de la siguiente manera: Mark

tiene tres amigos; Trump tiene cuatro amigos, Brad con dos amigos y Justin con cuatro amigos.

Nombre	Número de amigos x	Número total de amigos de sus amigos y	Promedio de amigos de sus amigos y/x
ANDRÉS	1	4	4
TRUMP	4	11	2.7
MARK	3	10	3.3
BRAD	2	7	3.5
JUSTIN	4	12	3
JORDAN	3	9	3
JACKSON	1	3	3
TOTAL	18	56	22.5
Promedio de amigos por personaje	2.5	--	--
Promedio de amigos de los amigos por personaje	--	--	3.1

Cuadro 4.1: Resumen del número de amigos y de la media de los amigos de los amigos para cada famoso.

Con la información obtenida en el cuadro 4.1, se puede decir que en promedio cada uno de los famosos tiene 2.5 amigos, en cambio los amigos de sus amigos tienen un promedio de 3.1.

En el caso de nuestra red, Trump y Justin son las personas con mayor número de amigos, con un total de 4 cada uno. Por otra parte, las personas que tienen una cantidad de amigos menor al promedio siempre estarán expuestas a encontrar fácilmente amistades que tiene más amigos que ellos, como es el caso de Andrés y Brad.

En términos matemáticos, esto se cumple para toda gráfica, si el personaje x tiene n amigos, entonces x se contará n veces al realizar el cálculo de sus amigos. Llamemos x_i al número de amigos del jugador i , y μ a la media de amigos de los n personajes. Entonces se puede decir que para calcular la media de las conexiones por vértice en la gráfica se tiene la siguiente fórmula, la cual se aplicará al ejemplo propuesto:

$$\mu = \frac{\sum_{i \in V} x_i}{n} = \frac{\sum_{i \in V} x_i}{|V|} = \frac{2|A|}{|V|} = \frac{2(9)}{7} = \frac{18}{7} = 2.5$$

Para calcular la media de amigos de los amigos, se toman los datos de la figura

4.2, en el orden que van apareciendo para cada persona de la lista, teniendo los siguientes:

$$4, 1, 3, 4, 3, 4, 2, 4, 3, 4, 4, 3, 2, 3, 4, 4, 1, 3$$

Reordenando los datos se ven de la siguiente forma:

$$4, 4, 4, 4, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 2, 2, 1, 1$$

Calculando la media :

$$\begin{aligned} & \frac{4 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 4 + 3 + 3 + 3 + 3 + 3 + 3 + 2 + 2 + 1 + 1}{18} = \\ & \frac{4 \cdot 4 + 4 \cdot 4 + 3 \cdot 3 + 3 \cdot 3 + 2 \cdot 2 + 1 \cdot 1 + 1 \cdot 1}{18} = \\ & \frac{4^2 + 4^2 + 3^2 + 3^2 + 2^2 + 1^2 + 1^2}{18} = \\ & \frac{56}{18} = 3.1 \end{aligned}$$

Por lo tanto la fórmula para calcular la media de amigos de los amigos es:

$$\mu' = \frac{\sum_{i \in V} (x_i)^2}{\sum_{i \in V} x_i}$$

Ahora falta demostrar que μ' es mayor o igual que μ , para dicha demostración se usarán conceptos de estadística que pueden encontrar en [5].

Demostración. Recordando la fórmula estadística de varianza para población que es:

$$\sigma^2 = \frac{1}{n} \sum_{i \in V} (x_i - \mu)^2$$

Desarrollando se tienen las siguientes expresiones:

$$\begin{aligned} & = \frac{1}{n} \sum_{i \in V} (x_i^2 - 2\mu x_i + \mu^2) \\ & = \frac{1}{n} \sum_{i \in V} x_i^2 - 2\mu \left(\frac{1}{n} \sum_{i \in V} x_i \right) + \mu^2 \end{aligned}$$

ya que $\mu = \frac{1}{n} \sum_{i \in V} x_i$ entonces se puede expresar de la siguiente manera:

$$= \left(\frac{1}{n} \sum_{i \in V} x_i^2 \right) - \mu^2$$

Esto implica que:

$$\sigma^2 = \frac{\sum_{i \in V} x_i^2}{n} - \mu^2$$

De aquí se obtiene que:

$$\sum_{i \in V} x_i^2 = (\mu^2 + \sigma^2) \cdot n$$

Se divide ambos miembros de la igualdad por $\sum_{i \in V} x_i$, obteniendo:

$$\frac{\sum_{i \in V} x_i^2}{\sum_{i \in V} x_i} = \frac{(\mu^2 + \sigma^2) \cdot n}{\sum_{i \in V} x_i}$$

Tomando en cuenta que $\mu \cdot n = \sum_{i \in V} x_i$ se sigue que:

$$\mu' = \frac{(\mu^2 + \sigma^2) \cdot n}{\mu \cdot n}$$

$$\mu' = \frac{\mu^2 + \sigma^2}{\mu}$$

$$\mu' = \mu + \frac{\sigma^2}{\mu}$$

$$\therefore \mu' \geq \mu$$

□

Puesto que la varianza siempre es positiva y la media de amigos también. Por lo tanto, no hay que sentirse mal por tener pocos amigos, es muy probable que tus amigos tengan más amigos que tú. Y esto tiene su origen debido a la variedad de grados, siempre habrá personas (vértices) que tienen muchos más amigos que el promedio, es muy probable que lleguemos a conectar con estas personas muy sociables, por lo tanto incrementará la media de amigos de nuestros amigos.

4.2. Gráficas versus epidemias

Retomando la sección anterior sobre la *paradoja de la amistad*, a continuación se hablará de una aplicación de dicha paradoja. La aplicación es sobre el control de enfermedades y epidemias.

El médico y sociólogo Nicholas Christakis, de la Universidad de Yale, y su colaborador James Fowler de la Universidad de California en San Diego, también investigador social, publicaron un artículo sobre su más reciente estudio en las gráficas sociales, en el cual mencionan que si los métodos de detección de epidemias utilizaran una propiedad de gráficas sociales, serían más eficientes, como por ejemplo: durante un brote contagioso, es probable que las personas en el centro de una gráfica se infecten antes que los miembros aleatorios de la población. Por lo tanto, la recolección cuidadosa de información de una muestra de individuos centrales, dentro de las gráficas sociales humanas, podría usarse para detectar brotes contagiosos antes de que ocurran en la población en general [8].

Ahora surge la pregunta: ¿cuándo un nodo es central? El grado es probablemente la medida más intuitiva de la centralidad: entenderemos que un nodo es central si está conectado a muchos otros nodos y también consideraremos que estos nodos a su vez tienen un número considerable de conexiones, de esta manera hace que un nodo sea más central.

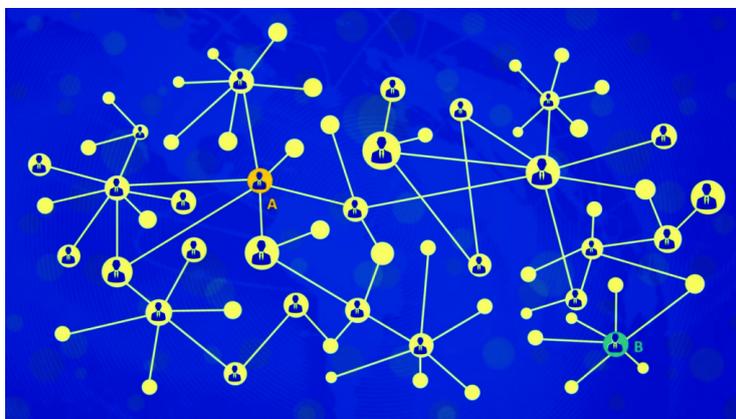


Figura 4.3: Gráfica social donde hay vértices centrales, el vértice A es más central que el vértice B .

Se puede observar en la figura 4.3, el vértice A y vértice B tienen el mismo grado de conexiones, pero hay que resaltar que el vértice A tiene conexiones con individuos que tienen mayor número de conexiones, a diferencia del vértice B .

Otro concepto importante para tener en cuenta es el *grado de separación*, el cual indica en una gráfica la distancia promedio que hay entre todos los vértices o específicamente de un vértice a todos los demás que componen la gráfica, a continuación un ejemplo y la forma de calcularlo.

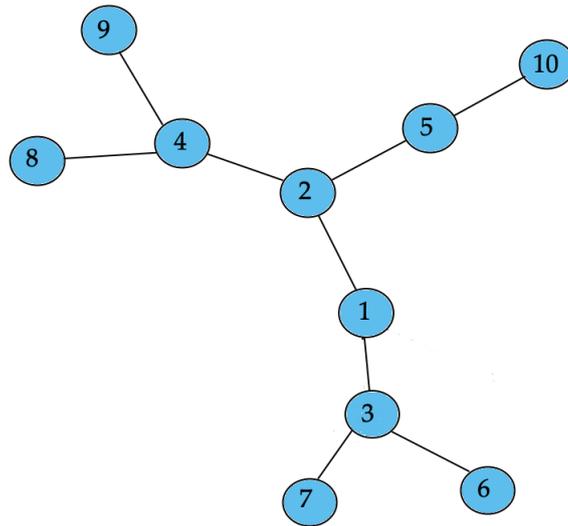


Figura 4.4: Gráfica social ejemplo.

Con ayuda de la figura 4.4, se harán los siguientes pasos:

- *Paso 1:* Calcular todos los posibles pares de vértices.

```

1,2 2,3 3,4 4,5 5,6 6,7 7,8 8,9 9,10
1,3 2,4 3,5 4,6 5,7 6,8 7,9 8,10
1,4 ..... 6,9 7,10
..... 5,9 6,10
..... 4,9 5,10
..... 3,9 4,10
.....2,9 3,10
1,9 2,10
1,10
    
```

Total de pares = 45

Cuando son gráficas más complejas o con un elevado número de vértices, los posibles pares de vértices se pueden usar calculando las combinaciones

$$\frac{n!}{k!(n-k)!} = \binom{n}{k}$$

- *Paso 2:* Encontrar la distancia más corta entre cada par de vértices, usando cualquier algoritmo de rutas más cortas como *Dijkstra*, *Floyd* y *Warshall*

Rutas más cortas				
1,2 = 1	2,4 = 1	3,7 = 1	5,6 = 4	7,9 = 5
1,3 = 1	2,5 = 1	3,8 = 4	5,7 = 4	7,10 = 5
1,4 = 2	2,6 = 3	3,9 = 4	5,8 = 3	8,9 = 2
1,5 = 2	2,7 = 3	3, 10 = 4	5,9 = 3	8,10 = 4
1,6 = 2	2,8 = 2	4,5 = 2	5,10 = 1	9,10 = 4
1,7 = 2	2,9 = 2	4,6 = 4	6,7 = 2	
1,8 = 3	2,10 = 2	4,7 = 4	6,8 = 5	
1,9 = 3	3,4 = 3	4,8 = 1	6,9 = 5	
1,10 = 3	3,5 = 3	4,9 = 1	6,10 = 5	
2,3 = 2	3,6 = 1	4,10 = 3	7,8 = 5	

Cuadro 4.2: Rutas más cortas de la gráfica que aparece en la figura 4.4.

- *Paso 3:* Sumar todas las rutas más cortas :
 $1+1+2+2+2+2+3+3+3+2+1+1+\dots = 127$ (Total)
- *Paso 4:* Sustituir los valores en la siguiente fórmula

$$\text{Grado de separación} = \frac{\text{Total}}{\text{Total de pares de vértices}} = \frac{127}{45} = 2.82222$$

Por lo tanto, se tiene que el grado de separación de esa gráfica es de 2.82, este mismo procedimiento se puede utilizar para saber el grado de separación de un solo vértice, se consideran todos los posibles pares que contengan ha dicho vértice de interés, posteriormente se calculan todas las rutas más cortas a ese vértice y al final se sustituyen los valores en la fórmula mencionada.

Como menciona Christakis en su artículo, si un contagio empieza aleatoriamente en algunas personas y se empieza a propagar persona a persona, en promedio llegará más rápido a personas que son vértices centrales de la red, que a las personas que están en la periferia, es decir en los alrededores de la red. Debido a que las personas centrales tienen un grado de separación menor a la que tiene un individuo promedio de la gráfica, un individuo central tiene mayor probabilidad de contagio, ya que es muy fácil llegar a él, porque las personas centrales tienen un número menor de grado

de separación del individuo promedio de la gráfica. Por lo tanto, al tener un monitoreo e información de estos individuos centrales, podría usarse para detectar brotes contagiosos antes de que ocurran en la población en general. Este método podría generalizarse a otros contagios biológicos, psicológicos, informativos o conductuales que se propagan en redes.

Hoy en día por ejemplo, una gripe se contagia entre personas que interactúan cara a cara, algunos contagios pueden propagarse a través del contacto incidental, la duración de la exposición entre personas con vínculos sociales es típicamente mucho más alto que entre extraños, lo que sugiere que la red social en sí será un conducto importante para la propagación de un brote.

Christakis realizó una variedad de ejercicios de modelación, uno de ellos fue con un brote de la gripe H1N1 en la universidad de Harvard en el otoño y el invierno de 2009, donde se tomó una muestra de 744 estudiantes de pregrado de Harvard College, se les pidió que dijeran el nombre de un amigo, para así ir modelando la red y tratar de tener el mayor número de personas, donde al final se obtuvo información de 1789 estudiantes, los cuales se consideraron para el estudio.

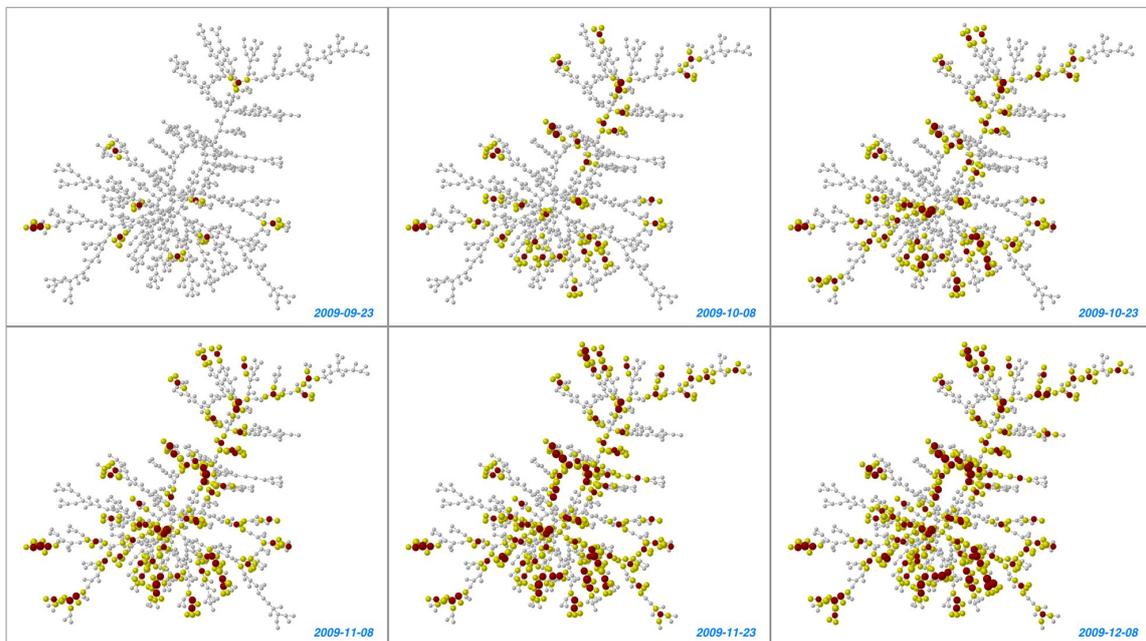


Figura 4.5: Red social de los alumnos de Harvard en el brote de influenza H1N1, imagen cortesía de [7].

En la figura 4.5 los individuos infectados son de color rojo, los amigos de los individuos infectados son de color amarillo y el tamaño del nodo es proporcional al número de amigos infectados, esto fue en un monitoreo de 3 meses. Se compararon dos grupos, el primero donde los individuos eran seleccionados al azar y otro compuesto por individuos que fueron nominados como amigos, por miembros del grupo aleatorio. Se observó que los individuos del grupo de amigos tendían a contraer la gripe antes que los del grupo aleatorio.

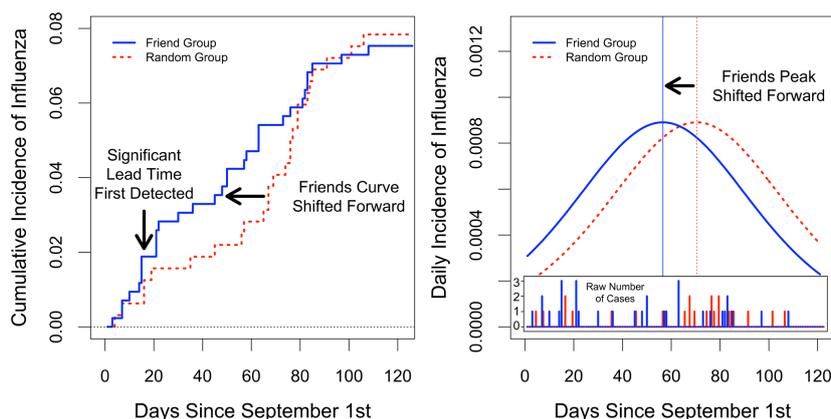


Figura 4.6: Diferencias empíricas en el contagio de gripe entre el grupo de amigos y los individuos elegidos al azar, *imagen cortesía de [8]*.

Como se muestra en la figura 4.6, el grupo aleatorio está de color rojo y el grupo de amigos en azul, la epidemia en el grupo de amigos tuvo un nivel alto de contagios en el día 13, la diferencia entre los dos grupos al llegar a su pico máximo de contagios fue de 16 días. Monitoreando el grupo de amigos se pudo tener una alerta 16 días antes de una epidemia inminente en esta población, incluso se puede analizar que cuando el grupo de los amigos inicio su nivel alto de propagación, hasta el punto máximo de contagios pasaron 42 días, es decir mes y medio, en el cual se pudo haber alertado la epidemia futura de una población en particular.

Como lo menciona Christakis este método no solo se limita a los gérmenes, sino también para predecir epidemias de todo tipo, cualquier cosa que se propague por contagio social, pueden ir desde ideas políticas, religión, compra de algún producto, estilos de vida, etc. La clave para que funcione el método es que debe haber una influencia interpersonal entre los individuos y esto depende del vínculo que los conecte, por ejemplo las infecciones de transmisión sexual se esparcen por los vínculos sexuales.

Otro de sus experimentos fue sobre la *inmunidad de grupo*, esto quiere decir que si se tiene una población de 1,000 personas y se busca hacer que está población sea inmune a un patógeno, no es necesario inmunizar a todos, da la misma eficacia inmunizar al 96 %, ya que si una o dos personas se infectan, no propagarían una epidemia, puesto que estarían rodeados de personas ya vacunadas.

Algunos científicos han estimado qué pasaría si se toma una muestra aleatoria del 30 % de estas 1,000 personas, es decir si se vacunan a 300, ¿se obtendría alguna inmunidad poblacional?, y la respuesta es no. Pero si se toma él 30 % de una población determinada aleatoriamente para vacunar y a ese porcentaje se le pregunta el nombre de un amigo, se vacunarán a esos amigos que mencionen, en el mejor de los casos se vacunará a otro 30 % de la población, teniendo vacunados a lo más el 60 % del total, aquí es donde entra la paradoja de la amistad, ya que el segundo grupo de vacunados es más probable que sean nodos centrales, es decir, personas con un mayor número de conexiones al ser vacunadas, es menos probable que lleguen a contagiar a más personas y así es muy probable que la población total quede protegida como si se hubiera vacunado el 95 %, hablando de virus que se transmiten de persona a persona.

A continuación se hará una pequeña simulación, en la cual se tomará un grupo real de 105 estudiantes y se hará su respectiva gráfica social donde cada arista representa un lazo de amistad.

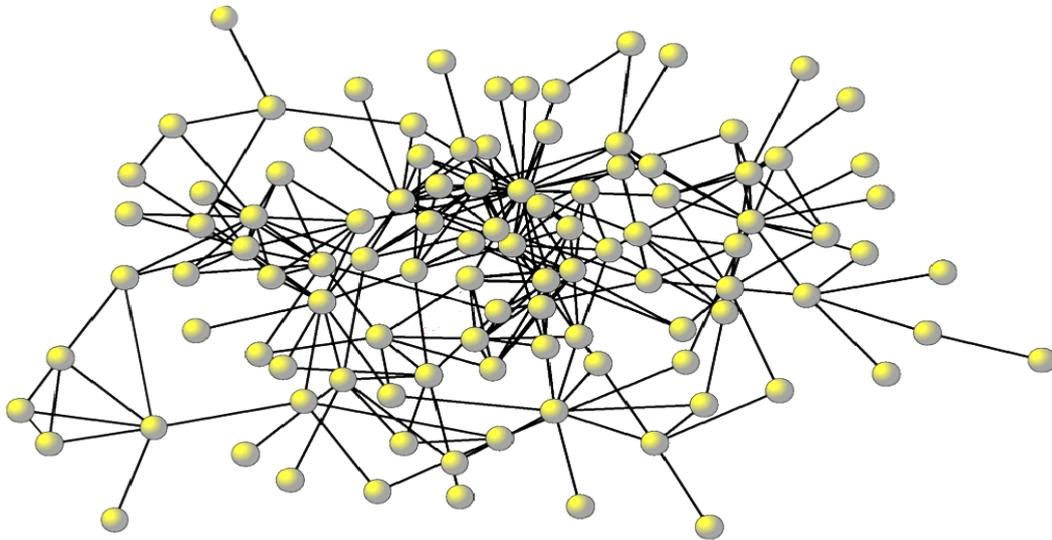


Figura 4.7: Gráfica social de 105 estudiantes.

Después se tomará una muestra aleatoria del 30% de personas para vacunar contra la influenza, que corresponde a 32 personas del total.

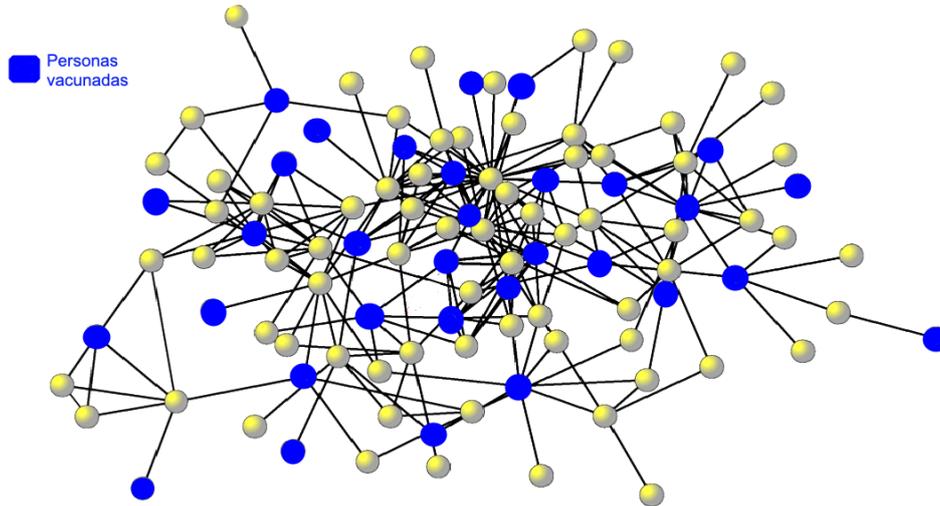


Figura 4.8: Personas escogidas aleatoriamente en la primera muestra para vacunar.

A cada una de las personas escogidas aleatoriamente se les preguntará el nombre de un amigo y el amigo mencionado se vacinará.

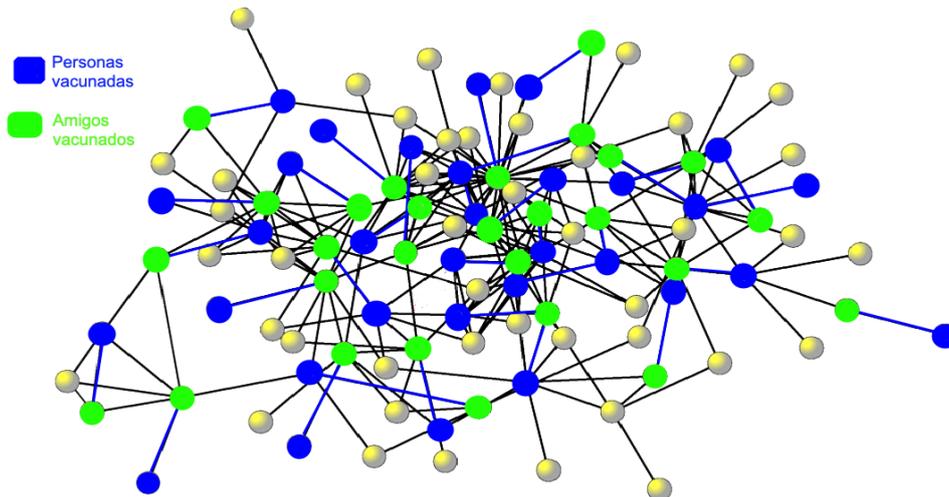


Figura 4.9: Gráfica con amigos vacunados.

Puede darse el caso que un amigo mencionado ya haya sido seleccionado en la primera muestra, en el mejor de los casos puede darse que no coincidan con ningún amigo de la primera muestra, entonces se obtendría el 30% de personas para la segunda muestra. En ejemplo propuesto se obtuvo 28 personas que corresponden el 26.6% del total. Ahora supongamos que una persona desarrolló influenza, veamos el origen y hasta dónde será capaz de contagiar.

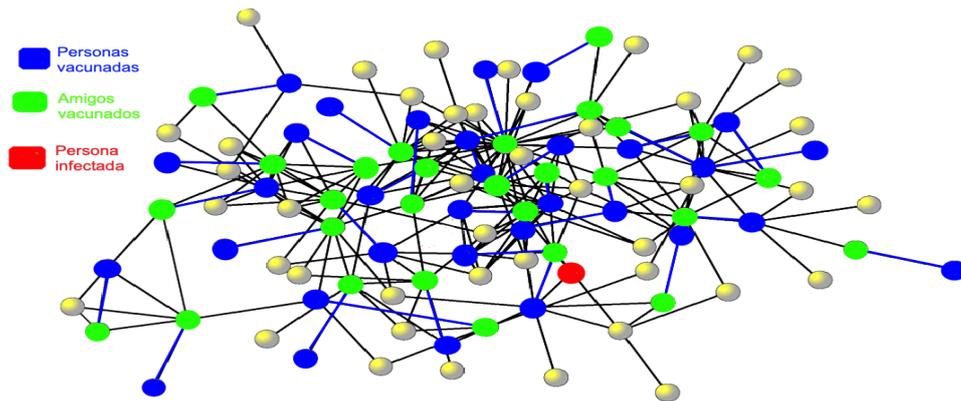


Figura 4.10: Gráfica donde aparece la primera persona infectada.

Como se puede ver en la gráfica de la figura 4.11, el virus avanzó contagiando a 3 personas más, pero se pudo frenar, esto representa el 3.8% de la población total. Por lo tanto se pudo inmunizar a esa población de estudiantes con una efectividad como si se hubiera vacunado el 95%.

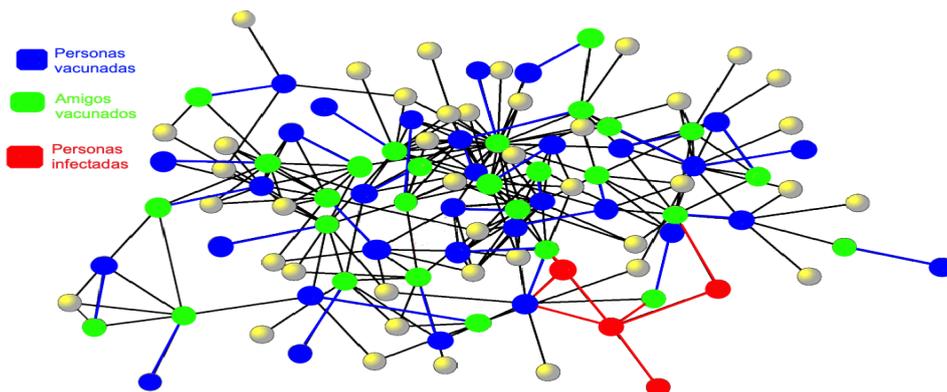


Figura 4.11: Gráfica con los nuevos infectados.

Otro caso que se puede considerar es, si el virus empieza simultáneamente en distintas personas, pongamos un ejemplo: el virus infectó a 8 personas, que corresponde el 7.6% del total.

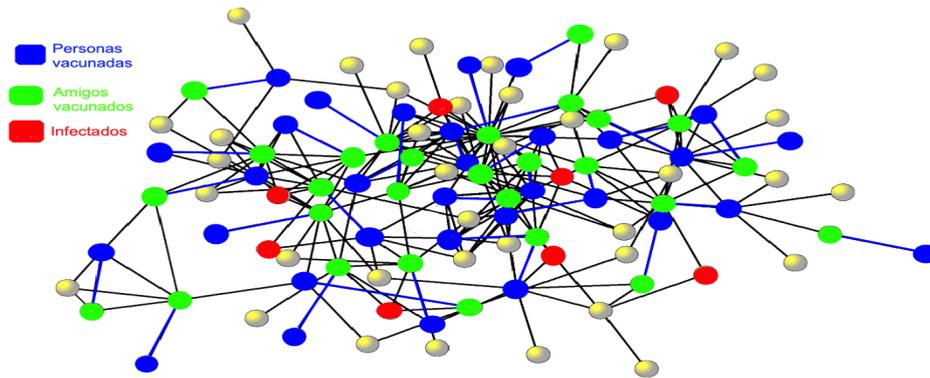


Figura 4.12: Gráfica con personas infectadas por el virus.

Ahora se verá todas las conexiones que tiene cada persona infectada y a quiénes puede llegar a infectar.

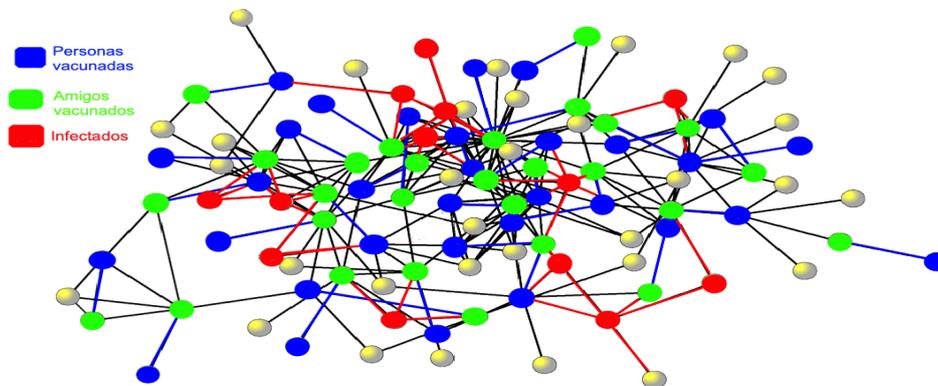


Figura 4.13: Gráfica con nuevos infectados.

En la gráfica 4.13, se observa que aumento el número de infectados, pero no fue tan alto, de 8 que se tenía originalmente subió a 13, además que cada una de las personas con influenza llego a contagiar a lo más 3 personas.

Para finalizar, la evidencia preliminar sugiere lo siguiente:

- Lo ideal sería vacunar a toda la población, pero en distintos países hay una gran escasez de vacunas por diversas cuestiones, como dinero, el tiempo que se tarda para realizar una vacuna puesto que es material biológico, no se puede acelerar su proceso [19] y la demanda puede llegar a ser alta, entonces esta es una alternativa para ahorrar costos y brindar lo más que se pueda de salud.
- Sin necesidad de vacunar el 95 % de la población y gracias a las diversas simulaciones de Christakis, empezar vacunando el 30 % de una población determinada y posteriormente al amigo más cercano con quién tenga más contacto, es el óptimo hablando de costo-beneficio, en el peor de los casos donde todas las personas que no tienen vacuna se les desarrolle el virus, al menos se tendría sana a casi el 60 % de la población.
- La velocidad de propagación de una enfermedad infecciosa depende de todo un abanico de factores biológicos, ambientales y sociales. No obstante, los epidemiólogos resumen el impacto de todos esos condicionantes en una cantidad que se llama: el **número básico de reproducción**, este representa el valor promedio de nuevos contagios que cabe esperar por cada persona infectada y va variando para cada enfermedad.
- En la práctica, la vacunación generalizada reduce el número reproductivo de la enfermedad. Si se vacuna una cantidad suficiente de personas, la cantidad promedio de contagios puede decrecer hasta uno, lo que garantizará que la infección se propagará solo de forma lineal [29].
- Un problema que puede intervenir en esta aplicación es el mapeo de la red social, ya que no es tan fácil como suena, se encuentran diversos problemas; es algo costoso, antiético por violación a la privacidad de las personas o demasiado complejo de realizar, entonces ¿cómo localizar a las personas centrales sin mapear la red?, es posible gracias a la paradoja de la amistad. Para más información consultar [9].
- Se sigue haciendo investigación sobre las gráficas sociales y su aplicación en la difusión de cualquier cosa transmisible dentro de la sociedad, puede ir desde una enfermedad, una moda, el comportamiento de las personas o hasta una ideología. Decidimos qué lazos crear y cuáles mantener o frenar; con cuántas personas nos conectamos, de qué manera y en qué medida [41].

Capítulo 5

Gráficas y datos

En el capítulo anterior se habló de gráficas sociales, de las cuales se siguen haciendo diversas investigaciones y su aplicación para la difusión de cualquier cosa que se pueda transmitir, para lograr esto es indispensable tener cierta información de las personas. Con el tiempo y avances tecnológicos, surgió la necesidad de almacenar información, para poder preservarla del espacio, deterioro y así poder consultarla cuando se necesite. Por ejemplo, a Facebook lo utilizan aproximadamente mil millones de personas en todo el mundo, por lo tanto debe haber un gran volumen de información. Actualmente con el aumento de las tecnologías de la información se ha logrado informatizar desde negocios, bancos, aerolíneas, empresas, transportes, sector educativo, administración pública, telecomunicaciones, etc. Aquí es donde entramos al mundo de los datos, con lo mencionado anteriormente podemos intuir que hay una inmensidad de ellos, por lo que debe haber una manera de poder almacenarlos.

Se define una **base de datos** como un almacén de datos de una parte seleccionada del mundo real, para ser utilizado con propósitos particulares [12]. En otras palabras, es un almacén con una serie de datos organizados y estructurados, los cuales son recolectados para consultas posteriores, seguridad, respaldo, recuperación, manipulación, etc.

Cuando se trata de una base de datos, donde habrá múltiples consultas por diversos clientes, es necesario el uso de un **servidor de base de datos**, en este se encuentran las bases de datos y se puede acceder a ellas desde distintas terminales, situadas en lugares diferentes con un programa llamado *cliente*.

Existen diversos tipos de software especiales para poder trabajar con las bases de datos dependiendo de lo que se requiera o necesite, se conocen como sistemas gestores de bases de datos. Con estos programas se permite almacenar y acceder de

forma rápida a los datos estructurados dentro de la base de datos. Hoy en día existe gran variedad de bases de datos que se aplican, se pueden clasificar de acuerdo con su utilidad, contexto, estructura y las necesidades que busquen solucionar.

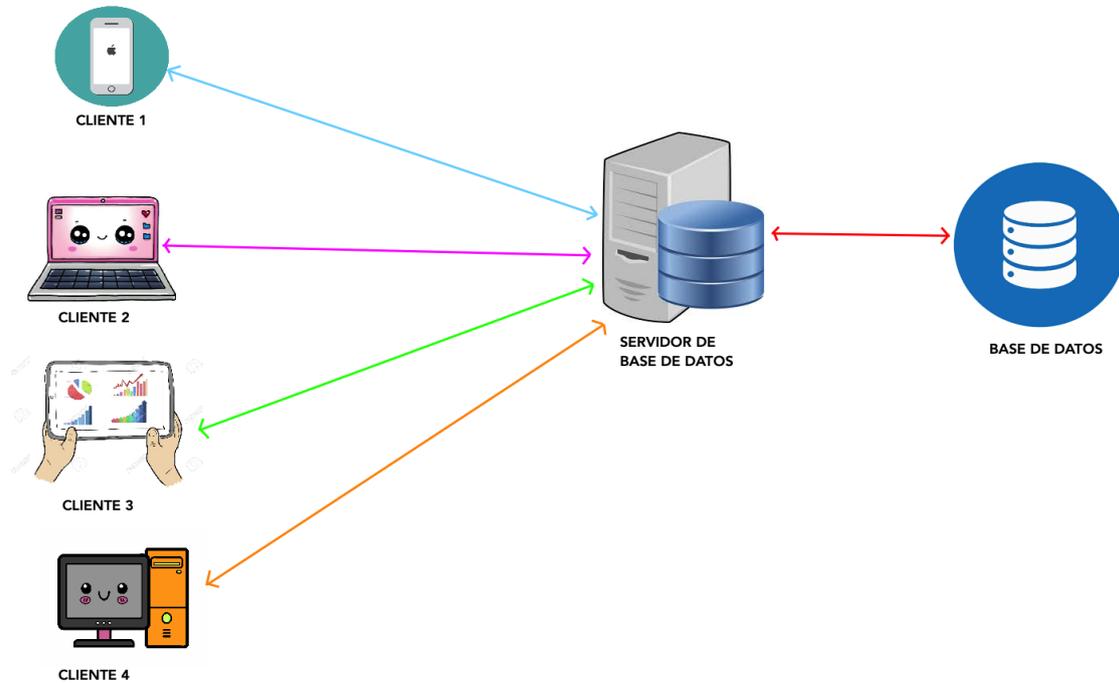


Figura 5.1: Modelo cliente-servidor para la gestión de bases de datos.

5.1. Bases de datos orientadas a gráficas

Como se mencionó anteriormente, existen diversos tipos de bases, donde destacan las relacionales y no relacionales. Las *bases de datos relacionales*, como MySQL, SQL Server y Oracle, utilizan el modelo relacional y siempre es mejor usarlas, cuando los datos son consistentes y ya tienes algo planificado. Por otro lado, las *bases de datos no relacionales* o mejor conocidas, como *bases de datos NoSQL* son una solución, para almacenar grandes cantidades de datos, no utilizan una estructura de datos en forma de tablas y relaciones entre ellas, por lo cual, se vuelve más fácil su manipulación, almacenando la información en otros formatos, como correspondencia de columnas, clave-valor, documentos o gráficas [10].

Para tener un poco más la idea, imaginemos un zoológico en el cual hay una variedad de animales, los cuales, dentro del mismo espacio, tienen sus hábitats respectivos, entonces cada especie interactúa distinto, no hay entendimiento con los otros animales, pero nadie afecta a nadie, ahora tomando un hábitat en específico, al ingresar un animal que no tiene relación con los demás, puede alterar a los animales que se encuentran dentro del hábitat.

En las bases relacionales se requiere primero, determinar la estructura de estos, si se cambia la estructura de uno de los datos, puede perjudicar todo el sistema, ya que las tablas están relacionadas, a diferencia de una base no relacional, donde puede tener su propia estructura sin afectar los demás y se puede agregar más campos a medida que avanza. Están conformadas por documentos, campos y datos del documento, además se almacenan por colecciones [32].

Aquí es donde entran las **bases de datos orientadas a gráficas** (*graph database*), representan la información como nodos de una gráfica y sus relaciones como las aristas de la misma. De esta forma, se puede usar la Teoría de Gráficas para recorrer la base de datos y describir atributos de los nodos (entidades) y de las aristas (relaciones). El beneficio del almacenamiento en gráficas y su principal objetivo, es el de tener un alto rendimiento y facilidad de escalabilidad [37]. Este tipo de bases de datos, no solamente da importancia a los datos, sino también a la conexión que pueden tener, un nodo puede tener tantas aristas ya sean dirigidas o no dirigidas como se necesiten, por lo tanto su aplicación puede ser en redes sociales, transporte público, sistemas de geoespaciales, gestión de redes, detección de fraudes bancarios, recomendaciones para comprar un producto en línea, etc. Se puede decir que permite un ajuste más cercano a los datos del mundo real con los que está trabajando, puesto que la composición utilizada es idéntica al pensamiento humano y la conexión de los datos, es muy intuitiva.

Empezaremos a tratar una aplicación interesante de las *graph database*. Hoy en día con el uso de las tecnologías, navegar por Internet y principalmente el uso de redes sociales, se pueden obtener datos como nuestros gustos, intereses, aficiones, a partir del registro del historial de búsqueda, aplicaciones en las que se pasa más tiempo, que tipo de contenido se ve, que música se escucha, que páginas o fotos nos gustan, etc. Todos estos datos los podemos meter en una gráfica, como se muestra en la figura 5.2.

Aquí se tiene el ejemplo de Valeria López la cual se puede decir que tiene gustos como ver Netflix; comparte fotos en sus redes de los dibujos que hace; sigue páginas de animales, puesto que es dueña de perritos chihuahua; le gusta comprar cosas en

línea, principalmente Amazon; los géneros de música que escucha y sus artistas preferidos. Toda esta información de su comportamiento les sirve a las empresas para analizarla, y crear un perfil completo del cliente, con el fin de mandarle publicidad personalizada. Por otro lado se puede observar, como esta gráfica puede crecer a otros nodos muy rápido, a partir de los amigos de los amigos, e ir creando sus perfiles de clientes. Podemos imaginar los millones y millones de datos que se van agregando a cada segundo. Incluso yo me atrevería a decir que navegar por la web es un mundo totalmente diferente para cada persona, por lo que las redes sociales pueden influir en todos los aspectos de nuestras vidas.

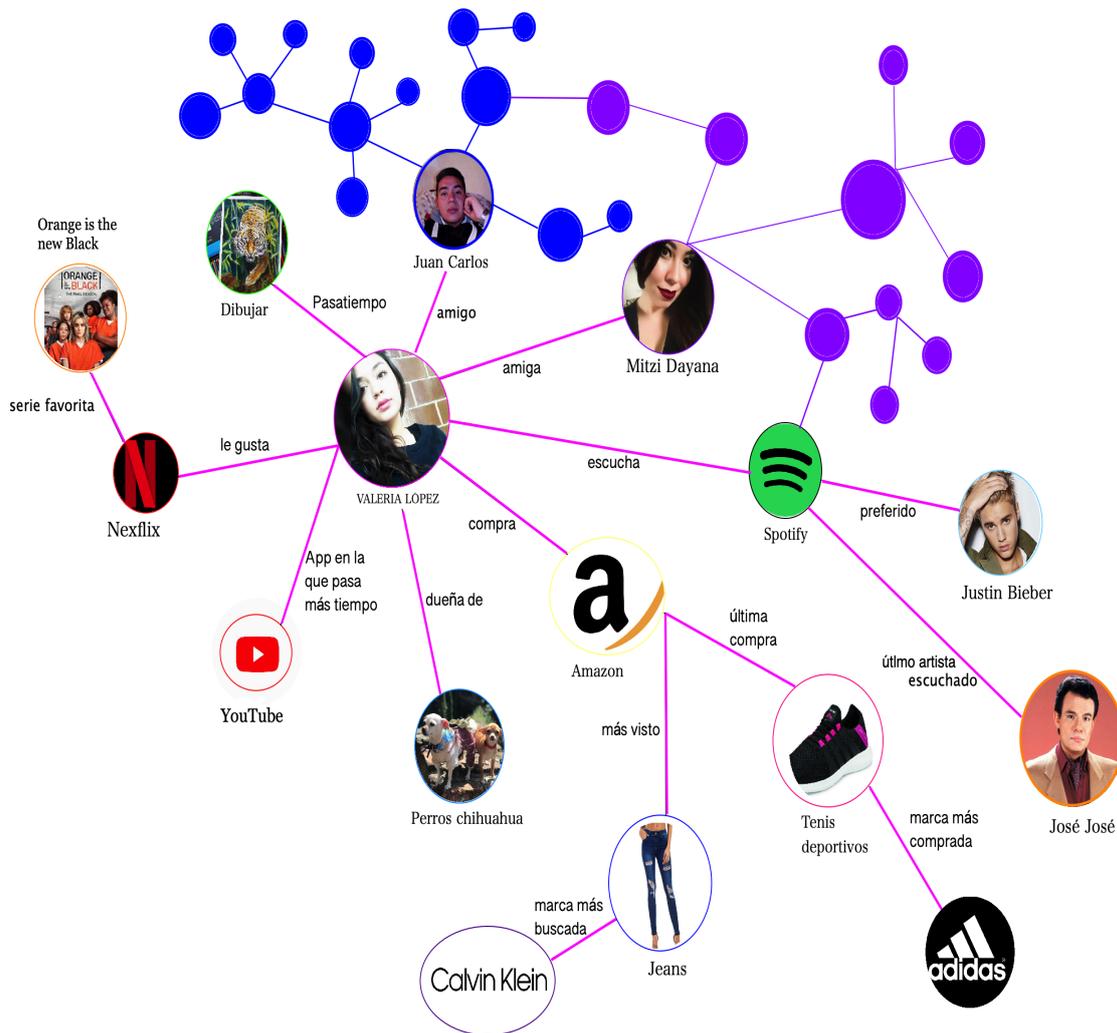


Figura 5.2: Ejemplo de *graph database* para crear un perfil de cliente.

5.2. Neo4j

Ahora bien, una vez que se ha decidido utilizar este tipo de bases de datos orientadas a gráficas, la pregunta es ¿cuál usar?

Existen diversos sistemas de bases de datos orientados a gráficas, así que hay que investigar la que se adapta más a nuestras necesidades, algunas son: *AllegroGraph*, *ArangoDB*, *InfiniteGraph*, *OrientDB*, *JanusGraph*, *Amazon Neptun* y la más conocida que se ha popularizado, **Neo4j**.

Es una base de datos de gráficas nativa, construida desde cero, para aprovechar no solo la información, sino también las relaciones que hay, conecta los datos a medida que se almacenan, lo que permite realizar casi cualquier tipo de consultas, a velocidades exorbitantes [24]. Algunas empresas que han usado a Neo4j son eBay, Walmart, Telnor, Cisco, HP, Airbnb para ofrecer mejor servicio a sus clientes o analizar sus datos. Brevemente se puede explicar algunos usos, eBay lo está utilizando para acelerar, optimizar y mejorar su popular servicio de entrega en el mismo día a los clientes finales de todo el mundo. Walmart de igual manera lo utiliza para dar sentido al comportamiento de los compradores en línea, a fin de poder optimizar y realizar ventas cruzadas, de las principales líneas de productos, en los mercados principales [25].

Para darse cuenta un poco más del potencial de *Neo4j*, en el libro *Neo4j in action*, el autor Vukotic y sus colegas [38] realizan un experimento, donde en una gráfica social con 1,000,000 de personas, además cada persona con aproximadamente 50 amigos, se busca encontrar a los amigos de los amigos a una profundidad máxima de cinco. Hace dos casos, primero con la información en una base de datos relacional, (*MySQL*) y otro con un recorrido en *Neo4j*, obteniendo los resultados siguientes:

Profundidad	Tiempo de ejecución (segundos) en MySQL	Tiempo de ejecución (segundos) en Neo4j	Registros devueltos
2	0.016	0.01	~2,500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	sin terminar	2.132	~800,000

Cuadro 5.1: Resultados de una base de datos relacional versus búsqueda en Neo4j.

En profundidad dos (amigos de amigos), tanto la base de datos relacional, como la base de datos de gráficas, funcionan lo suficientemente bien como para que consideremos su uso en un sistema en línea. Por otra parte en la profundidad tres, la

base de datos relacional tarda más de 30 segundos en lanzar resultados, lo cual es inaceptable para consultas en tiempo real. En la profundidad cinco, la base de datos relacional simplemente no termina en completar la consulta, mientras que *Neo4j* devuelve un resultado en unos dos segundos.

Otra cuestión que hay que considerar, es que, el modelo que propone *Neo4j*, llamado de gráficas etiquetadas y con propiedades, o más conocido como *labelled property graph model* que contiene las siguientes características [33]:

- Los nodos contienen propiedades (pares clave-valor).
- Los nodos se pueden etiquetar con una o más etiquetas.
- Las relaciones se nombran y dirigen, siempre tienen un nodo inicial y final.
- Las relaciones también pueden contener propiedades

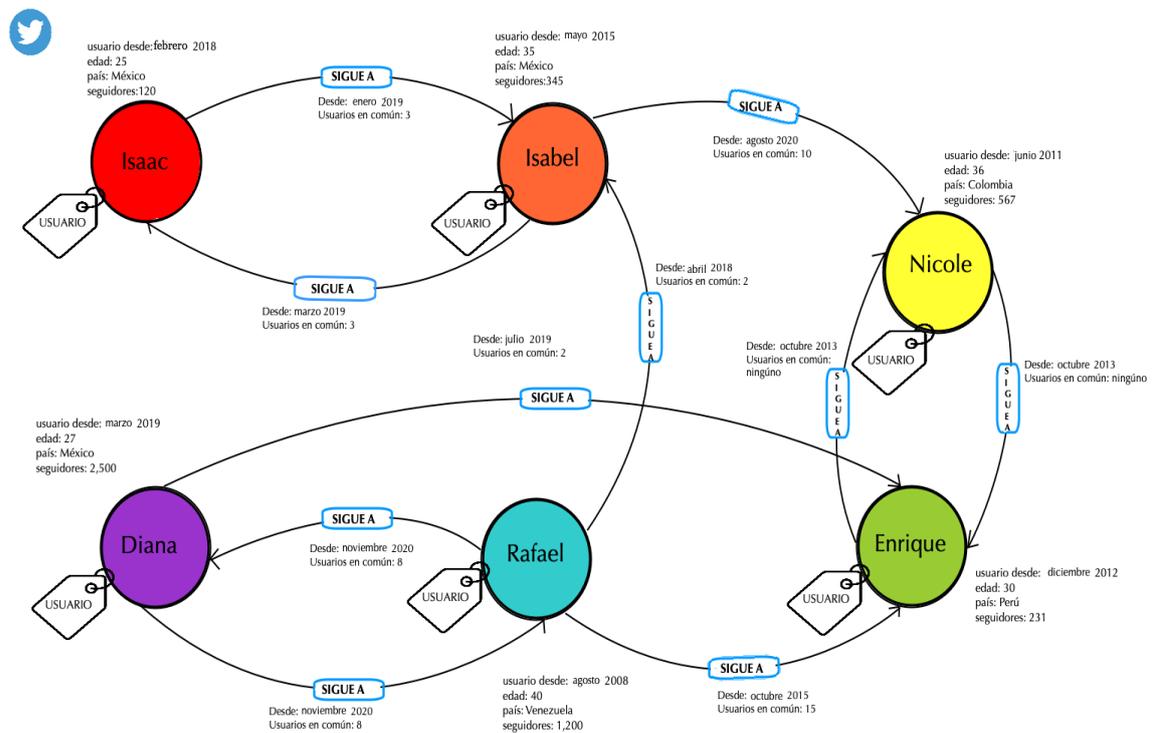


Figura 5.3: Ejemplo de *Labelled property graph model* en la red social Twitter.

En la figura 5.3 hay un ejemplo de la red social Twitter, donde los nodos tienen sus etiquetas que representan usuarios, además, se tienen otros atributos, como la

edad, país, número de seguidores, al igual en las aristas hay ciertos atributos los cuales sirven para poder hacer una mejor búsqueda de datos.

5.2.1. ¿Cuál es el secreto de *Neo4j*?

Con lo explicado anteriormente, se puede llegar a cuestionar, por qué es más rápido o eficiente *Neo4j*, pensando tal vez que éste usa un algoritmo veloz creado por los dueños, o que es producto de las tecnologías en que se basa, aunque después de todo está en java. El éxito se debe a la estructura de los datos, la naturaleza de las gráficas y la interconectividad de los datos, lo que permite hacerlo muy rápido [38].

Imaginemos que estamos en el Auditorio Nacional, en espera de un concierto de nuestro artista favorito y alguien nos pregunta: ¿Cuántas personas hay alrededor tuyo, a una distancia máxima de 10 metros? Intuitivamente empezaríamos a contar tan rápido como se pueda y acabaríamos en un tiempo considerable. Siguiendo el mismo orden de ideas, supongamos que estamos en un concierto en el Estadio Azteca, un lugar mucho más grande y con mayor capacidad de personas, de igual manera nos piden saber el número de personas que hay a nuestro alrededor a una distancia máxima de 10 metros, podemos intuir que nos tardaríamos en contar el mismo tiempo que tardamos en el Auditorio, ya que solo importa las personas a nuestro alrededor, no importan las personas que están del otro lado de la cancha ni en los palcos de arriba. Así funciona *Neo4j*, parte de un nodo inicial y visita los nodos conectados, sin importar si se van agregando más nodos a la gráfica. A diferencia de una base relacional que primero contaría todas las personas dentro del lugar y después descartaría una por una, las que no estén alrededor nuestro.

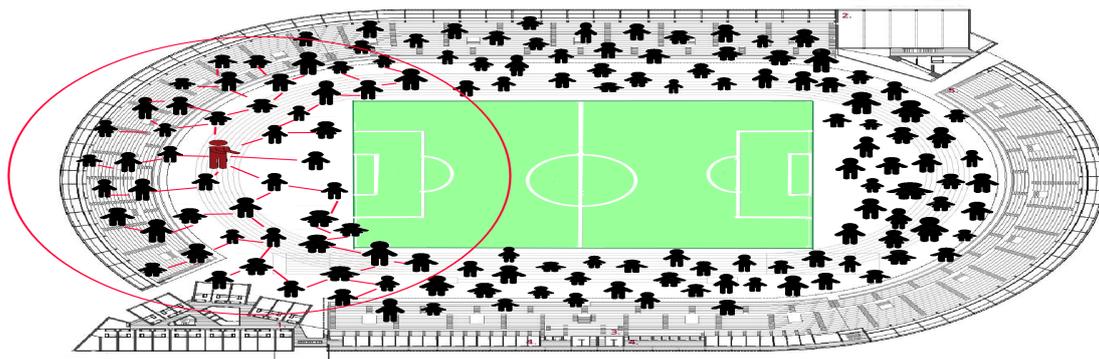


Figura 5.4: Ejemplo visual de como funciona *Neo4j*.

5.2.2. *Cypher*

Cypher es un lenguaje de consulta declarativo para gráficas. *Neo4j* utiliza el enfoque de gráfica con propiedades, donde las relaciones se almacenan junto con los datos en el modelo, y no se calculan hasta el momento de la consulta. Es un lenguaje que comprende y aprovecha estos elementos almacenados en conexiones. Al intentar encontrar patrones o conocimientos en los datos, las consultas cifradas suelen ser mucho más sencillas y más fácil de escribir, la naturaleza declarativa de coincidencia de patrones de *Cypher*, significa que puede consultar la gráfica, describiendo lo que necesita obtener de él [26]. Una ventaja que cabe mencionar, es que estas herramientas son desarrolladas en forma abierta, es decir, es código libre, que se puede descargar y manipular, sin ningún costo. A continuación se mostrará cómo crear una gráfica con *Cypher*.

Un ejemplo sencillo del uso de *Cypher*, en la figura 5.5, primero se crean los nodos con la etiqueta `singer` y `Group`, a cada nodo se le agregan sus propiedades, después se le agrega la relación que tendrán con la etiqueta de nombre `Miembrode` y también se le asigna una propiedad a esa relación, se pide que muestre la gráfica.

```
1 CREATE
2   (s:singer{name:"Chris Martin", born:1977}),
3   (g:Group{name:"Coldplay", genero:"Pop-Rock", exitos:"Yellow,Viva la vida, The scientist"})
4 WITH s, g
5 CREATE (s)-[p:Miembrode]-(g)
6 SET p.since=date("1996")
7 RETURN s, g, p
8
```

Figura 5.5: Código en *Cypher*.

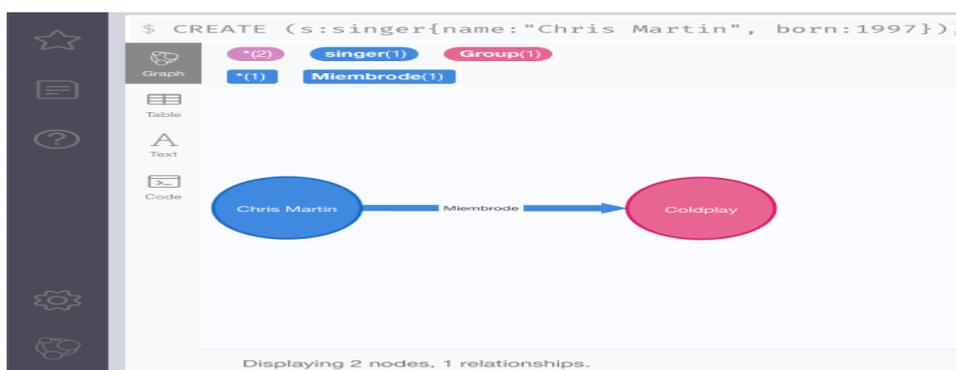


Figura 5.6: Gráfica hecha en *Neo4j*.

Una vez que *Neo4j* genera la gráfica, al dar clic en algunos de los nodos o en la flecha, lanza los demás datos de dicho elemento, como se observa en las figuras 5.7 y 5.8 donde está indicado con una flecha roja, teniendo que el cantante Chris Martin de nacionalidad británica, nacido en 1977 es miembro de la banda Coldplay desde 1996 que pertenece al género pop-rock, cuyos éxitos han sido Yellow, Viva la vida y The scientist.

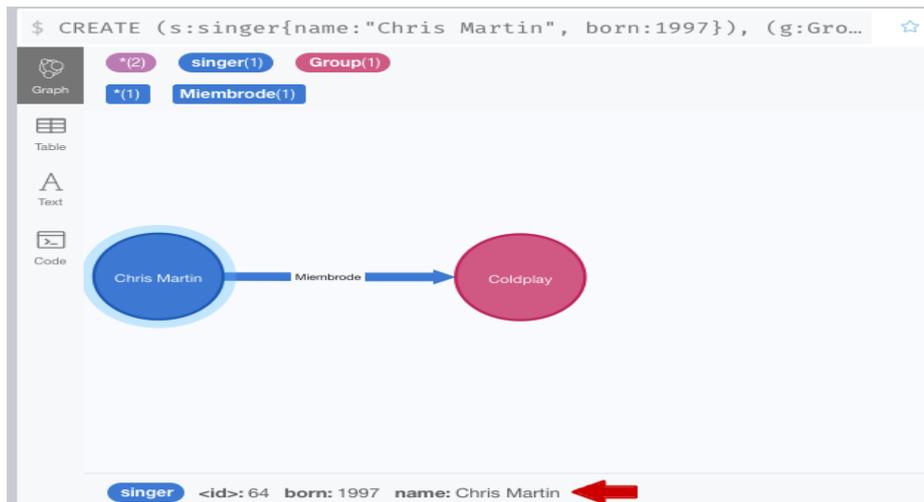


Figura 5.7: Gráfica en *Neo4j*.

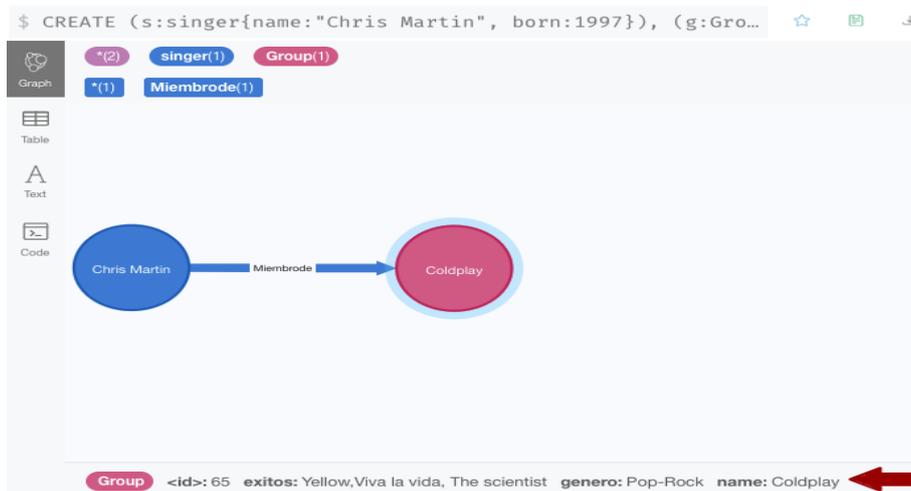


Figura 5.8: Gráfica en *Neo4j*.

5.3. Recomiéndame una película

En esta sección se verá una aplicación más directa de las bases de datos orientadas a gráficas, usando *Neo4j*. Las recomendaciones son algo que vemos en muchos lados, por ejemplo cuando se hace la compra de un producto en línea, al escuchar música, viendo un vídeo en YouTube, cuando Facebook nos recomienda amigos, páginas o viendo películas en alguna plataforma de Streaming.

Los algoritmos de recomendación establecen relaciones entre personas y cosas, ya sea otras personas, productos, servicios, contenido multimedia, etc. Las relaciones se establecen con base en los comportamientos del usuario cuando compra, reproduce, consume, califica o revisa los recursos en cuestión. El motor de recomendaciones puede identificar los recursos de interés para un individuo o grupo en particular o individuos y grupos que puedan tener algún interés en un recurso en particular [33].

La empresa **Mitflix** nos solicita hacer un sistema de recomendación a partir de un catálogo de películas en tiempo real, por lo que no solo se deberá considerar información histórica, sino la sesión actual del usuario. Hay una variedad de criterios para realizar una recomendación, para la aplicación que se plantea únicamente se usarán dos casos.

1. **Películas vistas por otros usuarios:** Se recomendará una película que el usuario no haya visto, pero que personas en una sesión parecida y con gustos similares ya vio y dio una buena calificación, esto acelera el acceso al contenido que el usuario posiblemente le gustará y así mejorará su experiencia.
2. **Género de película favorito:** Se recomendará al usuario las películas que no ha visto de su género favorito, el cual se puede deducir a partir del historial. Esta recomendación es oportuna para ampliar la exploración del usuario, permitiéndole conocer alternativas más adecuadas a las vistas habitualmente.

Lo primero será modelar la gráfica en *Neo4j*, para este ejemplo se hará con cuatro personas. En la figura 5.9 se puede observar, la gráfica que conecta a los usuarios con películas que ya vieron y el género al que pertenecen, en total aparecen 19 nodos y 32 relaciones entre ellos, la cual le toma a *Neo4j* hacerla en tan solo 6 milisegundos.

Código en Cypher

```

// crear nodos de los usuarios, películas y géneros
CREATE
(p1:Person {name:"Isabella Swan"}),
(p2:Person {name:"Jacob Black"}),
(p3:Person {name:"Alice Hale"}),
(p4:Person {name:"Edward Cullen"}),

(f1:film {title:"Guerra de los mundos"}),
(f2:film {title:"Spiderman"}), (f3:film {title:"Jurassic World"}),
(f4:film {title:"Harry Potter"}), (f5:film {title:"Fragmentado"}),
(f6:film {title:"Dragón Rojo"}), (f8:film {title:"El hoyo"}),
(f7:film {title:"Guerra Mundial Z"}), (f9:film {title:"Matrix"}),
(f10:film {title:"Rápidos y furiosos 5"}),
(f11:film {title:"Godzilla II"}), (f12:film {title:"X-Men"}),

(g1:Genero {name:"Ciencia ficción"}),
(g2:Genero {name:"Suspenso"}),
(g3:Genero {name:"Acción"}),

//crear las relaciones del género que pertenece cada película
y las películas que han visto los usuarios

(f1)-[:BELONG_TO]->(g1), (f2)-[:BELONG_TO]->(g1),
(f3)-[:BELONG_TO]->(g1), (f4)-[:BELONG_TO]->(g1),
(f5)-[:BELONG_TO]->(g2), (f6)-[:BELONG_TO]->(g2),
(f7)-[:BELONG_TO]->(g2), (f8)-[:BELONG_TO]->(g2),
(f9)-[:BELONG_TO]->(g3), (f10)-[:BELONG_TO]->(g3),
(f11)-[:BELONG_TO]->(g3), (f12)-[:BELONG_TO]->(g3),

(p1)-[:SEE]->(f1), (p1)-[:SEE]->(f2), (p1)-[:SEE]->(f3),
(p1)-[:SEE]->(f9), (p1)-[:SEE]->(f10), (p1)-[:SEE]->(f4),
(p2)-[:SEE]->(f5), (p2)-[:SEE]->(f9), (p2)-[:SEE]->(f7),
(p2)-[:SEE]->(f8), (p3)-[:SEE]->(f3), (p3)-[:SEE]->(f4),
(p3)-[:SEE]->(f7), (p3)-[:SEE]->(f10), (p3)-[:SEE]->(f8),
(p4)-[:SEE]->(f3), (p4)-[:SEE]->(f4), (p4)-[:SEE]->(f12),
(p4)-[:SEE]->(f8), (p4)-[:SEE]->(f10);

MATCH usuarios= ()-[:SEE]-() RETURN usuarios

```

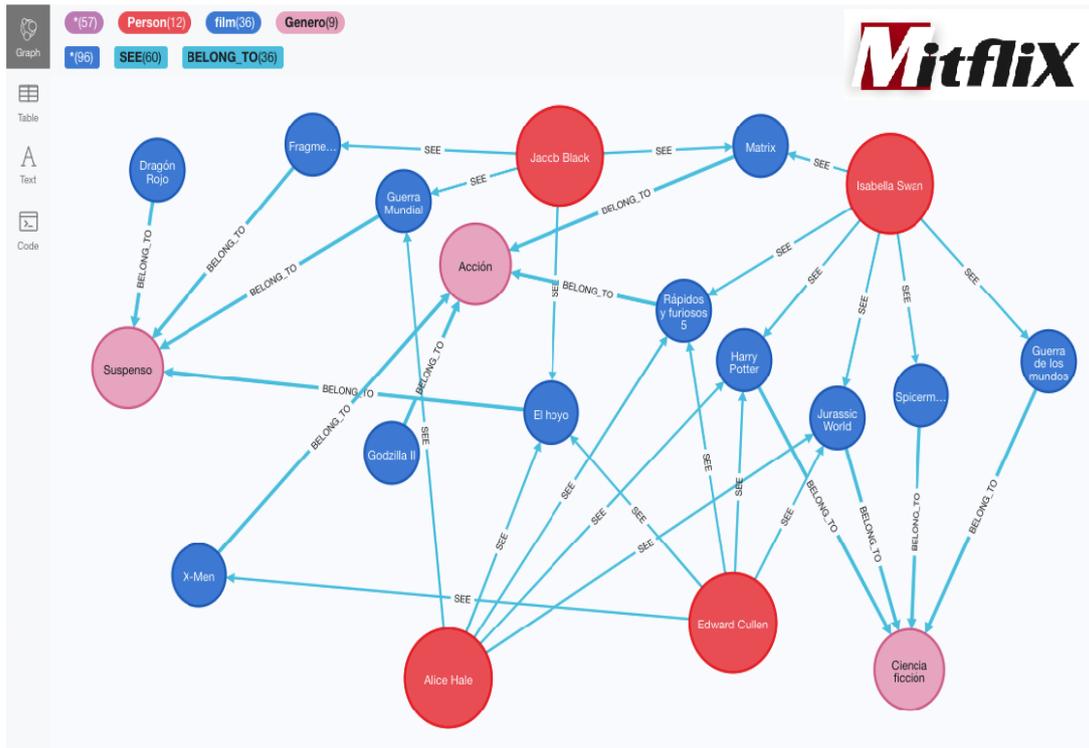


Figura 5.9: Gráfica en *Neo4j* para la empresa Mitflix.

Para el caso 1, la consulta se verá como en la figura 5.10 , por lo tanto, recomienda al usuario p1 todas las películas del usuario p2 que no se haya visto, ya que p1 y p2 pueden tener gustos parecidos por haber visto los dos la película film 1.

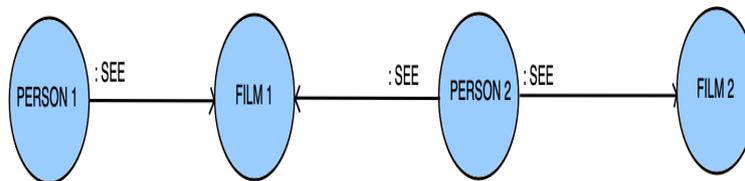


Figura 5.10: Patrón buscado en la gráfica.

Código en *Cypher* para consulta del caso 1

```
MATCH
(p1:Person)-[:SEE]->(f1:film)<-[:SEE]-(p2:Person)-[:SEE]->(f2:film)
WHERE not ((p1)-[:SEE]->(f2))
RETURN p1 as Persona1, p2 as Persona2, f2 as PeliculaRecomendada;
```

\$ MATCH (p1:Person)-[:SEE]->(f1:film)<-[:SEE]-(p2:Person)-[:SEE]...

Persona1	Persona2	PeliculasRecomendadas
{"name": "Edward Cullen"}	{"name": "Isabella Swan"}	{"title": "Matrix"}
{"name": "Edward Cullen"}	{"name": "Isabella Swan"}	{"title": "Spiderman"}
{"name": "Edward Cullen"}	{"name": "Isabella Swan"}	{"title": "Guerra de los mundos"}
{"name": "Alice Hale"}	{"name": "Isabella Swan"}	{"title": "Matrix"}
{"name": "Alice Hale"}	{"name": "Isabella Swan"}	{"title": "Spiderman"}
{"name": "Alice Hale"}	{"name": "Isabella Swan"}	{"title": "Guerra de los mundos"}
{"name": "Jacob Black"}	{"name": "Isabella Swan"}	{"title": "Harry Potter"}
{"name": "Jacob Black"}	{"name": "Isabella Swan"}	{"title": "Rápidos y furiosos 5"}
{"name": "Jacob Black"}	{"name": "Isabella Swan"}	{"title": "Jurassic World"}
{"name": "Jacob Black"}	{"name": "Isabella Swan"}	{"title": "Spiderman"}
{"name": "Jacob Black"}	{"name": "Isabella Swan"}	{"title": "Guerra de los mundos"}
{"name": "Edward Cullen"}	{"name": "Jacob Black"}	{"title": "Guerra Mundial Z"}
{"name": "Edward Cullen"}	{"name": "Jacob Black"}	{"title": "Matrix"}
{"name": "Edward Cullen"}	{"name": "Jacob Black"}	{"title": "Fragmentado"}
{"name": "Alice Hale"}	{"name": "Jacob Black"}	{"title": "Matrix"}
{"name": "Alice Hale"}	{"name": "Jacob Black"}	{"title": "Fragmentado"}
{"name": "Alice Hale"}	{"name": "Jacob Black"}	{"title": "Matrix"}
{"name": "Alice Hale"}	{"name": "Jacob Black"}	{"title": "Fragmentado"}
{"name": "Isabella Swan"}	{"name": "Jacob Black"}	{"title": "El hoyo"}
{"name": "Isabella Swan"}	{"name": "Jacob Black"}	{"title": "Guerra Mundial Z"}
{"name": "Isabella Swan"}	{"name": "Jacob Black"}	{"title": "Fragmentado"}
{"name": "Edward Cullen"}	{"name": "Alice Hale"}	{"title": "Guerra Mundial Z"}
{"name": "Jacob Black"}	{"name": "Alice Hale"}	{"title": "Rápidos y furiosos 5"}

Figura 5.11: Algunos resultados de recomendaciones.

Esta búsqueda se puede mejorar ya que a veces el tener un gusto en común con otra persona no garantiza coincidir en otro gusto, por lo que se puede cambiar el código de búsqueda y hacer que tengan al menos dos películas en común que los dos usuarios hayan visto.

Código en *Cypher* para consulta del caso 1

```
MATCH
(p1:Person)-[:SEE]->(f1:film)<-[:SEE]-(p2:Person)-[:SEE]->(f2:film)
WITH p1, p2, count(f1) as NPeliculas, collect(f1)
as PeliculasCompartidas, f2
WHERE not (p1)-[:SEE]->(f2) and NPeliculas >=2
RETURN p1.name as Persona1, p2.name as Persona2,
extract(x in PeliculasCompartidas | x.title)
as PeliculasCompartidas, f2 as PeliculaRecomendada;
```

"Persona1"	"Persona2"	"PeliculasCompartidas"	"PeliculaRecomendada"
"Edward Cullen"	"Isabella Swan"	["Harry Potter", "Rápidos y furiosos 5", "Jurassic World"]	{"title": "Matrix"}
"Edward Cullen"	"Isabella Swan"	["Harry Potter", "Rápidos y furiosos 5", "Jurassic World"]	{"title": "Spiderman"}
"Edward Cullen"	"Isabella Swan"	["Harry Potter", "Rápidos y furiosos 5", "Jurassic World"]	{"title": "Guerra de los mundos"}
"Alice Hale"	"Isabella Swan"	["Harry Potter", "Rápidos y furiosos 5", "Jurassic World"]	{"title": "Matrix"}
"Alice Hale"	"Isabella Swan"	["Harry Potter", "Rápidos y furiosos 5", "Jurassic World"]	{"title": "Spiderman"}
"Alice Hale"	"Isabella Swan"	["Harry Potter", "Rápidos y furiosos 5", "Jurassic World"]	{"title": "Guerra de los mundos"}
"Alice Hale"	"Jacob Black"	["El hoyo", "Guerra Mundial Z"]	{"title": "Matrix"}
"Alice Hale"	"Jacob Black"	["El hoyo", "Guerra Mundial Z"]	{"title": "Fragmentado"}
"Edward Cullen"	"Alice Hale"	["El hoyo", "Rápidos y furiosos 5", "Harry Potter", "Jurassic World"]	{"title": "Guerra Mundial Z"}
"Jacob Black"	"Alice Hale"	["El hoyo", "Guerra Mundial Z"]	{"title": "Rápidos y furiosos 5"}
"Jacob Black"	"Alice Hale"	["El hoyo", "Guerra Mundial Z"]	{"title": "Harry Potter"}
"Jacob Black"	"Alice Hale"	["El hoyo", "Guerra Mundial Z"]	{"title": "Jurassic World"}
"Isabella Swan"	"Alice Hale"	["Rápidos y furiosos 5", "Harry Potter", "Jurassic World"]	{"title": "El hoyo"}
"Isabella Swan"	"Alice Hale"	["Rápidos y furiosos 5", "Harry Potter", "Jurassic World"]	{"title": "Guerra Mundial Z"}
"Alice Hale"	"Edward Cullen"	["Rápidos y furiosos 5", "El hoyo", "Harry Potter", "Jurassic World"]	{"title": "X-Men"}
"Isabella Swan"	"Edward Cullen"	["Rápidos y furiosos 5", "Harry Potter", "Jurassic World"]	{"title": "El hoyo"}
"Isabella Swan"	"Edward Cullen"	["Rápidos y furiosos 5", "Harry Potter", "Jurassic World"]	{"title": "X-Men"}

Figura 5.12: Resultados de posibles recomendaciones.

Para el caso 2, se hace la consulta la cual busca el género que más se repite entre las películas del usuario, recomienda las que pertenezcan a ese género y que no se hayan visto antes.

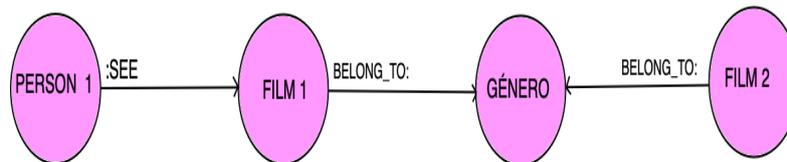


Figura 5.13: Patrón buscado en la gráfica.

Código en *Cypher* para consulta del caso 2

```

MATCH
(p:Person)-[:SEE]->(f1:film)-[:BELONG_TO]->(g:Genero)
<-[:BELONG_TO]->(f2:film)
WITH p,g,count(f1) as Generos,f2
WHERE NOT ((p)-[:SEE]->(f2)) AND Generos>1
RETURN p.name as Persona, g.name as Genero, collect(f2.title) as
PeliculasRecomendadas;
  
```

\$ MATCH (p:Person)-[:SEE]->(f1:film)-[:BELONG_TO]->(g:Genero) ...

"Persona"	"Genero"	"PeliculasRecomendadas"
"Edward Cullen"	"Ciencia ficció'n"	["Spiderman","Guerra de los mundos"]
"Alice Hale"	"Ciencia ficció'n"	["Spiderman","Guerra de los mundos"]
"Alice Hale"	"Suspenseo"	["Drag'ón Rojo","Fragmentado"]
"Jacob Black"	"Suspenseo"	["Drag'ón Rojo"]
"Edward Cullen"	"Acció'n"	["Godzilla II","Matrix"]
"Isabella Swan"	"Acció'n"	["X-Men","Godzilla II"]

Figura 5.14: Resultados de recomendaciones por género.

Esto sugiere que si se juntan varios tipos de recomendaciones, es muy probable que el resultado final sea muy compatible con el usuario. Finalmente, se ha visto en este capítulo algunas ventajas, usos de las bases de datos orientadas a gráficas y los múltiples beneficios que tienen, cabe mencionar que es importante saber el tipo de problema que se tiene; los datos a manipular y los resultados que se quieren obtener, para así decidir qué tipo de base usar y cuál es más conveniente.

Las bases de datos orientadas a gráficas y los múltiples algoritmos de gráficas, han evolucionado a un ritmo impresionante junto con el *big data*, conforme las empresas y organizaciones continúan con la explotación de datos, se ven en la necesidad de hacer consultas que llegan a ser muy complejas, de manera que las bases de datos orientadas a gráficas son una gran herramienta para cubrir las necesidades actuales y los éxitos del mañana.

Conclusiones

Todo lo expuesto anteriormente permite concluir lo siguiente, la *Teoría de gráficas* tuvo un origen modesto a comparación de otras ramas de las matemáticas, ya que no estaba enfocada a resolver problemas de medición, cálculo, movimiento, sino más bien por la curiosidad de resolver ciertos acertijos, poniendo a prueba el ingenio. Esto logró captar la atención de matemáticos, que poco a poco fue formando una disciplina nueva, la cual se ha ido forjando a partir de problemas que se han presentado en la vida cotidiana, aquí se expusieron algunas posibles aplicaciones prácticas, lo cual me ayudo a comprobar que el campo de aplicación es tan ilimitado que podemos modelar situaciones y aplicaciones tantas como nuestra creatividad e ingenio lo permita.

Es una teoría muy intuitiva que logra hacer una conexión entre nuestros pensamientos e ideas para así poder plasmarlos en algo que va más allá de la mente. A raíz de nuevos descubrimientos, se ha estimulado la investigación de nuevas aplicaciones en gráficas, puesto que la sociedad lo demanda, ya sea en medicina, psicología, economía, computación, entre otras. Sin duda es una rama que vale la pena conocer, explorar, aprender, sacar ese espíritu matemático de aceptar retos, pensar, preguntarse y solucionar problemas. No cabe duda que las aplicaciones matemáticas se presentan como un cúmulo de conocimientos y de nosotros depende darle sentido.



Bibliografía

- [1] BERMEO, A. Alerta mandatario hace dos meses riesgo de inundaciones en el Valle de México. *Reforma* (junio del 2007). Consultado el 18 de diciembre del 2020.
- [2] BIGGS, N., LLOYD, E.K., Y WILSON, R.J. *Graph Theory, 1736-1936*. Oxford University Press, 1986.
- [3] BONDY, J.A., Y MURTY, U.S.R. *Graph Theory with Applications*. American Elsevier, New York, 1976.
- [4] BURNS, J. El gran hedor: la ola de calor que tornó a Londres en una alcantarilla pestilente. *BBC News* (2018).
- [5] CÁCERES, R. A. *Estadística aplicada a las ciencias de la salud*. Ediciones Díaz de Santos, 2007.
- [6] CHARTRAND, G. *Introductory graph theory*. Courier Corporation, 1977.
- [7] CHRISTAKIS, N. A. Red social alumnos de harvard. <https://doi.org/10.1371/journal.pone.0012948.g004>. Consultado el 12 de febrero de 2021.
- [8] CHRISTAKIS, N., Y F. J. Social network sensors for early detection of contagious outbreaks. *PloS One* 5, 9 (2010).
- [9] CHRISTAKIS, N., Y FOWLER, J.H. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Hachette Book Group, 2009.
- [10] ESPINOZA, R. Y CUZCO, B. Análisis comparativo entre bases de datos relacionales con bases de datos no relacionales, 2013.
- [11] FELD, S. Why your friends have more friends than you do? *American Journal of Sociology* (1991).
- [12] FIDEL, R. Database design for information retrieval: a conceptual approach. New York, 1987.

-
- [13] GOLDFEDER, I. A. Hacia las gráficas: Una introducción básica. Instituto de Matemáticas, 2014.
- [14] GOLDFEDER, I. A. A través de las gráficas: Una introducción a la Teoría de las Gráficas con énfasis en la inducción. online, Noviembre 2018.
- [15] GONZÁLEZ, D. Introducción a la teoría de las gráficas. Departamento de Matemáticas Aplicadas y Sistemas, UAM-Cuajimalpa, 2017.
- [16] GROSS, L. J., Y YELLEN, J. *Graph theory and its applications*. Kenneth H. Rosen, 2006.
- [17] HILLIER FREDERICK S., Y LIEBERMAN GERALD J. *Introducción a la investigación de operaciones*, 9na ed. Mc Graw Hill, 2010.
- [18] LEANDRO, M. The path to the state as a form of political social organization. *Aequitas*, 11 (2018).
- [19] MACPHERSON, A. ¿Por qué faltan vacunas? *La Vanguardia* (marzo del 2017).
- [20] MALCOLM, D.G., ROSEBOOM, J.H., CLARK, C.E., Y FAZAR, W. Application of a technique for research and development program evaluation, operations Research, 1959.
- [21] MATHEMATICS. Tree : A tree is a connected, acyclic, undirected graph. <https://web.iitd.ac.in/~bspanda/Tree.pdf>. Consultado el 2 de abril del 2020.
- [22] MCNULTY, K. Graph theory helped the British become less stinky. *Towards Data Science* (2018).
- [23] MORALES, M. El teorema de los cuatro colores: la teoría de grafos al servicio del coloreado de mapas. Gaussianos, 2013. Consultado el 21 de marzo del 2021.
- [24] NEO TECHNOLOGY. What is neo4j? <https://neo4j.com>. Consultado el 16 de enero del 2021.
- [25] NEO TECHNOLOGY. eBay and Walmart adopt Neo4j: The graph is transforming retail. Consultado el 18 de enero del 2021.
- [26] NEO4J.INC. The neo4j operations manual v4.2. <https://neo4j.com>. Consultado el 18 de enero del 2021.
- [27] NEUMANN, M. C. Euler y la geometría de la posición. *Miscelánea matemática* 45: 87-96 (2007).

- [28] NUÑO, A. ¿extraterrestres haciendo pirámides? una egiptóloga explica qué se sabe realmente de su construcción. *El confidencial, alma corazón vida* (marzo del 2021). Consultado el 28 de marzo del 2021.
- [29] PATRICK, H. Las matemáticas de las epidemias (y de las vacunas). *Investigación y ciencia* (mayo, 2020).
- [30] PORTAL ESTADÍSTICA APLICADA. Gestión de proyectos: redes PERT-CPM. <https://www.estadistica.net/INVESTIGACION/PERT-CPM.pdf>. Consultado el 13 de junio del 2020.
- [31] PUNCH MAGAZINE. *The silent highwayman*, vol. 35. 1858.
- [32] RENDÓN, Y. Bases de datos relacionales vs. no relacionales, mayo del 2019.
- [33] ROBINSON, I., Y WEBBER, J. *Graph databases, new oportunities for conected data*, 2da ed. O'Reilly, 2015.
- [34] THE TIMES. Michael Faraday's letter observations on the filth of the Thames, 1855.
- [35] THOMAS, R. An update on the four-color theorem. *Notices of the AMS* 45, 7 (1998), 848–859.
- [36] TREJO, E. Introducción a la teoría de digráficas. Tesis no publicada, 2013.
- [37] VARGAS, W. V. Bases de datos orientadas a grafos y su enfoque en el mundo real.
- [38] VUKOTIC, A., WATT, N., ABEDRABBO, T., FOX, D. Y PARTNER, J. *Neo4j in action*, vol. 22. Manning Shelter Island, 2015.
- [39] WIKIBOOKS. Prehistoria,paleolítico,sociedad primitiva. https://es.wikibooks.org/wiki/Prehistoria/Paleolitico/Sociedad_primitiva, 2019. Consultado el 20 de mayo del 2020.
- [40] WIKIPEDIA. Imagen ejemplos de gráficas completas. https://es.wikipedia.org/wiki/Grafo_completo, consultado el 8 de marzo del 2020.
- [41] XYGALATAS, D. Nickolas a. christakis and james h. fowler (2009), connected: The surprising power of our social networks and how they shape our lives, little, brown, new york. 353 pages. *Journal of Cognition and Culture* 10 (01 2010), 3–4.