

Proyecto académico final: Caso de estudio de ciencia de redes

Vargas Naula, Francisco Oswaldo
Maestría en Ciencias y Tecnologías de la Computación
Universidad Técnica Particular de Loja
Loja, Ecuador
fovargas@utpl.edu.ec

Resumo—En este manuscrito, se presenta un caso de estudio basado en los conceptos fundamentales de la Ciencia de Redes. El foco se centra en el proceso de captación de estudiantes de la Universidad Técnica Particular de Loja (UTPL). Se analizan y describen las interacciones y características relacionadas con colegios, carreras y estudiantes, buscando comprender la complejidad inherente en el proceso de matrícula de estudiantes de primer ingreso universitario.

Palabras clave—redes, grafos, nodos, enlaces, captación, gephi, igrph, modelos

I. INTRODUCCIÓN

La Universidad Técnica Particular de Loja (UTPL) es una destacada institución educativa en Ecuador, que ofrece una amplia oferta de carreras en modalidades tanto presencial como a distancia. Sin embargo, el presente caso de estudio se enfocará exclusivamente en la modalidad presencial (Campus UTPL Loja). Se examinará la interacción y relaciones entre estudiantes, colegios, carreras y ciudades durante los últimos tres periodos de matrícula.

Preliminarmente, se ha observado que la mayoría de los estudiantes provienen de colegios ubicados en la Zona 7 del Ecuador, que comprende las provincias de Loja, Zamora y El Oro. Este análisis brinda la oportunidad no solo de confirmar este dato, sino también de identificar qué colegios específicos de esta zona tienen la mayor representación en términos de matriculados. Gracias a la Ciencia de Redes, se buscará obtener una comprensión más profunda de las dinámicas y estructuras involucradas en el proceso de captación de la UTPL [1].

El manuscrito se estructura en tres secciones. La primera ofrece una introducción al tema del documento. En la segunda sección, se discuten los conceptos clave abordados en la asignatura que han fundamentado el desarrollo del caso de estudio. Estos incluyen métricas de redes, grados de Euler y Hamilton, visualizaciones con herramientas como Gephi e iGraph para R, modelos de red y la teoría de los seis grados de separación.

II. DESARROLLO DEL PROYECTO

El proceso de construcción del caso de estudio se llevó a cabo en base a los contenidos que se abordaron semanalmente en la asignatura de Ciencia de Redes. En cada semana, se exploraron diferentes conceptos, herramientas y técnicas que

posteriormente se aplicaron para analizar y modelar la red de captación. En las siguientes subsecciones, se detalla cómo cada uno de estos temas contribuyó al desarrollo y comprensión de la red, permitiendo una visión integral y profunda de las interacciones y estructuras presentes.

A. Identificación de redes

Para el análisis de la red, se han identificado varios tipos de nodos y enlaces que representan las distintas entidades y sus interacciones. A continuación, se presenta una descripción detallada de estas entidades y sus propiedades.

1) **Nodos**: Se presentan seis tipos de nodos en la red. La Tabla I muestra cada tipo de nodo junto con una breve descripción.

TABLE I
TIPOS DE NODOS Y SU DESCRIPCIÓN

Nodo	Descripción
Estudiantes	Individuos que ingresan a una carrera en UTPL
Carreras	Programas educativos ofrecidos por UTPL
Colegios	Instituciones de educación secundaria de origen de los estudiantes
Cantones	Divisiones administrativas dentro de provincias
Provincias	Divisiones administrativas principales en Ecuador
Países	Naciones de origen de estudiantes internacionales

2) **Propiedades**: Cada nodo tiene asociadas ciertas propiedades que ofrecen información adicional sobre la entidad. La Tabla II lista las propiedades, a qué nodo pertenecen y una breve descripción.

TABLE II
PROPIEDADES, NODOS A LOS QUE PERTENECEN Y SU DESCRIPCIÓN

Propiedad	Pertenece a	Descripción
Nombres	Estudiantes	Nombres del estudiante
Apellidos	Estudiantes	Apellidos del estudiante
Ciudad	Colegios	Ciudad a la que pertenece el colegio
id	Todos los nodos	Identificador único del nodo
label	Todos los nodos	Nombre o etiqueta representativa del nodo
Tipo de nodo	Todos los nodos	Clasificación del nodo (e.g., estudiante, carrera)
Source	Todas las aristas	Nodo de origen del enlace
Target	Todas las aristas	Nodo destino del enlace
Tipo de enlace	Todas las aristas	Clasificación del enlace (e.g., enrolamiento)

Adicionalmente, es esencial mencionar que el análisis se centrará en los registros correspondientes a 1,611 matrículas de estudiantes nuevos. Estos datos provienen de los últimos procesos de captación de la Universidad Técnica Particular de Loja (UTPL) para la modalidad presencial. La consideración de esta muestra específica permitirá obtener una visión actualizada y representativa de las dinámicas de enrolamiento en la institución.

El análisis de esta red puede ayudar a identificar los colegios más influyentes, evaluar la efectividad de las estrategias de colaboración, comprender cómo los colegios recomiendan y seleccionan a los estudiantes, y mejorar los procesos de captación.

B. Métricas sobre la red

Una de las tareas fundamentales en el análisis de redes es la determinación de métricas que proporcionen una visión cuantitativa de las estructuras y relaciones presentes en la red. Estas métricas permiten comprender mejor las características, comportamientos y dinámicas de la red [3].

Comenzando con el **grado del nodo**, este representa el número de enlaces que un nodo tiene con otros nodos en la red. Si modelamos la relación entre carreras y colegios como una red completamente conectada, tendríamos que para cada carrera existirían 192 enlaces correspondientes a los 192 colegios y viceversa, para cada colegio, existirían 29 enlaces que representan las 29 carreras ofertadas por la universidad. Es importante mencionar que este tipo de configuración nos sitúa en el contexto de redes **bipartitas**, donde dos clases de nodos se enlazan entre sí, pero no dentro de su propia clase.

A continuación, presentamos una tabla con las métricas de red que se calcularán para el caso de estudio:

TABLE III
MÉTRICAS PRINCIPALES DE LA RED

Métrica	Descripción
Grado del nodo	Número de enlaces de un nodo con otros nodos
Grado promedio	Promedio de enlaces de todos los nodos en la red
Distribución de grado	Frecuencia de nodos para cada valor de grado
Matriz de adyacencia	Representación binaria de las conexiones entre nodos

La definición de las métricas de red tomó como base las lecciones adquiridas del famoso problema de los puentes de Königsberg estudiado en la asignatura. Aunque al principio no se establecen limitaciones para que los estudiantes de ciertos colegios se inscriban en carreras específicas, el análisis de datos reales revelará los grados de los nodos, identificando así colegios y carreras potenciales para optimizar estrategias de captación.

C. Relación con los grados de Euler y de Hamilton

Los caminos en la red representan las rutas que los estudiantes toman desde sus colegios hasta las carreras universitarias. Estas trayectorias son cruciales porque nos permiten comprender las inclinaciones de los estudiantes y cómo se distribuyen

entre las distintas carreras. Además, se puede conceptualizar la distancia dentro de la red de la siguiente manera: La distancia en la red puede interpretarse como la cantidad de pasos que un estudiante debe dar para transitar desde su colegio hacia una carrera específica.

Las definiciones fundamentales en este contexto son:

- **Grafo Euleriano:** Un grafo en el que existe un ciclo cerrado que pasa por cada arista exactamente una vez. [8]
- **Grafo Hamiltoniano:** Un grafo en el que existe un ciclo cerrado que pasa por cada vértice exactamente una vez. [9]

Para el caso de estudio:

- **Grafo Euleriano:** La red no califica como un grafo Euleriano. Para que un grafo sea Euleriano, cada vértice debe tener un grado par. En nuestra red, esto no sucede ya que cada colegio está vinculado a un número específico de carreras y viceversa. No hay certeza de que estos números sean pares.
- **Grafo Hamiltoniano:** La red tampoco se ajusta a la definición de un grafo Hamiltoniano. Para ser Hamiltoniano, debe existir un ciclo que visite cada vértice exactamente una vez. En la red, esto no es factible ya que los estudiantes no regresan a su colegio después de elegir una carrera. Aunque técnicamente es posible hacer la red Hamiltoniana agregando aristas adicionales que conecten todos los colegios y carreras, este enfoque no tiene sentido en el contexto de nuestro análisis y podría distorsionar las interpretaciones.

D. Relación con Isomorfismo, Conexión y Agrupamiento

A continuación, se presenta el análisis de la red con estos conceptos:

Isomorfismo: En la Ciencia de Redes, dos grafos se dicen isomorfos si hay una correspondencia uno a uno entre sus conjuntos de vértices y entre sus conjuntos de aristas de tal manera que la correspondencia conserva la relación de adyacencia [10]. En el caso de nuestro estudio, si tomamos dos redes de captación de diferentes periodos y estas resultan ser isomorfas, implicaría que la estructura de la relación entre colegios y carreras ha permanecido exactamente igual en ambos periodos, aunque los estudiantes y sus elecciones hayan cambiado.

Conexión: Una red es conexa si existe un camino entre cada par de nodos [11]. Para la UTPL, esto significa que desde cualquier colegio es posible llegar a cualquier carrera a través de una secuencia de relaciones. Si la red no fuera conexa, habría colegios o carreras "aisladas", lo que podría señalar falta de información o problemas en el proceso de captación.

Agrupamiento: Se refiere a la tendencia de los nodos a agruparse o formar comunidades densamente conectadas [12]. En el contexto de la UTPL, un alto coeficiente de agrupamiento podría sugerir que existen grupos de colegios cuyos estudiantes tienden a preferir las mismas carreras.

La representación de la lista de enlaces provee una visión simplificada de las conexiones, pero no captura de manera

inmediata las conexiones indirectas entre nodos o la ausencia de conexiones. En cambio, la matriz de adyacencia proporciona una descripción completa y estructurada de todas las conexiones entre nodos, permitiendo identificar rápidamente caminos y subgrafos específicos.

Para determinar cuántos caminos de longitud 3 existen entre dos nodos seleccionados en la red, uno podría elevar al cubo la matriz de adyacencia y luego examinar la entrada correspondiente a esos dos nodos. Esta entrada indicaría el número de dichos caminos. Contar ciclos de longitud 4 es más complejo y generalmente involucra el uso de algoritmos específicos o la descomposición de la matriz de adyacencia.

La matriz de adyacencia es esencial para resolver y comprender muchos de estos conceptos, ya que provee una representación integral y matricial de las conexiones en la red, facilitando cálculos y análisis.

E. Representación de la red con Gephi

Para visualizar y analizar la red de relaciones entre colegios y carreras, se utilizó la herramienta Gephi. Esta herramienta es ampliamente reconocida en la comunidad de Ciencia de Redes por su capacidad de procesamiento y representación gráfica de redes complejas [5]. A continuación, se presentan las métricas principales que Gephi nos ofreció para la red en estudio:

TABLE IV
MÉTRICAS DE UNA DE LAS REDES REPRESENTADAS CON GEPHI

Métrica	Valor	Descripción
Grado medio	6,335	Promedio de conexiones de cada nodo.
Diámetro de la red	1	Distancia más larga entre dos nodos en la red.
Densidad del grafo	0,015	Fracción de conexiones presentes en relación con las posibles conexiones.
Componentes conexos	1	Número de subredes donde cada nodo está conectado a al menos otro nodo.
Modularidad	0,262	Medida de la estructura de comunidad en la red.
Coefficiente medio de clustering	0	Media del coeficiente de clustering de todos los nodos en la red.
Longitud media del camino	1	Distancia promedio entre cualquier par de nodos en la red.
Número de comunidades	9	Número de grupos en los que la red puede dividirse basándose en la conectividad.

El proceso para representar las redes en Gephi inició con una fase importante de procesamiento de datos. Esta fase es fundamental para asegurar la correcta representación y análisis posterior de la red. Utilizamos la herramienta *Openrefine* para limpiar y transformar los datos [13]. Este software es conocido por su robustez y versatilidad en tareas de procesamiento y limpieza de datos, permitiendo gestionar grandes cantidades de información de manera eficiente.

Una vez que se tuvo una base de datos limpia y estructurada, se generaron archivos que contenían listados de nodos y aristas, adaptados al formato aceptado por Gephi. Específicamente, se creó un listado de nodos y aristas para cada tipo de enlace previamente definido entre los nodos.

Con estos archivos preparados, se procedió a la etapa de visualización en Gephi. La herramienta posee varias opciones de distribución y diseño que permiten una representación clara y significativa de la red. Se hizo uso del complemento especializado para visualizar redes bipartitas, lo cual facilitó la comprensión de las relaciones entre los distintos nodos, en nuestro caso, colegios y carreras [14].

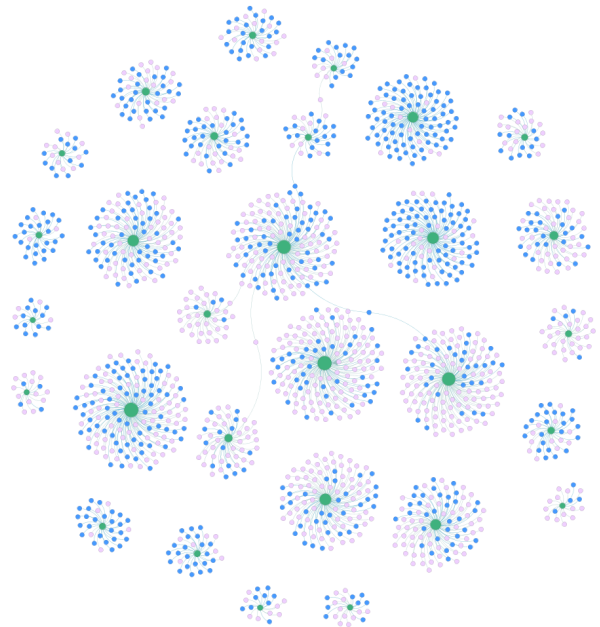


Fig. 1. Representación de red de estudiantes y carreras

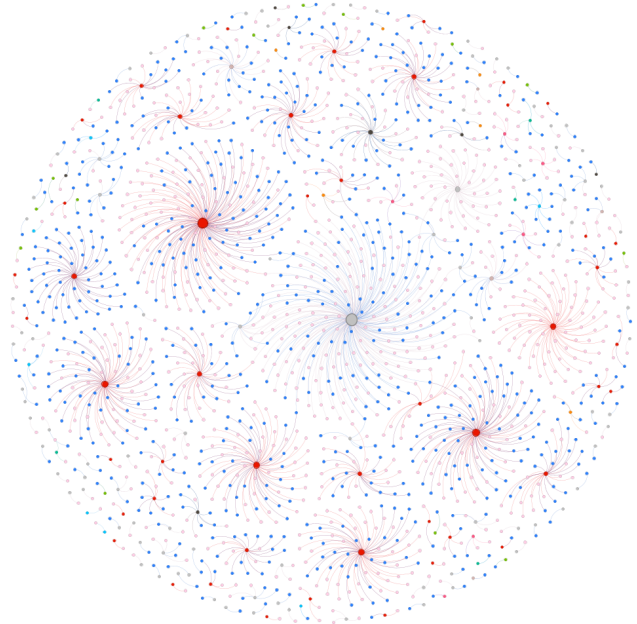


Fig. 2. Representación de red de estudiantes y colegios

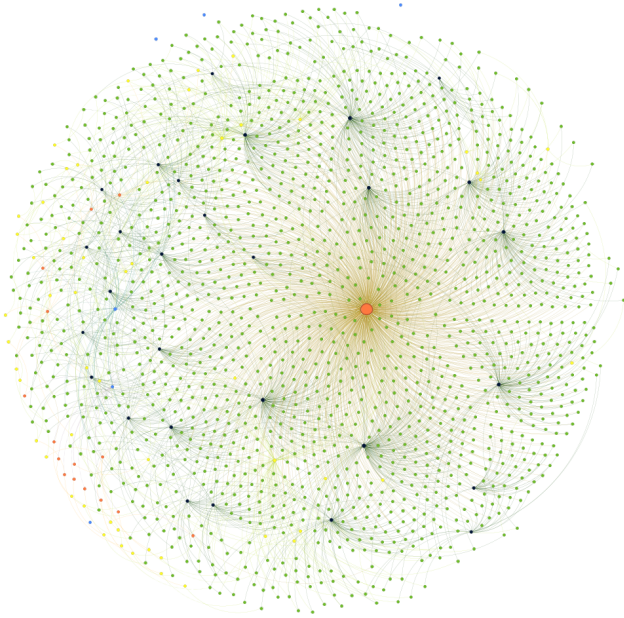


Fig. 3. Representación de red de estudiantes, carreras y cantones de residencia

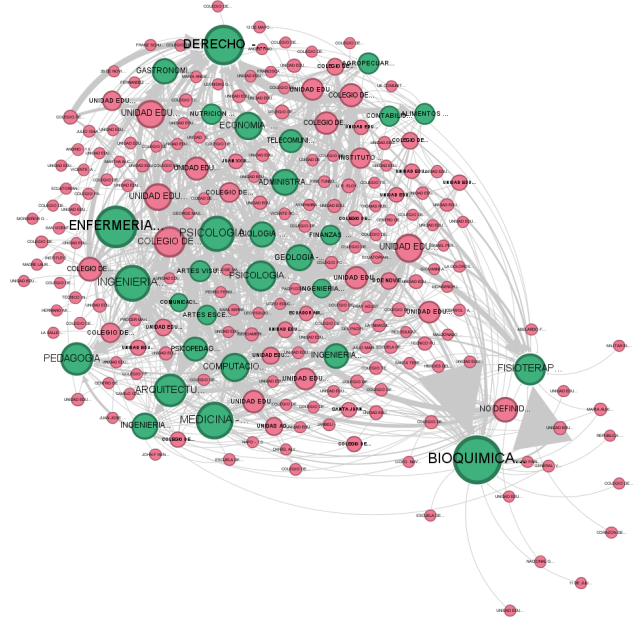


Fig. 5. Representación de red de colegios y carreras

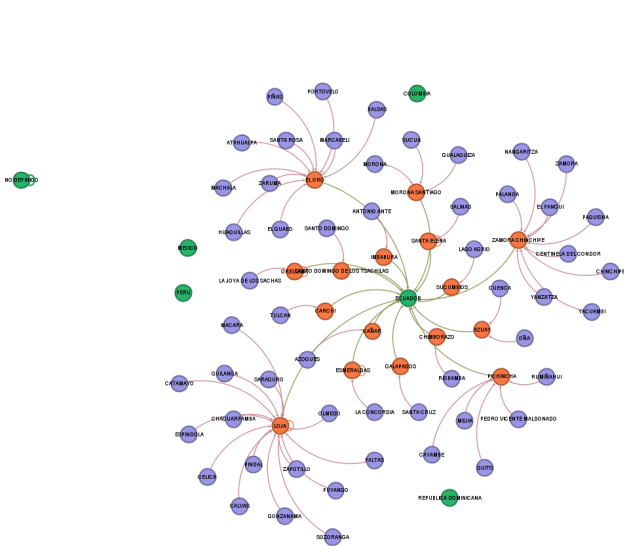


Fig. 4. Representación de red de cantones, provincias y países de los estudiantes



Fig. 6. Representación de red de colegios y carreras (Bipartita)

F. Resultados de Gephi

En el proceso de análisis, Gephi ofreció una serie de métricas derivadas de la red de colegios y carreras basada en el número de estudiantes como peso. Estas métricas proporcionan una visión detallada de la estructura y características de la red. A continuación, se presentan e interpretan las métricas

obtenidas:

- **Grado medio (3,394):** En promedio, cada nodo (ya sea colegio o carrera) en la red está conectado a aproximadamente 3,394 otros nodos. Esto sugiere que, en promedio, un colegio tiene estudiantes que han elegido alrededor de 3 o 4 carreras, o viceversa, una carrera recibe estudiantes de 3 o 4 colegios diferentes.

- **Grado medio con pesos (6,335):** Cuando se considera el peso (el número de estudiantes), la conexión entre colegios y carreras representa en promedio a 6,335 estudiantes. Esto sugiere que aunque hay aproximadamente 3 o 4 conexiones por nodo, el volumen de estudiantes entre esos nodos es significativo.
- **Diámetro de la red (1):** El diámetro indica la distancia más larga entre dos nodos en la red. Un diámetro de 1 sugiere que todos los nodos están directamente conectados entre sí. Esto implica que cualquier colegio está conectado directamente a cualquier carrera.
- **Densidad del grafo (0,015):** Esta métrica indica qué porcentaje de todas las conexiones posibles están presentes en la red. Un valor de 0,015 muestra que la red es bastante dispersa, es decir, solo el 1,5% de todas las posibles conexiones entre colegios y carreras existen.
- **Modularidad (0,262):** La modularidad refleja la estructura de comunidad dentro de la red. Un valor de 0,262 sugiere una estructura de comunidad moderada, lo que puede indicar que hay grupos de colegios y carreras que tienden a agruparse más entre sí que con otros nodos en la red.
- **Número de comunidades (9):** La red se puede dividir en nueve comunidades diferentes. Cada comunidad representa un conjunto de colegios y carreras que están más estrechamente interconectados entre sí que con los nodos de otras comunidades.

G. Representación de la red con iGraph

Adicionalmente a Gephi, también se realizó una representación de la red utilizando *iGraph* para R [15]. Aunque visualmente la representación no fue tan intuitiva y clara como en Gephi, debido a la densidad y cantidad de nodos, *iGraph* ofrece ventajas distintas.

iGraph es una biblioteca en R y Python que proporciona herramientas para la creación y análisis de redes complejas. Permite generar visualizaciones y calcular métricas relevantes para el estudio [6].

Una de las principales fortalezas de *iGraph* radica en su capacidad de calcular rápidamente estadísticas y métricas de la red, brindando un conjunto robusto de herramientas para el análisis de redes [16]. Asimismo, la flexibilidad que ofrece R, junto con las funcionalidades de *iGraph*, facilita la ejecución de análisis avanzados, abriendo un abanico amplio de posibilidades en cuanto a la interpretación y comprensión de las dinámicas y estructuras de la red.

H. Relación con las Características de los Tres Modelos de Red

Dentro de la teoría de redes, se han propuesto diversos modelos para entender y caracterizar las conexiones entre los nodos de una red. En la asignatura se han revisado al **Modelo de Red Aleatoria (Erdős-Rényi)**, el **Modelo de Mundo Pequeño (Watts-Strogatz)** y la **Red Libre de Escala (Barabási-Albert)**, los cuales se destacan en la literatura relacionada [17].

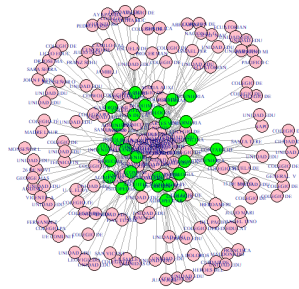


Fig. 7. Representación de red de colegios y carreras con iGraph para R

Considerando el caso de estudio, es evidente que algunos colegios y carreras son especialmente populares y atraen a un gran número de estudiantes. Sin embargo, también existen colegios y carreras que, por diversas razones, atraen a un número más reducido de estudiantes. Esta dinámica refleja una heterogeneidad en la distribución de grados, y el análisis muestra que esta distribución sigue una ley de potencia.

El principal rasgo de una **Red Libre de Escala** es precisamente una distribución de grado que sigue una ley de potencia, donde unos pocos nodos (llamados *hubs*) tienen muchos enlaces, mientras que la mayoría tiene solo unos pocos [18]. Considerando la evidencia y las características de la red en estudio, es razonable concluir que esta se asemeja a una Red Libre de Escala.

I. Relación con la Teoría de los Seis Grados de Separación

La **Teoría de los Seis Grados de Separación** postula que cualquier persona en el planeta puede estar conectada a cualquier otra persona a través de una cadena de conocidos que no tiene más de cinco intermediarios, haciendo un total de seis pasos o grados de separación [19]. La teoría, aunque se originó en el contexto de redes sociales humanas, ha encontrado aplicaciones en una amplia variedad de redes y sistemas, incluidos los sistemas académicos y organizacionales.

Dentro del caso de estudio que estamos analizando, la teoría puede aplicarse al examinar cómo los estudiantes de un colegio específico se conectan con carreras universitarias a través de redes de relaciones. Aunque en este contexto los seis grados de separación no se toman literalmente, la idea subyacente es que los caminos en la red universitaria pueden ser sorprendentemente cortos, y esto tiene implicaciones para cómo se distribuyen y fluyen la información, las preferencias y las tendencias entre los estudiantes y las instituciones.

Una implicación directa de esta teoría es que, si es cierto que estamos tan estrechamente conectados, las estrategias de marketing y promoción, por ejemplo, para atraer a estudiantes a ciertas carreras, pueden beneficiarse enormemente al dirigirse a nodos o individuos clave (o colegios en este caso) que actúan como intermediarios o puntos de conexión dentro de la red.

III. CONCLUSIONES Y TRABAJOS FUTUROS

IV. CONCLUSIONES

El presente documento intenta demostrar la potencia de utilizar grafos para representar y comprender problemas complejos. La relación entre colegios, carreras y la captación de estudiantes, representado a través de grafos, revela patrones, conexiones y estructuras que de otra manera podrían pasarse por alto.

Además, es fundamental el conocimiento de los principios básicos de las Ciencias de Redes. Con este conocimiento, se pueden aplicar métricas adecuadas para descifrar las mejores análisis.

La utilización de herramientas como Gephi e iGraph ha demostrado ser esencial en este proceso. Gephi, con su capacidad para visualizar claramente redes complejas, permite a los investigadores obtener una comprensión visual inmediata de las estructuras de la red.

Por otro lado, iGraph para R, aunque no necesariamente visual en naturaleza debido a la escala, ofrece capacidades analíticas poderosas, permitiendo cálculos y estadísticas rápidas y avanzadas.

Mirando hacia el futuro, este estudio podría beneficiarse de la inclusión de otras variables o características que enriquezcan la representación de la red de captación. Aspectos como la geolocalización de los colegios, las especialidades dentro de las carreras, las calificaciones promedio de los estudiantes o incluso factores socioeconómicos podrían incorporarse.

Estas adiciones no sólo harían la red más informativa, sino que también permitirían análisis más profundos y ofrecerían nuevas perspectivas sobre las dinámicas de captación de estudiantes.

Finalmente, resaltar que estas técnicas y herramientas ofrecen una gran ayuda para descifrar y comprender los sistemas complejos que definen nuestra sociedad.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), 167-256.
- [2] Barabási, A. L. (2016). *Network science*. Cambridge university press.
- [3] Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge university press.
- [4] Bollobás, B. (1998). *Modern graph theory*. Springer.
- [5] Bastian, M., Heymann, S., & Jacomy, M. (2009, March). Gephi: An open source software for exploring and manipulating networks. In *International AAAI Conference on Web and Social Media*.
- [6] Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems, 1695.
- [7] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440-442.
- [8] D.B. West, *Introduction to Graph Theory*, Prentice Hall, 2nd edition, 2001.
- [9] A. Schrijver, "On the history of the shortest path problem", *Documenta Mathematica*, vol. Extra Volume: Optimization Stories, pp. 155-167, 2012.
- [10] J.L. Gross and J. Yellen, *Handbook of Graph Theory*, CRC Press, 2004.
- [11] D.B. West, *Introduction to Graph Theory*, Prentice Hall, 2nd edition, 2001.
- [12] M.E.J. Newman, "The structure and function of complex networks", *SIAM Review*, vol. 45, no. 2, pp. 167-256, 2003.
- [13] Verborgh, R., & De Wilde, M. (2013). *Using OpenRefine*. Packt Publishing Ltd.

- [14] Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). *ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software*. *PloS one*, 9(6), e98679.
- [15] Csardi, G., & Nepusz, T. (2006). *The igraph software package for complex network research*. *InterJournal*, Complex Systems, 1695.
- [16] Kolaczyk, E. D., & Csárdi, G. (2014). *Statistical analysis of network data with R*. Springer.
- [17] Newman, M. E. J. (2018). *Networks*. Oxford university press.
- [18] Barabási, A. L., & Albert, R. (1999). *Emergence of scaling in random networks*. *Science*, 286(5439), 509-512.
- [19] Watts, D. J. (2003). *Six degrees: The science of a connected age*. WW Norton & Company.