



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

GENERACIÓN DE UNA METODOLOGÍA DE DETECCIÓN DE
WEBSITE KEYOBJECTS BASADO EN MÉTRICAS DE
CENTRALIDAD DE TEORÍA DE GRAFOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
INDUSTRIAL

JORGE DANIEL DUPRÉ CASANOVA

PROFESOR GUÍA:
SR. JUAN D. VELÁSQUEZ SILVA

MIEMBROS DE LA COMISIÓN:
SR. PEDRO MALDONADO ARBOGAST
SR. PABLO LOYOLA

SANTIAGO DE CHILE
2014

Financiado por el proyecto FONDEF CA12I10061 - AKORI

Resumen Ejecutivo

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL
POR : JORGE D. DUPRÉ CASANOVA
FECHA: 09/12/2014
PROF. GUIA: SR. JUAN D. VELÁSQUEZ

Este trabajo de título tiene como objetivo fundamental analizar el aporte que se pueda generar a una metodología al introducir un análisis con medidas de centralidad de Teoría de Grafos, para la detección de *Objetos Claves* dentro de un sitio web. Para lograr esto existen una serie de objetivos específicos que permitirán concluir finalmente con respecto a la hipótesis de investigación planteada, destacando que se pretende incorporar herramientas que hasta ahora no se habían utilizado para este campo.

Este trabajo es parte del procedimiento para optar al título de Ingeniero Civil Industrial, siendo parte de un proyecto FONDEF llamado “*Plataforma informática basada en web-intelligence y herramientas de análisis de exploración visual para la mejora de la estructura y contenido de sitios web (AKORI: Advanced Kernel for Ocular Research and Web Intelligence)*”. *AKORI* es un proyecto co-creado entre el departamento de Ingeniería Industrial y la Facultad de Medicina de la Universidad de Chile.

La justificación de este trabajo está relacionado con el crecimiento que ha sucedido en la web durante los últimos años. Este aumento exponencial de la cantidad de usuarios y páginas de Internet, ha llevado a que muchas empresas que basan su modelo de negocios en sus sitios web, tengan interés por destacar en la *nube* y tener los sitios optimizados para sus respectivos usuarios, produciendo una nueva rama de investigación sociológica, para ello se han desarrollado diversas metodologías que apuntan a mejorar el contenido y la estructura de los sitios web. Metodologías que tienen por lo general altos grados de subjetividad, siendo poco eficientes para definir exactamente las zonas u objetos que llaman más la atención para los distintos usuarios.

La hipótesis de investigación propuesta plantea que es posible mejorar la metodología creada por Juan D. Velásquez al evaluar con medidas de centralidad de teoría de grafos, outputs obtenidos por experimentos ya realizados a 23 sujetos que navegaron libremente en la página web del MBA del departamento de Ingeniería Industrial, experimento que utilizó herramientas de *eye-tracking*, *dilatación pupilar* y *EEG*. Al mejorar la detección de *Website Keyobjects* aumentando la eficiencia de la metodología con el modelo propuesto, se estaría creando un pilar fundamental para el proyecto *AKORI*.

Los *WKO* clasificados con esta metodología se compararon con otros trabajos previos y fueron evaluados finalmente por el webmaster determinando así la eficiencia de este nuevo trabajo y estableciendo las futuras líneas de acción a seguir. Si bien se demostró la hipótesis de investigación planteada, se cree que se podría desarrollar un mejor trabajo mejorando ciertas propuestas que se declaran en la parte final de este informe.

A mi pequeña amada, quién algún día dejará de ser pequeña pero nunca jamás dejará de ser Mía ... Te amo

Agradecimientos

Si tuviera que agradecer a todos quienes han hecho esto posible tendría que hacer un segundo tomo de esta memoria, esta es sin duda la página más difícil de escribir para mí.

En primera instancia quiero agradecerle a la persona que decidió estar a mi lado estos años, debe ser una de las personas con más paciencia que he conocido en mi vida. Estoy seguro que sin su compañía estos años de universidad hubiesen sido muy distintos, ha sido un largo y duro camino juntos que ha valido la pena.

Agradecerle de manera infinita a mi padre Jorge Dupré, a mi madre Patricia Casanova y a mis hermanos Nadia e Isaac. Ellos son el pilar de mi ser, me han formado, me han transmitido sus valores, creencias y amor infinito. Claramente sin la compañía y apoyo de ellos yo no estaría aquí.

Gracias a quien me dio la oportunidad de escoger esta memoria, gracias al profesor Juan que me entregó la confianza y la libertad que necesitaba para poder hacer un trabajo de título, esta memoria fue una muy grata experiencia por la forma en que se generó todo. Se me dio la posibilidad de trabajar con Pablo Loyola al comienzo de mi trabajo y fue gracias a él que esa nube oscura que se llamaba ‘informe de memoria del E’ se transformó en este informe final.

Agradezco también a mis amigos, ponerlos a todos aquí sería imposible. Le doy las gracias a Víctor Verdugo por todo el tiempo que invirtió en mí mientras hacía esta memoria, fácilmente se podría decir que un 50 % de los resultados y de la dirección que tomó este trabajo se debió a su ayuda. Le doy las gracias a cada uno de los principessos: Karl, perrito, Vitoco, Zúñi y Álvi. Les agradezco infinito todo lo que han hecho por mi en estos años, le agradezco toda la compañía, apoyo, cariño, buenos ratos, que me han entregado. Agradezco también la amistad incondicional de mi amiga Patrice, gracias totales por tu cariño.

Obviamente le agradezco a mis primos, gracias les doy por todos los buenos ratos que me dieron, por esa compañía que era eterna cuando niños. Soy una persona bendecida por la cantidad de primos que tengo, creo que he intentado aprovechar esto al máximo. Gracias a mis tíos, a varios de uds los quiero como que fueran mis hermanos. Gracias por que uds han seguido formando una familia gigante e increíble.

Gracias a mis abuelos por todo, espero poder lograr algún día formar una familia como la que uds crearon con tanto sacrificio y amor. Cierro con mis abuelos porque son lo máximo, sé que no he dado ni la centésima parte de amor, tiempo, paciencia, cariño y abrazos de lo que ellos alguna vez dieron por mí. Espero que la vida me permita tener más tiempo con ellos. Te amo mucho abuelo y espero que todo lo que hiciste por mí yo te lo pueda retribuir.

Tabla de Contenido

Resumen Ejecutivo	i
Agradecimientos	iii
1 Introducción	1
1.1 <i>AKORI</i>	2
1.2 Hipótesis de Investigación	3
1.3 Descripción del Proyecto y Justificación	5
1.4 Objetivos	7
1.4.1 Objetivo General	7
1.4.2 Objetivos Específicos	7
1.5 Metodología	8
1.6 Estructura de la Memoria	10
2 Marco Teórico	11
2.1 Atención Visual	11
2.1.1 Estructura del ojo humano	11
2.1.2 Movimientos Oculares	13
2.2 Eye Tracking	14
2.3 Cálculos básicos de Álgebra Lineal	16
2.4 Teoría Espectral de Grafos	17
2.4.1 Representación Matricial	21
2.4.2 Medidas de Centralidad de Teoría de Grafos	24
3 Análisis de los datos	33
3.1 Experimento	33
3.2 Preprocesamiento	34
3.2.1 Matlab	34
3.2.2 Generación y Selección de los Datos	35
3.2.3 Pre-procesamiento	36
3.2.4 Transformación	36
3.3 Cálculo de medidas de Centralidad	40
3.4 Consolidación de los datos	44
4 Resultados	48
4.1 Exploración Visual	48
4.1.1 Filtración de la Data	48
4.1.2 Primer y Segundo Objeto	50
4.2 Resultados Consolidados	56
4.2.1 Parte A: Análisis con todas las aristas observadas	57
4.2.2 Parte B: Análisis con aristas observadas mínimo 6 veces	63
4.2.3 Comparación Final	69
4.2.4 Comparación Resultados Parte A y Parte B	69
4.2.5 Comparación con <i>webmaster</i>	70
5 Conclusiones	73

5.1	Desarrollos Futuros	74
5.1.1	Sitio Web y Experimentos	74
5.1.2	Webmaster	75
5.1.3	Objetos	75
5.1.4	Comportamiento de los Sujetos	76
Bibliografía		77
Apéndices		80
Apéndice		80
A	Tablas	80
A .1	Tiempos Promedio por Objetos	80
A .2	Tiempos Promedio por Paginas	81
A .3	<i>WKO</i> detectados por Martínez	82
B	Figuras	85
C	Glosario Teoría de Grafos	90

Índice de Tablas

1.1	The Internet, A decade later	1
3.1	Estructura datos originales	36
3.2	Resumen Procesamiento Datos	38
3.3	Tabla Resumen con datos de cada sujeto	45
3.4	Extracto objetos visualizados por sujeto	46
3.5	Aristas Consolidadas Página 1	47
4.1	Datos Eliminados en Filtros	49
4.2	Primer y Segundo Objeto	51
4.3	Primer y Segundo Objeto tiempos	52
4.4	Correlación Parámetros no Normalizados	59
4.5	Correlación Parámetros Normalizados	59
4.6	Website KeyObjects - Indicadores A	61
4.7	Website KeyObjects - Indicadores B	62
4.8	Lista de Website Keyobject	63
4.9	Correlación Parámetros no Normalizados	65
4.10	Correlación Parámetros Normalizados	65
4.11	Website KeyObjects - Indicadores A2	66
4.12	Website KeyObjects - Indicadores B	67
4.13	Lista de Website Keyobject	68
4.14	Lista de Website Keyobject	71
4.15	Lista de Website Keyobject	72
5.1	Lista de Website Keyobject	82
5.2	Promedio Objetos Vistos por Página	83
5.3	Primer y Segundo Objeto por Página - sujeto 2	84

Índice de Figuras

1.1	Principales Áreas del Proyecto AKORI	3
1.2	Blogosfera – Gráfica Blogs Partidos Políticos EE.UU	6
2.1	Estructura del Ojo Humano	13
2.2	Herramientas de Eye Tracking	14
2.3	Ejemplo de Scanpath	15
2.4	Ejemplo de Mapa de Calor	16
2.5	Grafo de estrella $n = 5$	18
2.6	Grafo con 8 vértices	19
2.7	Grafos completos K_3 y K_4	20
2.8	Grafos isomorfos, A y B	21
2.9	Grafo Estrella con $n = 7$	29
3.1	Objetos observados por sujeto en página 1	42
3.2	Medidas de Centralidad, tiempo y Frecuencia - Páginas 1,2,3 y 4	44
4.1	Objetos 5 y 6	51
4.2	Tiempo Promedio por Objeto	53
4.3	Tiempo Promedio por Página	55
4.4	Scatterplots - Todas las aristas	58
4.5	Scatterplots - Más de 6 aristas	64
5.1	Tiempo Promedio por Objeto	80
5.2	Tiempo Promedio por Página	81
5.3	Top Five Key Objects	85
5.4	Top Five Key Objects	86
5.5	Top Five Key Objects	87
5.6	Top Five Key Objects	88
5.7	Top Five Key Objects	89

1 Introducción

En los últimos años el crecimiento que se ha vivido en la Web [1] ha desarrollado un especial interés en estudiar el comportamiento del ser humano al navegar en sitios. Las empresas por su parte, han invertido cada vez más dinero para mejorarlos en función de sus propios usuarios. En la Tabla 1.1 se muestran algunas cifras que cuantifican este crecimiento.

Información	Año 2002	Año 2012
Número Usuarios de Internet (millones)	569	2270
Cantidad de Páginas web (millones)	3.2	555.7
Tiempo promedio de navegación (min/día)	46	240

Tabla 1.1: The Internet, A decade later
Fuente: [2]

Las organizaciones que basan su modelo de negocio en sitios web han volcado su interés en los usuarios, para ello necesitan destacar en relación a su competencia, ya que ser parte de la nube es cada vez más necesario y rentable. Para aumentar la eficiencia de sitios web, se han desarrollado herramientas y metodologías que intentan recopilar información acerca de como navegan los usuarios en un sitio web o cuales son las zonas más llamativas al momento de visualizar un sitio, hasta ahora los procedimientos que se utilizan para conseguir esta información no son objetivos del todo. Algunas técnicas utilizadas actualmente son: entrevistas post navegación, encuestas de percepciones, navegación hablada, estudio de registro de uso, entre otras [2].

En el año 2004 Juan D. Velásquez en conjunto con profesores de la Universidad de Chile y de la Universidad de Tokio, generaron una metodología [3] que permite identificar los *Website Keywords*¹ de una página web. Esta metodología intenta detectar las palabras más importantes de un sitio web para los clientes o usuarios enfocándose en el contenido mismo y en cómo ellos navegan.

Luego Dujovne y Velásquez en [4], hacen referencia a los *Website Objects*² (*WO* de ahora en adelante). Esta metodología se basa en un concepto fundamental que define las páginas

¹Se define un *Website Keyword* como “una palabra o un conjunto de palabras que utilizan los usuarios en sus procesos de búsqueda y que finalmente caracterizan el contenido de una página web” [3].

²Se define un *Website Object* como “cualquier grupo estructurado de palabras o un recurso visual que pertenece a una página web que tiene meta-data y que describe el contenido de ésta.”

web y que permite estudiar el comportamiento de los usuarios en función de esto, los *Website KeyObjects* (*WKO* de ahora en adelante) [4]. Se detectan ahora los *WKO* de un sitio web en función de la cantidad de tiempo que un usuario dedica a observar un *objeto*, utilizando un algoritmo creado por los autores y en conjunto con una serie de resultados obtenidos a través de encuestas.

Hubo una nueva variación a este proyecto cuando González y Velásquez introdujeron la tecnología de eye tracking [5], por lo que ahora la detección de los *WKO* aumenta en eficiencia logrando un gran avance, dejando abierta la posibilidad de que esta metodología vaya mejorando su eficiencia al agregarle distintas técnicas o herramientas que aporten a la detección de los objetos claves. Este trabajo culmina con una publicación [6] en la Revista de Ingeniería de Sistemas en el año 2012 que resumía todo el trabajo realizado durante la memoria de González.

Ya han sido parte de este proyecto más de 7 alumnos de pregrado de la Universidad de Chile, cada uno de ellos ha analizado el impacto que pueda llegar a tener en la metodología señalada sus hipótesis de investigación, como por ejemplo el trabajo realizado por Claudio Aracena en [7] donde se investiga acerca de la correlación que existe entre la dilatación de la pupila y la emocionalidad, mediante análisis que incluyen herramientas de *minería de datos*, *eye tracking* y *dilatación pupilar*.

1.1 *AKORI*

Este trabajo de memoria está inmerso en un proyecto FONDEF llamado “*Plataforma informática basada en web intelligence y herramientas de análisis de exploración visual para la mejora de la estructura y contenido de sitios web*” (De ahora en adelante *AKORI*). Es un proyecto desarrollado conjuntamente entre la Facultad de Medicina y el departamento de Ingeniería Industrial de la Universidad de Chile y es liderado por los profesores Pedro Maldonado (Medicina) y Juan D. Velásquez (DII).

El equipo de trabajo del departamento de Ingeniería Industrial inicialmente esta compuesto por: Gustavo Martinez, Claudio Aracena, Joaquín Jadue, Rocío Ruiz, Gino Slanzi, Luis Salas y Jorge Dupré.

En la Facultad de Medicina el Proyecto es liderado por el Dr. Pedro Maldonado y dentro de su equipo de trabajo se encuentran profesionales del área de la salud y de ingeniería eléctrica. Ellos trabajan en el área de fisiología de la facultad que es donde están actualmente los equipos necesarios para realizar los experimentos de *eye tracking*, *pupil dilation* y *EEG*.

AKORI busca mejorar una metodología para la optimización de estructura y contenido de sitios web, utilizando herramientas de análisis de exploración visual y web-Intelligence. Se pretende que esta memoria incluya la teoría de Grafos al área de Web Intelligence, y que finalmente produzca una mejora en la metodología, dado que su impacto en áreas como la geografía [8] o Social Media [9] ya ha sido demostrado.

En la Figura 1.1 se describen las áreas recién mencionadas y sus características asociadas:



Figura 1.1: Principales Áreas del Proyecto AKORI
Fuente: Elaboración Propia

1.2 Hipótesis de Investigación

Lo que se plantea como hipótesis es la posibilidad de mejorar la metodología creada por Velásquez al incluir el análisis de medidas de centralidad de *Teoría de Grafos* a los datos ya obtenidos en experimentos de *eye-tracking*, para la detección de patrones que permitan modelar o predecir comportamientos de usuarios en un sitio web. Estas medidas permiten analizar el efecto que tienen objetos cercanos a la importancia relativa de un cierto objeto, característica no evaluada hasta el momento en la metodología mencionada.

Esta hipótesis se basa en el desarrollo realizado en [10], se cree factible mejorar el procedimiento utilizado al incluir medidas de centralidad como el *Eigenvector*, medida que ha demostrado en [11] tener mejores resultados que las tres medidas de centralidad principales (*Degree*, *Closeness* y *Betweenness*) que se muestran en [12], en especial cuando se pretende incluir la importancia que tienen los nodos cercanos a uno en particular.

Lo valioso de agregar este análisis con Teoría de Grafos tiene relación con una crítica que se le podría hacer a la metodología usada en [4] [13] [6], dado que los objetos se tratan de manera aislada. La evidencia muestra que el ojo percibe no sólo la existencia de la imagen sino que involucra lo que está en su entorno al momento de visualizar, la vista periférica se ve afectada por los colores, movimientos y otras características que desvían la atención al momento de navegar dentro de una página web por ejemplo [14].

Esta correlación que existe en la atención de eventos muestra que si uno mira un objeto A, el tiempo y la atención que se le presta al objeto se ven afectados por los otros objetos que están alrededor y por ende el movimiento hacia otro objeto B está en función de una serie de factores de redes que deben ser analizados.

Una de las disciplinas que evalúa esta interacción entre elementos es la *Teoría de Grafos*, que utiliza modelos y métricas que intentan explicar cómo funciona esta interacción, por ejemplo existen medidas de centralidad que le otorgan una importancia relativa a cada nodo dentro de un cierto grafo, interpretando de diversas maneras estos valores.

Además se pueden asignar pesos a las aristas y/o a los nodos dentro del grafo, lo que permite calificar diversas características de un cierto grafo, de manera tal que se pueda analizar el efecto de estas, por ejemplo, la distancia entre dos objetos dentro de una página web, puede ser evaluada asignándole un peso a la arista que los une, así desplazarse desde un nodo a otro depende de la distancia que existe entre estos. Esto es sumamente válido dado que si uno se desplaza de un objeto a otro colindante no debiera tener la misma ponderación que si uno recorre toda la diagonal de una página web para moverse desde un objeto a otro, el efecto que produjo el movimiento en estos dos casos debiera ser explicado por algún fenómeno y esto debiera ser respaldado con distintas asignaciones al valor de cada arista.

Este trabajo de memoria está supervisado directamente por el profesor guía Juan Velásquez. Sumado al apoyo de un equipo de trabajo compuesto por más de diez personas, se cuenta con la guía de un Ingeniero Civil Industrial egresado de la Universidad de Chile, PhD (c) de la Universidad de Tokio, Pablo Loyola, cuya experiencia en temas de optimización de páginas web en función de medidas de centralidad se puede apreciar en [15].

1.3 Descripción del Proyecto y Justificación

La importancia del proyecto *AKORI* es fundamental para comprender de una forma objetiva cómo reaccionamos los humanos al momento de navegar en la Web, ya que las herramientas hasta ahora utilizadas para la mejora de sitios web funcionan en base a preguntas

y cuestionarios, entre otros. Todos estos métodos dependen de cómo se exprese la persona, de que tanta facilidad tenga para explicar lo que está pasando por su cerebro, u otras variables. Se postula que al analizar sólo información objetiva se puede generar conocimiento aplicable a áreas impensadas hasta ahora como por ejemplo el diseño de páginas web.

Este trabajo en particular busca mejorar la metodología mencionada en un nivel distinto al de otras memorias que han sido parte de este proyecto, pues considera utilizar datos y resultados de experimentos de *eye-tracking* ya realizados en [10] y calcular con dicha información algunas medidas de centralidad de teoría de grafos para clasificar los *WO* y detectar los *WKO*, comparando finalmente estos resultados con los obtenidos anteriormente para determinar si es que se produce un aporte a la metodología mencionada.

Cabe destacar que dentro de las actividades de este proyecto se incluye el análisis a fondo de la literatura relacionada y el estado del arte en el uso de *eye-tracking*, como por ejemplo el documento publicado por Matsuda y Takeuchi [16] donde se muestra un estudio que sólo considera *Eye-tracking* para la detección de patrones que permitan optimizar sitios web, dejando la opción de realizar un análisis similar pero incluyendo herramientas como dilatación pupilar o teoría de grafos, variando el experimento o el enfoque a desarrollar. Aquí yace la importancia de este proyecto, pues en caso de validar la hipótesis, el proyecto *AKORI* se podría ver favorecido en su eficiencia al momento de detectar los *WKO*.

Un factor relevante que actúa como pilar de este trabajo está relacionado con el efecto que tienen los objetos cercanos en la importancia relativa de un cierto objeto dentro de un grafo. Esto es un factor que hasta ahora no ha sido evaluado, ya que las mediciones utilizadas ignoran este efecto. Para incorporarlo se pretende calcular una medida de centralidad de Teoría de Grafos llamada *Eigenvector* [17], que condiciona la importancia de un nodo dentro de un grafo en función de la relevancia que tienen todos los nodos que están conectados con él.

Estudios de diversas áreas han demostrado el efecto que tiene esta medida de centralidad, [18] es uno de los más conocidos y fue realizado durante las elecciones de EE.UU. del año 2004. En él se muestra una evaluación de los blogs asociados a los dos partidos políticos más grandes del país, se explica el efecto que logra en la importancia relativa dentro de la nube de las páginas web el hecho de estar linkeados a otros blogs cuya importancia relativa es mayor. Se usó el *Eigenvector* como unidad de medida para determinar la importancia de cada blog dentro del Grafo. La gráfica que se muestra a continuación tuvo gran repercusión en los medios dado el impacto visual del análisis de un sector no evaluado con estas medidas.

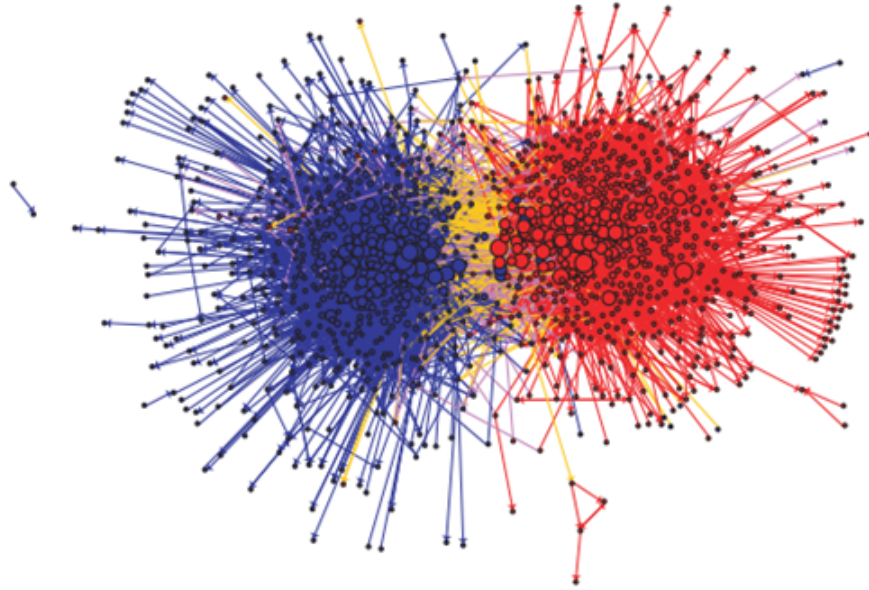


Figura 1.2: Blogosfera – Gráfica Blogs Partidos Políticos EE.UU
Fuente: Imagen obtenida de [18]

En la Figura 1.2 se muestra en color azul los blogs relacionados con el partido conservador, en color rojo se plasman los blogs involucrados con el partido liberal, en medio y de color amarillo se muestran los links que unen los blogs desde liberales a conservadores y en morado las conexiones desde conservadores a liberales. Las aristas que unen los nodos son referencias o links que utilizan los blogs para hacer publicidad unos con otros, así aumentan su relevancia dentro de la nube, y al mismo tiempo crece la posibilidad de ser vistos por más usuarios.

Este grafo muestra que la importancia que tiene un nodo dentro de una red, está en función de la cantidad de conexiones que pueda llegar a tener. No sólo la cantidad es importante, si no que también con quién está linkeado (calidad), dado que la probabilidad de que un usuario ingrese a un cierto blog depende de la cantidad de visitas que tengan los blogs que están linkeados con dicho sitio.

Dentro de las conclusiones se explica el efecto que tiene la conectividad excesiva a priori entre blogs de los mismos partidos, haciendo que todos aumente su importancia relativa ya que si uno aumenta en importancia, hace que todos los que están asociados a el aumenten su relevancia relativa.

Como paso final se concluirá con respecto a la hipótesis de investigación planteada y los resultados obtenidos, siendo un punto importante el destacar los mayores logros del trabajo y posibles propuestas de continuación del trabajo realizado.

1.4 Objetivos

Dado que este trabajo busca validar la hipótesis mencionada, se pretende delinear un objetivo general que permita responder a esta idea, por lo que este se plantea de la siguiente forma:

1.4.1 Objetivo General

Generar un modelo basado en diversas medidas de centralidad de Grafos que utilice data obtenida en experimentos de eye tracking para la detección de WKO para la mejora de una metodología ya existente.

1.4.2 Objetivos Específicos

Para lograr este objetivo propuesto, se plantean una serie de objetivos específicos que apoyarán la realización de este objetivo general. Cada uno de estos objetivos se puede representar como un entregable dentro de este trabajo de memoria.

1. Realizar una revisión de la literatura relacionada con nuevos avances en *eye-tracking*, medidas de centralidad de Teoría de Grafos e Interés Visual.
2. Realizar un análisis exploratorio de los datos
3. Detectar patrones utilizando métricas de grafos en data obtenida en experimentos.
4. Generar una metodología de agrupación de objetos, analizarla y comparar con resultados anteriores.
5. Analizar y concluir en base a los resultados y proponer futuras líneas de acción que continúen la investigación.

1.5 Metodología

El primer paso necesario para cumplir con los objetivos planteados es revisar el estado del arte de los tres tópicos fundamentales de este trabajo de título:

- Comportamiento del usuario web
- *Eye-tracking*
- Medidas de centralidad de Teoría de Grafo y Teoría espectral de grafos

Parte de la investigación del estado del arte se centra en papers que analicen conceptos de Teoría de Grafos, aplicaciones de variadas métricas de centralidad aplicadas en diversos ámbitos, entre otros.

Se necesitará analizar a fondo distintos trabajos efectuados que pudieran ser un aporte a este trabajo de memoria en función de los resultados o alguna novedad presentada.

En [16] se realiza un trabajo similar al presentado en este informe, con varias salvedades, incluyendo la no presencia de dilatación de la pupila y la ausencia de nuevas medidas de centralidad como por ejemplo *Eigenvector*.

Freeman en 1978 [12] fue uno de los primeros en hablar de las medidas de centralidad, es considerado la piedra inicial en muchas investigaciones que incluyen teoría de grafo en sus metodologías, incluye tres de las cuatro medidas de centralidad que se aplicarán en este proyecto. Cabe destacar que este trabajo de investigación ha sido citado en más de 4.000 artículos.

Este trabajo se cimenta en [10] ya que inicialmente se trabaja con la data obtenida en experimentos generados por el autor de este trabajo, por lo que analizar a fondo la forma y los resultados de su experimento son importantes para la calidad de las conclusiones que se puedan generar en el presente proyecto. Es por eso que es necesario detallar de forma resumida el procedimiento utilizado para desarrollar los experimentos, mencionando cantidad de personas que fueron parte de este proceso, tiempos dedicados por cada uno, entre otros.

Diversos protocolos existen para trabajar en Grafos, estos buscan generar diversas metodologías que encuentren focos importantes dentro de una imagen, un cuadro, una página web, entre otros. En [10] se muestra cómo se debe aplicar el procedimiento para la detección de nodos relevantes dentro de un grafo, utilizando la medida de centralidad *Eigenvector*,

explicando de manera algebraica como las matrices ayudan a efectuar un análisis dinámico y eficiente en grafos semejantes a los que se estudiarán en este trabajo.

Se pretende que el siguiente paso sea trabajar con los datos obtenidos en los experimentos realizados en [10] y analizar con diversas métricas de teoría de redes los outputs del *eye-tracking*, orden de vista y mapas de calor de manera de obtener patrones que permitan predecir o explicar el comportamiento de un usuario en un sitio web.

Una vez obtenidos los patrones se pretende generar un modelo basado en una función de probabilidad, que finalmente permita determinar la importancia relativa de un nodo. Para esto se pretende realizar un trabajo con un procedimiento similar a [6] para poder comparar finalmente ambas metodologías.

Finalmente se espera que en base a los patrones detectados con los resultados de los experimentos ya obtenidos y la hipótesis de investigación, se pueda concluir si es que este trabajo de memoria es un aporte a esta metodología y representa una mejora en la búsqueda de *Website Keyobjects* dentro de una página web.

Se agrega finalmente una última sección a este trabajo que consta en plantear posibles continuaciones del tema y mejoras que se pudiesen hacer en futuros trabajos de investigación, agregando además una propuesta de aplicación de los resultados obtenidos de manera de evaluar si es que este aporte logrado puede ser utilizado en algún otro rubro o área, dado la amplitud y lo novedoso del tema a investigar.

1.6 Estructura de la Memoria

Este trabajo de memoria está dividido en 5 Capítulos, estos muestran en cada uno de ellos una arista importante para el desarrollo final de la metodología planteada.

El primer capítulo muestra el trabajo inicial desarrollado durante la primera etapa de este proyecto. Acá se definieron items importantes como los objetivos, alcances, plazos, entregables, entre otros. La mayoría de estos item sufrieron modificaciones durante el desarrollo del proyecto, que se debieron principalmente a pivoteos producidos por nuevas lecturas o recomendaciones realizadas por gente experta en la materia.

En el Capítulo 2 se presenta un trabajo de investigación acerca del estado del arte de los tópicos fundamentales que generan los cimientos de este trabajo. Acá el lector podrá profundizar su conocimiento relacionado al comportamiento del ser humano al momento de navegar, el funcionamiento del ojo durante una observación y un resumen de Teoría de Grafos donde se detallan cálculos y medidas relevantes durante el desarrollo de esta memoria.

En el tercer capítulo se resumen cómo fue el proceso completo del experimento, se detalla además el procedimiento con el que se actuó al momento de manipular los datos, cómo estos fueron obtenidos, filtrados y procesados. Se menciona el cálculo de las medidas de centralidad y la consolidación de la data. Se incluye un análisis visual de los datos, que finalmente servirá de comparación con los resultados finales, para ver si realmente se obtiene un beneficio al considerar medidas de centralidad, o si sólo bastaba con saber básicamente a que objeto se le dedica más tiempo para determinar su importancia relativa dentro de una página web.

En el Capítulo 4 se muestran los resultados obtenidos, dividido en 2 partes principales: *La Exploración Visual* y *Resultados Consolidados*. Esto fundamentalmente con el objetivo de realizar dos análisis que luego de los resultados, puedan ser comparados entre ellos antes de concluir.

Finalmente en el último capítulo, se relatan las conclusiones obtenidas en función de la hipótesis de investigación. Además se incluyen algunas propuestas para obtener mejores resultados y futuras líneas de acción en caso de que algún memorista quisiera continuar con este trabajo de investigación.

2 Marco Teórico

En este capítulo se busca darle al lector una noción básica de los conceptos principales de este trabajo de memoria, de forma tal que pueda interiorizarse de los últimos avances en las distintas áreas y aprender conceptos que se mencionarán en los próximos capítulos.

En especial se profundizará en la atención visual de los seres humanos, las herramientas que se utilizaron durante los experimentos, y la materia base de análisis que se propone para la detección de *Website Keyobjects*.

2.1 Atención Visual

El ojo humano, y su capacidad de interpretar el entorno a través de la captación de estímulos luminosos, son de suma importancia en este trabajo. En este apartado se ahondará en conceptos como *atención visual*, *movimiento ocular* y *eye-tracking*, para relacionarlos con procesos de exploración visual, selección y decisión. Para ello se describen brevemente bases anatomofisiológicas en [19] [20].

2.1.1 Estructura del ojo humano

El ojo de los seres humanos, como en la mayoría de los vertebrados, posee una estructura esférica de 26-28 milímetros de diámetro, que contiene un material líquido denominado humor acuoso y otro de mayor consistencia, el humor vítreo, dentro de los cuales se generan fenómenos de refracción que modelan la transmisión de la luz y la generación de imágenes. La pared que constituye esta “*esfera*” está compuesta por capas, la esclerótica, coroides y retina (de exterior a interior) de distinta composición, de las que esta última es la protagonista.

Además se destacan en el ojo una serie compleja de lentes y mecanismos de regulación de la entrada de la luz, que permiten ver objetos a distancia, con poca luz, cambiar el foco de atención desde objetos lejanos a cercanos con rapidez o percibir el entorno desde lo central a lo periférico en forma constante.

En los próximos párrafos se detallarán algunas de las partes más relevantes del ojo humano, permitiendo así al lector ahondar su conocimiento relacionado con la estructura de este

órgano y el funcionamiento de sus componentes.

- **Esclerótica o esclerocórnea:** Es la capa más externa, altamente resistente y protege al globo ocular de traumatismos. Parte de esta capa es la córnea, la lente más potente del ojo humano. A través de esta capa transparente se puede ver la pupila y el color del iris.
- **Coroides:** Capa intermedia, originalmente descrita en la literatura como parte de la Úvea, junto con el Iris. Esta capa altamente vascularizada, nutre a la retina y está en contacto directo con ella. Se modifica en la parte anterior para formar el diafragma del iris. Este descansa sobre el cristalino y presenta una abertura central para la pupila, cuyo tamaño es variable gracias a las contracciones de los cuerpos ciliares musculares subyacentes.
- **Retina:** Es la capa más interna y delicada. Capa nerviosa adaptada para la recepción de estímulos de luz, pigmentada y que posee el púrpura visual que se blanquea al exponerse a una luz muy fuerte. Está formada por un patrón estratificado de células nerviosas, los foto-receptores, cuyas fibras conforman el nervio óptico.
- **Pupila:** Orificio de tamaño variable, cuyo diámetro normal está descrito entre los 2 y 4 mm, pudiendo alcanzar 8 mm en midriasis plena (dilatación máxima) y 0,5 mm en miosis. Regula la cantidad de luz que entra al ojo.
- **Cristalino:** Lente biconvexa traslúcida, que junto con la córnea se encarga de enfocar las imágenes en la retina. Su curvatura se modifica por los cambios de la tensión de sus ligamentos suspensorios para acomodarse a una visión cercana o a distancia.
- **Humor vítreo:** Porción del contenido interior del ojo de características gelatinosas, traslucido. Es atravesado por un canal central que se dirige desde la entrada del nervio óptico hacia la parte posterior del cristalino.
- **Humor acuoso:** Fluido que llena la cámara anterior, el espacio entre la córnea y el cristalino, y también la pequeña cámara posterior, la estrecha hendidura entre el cristalino y la parte posterior del iris. Igual que el líquido cefalorraquídeo, el humor acuoso se secreta y reabsorbe constantemente.

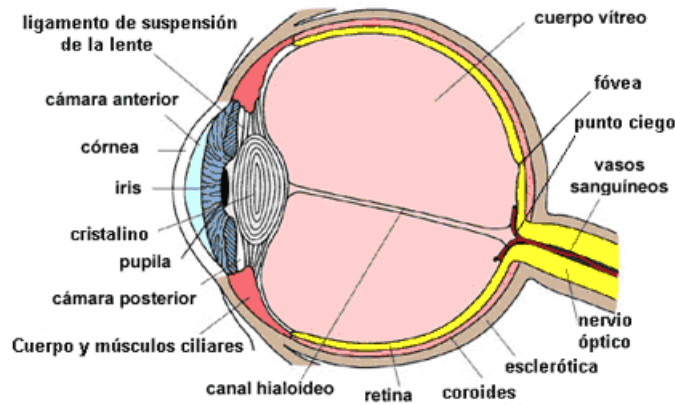


Figura 2.1: Estructura del Ojo Humano
Fuente: [19]

Diversos estudios como el RSVP – *Presentación Visual Serial Rápida* [21], estudian el comportamiento del ojo humano al leer textos o ver imágenes, intentando descifrar tiempos de procesamientos y diversas capacidades que pueda tener el ojo, siendo muchas de ellas indescifrables hasta hoy en día.

Importante es estimar el tiempo mínimo que requiere el cerebro para procesar una imagen que el ojo acaba de ver. Estudios realizados durante este año por científicos del MIT en [22] muestran que el cerebro requiere de 13 milisegundos para entender que es lo que está mirando, lo que contrasta con la conocida ley de los 80 milisegundos, que supone el tiempo mínimo que demora el cerebro para relacionar actividades diarias o escuchar sonidos determinados en su entorno.

2.1.2 Movimientos Oculares

El ojo humano percibe el entorno con permanentes movimiento de fijación (500 milisegundos), sacado (30-40 milisegundos [20]) o barrido de retorno (50 milisegundos) que fijan la atención en uno u otro punto, se detienen y enfocan. El ojo se mueve en sus ejes horizontal y verticalmente, en ángulos hacia afuera y adentro. Otros movimientos complejos se logran gracias a la contracción coordinada de varios músculos, como la convergencia, de aproximación, o divergencia, de alejamiento del objeto de visualización. Todo esto se utiliza por ejemplo en los denominados movimientos sacádicos, que elaboran un mapa mental de lo que nos rodea mediante la observación completa y dinámica de nuestro alrededor.

Como se mencionó anteriormente, la fovea o área central de la retina, altamente poblada por bastones (células fotorreceptoras) es capaz de percibir imágenes con mayor resolución que las áreas periféricas, y es este foco el que se desplaza de un lado a otro, buscando áreas de

interés que requieran, mayor atención y definición. Se han descrito diferencias en la atención del observador [14], dependiendo del origen del estímulo y expectativa en torno a éste. Es mucho más difícil evitar un cambio atencional provocado por una señal periférica que uno debido a una señal central, debido a la naturaleza automática de los cambios de atención con señales periféricas. Estas señales captan la atención del sujeto incluso cuando no son esperadas, no así las centrales.

Lo anterior, como muestra de cómo se percibe y obtiene información del medio, a través de la visión, fundamenta la búsqueda de herramientas que midan tales patrones, y modelen un correlato con el comportamiento humano. El estudio del movimiento ocular dirá qué se mira, por cuánto tiempo y cuándo se deja de mirar

2.2 Eye Tracking

El concepto hace referencia a la monitorización y registro de la forma en que un sujeto mira una escena o imagen. Registra en qué áreas fija su atención, por cuánto tiempo y qué orden sigue en su exploración visual. Esto es utilizado tanto como dispositivo de entrada o interacción o como herramienta para la evaluación objetiva de interfaces.

Como dispositivo de entrada o interacción no tiene tanta precisión como el mouse o el teclado, pero se ha utilizado en entornos de realidad virtual o por usuarios con discapacidad motriz. También existen distintos sistemas, como los eye-trackers que se colocan en la cabeza del sujeto o los remotos, que al estar camuflados en la pantalla o ser menos intrusivos, generan menos “ruido” en los procesos de evaluación, está demostrado que estos afectan de menor forma el comportamiento de los sujetos evaluados.

En la figura 2.2 se muestran dos imágenes relacionadas con dos herramientas de eye tracking, la primera es una menos invasiva, y en la segunda se muestra una que graba el posicionamiento de los ojos a una corta distancia.



Figura 2.2: Herramientas de Eye Tracking
Fuente: [20]

El sujeto al visualizar la escena va construyendo un Scanpath o recorrido (2.3), identificándose en él nodos o puntos en los que existe una fijación. Previamente calibrados, los aparatos van censando la rotación ocular, determinando así hacia donde se dirige la visión central (fóvea) y por ende el punto de visión más nítida y atención más clara, las denominadas fijaciones. Esta es la información más valiosa al recolectar los datos, pero su medición e identificación aún no está estandarizada, por lo que estudios que usasen medidas diferentes podrían tener diferentes resultados.

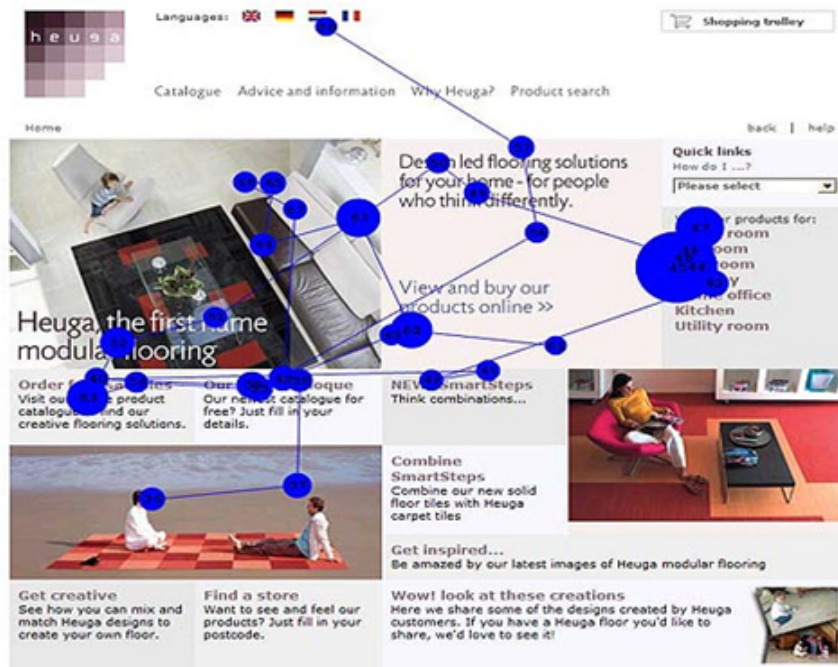


Figura 2.3: Ejemplo de Scanpath

Fuente: [20]

En cuanto a la representación de los datos, el Eye-Tracker graba los scanpath e interpreta por medio de distintos software, que generan representaciones del comportamiento visual, a través de mapas de calor que son representaciones animadas de un “punto de atención” sobre la interfaz de la escena, caminos o rutas sacádicas, mapas de zonas ciegas y mapas de calor, tal como se muestra en las imágenes 4 y 5. Con estos datos es posible definir “áreas de interés” en una interfaz [23].

En la figura 2.4 se puede apreciar un output de un experimento que presenta su resultado con un mapa de calor. En este caso es en función de un diario y en él se muestran las zonas que más llamaron la atención del sujeto, determinado en gran porcentaje por el tiempo de fijación en cada sección de la imagen.

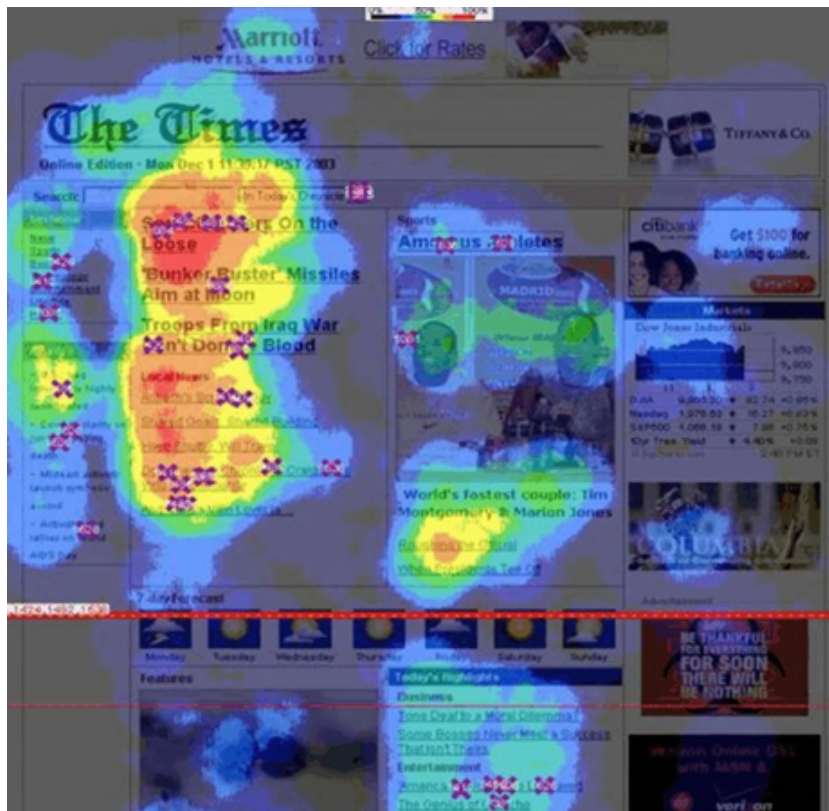


Figura 2.4: Ejemplo de Mapa de Calor
Fuente: [23]

2.3 Cálculos básicos de Álgebra Lineal

Para los diversos cálculos que se mencionarán en los capítulos siguientes, se mencionan a grandes rasgos los conceptos fundamentales que permiten el cálculo de las medidas de centralidad, en particular para calcular el *Eigenvector*, una de las medidas de centralidad, se requiere de pasos básicos de Álgebra Lineal que se explican a continuación.

Lo fundamental es detallar cómo es el cálculo de los valores y vectores propios, agregando condiciones y resultados que son necesarios para el cálculo final de medidas de centralidad que se explicarán pronto. La mayoría de los conceptos mencionados a continuación provienen de [24].

Sea A una matriz, λ_0 es un valor propio de A si existe un vector propio no nulo x tal que:

$$Ax = \lambda_0 x. \quad (2.1)$$

El conjunto de todos los valores propios distintos de A se denota como $\Lambda(A)$ y a este

conjunto se le denomina espectro de A . A x se le denomina vector propio de A asociado al valor propio λ_0 ,

$$S_{\lambda_0} = \{x \in C^m | Ax = \lambda_0 x\}, \quad (2.2)$$

Este subconjunto es un subespacio vectorial de C^m que se llama subespacio propio de A asociado a λ_0 . Como:

$$x \in S_{\lambda_0} \Leftrightarrow (\lambda_0 I_n - A)x = 0, \quad (2.3)$$

Se sabe que si los vectores propios de A están asociados al valor propio λ_i , entonces los vectores x_1, \dots, x_n son linealmente independientes.

El polinomio característico se obtiene al calcular el determinante de $\lambda I_n - A$, este polinomio de grado n se expresa en la ecuación 2.4,

$$p_A(\lambda) = \det(\lambda I_n - A) = \lambda^n + p_1 \lambda^{n-1} + \dots + p_{n-1} \lambda + p_n. \quad (2.4)$$

Al igualar el polinomio a cero, se obtiene la ecuación característica, con la que finalmente se obtienen los valores propios de la matriz, que vienen siendo las raíces de esta ecuación. Dado un valor propio λ_n , el *vector propio* de A es una solución distinta de cero que resuelve finalmente,

$$(\lambda_0 I_n - A)x = 0. \quad (2.5)$$

2.4 Teoría Espectral de Grafos

Si se considera un grupo de amigos y se quiere analizar el nivel de amistad, uno podría generar un análisis en función de la cantidad de llamados que se hacen entre ellos, y concluir por ejemplo que la persona más importante es la que está conectada constantemente con más integrantes del grupo en comparación con el resto. Esto sería una especie de intento de estimar la importancia relativa de alguien o algo dentro de una red, en este caso una red social.

Para complejizar un poco más el ejemplo anterior se podrían incluir más variables al análisis como por ejemplo decir que además de considerar la cantidad de llamados entre amigos es importante considerar con quién se tiene contacto, esto va muy de la mano con el ejemplo de los blogs, dado que no sirve de mucho estar conectado sólo con gente que tiene pocos amigos, esto no te volvería una persona popular en una escuela.

Al igual que el ejemplo anterior, la Teoría de Grafo analiza diversos grafos con distintas características con el objetivo de obtener información a partir de análisis que están sostenidos en variadas teorías. En este trabajo de memoria, el análisis está enfocado particularmente en las medidas de centralidad de teoría de Grafo, en particular se verán 4 medidas distintas: *Degree*, *Closeness*, *Betweenness* y *Eigenvector*.

A continuación se explicarán los conceptos fundamentales que más se mencionan en este trabajo, junto con incluir en Anexos un diccionario con gran cantidad de conceptos claves a nivel general de la Teoría de Grafos.

Un *grafo* consiste en un conjunto de *nodos* y *aristas* que los conectan entre sí, por lo general modelan un conjunto de objetos que tienen un grado de conexión como por ejemplo un conjunto de hombres y mujeres como nodos y la relación de amigos entre ellos reflejada como las aristas que los conectan.

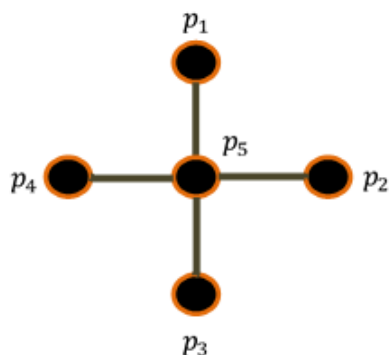


Figura 2.5: Grafo de estrella $n = 5$
Fuente: Elaboración Propia

Pueden ser muy simples como el típico grafo estrella que muestra Freeman en [12] que se muestra en la Figura 2.5, el cual es uno de los grafos básicos para entender la interpretación que le da el autor al concepto de centralidad.

Alguna de las siguientes definiciones fueron obtenidas de [17] y de [25]:

Un *grafo* G está formado por un par $(V(G), E(G))$, siendo $V(G)$ un conjunto finito de

vértices y $E(G)$ un conjunto de aristas que por lo general se denotan i, j o (i, j) .

En particular el número de *vértices* se denota como $n = |V(G)|$ y el número de aristas se escribe $m = |E(G)|$, un grafo debe a lo menos tener un vértice y no necesariamente tener aristas, en este caso se denomina *grafo trivial*. En la Figura 2.5, podemos ver que $n = 5$ y que $m = 4$.

Un grafo puede tener una o más aristas que conecte dos puntos, en caso de que hayan más de una arista entre dos puntos i, j , se dice que el grafo G es un *multigrafo* y en caso de que exista una arista que conecte a un punto i con el mismo (*loop*), el grafo se denomina *pseudografo*.

Dos vértices i, j son *adyacentes* si $(i, j) \in E(G)$, en este caso se dice que i y j son los *extremos* de la arista. En caso de que un vértice i esté conectado con al menos un vértice diferente, existe un conjunto que abarca a todos los puntos que son adyacentes a i , este conjunto se llama vecindario del vértice i y se escribe como $N_G(i)$, el vecindario cerrado de i es $N_G[i] = N_G(i) \cup i$.

En la figura 2.6 se muestra un grafo con dos vecindarios, uno normal y el otro cerrado.

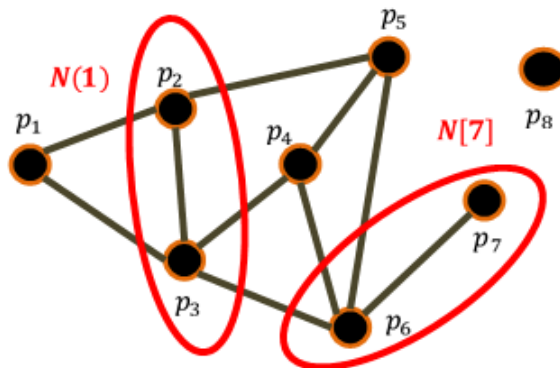


Figura 2.6: Grafo con 8 vértices
Fuente: Elaboración Propia

El *grado* de un vértice i es la cardinalidad del conjunto $N_G(i)$ y se denota por $d_G(i)$, además existe notación para el grado mínimo y máximo de un grafo G , como ejemplo en la Figura 2.6 se observa que el grado del vértice 4 es 3, dado que está conectado con 3 puntos (p_3, p_5, p_6), el grado mínimo del grafo es $\delta(G) = 0$ (grado del vértice 8) y el grado máximo del grafo es $\Delta = 4$ (grado del vértice 6).

En caso de que un vértice no sea adyacente a ningún otro punto del grafo, se dice que ese vértice es un vértice *aislado*, o sea que $d(i) = 0$. Por el contrario, si un vértice está conectado

con todos los otros vértices del grafo se dice que ese vértice es *universal* y que $d(i) = n - 1$.

En caso de que en el grafo todos sus vértices tengan el mismo grado, este grafo se dice *regular*. El *complemento* del grafo G se denota por \bar{G} , este grafo tiene el mismo conjunto de vértices que G y posee el conjunto de aristas que le faltan a G para ser un grafo regular de grado $(n - 1)$. Un grafo con todos sus n vértices adyacentes se conoce como grafo *completo* y se denomina K_n .

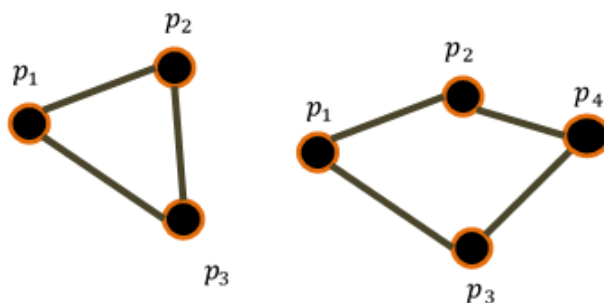


Figura 2.7: Grafos completos K_3 y K_4
Fuente: Elaboración Propia

Un *camino* o *trayectoria* es un conjunto de aristas que unen dos vértices i, j distintos. Por ejemplo en la Figura 2.7, en la imagen de la derecha, se muestra un grafo que muestra dos caminos de largo 2 entre los puntos p_1 y p_4 , uno que pasa por el punto p_2 y otro que pasa por el punto p_3 .

Grafo *conexo* G es tal que cualquier vértice $i \in V(G)$ tiene al menos un camino para alcanzar otro punto del grafo. Esta definición es recomendable recordar para el cálculo de las medidas de centralidad pues es una condición para alguna de ellas.

La *distancia* de un Grafo G es la longitud de una trayectoria entre dos puntos $i, j \in V(G)$ y la más corta entre dos puntos $i, j \in V(G)$ se denomina ruta geodésica o camino geodésico.

La *excentricidad* de un vértice $i \in V(G)$ es la máxima distancia entre dos puntos distintos, sin que se repitan aristas dentro de la trayectoria.

Dos grafos son *isomorfos* si tiene la misma cantidad de objetos, la misma cantidad de aristas, el mismo grado en sus nodos, pero conectados de maneras distintas. En la Figura 2.8 se muestran dos grafos isomórficos.

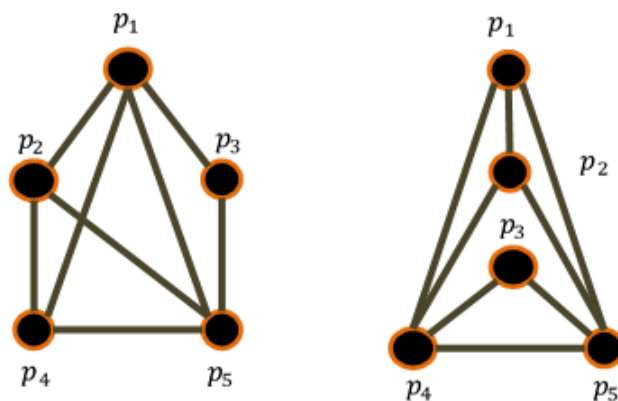


Figura 2.8: Grafos isomorfos, A y B
Fuente: Elaboración Propia

Cabe destacar que en el caso de que no se entiendan ciertos conceptos que no son explicados dentro de este capítulo, en la sección de anexos se encuentra un *Glosario* con los conceptos más relevantes de teoría de Grafos que permitirán una mejor lectura en caso de que el lector no esté familiarizado con esta materia.

2.4.1 Representación Matricial

En teoría de grafo se utilizan diversas matrices que permiten modelar y desarrollar cálculos necesarios para el entendimiento de los conceptos más utilizados en esta área, a continuación se detallan matrices que se utilizarán durante el desarrollo de este trabajo y su posible utilidad en el cálculo de las medidas de centralidad.

Tal como se mencionó, $G = (V, E)$ es un grafo con $|V| = n$ y $|E| = m$, donde n y m son el número de vértices y aristas respectivamente. Asumiremos para las siguientes interpretaciones que G es un grafo sin peso, no dirigido y que no posee ni ciclos (*loops*) ni múltiples aristas.

Matriz de Adyacencia

La matriz de adyacencia permite modelar un grafo, en resumen explica que nodos están conectados entre sí, cuales nodos están aislados, incluso se puede representar la cantidad de veces un nodo puede estar conectado con otro, asignándole un valor mayor a uno en el caso de que hayan múltiples aristas.

Dado un grafo G , la matriz de adyacencia $A = A_G$ es la matriz $n \times n$ siguiente:

$$A_{ij} = \begin{cases} 1 & \text{si } (i, j) \in E \\ 0 & \text{si no} \end{cases} \quad (2.6)$$

Claramente en el caso de grafos no dirigidos, la matriz A_G es una matriz simétrica y binaria.

A continuación se muestra la matriz de adyacencia del grafo A de la Figura 2.8. Recordando que $A = (V, E)$ y:

$$V = \{1, 2, 3, 4, 5\},$$

$$E = \{(1, 2), (1, 3), (1, 4), (1, 5), (2, 4), (2, 5), (3, 5), (4, 5)\}.$$

$$A_A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Matriz de Incidencia

Dado el grafo $G = (V, E)$ se define a la matriz de incidencia del grafo de dimensiones $|V| \times |E|$ como:

$$b_G = \begin{cases} 1 & \text{si } (i, j) \in e_j \\ 0 & \text{si no} \end{cases} \quad (2.7)$$

Con $e_j \in E$. El producto punto entre filas de la matriz B_G indica que los vértices asociados a las filas no son adyacentes y la suma de los valores de la fila muestra el grado del vértice.

En la siguiente matriz se muestra la matriz de incidencia del grafo A de la Figura 2.7. Recordando que el grafo $A = (V, E)$ y:

$$V = \{1, 2, 3, 4, 5\},$$

$$E = \{(1, 2), (1, 3), (1, 4), (1, 5), (2, 4), (2, 5), (3, 5), (4, 5)\}.$$

$$B_A = (b_1, b_2, \dots, b_8) = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Laplaciano

El Laplaciano está relacionado con el cálculo algebraico que permite obtener los valores y vectores propios de una matriz de adyacencia, este tiene una representación similar y se explica a continuación:

Dado un grafo sin peso G , el Laplaciano $L = L_G$ es la matriz $n \times n$ siguiente:

$$L_{ij} = \begin{cases} -1 & \text{si } (i, j) \in E \\ d_i & \text{si } i = j \\ 0 & \text{en caso contrario} \end{cases} \quad (2.8)$$

Con d_i una diagonal donde su valor en la posición i son el grado del vértice i -ésimo.

Para grafos sin peso y no direccionados, el Laplaciano es una matriz simétrica, y una definición equivalente para esta matriz es:

$$L_G = D_G - A_G \quad (2.9)$$

Donde D_G es la matriz diagonal de dimensión $|V| \times |V|$ que tiene en su término i -ésimo el valor del grado del nodo v_i y A_G es la matriz de adyacencia. Se suele decir que el Laplaciano es una de las matrices más importantes, dado la cantidad de información que se puede obtener de él.

El Laplaciano se puede obtener también en función de la matriz de adyacencia A_G y de la matriz de incidencia B_G , según la fórmula:

$$L_G = B_G \times B_G^T - 2A_G \quad (2.10)$$

Nuevamente se muestra una matriz en función del grafo A de la Figura 2.7, esta vez muestra el Laplaciano, el cual se puede calcular con las fórmulas 2.9 y 2.10.

$$L_G = D_G - A_G = B_G \times B_G^T - 2A_G = \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 3 & 0 & -1 & -1 \\ -1 & 0 & 2 & 0 & -1 \\ -1 & -1 & 0 & 3 & -1 \\ -1 & -1 & -1 & -1 & 4 \end{pmatrix}$$

El espectro del Laplaciano es el conjunto de los valores propios: $\lambda_0 = 0$, $\lambda_1 = 2$, $\lambda_2 = 4$, $\lambda_3 = 5$ y $\lambda_4 = 5$. En este caso tenemos dos valores propios mayores por lo que tenemos dos vectores propios asociados a los valores propios 3 y 4 respectivamente:

$$X_3 = \begin{pmatrix} -0.2236 \\ -0.2236 \\ -0.2236 \\ -0.2236 \\ 0.8944 \end{pmatrix}$$

$$X_4 = \begin{pmatrix} -0.8660 \\ 0.2887 \\ 0.2887 \\ 0.2887 \\ 0 \end{pmatrix}$$

En teoría de grafos, se busca obtener información de distintas redes de objetos conectados de manera de poder optimizar ciertos procesos o prevenir ciertas transmisiones, esto dependiendo del campo del que se está hablando. El mapeo sobre ciertos conjuntos de vértices consiste en definir capacidades de los nodos como poder, importancia, confiabilidad, costo, etc. Para esto existen diversas herramientas como Medidas de Centralidad, ponderación o etiquetado de grafos que buscan explicar estas capacidades.

2.4.2 Medidas de Centralidad de Teoría de Grafos

En teoría de Grafos existen diversas medidas que buscan explicar o representar la importancia relativa que un nodo tiene dentro del grafo, esta importancia está en función de

diversos factores como: la posición dentro del grafo, los nodos que tiene a su alrededor, el número de veces que se ubica entre caminos geodésicos, entre otros.

Dependiendo del autor, las explicaciones del efecto que tiene cada medida difieren, incluso hay distintas variaciones de las medidas de centralidad más utilizadas, que en función del campo en el que se está trabajando, tienen un mayor poder explicativo o un mejor modelamiento del fenómeno estudiado.

Las primeras medidas de centralidad que se mencionaron fueron realizadas por Freeman en el año 1978 [12] y son tres medidas fundamentales que han derivado en más de 30 medidas hasta hoy en día. A continuación se detallan sus cálculos, lo que explican y algunos pros y contras que tiene cada una de ellas.

Degree

Es una medida que interpreta la importancia relativa que tiene un nodo dentro de una red, esta métrica se basa en la cantidad de conexiones que tiene dicho nodo en el grafo. Su origen es ambiguo dado que muchos autores propusieron diversas medidas como Rogers en 1974, que propuso una medida simple en comparación a otras pero que sólo aplicaba a aplicaciones particulares o la que propusieron Kajitani y Maruyama en 1976, medida que a pesar de ser matemáticamente adecuada, tenía un nivel computacional de desarrollo que no se adecuaba a la época. Por lo que la medida de Nieminen's [26] fue la que tomó más peso y es hasta hoy en día la que se considera como medida base.

El grado según Nieminen es la suma de todas las adyacencias que tiene un punto p_k , tal como se expresa a continuación:

$$C_D(p_k) = \sum_{i=1}^n a(p_i, p_k) \quad (2.11)$$

Donde:

$$a(p_i, p_k) = \begin{cases} 1 & \text{si el punto } p_i \text{ está conectado con el punto } p_k \\ 0 & \text{en caso contrario} \end{cases} \quad (2.12)$$

Cabe mencionar que $C_D(p_k) = 0$ en caso de que el punto p_k sea un punto aislado.

Claramente esta medida de centralidad está en función del grafo, en particular de la cantidad de nodos que contenga este, dado que pueden haber dos puntos p_A y p_B , pertenecientes a dos grafos A y B respectivamente, ambos con *Degree* igual a 5, pero resulta que el grafo

A tiene 6 nodos, y por ende la máxima cantidad de nodos adyacentes que puede tener un punto p_k cualquiera es $n - 1 = 5$, en cambio el grafo B tiene 100 nodos, por lo que puede tener nodos con grado igual a 10, representando este un bajo porcentaje de conectividad.

Para ajustar esta medida de centralidad y hacerla comparable entre nodos de distintos grafos, se propuso normalizar por el número máximo de adyacencia que puede tener un grafo: $(n - 1)$. Por lo que el grado normalizado sería:

$$C_D(p_k) = \frac{\sum_{i=1}^n a(p_i, p_k)}{n - 1} \quad (2.13)$$

Lo que intenta explicar esta medida es el potencial de comunicación que tiene un punto dentro de un grafo, por ejemplo que tan importante es una persona dentro de su grupo de amistad al momento de hacer circular una información.

Betweenness

Para el cálculo de esta segunda medida de centralidad, el camino geodésico es fundamental. Tal como fue descrito, es el camino óptimo o uno de los caminos óptimos entre dos puntos de un grafo. Ahora bien, si uno hace la analogía de que cada punto es una persona contagiada o un centro de transmisión de energía, mientras más veces un nodo esté ubicado en un camino geodésico, más influyente se transforma dado que de él depende la velocidad con la que se propaga un virus o la eficiencia con la que se transporta la energía. En el fondo, se puede asumir que cada vez que un punto está situado en un geodésico, este ve aumentado su importancia relativa en el grafo, a esto se le llama Betweenness, y se detallada a continuación.

En [27] y [28] se dice que: *“Una persona estratégicamente localizada en un camino de comunicación entre dos puntos, es una persona central, una persona en tal posición puede influenciar el grupo manteniendo o distorsionando la transmisión de la comunicación”*.

Este potencial de control de información que posee un punto que está dentro del geodésico se intenta modelar con esta medida de centralidad que se calcula de diversas formas que fueron desarrolladas por autores como Anthonisse en 1971 y Freeman en 1977.

El cálculo de esta medida es simple siempre y cuando un punto A está en el único camino geodésico, en el caso de que A esté en uno de varios caminos geodésicos o que al menos no esté en el 100% de los caminos geodésicos, el problema se torna un poco más complicado, esto es lógico dado que si volvemos a la analogía planteada al comienzo de esta página, si una

persona se ubica en el camino óptimo entre dos personas, una contagiada de una enfermedad con una alta tasa de transmisión y una sana, la probabilidad de que la persona sana se enferme o la velocidad con la que esta persona se enferme, está muy condicionada al punto A . En cambio si esa persona está en uno de los caminos óptimos entre una persona sana y una contagiada, el potencial de propagación se ve disminuido dado que existe otro canal con la misma probabilidad de transmisión que él.

En caso de que existan dos o más caminos óptimos, A pierde importancia ya que la transmisión no depende tanto de él, en el caso de que el fuera inmune o que tenga una fuerte resistencia a la enfermedad.

Para solucionar el problema de varios caminos geodésicos, se puede abordar el problema trabajando con probabilidades. Claramente existe un supuesto fuerte y es que se asumen que en el caso de haber varios caminos geodésicos, estos tienen probabilidad igual a:

$$\frac{1}{g_{ij}} \quad (2.14)$$

Donde g_{ij} es el número de caminos geodésicos que unen ambos puntos i y j .

Sea $g_{ij}(p_k)$ el número de geodésicos que unen los puntos i y j , que contienen a p_k . Entonces calcular la probabilidad de que el punto p_k esté aleatoriamente en el geodésico entre p_i y p_j sería igual a:

$$b_{ij}(p_k) = \frac{g_{ij}(p_k)}{g_{ij}} \quad (2.15)$$

Por lo que para calcular la probabilidad de k para todos los puntos i y j pertenecientes al grafo hay que calcular la sumatoria que considere todos los puntos, así el cálculo de esta medida de centralidad sería el siguiente:

$$C_B(p_k) = \sum_{i \leq j} \frac{g_{ij}(p_k)}{g_{ij}} \quad (2.16)$$

Considerando la restricción de que i debe ser menor estricto que j , para así no calcular probabilidades más de una vez, y siendo n la cantidad total de puntos en el grafo.

Cada vez que un punto p_k cae en el único geodésico entre un par de puntos, el valor de

esta medida aumenta en 1, si p_k está en uno de varios geodésicos entre varios puntos, entonces el valor de $C_B(p_k)$ aumenta inversamente proporcional a la cantidad de caminos geodésicos que hay en el grafo.

Al igual que $C_D(p_k)$, $C_B(p_k)$ también depende del tamaño de la red y debe ser normalizado de manera tal que pueda ser comparado con nodos que estén en otros grafos. Esta normalización es relativa al máximo valor en términos de números de puntos en el grafo.

Freeman estableció que el máximo valor tomado por $C_B(p_k)$ es alcanzado sólo por la posición central en una estrella (recordar Figura 2.5), esto es:

$$\frac{n^2 - 3n + 2}{2} \quad (2.17)$$

Esto es simple de demostrar dado que si suponemos cualquier grafo estrella con n nodos, en la figura 2.9 se muestra un grafo estrella con $n = 7$, tenemos que $C_B(p_k)$ del nodo central p_k que está en los $n - 1$ geodésicos es uno. Ahora extendemos este cálculo a los nodos restantes para obtener cuál sería el mayor valor posible de un nodo cualquiera dentro de un grafo de n nodos:

$$\begin{aligned} C_B(p_k) &= \sum_{i \leq j} 1 & (2.18) \\ &= \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} 1 \\ &= \sum_{i=1}^{n-2} (n - 1 - (i + 1 + 1)) \\ &= \sum_{i=1}^{n-2} (n - i - 1) \\ &= \sum_{i=1}^{n-2} (n) - \sum_{i=1}^{n-2} (i) - \sum_{i=1}^{n-2} (n) \\ &= n(n - 2) - (n - 2)(n - 1) - (n - 2) \\ &= (n - 2) \left[n - \frac{(n - 1)}{2} - 1 \right] \\ &= n(n - 2) \frac{(2n - n + 1 - 2)}{2} \\ &= \frac{(n - 2)(n - 1)}{2} \end{aligned}$$

$$\therefore C'_B(p_k) = C_B(p_k) \times \frac{n^2 - 3n + 2}{2} \quad (2.19)$$

Cabe destacar que tanto $C_B(p_k)$ como $C'_B(p_k)$ pueden ser determinados en grafos simétricos conectados o no. La medida de Betweenness es útil para tener una medida del control potencial de comunicación que puede tener un punto.

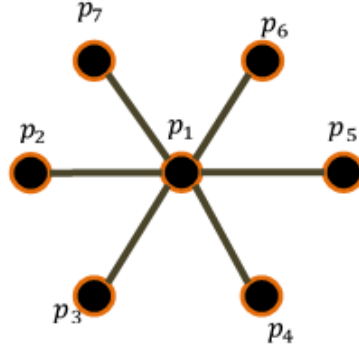


Figura 2.9: Grafo Estrella con $n = 7$
Fuente: Elaboración Propia

Closeness

Esta medida de centralidad tiene un foco distinto a las dos anteriores, puesto que evalúa la importancia relativa de un punto en función de todos los otros puntos del grafo y además se puede considerar, tal como lo mencionó su autor Sabidussi's en 1966 [29], que es una medida de descentralidad o centralidad inversa dado que mientras más apartado del resto está un punto, mayor es el valor de su *Closeness*.

Se considera nuevamente el camino geodésico entre un punto p_k y el resto de los puntos del grafo.

Sea $d(p_i, p_k)$: El número de caminos geodésicos entre p_i y p_k (recordando que puede haber más de uno). Entonces la medida según Sabidussi's estaría expresado en la ecuación 2.20:

$$C_C(p_k)^{-1} = \sum_{i=1}^n d(p_i, p_k) \quad (2.20)$$

El hecho de que esté elevado a -1 se explica porque esta medida aumenta en función de la distancia entre nodos de un grafo, lo importante es su interpretación natural, de forma tal que sea más simple compararla con las dos medidas anteriormente mencionadas.

Claramente esta medida tiene una restricción en relación al grafo, dado que en caso de que en el grafo existan nodos aislados, la distancia entre cualquier nodo p_i y ese nodo aislado es infinita o no está establecida, esto varía dependiendo del autor. Por lo que se restringe a sólo grafos conectados, esto quiere decir que siempre existe un camino desde un punto p_i a un punto p_k .

Al igual que las otras medidas mencionadas, *Closeness* también está en función de la cantidad de nodos del grafo, por lo que es necesario normalizarla de manera tal que sea comparable entre nodos de distintos grafos. Esto fue resuelto por Beauchamp en 1965, y se muestra en la ecuación 2.21 :

$$\begin{aligned} C'_C(p_k)^{-1} &= \left[\frac{\sum_{i=1}^n d(p_i, p_k)}{n-1} \right]^{-1} \\ &= \frac{n-1}{\sum_{i=1}^n d(p_i, p_k)} \end{aligned} \quad (2.21)$$

Eigenvector

Para el cálculo del *Eigenvector* existen dos formas que permiten llegar al resultado. La primera explicada a continuación utiliza *El Laplaciano* de la matriz de adyacencia junto con la diagonal cuyos valores tienen el grado de cada nodo, tal como se muestra a continuación:

Para el cálculo del Laplaciano, basta con restarle a D la matriz de adyacencia, tal como se muestra en 2.22.

$$L = D - A \quad (2.22)$$

Llamaremos a la matriz de adyacencia A , *El Laplaciano* será la matriz L (matriz positiva semidefinida) y la matriz D cuyos elementos son la suma de los pesos. Procederemos a calcular el *Laplaciano normalizado*:

$$L' = D^{-\frac{1}{2}} \times L \times D^{-\frac{1}{2}} \quad (2.23)$$

Luego se calculan los valores propios del *Laplaciano normalizado* y se determina el vector

propio asociado al mayor valor propio, en este vector se guardan los valores de las medidas de centralidad de los nodos del grafo, en particular en este trabajo se determinarán las medidas de importancia relativa de los objetos que el usuario registró en cada página dentro del experimento.

Supongamos que M es una matriz de $n \times n$ y que $Mx = \lambda x$. Esto para todo $x \in R^n$ con $x \neq 0$ y $\lambda \in R$. Entonces se llamará a x el vector propio (*eigenvector*) y a λ el valor propio (*eigenvalue*) de M .

Para proceder con el cálculo de esta medida de centralidad, es necesario comprobar que la matriz M cumple ciertas condiciones que se detallan en la siguiente proposición:

Proposición [30]: Sea M una matriz de $n \times n$, entonces si v y w son vectores propios de M con distintos valores propios, entonces v y w son ortogonales, o sea el producto punto entre estos dos vectores es cero.

Si v y w son vectores propios de M con el mismo valor propio, entonces si $q = av + bw$, entonces estos vectores propios no son ortogonales si tienen el mismo valor propio. Si M es una matriz de vectores ortogonales v_1, \dots, v_n , todos los valores propios y vectores propios son reales.

Luego, M es diagonalizable si

$$M = V\Lambda V^T, \tag{2.24}$$

donde Λ es una diagonal con los valores propios correspondientes a M , por lo que

$$M = \sum_{i=1}^n \lambda_i v_i v_i^T. \tag{2.25}$$

Cabe destacar que en esta proposición, M es simétrica.

La matriz de adyacencia de un grafo no dirigido es simétrica, y tiene un completo conjunto de valores propios reales y un vector propio ortogonal. El conjunto de valores propios de un grafo es el *espectro* del grafo.

Si A es la matriz de adyacencia de un grafo dirigido o no dirigido G , entonces la matriz

A^n tiene la siguiente interpretación: El elemento en la fila i y la columna j entrega el número de caminos de largo n desde el vértice i hasta el vértice j .

La diagonal de una matriz de adyacencia tiene sólo valores cero, indicando que no existen *loops* en los vértices.

3 Análisis de los datos

En el presente capítulo describiremos el proceso completo al que fueron sometidos los datos para finalmente obtener los resultados que se explicarán en el siguiente capítulo. Se describe el proceso previo a esta memoria, tanto con el experimento como también con un pre-procesamiento de la data, hasta un análisis visual paralelo al cálculo de las medidas de centralidad.

3.1 Experimento

Antes de explicar el análisis realizado durante este trabajo, se contextualizará al lector acerca del experimento, para así tener una mejor comprensión en cuanto al análisis y la interpretación de los resultados.

Los datos a utilizar durante este trabajo de memoria son producto de un experimento realizado durante el año 2013 en el laboratorio de neurología de la Facultad de Medicina de la Universidad de Chile, el encargado fue Gustavo Martínez, siendo el experimento parte del trabajo de su memoria [10], el que tuvo una constante supervisión del Doctor Pedro Maldonado.

Este experimento fue realizado a 23 voluntarios que debían cumplir los siguientes requisitos, de manera tal que los resultados pudieran ser extrapolados al resto de la población o grupo objetivo del proyecto:

1. Rango de edad entre 16 y 50 años.
2. Haber usado y conocer cómo usar un navegador web.
3. Haber respondido una pequeña encuesta que se puede ver en la sección de Anexos.
4. Información acerca de la expertiz del usuario.

En este experimento se les solicita a los usuarios que naveguen libremente una página web, específicamente la del MBA del Departamento de Ingeniería Industrial de la Universidad de Chile. Básicamente se escogió este sitio por dos motivos: el primero es por el libre acceso a los datos de la página y segundo es que este sitio tiene parámetros que se ajustan a lo que

requiere para un experimento de este tipo, como:

1. El número de páginas del sitio.
2. El número de objetos en el sitio.
3. La cantidad de visitas.

Cabe destacar que dentro del alcance de este experimento, se solicita a los sujetos que simulen navegar dentro de la página web, ya que realmente sólo están observando pantallazos de cada una de las 32 páginas que en ese momento la plataforma del MBA tenía. Este es un supuesto fuerte y es posible que las conclusiones que se puedan realizar se vean afectadas por este detalle, ya que puede ser que el comportamiento de los usuarios no sea el real al no pueden hacer scroll ni click mientras navegan durante el experimento.

Para más detalle del experimento y sus etapas se puede revisar [10], acá se especifica acerca de la segmentación de los objetos, que fue realizada en conjunto con el webmaster y las herramientas utilizadas durante el proceso, tanto para la grabación de los experimentos como los software utilizados para la extracción de la data.

3.2 Preprocesamiento

Antes de explicar cómo fue el proceso completo de procesamiento de los datos obtenidos, se hará una mención al software utilizado, pieza fundamental en el desarrollo de este trabajo de memoria, donde se realizó cerca del 90% del procesamiento de los datos.

3.2.1 Matlab

La herramienta utilizada para la limpieza y procesamiento de los datos fue *Matlab*, versión R2012a, software optimizado para el cálculo avanzado de vectores y matrices. El motivo para escoger este programa es por la facilidad que implica en el manejo de los datos, que en este trabajo en particular son matrices con cientos de miles de filas que necesitan un software adecuado que procese los datos un lapso de tiempo prudente.

Es importante recalcar que este programa tiene una diversidad de fuentes en la web, que permiten investigar de manera sencilla y utilizar scripts que son de libre acceso, dado que los mismos usuarios van liberando sus códigos en diversas plataformas, es en foros donde se genera el mayor canal de traspaso de conocimiento entre novatos y usuarios avanzados. Con estos scripts, el manejo de los datos es automatizable y no es necesario realizar los comandos

y/o generar códigos cada vez que se procesa un archivo, esto puede ser considerado un aporte para el manejo de futuros datos que tengan una estructura similar.

Otro factor importante es la facilidad para generar gráficos a partir de la data obtenida en los algoritmos, es por eso que el uso de *Matlab* se prefirió por sobre otras herramientas que requerían un lenguaje de programación más avanzado.

3.2.2 Generación y Selección de los Datos

El proceso de captura de la data se hizo mediante técnicas de *eye tracking* de manera de generar información cuantitativa acerca del tiempo que se pasa en cada objeto.

El sitio web analizado contiene 32 páginas y cada una de ellas fue dividida en objetos, siguiendo el mismo patrón que se venía gestando por autores de memorias y trabajos pasados que apuntan a la misma metodología, así al momento de comparar es más simple concluir en base a los resultados obtenidos. La cantidad de objetos por página varía, siendo una página descartada luego por la poca cantidad de objetos presentes. En la sección B de Anexos se muestran pantallazos de algunas páginas y la cantidad de objetos en cada una.

La primera transformación de los datos fue realizada en Medicina por un alumno de ingeniería eléctrica que trabaja en el laboratorio. Por lo que este trabajo comienza con el manejo de estos datos, que básicamente son 23 documentos formato *.txt* cuyos pesos hacen que no sea sencillo trabajarlos en *Excel*, es por esto que se usa el ya mencionado software *Matlab*. El script desarrollado está compartido en una carpeta en *Dropbox* que todo el equipo de *AKORI* dispone.

La estructura de los archivos *.txt* resume la información de cada experimento en ocho columnas, que muestran la información recabada cada 2 milisegundos. A continuación se muestra en la Tabla 3.1 un extracto de la información obtenida en cada uno de estos archivos:

En la primera columna se muestra cada cuanto tiempo el experimento guardaba datos del usuario mientras navegaba (2 milisegundos), en la segunda se muestra qué objeto está mirando el ojo derecho, en la tercera columna se muestra qué objeto está mirando el ojo izquierdo, cabe destacar que no siempre estos dos objetos son el mismo (se explica más adelante porque puede llegar a suceder), luego en la columna cuatro se muestra en qué página está la persona en ese determinado momento.

Las siguientes cuatro columnas muestran las velocidades de los ojos en los ejes *X* e *Y* respectivamente, están estructurados con el protocolo que entrega el mismo software y en este

Tiempo	Objeto d	Objeto i	Id Página	Vel d x	Vel d y	Vel i x	Vel i y
1856	5	5	1	-27.1	120.5	-36.4	137.6
1858	5	5	1	-23.4	148.2	-35.5	162.9
1860	5	5	1	-15.2	167.8	-31.2	180.6
1862	0	5	1	-5.3	174.9	-24.4	186.2
1864	6	6	1	3.4	167.7	-15.9	177.5
1866	6	6	1	8.4	145.5	-6.7	153.8
1868	0	6	1	9.2	111.1	3	118.3
1870	0	0	1	6.8	68.9	12.3	74.9
1872	0	0	1	2.9	24.1	20.3	29.4
1874	0	6	1	-3.6	-50	28.7	-43.9

Tabla 3.1: Estructura datos originales
Fuente: Elaboración Propia

proyecto no son utilizados en absoluto, por lo que en el proceso de filtrado son los primeros datos en ser eliminados.

3.2.3 Pre-procesamiento

Los outputs que se obtienen de los experimentos, proviene del software que graba el *Eye-Tracking*, éste entrega los resultados en formato *.asc*, dichos formatos no son sencillos de trabajar y para su manipulación se requiere del software con licencia que sólo se encuentra en la Facultad de Medicina. Se requiere por ende que estos datos sean procesados de forma tal que se puedan trabajar en programas de acceso más simple como *Matlab*, *Excel* o *Python*.

Los resultados están en función de las coordenadas de los objetos dentro de cada página web, es necesario entonces que se haga una convención de coordenadas a objetos. Esto quiere decir que el paso de decir en cada fila de muestra, se muestren los objetos que el sujeto está mirando en vez de la posición dentro de la página.

Este trabajo de preprocesamiento se realiza en el área de neurociencia y el output de este proceso son archivos *.txt* que tienen 8 columnas, tal como se mostró en la tabla 3. El script asociado a este procesamiento también se encuentra libre para todo el equipo.

3.2.4 Transformación

Los datos obtenidos fueron almacenados y su manipulación fue directamente desde *Matlab*, lo que quiere decir que los datos originales no han sido modificados en lo más mínimo y se han utilizado constantemente para verificar el funcionamiento de distintas secciones del algoritmo.

A continuación se detallan los pasos realizados durante la transformación y limpieza de los datos, se argumentan alguno de los pasos en relación a lo descrito en el estado del arte y se analizan ciertos datos obtenidos durante el desarrollo de esta etapa:

Lo primero que se hizo, fue eliminar las columnas 5, 6, 7 y 8 de los archivos dado que no aportaban información útil para este análisis (las velocidades de cada ojo en el eje x e y), estas columnas son las últimas 4 de cada archivo. Este algoritmo inicial del código tiene la facilidad que permite escoger dentro de las carpetas el archivo *.txt* a procesar, facilitando el proceso dado que no es necesario modificar el código cada vez que se itera con un archivo.

Luego se generó una columna nueva que tenía el delta tiempo de cada fila, información necesaria que luego serviría para calcular los tiempos totales entre objetos una vez terminada la limpieza. Ahora el cálculo de los subtotales para cada objeto es más simple de realizar dado que luego de los filtros los datos quedan con un delta t asociado que se suma por cada objeto, indicando finalmente el tiempo que el sujeto estuvo en cada objeto.

El tiempo exacto en que cada usuario está observando cada objeto es obviado y sólo se considera de ahora en adelante el delta tiempo en cada objeto, relacionado con el tiempo que se realiza cada toma de datos. Esto puede jugar en contra en caso de que se quiera comparar estos resultados con alguna otra herramienta y analizar patrones en mismos momentos. Afortunadamente en este trabajo de memoria sólo se trabajará con el *Eye-Tracking*, así que no es un inconveniente.

El primer filtro realizado fue eliminar las filas que tuvieran en la segunda o tercera columna el valor cero (el valor cero significa que el usuario no estaba viendo un objeto en ese preciso momento). El sujeto puede haber estado observando áreas entre objetos o su vista estaba fuera de la pantalla.

Cada vez que se aplica un filtro en los datos, se guarda en un *Excel* la cantidad de filas de la matriz resultante, esto con el objetivo de capturar el efecto del filtro realizado, para finalmente realizar conclusiones a partir de estos procesos. En particular en este primer paso el porcentaje de datos que se filtran en promedio es de: 73%, siendo el proceso que elimina más filas dentro de todo el preprocesamiento. Siendo un resultado no menor, se analizará el impacto de este filtro en las conclusiones.

Luego se eliminan de esta matriz resultante, las filas cuyos elementos en la columna dos y tres son distintos, o sea que los ojos están mirando distintos objetos en el mismo momento. Se puede deber a que existen fallas en la toma de muestra o se asocia por lo general a pequeñas vibraciones oculares que en los límites de los objetos colindantes distorsiona los resultados.

En esta etapa se filtran en promedio el 7,2% de los datos.

En el proceso siguiente existe un supuesto fuerte: Si un usuario está viendo un objeto, luego observa un objeto vacío o simplemente fuera de la pantalla, y vuelve a posicionarse en el mismo objeto, se asume que el usuario estuvo viendo de manera continua, sin considerar el tiempo que estuvo fuera de este objeto. No se considera el tiempo perdido, pero si se asume que el usuario estuvo viendo todo el tiempo dicho objeto.

Luego se eliminan todas las filas cuyos tiempos de observación son menores a 14 milisegundos, esto se basa en [31], estudio realizado por investigadores del MIT que habla acerca de que una visualización tarda a lo menos 14 milisegundos para que el cerebro logre procesar la información entregada por el ojo. Esto no quiere decir que el observador pueda entender de qué trata o pueda leer un texto, sólo se puede interpretar que el cerebro es capaz de procesar un conjunto de imágenes para finalmente escoger observar por más tiempo la que le llamó la atención en ese pequeño lapso de tiempo.

Finalmente se calculan los subtotales de cada objeto observado, obteniendo finalmente en esta etapa una matriz que muestra la cantidad de objetos que cada sujeto ver por cada página con sus tiempos asociados. Por ende, si se considera toda la matriz se puede ver cuántos objetos observó durante todo el experimento. En la sección B de Anexos se muestra una tabla de uno de los sujetos que muestra lo recién descrito.

En la Tabla 3.2 se muestra un resumen del proceso explicado, donde se detalla la cantidad de filas promedio en cada una de las etapas de procesamiento, junto con el porcentaje de datos que se eliminan en cada una de los filtros mencionados.

	Matriz Original	Objetos = 0 Eliminados	%	Objetos Distintos Eliminados	%
Promedio	520304.8	140242.8	27.0%	130716.7	92.8%
Desv. Est.	176142.5	74849.3	9.7%	71864.1	4.0%

Tabla 3.2: Resumen Procesamiento Datos
Fuente: Elaboración Propia

En promedio los experimentos duraban un poco más de 17 minutos, se puede apreciar por la desviación estándar (5,3 minutos) que estos tiempos dependen mucho del sujeto en cuestión. En ciertos casos algunos sujetos utilizaron tres veces más tiempos que otros.

Si bien en las páginas hay en promedio 31.9 objetos, se puede apreciar que en promedio cada sujeto observa 355 objetos por experimento, lo que significa que posiciona su mirada al menos 14 milisegundos en 11 objetos por página en promedio. Estos objetos pueden ser visto

más de una vez por cada sujeto, por lo mismo más adelante se verá que el número promedio de nodos en cada grafo o el número promedio de objetos vistos al menos una vez por página es cercano a 6.

Para la parte final de esta etapa de procesamiento se necesitó de un código extraído desde *Mathworks*, plataforma que permite el aprendizaje directo del uso de *Matlab* y además la descarga de códigos libres que otros usuarios han subido a la plataforma. Esta es una ventaja importante de este lenguaje, ya que existen diversos usuarios que en plataformas de ayudas y foros suben sus códigos para evitar que futuros usuarios que vivan problemas que ellos ya han vivido. En la parte A de Anexos, se muestra este script descargado que permite generar una matriz que resume los tiempos totales por objeto por cada página.

Luego de calcular los tiempos totales por objeto en cada página, se procede a actuar de forma similar, pero esta vez determinando la frecuencia por objeto por cada página y la frecuencia por objeto por experimento. Estos datos también son almacenados en la matriz resumen, que llamaremos de ahora en adelante: H .

Las medidas de centralidad escogidas, que pronto serán detalladas, son medidas que para ser comparables necesitan también ser normalizadas (tal como se explica en el capítulo anterior). El objetivo es comparar estas medidas de centralidad con el tiempo y con la frecuencia en cada objeto, para así poder en una primera instancia determinar qué objetos son más relevante en función de estos valores.

Con el objetivo de hacer comparable los resultados entre distintos experimentos, se normalizan estos datos, haciendo así que no afecte el tiempo que estuvo la persona en cada página y que lo que importe realmente sea un porcentaje del tiempo total. Esto con el objetivo de evitar que dos nodos de grafos distintos que tengan igual valor, en el grado por ejemplo, se interpreten sus valores en función del tamaño de cada Grafo.

Sucedió que de todos los archivos *.txt*, 4 no tenían toda la información correspondiente al experimento del sujeto. Se puede asumir que la falta de datos puede deberse a dos motivos:

1. Que al momento de grabar el experimento, este no haya capturado la totalidad de este, dado que algunos tienen datos desde cierta página en adelante.
2. Que al momento de transformar los datos inicialmente desde el formato *.asc* a *.txt* haya habido un error dentro de estos datos, esto dado que existen datos inexistentes en páginas intermedias.

Es por esto que fueron eliminados de este análisis estos 4 archivos, ya que la información que aportaban venía modificada y/o no mostraba el 100% de la observación del sujeto. Dicha decisión fue conversada y analizada en conjunto con quien estuvo a cargo del experimento como también con el encargado de la primera manipulación de los datos originales.

Las fallas que mostraron los cuatro archivos mencionados se describen a continuación:

1. Datos guardados desde la página número 11 hasta la 32 sin hojas vacías entre medio.
2. Datos guardados desde la página 1 hasta la página 20 sin hojas vacías entre medio.
3. Datos faltantes desde la página 11 hasta la página 20. Con los datos completos en ambos extremos.
4. Datos guardados desde la página 1 hasta la página 10, sin hojas vacías entre medio.

3.3 Cálculo de medidas de Centralidad

Ahora con los datos procesados y filtrados se procede a hacer los cálculos explicados en Teoría Espectral de Grafos, y se considerará uno de los experimentos para ir explicando los resultados a priori. Esto con el objetivo de que el lector tenga una sencilla explicación de lo que se hizo con cada uno de los datos.

Lo primero que hay que hacer para calcular las diversas medidas de centralidad es generar una matriz de adyacencia por cada página, que muestre la relación entre los nodos que cada persona observó en cada página. Esto es importante de destacar dado que no todos los usuarios vieron todos los objetos de cada página y en algunos casos la diferencia de los tiempos en objetos y cantidad de estos observados varía notoriamente.

En la matriz de adyacencia A_G se muestra la relación de los objetos de la página 1 que uno de los usuarios observó durante su experimento:

$$A_G = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Esta matriz muestra la conexión que existe entre los objetos presentes en la página 1, sólo se muestran aquellos objetos que fueron observados por la persona y se consideran sólo aquellos en los que la observación tuvo un tiempo superior a los 14 milisegundos.

Cabe destacar que los valores que están en la diagonal de esta matriz de adyacencia son iguales a cero, lo que se debe a que uno de los requisitos del Grafo para poder hacer un análisis con alguna de las medidas a calcular es que sea un Grafo no dirigido y que no tenga loops. Esto último quiere decir que desde un objeto no puedo pasar a ese mismo objeto, no existen ciclos.

En la Figura 3.1 se puede apreciar exactamente la misma imagen que el sujeto observó durante el experimento, dándole al lector una idea más tangible de lo que se está hablando. Cabe destacar que se trata de la primera página del experimento y que no se muestra en el esquema el sentido de las conexiones entre dos nodos, dado que es un Grafo no dirigido.

Este sujeto en particular durante el experimento puso su mirada en 7 objetos, ordenados según su id serían: 1, 4, 5, 6, 22, 35 y 39. Lo que se busca con este análisis es analizar los tiempos que dedicó la persona al observar cada objeto, su frecuencia y una serie de medidas de centralidad.



Figura 3.1: Objetos observados por sujeto en página 1
Fuente: Elaboración Propia

Para calcular el grado, basta con sumar las columnas (o filas, ya que es una matriz simétrica) de la matriz de adyacencia. La suma indica el grado asociado a cada nodo, siendo estos ordenados de mayor a menor. Ej: la suma de la tercera columna es tres, esto quiere decir que el grado asociado al tercer objeto, que en este caso es 5, tiene ese valor.

Tal como se ha explicado, este valor sólo indica la cantidad de nodos que están conectados con este objeto. Para poder realizar un análisis que permita comparar valores del grado de distintos objetos, este debe ser normalizado en función de la cantidad de nodos en el grafo.

Para calcular el *Eigenvector* se utiliza también la matriz de adyacencia A_G . A diferencia del cálculo recién detallado, se procede tal como se explicó en la sección de medidas de

centralidad - Eigenvector del Capítulo 2, por lo que el Laplaciano asociado a este sujeto sería:

$$L_G = D_G - A_G = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 2 & -1 & 0 & 0 & 0 & -1 \\ -1 & -1 & 3 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 3 & 1 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 & 2 \end{pmatrix}$$

Esta matriz cumple las condiciones necesarias que permiten calcular los valores propios para luego escoger el vector propio asociado al mayor valor propio, que finalmente será nuestro *Eigenvector*.

En el siguiente vector se muestran los valores asociados a cada uno de los objetos que el sujeto observó en la página 1 durante el experimento, estos datos corresponden al vector propio asociado al mayor valor propio ($\lambda_7 = 4.6675$).

$$L_G = \begin{pmatrix} -0.1668 \\ -0.2090 \\ 0.6116 \\ -0.6441 \\ 0.3533 \\ 0.1090 \\ 0.0541 \end{pmatrix}$$

Se puede apreciar que los objetos con mayor valor en esta medida son los ubicados en la tercera, cuarta y quinta posición. Estos son los objetos 5, 6 y 22 respectivamente. Claramente los objetos 5 y 22 son importantes por la cantidad de nodos adyacentes que poseen, lo novedoso es la importancia relativa que se le asocia al objeto 6, que sólo tiene 2 nodos adyacentes.

La diferencia que tiene el nodo 6, con el 4 y con el 39, no es la cantidad de nodos adyacentes, dado que los tres tienen dos. Lo que lo hace destacar son los nodos adyacentes, puesto que en este caso el nodo 6 está conectado con los nodos 5 y 22, los nodos más importantes del grafo. Es por esto, que la relevancia que pueda tener el nodo 6 está estrechamente ligada

con la importancia que tienen los nodos adyacentes, esta es la gran ventaja de calcular el eigenvector por sobre otras medidas.

Para graficar este vector se consideran los valores absolutos, donde además de las medidas mencionadas, se incluye en el gráfico el tiempo (normalizado) que estuvo el usuario en cada objeto más la frecuencia o cantidad de veces que el sujeto observa un objeto.

3.4 Consolidación de los datos

Luego de calcular las medidas de centralidad a los datos de cada uno de los sujetos, se pretende realizar un análisis consolidado que permita determinar finalmente los *Website KeyObjects*. Este análisis previo realizado permite detectar a priori cómo se comportan estas medidas entre sí, lo que se ve reflejado en la Figura 3.2 donde se grafican las medidas de centralidad junto con el tiempo y la frecuencia de uno de los sujetos en las primeras cuatro páginas del experimento.

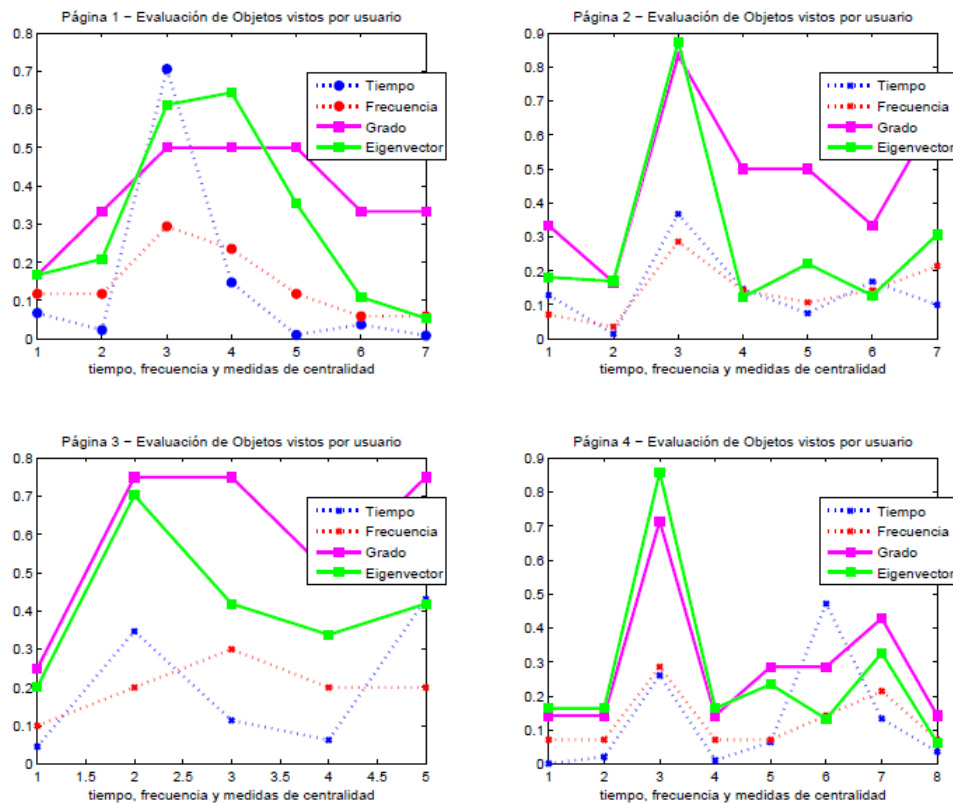


Figura 3.2: Medidas de Centralidad, tiempo y Frecuencia - Páginas 1,2,3 y 4
Fuente: Elaboración Propia

En la figura 3.2 se puede apreciar cómo se relacionan las dos medidas de centralidad

finalmente escogidas con el tiempo y la frecuencia, en especial en los peaks de las curvas. Esto da un indicio de la correlación que puedan tener estos parámetros y permiten en conjunto con otras herramientas que se mencionarán determinar el indicador final que detecte los *WKO*. Toda esta información es almacenada automáticamente por el script realizado en *Matlab*.

1	1	404	1	0.01993093	0.07692308	0.66666667	0.28867513
1	5	14756	6	0.72797237	0.46153846	1	0.8660254
1	6	5094	5	0.25130735	0.38461538	0.66666667	0.28867513
1	35	16	1	0.00078934	0.07692308	0.33333333	0.28867513
2	1	984	3	0.24538653	0.3	1	0.79091623
2	4	196	1	0.04887781	0.1	0.66666667	0.09733813
2	5	2308	4	0.5755611	0.4	1	0.59623997

Tabla 3.3: Tabla Resumen con datos de cada sujeto
Fuente: Elaboración Propia

El código hasta ahora detallado guardaba la información relevante de cada uno de los 19 sujetos en un archivo *Excel*, la estructura con la que se almacena esta información es tal como se ve en la Tabla 3.3, en donde se puede apreciar la siguiente información por cada columna:

1. id página.
2. id objeto.
3. tiempo objeto.
4. frecuencia objeto.
5. tiempo objeto normalizado.
6. frecuencia objeto normalizada.
7. grado normalizado.
8. eigenvector.

Sólo dos medidas de centralidad fueron utilizadas durante este desarrollo, el grado y el eigenvector, descartando las medidas de Closeness y Betweenness, debido a que no se ajustaban bien a grafos con muy pocos nodos. La información que pudiera entregar una medida que depende de los caminos geodésicos en un grafo con 3 o 4 nodos es mínima.

Se generó además una pestaña en los archivos Excel que almacenaba los objetos que cada sujeto observó durante su experimento, considerando el orden de cada uno de estos. Así con

esta información se podrán generar las aristas que unen los nodos. Más adelante se explicará el procedimiento utilizado y su aporte en el desarrollo de esta memoria.

En la tabla 3.4 se muestra un extracto de la data almacenada de uno de los sujetos relativa al orden de observación de los objetos por cada página. Se puede apreciar que en las tres columnas están de manera ordenada:

1. id página.
2. id objeto.
3. tiempo objeto.

A diferencia de la Tabla 3.3 que guardaba la información global de cada objeto por página, la Tabla 3.4 muestra los datos de cada objeto en el mismo orden que el sujeto posicionó su mirada por al menos 14 milisegundos.

1	5	1548
1	6	1256
1	5	234
1	6	166
1	1	404
1	5	4856
1	35	16
2	5	1368
2	6	282
2	1	136
2	4	196

Tabla 3.4: Extracto objetos visualizados por sujeto
Fuente: Elaboración Propia

Con los datos de la Tabla 3.4 se construyeron las aristas visualizadas por los sujetos, luego estas aristas se contabilizan y se consolidan por página, de modo tal que ahora se cuenta con una matriz que tiene dos columnas, en la primera se muestran las aristas y en la otra se muestra cuantos de los 19 sujetos visualizaron dicha arista.

En la tabla 3.5 se muestran los datos resumidos de la página 1, pudiendo apreciar todas las aristas que observaron los 19 sujetos que fueron parte del experimento y además la cantidad de sujetos que observó cada una de ellas, siendo 1 la menor cantidad de sujetos que pueden ver una arista que conecte dos puntos adyacentes y 19 el máximo.

5 - 6	18
1 - 6	2
1 - 5	12
5 - 35	14
6 - 22	13
6 - 35	9
22 - 35	13
22 - 39	4
4 - 39	1
4 - 5	6
5 - 22	6
23 - 39	6
5 - 23	2
6 - 23	2
5 - 39	1
23 - 35	2
22 - 23	2
1 - 3	2
3 - 5	1
35 - 39	2
1 - 22	1

Tabla 3.5: Aristas Consolidadas Página 1

Fuente: Elaboración Propia

Se puede apreciar de la Tabla 3.5 que la arista que une los objetos 5 y 6 es la arista que más se repite entre los sujetos y que sólo uno de los 19 no recorrió. Por el contrario, las aristas que unen los objetos: 4-39, 5-39, 3-5 y 1-22 la utilizó un solo sujeto. Es por esto mismo que el análisis siguiente se dividió en dos etapas, una consideró todas las aristas por cada página, independiente de cuantos sujetos pasaron de un nodo al otro, y la otra establece un mínimo de sujetos (un tercio) que hayan usado dicha arista para considerarla en el análisis.

4 Resultados

En el presente capítulo se presentarán los resultados obtenidos durante el desarrollo de esta investigación. Está dividido en cuatro partes principales:

1. *Exploración Visual*: se identifican ciertos patrones sólo en base a la observación de la data obtenida.
2. Cálculo de las medidas de centralidad a nivel consolidado.
3. Clasificación de los *Website KeyObject* con diversos indicadores desarrollados.
4. Análisis comparativo con los *WKO* detectados.

Se busca generar un análisis robusto de los resultados de manera que se pueda concluir en el próximo capítulo, dado que la metodología utilizada es estrictamente correcta. Es por ello que el procedimiento utilizado para este trabajo se basa en recomendaciones realizadas por un profesor y un alumno del departamento de Ingeniería Industrial.

4.1 Exploración Visual

Dentro del script creado en *Matlab* se hizo una función que guardaba en un archivo *Excel* datos importantes acerca de los experimento cada vez que se ejecutaba un archivo *.txt*. Estos datos presentados a continuación forman parte de este análisis previo, que intenta descubrir patrones que quizás a simple vista no son triviales.

Además se espera que estos resultados sean distintos al resultado final, de no ser así se podría establecer que el aporte de esta memoria es nulo y que la Teoría de Grafos no aporta a la detección de *WKO*, dado sólo bastaría un buen análisis exploratorio para detectar los objetos relevantes en cada una de las páginas del experimento.

4.1.1 Filtración de la Data

El primer punto de este análisis se basa en los procesos de filtrado de los datos, donde a los datos originales se les van eliminando ciertas filas que no cumplen requisitos necesarios para los objetivos propuestos. Se busca analizar el porcentaje de filtrado en cada proceso,

para dimensionar el impacto dentro de todo el proyecto.

Sujeto	Matriz Original	Tiempo (min) Experimento	Objetos = 0 Eliminados	%	Objetos Distintos Eliminados	%
1	804221	26.8	97688	12.1%	94649	96.9%
2	560263	18.7	214921	38.4%	204393	95.1%
3	268494	8.9	51279	19.1%	48340	94.3%
4	463387	15.4	83666	18.1%	67672	80.9%
5	414793	13.8	73786	17.8%	69768	94.6%
6	888555	29.6	354010	39.8%	339094	95.8%
7	398518	13.3	162881	40.9%	149316	91.7%
8	579029	19.3	123411	21.3%	110998	89.9%
9	701931	23.4	247619	35.3%	232563	93.9%
10	413264	13.8	81214	19.7%	77095	94.9%
11	722773	24.1	187671	26.0%	173595	92.5%
12	492577	16.4	109207	22.2%	92187	84.4%
13	263906	8.8	112856	42.8%	104842	92.9%
14	572359	19.1	151538	26.5%	143276	94.5%
15	422830	14.1	77544	18.3%	73387	94.6%
16	678562	22.6	198219	29.2%	184531	93.1%
17	298378	9.9	115646	38.8%	110513	95.6%
18	496209	16.5	147387	29.7%	136494	92.6%
19	445743	14.9	74071	16.6%	70905	95.7%
Promedio	520304.8	17.3	140242.8	27.0%	130716.7	92.8%
Desv. Est.	176142.5	5.9	74849.3	9.7%	71864.1	4.0%

Tabla 4.1: Datos Eliminados en Filtros

Fuente: Elaboración Propia

En la tabla 4.1 se muestra la cantidad de datos que tenía cada archivo original, los datos no eliminados en cada filtro y el porcentaje respectivo de datos que se mantuvieron. Importante es destacar que estos datos dependen de cuánto tiempo estuvo el sujeto en el experimento y cuantos objetos observó por página.

Se puede apreciar inicialmente una diferencia enorme entre los tiempos que cada sujeto le dedicó en total al experimento. Por ejemplo, el sujeto que más tiempo estuvo frente al pc durante el experimento casi alcanzó los 29,6 minutos, más de tres veces del tiempo que tomó el sujeto que menos tiempo estuvo navegando (8,8 minutos).

Preocupante por así decirlo podría ser el caso del sujeto 1, ya que durante el primer proceso de filtrado se eliminan más del 87% de sus datos, dado que en sus columnas 2 o 3 se ubicaba un cero. Esto puede deberse a que el sujeto estuvo mucho tiempo distraído mirando fuera de la pantalla o a que algo falló durante el experimento.

Durante este primer proceso, la desviación estándar cercana al 10%, muestra un valor elevado y explica la gran diferencia de los porcentajes eliminados a cada uno de los sujetos. Explicar realmente este factor es complicado y se presume que el comportamiento de los sujetos durante el experimento depende de cuánto tiempo del total estuvieron mirando el pc.

Tal como se esperaba, la cantidad de datos eliminados por presentar valores distintos en las columnas dos y tres es baja, cercana al 7% de los datos ya procesados. Esto quiere decir que las posibles fallas mencionadas son mínimas en comparación a la cantidad de datos eliminados en el proceso anterior.

Si bien en promedio las páginas tienen 31.9 objetos, los sujetos durante el experimento observan sólo 5 objetos por página, por lo que una gran cantidad de objetos no son observados por los sujetos. La cantidad total de objetos que hay en todas las páginas es 359, muchos de estos están en visibles en más de una página y se puede observar de la data que estos aparecen en total 1014 veces. En cuanto al tiempo de observación dedicado a los objetos, se puede apreciar que el promedio de tiempo que cada sujeto le dedica a un objeto es 4,8 segundos. Se comprobó además que el primer objeto visualizado tiene una importancia relativa y los sujetos tienden a cerca de 1,5 veces más de tiempo que al siguiente objeto.

4.1.2 Primer y Segundo Objeto

La importancia de lo primero que llama la atención a un observador es calificada como determinante por diseñadores, en cuanto a la importancia que tiene aquel objeto que llamó la atención al observador. Es por esto que el siguiente dato almacenado está relacionado con cuáles fueron los primeros y segundos objetos que cada sujeto observó al momento de navegar, junto con el tiempo que estuvo en cada uno de estos.

En la Tabla 4.2 se muestra un resumen de los dos primeros objetos que capturaron la atención de un sujeto en particular. En la sección A de Anexos se puede apreciar la Tabla completa.

Es llamativo que en la mayoría de los casos este sujeto comenzaba la observación de cada página en el objeto 6 o en el 5, ocurriendo esto en 23 de las 32 páginas. Esto se debe básicamente a que estos dos objetos en particular se repiten en todas las páginas y tienen una posición privilegiada.

En la Figura 4.1 se muestra el objeto 5: es la imagen que se ubica al costado derecho del logo del MBA y el objeto 6: es la barra de menú. Es lógico que estos sean dos de los objetos más vistos, tanto por el tamaño, la posición y la recurrencia que tiene un navegador

Id Página	Primer Objeto	t (miliseg)	Segundo Objeto	t (miliseg)
1	6	158	5	2960
2	6	334	5	228
3	59	258	6	210
4	5	392	59	1062
5	5	404	6	2054
6	6	1166	87	26
7	5	606	6	22
8	5	1382	59	200
9	6	298	51	190
10	59	668	51	134
.
.
.
32	359	300	59	442

Tabla 4.2: Primer y Segundo Objeto
Fuente: Elaboración Propia

web hacia el menú, que es el lugar sector que uno recurre para saber que ofrece la página.

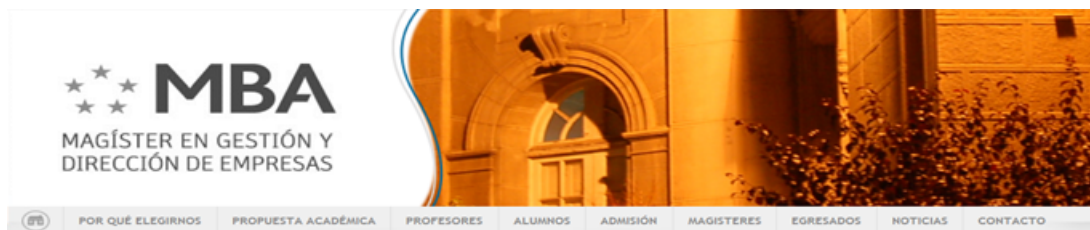


Figura 4.1: Objetos 5 y 6
Fuente: Elaboración Propia

Si bien lo que se busca en este trabajo es la identificación de objetos dentro de una página web, observar a nivel individual el comportamiento es simplemente para tener una idea. Es por esto que además se realiza un análisis consolidado previo a la metodología a desarrollar.

A continuación se muestra en la Tabla 4.3 parte de los datos obtenidos del análisis de los primeros objetos visualizados por todos los sujetos durante sus experimentos, esto se considerará como un primer approach para algunas conclusiones acerca de la exploración visual. En la columna 2 y 3 se presenta el tiempo promedio que le dedican los sujetos al observar el primer y segundo objeto por cada una de las 32 páginas, se agrega una cuarta columna con el delta tiempo entre el primer y segundo objeto.

De la Tabla 4.3 se desprende algo no trivial, ya que en 22 de las 32 páginas, el tiempo

Id Página	Primer Objeto	Segundo Objeto	Delta tiempo
1	763.1	872.7	-109.7
2	1492.5	228.0	1264.5
3	618.9	528.0	90.9
4	1353.7	405.6	948.1
5	722.5	515.1	207.5
6	445.5	535.3	-89.8
7	777.1	367.2	409.9
8	557.1	186.5	370.5
9	527.4	271.7	255.7
10	615.8	324.6	291.2
.	.	.	.
.	.	.	.
.	.	.	.
32	506.2	262.6	243.6

Tabla 4.3: Primer y Segundo Objeto tiempos
Fuente: Elaboración Propia

promedio que cada sujeto le dedica a observar el primer objeto es mayor que el tiempo que le otorga al segundo objeto. De hecho en cinco páginas el tiempo promedio que se posiciona la vista en el primer objeto es medio segundo más.

Se podría asumir que el primer objeto genera una mayor atención al sujeto que el segundo, si bien es un supuesto. Se podría asumir como resultado de este trabajo que el primer objeto que el usuario observa tiende a ser observado más que el segundo.

En la Tabla 4.3 se los *Website KeyObjets* en función de los primeros y segundos objetos. Estos serían los objetos: 5, 59, 6, 51, 121, 205, 267, 309, 356 y 359.

En paralelo se realizó un análisis en función del tiempo. Este enfoque se hace con el objetivo de comparar los *key objects* ya detectados con los que se detecten con las medidas de centralidad.

Se almacenó en un archivo *Excel* los tiempos de observación que cada sujeto le dedicó en promedio a los objetos y páginas, con el objetivo de identificar algún patrón hasta ahora no observado. En las Tablas 5.1 y 5.2 se muestra lo recién mencionado.

Se puede detectar claramente cuáles son los objetos con más tiempo dedicado por parte de los sujetos, estos tiempos están en función de la cantidad de veces que aparecen en cada página, por ende no se puede asumir que sean *Website KeyObjetc* sólo por el hecho de tener

Objeto	Promedio	Objeto	Promedio	Objeto	Promedio	Objeto	Promedio
1	7407	86	203	208	49	276	67
3	342	87	216	209	15	277	197
4	12579	88	2295	211	616	278	662
5	129611	89	553	212	749	279	8
6	15951	94	975	213	61	280	260
11	276	95	6	214	935	281	15
12	149	96	689	215	6608	282	29
14	14	103	81	216	154	294	8
15	8	104	304	218	657	295	1533
22	514	105	4	219	52	296	447
23	519	121	407	221	88	304	38
35	767	122	332	225	2	306	4702
39	399	123	177	229	4	307	2692
51	3566	124	116	238	22	308	1181
52	1141	125	185	242	5	309	2471
54	128	126	160	243	2347	311	471
57	1	127	120	244	5120	312	651
59	9872	128	23	245	7288	315	53
60	4383	129	15	246	16	341	406
61	248	130	9	247	84	342	16
70	16	134	1857	264	253	343	223
71	9	138	867	265	2760	347	377
73	1415	195	5035	266	749	349	107
74	1604	197	122	267	5014	356	1701
75	1812	201	8	268	54	357	1434
76	103	205	173	269	1954	358	1
84	3044	206	74	271	1607	359	546
85	6	207	37	272	1		

Figura 4.2: Tiempo Promedio por Objeto
Fuente: Elaboración Propia

tiempos acumulados elevados. De asumir que los objetos con más tiempo promedio por sujeto son los más relevantes se podría estar obviando algún detalle que haría de las conclusiones sin fundamentos.

Incluso si se estimara el tiempo promedio por página de cada objeto no se estaría evaluando lo que sucede cuando un sujeto ve un mismo objeto en reiteradas ocasiones. Este es un punto relevante a considerar al momento de las conclusiones y futuras líneas de acción, quizá plantear un nuevo experimento donde los objetos no se repitan por cada página o que se repitan de forma homogénea, pero no que suceda que un objeto esté en todas las páginas y que hayan otros que sólo se encuentren en una o dos, podría hacer de las conclusiones y resultados más fuertes y robustos.

Los objetos con mayores tiempos promedios son: 1, 4, 5, 6, 51, 59, 60, 84, 195, 215, 244,

245, 267 y 306.

Las páginas con mayores tiempos promedios son: 1, 2, 5, 13, 20, 25, 28 y 32.

Una interpretación válida con este resultado es que no se puede ponderar de igual forma los *objetos claves* detectados que pertenecen a páginas con tiempos promedios menores que a los *WKO* detectados en las páginas con tiempos promedios mayores recién declaradas.

A priori no se puede concluir algo en base a este factor, pero es aceptable destacar que no tiene la misma relevancia en un sitio web, el *objeto clave* de la página con mayor tiempo promedio, dado que este podría estar compitiendo con varios otros objetos que llamaron mucho la atención a la mayoría de los sujetos. En cambio el objeto de la página 18 por ejemplo, es un supuesto *Website KeyObject* pero no tiene mucha competencia, dado que en esta página no hay muchos otros objetos, entonces no es trivial concluir que sólo el tiempo promedio en cada objeto basta para establecer importancias relativas dentro de un sitio web.

Id_pagina	Total página
1	16609
2	17312
3	5580
4	8957
5	12236
6	3794
7	10929
8	3132
9	3089
10	5658
11	6069
12	10468
13	18764
14	1708
15	3138
16	8509
17	3942
18	442
19	8894
20	10394
21	5958
22	8394
23	9837
24	6293
25	10136
26	9909
27	8183
28	13387
29	9860
30	6526
31	7531
32	12846

Figura 4.3: Tiempo Promedio por Página
Fuente: Elaboración Propia

4.2 Resultados Consolidados

El procedimiento para determinar los *Website KeyObjects* es lo fundamental en este trabajo de memoria, es por eso que seguir una rigurosa metodología de trabajo es lo que podría hacer que los resultados finales puedan ser robustos o que no tengan validez absoluta en caso de errar el camino o sustentar supuestos en ideas con carencia de sentido.

Consolidar toda la información de los experimentos de cada uno de los sujetos es relevante para determinar finalmente los *WKO*. Existen diversos métodos en Teoría de Grafos que permiten agrupar información de varios Grafos de forma tal que se pueda evaluar un sólo Grafo que entregue la información resumida, a continuación se detallará el procedimiento escogido.

Se determinó trabajar con las aristas generadas por cada uno de los 19 sujetos en las 32 páginas que observaron durante su experimento. Se almacenaron las aristas utilizadas en función de los recorridos que realizaban los sujetos entre todos los objetos dentro de una misma página, y se cargaban en un archivo *.xls* (revisar Capítulo 3 para más detalle).

Estos conjuntos de aristas junto con los objetos generan en cada página un grafo por cada sujeto. Para consolidar dicha información y poder establecer grafos que resuman todo el experimento se optó por contabilizar la cantidad de veces que cada arista fue utilizada en cada página por todos los sujetos. Ejemplo: Si en la página 1 cinco sujetos pasaron del objeto 5 al 6, entonces la arista 5-6 (que une ambos objetos) tiene un peso igual a 5. En el fondo se genera un mega Grafo compuesto por cada uno de los grafos que los sujetos crearon al momento de navegar en las 32 páginas.

Podrían haber aristas que tengan un peso igual a cero, osea que ningún sujeto haya pasado de un cierto objeto a otro en alguna página. Caso contrario podría darse que los 19 sujetos usen una arista que une dos objetos. Entonces hay que considerar de forma distinta a estas aristas, es por ello que el análisis siguiente se divide en dos.

La primera parte evalúa todo el Grafo consolidado considerando todas las aristas que tengan peso mayor o igual a uno, en simples palabras si ninguno de los sujetos paso del objeto A al objeto B, entonces la arista AB no será parte de este grafo final.

En la segunda parte se optó por trabajar con un número de corte mínimo para determinar si una arista será finalmente parte de este Grafo consolidado. Para escoger el peso óptimo para que una arista sea parte de esta sección se realizaron 3 simulaciones distintas con tres pesos mínimos distintos: 6, 9 y 12. Resultando finalmente que el número 6 era el

mejor dado que con 9 y 12 se perdía mucha información y quedaban Grafos con 3 nodos en promedio por página.

A estos dos grafos finales (pesos mínimos 0 y 6) se les calcularon las medidas de centralidad y los coeficientes de correlación, luego se compararon los resultados y finalmente se concluyó en base a la hipótesis de investigación propuesta.

Se usó el coeficiente de correlación lineal de Pearson para calcular la dependencia entre las variables, asumiendo que la data obtenida sigue una distribución Normal. Este supuesto puede ser considerado uno de los puntos a evaluar dentro de las posibles líneas de acción, en caso de no serlo se podrían considerar otras correlaciones.

Lo que sigue en este análisis de ahora en adelante está dividido en dos partes:

1. Considerando todas las aristas registradas por los sujetos.
2. Considerando sólo aquellas aristas que fueron vistas por al menos un tercio de los sujetos.

4.2.1 Parte A: Análisis con todas las aristas observadas

Correlación de los parámetros

Para tener una primera impresión acerca del comportamiento de los parámetros que se utilizarán para la detección de los *Website KeyObjects*, se realizó un análisis para saber que tan correlacionados se encuentran las 7 medidas que se utilizarán:

1. Tiempo promedio.
2. Tiempo promedio normalizado.
3. Frecuencia promedio.
4. Frecuencia promedio normalizada.
5. Grado.
6. Grado normalizado.
7. Eigenvector.

Lo primero que se hizo fue graficar los Scatterplots correspondiente a parejas de estas medidas para comprender como se comportan entre ellas. En la Figura 4.4 se pueden apreciar

6 gráficos que muestran dicha relación.

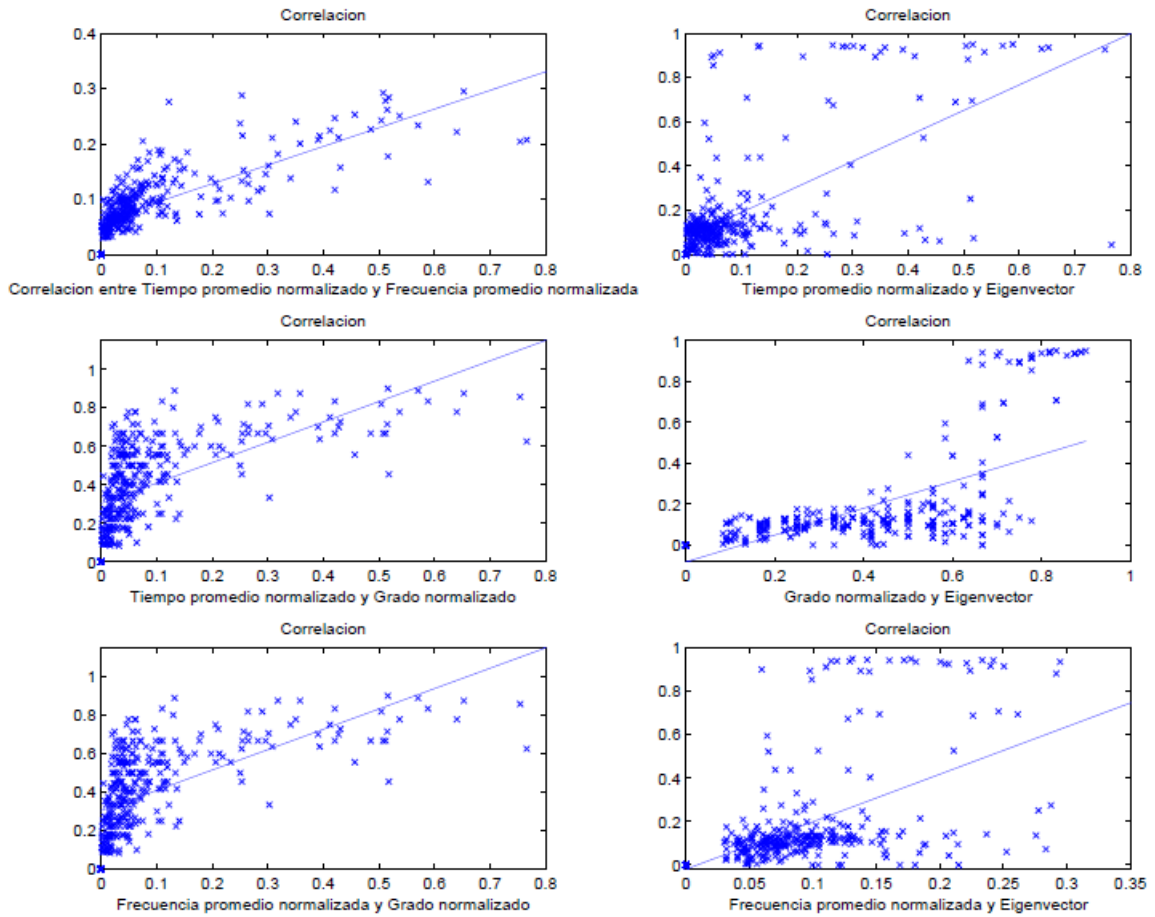


Figura 4.4: Scatterplots - Todas las aristas
Fuente: Elaboración Propia

Sólo con la observación lo representado en la Figura 4.4 no se puede concluir nada, pero si se puede tener una intuición acerca del comportamiento de los parámetros, en especial al mirar los gráficos: *Tiempo promedio normalizado-Frecuencia promedio normalizada* y *Frecuencia promedio normalizada-Grado normalizado*. Estos parecieran ser pares de medidas que a priori tienen una correlación alta.

La correlación de Pearson de estos duetos de parámetros muestra la relación que puede haber entre las variables, mientras más cercana a uno significa que los parámetros se comportan de manera similar al aumentar o disminuir su valor. En las Tabla 4.4 y 4.5 se pueden apreciar lo mencionado.

Se puede ver de ambas Tablas, que la correlación aumenta considerablemente cuando se utilizan los valores normalizados, en especial la correlación de Pearson entre la Frecuencia y

Tiempo Promedio	Frecuencia Promedio	Grado	Eigenvector
1	1	1	1
0.705510751	0.44141094	0.585729538	
0.49962348	0.376889327		
0.575026291			

Tabla 4.4: Correlación Parámetros no Normalizados
Fuente: Elaboración Propia

Tiempo Promedio Norm	Frecuencia Promedio Norm	Grado Norm	Eigenvector
1	1	1	1
0.770800831	0.610833991	0.625146779	
0.593179263	0.501099997		
0.623605802			

Tabla 4.5: Correlación Parámetros Normalizados
Fuente: Elaboración Propia

el Grado (ambos normalizados).

Era esperable que el par con el mayor índice de correlación fuera el par: Tiempo promedio normalizado - Frecuencia promedio normalizada. Esto dado que en el análisis visual, estos pares de datos eran los que tenían un alto valor en los mismos objetos, lógicamente estos parámetros están relacionados, puesto que si un objeto tiene una alta frecuencia de observación, debería suceder que además tuviera un alto tiempo de observación, esto suponiendo que el tiempo promedio de observación tiene una baja variabilidad (lo que es real en este caso).

Si se observa la Tabla 4.5 se puede ver que los pares de parámetros que tienen una mayor correlación son:

1. Tiempo promedio normalizado \times Frecuencia promedio normalizada.
2. Grado normalizado \times Eigenvector.
3. Tiempo promedio normalizado \times Eigenvector.
4. Frecuencia promedio normalizada \times Grado normalizado.

Será uno de estos cuatro pares de medidas a calcular el que finalmente decidirá cuáles serán los *Website KeyObjects*. El motivo de escoger estos indicadores es que se están ponderando valores correlacionados, aumentando la brecha que existe entre los objetos que destacan con

los que tienen los mínimos valores. Es decir que se le está dando mayor relevancia a datos con altos índices de forma tal de poder ver entre los que destacan cuáles se alejan del promedio, afirmando finalmente que estos serán los *KeyObjects* de esta nueva metodología.

A estas cuatro indicadores se le agregaron cuatro más para probar básicamente que resultados arrojaría y ver si podrían finalmente detectar con una mejor precisión los *WKO*. El motivo de agregar estos nuevos grupos de parámetros es ver si al ponderar por otro indicador más el efecto se ve aumentado en caso de haber correlación o si es que el supuesto establecido carece de sentido. De hecho es esperable que el último indicador, el que multiplica las cuatro medidas, no entregue resultados coherentes pues está multiplicando medidas con alta y baja correlación al mismo tiempo. Esto debería disuadir el efecto de ponderar las variables.

Los indicadores agregados son los siguientes:

1. Tiempo promedio normalizado \times Frecuencia promedio normalizada \times Grado normalizado.
2. Tiempo promedio normalizado \times Frecuencia promedio normalizada \times Eigenvector.
3. Tiempo promedio normalizado \times Grado normalizado \times Eigenvector.
4. Tiempo promedio normalizado \times Frecuencia promedio normalizada \times Grado normalizado \times Eigenvector.

En las Tablas 4.6 y 4.7 se muestran los objetos con mayor valor en cada uno de estos 8 indicadores. Estos vendrían siendo los *Website KeyObjects* generados en base a cada una de las funciones propuestas.

En estas dos Tablas se muestran los 25 objetos más importantes según los 8 indicadores mencionados, para escoger finalmente la función que determinará los *Website KeyObjects* se realizó un análisis comparativo de los 25 objetos que arrojó cada columna y se ponderó por quién entregaba los más valiosos según el valor asociado.

La columna que entregó los mejores resultados fue la columna número 6, esta columna está determinada por la función: *Tiempo promedio normalizado \times Frecuencia promedio normalizada \times Eigenvector*. Por lo tanto los *WKO* que se generen con esta metodología serán los que se compararán con el análisis exploratorio realizado, memorias previas y el análisis del *webmaster*.

Se compararán *Website keyObject* detectados con este indicador con los detectados por Gustavo Martinez en [10]. Cabe destacar que en el desarrollo de su trabajo sólo se detectaron

T prom norm F prom norm	Grado norm Eigen	T prom norm Eigen	F prom norm Grado norm
215	215	215	215
244	245	245	5
306	60	269	306
5	269	5	243
245	84	60	244
243	86	84	60
267	5	195	267
265	73	73	86
307	195	307	245
73	88	88	73
60	307	86	269
269	89	243	96
195	280	280	307
94	138	265	266
356	243	89	309
96	357	75	195
214	211	244	94
212	278	278	356
309	6	94	87
75	221	267	265
218	343	306	89
84	104	96	211
247	75	211	88
74	96	218	84
88	349	214	212

Tabla 4.6: Website KeyObjects - Indicadores A

Fuente: Elaboración Propia

20 Objetos Claves, por lo que ahora sólo se considerarán los 20 primeros *WKO*, de modo que la comparación entre ambas metodologías sea justa. En la Tabla ubicada en Anexos se muestran los Key Object detectados por Martínez.

Hasta ahora no se ha contado con la opinión del webmaster, es por esto que la última columna de la matriz no está completa.

Comparación con Martínez [10]

Realizando una comparación con los objetos detectados por Martínez, se puede decir que:

1. De los 20 objetos declarados por Martinez, se tienen en común 10 objetos.
2. Ninguno de los 4 objetos que no son keyobjects y que Martinez había declarado WKO

T norm F norm G norm	T norm F norm Eigen	T norm G norm Eigen	T norm F norm G norm Eigen
215	215	215	215
244	245	245	245
245	5	269	5
306	60	60	60
5	269	5	269
267	73	84	73
243	195	73	195
265	243	195	84
60	84	307	307
307	307	88	243
269	244	86	86
73	86	265	88
195	88	243	244
94	306	280	306
96	89	89	265
214	265	75	89
356	94	94	267
212	267	96	94
84	75	267	96
309	280	214	75
75	96	244	280
88	211	211	211
86	356	306	356
266	212	278	214
74	309	212	212

Tabla 4.7: Website KeyObjects - Indicadores B
Fuente: Elaboración Propia

están dentro de los 25 objetos detectados.

- Hay 10 objetos que deben ser revisados por el webmaster, para determinar si son o no WKO.

En la sección de Anexos se muestra en la Tabla 5.1 los *Objetos Principales* detectados por Martinez, esta es la base usada para la comparación detallada.

Se mostrará a continuación un análisis similar al detallado, sólo que ahora se considerarán sólo aquellas aristas que hayan sido observadas por al menos un tercio de los sujetos. Esto quiere decir que si 5 sujetos pasan del objeto 5 al 39, no se considerará esa arista en el procedimiento y a menos que otra arista contenga a estos objetos, no serán parte del análisis.

Object Number	Website Object	Name	Is Keyobject?
1	Objeto 215	paragraph propuesta 1	Si
2	Objeto 245	paragraph perfil egresados 1	Si
3	Objeto 5	header picture	Si
4	Objeto 60	admission paragraph 1	Incierto
5	Objeto 269	paragraph quienes 2	Incierto
6	Objeto 73	plan de estudios chart	Incierto
7	Objeto 195	paragraph detalle noticia 1	Si
8	Objeto 243	paragraph plataforma 1	Si
9	Objeto 84	paragraph 1 doble grado	Incierto
10	Objeto 307	paragraph metodologia 2	Si
11	Objeto 244	image plataforma 1	Si
12	Objeto 86	paragraph 1 folleto	Incierto
13	Objeto 88	paragraph 1 características	Incierto
14	Objeto 306	paragraph metodologia 1	Si
15	Objeto 89	paragraph 2 características	Incierto
16	Objeto 265	paragraph corporacion 1	Si
17	Objeto 94	paragraph 1 contacto	Incierto
18	Objeto 267	paragraph quienes 1	Si
19	Objeto 75	paragraph 2 financiamiento	Incierto
20	Objeto 280	paragraph contenido 2	Incierto

Tabla 4.8: Lista de Website Keyobject
Fuente: [Elaboración Propia]

4.2.2 Parte B: Análisis con aristas observadas mínimo 6 veces

Tal como se aclaró, a continuación se actuará con el mismo protocolo que en la Parte A sólo que utilizando aquellas aristas observadas por a lo menos un tercio de los sujetos participantes. Esto con el objetivo de comparar luego los *Website KeyObjects* y ver si es que existe un cambio significativo al generar esta modificación.

Correlación de los parámetros

Al igual que en la parte A, se trabajó con los mismos 7 parámetros y se procedió a realizar los mismos procedimientos para ver cómo estos se comportaban entre ellos, para finalmente establecer cuáles serían parte de las medidas finales para la clasificación de los *Website KeyObjects*.

Se graficaron los Scatterplots correspondiente a parejas de estas medidas para comprender como se comportan entre ellas. En la Figura 4.5 se pueden apreciar 6 gráficos que muestran dicha relación. Es necesario recordar que no son los mismos diagramas que en la parte A.

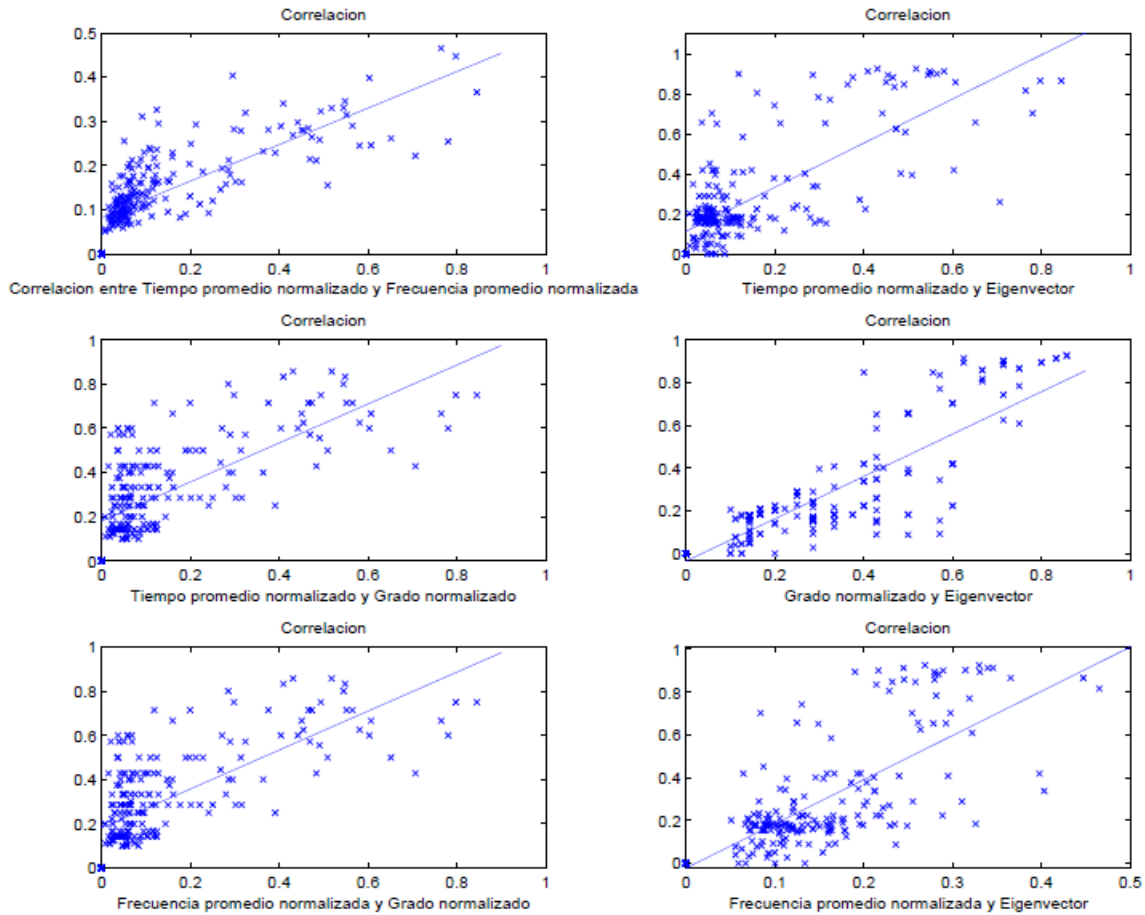


Figura 4.5: Scatterplots - Más de 6 aristas
Fuente: Elaboración Propia

En este caso el gráfico asociado al par: Grado normalizado - Eigenvector pareciera ser el que muestra una mayor relación lineal entre ambas medidas. Por el contrario la frecuencia promedio normalizada pareciera no tener ninguna relación con el Eigenvector, pues los puntos en el diagrama están distribuidos en el gráfico sin mostrar un comportamiento relevante. Esto se confirmará al igual que en la Parte A con un cálculo del coeficiente de Correlación de Pearson.

En las Tabla 4.9 y 4.10 se pueden apreciar estos valores, tanto para los valores normalizados y no normalizados.

Nuevamente sucede algo similar a la Parte A, dado que al estar los datos normalizados, los coeficientes aumentan considerablemente. En este caso el coeficiente de Pearson del par Grado normalizado y Eigenvector es el más alto, llegando a un valor superior a 0,8.

En este caso los coeficientes asociados a los pares de todas las variables normalizadas son

Tiempo Promedio	Frecuencia Promedio	Grado	Eigenvector
1			
0.681721326	1		
0.626264127	0.559019402	1	
0.615549082	0.491377101	0.746625182	1

Tabla 4.9: Correlación Parámetros no Normalizados
Fuente: Elaboración Propia

Tiempo Promedio Norm	Frecuencia Promedio Norm	Grado Norm	Eigenvector
1			
0.771708292	1		
0.680022038	0.647561389	1	
0.726813708	0.696336172	0.810439504	1

Tabla 4.10: Correlación Parámetros Normalizados
Fuente: Elaboración Propia

superiores a 0.6, esto quiere decir que todos presentan un comportamiento similar, pero no necesariamente se puede concluir que existe una implicancia entre estos valores.

Si se observa la Tabla 4.10 se puede ver que los pares de parámetros que tienen una mayor correlación son:

1. Tiempo promedio normalizado \times Frecuencia promedio normalizada.
2. Grado normalizado \times Eigenvector.
3. Eigenvector \times Tiempo promedio normalizado.
4. Grado normalizado \times Tiempo promedio normalizado.

Serán estos 4 pares las medidas a calcular que finalmente decidirán cuáles serán los *Website KeyObjects*.

Al igual que en la Parte A se agregarán a estas cuatro indicadores los indicadores:

1. Tiempo promedio normalizado \times Frecuencia promedio normalizada \times Grado normalizado.
2. Tiempo promedio normalizado \times Frecuencia promedio normalizada \times Eigenvector.
3. Tiempo promedio normalizado \times Grado normalizado \times Eigenvector.

4. Tiempo promedio normalizado \times Frecuencia promedio normalizada \times Grado normalizado \times Eigenvector.

En las Tablas 4.11 y 4.12 se muestran los objetos con mayor valor en cada uno de estos 8 indicadores. Estos vendrían siendo los *Website KeyObjects* generados en base a cada una de las funciones propuestas.

T prom norm F prom norm	Grado norm Eigen	T prom norm Eigen	F prom norm Grado norm
244	94	215	215
215	306	306	244
245	88	267	306
5	5	5	245
306	267	94	5
243	215	244	267
267	359	212	265
214	211	265	94
212	212	245	60
265	244	88	73
211	265	211	212
96	356	195	88
94	309	214	195
73	195	243	243
60	96	96	96
269	243	356	356
307	73	309	211
84	245	73	309
356	271	75	214
309	134	295	271
195	60	84	84
75	59	278	269
266	278	307	307
88	51	269	75
74	343	60	74

Tabla 4.11: Website KeyObjects - Indicadores A2
Fuente: Elaboración Propia

En estas dos Tablas se muestran los 25 objetos más importantes según los 8 indicadores mencionados, para escoger finalmente la función que determinará los *Website KeyObjects* se realizó un análisis comparativo de los 25 objetos que arrojó cada columna y se ponderó por quién entregaba los más valiosos según el valor asociado.

La columna que entregó los mejores resultados fue la columna número 7, esta columna está determinada por la función:

T norm F norm G norm	T norm F norm Eigen	T norm G norm Eigen	T norm F norm G norm Eigen
244	215	215	215
215	306	306	306
306	5	5	5
5	244	267	244
245	267	94	267
267	212	244	94
243	94	88	212
212	245	212	211
94	211	265	245
265	243	245	243
73	265	211	265
211	214	195	88
60	96	356	96
96	88	243	356
214	356	96	309
356	195	309	73
309	309	73	195
88	73	214	214
195	307	60	60
84	84	75	84
269	60	271	6
307	269	84	307
266	75	278	269
75	295	307	75
271	6	295	271

Tabla 4.12: Website KeyObjects - Indicadores B
Fuente: Elaboración Propia

Tiempo promedio normalizado \times *Frecuencia promedio normalizada* \times *Eigenvector*,

Los *WKO* que se generen con esta metodología serán los que se compararán con el análisis exploratorio realizado, memorias previas y el análisis del *webmaster*.

Compararemos los *Website keyObject* detectados con este indicador con los detectados por Gustavo Martinez en [10]. Cabe destacar que en [10] sólo se detectaron 20 Objetos, por lo que ahora sólo se considerarán los 20 primeros *WKO*. En la Tabla ubicada en Anexos se muestran los Key Object detectados por Martinez.

Object Number	Website Object	Name	Is Keyobject?
1	215	paragraph propuesta 1	Si
2	306	paragraph metodologia 1	Si
3	5	header picture	Si
4	267	paragraph quienes 1	Si
5	94	paragraph 1 contacto	Incierto
6	244	image plataforma 1	Si
7	88	paragraph 1 características	Incierto
8	212	title reglamento 2	Incierto
9	265	paragraph corporacion 1	Si
10	245	paragraph perfil egresados 1	Si
11	211	title reglamento 1	Incierto
12	195	paragraph detalle noticia 1	Si
13	356	picture detalle profe 1	Si
14	243	paragraph plataforma 1	Si
15	96	paragraph 1 generaciones	Incierto
16	309	paragraph noticia 1	Incierto
17	73	plan de estudios chart	Incierto
18	214	paragraph reglamento 2	Incierto
19	60	admission paragraph 1	Incierto
20	75	paragraph 2 financiamiento	Incierto

Tabla 4.13: Lista de Website Keyobject
Fuente: [Elaboración Propia]

Comparación con Martínez [10]

Antes de comparar los objetos claves con la opinión del webmaster se detallará un breve análisis comparativo con lo realizado por Martínez, en él se podrá tener una primera impresión de los resultados como la cantidad de objetos en común, los WKO que se repiten y cuáles serían los objetos claves que harían la diferencia entre estas dos metodologías.

1. De los 20 objetos claves declarados por Martinez, se tienen en común 11 objetos todos ellos son WKO.

Este punto es el más llamativo y si bien más del 50% de los objetos claves se repiten, todos estos son *WKO*. Al observar los objetos, tal como se puede ver en anexos, se puede apreciar que los objetos en común tienen características similares, osea que su importancia relativa puede deberse al tamaño o al formato del objeto, ej: que sea un texto y haya que dedicarle tiempo para leerlo.

Los objetos que difieren son los que deberían resaltar el efecto que capturan las medidas de centralidad, el efecto de estar anexado a objetos importante.

2. Ninguno de los 4 objetos que no son keyobjects y que Martinez había declarado WKO,

están dentro de los 20 objetos detectados.

Este factor es clave, dado que de los objetos que el webmaster rechazó, hay dos que se repiten en casi todas las páginas y según la opinión del webmaster estos no se debieran considerar como WKO dado que no son parte de lo fundamental de cada página sino que forman la base de cada una de las páginas.

Se podría decir que las medidas de centralidad pudieron representar el efecto por sobre el tiempo, una ponderación que tiene un efecto claro, no considerar como WKO objetos por el sólo hecho de capturar más la atención del sujeto, hay argumentos que van más allá y estos son los que se intentan explicar con la medida Eigenvector.

3. Hay 9 objetos que deben ser revisados por el webmaster, para determinar si son o no WKO.

Finalmente, estos 9 objetos revisados por el webmaster son los que determinarán la eficiencia del modelo planteado. Esta etapa es un tanto ambigua dado que todo el proyecto consistía en darle objetividad al asunto de la determinación de cuáles objetos eran y cuáles no WKO. Se plantea en el Capítulo de Conclusiones que esta sea una etapa a excluir de la metodología puesto que ya se tienen diversas líneas bases, incluyendo esta, que permiten comparar futuros proyectos. Así se podría establecer si nuevos trabajos aportan o no a esta metodología planteada.

4.2.3 Comparación Final

Este análisis se divide en comparar el trabajo realizado durante este trabajo de memoria con la exploración visual y con la opinión del web master.

El primer punto sirve para tener un primer approach acerca de los resultados objetivos que se pretende consolidar en este trabajo, siendo finalmente la opinión del webmaster quién determine el nivel de efectividad de todo este desarrollo. Se acota tanto en los alcances como en los trabajos propuesto a futuro acerca de esto y como se plantea una línea de acción que optimice esta etapa.

4.2.4 Comparación Resultados Parte A y Parte B

Lo importante de realizar una comparación entre estas dos etapas es determinar si finalmente la intuición acerca de excluir las aristas que tienen una baja frecuencia de visualización aporta información o si mejora los resultados en comparación con la primera parte que utiliza todas las aristas, independiente de la cantidad de veces que se observó. Recordando que no se les asigna un peso por frecuencia en ninguna de las dos Partes.

1. De los 20 Objetos detectados, 14 coinciden en ambos procedimientos.
Los objetos coincidentes son: {5, 60, 73, 75, 88, 94, 195, 215, 243, 244, 245, 265, 267y306}.
Esto muestra que el cambio, aunque mínimo, representa el 30% de los objetos claves.
Lo importante es ver si estos objetos distintos son los que finalmente se consideran KeyObjects.
2. En cuanto a los objetos distintos, en la parte A se clasificaron como *Website KeyObject* los siguientes objetos: {84, 86, 89, 269, 280y307}.
En cambio, los objetos clasificados como *WKO* en la parte B son: {96, 211, 212, 214, 309y356}.
3. La efectividad de ambas partes se retomará una vez mencionada la opinión del webmaster.

4.2.5 Comparación con *webmaster*

Esta es la sección más relevante del trabajo pues con la opinión del webmaster se pudo finalmente determinar la efectividad de todo este trabajo y además determinar cuál de las dos partes es finalmente la que será la base de esta memoria.

La parte que tenga una mayor eficiencia dependiendo de cuáles son los *WKO* será la que defina los objetos claves finales, esta es la que se compara con Martínez y es la que servirá de base para trabajos futuros.

En las Tablas 4.14 y 4.15 se puede apreciar la opinión final del webmaster en función de los dos conjuntos de objetos escogidos.

De los 20 objetos escogidos en ambas partes, 3 y 5 de ellos fueron rechazados respectivamente por el webmaster, produciendo una eficiencia en la detección de Website KeyObjects en la Parte A del 85% y una eficiencia igual a 75% en la Parte B. Este valor se ve afectado por la experiencia del webmaster, quién a diferencia de otras memorias modificó algunos aspectos al momento de decidir cuándo un objeto es clave y cuando no.

En particular los dos objetos que pasaron de ser claves a objetos comunes son los objetos 5 (header picture) y 244 (image plataforma 1).

Si sólo se consideran los objetos determinados en la Parte A, y se comparan con la exploración visual se puede notar que 8 de los objetos definidos y aceptados por el webmaster como claves pertenecen a los objetos con mayor tiempo de observación promedio. Esto es análogo a decir que de los 14 objetos que tienen mayores tiempo de observación promedio, 6 no son *WKO*.

Object Number	Website Object	Name	Is Keyobject?
1	Objeto 215	paragraph propuesta 1	Si
2	Objeto 245	paragraph perfil egresados 1	Si
3	Objeto 5	header picture	NO
4	Objeto 60	admission paragraph 1	Si
5	Objeto 269	paragraph quienes 2	Si
6	Objeto 73	plan de estudios chart	Si
7	Objeto 195	paragraph detalle noticia 1	Si
8	Objeto 243	paragraph plataforma 1	Si
9	Objeto 84	paragraph 1 doble grado	Si
10	Objeto 307	paragraph metodologia 2	Si
11	Objeto 244	image plataforma 1	No
12	Objeto 86	paragraph 1 folleto	Si
13	Objeto 88	paragraph 1 características	Si
14	Objeto 306	paragraph metodologia 1	Si
15	Objeto 89	paragraph 2 características	Si
16	Objeto 265	paragraph corporacion 1	Si
17	Objeto 94	paragraph 1 contacto	Si
18	Objeto 267	paragraph quienes 1	Si
19	Objeto 75	paragraph 2 financiamiento	Si
20	Objeto 280	paragraph contenido 2	No

Tabla 4.14: Lista de Website Keyobject
Fuente: [Elaboración Propia]

De los 15 objetos en común que arrojó la comparación entre las Partes A y B, resulta que de los 15 objetos sólo 2 no son considerados como *Website KeyObjects*. Se destaca además que los únicos dos objetos rechazados por el webmaster son el objeto 5 (header picture) y el objeto 244 (image plataforma 1), estos son justamente los 2 objetos que en opiniones previas si eran considerados claves y que esta vez no corrieron la misma suerte.

Object Number	Website Object	Name	Is Keyobject?
1	215	paragraph propuesta 1	Si
2	306	paragraph metodologia 1	Si
3	5	header picture	No
4	267	paragraph quienes 1	Si
5	94	paragraph 1 contacto	Si
6	244	image plataforma 1	No
7	88	paragraph 1 características	Si
8	212	title reglamento 2	No
9	265	paragraph corporacion 1	Si
10	245	paragraph perfil egresados 1	Si
11	211	title reglamento 1	Si
12	195	paragraph detalle noticia 1	Si
13	356	picture detalle profe 1	No
14	243	paragraph plataforma 1	Si
15	96	paragraph 1 generaciones	Si
16	309	paragraph noticia 1	Si
17	73	plan de estudios chart	Si
18	214	paragraph reglamento 2	No
19	60	admission paragraph 1	Si
20	75	paragraph 2 financiamiento	Si

Tabla 4.15: Lista de Website Keyobject

Fuente: [Elaboración Propia]

5 Conclusiones

El propósito de este trabajo de memoria es trabajar en base a una metodología previa que ha sido desarrollada por diversos autores en trabajos que datan desde ya más de 10 años. Se propone que al agregarle a esta metodología de detección de *Website KeyObjects* la Teoría de Grafos para la detección de nodos claves dentro de una red puede existir una mejora en la eficiencia de detección.

Calcular cuáles son los objetos con mayor importancia relativa en base a diversas medidas de centralidad de Teoría de Grafos es la tarea desarrollada, esto luego de modelar los resultados de experimentos pasados. Es similar a suponer que cada página web de un sitio es una red compuesta de nodos que son los objetos que la componen, siendo los desplazamientos entre estos al momento de navegar las aristas que terminan definiendo el Grafo.

Las medidas de centralidad, como el *Eigenvector*, tienen la facultad de asignar una importancia relativa a un nodo (objeto) en función de los objetos con los que se relaciona, en el fondo la probabilidad de que un objeto sea *clave* depende con cuántos otros objetos claves se relacione.

Los resultados fueron mejores de lo esperado, gracias a que se pudo mejorar la eficiencia previa que habían logrado alcanzar trabajos como los de Gustavo Martínez o Larry Gonzalez. Esto básicamente explicado por el aporte que genera el uso de las medidas de centralidad, cuyo aporte ya ha sido más que validado en campos como la ciencia o la construcción. A pesar de que el webmaster cambió su opinión en la definición de objetos claves en dos de estos, siendo un detalle no menor en los resultados finales de este proyecto.

Al comparar este trabajo con lo realizado por Martínez, no se consideran como *WKO* los objetos 2 y 6 que el webmaster rechazó en otro trabajo, estos dos objetos son 2 de los cinco objetos que tienen mayores tiempos promedios de visualización. Este factor clave demuestra que si bien el *Eigenvector* está relacionado con el tiempo, pueden diferir en la significancia entre algunos nodos, en especial los más relevantes.

Si bien para consolidar la información de los experimentos se definieron dos métodos alternativos, cabe destacar que el restringir la información estableciendo un número mínimo de corte para la selección de las aristas que definen el Grafo final, no aporta una mejora y

si bien se pierde información al no considerar algunas, el cambio que se produjo no es tan relevante si se compara la opinión del webmaster en ambos casos.

Se podría sentar un precedente con este trabajo dado que futuros trabajos de desarrollo de diseño de sitios webs o estudios de optimización de páginas web podrían basarse en estas herramientas que procesan información objetiva y que concluye con una muy buena precisión cuáles son realmente los objetos que más llaman la atención a los usuarios.

Lo novedoso es que hasta ahora no se ha mezclado el análisis de Teoría de Grafos con objetos en páginas web estáticas para detectar los puntos principales o con mayor importancia relativa. Este trabajo de título tiene una particular relevancia si se llega a concluir un verdadero aporte a la metodología mencionada.

Claramente este trabajo es muy básico como para entregar una herramienta que se empiece a trabajar para la mejora de futuros sitios webs, es por esto que para llevar a cabo y desarrollar un modelo que entregue un resultado final se requiere mejorar una serie de puntos que automaticen el proceso y estandaricen el trabajo para cualquier sitio web. Para ello se incluye una serie de mejoras como el mapeo del sitio, la estandarización de los objetos dentro de un sitio, la opinión de un webmaster, entre otros. Mejorando estos pequeños detalles se podría presentar un producto que trabaje para la optimización de sitios web.

5.1 Desarrollos Futuros

Para nuevos trabajos que se basen en esta memoria, se plantea realizar un análisis con páginas web dinámicas para verificar si el comportamiento en páginas estáticas, tal como fue demostrado acá, tiene resultados similares. Esta y las siguientes recomendaciones se basan en lo aprendido durante esta investigación.

5.1.1 Sitio Web y Experimentos

Lo importante es que el sitio sea un sitio web con objetos que no se repitan mucho en las páginas, dado que esto afecta el comportamiento del usuario al navegar. Para que realmente pueda navegar se recomienda que los experimentos sean en sitios web reales que no afecten el modo de navegar, permitiendo hacer scroll, utilizar el mouse y hacer click en los objetos que el sujeto determine llamativos. De acá podrían salir nuevos patrones no considerados hasta ahora.

El sitio web debe ser escogido en función de varios factores que afectan los tiempos totales

de la investigación como la calidad de los resultados. Es por esto que establecer plataformas que no tengan muchas páginas, ni muchos objetos es fundamental para un mejor desarrollo, dado que el mapeo de los objetos es lento y definir que y cuáles son los objetos de cada página toma tiempo y afecta de sobremanera los resultados finales.

Tal como se mencionó dentro del alcance de esta memoria, incluir un análisis en el cual se incluya el tiempo al momento de realizar los grafos, que representan la navegación de los sujetos en el experimento, es un detalle interesante de analizar puesto que se muestran en diversos trabajos Grafos que evolucionan con el tiempo, estos podrían explicar la importancia relativa de objeto dentro de sitios webs dinámicos. Esto podría ser una muy buena opción para continuar con este trabajo de memoria.

Es relevante considerar que al momento de diseñar futuros experimentos se tome en consideración la cantidad de información que se pierde luego al momento de los filtros, esta pérdida de datos podría deberse a malos procedimientos en la toma de la muestra, lo que afecta sensiblemente a los resultados finales.

5.1.2 Webmaster

Uno de los pilares fundamentales que sostienen los resultados de la Teoría de Grafos en campos como la Salud, la Ciencia o la Ingeniería, es que hasta ahora ha demostrado tener buenos resultados trabajando información objetiva. Lo que permite que las conclusiones que se toman en base a estos desarrollos sean basadas en hechos y no en supuestos u opiniones, es por ello que se plantea que para próximos trabajos no se compare todo el trabajo desarrollado y la eficiencia de la metodología con la opinión de una persona, en este caso el web master.

Puede que si el trabajo está orientado a un cliente en particular, este pueda dar su opinión en base a lo que el considera que es lo que le gustaría que fuera lo más llamativo de su plataforma, pero no que finalmente lo que el determine sea la eficiencia de la metodología. Menos aún si la opinión del webmaster varía en función del tiempo, esto hace que el análisis comparativo entre diversos trabajos pierda sentido.

5.1.3 Objetos

Definir que es un objeto de manera estandarizada, permitirá que análisis como este sean estándar y así permitirá realizar comparaciones que generen sinergia entre futuras investigaciones. Además se debe normalizar dentro de un experimento que puede y que no puede ser un objeto, incluyendo en esta definición tamaño, ubicación, forma, colores, estructura, etc. Todos esos factores afectan la importancia que un sujeto le otorgue a un objeto dado.

Además de las características de los objetos, la funcionalidad que estos tengan afectará el tiempo que se le dedique y la importancia relativa que le designe cada sujeto. Dado que tal como se mencionó, un sujeto observa de formas distintas un menú de navegación, un banner, un texto, una imagen, entre otros. Por ejemplo, si una persona observa una imagen que se abre sólo y que tiene como objetivo la publicidad, el tiempo que le dedique variará significativamente si ese sujeto es parte del grupo objetivo de esa publicación. Esto distorsiona los datos en caso de que se pretenda generalizar los resultados obtenidos.

5.1.4 Comportamiento de los Sujetos

Según investigaciones mencionadas en esta memoria, el comportamiento del sujeto al navegar mientras está en un sitio web está afectado por diversos factores, que se deben considerar al momento de seleccionar a los participantes. Dado que está demostrado que la navegación en una página web depende de si el sujeto ha visitado alguna vez el sitio o no. Afecta además el comportamiento del sujeto si tiene una misión particular en cada página versus una visualización libre en la cual puede hacer lo que prefiera, esto distorsiona los tiempos y la frecuencia de observación, por lo que se recomienda tener una particular atención a este punto.

Considerar la fatiga que podría generar en un sujeto una página web en particular es fundamental, experimentar a futuro en sitios web que no tengan pocos objetos muy repetidos y muy pocos objetos una o dos veces es necesario para poder establecer resultados concluyentes que no estén influenciados por la cantidad de veces que un sujeto visualiza un objeto. Esto podría mejorar aún la explicación que pueden generar las medidas de centralidad en un análisis como el planteado durante este trabajo de memoria.

Detectar patrones en función de los resultados, variará dependiendo del tipo de sujeto que fue considerado durante los experimentos, ya que no se comportan de la misma forma abuelos y niños, ni empresarios jóvenes con universitarios, etc. Darle importancia a esto al momento de determinar quiénes formarán parte del experimento es recomendable para futuros trabajos.

Bibliografía

- [1] N. Company, “nielsen ratings.” http://www.nielsen-online.com/pr/pr_030220.pdf. Visto en 13/11/2013.
- [2] M. M. Group, “estadísticas de uso de internet y sitios web.” <http://www.internetworldstats.com/stats.htm>, 2013. Visto en 10/11/2013.
- [3] J. Velásquez, R. Weber, H. Yasuda, and T. Aoki, “a methodology to find web site keywords,” *IEEE Int. Conf. On e-Technology, e-Commerce and e-Service*, 2004.
- [4] L. E. Dujovne and J. D. Velásquez, “design and implementation of a methodology for identifying website keyobjects,” in *Proceedings of the 13th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pp. 301–308, KES, 2009.
- [5] L. González, “mejoramiento de una metodología para la identificación de website keyobjects mediante la aplicación de tecnologías eye-tracking y algoritmos de web mining,” *Universidad de Chile*, 2011.
- [6] L. Gonzalez and J. Velásquez, “una aplicación de herramientas de eye-tracking para analizar preferencias de contenido de los usuarios de sitios web,” *Revista Ingeniería de Sistemas*, 2012.
- [7] C. Aracena, “estudio de la relación entre neurodatos, dilatación pupilar y emocionalidad basado en técnicas de minería de datos,” *Universidad de Chile*, 2014.
- [8] P. S. Leo Spizzirri, “justification and application of eigenvector centrality,” *Algebra in Geography: Eigenvectors of Networks*, pp. 37 – 43, 2001.
- [9] G. S. Canright and K. Engo-Monsen, “some relevant aspects of network analysis and graph theory,” in *Handbook of Network and System Administration*, pp. 361 – 417, Elsevier B.V., 2007.
- [10] G. Martínez, “mejoramiento de una metodología para la identificación de website keyobjects mediante la aplicación de tecnologías eye-tracking, análisis de dilatación pupilar y algoritmos de web mining,” *Universidad de Chile*, 2014.

- [11] P. Bonacich, “some unique properties of eigenvector centrality,” *Elsevier - Social Neetworks*, 2007.
- [12] L. C. Freeman, “centrality in social networks conceptual clarification,” *Elsevier Sequoia S.A. - Social Networs*, 1978.
- [13] J. D. Velásquez, “Combining eye-tracking technologies with web usage mining for identifying website keyobjects,” *Engineering Applicationes of Artificial Intelligence.*, v2013.
- [14] J. Colmenero, A. Catena, and L. J. Fuentes, “atención visual: Una revisión sobre las redes atencionales del cerebro,” *Anales de Psicología*, pp. 45–67, 2001.
- [15] P. Loyola, “characterizing web user visual gaze patterns: a graph theory inspired approach,” *Universidad de Chile*, 2014.
- [16] N. Matsuda and H. Takeuchi, “joint analysis of static and dynamic importance in the eye-tracking records of web page readers,” *Journal of Eye Movement Research*, 2011.
- [17] F. Bonomo, “introducción a la teoría de grafos.” http://www-2.dc.uba.ar/personal/fbonomo/grafos/curso_grafos_handout080909.pdf. Visto en 09/05/2014.
- [18] L. A. Adamic and N. Glance, “the political blogosphere and the 2004 u.s. election: Divided they blog,” *Proceedings of the 3rd international workshop on Link discovery*, 2004.
- [19] D. Vay, *El Ojo: Anatomía y Fisiología Humana*. Paidotribo, 2004.
- [20] R. Ferreruela, “instituto ilo oftalmología - la visión y el ojo.” <http://redined.mecd.gob.es/xmlui/bitstream/handle/11162/44800/01520123000294.pdf?sequence=1>. Visto en 01/12/2013.
- [21] O. de Bruijn and R. Spence, “rapid serial visual presentation: a space-time trade-off in information presentation,” *Bibliometrics*, pp. 189–192, 2002.
- [22] M. C. Potter, B. Wyble, C. E. Hagmann, and E. S. McCourt, “detecting meaning in rsvp at 13 ms per picture,” *Attention, Perception y Psychophysics*, 2014.
- [23] Y. Hassan Montero and H. Solana, “no solo usabilidad journal.” <http://www.nosolousabilidad.com/articulos/eye-tracking.htm#jacobk>. Visto en 15/12/2013.
- [24] I. Zaballa, “el Álgebra de los valores propios.” http://www.ehu.es/izaballa/Ana_Matr/Apuntes/lec9.pdf. Visto en 27/04/2014.

- [25] J. E. Pezoa, “apuntes de teoría de grafos.” <http://cursos.die.udec.cl/~jpezoa/papers/ApunteTeoriaGrafos.pdf>. Visto en 09/05/2014.
- [26] J. Nieminen, “on the centrality graph,” *Scandinavian Journal of Psychology*, pp. 15:322–336, 1974.
- [27] A. Bavelas, “a mathematical model for group structures,” *Human Organization*, 1948.
- [28] M. E. Shaw, “group structure and behavior of individuals in small groups,” *Journal of Psychology*, pp. 38:139–149, 1954.
- [29] G. Sabidussi, “the centrality index of a graph,” *Psychometrika*, pp. 31:581–603, 2002.
- [30] J. Geneson, “an algorithmists toolkit - lecture 1.” <http://redined.mecd.gob.es/xmlui/bitstream/handle/11162/44800/01520123000294.pdf?sequence=1>. Visto en 07/03/2014.
- [31] H. Prendinger and M. Ishizuka, “eye movements as indices for the utility of life-like interfaces agents: a pilot study,” *Interacting with Computers*, pp. 281–292, 2007.

Apéndices

A Tablas

A .1 Tiempos Promedio por Objetos

Objeto	Promedio	Objeto	Promedio	Objeto	Promedio	Objeto	Promedio
1	7407	86	203	208	49	276	67
3	342	87	216	209	15	277	197
4	12579	88	2295	211	616	278	662
5	129611	89	553	212	749	279	8
6	15951	94	975	213	61	280	260
11	276	95	6	214	935	281	15
12	149	96	689	215	6608	282	29
14	14	103	81	216	154	294	8
15	8	104	304	218	657	295	1533
22	514	105	4	219	52	296	447
23	519	121	407	221	88	304	38
35	767	122	332	225	2	306	4702
39	399	123	177	229	4	307	2692
51	3566	124	116	238	22	308	1181
52	1141	125	185	242	5	309	2471
54	128	126	160	243	2347	311	471
57	1	127	120	244	5120	312	651
59	9872	128	23	245	7288	315	53
60	4383	129	15	246	16	341	406
61	248	130	9	247	84	342	16
70	16	134	1857	264	253	343	223
71	9	138	867	265	2760	347	377
73	1415	195	5035	266	749	349	107
74	1604	197	122	267	5014	356	1701
75	1812	201	8	268	54	357	1434
76	103	205	173	269	1954	358	1
84	3044	206	74	271	1607	359	546
85	6	207	37	272	1		

Figura 5.1: Tiempo Promedio por Objeto
Fuente: Elaboración Propia

A .2 Tiempos Promedio por Paginas

Id_pagina	Total página
1	16609
2	17312
3	5580
4	8957
5	12236
6	3794
7	10929
8	3132
9	3089
10	5658
11	6069
12	10468
13	18764
14	1708
15	3138
16	8509
17	3942
18	442
19	8894
20	10394
21	5958
22	8394
23	9837
24	6293
25	10136
26	9909
27	8183
28	13387
29	9860
30	6526
31	7531
32	12846

Figura 5.2: Tiempo Promedio por Página
Fuente: Elaboración Propia

A .3 *WKO* detectados por Martínez

Object Number	Website Object	Name	Is Keyobject?
1	Object 5	header picture	Yes
2	Object 356	picture detalle profe 1	Yes
3	Object 6	navigation menu	Yes
4	Object 195	paragraph detalle noticia 1	Yes
5	Object 4	logo mba	No
6	Object 267	paragraph quienes 1	Yes
7	Object 306	paragraph metodologia 1	Yes
8	Object 244	image plataforma 1	Yes
9	Object 1	ingenieria industrial logo	No
10	Object 265	paragraph corporacion 1	Yes
11	Object 307	paragraph metodologia 2	Yes
12	Object 215	paragraph propuesta 1	Yes
13	Object 243	paragraph plataforma 1	Yes
14	Object 277	title contenido 1	Yes
15	Object 245	paragraph perfil egresados 1	Yes
16	Object 130	title elegirnos 10	No
17	Object 3	search form	No
18	Object 59	titulo cuerpo	Yes
19	Object 23	main post text 1	Yes
20	Object 22	main post picture	Yes

Tabla 5.1: Lista de Website Keyobject
Fuente: Gustavo Martinez

Id Página	Cantidad Promedio Objetos Vistos
Página 1	5.611111111
Página 2	6.526315789
Página 3	3.894736842
Página 4	5.052631579
Página 5	5.578947368
Página 6	4.052631579
Página 7	6.000000000
Página 8	4.210526316
Página 9	4.157894737
Página 10	5.263157895
Página 11	7.473684211
Página 12	5.631578947
Página 13	6.105263158
Página 14	4.052631579
Página 15	4.842105263
Página 16	4.526315789
Página 17	5.894736842
Página 18	1.176470588
Página 19	4.263157895
Página 20	5.526315789
Página 21	5.105263158
Página 22	5.421052632
Página 23	5.631578947
Página 24	5.526315789
Página 25	5.789473684
Página 26	5.947368421
Página 27	5.052631579
Página 28	5.263157895
Página 29	4.000000000
Página 30	3.263157895
Página 31	4.789473684
Página 32	4.052631579
Total	5.001652893

Tabla 5.2: Promedio Objetos Vistos por Página
Fuente: Elaboración Propia

Id Página	Primer Objeto	Tiempo (miliseg)	Segundo Objeto	Tiempo (miliseg)
1	6	158	5	2960
2	6	334	5	228
3	59	258	6	210
4	5	392	59	1062
5	5	404	6	2054
6	6	1166	87	26
7	5	606	6	22
8	5	1382	59	200
9	6	298	51	190
10	59	668	51	134
11	5	336	59	364
12	6	18	59	218
13	195	186	59	780
14	5	300	59	266
15	5	608	211	42
16	6	760	215	708
17	51	98	359	184
18	238	134	359	618
19	5	402	359	646
20	6	204	5	328
21	6	14	5	642
22	6	44	5	112
23	5	1036	59	268
24	59	262	277	694
25	5	278	59	286
26	6	116	59	62
27	5	352	59	344
28	311	230	309	14
29	5	860	59	374
30	5	2990	59	1130
31	356	200	59	2028
32	359	300	59	442

Tabla 5.3: Primer y Segundo Objeto por Página - sujeto 2
Fuente: Elaboración Propia

B Figuras

A continuación se muestran las figuras que representan a los 5 objetos claves principales que arrojó este trabajo de título. Se muestra además la página en la que estaba y la ubicación exacta.



Figura 5.3: Top Five Key Objects
Fuente: Elaboración Propia

Suscríbete RSS 

INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE

buscar... 



MBA

MAGÍSTER EN GESTIÓN Y DIRECCIÓN DE EMPRESAS




[POR QUÉ ELEGIRNOS](#)
[PROPUESTA ACADÉMICA](#)
[PROFESORES](#)
[ALUMNOS](#)
[ADMISIÓN](#)
[MAGISTERES](#)
[EGRESADOS](#)
[NOTICIAS](#)
[CONTACTO](#)

ADMISIÓN

Modalidad

Versión Executive Profesionales
Se imparte en la modalidad Part Time y está diseñado para profesionales de diversas formaciones universitarias. Tiene una extensión 18 meses, distribuidos en 8 bimestres y con aproximadamente 750 horas de clases.

Versión Executive ICHC
También se imparte en la modalidad Part Time y está dirigido a Ingenieros Civiles Industriales y Comerciales de universidades chilenas, que puedan convalidar un conjunto de cursos de la fase de Homogeneización del Programa. Tiene una duración de 14 meses y aproximadamente 620 horas de clases.

En ambas versiones, su formato es vespertino de 18:30 a 21:30 horas en días hábiles, tres a cuatro veces a la semana. Además tiene actividades un sábado al mes, o dos medios sábados al mes, según su programación durante cuatro bimestres, dirigidas principalmente a talleres del área de Habilidades Directivas.

¿Quiénes pueden postular?

Todos aquellos profesionales con más de 3 años de experiencia relevante de jornada completa que hayan obtenido, en Chile o en el extranjero, una licenciatura o un título profesional cuyo nivel, contenido y duración de estudios correspondan a una formación equivalente a la del grado de Licenciado de la Universidad de Chile.

¿Existe prueba de admisión?

Se exige al postulante rendir una prueba de admisión.

- Test de aptitudes y conocimientos (matemáticas, G... su reemplazo se acepta GMAT con una antigüedad...
- ... economía), o en ... de postular

TESTIMONIOS



Teodoro Wigodski, MBA
Loyola College Maryland EEUU, Magister en Derecho de Empresas PUC Chile, Magister en Economía Aplicada Georgetown University EEUU.

"El MBA en Ingeniería Industrial de la Facultad de Ingeniería de la U. de Chile, ha liderado las innovaciones en el estado del arte del conocimiento y el desarrollo de las habilidades directivas." [=>continuar=>]



Maritza Espinal, Gerente General Catálogo Virtual de Bolivia, egresada MBA.

"Gracias al MBA, logré darle un marco a aquello que me interesaba y quería estudiar. Entonces así pude complementar lo teórico con lo práctico, tras lo cual creo haber alcanzado un resultado eficaz." [=>continuar=>]

60

Pag. 2

Figura 5.4: Top Five Key Objects
Fuente: Elaboración Propia

Suscríbete RSS

INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE

buscar...

MBA
MAGÍSTER EN GESTIÓN Y
DIRECCIÓN DE EMPRESAS

[POR QUÉ ELEGIRNOS](#)
[PROPUESTA ACADÉMICA](#)
[PROFESORES](#)
[ALUMNOS](#)
[ADMISIÓN](#)
[MAGISTERES](#)
[EGRESADOS](#)
[NOTICIAS](#)
[CONTACTO](#)

PROPUESTA ACADÉMICA

En el mundo competitivo y global en que se desenvuelve la empresa actualmente, resulta cada vez más importante contar con líderes innovadores, que tengan una visión transcultural de los mercados, capaces de adaptarse a los cambios y prepararse para conducir de forma motivadora a toda la organización.

Formar líderes que desde la dirección o creación empresarial influyan en el desarrollo socio económico con los conocimientos, valores y creatividad que exige el escenario empresarial de hoy -competitivo, globalizado e innovador- constituye todo un desafío. En esta perspectiva, Ingeniería Industrial de la Universidad de Chile creó este programa de Magíster.

↓
folleto

[recomendar](#)
[compartir](#)
[twittear](#)
[compartir](#)
[más información](#)

TESTIMONIOS

Mario Conca Rosende, Gerente General Chilevisión, visión del Experto.

"El mundo que enfrentan las empresas de hoy es altamente competitivo y crecientemente globalizado. Ello impone un conjunto de requerimientos sobre sus ejecutivos."
[=continuar=]

David Plaza E., Subgerente de Riesgo Financiero, Banco Penta, egresado MBA.

"El incorporar conocimientos sobre cómo incentivar y dirigir a un conjunto de personas es el punto fuerte que me dejó este Programa Académico."
[=continuar=]

Pag. 16

Figura 5.5: Top Five Key Objects
Fuente: Elaboración Propia

245

INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE

Suscríbete RSS

buscar...

MBA
MAGÍSTER EN GESTIÓN Y DIRECCIÓN DE EMPRESAS

POR QUÉ ELEGIRNOS PROPUESTA ACADÉMICA PROFESORES ALUMNOS ADMISIÓN MAGISTERES EGRESADOS NOTICIAS CONTACTO

PERFIL EGRESADOS

El alto nivel de exigencia del proceso educativo, la calidad del cuerpo académico, la rigurosa selección y la diversidad de los estudiantes imprimen un sello a los egresados, que los caracteriza por:

- Su sólida formación en las distintas áreas de gestión
- Capacidad de reflexión rigurosa
- Capacidad de desempeñarse dentro de distintas culturas empresariales
- Adaptabilidad a cualquier tipo de organización
- **Versatilidad para desempeñar distintas funciones de responsabilidad en la empresa**
- Hábito de trabajo en equipo
- Potencialidad para ser agentes de cambio e innovación
- Capacidad de conducir y motivar personas, grupos y organizaciones.

Los siguientes gráficos nos muestran el perfil de nuestros egresados, el origen de los datos corresponde a la actualización de datos realizada durante el año 2011, la cual alcanza a un 90%.

Evolución de Cargos

Porcentaje de Gerentes y Directores

Año	Blue Line (%)	Red Line (%)
2008	28%	27%
2009	32%	26%
2010	34%	35%
2011	53%	42%

TESTIMONIOS

Carlos Vignolo, Ph.D.(c) en Economía Política, Universidad de Sussex, Inglaterra.
"En el transcurso del tiempo hemos logrado consolidar un grupo de investigadores y profesores de alto nivel en el ámbito de las habilidades directivas" [-continuar-]

Ramón Serrano, Gerente General, Specialty S.A. Panamá, egresado MBA.
"Las expectativas que tenía fueron plenamente satisfechas: es un programa del más alto nivel, con un enfoque integral, actualizado y exigente, que invita a los alumnos a explorar y expandir sus competencias personales." [-continuar-]

David Plaza E., Subgerente de Riesgo Financiero, Banco Penta, egresado MBA.
"El incorporar conocimientos sobre cómo incentivar y dirigir a un conjunto de personas es el punto fuerte que me dejó este Programa Académico." [-continuar-]

Pag. 20

Figura 5.6: Top Five Key Objects
Fuente: Elaboración Propia

Suscríbete RSS 





MBA
MAGÍSTER EN GESTIÓN Y
DIRECCIÓN DE EMPRESAS




[POR QUÉ ELEGIRNOS](#)
[PROPUESTA ACADÉMICA](#)
[PROFESORES](#)
[ALUMNOS](#)
[ADMISIÓN](#)
[MAGISTERES](#)
[EGRESADOS](#)
[NOTICIAS](#)
[CONTACTO](#)

QUIÉNES PARTICIPAN

El Programa está diseñado para profesionales de diversas formaciones universitarias, provenientes de Chile y el extranjero, que aspiren a desarrollar sus potencialidades personales de liderazgo, su creatividad y su capacidad de análisis. Los alumnos se integran a grupos de trabajo conformados por profesionales de diferentes nacionalidades, grados de experiencia, formación académica y edades, en un ambiente que propicia la experimentación, el aprendizaje activo y la interacción.

¿Quiénes pueden postular?

Todos aquellos profesionales con más de 3 años de experiencia relevante de jornada completa que hayan obtenido, en Chile o en el extranjero, una licenciatura o un título profesional cuyo nivel, contenido y duración de estudios correspondan a una formación equivalente a la del grado de Licenciado de la Universidad de Chile.

 recomendar
  compartir
  twittear
  compartir
  más información

TESTIMONIOS



Mario Conca Rosende,
Gerente General Chilevisión,
visión del Experto.

"El mundo que enfrentan las empresas de hoy es altamente competitivo y crecientemente globalizado. Ello impone un conjunto de requerimientos sobre sus ejecutivos."
[=continuar=]



Juan Pablo Gutiérrez,
Gerente de Ventas, Business
Soft Nicaragua, egresado
MBA.

"Siempre había catalogado a la Universidad de Chile entre las mejores dentro de los rankings. Sin embargo, cuando empecé a cursar este Programa me di cuenta de que, entre todas, sin duda era la

Pag. 22

Figura 5.7: Top Five Key Objects
Fuente: Elaboración Propia

C Glosario Teoría de Grafos

Adyacencia: En un grafo, los vertices son adyacentes si están unidos mediante una arista.

Árbol: Un árbol es un grafo conexo simple acíclico. Algunas veces, un vértice del árbol es distinguido llamándolo raíz. Los árboles se usan frecuentemente como estructuras de datos en ciencias de la computación.

Arista: Una arista o arco es una relación matemática que conecta dos vértices. Una arista dirigida es una arista de un digrafo y tiene una dirección asociada consigo, esto es, posee un vértice inicial y un vértice final. Una arista no dirigida es una donde no se distingue un vértice inicial ni uno final.

Bosque: Un bosque es un conjunto de árboles; o de forma equivalente, un bosque es un grafo acíclico.

Bucle: Un bucle o lazo (loop en inglés) en un grafo o digrafo es una arista que conecta al mismo vértice consigo mismo. Un grafo simple no puede tener bucles.

Búsqueda en anchura: La búsqueda en anchura o BFS (Breadth First Search) es un algoritmo que permite recorrer todos los vértices de un árbol de manera ordenada, recorriendo primero los vértices vecinos al inicial, luego los vértices vecinos a los recorridos en el paso anterior y así sucesivamente hasta agotar la gráfica.

Búsqueda en profundidad: La búsqueda en profundidad o DFS (Depth First Search) es un algoritmo que permite recorrer todos los vértices de un árbol de manera ordenada, avanzando sobre cada rama hasta que no haya posibilidad de continuar y luego se retrocede hasta la última bifurcación para seguir por otra rama. Puede usarse para recorrer un grafo cualquiera si se usa un árbol generador del grafo.

Camino: Un camino es una sucesión de vértices tal que de cada uno de sus vértices existe una arista hacia el vértice sucesor. Un camino simple es aquel que no repite vértices en su recorrido. Dos caminos son ajenos o independientes si no tienen ningún vértice en común excepto el primero y el último. La longitud de un camino es el número de aristas que usa dicho camino, contando aristas recorridas varias veces el mismo número de veces que las recorramos. En el ejemplo, (1, 2, 5, 1, 2, 3) es un camino con longitud 5, y (5, 2, 1) es un camino simple de longitud 2..

Camino euleriano: Un camino euleriano en un grafo es un camino que usa cada arista

una y sólo una vez. Si existe tal camino decimos que el grafo es euleriano. Esta definición es dual a la de camino hamiltoniano.

Camino hamiltoniano: Existe un concepto dual al de camino/ciclo Euleriano. Un camino hamiltoniano en un grafo es un camino que visita cada vértice una y sólo una vez.

Ciclo: Un Ciclo (o circuito) es un camino que empieza y acaba en el mismo vértice. Los ciclos de longitud 1 se denominan lazos o bucles. Un ciclo simple es un ciclo que tiene como longitud al menos 3 y en el que el vértice inicial coincide con el vértice final.

Ciclo euleriano: Un ciclo euleriano en un grafo es un ciclo que usa cada arista una y sólo una vez.

Ciclo hamiltoniano: Un ciclo hamiltoniano en un grafo es un ciclo que visita cada vértice una y sólo una vez

Circunferencia: La circunferencia de un grafo es la longitud de su ciclo simple más largo.

Clique: Una clique en un grafo es un conjunto de vértices tal que para todo par de vértices, existe una arista que las conecta. En el ejemplo, los vértices 1, 2 y 5 forman una clique de tamaño 3. En otras palabras, es un subgrafo completo K_n .

Cobertura de vértices: La cobertura de vértices, covering o recubrimiento de vértices de un grafo es un conjunto de vértices cuyos elementos son adyacentes a todos los demás vértices del grafo.

Coloración de grafos: La coloración de grafos es quizá el problema NP-completo más afamado de la teoría de grafos, y consiste en asignarle distintos colores o marcas a los vértices de un grafo, de manera que ningún par de vértices adyacentes compartan el mismo color o marca.

Contracción (de aristas): La contracción es una operación que elimina una arista del grafo al mismo tiempo que fusiona los dos vértices extremos. La contracción es una operación fundamental en la teoría de grafos.

Componente fuertemente conexo: Un componente fuertemente conexo es un grafo tal que para cada par de vértices, existe un camino de uno hacia el otro, y viceversa. Los componentes fuertemente conexos de un grafo dirigido son sus subgrafos máximos fuertemente conexos. Estos subgrafos forman una partición del grafo.

Conjunto independiente: Un conjunto independiente en un grafo es un conjunto de vértices tal que ninguno es adyacente a otro. En el ejemplo, los vértices 1, 3, y 6 forman un conjunto tal y los 3, 5, y 6 forman otro conjunto independiente.

Digrafo: Es un grafo cuyas aristas son dirigidas, es decir, cada arista posee un vértice inicial y uno final.

Distancia: Se denomina distancia entre dos vértices de un grafo al número de vértices mínimo que debe recorrerse para unirlos. La distancia entre dos nodos de un grafo es la longitud del camino más corto

Girth: El girth o cintura de un grafo es la longitud del ciclo simple más corto en el grafo. El girth de un grafo acíclico se define como infinito.

Grado: El grado o valencia de un vértice es el número de aristas incidentes en él. Para un grafo con bucles, éstos son contados por dos. En el ejemplo, los vértices 1 y 3 tienen grado 2; los vértices 2, 4 y 5, grado 3; y el vértice 6, grado 1. En un dígrafo, podemos distinguir el grado saliente (el número de aristas que dejan el vértice) y el grado entrante (el número de aristas que entran en un vértice). El grado de un vértice sería la suma de ambos números.

Grafo: Un grafo es un conjunto de vértice o nodos unidos por aristas o arcos.

Grafo acíclico: Un grafo se dice acíclico si no contiene ningún ciclo simple.

Grafo bipartito: Un grafo bipartito es cualquier grafo cuyos vértices pueden ser divididos en dos conjuntos, tal que no haya aristas entre los vértices del mismo conjunto. Se ve que un grafo es bipartido si no hay ciclos de longitud impar. Un grafo k -partido o grafo k -colorable es un grafo con cuyos vértices se puede hacer una partición en k subconjuntos disjuntos tal que no haya aristas entre vértices del mismo subconjunto. Un grafo 2-partido es lo mismo que un grafo bipartido. Un grafo k -partido se dice semiregular si cada partición tiene un grado uniforme.

Grafo completo: Un grafo completo es un grafo simple en el que cada vértice es adyacente a cualquier otro vértice. El del ejemplo no es completo. El grafo completo en n vértices se denota a menudo por K_n . Tiene $n(n - 1)/2$ aristas (correspondiendo a todas las posibles elecciones de pares de vértices).

Grafo conexo: Si es posible formar un camino desde cualquier vértice a cualquier otro en el grafo, decimos que el grafo es conexo. Si es posible hacer esto incluso tras quitar $k - 1$

vértices, decimos que el grafo es k -conexo. Un grafo es k -conexo si y sólo si contiene k caminos independientes entre cualesquiera dos vértices.

Grafo denso: Un grafo denso es un grafo en el que el número de aristas está cercano al número de máximo de aristas. Lo opuesto, un grafo con solo algunas aristas, es un grafo disperso.

Grafo dirigido: Es un conjunto de vértices V y un conjunto de aristas E tal que para cada arista perteneciente al conjunto de aristas E se asocia con dos vértices en forma ordenada.

Grafo nulo: El grafo nulo es el grafo cuyos conjuntos de aristas y de vértices son vacíos.

Grafo plano: Un grafo plano es uno que es posible dibujar en el plano sin que ningún par de aristas se interseque.

Grafo ponderado: Un grafo ponderado asocia un valor o peso a cada arista en el grafo. El peso de un camino en un grafo con pesos es la suma de los pesos de todas las aristas atravesadas.

Grafo regular: Un grafo regular es un grafo cuyos vértices tienen todos el mismo grado.

Grafo simple: Un grafo simple es un grafo o digrafo que no tiene bucles, y que no es un multigrafo.

Grafo trivial: Un grafo trivial es un grafo vacío con un único vértice.

Grafo universal: Un grafo universal en una clase K de grafos es un grafo en el que puede incluirse como subgrafo todo elemento de K .

Grafo vacío: Un grafo vacío es el grafo cuyo conjunto de aristas es vacío.

Hipergrafo: Un hipergrafo es una generalización de un grafo, cuyas aristas aquí se llaman hiperaristas, y pueden relacionar a cualquier cantidad de vértices, en lugar de sólo un máximo de dos como en el caso particular.

Isomorfismo: Un Isomorfismo de grafos entre dos grafos G y H es una biyección f entre los conjuntos de sus vértices $f : V(G) \rightarrow V(H)$ que preserva la relación de adyacencia. Es decir, cualquier par de vértices u y v de G son adyacentes si y solo si lo son sus imágenes, $f(u)$ y $f(v)$, en H .

Lista de adyacencia: Una lista de adyacencia es una representación de todas las aristas o arcos de un grafo mediante una lista.

Matriz de adyacencia: Una matriz de adyacencia es una matriz de $n \times n$ que permite representar un grafo o digrafo finito, donde cada valor en la posición (i, j) representa el número de aristas desde el vértice i -ésimo al j -ésimo.

Orden: Se llama orden del grafo G a su número de vértices, designado como $|V|$.

Puente: Un puente a es una arista tal que si la quitamos nos quedamos con un grafo con una componente conexa más que el original.

Subárbol: Un subárbol de un grafo G es un subgrafo que es además un árbol.

Subgrafo: Un subgrafo de un grafo G es un grafo cuyo conjunto de vértices es un subconjunto del de G , cuyo conjunto de aristas es un subconjunto del conjunto de las aristas de G , y tal que la aplicación w es la restricción de la aplicación de G .

Subgrafo de expansión: Un subgrafo de expansión de un grafo G es un subgrafo con el mismo conjunto de vértices que G . Un árbol expansión es un subgrafo expansión que es un árbol. Cada grafo tiene un árbol de expansión.

Teoría espectral: La teoría espectral es aquella que estudia las relaciones entre las propiedades de la matriz de adyacencia y las de su grafo.

Torneo: Un torneo es un grafo dirigido completo, simple, no generalizado, no degenerado y sin dígonos.

Vecindad: Dos vértices son vecinos, adyacentes o incidentes si existe una arista entre ellos. En el ejemplo, el vértice 1 tiene dos vecinos: el vértice 2 y el 5. Para un grafo simple, el número de vecinos de un vértice es igual a su grado.

Vértice: Un vértice o nodo es la unidad fundamental de la que están formados los grafos.

Vértice de corte: Un vértice de corte es un vértice tal que si lo quitamos nos quedamos con un grafo con más componentes conexas que el original.