



INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA

“Detección de clusters en la red de pagos”

T E S I S

PARA OBTENER EL GRADO DE MAESTRO EN
SISTEMAS COMPUTACIONALES

P R E S E N T A

ISC. RAMIRO BASTAR GONZÁLEZ

ASESOR

DR. RAJESH ROSHAN BISWAL

CO-ASESOR

DR. EDDY SÁNCHEZ DE LA CRUZ

Misantla, Veracruz

Agosto, 2018

AGRADECIMIENTOS

Después de poco más de dos años y medio logré finalizar mi maestría. Y esto no fue solo por mérito propio, si no con apoyo de las personas a quienes estoy a punto de agradecerles a continuación:

Mamá y Papá: ustedes que hasta la fecha me siguen guiando y protegiendo tal cual si fuera su hijo el menor, muchas gracias por su apoyo en cada momento de mi vida hasta ahora. Sin sus regaños no sería lo que soy ahora. ¡Los amo!

Hermanos Edwin, Wilmer y Eira, ya al fin terminé la tesis, ya no los molestaré con que dejen de hacer ruido porque estoy trabajando en ella. Gracias por respetar mi espacio y espero siempre seguir siendo un buen ejemplo para ustedes.

Abuelos Tita, Papo y Esther. Gracias por sus sabias palabras en cada momento que me veían preocupado, cada bendición, cada regaño, cada dicho, o cada alimento que me compartían, es una bendición a su lado. Los tres desde el kínder siempre han estado pendiente de mi educación y hoy que me título de maestría les agradezco el acompañarme en este largo trayecto de 22 años de estudio.

Abby Echeverria, amor, compañera de estudios de licenciatura, maestría, y hoy, mi prometida. Estamos culminando una etapa más juntos la cual creíamos casi imposible al inicio, la cual, estuvimos casi a punto de caer, pero nunca nos rendimos. Sigamos adelante. Hoy estamos un paso más cerca de nuestro objetivo juntos. Te amo.

Muchas gracias Dra. Biliانا Alexandrova-Kabadjova, hace 5 años cuando vi su nombre en la lista de doctores investigadores para elegir en el programa de verano de investigación científica, nunca pensé que pudiese llegar a terminar haciendo mi tesis de maestría siendo apoyado por usted. Gracias por abrirme las puertas una vez más para colaborar con usted y

muchas gracias por todo su conocimiento y respaldo en este trabajo de tesis, sin su ayuda, nada de esto hubiera sido posible.

Dr Rajesh Roshan Biswal, usted nos ha demostrado muchas gracias por su apoyo en este trabajo de tesis, además de ser nuestro asesor, sabemos que en usted encontramos un buen amigo con quien aprendimos a separar el trabajo y el estudio, los momentos de convivir y de estudiar, los aspectos emocionales y los académicos, gracias por su conocimiento, apoyo y confianza, además, por levantar la mano el día que no teníamos asesor.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) ya que me becó durante los dos años que dura el plan de estudios del posgrado. Muchas gracias.

Gracias familia, amigos, maestros y alumnos que confiaron en mí y hoy estoy aquí.

¡Gracias a Dios sobre todas las cosas!

Ramiro Bastar González

RESUMEN

El presente trabajo muestra la clusterización de los 38 bancos participantes de manera inicial en los inicios del SPEI en el periodo agosto 2004 a diciembre 2005, utilizando el algoritmo k-means, así como el pre procesamiento, procesamiento y análisis de los datos. Se utilizaron indicadores de red para poder hacer el agrupamiento de los bancos, tales como son *degree*, *degree in*, *degree out*, montos in, montos out, *closeness*, *betweenness* y *affinity*. De manera inicial se realizó una limpieza de los datos, la cual implicó quitar ruido de la base de datos, quitando variables innecesarias, así como bancos inexistentes en ese periodo de tiempo. Una vez seleccionados los indicadores, y solo con los bancos que se trabajarán, se procedió a seleccionar el algoritmo y su configuración el cual fue k-means, y con la técnica de WCSS se seleccionó el mejor valor para **k** el cual fue 5. Antes de la ejecución de k-means, se normalizaron los datos, para que no exista tendencia hacia ninguna variable en el agrupamiento. La ejecución del algoritmo, arrojó 5 grupos, los cuales son volátiles a través del tiempo, pero que también muestran características muy similares tales como bancos que nunca cambian de grupo, o bien que persisten más del 70% del periodo de estudio en uno mismo, son parte de los resultados que se presentan en este trabajo de investigación, el cual además complementa con gráficas y de manera estadística la ejecución del algoritmo de k-means.

ÍNDICE GENERAL

RESUMEN.....	1
ÍNDICE GENERAL.....	4
ÍNDICE DE FIGURAS.....	8
INDICE DE TABLAS.....	11
INTRODUCCIÓN.....	13
CAPÍTULO 1:.....	15
GENERALIDADES.....	15
1.1 PLANTEAMIENTO DEL PROBLEMA.....	16
1.2 PROPUESTA DE SOLUCIÓN.....	18
1.3 OBJETIVOS.....	19
1.4 ALCANCES Y LIMITACIONES.....	20
1.5 ESTRUCTURA DE LA TESIS.....	21
CAPÍTULO 2.....	22
MARCO TEÓRICO.....	22
2.1 BANCO DE MÉXICO.....	23
2.2 SISTEMAS DE PAGO.....	24
2.3 SPEI.....	26
2.4 MINERÍA DE DATOS.....	27
2.5 APRENDIZAJE SUPERVISADO.....	28
2.6 APRENDIZAJE NO SUPERVISADO.....	28
2.7 CLUSTERING.....	29
2.8 ALGORITMOS DE CLUSTERING.....	30
2.9 MATLAB.....	34
2.10 PYTHON.....	35

2.11 INDICADORES DE RED.....	38
CAPÍTULO 3:.....	41
ESTADO DEL ARTE.....	41
3.1 CLUSTERING AUSTRIAN BANKS' BUSINESS MODELS AND PEER GROUPS IN THE EUROPEAN BANKING SECTOR.....	42
3.2 UKRAINIAN BANKS' BUSINESS MODELS CLUSTERING: APPLICATION OF KOHONEN NEURAL NETWORKS.....	42
3.3 STRATEGIC GROUPS IN AUSTRIAN BANKING	43
3.4 CLUSTERING OF RUSSIAN BANKS: BUSINESS MODELS OF INTERACTION OF THE BANKING SECTOR AND THE REAL ECONOMY	43
3.5 THE CLUSTER ANALYSIS OF THE BANKING SECTOR IN EUROPE .	44
3.6 AGRUPACIÓN DE INSTITUCIONES BANCARIAS A PARTIR DEL ANÁLISIS DE CLUSTER: UNA APLICACIÓN AL CASO DE CHILE	44
CAPÍTULO 4.....	46
EXPERIMENTACIÓN	46
4.1 INTRODUCCIÓN.....	47
4.2. DATOS.....	47
4.3 PREPROCESAMIENTO	49
4.4 CÁLCULO DE K.....	60
CAPÍTULO 5.....	64
ANÁLISIS DE RESULTADOS.....	64
5.1 INTRODUCCIÓN	65
5.2 Análisis mes de agosto 2004	65
5.3 Análisis mes de septiembre 2004	67
5.4 Análisis mes de octubre 2004.....	69
5.5 Análisis mes de noviembre 2004	70

5.6 Análisis mes de diciembre 2004	72
5.7 Análisis mes de enero 2005.....	73
5.8 Análisis mes de febrero 2005.....	75
5.9 Análisis mes de marzo 2005	77
5.10 Análisis mes de abril 2005	79
5.11 Análisis mes de mayo 2005	81
5.12 Análisis mes de junio 2005	83
5.13 Análisis mes de julio 2005.....	85
5.14 Análisis mes de agosto 2005	87
5.15 Análisis mes de septiembre 2005	89
5.16 Análisis mes de octubre 2005	91
5.17 Análisis mes de noviembre 2005	93
5.18 Análisis mes de diciembre 2005	95
5.19 Reacomodo de los clúster.....	97
5.20 Análisis de nuevos clusters.....	97
5.21 CONCLUSIONES Y TRABAJO FUTURO.....	102
ANEXO 1	103
ÍNDICE DE BANCOS TRABAJADOS EN LA TESIS CORRESPONDIENTES A LOS ORIGINALES DADOS POR EL BANCO DE MÉXICO	103
ANEXO 2	104
GRAFICAS OBTENIDAS AL EJECUTAR WCSS PARA LA BÚSQUEDA DE EL VALOR “K”.....	104
ANEXO 3	112
SALIDA OBTENIDA EN LA EJECUCION DE KMEANS	112
ANEXO 4	116

EVOLUCIÓN DE LOS BANCOS A TRAVÉS DEL TIEMPO EN LOS GRUPOS
“A” “B” “C” “D” “E” 116

ANEXO 5 136

GRAFICA MULTIDIMENSIONAL POR MES 136

Bibliografía..... 153

ÍNDICE DE FIGURAS

Figura 2.1 Ejemplo de ejecución de Kmeans.....	28
Figura 2.2 Ejemplo de ejecución de Clustering Jerárquico	31
Figura 2.3 Software Matlab.....	32
Figura 4.1 Archivos entregados por el banco	45
Figura 4.2 Contenido de los archivos entregados por el banco	46
Figura 4.3 Código para generar estructura	47
Figura 4.4 Código para generar estructura con 38 bancos.....	48
Figura 4.5 Matriz de adyacencia W_b	49
Figura 4.6 Matriz de adyacencia W_l	49
Figura 4.7 Matriz de adyacencia W	50
Figura 4.8 Degree representativo a la imagen 4.8.....	50
Figura 4.9 Código para seleccionar solo indicadores	51
Figura 4.10 Código para generar base de datos diaria.....	51
Figura 4.11 Imagen de muestra de archivos diarios de Matlab	52
Figura 4.12 Código usado para exportar a csv	53
Figura 4.13 Ejemplo de salida en carpetas.....	53
Figura 4.14 Ejemplo de contenido en carpetas.....	54
Figura 4.15 Promediado mensual en Python.....	55
Figura 4.16 Contenido del archivo promediado	55
Figura 4.17 Código para normalizar los valores.....	56

Figura 4.18 Valores normalizados entre 0 y 1 mismos que la figura 4.16	56
Figura 4.19 Meses escalados	57
Figura 4.20 WCSS septiembre 2004	59
Figura 4.21 WCSS junio 2005.....	59
Figura 4.22 Código para calcular el valor de k en cada mes	60
Figura 4.23 Ejecución del algoritmo de Kmeans.....	61
Figura 5.1 Cantidad de bancos por clúster en agosto 2004.....	63
Figura 5.2 Cantidad de bancos por clúster en septiembre 2004.....	65
Figura 5.3 Cantidad de bancos por clúster en octubre 2004	67
Figura 5.4 Cantidad de bancos por clúster en noviembre 2004.....	69
Figura 5.5 Cantidad de bancos por clúster en diciembre 2005	70
Figura 5.6 Cantidad de bancos por clúster en enero 2005	72
Figura 5.7 Cantidad de bancos por clúster en febrero 2005	73
Figura 5.8 Cantidad de bancos por clúster en marzo 2005	75
Figura 5.9 Cantidad de bancos por clúster en abril 2005.....	77
Figura 5.10 Cantidad de bancos por clúster en mayo 2005.....	79
Figura 5.11 Cantidad de bancos por clúster en junio 2005.....	81
Figura 5.12 Cantidad de bancos por clúster en julio 2005	83
Figura 5.13 Cantidad de bancos por clúster en agosto 2005.....	85
Figura 5.14 Cantidad de bancos por clúster en septiembre 2005.....	87
Figura 5.15 Cantidad de bancos por clúster en octubre 2005	85

Figura 5.16 Cantidad de bancos por clúster en noviembre 2005.....	91
Figura 5.17 Cantidad de bancos por clúster en diciembre 2005.....	93
Figura 5.18 Frecuencia de participación de los bancos en el clúster A ...	96
Figura 5.19 Frecuencia de participación de los bancos en el clúster B ...	97
Figura 5.20 Frecuencia de participación de los bancos en el clúster C ...	97
Figura 5.21 Frecuencia de participación de los bancos en el clúster D ...	98
Figura 5.22 Frecuencia de participación de los bancos en el clúster E ...	99

INDICE DE TABLAS

Tabla 1 Participantes por clúster en agosto 2004	65
Tabla 2 Análisis centroide agosto 2004	66
Tabla 3 Análisis medias y desviaciones estándar agosto 2004	66
Tabla 4 Participantes por clúster en septiembre 2004	67
Tabla 5 Análisis centroide septiembre 2004	68
Tabla 6 Análisis media y desviaciones estándar septiembre 2004	68
Tabla 7 Participantes por clúster en octubre 2004	69
Tabla 8 Análisis centroide octubre 2004	70
Tabla 9 Análisis medias y desviaciones estándar octubre 2004	70
Tabla 10 Participantes por clúster en noviembre 2004	71
Tabla 11 Análisis centroide noviembre 2004	71
Tabla 12 Análisis medias y desviaciones estándar noviembre 2004	72
Tabla 13 Participantes por clúster en diciembre 2005	73
Tabla 14 Análisis centroide diciembre 2005	73
Tabla 15 Análisis medias y desviaciones estándar diciembre 2005	73
Tabla 16 Participantes por clúster en enero 2005.....	74
Tabla 17 Análisis centroide enero 2005	74
Tabla 18 Análisis medias y desviaciones estándar enero 2005.....	75
Tabla 19 Participantes por clúster en febrero 2005	76
Tabla 20 Análisis centroide febrero 2005.....	76
Tabla 21 Análisis de medias y desviaciones estándar febrero 2005.....	77
Tabla 22 Participantes por clúster en marzo 2005.....	78
Tabla 23 Análisis centroide marzo 2005	78
Tabla 24 Análisis de medias y desviaciones estándar marzo 2005	79
Tabla 25 Participantes por clúster en abril 2005.....	80
Tabla 26 Análisis centroide abril 2005	80
Tabla 27 Análisis de medias y desviaciones estándar abril 2005	81
Tabla 28 Participantes por clúster en mayo 2005	82
Tabla 29 Análisis centroide mayo 2005	82
Tabla 30 Análisis de medias y desviaciones estándar mayo 2005	83

Tabla 31 Participantes por clúster en junio 2005	84
Tabla 32 Análisis centroide junio 2005	84
Tabla 33 Análisis de medias y desviaciones estándar junio 2005	85
Tabla 34 Participantes por clúster en julio 2005	86
Tabla 35 Análisis centroide julio 2005.....	86
Tabla 36 Análisis de medias y desviaciones estándar julio 2005	87
Tabla 37 Participantes por clúster en agosto 2005	88
Tabla 38 Análisis centroide agosto 2005	88
Tabla 39 Análisis de medias y desviaciones estándar agosto 2005	89
Tabla 40 Participantes por clúster en septiembre 2005	90
Tabla 41 Análisis centroide septiembre 2005	90
Tabla 42 Análisis de medias y desviaciones estándar septiembre 2005	91
Tabla 43 Participantes por clúster en octubre 2005.....	92
Tabla 44 Análisis centroide octubre 2005	92
Tabla 45 Análisis de medias y desviaciones estándar octubre 2005	93
Tabla 46 Participantes por clúster en noviembre 2005	94
Tabla 47 Análisis centroide noviembre 2005	94
Tabla 48 Análisis de medias y desviaciones estándar noviembre 2005	95
Tabla 49 Participantes por clúster en diciembre 2005	96
Tabla 50 Análisis centroide diciembre 2005	96
Tabla 51 Análisis de medias y desviaciones estándar diciembre 2005	96
Tabla 52 Reacomodo de los clúster.....	97

INTRODUCCIÓN

Al día de hoy, en nuestro país el Banco de México es la institución encargada de la correcta administración de los sistemas de pago, además del poder adquisitivo de la moneda nacional. Es desarrollador y operador del “Sistema de Pagos Electrónicos Interbancarios” (SPEI) desde sus inicios en agosto del 2004. Este sistema, inició su funcionamiento con 38 instituciones, y en la actualidad cuenta con más de 120 bancos participantes, que día con día hacen miles de transferencias entre ellos.

Tomando en cuenta la cantidad de datos que se generan en el SPEI desde sus inicios, hace que el análisis de esta información no pueda realizarse de manera trivial, y se deban utilizar técnicas modernas de computación para su tratado, y es aquí donde la minería de datos tiene un área de oportunidad.

Según (González Bernal, 2017, pág. 7) define a la minería de datos como el proceso de descubrir conocimiento interesante de grandes cantidades de datos almacenadas en bases de datos, datawarehouses u otro repositorio de información.

Entre varias áreas de la minería de datos por explorar, tenemos el *clustering*, también conocido como agrupamiento, el cual, es una técnica que consiste en la división de los datos en grupos de objetos similares. Cuando se representan la información obtenida a través de *clusters* se pierden algunos detalles de los datos, pero a la vez se simplifica dicha información.

El análisis de clúster determina un agrupamiento de entidades (en este caso de bancos), con la finalidad de que cada participante en el clúster tenga características lo más similares posibles, siendo los distintos grupos entre ellos tan disimilares como sea posible.

En un sistema bancario con un número alto de participantes como el de México, surgen algunas interrogantes, que apoyados de la minería de datos podemos intentar resolver, como lo son: ¿Es posible la agrupación de

instituciones bancarias respecto a características similares? ¿Dichos grupos son lo suficientemente homogéneos para que se permita posteriormente la comparación en cada grupo? ¿Las agrupaciones formadas son estables en el tiempo?

El siguiente trabajo, propone la implementación de técnicas basadas en el análisis de clúster, sobre un *dataset* que abarca los periodos agosto 2004 a diciembre 2005, en el cual se analizan los 38 bancos mexicanos que laboraban en el SPEI en ese periodo.

CAPÍTULO 1:
GENERALIDADES

1.1 PLANTEAMIENTO DEL PROBLEMA

El SPEI, es el sistema de pagos más importante de la República Mexicana. En él desde su primer día de funcionamiento, se efectuaron cerca de 147 envíos, y en tan solo un año, el número de envíos logro superar los 47,000. Al día de hoy, esa cantidad se ha quedado atrás, procesando más de 1,000,000 de pagos diarios, lo cual implica una cantidad muy grande de información.

Si tomamos en cuenta que, por cada transferencia enviada, de esta se tienen datos como emisor, receptor, fecha, hora, monto, entre otras características, la base de datos del SPEI puede ser tan grande que una computadora común no podría operar sobre ella.

Los científicos de datos del banco de México, día con día analizan la información del SPEI de distintas formas, estudian distintas variables como lo son, montos de envío y recepción, fechas, lugar de pago, o bien variables no implícitas en la transacción, pero si programables, como lo son los indicadores de red, los cuales le ayudan a medir el comportamiento general de la red, además de dar a conocer más detalles de interés como lo son la liquidez entre bancos, riesgo sistémico, etc.

También conocido como riesgo de sistema, el riesgo sistémico se materializa cuando, sea cual fuere el origen del fallo que afecte a un sistema, el incumplimiento de las obligaciones por parte de una entidad participante en dicho sistema provoca que otras, a su vez, no puedan cumplir con las suyas, generando una cadena de fallos que puede terminar colapsando todo el funcionamiento del mecanismo. (López Domínguez, 2017)

El hecho de conocer grupos de bancos de comportamientos similares en la red de pagos SPEI, podría permitir al Banco de México conocer cuáles son los bancos más grandes de la red, con quienes y cuantos bancos se les relaciona, o quiénes son los bancos más pequeños, si existen bancos que durante el tiempo han cambiado su comportamiento y han ido creciendo hasta

parecerse a otros, entre varios beneficios más que pueden explorarse, y así poder modificar si así se desease sus políticas en beneficio de la red y con esto evitar un posible riesgo sistémico o algún otro inconveniente futuro.

1.2 PROPUESTA DE SOLUCIÓN

Como propuesta de solución, se plantea, analizar la red de pagos del SPEI desde su inicio en agosto 2004 hasta diciembre 2005, para *clusterizar* los bancos participantes respecto a distintos indicadores de red y con esto, poder agrupar a los bancos respecto a aquellos con los que tengan características similares, y separarlos de aquellos con características distintas, todo esto en 5 fases distintas:

Fase 1: Preprocesar los datos en Matlab, con la finalidad de limpiarlos y quitarles el ruido que pudiesen traer.

Fase 2: Ejecutar algoritmos de indicadores de red, los cuales serán sugeridos por el banco de México.

Fase 3: Separar y promediar de manera mensual los datos para poder obtener una base de datos final

Fase 4: Aplicar un algoritmo de clustering en el Python

Fase 5: Analizar los grupos obtenidos, con la finalidad de identificar sus características.

1.3 OBJETIVOS

1.3.1 OBJETIVO GENERAL

Clusterizar bancos mexicanos y analizar las características de los grupos obtenidos, con base en indicadores de red programados a partir de la información obtenida del SPEI en el periodo agosto 2004 a diciembre 2005.

1.3.2 OBJETIVOS ESPECÍFICOS

- Preprocesar los datos provenientes del SPEI
- Analizar y elegir algoritmo de *clustering* a aplicar
- Implementar algoritmo de clustering
- Analizar y validar los resultados obtenidos

1.4 ALCANCES Y LIMITACIONES

1.4.1 ALCANCES

Implementar un algoritmo de *clustering* para la agrupación de los 38 bancos mexicanos que estuvieron desde el inicio del SPEI en el periodo agosto 2004 a diciembre 2005, sobre indicadores de red programados a partir de los datos obtenidos de Matlab, utilizando Python.

1.4.2 LIMITACIONES

Se utilizará el algoritmo de *kmeans* el cual ya está implementado en Python, utilizando la biblioteca scikit.

Los datos proporcionados por el Banco de México para este trabajo son de todos los días de funcionamiento de la red de pagos SPEI del periodo agosto 2004 a diciembre 2005.

Solo se trabajará con los 38 bancos participantes en las fechas antes mencionadas, en la actualidad existen más bancos en el SPEI.

1.5 ESTRUCTURA DE LA TESIS

Este documento se encuentra organizado en seis capítulos los cuales se describen a continuación:

En el capítulo 1, se explica el marco metodológico, es aquí donde se describe a fondo el planteamiento del problema y la propuesta de solución, además se marca un objetivo general y objetivos específicos para en este trabajo de tesis, así como los alcances que se tienen y sus respectivas limitaciones.

En el capítulo 2, el marco teórico consiste en mostrar la teoría que fundamenta el proyecto, conceptos tales como minería de datos, clusterización, aprendizaje supervisado y no supervisado, sistemas de pago, entre otros se explican en este capítulo.

El capítulo 3 contiene el estado del arte, el cuál analiza trabajos sobre clusterización de bancos en otros países.

El capítulo 4 muestra información sobre los datos, la metodología que se realizó para el pre procesamiento de ellos, el algoritmo utilizado para la elección del valor de k , necesario para aplicar *kmeans* y la descripción de la ejecución de dicho algoritmo, así como todo el proceso intermedio que se realizó.

El capítulo 5 muestra los resultados obtenidos posterior a la ejecución del algoritmo de *kmeans* a través de gráficas y tablas comparativas mensualmente, las cuales muestran características de los clusters obtenidos tales como media, desviación estándar, bancos incluidos en él, características del centroide, entre otros.

CAPÍTULO 2

MARCO TEÓRICO

2.1 BANCO DE MÉXICO

El Banco de México es el banco central del Estado Mexicano. Por mandato constitucional, es autónomo en sus funciones y administración. Su finalidad es proveer a la economía del país de moneda nacional y su objetivo prioritario es procurar la estabilidad del poder adquisitivo de dicha moneda. Adicionalmente, le corresponde promover el sano desarrollo del sistema financiero y propiciar el buen funcionamiento de los sistemas de pago. (Banco de México, 2017)

Banco de México considera que es de suma importancia mejorar la comprensión pública de qué es y qué hace la banca central en nuestro país, en particular respecto de sus acciones encaminadas a mantener la estabilidad de precios, procurar el sano desarrollo del sistema financiero, garantizar el buen funcionamiento de los sistemas de pago y proveer un medio de intercambio seguro y confiable para que las personas puedan realizar sus transacciones económicas.

En este sentido, el portal de Divulgación de Banco de México tiene el propósito de fomentar una mayor cultura económica y financiera entre la población. Una mejor educación en estas materias se traduce en un mayor bienestar para los individuos puesto que se encuentran en posibilidad de tomar mejores decisiones económicas. (Banco de México, 2017)

2.2 SISTEMAS DE PAGO

Los sistemas de pagos están constituidos por un conjunto de instrumentos, procedimientos y normas para transferir recursos financieros entre sus participantes. Dichos sistemas son indispensables para que el sistema financiero funcione eficientemente. Algunos de ellos son especialmente críticos ya que, si su diseño no es adecuado, pueden magnificar la transmisión de problemas de liquidez de un participante a los demás y perturbar la estabilidad del sistema financiero. Por estas razones, uno de los objetivos del Banco de México es propiciar el buen funcionamiento de los sistemas de pago del país.

Los billetes y monedas son muy útiles como medios de pago. Sin embargo, en algunos casos éstos tienen algunos inconvenientes. Es necesario protegerlos en cajas de seguridad, vehículos blindados, cerraduras o guardias ya que hay delincuentes que podrían asaltar a quien transporte o guarde una gran cantidad de dinero en billetes. Para muchas transacciones, como comprar un refrigerador o una casa, es además poco práctico contar y transportar billetes. Con el propósito de hacer más fácil las operaciones comerciales y financieras, los bancos han creado instrumentos de pago, como, por ejemplo, las tarjetas de crédito o débito y los cheques. Estos instrumentos de pago permiten pagar cantidades exactas sin la necesidad de llevar billetes. Asimismo, se han desarrollado otras formas de pago, como las transferencias por Internet que logran hacer pagos sin tener que ir a un banco. Con los instrumentos de pago electrónicos, el dinero pasa directamente de una cuenta bancaria a otra.

Para que la gente pueda disponer de instrumentos de pagos seguros y fáciles de usar es indispensable el buen funcionamiento de la infraestructura, reglas y medios eficientes que permitan la transferencia de recursos, lo cual se conoce como sistemas de pagos. Es necesario que estos sistemas sean muy confiables para facilitar la actividad económica de un país. En México, el

Banco de México es la institución por mandato constitucional encargada de cumplir con dicha tarea, es decir, “propiciar el buen funcionamiento de los sistemas de pagos”. (Banco de México, 2017)

Para que la economía funcione bien, es necesario que el conjunto de instrumentos y procesos bancarios que se usan para transferir dinero funcionen bien y sean muy seguros. Este conjunto de instrumentos y procesos bancarios se le conoce como sistemas de pagos. El buen funcionamiento y la seguridad de los sistemas de pago hacen que la gente tenga confianza y con ello se facilitan las actividades económicas. El Banco de México es responsable de que los sistemas de pago sean seguros y eficientes. Un sistema de pagos es seguro en la medida que los riesgos por participar en él estén bien administrados, esto es, los riesgos sean eliminados, atenuados, y/o distribuidos de forma justa entre sus participantes.

Un sistema de pagos es eficiente cuando al utilizarlo es práctico para sus participantes y los costos asociados a su participación son relativamente bajos, considerando los beneficios operativos y/o de reducción de riesgos obtenidos por participar en él.

Los sistemas de pago interbancarios se basan en un contrato que los participantes firman con el operador del sistema. Además del operador del sistema y de los participantes, interviene un agente liquidador que lleva las cuentas de los participantes donde carga y abona el importe correspondiente a los pagos para liquidarlos. El banco central es frecuentemente el agente liquidador de un sistema de pagos ya que generalmente todos los bancos tienen cuentas en él. En México, el agente liquidador para varios sistemas es el Banco de México. (Banco de México, 2017)

2.3 SPEI

SPEI® es un sistema desarrollado y operado por el Banco de México que permite al público en general realizar en cuestión de segundos pagos electrónicos, también llamados transferencias electrónicas, a través de la banca por internet o de la banca móvil. Este sistema permite transferir dinero electrónicamente entre cuentas de depósito de los bancos de manera casi instantánea. (Banco de México, 2017)

El SPEI es un sistema de transferencias electrónicas de fondos que pertenece a y es operado por el Banco de México, el banco central de México. Este sistema se desarrolló con el objetivo de facilitar los pagos entre las instituciones financieras, además de habilitarlas para ofrecer a la población servicios de pago al menudeo de forma segura y eficiente.

El SPEI permite a sus participantes realizar pagos en pesos mexicanos por cuenta propia y a nombre de sus cuentahabientes, prácticamente en tiempo real, las 24 horas del día, todos los días del año. El funcionamiento del SPEI se sustenta en un marco jurídico basado en la Ley de Sistemas de Pagos, de la cual se deriva que la compensación y liquidación de las órdenes de transferencia aceptadas por este sistema son firmes, irrevocables, exigibles y oponibles frente a terceros, con lo que se asegura la finalidad de las operaciones.

El gobierno corporativo del SPEI está alineado al marco normativo y a las políticas del Banco de México. En este marco y políticas están contenidos los objetivos y funciones de la institución, su estructura orgánica, las atribuciones del Gobernador y de la Junta de Gobierno del Banco de México, así como de las unidades administrativas que dependen de estos y los mecanismos de rendición de cuentas, entre otros aspectos que son relevantes para el funcionamiento del SPEI.

2.4 MINERÍA DE DATOS

La definición formal de *data mining* o minería de datos según (Frawley, 1991) citado por (Paniagua Arís & López Ayuso, 2007) es: la extracción no trivial de información implícita, previamente desconocida y potencialmente útil a partir de datos. Otra manera de definirlo podría ser: la exploración y el análisis -por medios automáticos o semiautomáticos- de grandes cantidades de datos con el fin de descubrir patrones con significado.

El *data mining* nació con la idea de aprovechar dos cosas: la ingente cantidad de datos que se almacenaban en áreas como el comercio y la potencia de los nuevos ordenadores para realizar operaciones de análisis sobre esos datos.

El ámbito de la investigación las técnicas de *data mining* pueden ayudar a los científicos a clasificar y segmentar datos y a formar hipótesis. El *data mining* permite encontrar información escondida en los datos que no siempre resulta aparente, ya que, dado el gigantesco volumen de datos existentes, gran parte de ese volumen nunca será analizado.

Las técnicas de data mining pueden ser de dos tipos:

Métodos descriptivos- Buscan patrones interpretables para describir datos. Son los siguientes: clustering, descubrimiento de reglas de asociación y descubrimiento de patrones secuenciales. Se han utilizado, por ejemplo, para ver que productos suelen adquirirse conjuntamente en el supermercado.

Métodos predictivos- Usan algunas variables para predecir valores futuros o desconocidos de otras variables. Son los siguientes: clasificación, regresión y detección de la desviación. (Ingeniesia Desarrollo Cloud, 2015)

2.5 APRENDIZAJE SUPERVISADO

Es una técnica para deducir una función que sea capaz de predecir el valor correspondiente de un dato de entrada a partir de unos datos de entrenamiento, es decir unos ejemplos previamente estudiados. (del Cerro Sánchez & Novalbos Laina, 2005).

Se le otorgan ejemplos al sistema acerca de qué debe hacer exactamente en esos casos, de tal forma que el sistema aprenda de ellos a generalizar situaciones nuevas. (Sánchez-Montañés, Lago, & González, 2017)

2.6 APRENDIZAJE NO SUPERVISADO

Según (Cáceres Tello, 2017), el agrupamiento es una forma de clasificación no supervisada en la que, a diferencia de la supervisada, no se conocen las etiquetas de las clases (no hay clases predefinidas) y puede que tampoco se conozca el número de grupos.

En el aprendizaje no supervisado, los datos de entrenamiento solo contienen información de la entrada $x \in X$. El objetivo es obtener información sobre la estructura del dominio de salida, Y . En problemas de clasificación, esta información se refiere a la (posible) estructura en clases de los datos $x \in X$. En este caso, el problema se conoce como agrupamiento o “clustering”

El aprendizaje no supervisado consiste en que la red descubra por sí misma características, regularidades, correlaciones o categorías en los datos de entrada y se obtengan de forma codificada en la salida. En algunos casos, la salida representa el grado de similitud entre la información que se le está presentado en la entrada y la que se le ha mostrado en el pasado.

2.7 CLUSTERING

Clustering es el proceso de agrupar datos en clases o clusters de tal forma que los objetos de un clúster tengan una similitud alta entre ellos, y baja (sean muy diferentes) con objetos de otros clusters. La medida de similaridad está basada en los atributos que describen a los objetos. Los grupos pueden ser exclusivos, con traslapes, probabilísticos, jerárquicos. El clustering puede ser aplicado, por ejemplo, para caracterizar clientes, formar taxonomías, clasificar documentos, etc.

Existe una gran cantidad de algoritmos de clustering (solo vamos a ver algunos). En particular existen diferentes algoritmos basados en: (Morales & Escalante, 2017)

- Particiones
- Jerárquicos
- Densidades
- Rejillas
- Modelos
- Teoría de grafos
- Búsqueda Combinatoria
- Técnicas Fuzzy
- Redes Neuronales
- Kernels
- Para Datos Secuenciales
- Para grandes conjuntos de datos.

2.8 ALGORITMOS DE CLUSTERING

2.8.1 KMEANS

El algoritmo K-means, creado por MacQueen en 1967 es el algoritmo de clustering más conocido y utilizado ya que es de muy simple aplicación y eficaz. Sigue un procedimiento simple de clasificación de un conjunto de objetos en un determinado número K de clústeres, K determinado a priori. El nombre de K-means viene porque representa cada uno de los clusters por la media (o media ponderada) de sus puntos, es decir, por su centroide. La representación mediante centroides tiene la ventaja de que tiene un significado gráfico y estadístico inmediato. Cada cluster por tanto es caracterizado por su centro o centroide que se encuentra en el centro o el medio de los elementos que componen el clúster. Kmeans es traducido como K-medias. (García Cambroneró & Gómez Moreno, 2017)

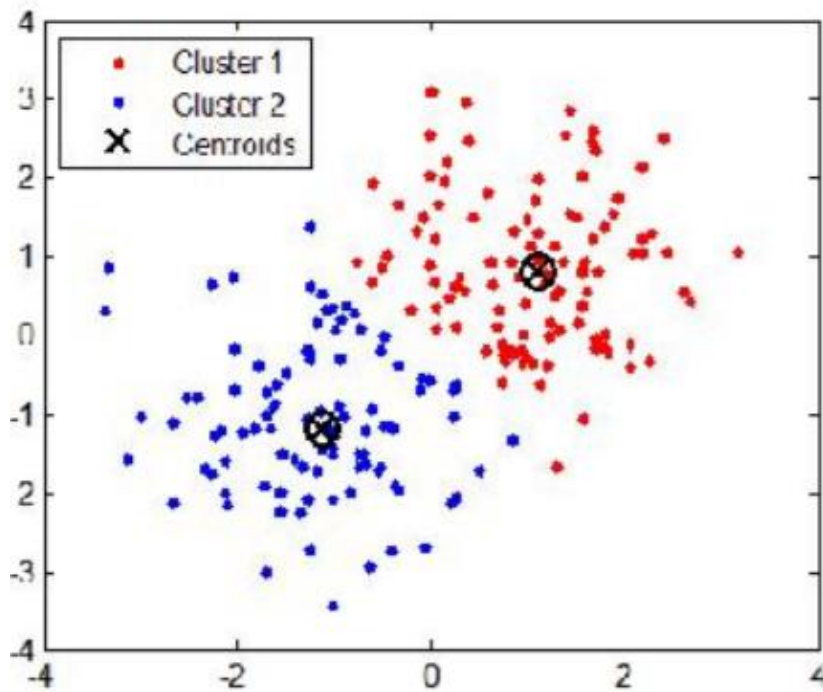


Figura 2.1 Ejemplo de ejecución de KMeans

Sea un conjunto de objetos $D_n = (x_1, x_2, \dots, x_n)$, para todo el i , x_i reales y k, v_1 , los centros de los K clúster. El algoritmo del K-means se realiza en 4 etapas:

Etapa 1: Elegir aleatoriamente K objetos que forman así los K clusters iniciales. Para cada clúster k , el valor inicial del centro es $= x_i$, con los x_i únicos objetos de D_n pertenecientes al clúster.

$$\hat{s} = \operatorname{argmin} \|u_k - x\|^2$$

Etapa 2: Reasigna los objetos del clúster. Para cada objeto x , el prototipo que se le asigna es el que es más próximo al objeto, según una medida de distancia, (habitualmente la medida euclidiana).

Etapa 3: Una vez que todos los objetos son colocados, recalculan los centros de K clúster. (los baricentros).

Etapa 4: Repetir las etapas 2 y 3 hasta que no se hagan más reasignaciones. Aunque el algoritmo termina siempre, no se garantiza el obtener la solución óptima. En efecto, el algoritmo es muy sensible a la elección aleatoria de los K centros iniciales. Esta es la razón por la que, se utiliza el algoritmo del K-means numerosas veces sobre un mismo conjunto de datos para intentar minimizar este efecto, sabiendo que a centros iniciales lo más espaciados posibles dan mejores resultados.

2.8.2 DBSCAN

Es un algoritmo basado en densidad, se definen los conceptos de punto central (puntos que tienen en su vecindad una cantidad de puntos mayor o igual que un umbral especificado), borde y ruido.

El algoritmo comienza seleccionando un punto p arbitrario, si p es un punto central, se comienza a construir un grupo y se ubican en su grupo todos

los objetos denso-alcanzables desde p . Si p no es un punto central se visita otro objeto del conjunto de datos. El proceso continúa hasta que todos los objetos han sido procesados. Los puntos que quedan fuera de los grupos formados se llaman puntos ruido, los puntos que no son ni ruido ni centrales se llaman puntos borde. De esta forma DBSCAN construye grupos en los que sus puntos son o puntos centrales o puntos borde, un grupo puede tener más de un punto central. (Ester, Krieger, Sander, & Xu, 1996)

2.8.3 OPTICS

La motivación para la realización de este algoritmo se basa en la necesidad de introducir parámetros de entrada en casi todos los algoritmos de agrupamiento existentes que en la mayoría de los casos son difíciles de determinar, además en conjuntos de datos reales no existe una manera de determinar estos parámetros globales, el algoritmo OPTICS trata de resolver este problema basándose en el esquema del algoritmo DBSCAN creando un ordenamiento de la base de datos para representar la estructura del agrupamiento basada en densidad, además puede hacer una representación gráfica del agrupamiento incluso para conjuntos de datos grandes. (Pascual, 2017)

2.8.4 CLUSTERING JERÁRQUICO

Este método permite la construcción de un árbol de clasificación, que recibe el nombre de dendograma un ejemplo se muestra en la Figura 2.2, en el cual se puede seguir de forma gráfica el procedimiento de unión seguido, mostrando que grupos se van uniendo, en qué nivel concreto lo hacen, así como el valor de la medida de asociación entre los grupos cuando éstos se agrupan (valor que llamaremos nivel de fusión).

En resumen, la forma general de operar de estos métodos es bastante simple. Por ejemplo, en los métodos aglomerativos se parte de tantos grupos como individuos haya. A continuación, se selecciona una medida de similitud, agrupándose los dos grupos o clusters con mayor similitud. Así se continúa hasta que:

1. Se forma un solo grupo.
2. Se alcanza el número de grupos prefijado.
3. Se detecta, a través de un contraste de significación, que hay razones estadísticas para no continuar agrupando clusters, ya que los más similares no son lo suficientemente homogéneos como para determinar una misma agrupación. (Gallardo, 2017)

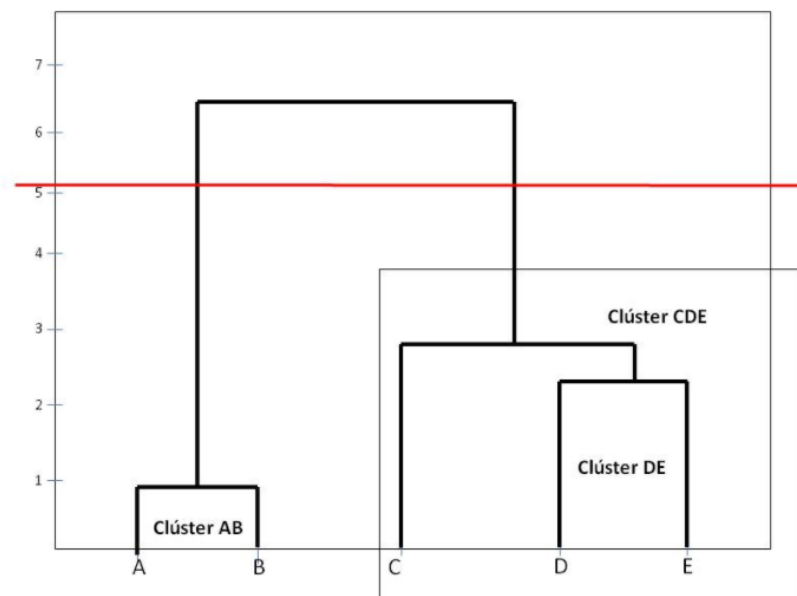


Figura 2.2 Ejemplo de ejecución de Clustering Jerárquico

2.9 MATLAB

Matlab es un sistema de programación y cálculo basado en la manipulación de matrices. El nombre mismo del sistema o paquete de cómputo proviene de la abreviación MATriz LABoratory o Laboratorio de Matrices.

MATLAB es un gran programa de cálculo técnico y científico. Para ciertas operaciones es muy rápido, cuando puede ejecutar sus funciones en código nativo con los tamaños más adecuados para aprovechar sus capacidades de vectorización.

MATLAB dispone de un código básico y de varias librerías especializadas (toolboxes). En estos apuntes se hará referencia exclusiva al código básico. MATLAB incluye una gran cantidad de funciones predefinidas, que ayudan a realizar cálculos de todo tipo así como para visualización de datos y resultados. (Zúñiga, 2008)



Figura 2.3 Software Matlab

2.10 PYTHON

2.10.1 Lenguaje Python

Python es un lenguaje de programación creado por Guido van Rossum a principios de los años 90 cuyo nombre está inspirado en el grupo de cómicos ingleses “Monty Python”.

Es un lenguaje similar a Perl, pero con una sintaxis muy limpia y que favorece un código legible. Se trata de un lenguaje interpretado o de script, con tipado dinámico, fuertemente tipado, multiplataforma y orientado a objetos.

Python tiene, muchas de las características de los lenguajes compilados, por lo que se podría decir que es semi-interpretado. En Python, como en Java y muchos otros lenguajes, el código fuente se traduce a un pseudocódigo máquina intermedio llamado bytecode la primera vez que se ejecuta, generando archivos .pyc o .pyo (bytecode optimizado), que son los que se ejecutarán en sucesivas ocasiones.

Su sintaxis simple, clara y sencilla; el tipado dinámico, el gestor de memoria, la gran cantidad de librerías disponibles y la potencia del lenguaje, entre otros, hacen que desarrollar una aplicación en Python sea sencillo, muy rápido. Python no es adecuado sin embargo para la programación de bajo nivel o para aplicaciones en las que el rendimiento sea crítico. (González Duque, 2017)

2.10.2 Spyder

Spyder (anteriormente Pydee) es un entorno de desarrollo integrado y multiplataforma de código abierto (IDE) para programación científica en el lenguaje Python. Este IDE se liberó bajo la licencia de MIT. Spyder es extensible con complementos. Incluye soporte de herramientas interactivas para la inspección de datos e incorpora controles de calidad específicos de Python e instrumentos como Pyflakes, Pylint y Rope.

Como ya he dicho, es un IDE multiplataforma a través de Anaconda, en Windows con WinPython y Python (x, y), en macOS a través de MacPorts. También está disponible para las principales distribuciones de Gnu/Linux como Arch Linux, Debian, Fedora, Gentoo Linux, openSUSE y Ubuntu.

Desde mediados de noviembre de 2017, Anaconda ha dejado de financiar el desarrollo de este IDE, después de hacerlo durante los últimos 18 meses. Debido a esto, el desarrollo se centrará desde ahora en mantener Spyder 3 a un ritmo mucho más lento que antes, aun que esto no supone que abandonen el proyecto. Se puede saber más acerca de esta noticia en el siguiente enlace. (Amoedo, 2017)

2.10.3 scikit-learn

Scikit-learn es una librería especializada en algoritmos para data mining y machine learning.

Proporciona funciones para animaciones en estadística, cubriendo temas en teoría de probabilidad, estadística matemática, estadísticas multivariantes, estadísticas no paramétricas, encuestas de muestreo, modelos lineales, series de tiempo, estadísticas computacionales, minería de datos y aprendizaje automático. Estas funciones pueden ser útiles en la enseñanza de estadísticas y análisis de datos.

Algunos de los problemas que podemos resolver utilizando las herramientas de Scikit-learn, son:

- **Clasificaciones:** Identificar las categorías a que cada observación del conjunto de datos pertenece.
- **Regresiones:** Predecir el valor continuo para cada nuevo ejemplo.
- **Agrupaciones:** Agrupación automática de objetos similares en un conjunto.
- **Reducción de dimensiones:** Reducir el número de variables aleatorias a considerar.
- **Selección de Modelos:** Comparar, validar y elegir parámetros y modelos.
- **Pre procesamiento:** Extracción de características a analizar y normalización de datos. (López Briega, 2014)

2.11 INDICADORES DE RED

Para poder llegar a los objetivos de este trabajo, se requiere del uso de varios indicadores de red, algunos de ellos implícitos en la base de datos original, otros de ellos deberán ser programados para su obtención. Según (Bravo-Benitez, Kabadjova, & Martinez-Jaramillo, 2014) Estos indicadores nos ayudarán a entender mejor el comportamiento del sistema.

2.11.1 Degree

El grado de un nodo (*degree*) en una red es una medida simple pero muy útil. Esta medida captura la cantidad de nodos a los que se conecta. El grado, $d(i)$, de un vértice, i , en un grafo se define como:

$$d_i = \sum_{j \in N(i)} a_{ij}$$

donde $N(i)$ es el conjunto de vecinos del vértice i ; es decir, el conjunto de vértices que tienen una conexión con el vértice i .

El grado interno (*inner degree*), d_i^- y el grado externo (outer degree), d_i^+ de un nodo, i , en un dígrafo se definen como:

$$d_i^- = \sum_{j \in N^-(i)} a_{ij} \quad d_i^+ = \sum_{j \in N^+(i)} a_{ij}$$

donde $N^-(i)$ es el conjunto de vecinos internos del vértice i , que es el conjunto de nodos que tienen un arco que termina en el nodo i . $N^+(i)$ es el conjunto de

vecinos externos de vértice i , el conjunto de nodos que tienen un arco que comienza en el nodo i .

2.11.2 Closeness

La centralidad de proximidad (*Closeness centrality*) tiene una interpretación de la independencia en las redes sociales en términos de control de comunicación. Un nodo con alto *closeness* dependería menos de otros nodos intermediarios para recibir mensajes. En el contexto del riesgo sistémico y el contagio financiero, esta medida puede ser asociado con la capacidad de un nodo para propagar el contagio, como tal un nodo está cerca del resto de la red. Se define como:

$$C_C(v) = \sum_{j \in V \setminus \{v\}} \frac{1}{d_G(v, j)}.$$

En el que $d_G(v, j)$ indica la longitud de la ruta más corta que une v y j .

2.11.3 Betweenness

La centralidad de intermediación (*Betweenness*), en las redes sociales, se asocia con ser estratégicamente ubicado en las rutas de comunicación de muchos otros nodos en la red. Un nodo con alta centralidad de intermediación tendría una importante influencia en otros nodos, ya que puede detener o distorsionar la información que pasa a través de él. Esta medida de centralidad es particularmente relevante en redes de sistemas de pago.

Estableciendo $\sigma_{ij} = \sigma_{ji}$ y denotando el número total de caminos más cortos entre

i y j . Y que $\sigma_{ij}(v)$ sea el número total de caminos más cortos entre i y j que pasan a través del vértice v , entonces

$$C_B(v) = \sum_{i \neq v \neq j \in V} \frac{\sigma_{ij}(v)}{\sigma_{ij}}.$$

2.11.4 Affinity

La afinidad (*affinity*) es una medida que, en función del grado de un nodo, describe el tipo de nodos a los que dicho nodo tiende a tener un enlace.

$$a_i = \frac{1}{d_i} \sum_{j \in \mathcal{N}(i)} d_j$$

Si a_i está aumentando con d_i , entonces los nodos con alto grado tienden a tener relaciones con nodos que poseen un grado similar. Si a_i disminuye con d_i , entonces la mayoría de los vecinos de nodos de alto grado tienen menor grado.

Por el contrario, los nodos con bajo grado tienden a tener relaciones con nodos de alto grado.

Esta medida describe si los nodos en una red tienden a tener relaciones con nodos de grado similar o nodos con un grado diferente.

CAPÍTULO 3:
ESTADO DEL ARTE

3.1 CLUSTERING AUSTRIAN BANKS' BUSINESS MODELS AND PEER GROUPS IN THE EUROPEAN BANKING SECTOR

En este trabajo se muestra el agrupamiento de grandes bancos europeos para identificar clusters de bancos seleccionados. El análisis se basa en datos de los estados financieros de los bancos. Se utilizó un método de agrupamiento de k-means basado en la distancia de Mahalanobis para asignar los bancos a grupos que representan modelos de negocios similares. Se encontró que los modelos de los bancos pueden agruparse en cinco modelos de negocio distintos. El artículo proporciona una lista de bancos con su respectivo modelo de negocio al que pertenecen, además de que resaltan la participación de los principales bancos de Austria.

3.2 UKRAINIAN BANKS' BUSINESS MODELS CLUSTERING: APPLICATION OF KOHONEN NEURAL NETWORKS

En este artículo, para poder clusterizar a los bancos, se utilizaron redes neuronales Kohonen, con ellas, identificaron los modelos de negocio bancarios que prevalecieron en Ucrania, separándolos en HTC, Retail, Universal, Corporativo, Inversión, Venta al por mayor y a los bancos congelados/indecisos. También se comenta que estos modelos surgen a partir de la crisis bancaria, pues demuestran que más de la mitad de HTC y de los bancos congelados/indecisos, quebraron, lo que indicaba que esos modelos eran más riesgosos.

Como complemento del análisis realizado, construyeron un mapa de riesgo, basado en el conjunto de indicadores de riesgo: Liquidez, Pro fitness, Crédito, Concentration y Money Laundering.

3.3 STRATEGIC GROUPS IN AUSTRIAN BANKING

Este artículo utiliza kmeans para clisterizar a los 35 más importantes bancos Austriacos entre el periodo de los años 1995 y 2000, identificando grupos estratégicos de instituciones crediticias tan parecidas como su pudiese- El autor, eligió como parámetros de entrada el índice de consolidación, fondos disponibles, depósitos, préstamos interbancarios, y pérdidas operacionales. Como resultado se obtuvo cinco grupos de bancos, de los cuales solo tres son considerados como estratégicos. Los otros dos clusters son muy volátiles a través del tiempo así que no son considerados como clúster formalmente en esa investigación.

3.4 CLUSTERING OF RUSSIAN BANKS: BUSINESS MODELS OF INTERACTION OF THE BANKING SECTOR AND THE REAL ECONOMY

Este trabajo, clusteriza bancos de Rusia, identificando modelos de negocio de instituciones crediticias, considerando la interconexión de los bancos en cada uno de los grupos, todo esto basado en la interacción que se tiene día con día. El documento, se enfoca en el método de gestión de liquides, ya que está centrado en el mercado de los préstamos interbancarios, ya que esto es lo que más refleja el comportamiento de una institución bancaria. El cálculo de los índices se realizó basándose en los datos actuales de las entidades de crédito rusas, quienes revelan sus declaraciones al banco de

Rusia. Confirman que el análisis es actualmente uno de los más populares y avanzados métodos de agrupación matemática tanto en la economía y otras ciencias existentes, permitiendo clasificar los bancos por muchas de sus propiedades al mismo tiempo. Utilizan una variante de kmeans, ya que consideran que es el método más adecuado para la clusterización de sus bancos. Para elegir el valor de k, toman en cuenta la variabilidad interna mínima, haciendo que el algoritmo en cada iteración mejore el número de k y forme nuevas agrupaciones. Utilizan distancia euclidiana.

3.5 THE CLUSTER ANALYSIS OF THE BANKING SECTOR IN EUROPE

Este trabajo se enfoca en analizar los bancos participantes en la Union Europea durante la crisis en el 2008 mediante rangos de indicadores bancarios consolidados por el Banco Central Europeo. El estudio busca determinar patrones similares según las proporciones del sector bancarios y los cambios detectados en la crisis financiera. Los resultados confirman que los países en áreas geográficas similares y con una mayor asociación económica tienden a agruparse en grupos de manera similar.

3.6 AGRUPACIÓN DE INSTITUCIONES BANCARIAS A PARTIR DEL ANÁLISIS DE CLUSTER: UNA APLICACIÓN AL CASO DE CHILE

El documento aplica técnicas multivariadas para la clasificación de las instituciones financieras de acuerdo a su grado de similitud, conocida como análisis de clúster. Haciendo uso de información de balance para el período 2008-2013, sobre un conjunto de 23 bancos en Chile, encontraron que la que

la industria bancaria puede ser agrupada en siete grupos de instituciones homogéneas a) multi-bancos grandes, b) multi-bancos medianos, c) banca especializada mediana, d) banca de consumo, d) bancos de tesorería, e) bancos de comercio exterior y f) banca dedicada a los servicios financieros. Adicionalmente, muestra que la clasificación de los bancos es estable a través del tiempo y que puede contribuir a un mejor monitoreo de los riesgos del sistema bancario chileno.

CAPÍTULO 4

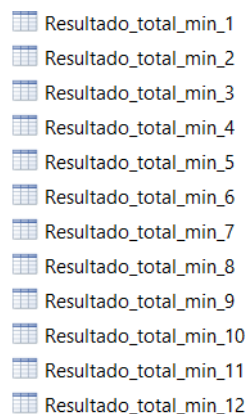
EXPERIMENTACIÓN

4.1 INTRODUCCIÓN

Este capítulo, muestra el diseño y desarrollo del pre procesamiento de los datos, paso a paso desde su forma original, reducción de campos, unión de archivos, reducción de bancos al tamaño necesitado, elaboración de indicadores de red, normalización de datos, cálculo de K, y ejecución del algoritmo de Kmeans, así como todo lo necesario para poder llegar a los resultados deseados.

4.2. DATOS

El Banco de México entregó un conjunto de archivos que extrajeron del SPEI y enviaron a Matlab. En total, se tienen 17 archivos con el nombre Resultado_total_min_X, donde X es el número del mes correspondiente. Se presentan en cinco archivos del año 2004, y similarmente doce archivos para el 2005. La siguiente Figura muestra dichos archivos del año 2005.



Resultado_total_min_1
Resultado_total_min_2
Resultado_total_min_3
Resultado_total_min_4
Resultado_total_min_5
Resultado_total_min_6
Resultado_total_min_7
Resultado_total_min_8
Resultado_total_min_9
Resultado_total_min_10
Resultado_total_min_11
Resultado_total_min_12

Figura 4.1 Archivos entregados por el banco

Todos los archivos contenían la misma estructura, la cual es de tamaño 1 X N dónde, dónde N es el número de días que funcionó el SPEI en ese mes. El archivo contiene 10 campos los cuales son:

- Saldos

- Mínimo
- Sumas
- Payments
- Payments_num
- Minimos_hora
- Suma_Hora
- S_Institution
- Reserve
- Nombre_i

Entre los campos que destacan, Sumas contiene un arreglo diario de 130 x 1 con la suma de envíos realizados por un banco hacia todos los demás, payments contiene matrices diarias de tamaño 130 x 130 de bancos contra bancos con el monto total que cada banco envió en el día hacia los demás, payments_num, es similar a payments, solo que esta contiene información de la cantidad de pagos, nombre_i contiene el nombre del día que se tiene almacenado del SPEI. Para fines de este trabajo se utilizarán los campos de payments y nombre_i. La Figura siguiente muestra parte de la estructura original del archivo.

Fields	saldos	minimo	sumas	payments	payments_num
1	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double
2	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double
3	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double
4	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double

minimos_hora	suma_hora	S_Institution	reserve	nombre_i
130x24 double	130x24 double	1x130 struct	1x1 struct	'01-02-2005....
130x24 double	130x24 double	1x130 struct	1x1 struct	'02-02-2005....
130x24 double	130x24 double	1x130 struct	1x1 struct	'03-02-2005....
130x24 double	130x24 double	1x130 struct	1x1 struct	'04-02-2005....

Figura 4.2 Contenido de los archivos entregados por el banco

4.3 PREPROCESAMIENTO

4.3.1 Unión de archivos diarios

La primera actividad a realizar para poder obtener la base de datos final, fue la elaboración de una nueva estructura, la cual, deberá unir todos los archivos antes mencionados, además de solo mantener las dos columnas relevantes para nuestro trabajo, payments y nombre_i. Dicha estructura se realizará en Matlab. El código se muestra a continuación

```
1 - datosEntrada='E:\Programacion\Matlab\29 marzo\datos';
2 - salida='E:\Programacion\Matlab\salida';
3 - cd(datosEntrada);
4 - Dir1_meses=dir;
5 - numMeses=length(Dir1_meses);
6 - num_inst=98;
7 - k=1;
8 - for mes=3:numMeses
9 -     %Accedo ya con esto a la carpeta (año)
10 -     nombrei=Dir1_meses(mes).name;
11 -
12 -     load(nombrei);
13 -     numd=size(Resultado_total_min,2);
14 -
15 -     aux=100;
16 -     for x=1:numd
17 -
18 -         nombre=strcat('a',Resultado_total_min(x).nombre_i);
19 -         s=Resultado_total_min(x).payments_num;
20 -         estructura2(k)=struct('payments',h,'dia',nombre);
21 -         k=k+1;
22 -         aux=aux+1;
23 -     end
24 -     clear numd;
25 -
26 - end
27 - cd(salida);
28 - save(strcat('estructura_agosto2004_dic_2005_aumentada'),'estructura
```

Figura 4.3 Código para generar estructura

Como ya se mencionó, posterior a la ejecución de este código se tiene una estructura con solo dos campos que son los necesarios para el siguiente paso.

4.3.2 LIMPIEZA DE RUIDO, REDUCCIÓN DE BANCOS

El siguiente paso en el pre procesamiento de nuestros datos es la limpieza del ruido. El único tipo de ruido reconocible, son el excedente de los bancos, ya que la estructura original cuenta con matrices de tamaño 130, y muchos valores de ellos son cero, puesto que no todos los bancos laboraban en el SPEI en el periodo de tiempo que estamos trabajando. Las columnas y filas con datos que nos interesan son del 1 al 32 y del 56 al 61 ya que son aquellas que contienen información.

Para poder hacer la limpieza de los datos antes mencionada se programó y ejecutó el siguiente código.

```
1 - load('estructura_agosto2004_dic_2005_aumentada.mat');
2 - datos=estructura2;
3 - a=1:32
4 - b=56:61
5 - c=[a,b];
6 - for i=1:355
7 -     pag=datos(i).payments(c,c);
8 -     dia=datos(i).dia;
9
10 -     datos_res(i)=struct('pagos',pag,'dias',dia);
11
12 - end
13 - save(strcat('nueva_estructura_38'),'datos_res');
..
```

Figura 4.4 Código para generar estructura con 38 bancos

El código anterior, toma el archivo obtenido con anterioridad y selecciona las filas y columnas de los bancos con información, las une en un vector, y utilizando un ciclo recorre cada uno de los días del SPEI reduciendo las matrices de tamaño 130 a solo 38 para finalmente guardar la nueva estructura.

4.3.3 PROGRAMACIÓN DE INDICADORES

Una vez obtenida una estructura, la cual contiene solo a los bancos que nos interesan, y con los campos de payments y el nombre del día, la siguiente tarea es programar los indicadores de red necesarios. Para esto, se deberá crear primero una matriz de adyacencias, la cual contendrá un 1 para aquellos bancos quienes el banco emisor hace un envío a otro, y un 0 a aquellos bancos con los que no se comunica. Dicha matriz será denominada W_b , su aspecto será como la siguiente figura.

Emisores	0	1	1	0	0	1
	1	0	1	0	1	1
	0	1	0	1	0	1
	0	1	1	0	1	0
	1	1	1	0	0	1
	0	1	1	0	0	0
	Receptores					

Figura 4.5 Matriz de adyacencia W_b

Ahora se procede a calcular la matriz de adyacencia W_l , que es la matriz de recepción de pagos. $W_l = W_b^T$.

Receptores	0	1	0	0	1	0
	1	0	1	1	1	1
	1	1	0	1	1	1
	0	0	1	0	0	0
	0	1	0	1	0	0
	1	1	1	0	1	0
	Emisores					

Figura 4.6 Matriz de adyacencia W_l

Finalmente, con las matrices W_l y W_b procederemos a calcular la matriz W quien será la matriz de pesos que indique el número de relaciones que tiene una entidad en ese día. Se expresa por $W = W_l + W_b$,

Participante	0	1	1	0	1	1
	1	0	1	1	1	1
	1	1	0	1	1	1
	0	1	1	0	1	0
	1	1	1	1	0	1
	1	1	1	0	1	0
	Participante comunicado					

Figura 4.7 Matriz de adyacencia W

Una vez calculados W , W_l y W_b se procede a calcular los *degree*. Como ya me mencionó en el capítulo dos, el degree es una medida usada para calcular el número de nodos con que un elemento está conectado. El *degree* se encuentra definido cómo:

$$d_i = \sum_{j \in N(i)} a_{ij}$$

Donde $N(i)$ es el conjunto de vértices vecinos i (vértices que tienen una arista hacia i). Relacionado con la figura 4.8 la siguiente imagen nos muestra el degree del ejemplo.

Degree	4
	5
	5
	3
	5
	4

Figura 4.8 Degree representativo a la imagen 4.8

Con apoyo del degree, y con base en el conocimiento sustentado en el marco teórico, se programaron los demás indicadores necesarios para la siguiente etapa, como lo son *inner degree*, *outer degree*, montos in, montos out, *betweenness*, *affinity* y *closeness*, y se almacenaron en una nueva estructura de Matlab como se muestra en la siguiente figura.

```

datos=D_Resultado_matrices_2013_01_95;
for i=1:355
nombre=extractBefore(extractAfter(datos(i).G.dias,1),11);
degree=datos(i).D;
degree_in=datos(i).D_in;
degree_out=datos(i).D_out;
montos_in=datos(i).S1;
montos_out=datos(i).Sb;
affinity=datos(i).Aff;
betweenness=datos(i).Betweenness;
closeness=datos(i).Closeness;
salida(i)=struct('nombre',nombre,'degree',degree,'degree_in',degree_in,'degree_out',...
degree_out,'montos_in',montos_in,'montos_out',montos_out,'affinity',affinity,'betweenness',betweenness,'closeness',closeness);
end
save(strcat('testbdfinalabril'),'salida');

```

Figura 4.9 Código para seleccionar solo indicadores

Finalmente, se deberá dar el formato de filas y columnas esperado como parámetro de entrada en Python, por lo tanto, se utilizará el siguiente código el cual generará un archivo .mat por cada día que laboró el SPEI.

```

for i=1:355
    fecha=datos(i).nombre;
    degree=datos(i).degree;
    degree_in=datos(i).degree_in;
    degree_out=datos(i).degree_out;
    montos_in=datos(i).montos_in;
    montos_out=datos(i).montos_out;
    affinity=datos(i).affinity;
    betweenness=datos(i).betweenness;
    closeness=datos(i).closeness;

    salidasIn(:,1)=degree';
    salidasIn(:,2)=degree_in';
    salidasIn(:,3)=degree_out';
    salidasIn(:,4)=montos_in';
    salidasIn(:,5)=montos_out';
    salidasIn(:,6)=affinity';
    salidasIn(:,7)=betweenness';
    salidasIn(:,8)=closeness';

    save(strcat('bdd_',int2str(i),'_',fecha),'salidasIn');
end

```

Figura 4.10 Código para generar base de datos diaria

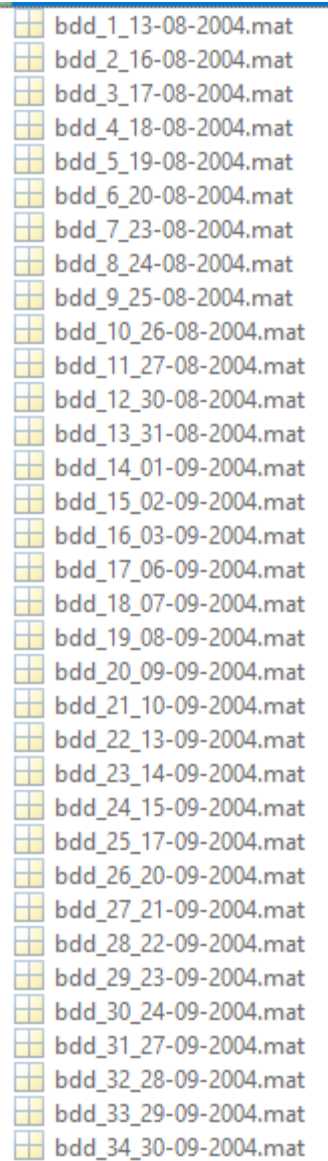


Figura 4.11 Imagen de muestra de archivos diarios de matlab

EXPORTACIÓN A CSV

Para el uso de los datos en Python, estos, se deben exportar en formato csv, lo cual se realizó con el siguiente código

```

ruta='D:\Programacion\Matlab\abril\bdram';
cd(ruta);
directorio=dir;
for i=3:357
    nombre=directorio(i).name;
    load(nombre);
    largo=length(nombre)-3;
    nombreNuevo=strcat(extractBefore(nombre,largo),'.csv');
    csvwrite(nombreNuevo,salidasIn);
end

```

Figura 4.12 Código usado para exportar a csv

Finalmente se deberán agrupar los valores por mes en distintas carpetas

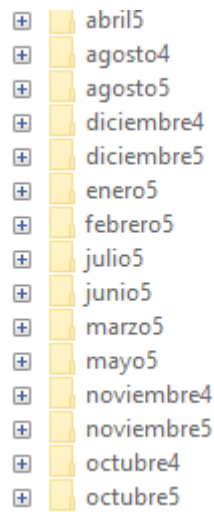


Figura 4.13 Ejemplo de salida en carpetas

Y su contenido será:

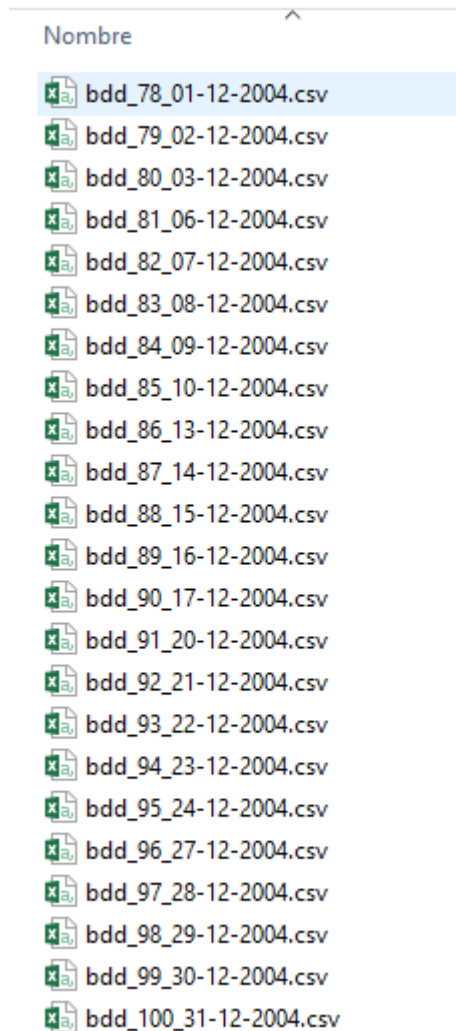


Figura 4.14 Ejemplo de contenido en carpetas

El cual ejemplifica todos los días que trabajó el SPEI en ese mes.

PYTHON

PROMEDIO MENSUAL

En Python se elaboró un script el cual promediará los archivos mensuales, el cual hace uso de librerías tales como **numpy**, **pandas** y **os**.

```

8 import numpy as np
9 import pandas as pd
10 import os
11
12
13 encabezados=["degree", "degreeIn", "degreeOut", "montosIn", "montosOut", "affinity", "betweenness", "closeness"]
14 mes="diciembre5"
15 nombreSalida="diciembre 2005.csv"
16 x=np.zeros(shape=(38,8))
17 contador=0
18
19 archivos= os.listdir(mes)
20 datatotal=pd.DataFrame(x,columns=encabezados)
21
22
23 for dia in archivos:
24     ruta=mes+"/"+dia
25     temp = pd.read_csv(ruta,names=encabezados)
26     datatotal=datatotal+temp
27     contador=contador+1
28
29 datatotal=datatotal/contador
30
31 datatotal.to_csv(nombreSalida)
32

```

Figura 4.15 Promediado mensual en python

La salida de este proceso, otorga un csv por cada mes con contenido como el siguiente

	A	B	C	D	E	F	G	H	I
1		degree	degreeIn	degreeOut	montosIn	montosOut	affinity	betweenness	closeness
2	0	31.7142857	31.5714286	30.5238095	1.2363E+11	1.2496E+11	15.1462381	55.525081	31.8571428
3	1	0	0	0	0	0	0	0	0
4	2	31.2380952	30.8095238	30.7142857	8.5845E+10	7.8697E+10	15.4335714	18.7103905	31.6190476
5	3	30.047619	29.3809524	27.2857143	8.6014E+10	8.4815E+10	15.8271905	25.0083857	31.0238095
6	4	28.8571429	27.6190476	25.6666667	2.432E+10	2.7212E+10	16.2228095	19.4204714	30.4285714
7	5	5.71428571	4.47619048	4.23809524	901367619	895730476	28.5723333	0	18.8571428
8	6	19.6666667	17	15.7142857	3327380952	3253352381	19.650381	5.65530952	25.8333333
9	7	21.1904762	18.5714286	16	3323585714	3365461905	18.8058095	9.4850619	26.5952381
10	8	20.952381	18.3333333	14.6190476	1.2211E+10	1.3091E+10	18.7779048	10.5819	26.452381
11	9	11.5238095	9.19047619	9.28571429	674840000	546930476	24.13	0.3872559	21.7380952
12	10	14.5238095	11.3333333	10.7619048	785130000	795193810	21.5110952	3.31825714	23.2619048
13	11	28	26.6190476	23.047619	1.139E+10	1.0553E+10	16.4992381	36.848619	30
14	12	16.952381	15.4285714	11.8571429	873879524	892798095	20.7805714	4.06559048	24.4761905
15	13	14.952381	11.8095238	11.4285714	1254221905	1166411429	21.9521905	2.6213481	23.4761905
16	14	13.1428571	10	10.5238095	476294810	478694429	22.5838095	2.0981981	22.5714286
17	15	14	12.1428571	10.6190476	818702857	860089524	21.6688571	2.41033095	23
18	16	27.8095238	26.6190476	23.8571429	1.4273E+10	1.4000E+10	16.5063333	41.8273333	28.8047619

Figura 4.16 Contenido del archivo promediado

ESCALAMIENTO

Para obtener mejores resultados en el algoritmo de clustering a aplicar, y los datos no tiendan hacia una variable, los datos se deben escalar

de manera igualitaria, por lo tanto, se programó el script escalador, el cual transforma todos los valores que se tenían a valores entre 0 y 1.

```

7
8 import pandas as pd
9 import os
10 from sklearn.preprocessing import MinMaxScaler
11
12
13 encabezados=["degree","degreeIn","degreeOut","montosIn","montosOut","affinity","betweenness","closeness"]
14 carpeta="resumen mensual"
15 archivos= os.listdir(carpeta)
16
17 escalador = MinMaxScaler()
18
19 for mes in archivos:
20     ruta=carpeta+"/"+mes
21     temp = pd.read_csv(ruta)
22     temp=temp.drop("Unnamed: 0",1)
23     df_escalado = escalador.fit_transform(temp)
24     df_escalado=pd.DataFrame(df_escalado, columns=encabezados);
25     nombre_salida=mes[:-4]+"_escalado.csv"
26     df_escalado.to_csv(nombre_salida)
27

```

Figura 4.17 Código para normalizar los valores

Dicho script después de su ejecución, arrojo archivos de la siguiente manera:

	A	B	C	D	E	F	G	H	I
1		degree	degreeIn	degreeOut	montosIn	montosOut	affinity	betweenness	closeness
2	0	0.99848714	1	0.97392638	1	1	0.53169535	0.67345188	0.99924869
3	1	0.00151286	0.00151976	0	0	7.37E-08	0.01982095	0	0.0202855
4	2	1	0.98632219	1	0.70406599	0.64145231	0.53285323	0.727614	1
5	3	0.95310136	0.93009119	0.87576687	0.81007553	0.80421881	0.55138912	0.28775436	0.97670924
6	4	0.92738275	0.89057751	0.8404908	0.1643859	0.18137938	0.55918701	0.4099366	0.96393689
7	5	0.18305598	0.13829787	0.15184049	0.00768326	0.00760225	1	0	0.59403907
8	6	0.60816944	0.54255319	0.5	0.02905596	0.02943102	0.6889779	0.13328432	0.80540947
9	7	0.66717095	0.56838906	0.52453988	0.02374845	0.02465448	0.66198342	0.17060962	0.8344598
10	8	0.66717095	0.56534954	0.52453988	0.12397691	0.12913012	0.65939843	0.22409966	0.8344598
11	9	0.36762481	0.2887538	0.30368098	0.00469645	0.00378354	0.84104093	0.0149656	0.68569947
12	10	0.45083207	0.33738602	0.33128834	0.00722589	0.00717939	0.74941693	0.08054051	0.72702179
13	11	0.90468986	0.84954407	0.73619632	0.08058855	0.07977192	0.5673136	0.9095885	0.95266717
14	12	0.5688351	0.51215805	0.41411043	0.00683164	0.00707972	0.70781771	0.11419327	0.78562434
15	13	0.4720121	0.38601824	0.37576687	0.0085974	0.00969807	0.77137855	0.05642459	0.7375402
16	14	0.40847201	0.31914894	0.32515337	0.00453361	0.00453712	0.80041624	0.02838632	0.70598497
17	15	0.41603631	0.3662614	0.33895706	0.00617032	0.00655594	0.77554755	0.04438378	0.70974155
18	16	0.89712557	0.87537994	0.73312883	0.12938567	0.13449287	0.57113244	1	0.94891059
19	17	0.34493192	0.29483283	0.30828221	0.11062131	0.11149552	0.84263487	0.01475445	0.67468069
20	18	0.29954614	0.2325228	0.24233129	0.00062769	0.00208873	0.88153844	0.00375617	0.65189031
21	19	0.54462935	0.45896657	0.50460123	0.24088676	0.24796394	0.71949225	0.10831836	0.77385424
22	20	0.30257186	0.23860182	0.2607362	0.00614709	0.00642865	0.88289782	0.00230883	0.65339294

Figura 4.18 Valores normalizados entre 0 y 1 mismos que la figura 4.16

Una vez escalados todos los archivos, concluimos con la finalización de la fuente de datos final, la cual consta de 17 archivos (1 por cada mes) los cuales están compuestos de 9 columnas (8 indicadores más una de numeración) por 39 filas (38 bancos más una de encabezados) con valores escalados 0 y 1.

Nombre

















-  01_agosto 2004 escalado.csv
-  02_septiembre 2004 escalado.csv
-  03_octubre 2004 escalado.csv
-  04_noviembre 2004 escalado.csv
-  05_diciembre 2004 escalado.csv
-  06_enero 2005 escalado.csv
-  07_febrero 2005 escalado.csv
-  08_marzo 2005 escalado.csv
-  09_abril 2005 escalado.csv
-  10_mayo 2005 escalado.csv
-  11_junio 2005 escalado.csv
-  12_julio 2005 escalado.csv
-  13_agosto 2005 escalado.csv
-  14_septiembre 2005 escalado.csv
-  15_octubre 2005 escalado.csv
-  16_noviembre 2005 escalado.csv
-  17_diciembre 2005 escalado.csv

Figura 4.19 Meses escalados

4.4 CÁLCULO DE K

Para algunas aplicaciones, puede ser sencillo calcular el número de clusters “K” de acuerdo al contexto en el que se esté trabajando. Pero para nuestra situación, al no conocer el comportamiento de los datos, **K** es un valor desconocido.

El algoritmo de *kmeans* necesita el valor “K” como parámetro de entrada, y si deseamos tener buenos resultados, este valor está fuertemente ligado con su correcta elección.

Una clusterización con demasiados grupos complicaría los resultados, puesto que los haría difíciles de analizar o interpretar, y en contraparte, el tener pocos clusters llevaría pérdida de información lo cual nos orientaría a una mala toma de decisiones.

Según (Morales & Escalante, 2017) al problema de la determinación del número de clusters se le conoce como “Problema fundamental de la validez de un clúster”.

Algunos métodos que se han utilizado para encontrar el número adecuado del clúster se enlistan a continuación:

- Visualización del conjunto de datos, lo que funciona bien para dos dimensiones, pero generalmente nuestros conjuntos de datos son mucho más complicados.
- Construcción de índices (o reglas de paro). En este caso se utilizan índices para enfatizar la compactes intra-cluster e isolación inter-clúster considerando efectos tales como: el error cuadrático, propiedades geométricas o estadísticas de los datos, el número de patrones, la disimilaridad o similaridad, número de clúster.
- Optimización de alguna función de criterio bajo el marco del modelo de mezcla de probabilidades. En este caso se utiliza el algoritmo EM

(usualmente), para encontrar el valor de "K" que maximice o minimice el criterio definido como óptimo.

- Criterio de Información de Akaike (AIC).
- Criterio de Inferencia Bayesiana.

Para este trabajo de tesis, se utiliza el método **WCSS (within-cluster sums of squares)** el cual busca minimizar la suma de los cuadrados dentro de cada grupo. A continuación, se muestran dos ejemplos de la gráfica obtenida en diferentes meses, además del código utilizado para su ejecución, el conjunto completo de gráficas se muestra en el apartado anexos.

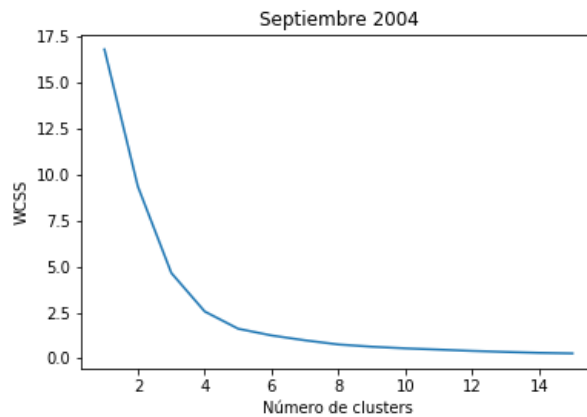


Figura 4.20 WCSS Septiembre 2004

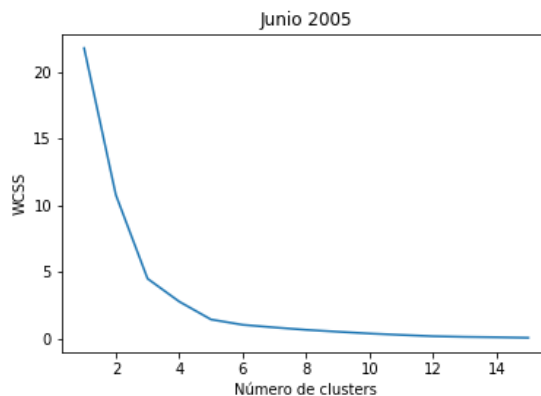


Figura 4.21 WCSS Junio 2005

```

import matplotlib.pyplot as plot
import pandas as pd
import os
from sklearn.cluster import KMeans

carpeta="resumen mensual escalado"
archivos= os.listdir(carpeta)

encabezados=["Mes", "Año", "Clusters"]

df=pd.DataFrame(columns=encabezados)
conta = 0

for ar in archivos:
    nombre=carpeta+"/"+ar
    temp = pd.read_csv(nombre)
    x=temp.drop("Unnamed: 0",1)

    wcss = []
    for i in range (1,16):
        kmeans=KMeans(n_clusters=i, init='k-means++',random_state=0)
        kmeans.fit(x)
        wcss.append(kmeans.inertia_)

    miMes=ar[:-13]
    anio=miMes[-4:]
    mes=miMes[:-5]
    #Número de clusters
    for z in range(1,14):
        resta=wcss[z-1]-wcss[z]
        if (resta<1):
            df.set_value(index=conta, col="Mes", value=mes)
            df.set_value(index=conta, col="Año", value=anio)
            df.set_value(index=conta, col="Clusters", value=z)
            conta = conta + 1
            break

    plot.plot(range(1,16),wcss)
    plot.title(mes[3:].capitalize()+" "+anio)
    plot.xlabel('Número de clusters')
    plot.ylabel('WCSS')
    plot.savefig(mes+" "+anio+".png")
    plot.show()

```

Figura 4.22 Código para calcular el valor de k en cada uno de los meses

Como se puede observar, los valores de K, a partir de que llegan a 5 tienden ya no variar tanto, por lo tanto, ese será el parámetro que se tomará.

4.5 EJECUCION KMEANS

Ya obtenido el valor de k, necesario para la aplicación del algoritmo de kmeans, se procede a la ejecución de dicho algoritmo como se muestra en el siguiente código, la clusterización obtenida se muestra en el anexo 3.

```
bancos=["Banco 1","Banco 2","Banco 3","Banco 4","Banco 5","Banco 6","Banco 7",
        "Banco 8","Banco 9","Banco 10","Banco 11","Banco 12","Banco 13",
        "Banco 14","Banco 15","Banco 16","Banco 17","Banco 18","Banco 19",
        "Banco 20","Banco 21","Banco 22","Banco 23","Banco 24","Banco 25",
        "Banco 26","Banco 27","Banco 28","Banco 29","Banco 30","Banco 31",
        "Banco 32","Banco 33","Banco 34","Banco 35","Banco 36","Banco 37",
        "Banco 38"]
encabezados=["Ago04","Sep04","Oct04","Nov04","Dic04","Ene05","Feb05","Mar05",
             "Abr05","May05","Jun05","Jul05","Ago05","Sep05","Oct05","Nov05",
             "Dic05"]

df=pd.DataFrame()
dfCentro=pd.DataFrame()

carpeta="resumen mensual escalado"
archivos= os.listdir(carpeta)

con = 0
for ar in archivos:
    nombre=carpeta+"/"+ar
    temp = pd.read_csv(nombre)
    x=temp.drop("Unnamed: 0",1)
    kmeans=KMeans(n_clusters=5, init='k-means++',random_state=0)

    y=kmeans.fit_predict(x)
    centroid=kmeans.cluster_centers_

    closest, _ = pairwise_distances_argmin_min(centroid, x)

    df.insert(loc=con, column=encabezados[con], value=y)
    dfCentro.insert(loc=con,column=encabezados[con],value=closest)

    x.insert(loc=8, column="grupo", value=y)

    matriz=df.as_matrix()

    con = con +1
```

Figura 4.23 Ejecución del algoritmo de Kmeans

CAPÍTULO 5

ANÁLISIS DE RESULTADOS

5.1 INTRODUCCIÓN

En este capítulo analizaremos a través de gráficas y tablas, los resultados obtenidos en el capítulo anterior, en el cual se ejecutó el algoritmo de *kmeans*.

5.2 Análisis mes de agosto 2004

La ejecución del algoritmo, en el mes de agosto 2004 asigna 9 bancos al clúster 0 y al clúster 4, 8 bancos al clúster 1 y al clúster 2, y 4 bancos al clúster 3, como se muestra en la siguiente gráfica.

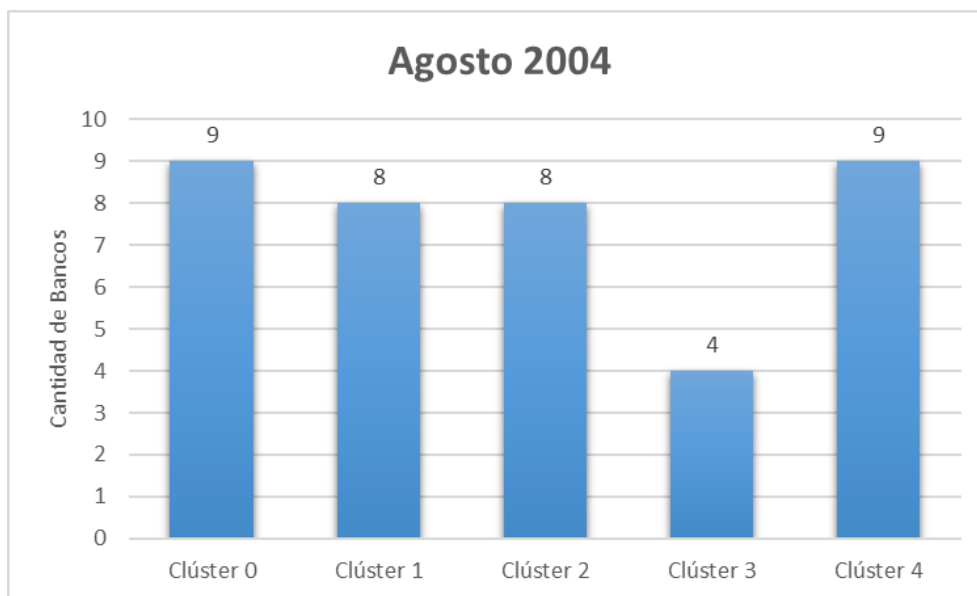


Figura 5.1 Cantidad de bancos por clúster en Agosto 2004

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	5,19,23,26,28,30,31,32,37
Clúster 1	4,7,8,11,16,21,25,35
Clúster 2	6,10,13,14,15,27,29,33
Clúster 3	0,1,2,3
Clúster 4	9,12,17,18,20,22,24,34,36

Tabla 1 Participantes por clúster en agosto 2004

La siguiente tabla muestra el análisis del centroide de cada uno de los clusters obtenidos

Mes de Agosto 2004					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 37	Banco 35	Banco 13	Banco 0	Banco 22
Degree	0.076923077	11.38461538	5.307692308	10.84615385	1.769230769
DegreeIn	0.076923077	7.076923077	2.692307692	8.230769231	1.615384615
DegreeOut	0.076923077	8.230769231	3.769230769	6.461538462	0.307692308
MontosIn	*****				
MontosOut	*****				
Betweenness	0.538461538	9.692761538	9.940361538	10.52274615	9.75
Closeness	0	40.50326923	13.28269231	25.35718462	0.230992308
Affinity	0.839769231	18.66669231	14.76292308	18.50646154	10.35256154

Tabla 2 Análisis centroide agosto 2004

Y la siguiente tabla, nos muestra las medias y desviaciones estándar por variable en cada uno del clúster.

Agosto 2004		Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Degree	Media	0.1368	11.7019	5.0865	11.3269	1.8718
	Desv. Est.	0.2012	1.0092	0.9951	1.2761	0.9669
DegreeIn	Media	0.1282	7.5	2.7404	9.4231	1.1538
	Desv. Est.	0.1986	1.6721	0.8904	1.4334	0.8741
DegreeOut	Media	0.0513	8.2404	3.5288	6.9808	0.9573
	Desv. Est.	0.0811	1.2747	0.8544	1.6733	0.6219
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	1.0641	10.0286	9.3756	10.2438	8.94
	Desv. Est.	1.4852	0.4629	1.0405	0.2849	2.3178
Betweenness	Media	0.0218	46.9382	13.7589	34.8488	3.2664
	Desv. Est.	0.056	28.3926	8.6382	7.5332	5.7443
Closeness	Media	1.1601	18.8974	14.2103	18.7676	10.3699
	Desv. Est.	1.5926	0.5849	0.9558	0.6804	2.2247

Tabla 3 Análisis medias y desviaciones estándar agosto 2004

5.3 Análisis mes de septiembre 2004

El mes de septiembre 2004 asigna 9 bancos al clúster 0, 7 bancos al clúster 1, 18 bancos al clúster 2, 1 banco al clúster 3, y 3 bancos al clúster 4 como se muestra en la siguiente gráfica.

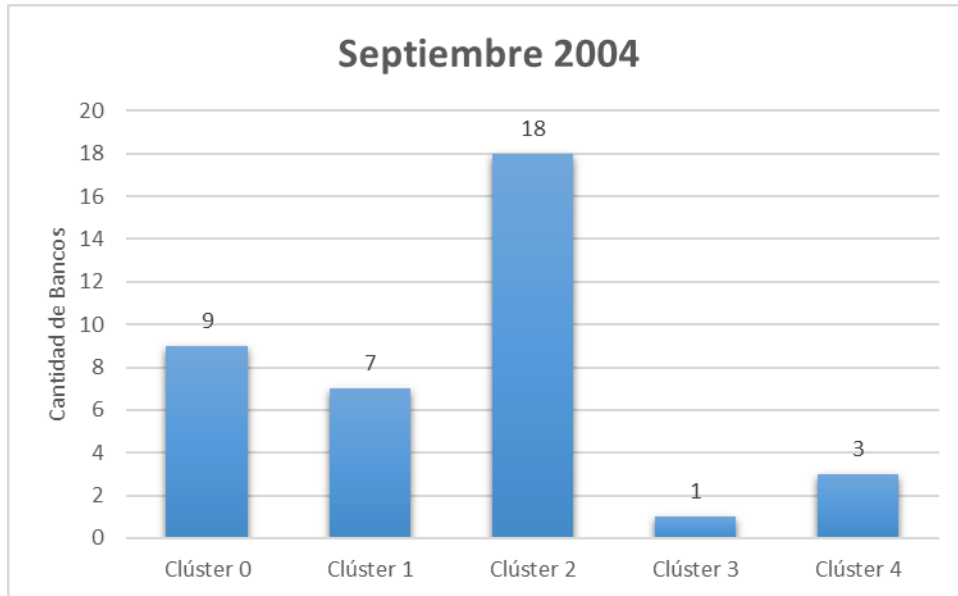


Figura 5.2 Cantidad de bancos por clúster en Septiembre 2004

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	3,4,7,8,11,16,19,25,35
Clúster 1	5,23,26,28,30,31,37
Clúster 2	6,9,10,12,13,14,15,17,18,20,22,24,27,29,32,33,34,36
Clúster 3	2
Clúster 4	0,1,21

Tabla 4 Participantes por clúster en septiembre 2004

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Septiembre 2004					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 3	Banco 28	Banco 33	Banco 2	Banco 1
Degree	12.76190476	0.80952381	6.285714286	25.04761905	16
DegreeIn	9.904761905	0.238095238	2.904761905	20.47619048	13.52380952
DegreeOut	7.952380952	0.619047619	4.571428571	20.47619048	10.42857143
MontosIn	*****				
MontosOut	*****				
Betweenness	13.16109524	3.514666667	14.08833333	9.772285714	12.25812381
Closeness	35.42088048	3.91847619	4.715270952	183.9484762	24.65178095
Affinity	21.43642857	3.214285714	17.31347619	27.74595238	23.13085714

Tabla 5 Análisis centroide septiembre 2004

Y la siguiente tabla, nos muestra las medias y desviaciones estándar por variable en cada uno de los clúster.

Septiembre 2004		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Degree	Media	13.0741	0.3061	5.1217	25.0476	16.3492
	Desv. Est.	1.9492	0.3498	2.0736	0	0.2469
DegreeIn	Media	9.3968	0.1769	2.8095	20.4762	12.5556
	Desv. Est.	2.4553	0.2048	1.4472	0	1.4372
DegreeOut	Media	8.5397	0.1701	3.3968	20.4762	11.619
	Desv. Est.	1.2627	0.2143	2.1841	0	1.0353
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	13.0199	2.5497	14.7225	9.7723	12.0417
	Desv. Est.	0.8532	3.0701	1.9786	0	0.1745
Betweenness	Media	31.1034	0.6027	6.3554	183.9485	37.8615
	Desv. Est.	15.5806	1.3576	6.6519	0	14.2724
Closeness	Media	21.5163	2.5425	16.5732	27.746	23.2962
	Desv. Est.	1.15	3.009	2.1156	0	0.117

Tabla 6 Análisis media y desviaciones estándar septiembre 2004

5.4 Análisis mes de octubre 2004

El mes de octubre 2004, asigna 12 bancos en el clúster 0, 2 bancos en el clúster 1, 5 bancos en el clúster 2, 8 bancos en el clúster 3 y 11 bancos en el clúster 4, como se muestra en la siguiente gráfica.

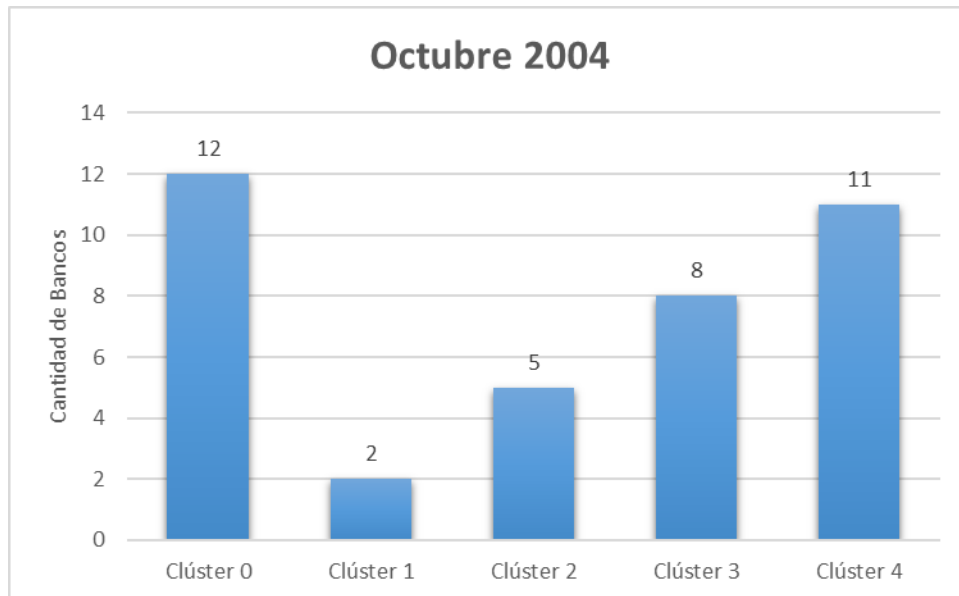


Figura 5.3 Cantidad de bancos por clúster en octubre 2004

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	6,7,8,10,12,13,14,17,25,27,32,33
Clúster 1	0,2
Clúster 2	5,26,30,31,37
Clúster 3	1,3,4,11,16,19,21,35
Clúster 4	9,15,18,20,22,23,24,28,29,34,36

Tabla 7 Participantes por clúster en octubre 2004

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Octubre 2004					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 32	Banco 2	Banco 26	Banco 3	Banco 34
Degree	12	26.85714286	0	18.66666667	4.714285714
DegreeIn	7.047619048	24.66666667	0	13.61904762	1.523809524
DegreeOut	8.952380952	21.38095238	0	14.0952381	3.714285714
MontosIn	*****				
MontosOut	*****				
Betweenness	16.111	12.37255714	0	14.065	18.37042857
Closeness	14.91528429	115.876619	0	41.78091429	0.759608571
Affinity	21.65871429	29.28571429	0	25.1427619	17.627

Tabla 8 Análisis centroide octubre 2004

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Octubre 2004		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Degree	Media	11.0278	22.8889	0.0571	18.1224	4.2684
	Desv. Est.	2.1309	2.9825	0.1143	0.669	2.1428
DegreeIn	Media	7.4524	20.7619	0.0381	14.6259	2.1429
	Desv. Est.	2.2016	2.9023	0.0762	1.2631	1.1745
DegreeOut	Media	8.2183	18.619	0.019	12.8367	3.039
	Desv. Est.	2.1015	2.2551	0.0381	2.5427	2.03
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	16.5586	13.3109	0.9429	14.7517	17.7561
	Desv. Est.	1.2745	0.7069	1.8857	0.4215	2.6877
Betweenness	Media	10.0677	68.8321	0	34.5393	1.5009
	Desv. Est.	7.8451	35.9816	0	16.4651	2.8372
Closeness	Media	21.1838	27.2778	0.7095	24.89	16.5887
	Desv. Est.	1.1542	1.5172	1.419	0.3466	2.5623

Tabla 9 Análisis medias y desviaciones estándar octubre 2004

5.5 Análisis mes de noviembre 2004

El mes de noviembre 2004 asigna 14 bancos al clúster 0, 8 bancos al clúster 1, 9 bancos al clúster 2, 3 bancos al clúster 3 y 4 bancos al clúster 4, como se muestra en la siguiente gráfica.

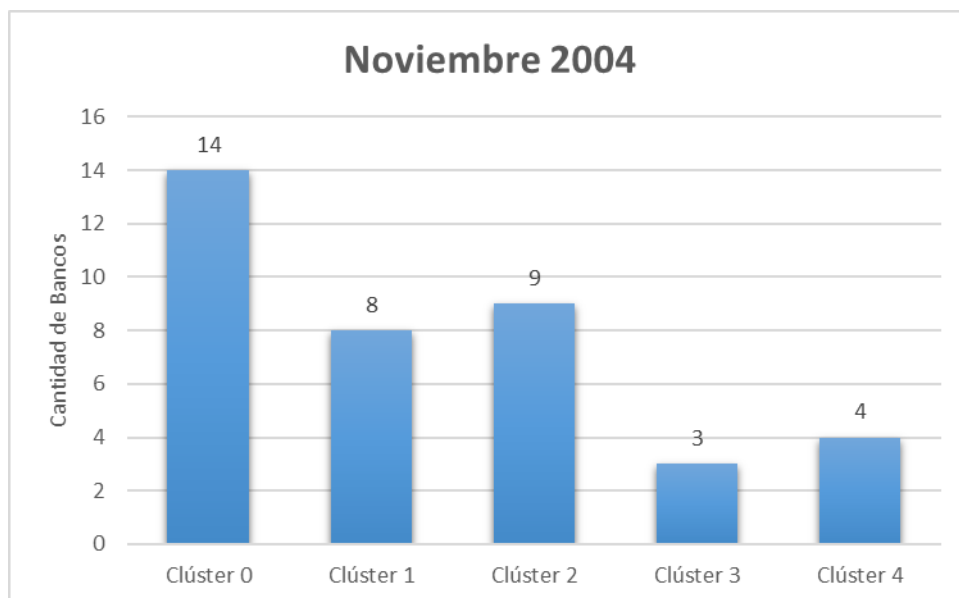


Figura 5.4 Cantidad de bancos por clúster en noviembre 2004

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	7,8,10,12,13,14,15,17,20,25,27,29,32,33
Clúster 1	3,4,6,11,16,19,21,35
Clúster 2	5,9,18,22,23,24,28,34,36
Clúster 3	0,1,2
Clúster 4	26,30,31,37

Tabla 10 Participantes por clúster en noviembre 2004

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Noviembre 2004					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 13	Banco 11	Banco 9	Banco 0	Banco 26
Degree	9.000	19.091	3.000	22.364	0.045
DegreeIn	6.182	14.955	2.136	20.455	0.045
DegreeOut	6.091	12.182	1.273	17.682	0.000
MontosIn	*****				
MontosOut	*****				
Betweenness	17.338	14.365	15.819	13.144	0.500
Closeness	12.009	44.104	0.824	63.248	0.000
Affinity	20.205	25.651	16.364	27.318	0.602

Tabla 11 Análisis centroide noviembre 2004

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Noviembre 2004		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	9.7208	17.4886	2.697	23.2121	0.1818
	Desv. Est.	2.1637	1.7173	1.4313	2.1632	0.2893
DegreIn	Media	6.1266	14.2045	1.404	21.2727	0.0114
	Desv. Est.	2.3987	1.8678	0.4921	2.5169	0.0197
DegreeOut	Media	7.6006	12.1136	1.8586	18.3939	0.1705
	Desv. Est.	1.8815	2.2714	1.3664	1.6911	0.2952
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	17.4106	14.6895	16.6397	13.0462	1.0644
	Desv. Est.	1.4523	0.8494	3.1247	0.4723	1.5682
Betweenness	Media	8.0734	42.1007	0.8281	72.47	0.009
	Desv. Est.	6.0795	29.4651	1.1745	26.1989	0.0156
Closeness	Media	20.7318	24.8125	14.664	27.7411	1.3229
	Desv. Est.	1.1345	0.8872	3.1621	1.1007	1.9591

Tabla 12 Análisis medias y desviaciones estándar noviembre 2004

5.6 Análisis mes de diciembre 2004

El mes de diciembre 2004 asigna 10 bancos al clúster 0, 3 bancos al clúster 1, 4 bancos al clúster 2, y 3, y 17 bancos al clúster 4, como se muestra en la siguiente gráfica.

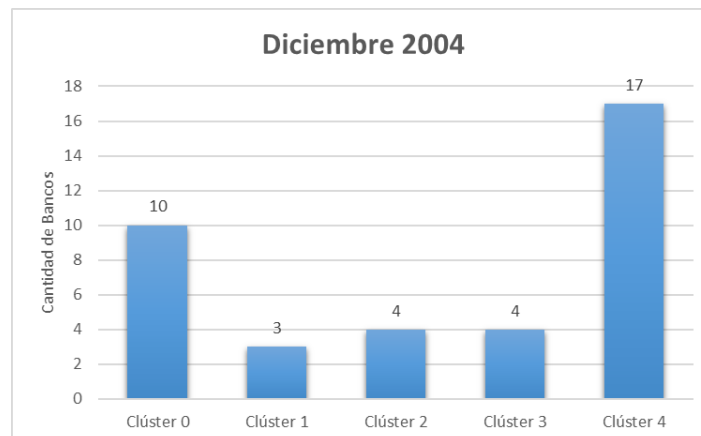


Figura 5.5 Cantidad de bancos por clúster en diciembre 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	4,6,7,8,12,17,19,21,25,27
Clúster 1	0,1,2
Clúster 2	24,26,30,31
Clúster 3	3,11,16,35
Clúster 4	5,9,10,13,14,15,18,20,22,23,28,29,32,33,34,36,37

Tabla 13 Participantes por clúster en diciembre 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Diciembre 2004					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 8	Banco 0	Banco 24	Banco 35	Banco 34
Degree	14.478	23.130	0.000	18.870	4.565
DegreeIn	10.217	21.000	0.000	14.043	2.391
DegreeOut	12.217	18.609	0.000	15.087	3.435
MontosIn	*****				
MontosOut	*****				
Betweenness	17.117	13.192	0.000	14.459	20.043
Closeness	10.753	92.661	0.000	57.726	0.225
Affinity	23.413	27.812	0.000	25.638	17.942

Tabla 14 Análisis centroide diciembre 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Tabla 15 Análisis medias y desviaciones estándar diciembre 2005

5.7 Análisis mes de enero 2005

El mes de enero 2005 asigna 3 bancos al clúster 0, 18 bancos al clúster 1, 5 bancos al clúster 2, 9 bancos al clúster 3 y 3 bancos al clúster 4, como se muestra en la siguiente gráfica.

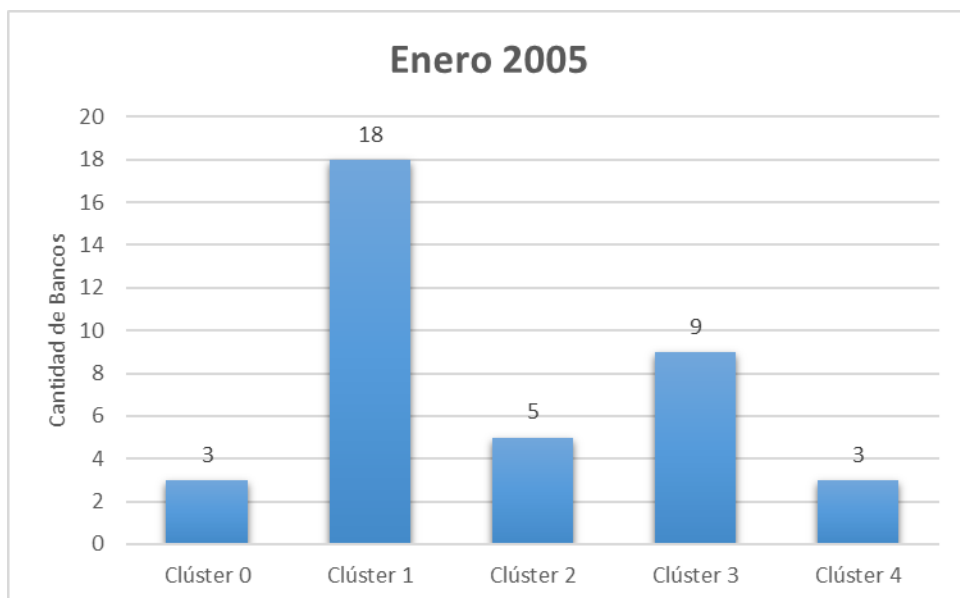


Figura 5.6 Cantidad de bancos por clúster en enero 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	0,2,3
Clúster 1	5,9,10,13,14,15,18,20,22,23,27,28,29,32,33,34,36,37
Clúster 2	1,24,26,30,31
Clúster 3	4,6,7,8,12,17,19,21,25
Clúster 4	11,16,35

Tabla 16 Participantes por clúster en enero 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Enero 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 3	Banco 36	Banco 1	Banco 8	Banco 16
Degree	24.857	5.571	0.000	14.286	20.000
DegreeIn	21.762	1.571	0.000	11.524	16.429
DegreeOut	20.095	4.476	0.000	10.381	14.714
MontosIn	*****				
MontosOut	*****				
Betweenness	12.279	18.792	0.000	16.196	13.967
Closeness	89.925	1.084	0.000	10.362	53.172
Affinity	28.365	18.440	0.000	23.008	25.921

Tabla 17 Análisis centroide enero 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Enero 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	23.8889	5.6799	0	13.9312	20.2063
	Desv. Est.	1.3358	2.5561	0	2.1086	1.5427
DegreIn	Media	21.7302	3.2275	0	11.4339	15.4127
	Desv. Est.	0.9334	1.844	0	1.8874	1.3043
DegreeOut	Media	18.6349	4.1693	0	10.4921	15.6825
	Desv. Est.	1.4181	2.1651	0	1.8488	1.2371
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	12.5058	18.0499	0	16.2478	13.4532
	Desv. Est.	0.2343	2.4343	0	1.2665	0.4533
Betweenness	Media	82.8358	2.7996	0	13.9978	62.0678
	Desv. Est.	6.1984	3.5502	0	10.4597	9.0632
Closeness	Media	27.8624	18.1304	0	22.8072	26.0198
	Desv. Est.	0.6997	1.7735	0	1.1002	0.7888

Tabla 18 Análisis medias y desviaciones estándar enero 2005

5.8 Análisis mes de febrero 2005

El mes de febrero 2005 asigna 10 bancos al clúster 0, 3 bancos al clúster 1, 5 bancos al clúster 2, 3 bancos al clúster 3 y 17 bancos al clúster 4, como se muestra en la siguiente gráfica.

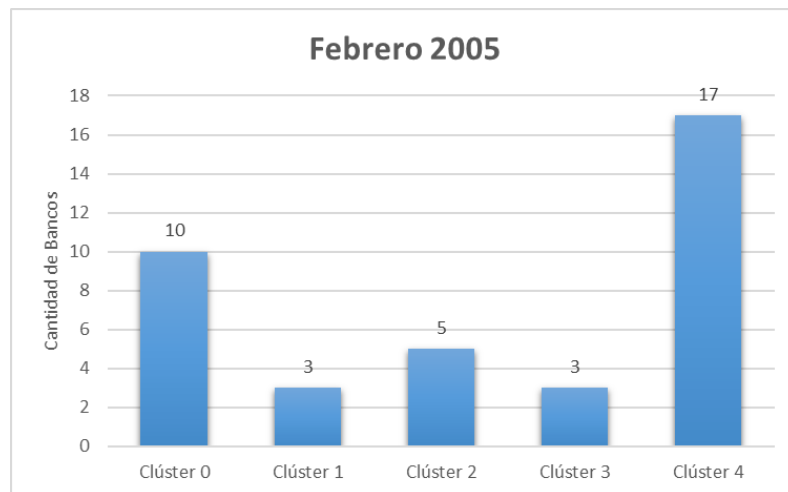


Figura 5.7 Cantidad de bancos por clúster en febrero 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	4,6,7,8,10,12,17,19,21,25
Clúster 1	0,2,3
Clúster 2	1,24,26,30,31
Clúster 3	11,16,35
Clúster 4	5,9,13,14,15,18,20,22,23,27,28,29,32,33,34,36,37

Tabla 19 Participantes por clúster en febrero 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Febrero 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 8	Banco 3	Banco 1	Banco 16	Banco 22
Degree	15.550	25.550	0.000	22.150	5.300
DegreeIn	11.500	22.750	0.000	19.350	3.650
DegreeOut	10.650	20.900	0.000	16.100	3.000
MontosIn	*****				
MontosOut	*****				
Betweenness	15.206	12.745	0.000	14.027	18.694
Closeness	15.446	85.612	0.000	49.504	5.735
Affinity	23.575	28.617	0.000	26.925	17.888

Tabla 20 Análisis centroide febrero 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Febrero 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	14.335	24.9667	0	20.8333	5.7882
	Desv. Est.	2.2031	1.0819	0	2.0778	2.7331
DegreIn	Media	11.015	22.85	0	17.0333	3.3029
	Desv. Est.	3.1163	1.0231	0	2.134	2.0603
DegreeOut	Media	10.825	18.85	0	15.9167	4.3176
	Desv. Est.	1.5702	1.6538	0	1.3331	2.1773
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	16.8229	12.939	0	13.9474	19.367
	Desv. Est.	1.1945	0.1796	0	0.6315	2.0587
Betweenness	Media	10.3595	76.9247	0	65.4689	1.6177
	Desv. Est.	5.2345	8.9223	0	18.4071	2.0784
Closeness	Media	22.9833	28.3444	0	26.2639	18.0725
	Desv. Est.	1.1141	0.5327	0	1.0307	2.0014

Tabla 21 Análisis de medias y desviaciones estándar febrero 2005

5.9 Análisis mes de marzo 2005

El mes de marzo 2005 asigna 9 bancos al clúster 0, 3 bancos al clúster 1, 5 bancos al clúster 2, 3 bancos al clúster 3 y 18 bancos al clúster 4, como se muestra en la siguiente gráfica.

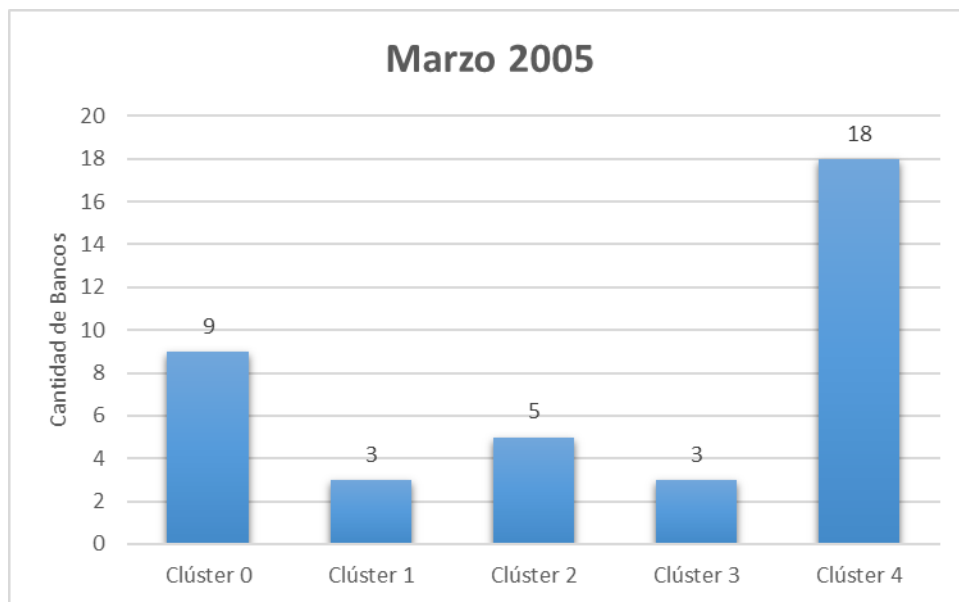


Figura 5.8 Cantidad de bancos por clúster en marzo 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	6,7,8,10,12,13,19,21,35
Clúster 1	0,2,3
Clúster 2	1,24,26,30,31
Clúster 3	4,11,16
Clúster 4	5,9,14,15,17,18,20,22,23,25,27,28,29,32,33,34,36,37

Tabla 22 Participantes por clúster en marzo 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Marzo 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 6	Banco 2	Banco 24	Banco 11	Banco 18
Degree	18.950	30.750	0.000	26.850	8.050
DegreeIn	16.100	30.200	0.000	24.100	5.750
DegreeOut	14.450	29.750	0.000	23.100	6.500
MontosIn	*****				
MontosOut	*****				
Betweenness	18.391	14.630	0.000	15.654	24.813
Closeness	10.861	38.663	0.000	43.488	0.110
Affinity	25.517	31.425	0.000	29.458	20.008

Tabla 23 Análisis centroide marzo 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Marzo 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	17.6556	30.0167	0.03	27.1167	8.5278
	Desv. Est.	2.3237	0.8663	0.06	0.2248	2.4917
DegreIn	Media	14.4889	29.4	0.03	25.5167	6.1167
	Desv. Est.	1.9841	1.0271	0.06	1.0119	2.4513
DegreeOut	Media	14.0722	27.9	0	22.7	7.0528
	Desv. Est.	2.2988	1.5105	0	0.432	2.0444
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	19.3759	14.7782	0.63	15.6798	24.3336
	Desv. Est.	1.2237	0.246	1.26	0.0534	1.7526
Betweenness	Media	9.3841	40.1928	0	38.9698	1.2936
	Desv. Est.	6.4422	12.5748	0	11.2925	1.9844
Closeness	Media	24.8602	31.0583	0.45	29.6	20.2092
	Desv. Est.	1.1691	0.4332	0.9	0.1157	1.4039

Tabla 24 Análisis de medias y desviaciones estándar marzo 2005

5.10 Análisis mes de abril 2005

El mes de abril 2005 asigna 10 bancos al clúster 0, 3 bancos al clúster 1, 5 bancos al clúster 2, 3 bancos al clúster 3 y 17 bancos al clúster 4, como se muestra en la siguiente gráfica.

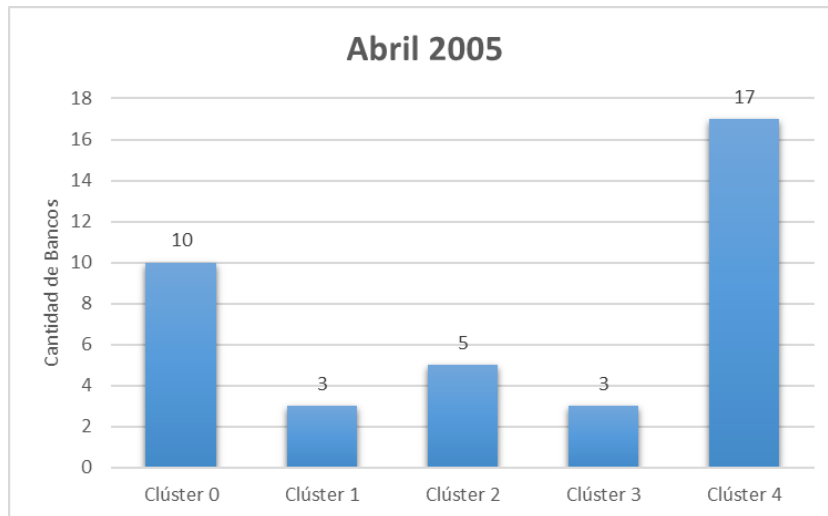


Figura 5.9 Cantidad de bancos por clúster en abril 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	6,7,8,10,12,13,19,21,32,35
Clúster 1	0,2,3
Clúster 2	1,24,26,30,31
Clúster 3	4,11,16
Clúster 4	5,9,14,15,17,18,20,22,23,25,27,28,29,33,34,36,37

Tabla 25 Participantes por clúster en abril 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Abril 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 6	Banco 2	Banco 24	Banco 11	Banco 18
Degree	19.333	31.381	0.000	27.333	9.143
DegreeIn	16.190	30.714	0.000	25.429	6.667
DegreeOut	15.667	31.143	0.000	23.286	7.000
MontosIn	*****				
MontosOut	*****				
Betweenness	19.628	15.233	0.000	16.516	25.289
Closeness	5.499	29.946	0.000	35.680	0.423
Affinity	25.667	31.690	0.000	29.643	20.556

Tabla 26 Análisis centroide abril 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Abril 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	17.9667	30.9048	0.019	28.1746	9.028
	Desv. Est.	2.3347	0.7074	0.0381	0.7002	2.7578
DegreIn	Media	14.6143	30.3333	0.019	26.9048	6.577
	Desv. Est.	2.4658	1.0087	0.0381	1.0838	2.6943
DegreeOut	Media	14.2619	29.6349	0	23.6349	7.4902
	Desv. Est.	2.1802	1.5378	0	1.0784	2.2631
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	20.1804	15.3809	0.4286	16.276	24.9194
	Desv. Est.	1.1999	0.2728	0.8571	0.2015	1.6934
Betweenness	Media	7.8926	29.1886	0	35.417	0.6024
	Desv. Est.	6.0249	5.763	0	9.404	0.5564
Closeness	Media	24.9833	31.4524	0.2905	30.0714	20.4117
	Desv. Est.	1.1673	0.3537	0.581	0.3499	1.6498

Tabla 27 Análisis de medias y desviaciones estándar abril 2005

5.11 Análisis mes de mayo 2005

El mes de mayo 2005 asigna 3 bancos al clúster 0, 14 bancos al clúster 1, 5 bancos al clúster 2, 4 bancos al clúster 3 y 10 bancos al clúster 4, como se muestra en la siguiente gráfica.

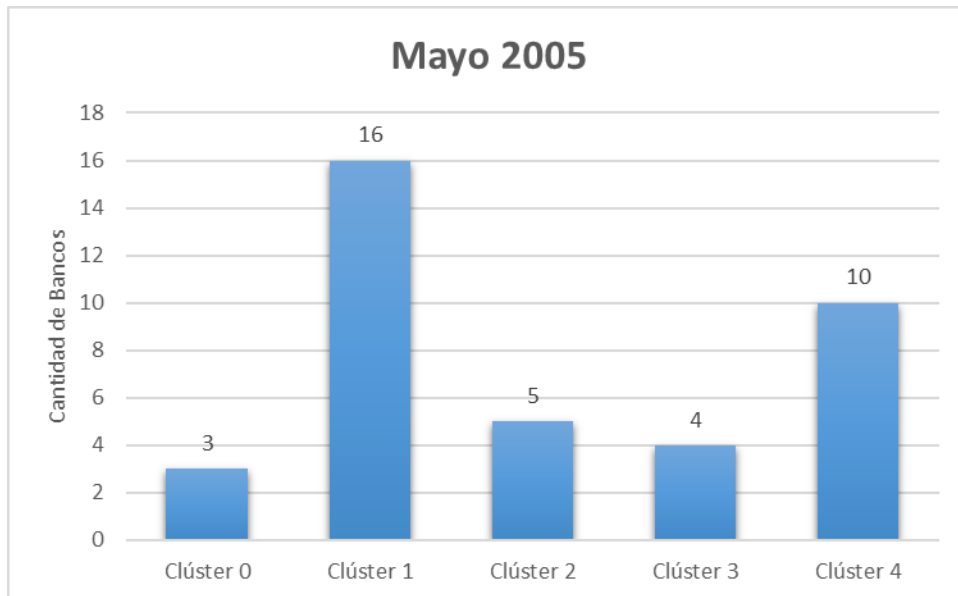


Figura 5.10 Cantidad de bancos por clúster en mayo 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	0,2,3
Clúster 1	5,9,14,17,18,20,22,23,25,27,28,29,33,34,36,37
Clúster 2	1,24,26,30,31
Clúster 3	4,11,16,35
Clúster 4	6,7,8,10,12,13,15,19,21,32

Tabla 28 Participantes por clúster en mayo 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Mayo 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 3	Banco 22	Banco 24	Banco 11	Banco 6
Degree	29.682	9.091	0.000	28.545	18.273
DegreeIn	28.727	6.182	0.000	26.955	15.727
DegreeOut	27.318	7.682	0.000	23.318	14.773
MontosIn	*****				
MontosOut	*****				
Betweenness	16.029	24.837	0.000	16.328	20.204
Closeness	16.661	0.621	0.000	39.015	4.751
Affinity	30.864	20.568	0.000	30.295	25.159

Tabla 29 Análisis centroide mayo 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Mayo 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	31	8.9915	0.0091	26.9091	17.0864
	Desv. Est.	0.9323	2.549	0.0182	2.8408	2.1201
DegreIn	Media	30.4091	6.4403	0.0091	24.4659	14.2318
	Desv. Est.	1.2009	2.3491	0.0182	4.7605	2.1462
DegreeOut	Media	29.6061	7.4744	0	22.3977	13.65
	Desv. Est.	1.6731	2.1802	0	2.252	1.8984
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	15.5062	25.3612	0.1818	16.7207	20.7784
	Desv. Est.	0.3695	2.0982	0.3636	0.6887	1.0547
Betweenness	Media	29.7399	0.5445	0	32.2617	5.2339
	Desv. Est.	13.3766	0.4425	0	9.8969	2.7575
Closeness	Media	31.5227	20.5123	0.1348	29.4697	24.5652
	Desv. Est.	0.4661	1.2777	0.2697	1.4199	1.0612

Tabla 30 Análisis de medias y desviaciones estándar mayo 2005

5.12 Análisis mes de junio 2005

El mes de junio 2005 asigna 3 bancos al clúster 0, 16 bancos al clúster 1, 5 bancos al clúster 2, 4 bancos al clúster 3 y 10 bancos al clúster 4, como se muestra en la siguiente gráfica.

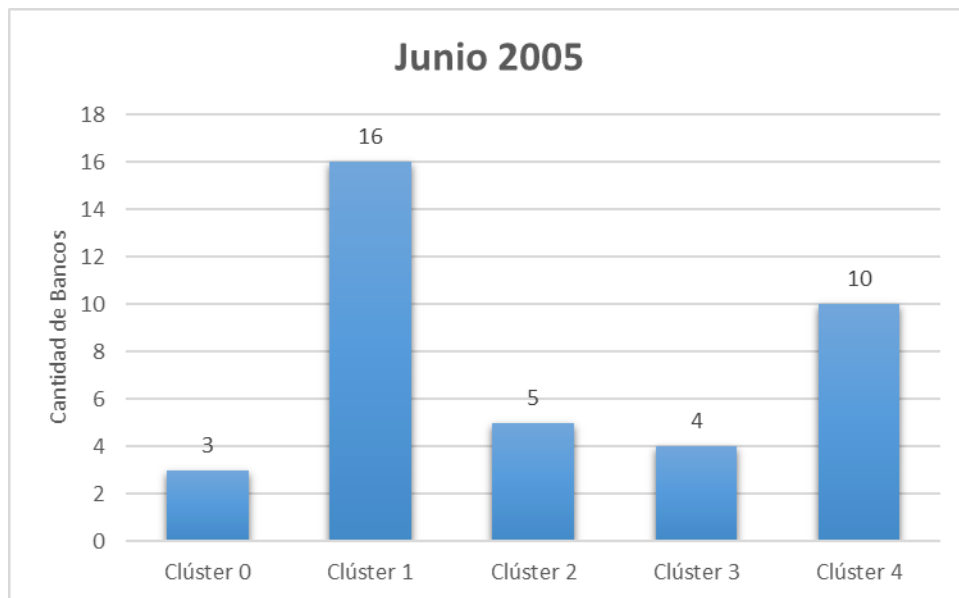


Figura 5.11 Cantidad de bancos por clúster en junio 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	0,2,3
Clúster 1	5,9,14,17,18,20,22,23,25,27,28,29,33,34,36,37
Clúster 2	1,24,26,30,31
Clúster 3	4,11,16,35
Clúster 4	6,7,8,10,12,13,15,19,21,32

Tabla 31 Participantes por clúster en junio 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Junio 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 2	Banco 18	Banco 1	Banco 11	Banco 12
Degree	31.364	9.000	0.000	27.545	17.136
DegreeIn	31.000	6.455	0.000	25.500	15.727
DegreeOut	30.864	7.364	0.000	23.182	12.409
MontosIn	*****				
MontosOut	*****				
Betweenness	15.419	25.114	0.000	16.632	20.748
Closeness	40.505	0.296	0.000	36.696	4.541
Affinity	31.614	20.432	0.000	29.705	24.500

Tabla 32 Análisis centroide junio 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Junio 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	30.7273	9.071	0	26.6591	17.1773
	Desv. Est.	1.0312	2.9359	0	2.5098	2.596
DegreIn	Media	30.2424	6.4801	0	24.0114	14.3
	Desv. Est.	1.3368	2.593	0	4.8476	2.5481
DegreeOut	Media	29.2121	7.4347	0	22.7614	13.5818
	Desv. Est.	1.9329	2.4024	0	2.5376	2.2011
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	15.645	24.7143	0	16.8329	20.6196
	Desv. Est.	0.4139	1.4723	0	0.6332	1.0779
Betweenness	Media	31.4632	0.6156	0	31.6523	5.1576
	Desv. Est.	9.8468	0.5578	0	8.7708	2.3644
Closeness	Media	31.2955	20.3314	0	29.2614	24.5205
	Desv. Est.	0.5156	1.86	0	1.2549	1.298

Tabla 33 Análisis de medias y desviaciones estándar junio 2005

5.13 Análisis mes de julio 2005

El mes de julio 2005 asigna 3 bancos al clúster 0, 16 bancos al clúster 1, 5 bancos al clúster 2, 4 bancos al clúster 3 y 10 bancos al clúster 4, como se muestra en la siguiente gráfica.

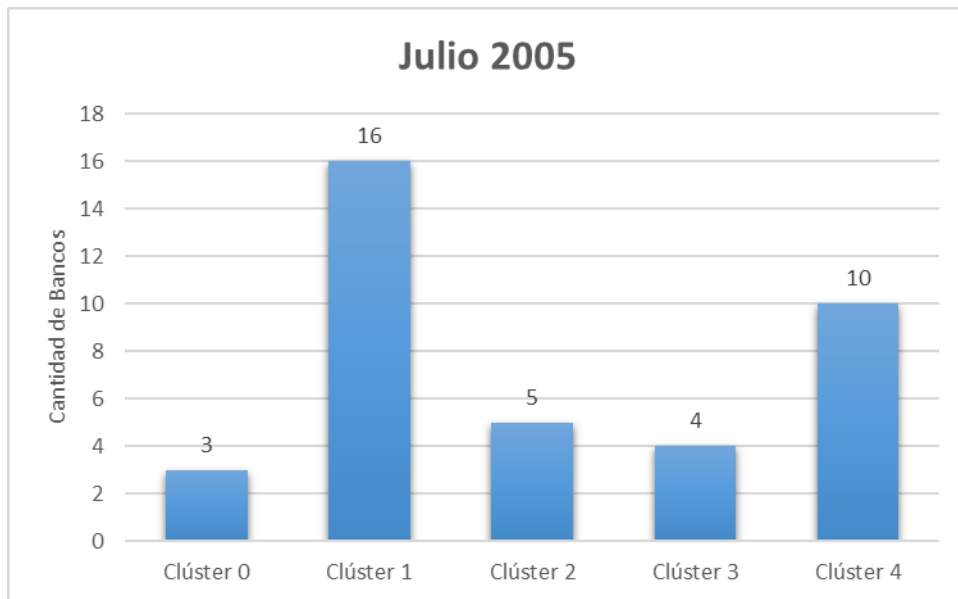


Figura 5.12 Cantidad de bancos por clúster en julio 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	0,2,3
Clúster 1	5,9,14,17,18,20,22,23,25,27,28,29,33,34,36,37
Clúster 2	1,24,26,30,31
Clúster 3	4,11,16,35
Clúster 4	6,7,8,10,12,13,15,19,21,32

Tabla 34 Participantes por clúster en julio 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Julio 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 2	Banco 18	Banco 1	Banco 11	Banco 12
Degree	31.381	8.762	0.000	28.048	17.429
DegreeIn	30.905	6.714	0.000	26.286	15.048
DegreeOut	30.810	7.238	0.000	23.524	13.381
MontosIn	*****				
MontosOut	*****				
Betweenness	15.067	25.713	0.000	16.125	20.638
Closeness	35.877	0.235	0.000	37.948	4.627
Affinity	31.667	20.357	0.000	30.000	24.690

Tabla 35 Análisis centroide julio 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Julio 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	30.9365	8.6696	0	26.9643	16.5571
	Desv. Est.	0.7672	2.528	0	2.4528	2.5877
DegreIn	Media	30.4286	6.3423	0	24.4167	13.6333
	Desv. Est.	0.9556	2.4091	0	4.4707	2.4117
DegreeOut	Media	29.5079	7.244	0	22.869	13.0857
	Desv. Est.	1.7407	2.1947	0	2.3967	2.2829
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	15.2025	25.5422	0	16.4304	20.7043
	Desv. Est.	0.2978	1.7374	0	0.6478	1.2022
Betweenness	Media	32.5834	0.3612	0	32.5316	5.1771
	Desv. Est.	8.78	0.3734	0	9.8219	2.7456
Closeness	Media	31.4444	20.2589	0	29.4524	24.2524
	Desv. Est.	0.3836	1.4104	0	1.2216	1.2937

Tabla 36 Análisis de medias y desviaciones estándar julio 2005

5.14 Análisis mes de agosto 2005

El mes de agosto 2005 asigna 20 bancos al clúster 0, 3 bancos al clúster 1, 5 bancos al clúster 2, 7 bancos al clúster 3 y 3 bancos al clúster 4, como se muestra en la siguiente gráfica.

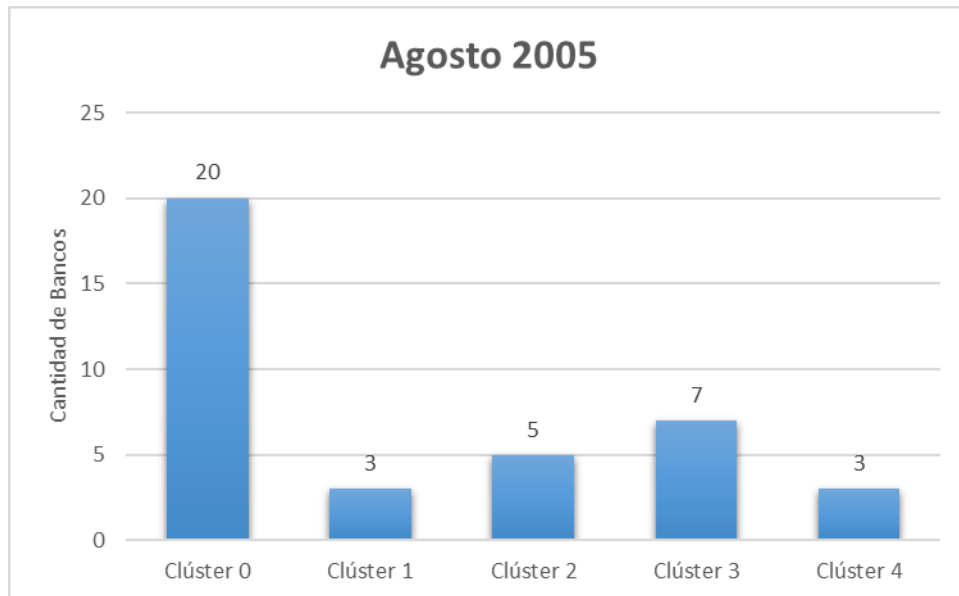


Figura 5.13 Cantidad de bancos por clúster en agosto 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	5,9,10,13,14,15,17,18,20,22,23,25,27,28,29,,32,33,34,36,37
Clúster 1	0,2,3
Clúster 2	1,24,26,30,31
Clúster 3	6,7,8,12,19,21,35
Clúster 4	4,11,16

Tabla 37 Participantes por clúster en agosto 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Agosto 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 18	Banco 2	Banco 1	Banco 8	Banco 11
Degree	9.435	31.478	0.000	21.391	28.348
DegreeIn	7.043	30.652	0.000	17.391	26.783
DegreeOut	6.957	31.174	0.000	16.087	23.565
MontosIn	*****				
MontosOut	*****				
Betweenness	25.099	15.186	0.000	18.656	16.271
Closeness	0.304	35.425	0.000	11.446	39.506
Affinity	20.717	31.739	0.000	26.696	30.130

Tabla 38 Análisis centroide agosto 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Agosto 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	9.7652	31.058	0	19.4969	28.4928
	Desv. Est.	3.0694	0.6887	0	1.8172	0.3355
DegreIn	Media	7.3065	30.3768	0	16.2422	27.058
	Desv. Est.	2.961	0.8567	0	1.1053	0.3019
DegreeOut	Media	7.9348	29.7681	0	15.8385	24.4203
	Desv. Est.	2.3942	1.6331	0	1.6003	1.0886
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	25.0578	15.3216	0	19.5205	16.1781
	Desv. Est.	2.3797	0.2734	0	1.0402	0.1289
Betweenness	Media	0.8187	29.2069	0	9.8643	35.3596
	Desv. Est.	0.9902	8.4193	0	6.7832	10.344
Closeness	Media	20.8804	31.529	0	25.7484	30.2319
	Desv. Est.	1.5382	0.3443	0	0.9086	0.1751

Tabla 39 Análisis de medias y desviaciones estándar agosto 2005

5.15 Análisis mes de septiembre 2005

El mes de septiembre 2005 asigna 3 bancos al clúster 0, 20 bancos al clúster 1, 5 bancos al clúster 2, 7 bancos al clúster 3 y 3 bancos al clúster 4, como se muestra en la siguiente gráfica.

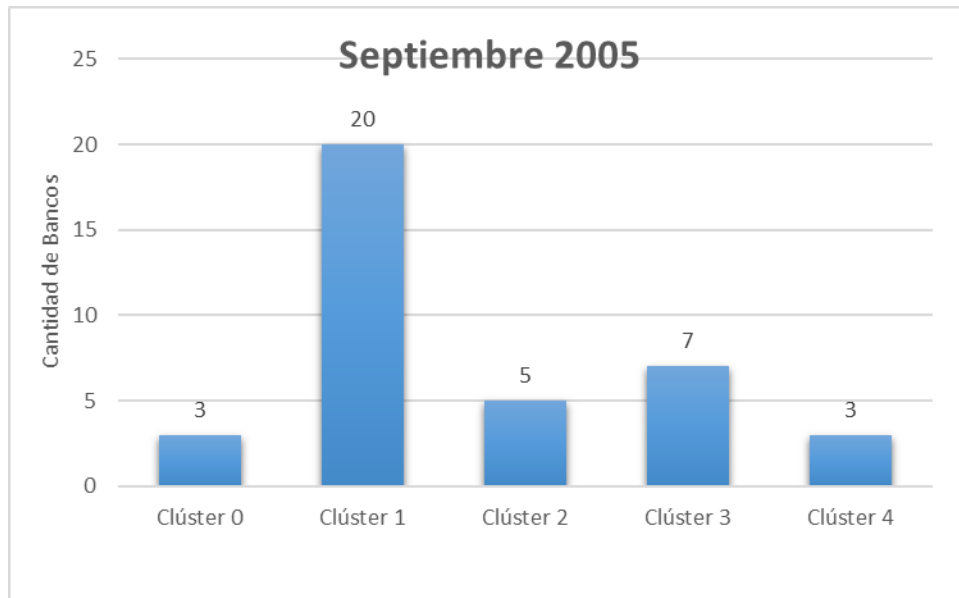


Figura 5.14 Cantidad de bancos por clúster en septiembre 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	0,2,3
Clúster 1	5,9,10,13,14,15,17,18,20,22,23,25,27,28,29,32,33,34,36,37
Clúster 2	1,24,26,30,31
Clúster 3	6,7,8,12,19,21,35
Clúster 4	4,11,16

Tabla 40 Participantes por clúster en septiembre 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Septiembre 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 2	Banco 18	Banco 1	Banco 21	Banco 11
Degree	31.190	10.143	0.000	18.143	28.190
DegreeIn	30.714	7.714	0.000	15.857	26.905
DegreeOut	30.810	7.524	0.000	14.714	23.238
MontosIn	*****				
MontosOut	*****				
Betweenness	15.288	24.729	0.000	20.260	16.258
Closeness	28.397	0.506	0.000	9.051	39.858
Affinity	31.524	21.000	0.000	25.000	30.000

Tabla 41 Análisis centroide septiembre 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Septiembre 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	30.8254	9.8119	0	19.3946	28.4127
	Desv. Est.	0.7642	3.0009	0	1.7822	0.3143
DegreIn	Media	30.3651	7.2881	0	16.1837	27.2857
	Desv. Est.	0.8159	2.9511	0	1.4817	0.2722
DegreeOut	Media	29.5556	8.0262	0	15.7483	24.1905
	Desv. Est.	1.512	2.3635	0	1.4477	1.6237
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	15.3849	24.5241	0	19.5309	16.1843
	Desv. Est.	0.2894	1.6918	0	0.975	0.1129
Betweenness	Media	30.6247	0.8396	0	9.3994	33.8157
	Desv. Est.	8.8404	0.8568	0	5.9158	12.8512
Closeness	Media	31.3413	20.7214	0	25.6224	30.127
	Desv. Est.	0.3821	1.8575	0	0.8878	0.163

Tabla 42 Análisis de medias y desviaciones estándar septiembre 2005

5.16 Análisis mes de octubre 2005

El mes de octubre 2005 asigna 3 bancos al clúster 0, 19 bancos al clúster 1, 5 bancos al clúster 2, 4 bancos al clúster 3 y 7 bancos al clúster 4, como se muestra en la siguiente gráfica.

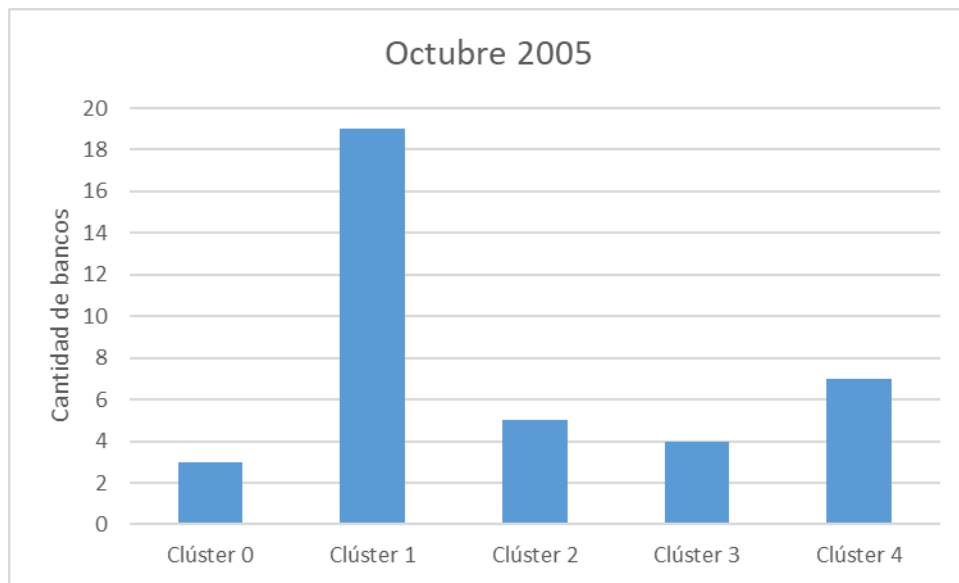


Figura 5.15 Cantidad de bancos por clúster en octubre 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	0,2,3
Clúster 1	5,9,10,14,15,17,18,20,22,23,25,27,28,29,32,33,34,36,37
Clúster 2	1,24,26,30,31
Clúster 3	4,11,16,35
Clúster 4	6,7,8,12,13,19,21

Tabla 43 Participantes por clúster en octubre 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Octubre 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 0	Banco 20	Banco 24	Banco 11	Banco 6
Degree	31.429	9.524	0.000	28.476	19.143
DegreeIn	31.333	7.476	0.000	26.619	17.000
DegreeOut	30.238	8.095	0.000	22.857	15.524
MontosIn	*****				
MontosOut	*****				
Betweenness	15.329	25.454	0.000	16.355	19.863
Closeness	30.033	0.103	0.000	40.564	5.944
Affinity	31.667	20.706	0.000	30.190	25.524

Tabla 44 Análisis centroide octubre 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Octubre 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	30.9683	9.7619	0.0095	27	18.4558
	Desv. Est.	0.6849	2.877	0.019	2.8534	2.0176
DegreIn	Media	30.4603	7.0476	0.0095	24.4643	15.8435
	Desv. Est.	0.9479	2.7876	0.019	4.9632	1.8873
DegreeOut	Media	29.4921	8	0	22.4762	14.8163
	Desv. Est.	1.6607	2.3088	0	2.8148	1.6765
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	15.5289	24.7086	0.1143	16.719	20.2456
	Desv. Est.	0.2601	1.772	0.2286	0.7124	1.0199
Betweenness	Media	25.1046	1.1693	0	32.9788	6.3649
	Desv. Est.	8.7335	1.7835	0	10.3535	2.3463
Closeness	Media	31.4365	20.7072	0.1286	29.4524	25.1757
	Desv. Est.	0.3425	1.6995	0.2571	1.4267	1.0083

Tabla 45 Análisis de medias y desviaciones estándar octubre 2005

5.17 Análisis mes de noviembre 2005

El mes de noviembre 2005 asigna 3 bancos al clúster 0, 17 bancos al clúster 1, 5 bancos al clúster 2, 10 bancos al clúster 3 y 3 bancos al clúster 4, como se muestra en la siguiente gráfica.

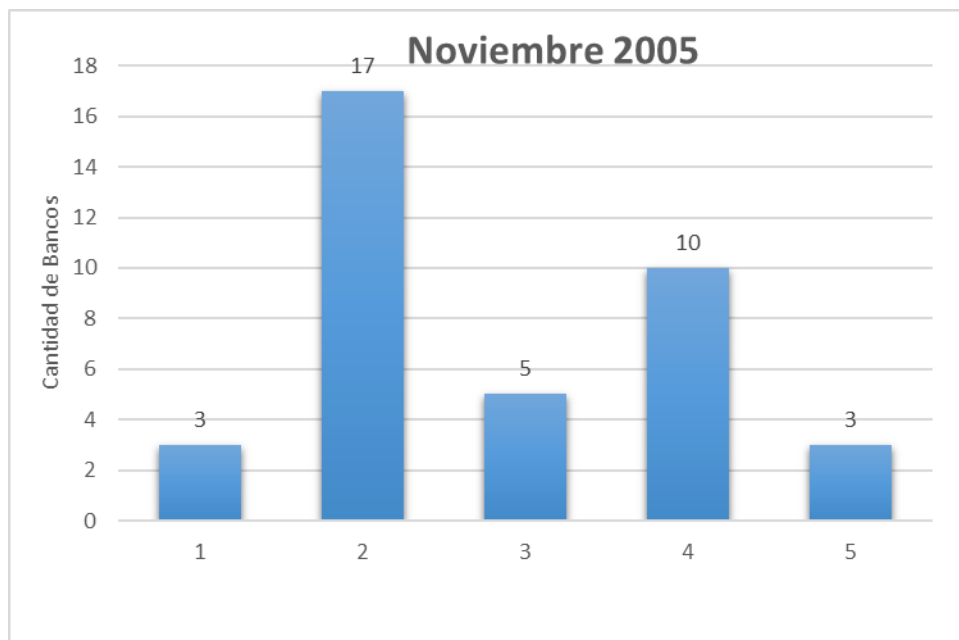


Figura 5.16 Cantidad de bancos por clúster en noviembre 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	0,2,3
Clúster 1	5,9,14,17,18,20,22,23,25,27,28,29,32,33,34,36,37
Clúster 2	1,24,26,30,31
Clúster 3	6,7,8,10,12,13,15,19,21,35
Clúster 4	4,11,16

Tabla 46 Participantes por clúster en noviembre 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Noviembre 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 2	Banco 20	Banco 1	Banco 6	Banco 11
Degree	31.1818	8.6818	0.0000	18.9545	28.1364
DegreeIn	30.8636	6.6364	0.0000	16.0909	27.0000
DegreeOut	30.6364	7.2727	0.0000	16.0909	23.8182
MontosIn	*****				
MontosOut	*****				
Betweenness	15.2923	25.5687	0.0000	19.7684	16.2592
Closeness	24.4591	0.1862	0.0000	4.5324	38.4418
Affinity	31.5682	20.3182	0.0000	25.4545	30.0455

Tabla 47 Análisis centroide noviembre 2005

Finalmente se muestra el análisis de media y desviación estándar de cada uno de los clúster.

Noviembre 2005		Cluster	Cluster	Cluster	Cluster	Cluster
		0	1	2	3	4
Degree	Media	30.803	9.0909	0	17.7045	28.3788
	Desv. Est.	0.6346	2.6854	0	2.4455	0.3428
DegreIn	Media	30.2576	6.4733	0	14.4727	27.2424
	Desv. Est.	0.8894	2.6485	0	2.2499	0.1754
DegreeOut	Media	29.3333	7.5187	0	13.9682	23.9242
	Desv. Est.	1.8107	2.2622	0	2.1648	1.6162
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	15.3983	24.963	0	20.2086	16.1874
	Desv. Est.	0.2635	1.8236	0	1.2053	0.1024
Betweenness	Media	33.2756	0.6124	0	7.2792	35.0083
	Desv. Est.	17.0637	0.5848	0	5.9892	10.7176
Closeness	Media	31.3636	20.4715	0	24.8273	30.1667
	Desv. Est.	0.3219	1.4814	0	1.2195	0.1714

Tabla 48 Análisis de medias y desviaciones estándar noviembre 2005

5.18 Análisis mes de diciembre 2005

El mes de diciembre 2005 asigna 11 bancos al clúster 0, 3 bancos al clúster 1, 5 bancos al clúster 2, 4 bancos al clúster 3 y 15 bancos al clúster 4, como se muestra en la siguiente gráfica.

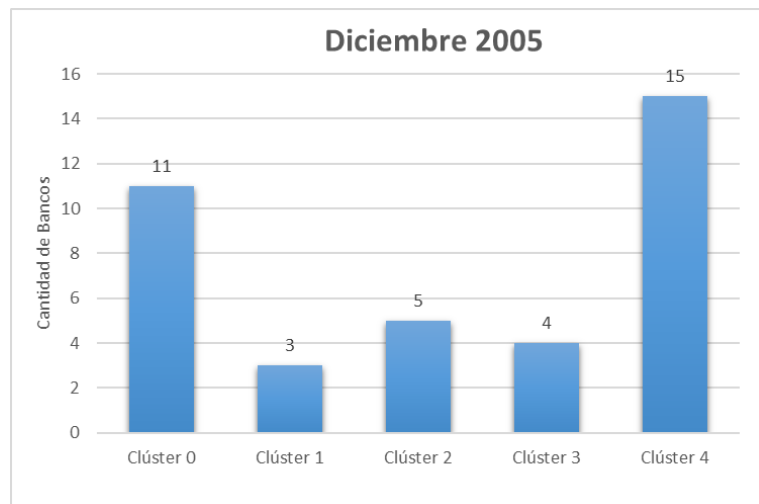


Figura 5.17 Cantidad de bancos por clúster en diciembre 2005

Los bancos que se encuentran dentro del clúster son:

Participantes en clúster	
Clúster 0	6,7,8,10,12,13,14,15,19,21,29
Clúster 1	0,2,3
Clúster 2	1,24,26,30,31
Clúster 3	4,11,16,35
Clúster 4	5,9,17,18,20,22,23,25,27,28,32,33,34,36,37

Tabla 49 Participantes por clúster en diciembre 2005

La siguiente tabla muestra el análisis del centroide de cada uno de los clúster obtenidos

Mes de Diciembre 2005					
Clúster	Clúster 0	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Centroides	Banco 12	Banco 3	Banco 1	Banco 11	Banco 27
Degree	16.9524	30.0476	0.0000	28.0000	8.3333
DegreeIn	15.4286	29.3810	1.0000	26.6190	5.7143
DegreeOut	11.8571	27.2857	0.0000	23.0476	7.0952
MontosIn	*****				
MontosOut	*****				
Betweenness	20.7806	15.8272	0.0000	16.4992	24.5488
Closeness	4.0656	25.0084	0.0000	36.8486	0.3955
Affinity	24.4762	31.0238	0.0000	30.0000	20.1667

Tabla 50 Análisis centroide diciembre 2005

Diciembre 2005		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Degree	Media	16.619	31	0	26.6905	8.7873
	Desv. Est.	2.7718	0.7009	0	2.6822	2.4607
DegreeIn	Media	13.8442	30.5873	0	24.119	6.0254
	Desv. Est.	2.9865	0.908	0	4.9244	2.2919
DegreeOut	Media	12.987	29.5079	0	22.4167	7.3238
	Desv. Est.	2.1732	1.5733	0	2.7728	2.2295
MontosIn	Media	*****				
	Desv. Est.	*****				
MontosOut	Media	*****				
	Desv. Est.	*****				
Affinity	Media	20.8407	15.469	0	16.7369	25.2659
	Desv. Est.	1.2746	0.2791	0	0.538	1.8153
Betweenness	Media	5.0455	33.0813	0	31.5951	0.5225
	Desv. Est.	2.833	16.0771	0	8.2325	0.4488
Closeness	Media	24.3074	31.5	0	29.3452	20.3868
	Desv. Est.	1.3825	0.3505	0	1.3411	1.2356

Tabla 51 Análisis de medias y desviaciones estándar diciembre 2005

5.19 Reacomodo de los clúster

Al hacer análisis clúster con *kmeans*, los grupos pueden moverse mes con mes, siendo estos los mismos, pero con diferente nombre, por ejemplo el **clúster 0**, al siguiente mes, el algoritmo lo toma como el **clúster 4**, y posteriormente lo tomará como el **clúster 3**, etc. esto es debido a el comportamiento aleatorio interno del algoritmo. Para fines de poder identificar el comportamiento de los bancos a través del tiempo, los clúster se reacomodaron de la siguiente manera, con base en los resultados previamente mostrados:

MES	NOMBRE DEL CLUSTER									
	ORIGINAL	NUEVO	ORIGINAL	NUEVO	ORIGINAL	NUEVO	ORIGINAL	NUEVO	ORIGINAL	NUEVO
ago-04	0	A	1	E	2	B	3	D	4	C
sep-04	0	E	1	A	2	B	3	D	4	C
oct-04	0	C	1	D	2	A	3	E	4	B
nov-04	0	C	1	E	2	B	3	D	4	A
dic-04	0	C	1	D	2	A	3	E	4	B
ene-05	0	D	1	B	2	A	3	C	4	E
feb-05	0	C	1	D	2	A	3	E	4	B
mar-05	0	C	1	D	2	A	3	E	4	B
abr-05	0	C	1	D	2	A	3	E	4	B
may-05	0	D	1	B	2	A	3	E	4	C
jun-05	0	D	1	B	2	A	3	E	4	C
jul-05	0	D	1	B	2	A	3	E	4	C
ago-05	0	B	1	D	2	A	3	C	4	E
sep-05	0	D	1	B	2	A	3	C	4	E
oct-05	0	D	1	B	2	A	3	E	4	C
nov-05	0	D	1	B	2	A	3	C	4	E
dic-05	0	C	1	D	2	A	3	E	4	B

Tabla 52 Reacomodo de los clúster

5.20 Análisis de nuevos clusters

A continuación se muestra el análisis de los clúster **A,B,C,D,E**, en los 17 meses.

Clúster A: Este clúster se forma solamente por los bancos 5,19,23,26,28,30,31,32,37, en las frecuencias que se muestran a continuación.

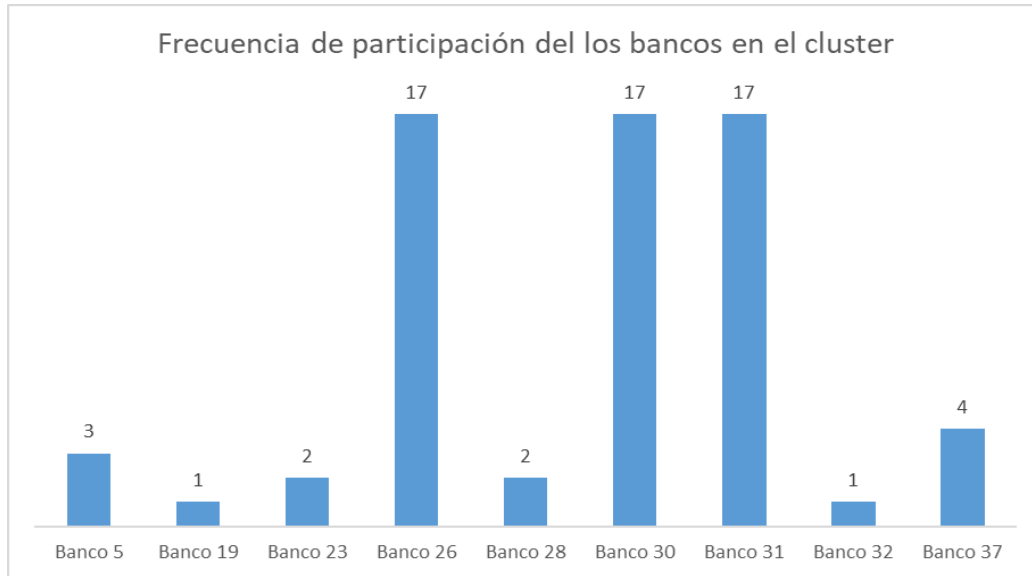


Figura 5.18 Frecuencia de participación de los bancos en el clúster A

Se observa que los bancos 26,30,31 estuvieron siempre en ese clúster, y hubo varios bancos que transitaron en cuatro o menos ocasiones. Por lo regular, el centroide de este clúster se ubica en un banco con inactividad en ese mes, además de aportar muy poco a la red.

Clúster B: Siendo el clúster más activo, contando con la participación del 50% de los bancos, el clúster B es un clúster cuyas características principales son que por lo regular su centroide, tiene un *betweenness* mayor que el de todos los centroides, los primeros 7 meses de operación, sus bancos manejaron montos de envío y recepción en promedio entre 100 y 300 millones de pesos, y posterior a eso, entre 1,000 y 4,000 millones. A continuación se muestra la gráfica de participación por banco en el clúster, la cual indica que los bancos 9,14,15,17,18,20,22,27,29,33,34,36 estuvieron en el clúster mas del 70% del periodo analizado.

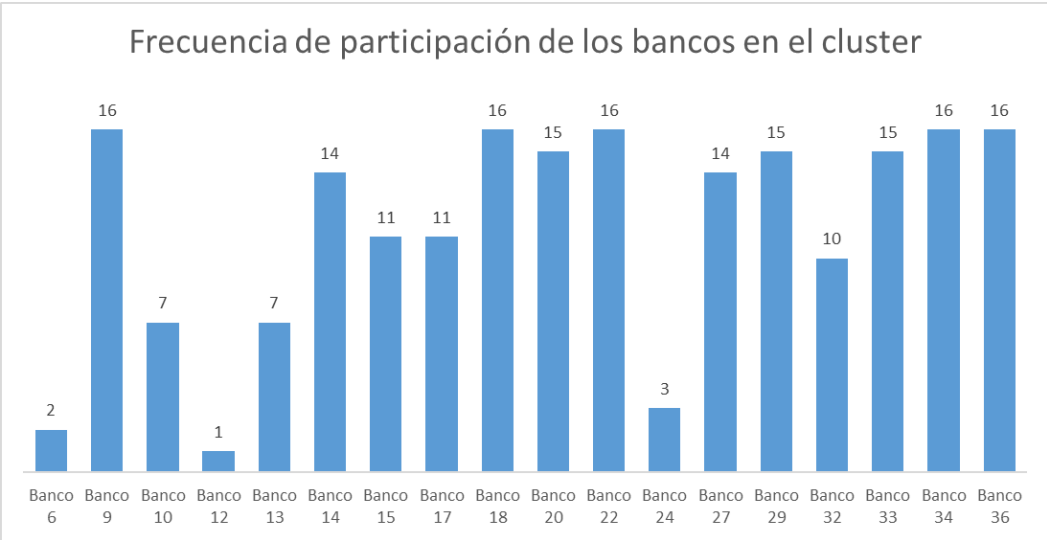


Figura 5.19 Frecuencia de participación de los bancos en el clúster B

Clúster C: Es un clúster muy volátil, ya que solo el banco 12 y el banco 21 estuvieron en el mas del 70% del periodo analizado, resaltando que en sus primeros 7 meses fue un grupo que tuvo un degree en promedio de 14, y manejo de montos aproximadamente de 6,000 millones de pesos, y posterior a ese periodo, llego a alcanzar en ocasiones un degree de 20 y manejando hasta 8,000 millones de pesos en promedio.

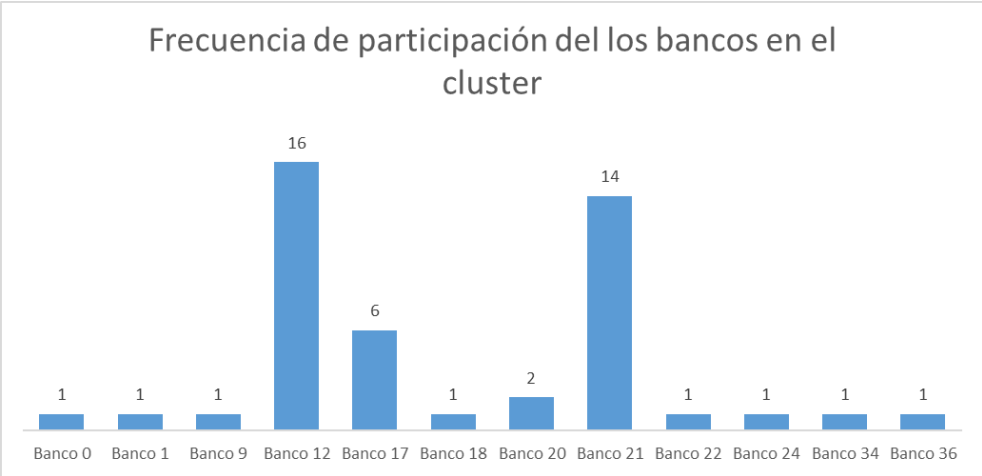


Figura 5.20 Frecuencia de participación de los bancos en el clúster C

Clúster D: Su principal particularidad de este clúster es que los bancos 0,2 y 3, pertenecieron a él al menos un 75% del tiempo. Es el clúster con más alto degree ya que se comunicó más del 90% de las veces con arriba de 30 de 38 bancos participantes, además de ser el clúster con mayores montos de envío y recepción.

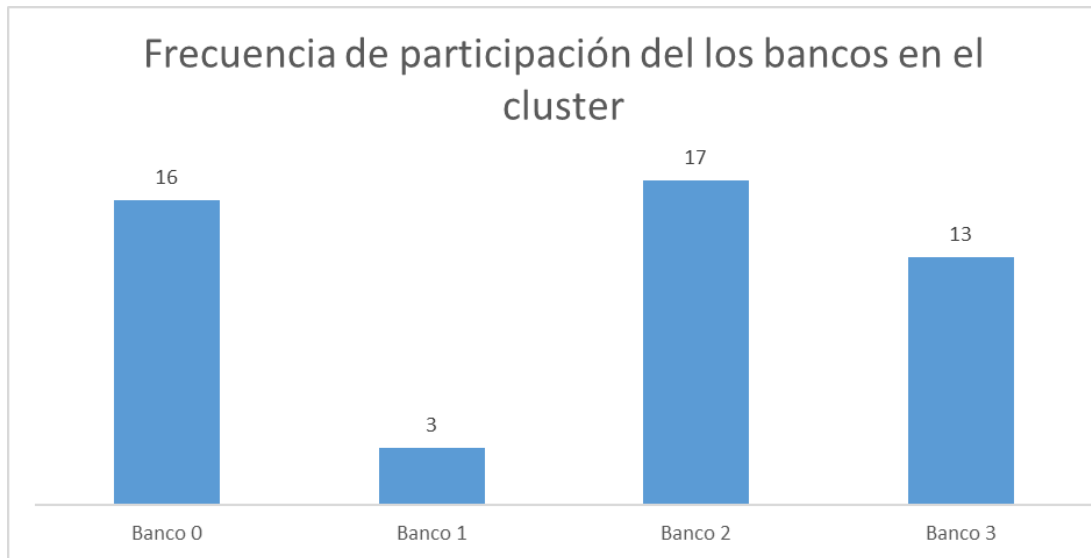


Figura 5.21 Frecuencia de participación de los bancos en el clúster D

Clúster E: En este grupo predominan los bancos 11 y 16, quienes estuvieron en todo el periodo de tiempo, además de los bancos 14 y 12 que estuvieron al menos el 70%. Antes de marzo 2005, se comunicaba con aproximadamente 20 bancos y manejaba montos en promedio de 3,500 millones de pesos, posterior a ese periodo, se comunica con al menos 27 bancos y hace transacciones en promedio por 13 mil millones de pesos.

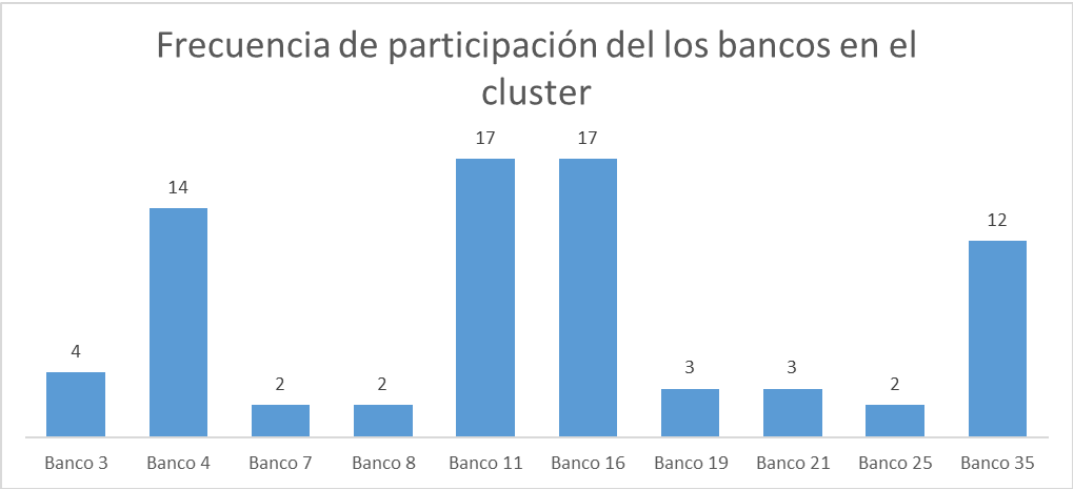


Figura 5.22 Frecuencia de participación de los bancos en el clúster E

Finalmente, se en el anexo 4 se muestra la evolución a través del tiempo de cada uno de los 38 bancos participantes.

5.21 CONCLUSIONES Y TRABAJO FUTURO

Con base en el trabajo realizado, se determina que es posible agrupar a los 38 bancos participantes que iniciaron en el SPEI, en cinco clusters. Dichos clusters son volátiles a través del tiempo, ya que sus participantes son dinámicos, pero también son muy similares ya que en algunos casos hay bancos que siempre pertenecen al mismo clúster. Al hacer el análisis de centroide en cada uno de los meses se puede observar que banco domina cada uno de los grupos y con esto, saber las características generales del clúster. Además, se obtuvieron gráficas para observar en que clúster se encuentra cada banco a través del tiempo.

Con los resultados que se obtuvieron, se proyectan los siguientes trabajos a futuro, ya sea corto o mediano plazo:

- Clusterización de datos en un periodo de tiempo mayor a los 17 meses utilizados en esta tesis
- Detección de patrones de comportamiento en los grupos de bancos ya marcados en este trabajo de tesis
- Aplicación de algún algoritmo de redes neuronales evolutivas sobre los grupos detectados
- Modelado de sistemas multi-agente sobre los grupos y su impacto en el riesgo sistémico

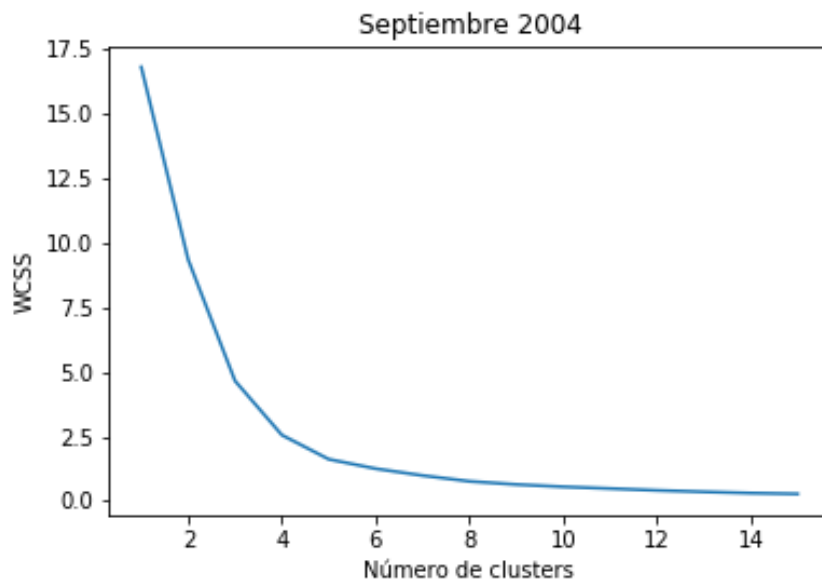
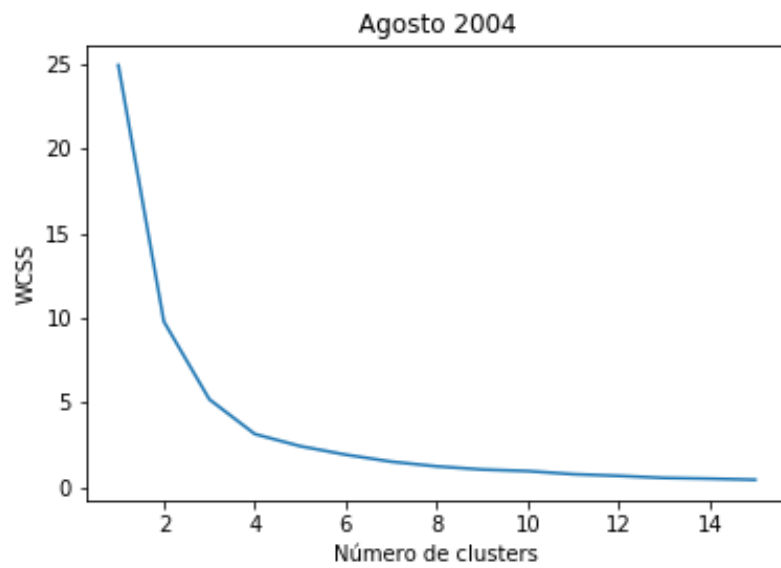
ANEXO 1

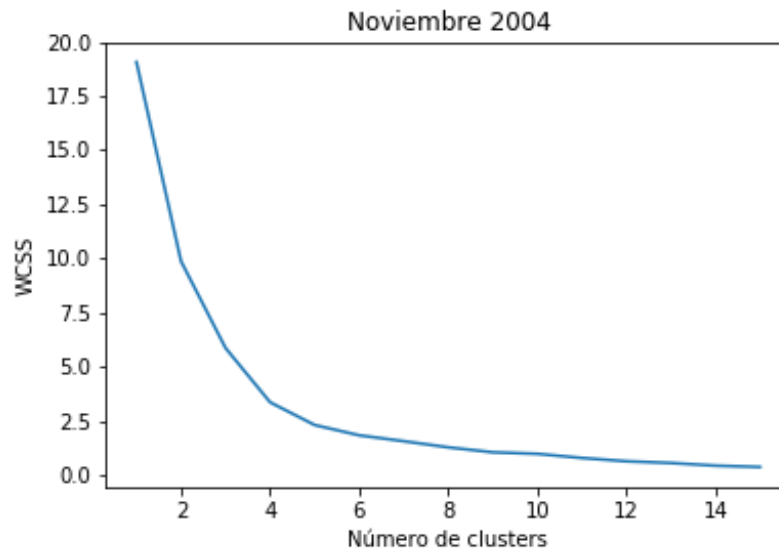
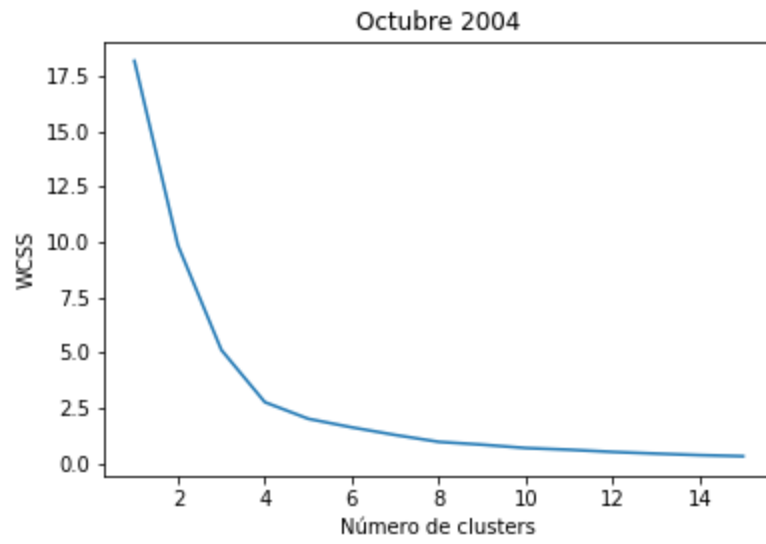
ÍNDICE DE BANCOS TRABAJADOS EN LA TESIS CORRESPONDIENTES A LOS ORIGINALES DADOS POR EL BANCO DE MÉXICO

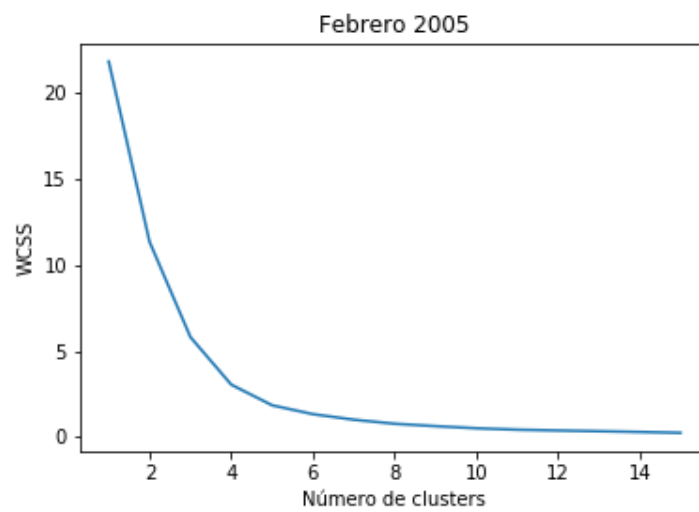
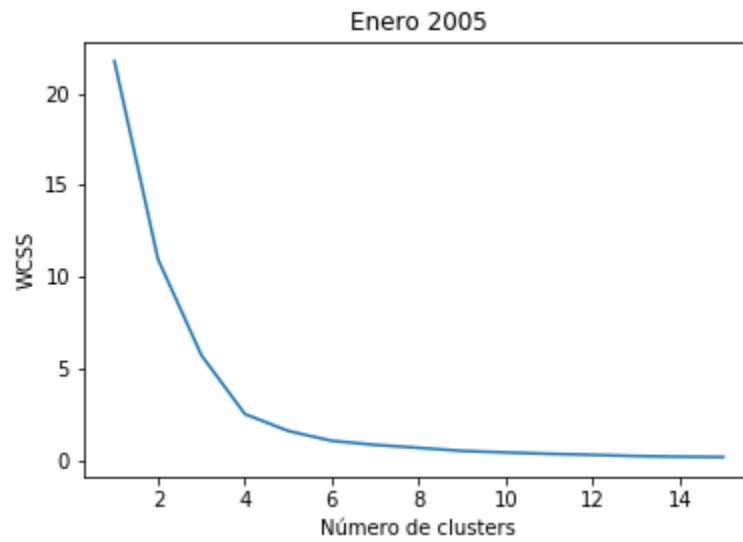
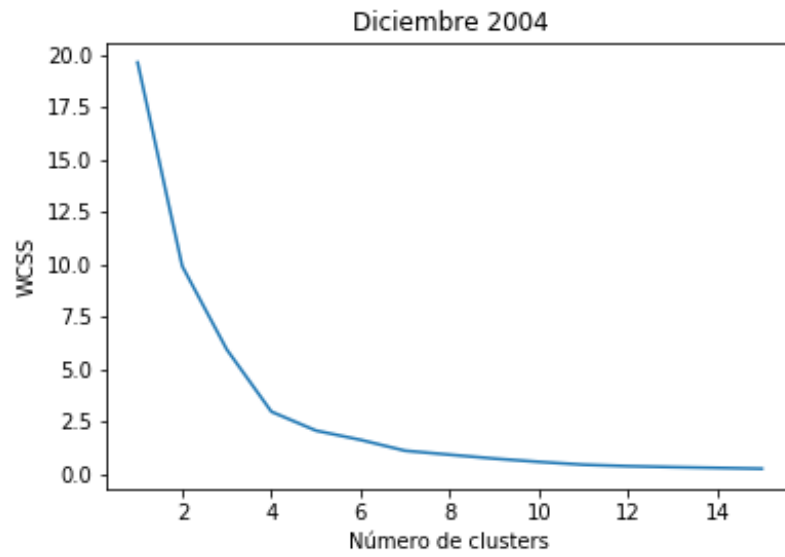
Relación de nomenclatura de bancos			
Banco en tesis	Banco original	Banco en tesis	Banco original
0	1	19	20
1	2	20	21
2	3	21	22
3	4	22	23
4	5	23	24
5	6	24	25
6	7	25	26
7	8	26	27
8	9	27	28
9	10	28	29
10	11	29	30
11	12	30	31
12	13	31	32
13	14	32	56
14	15	33	57
15	16	34	58
16	17	35	59
17	18	36	60
18	19	37	61

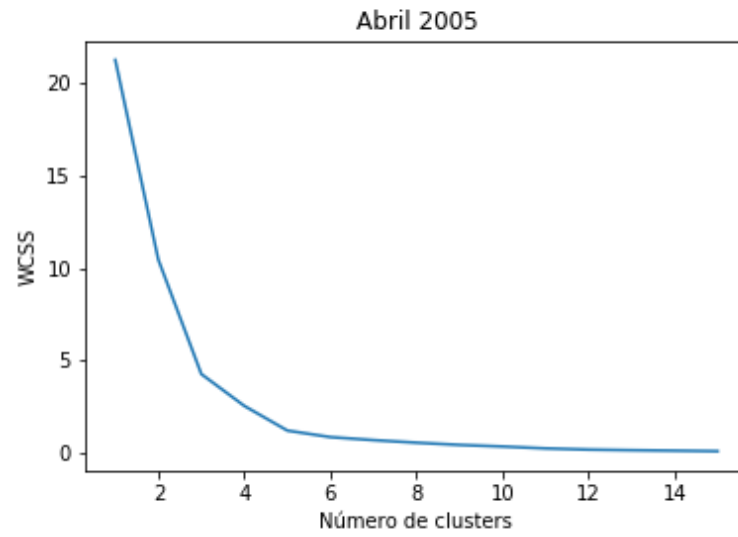
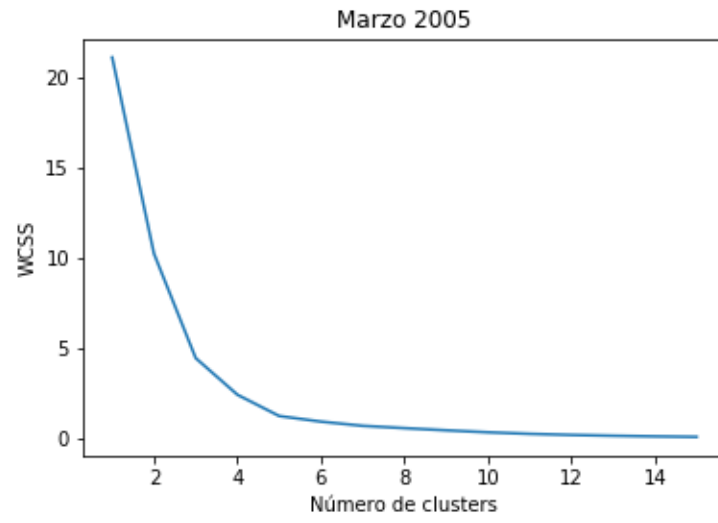
ANEXO 2

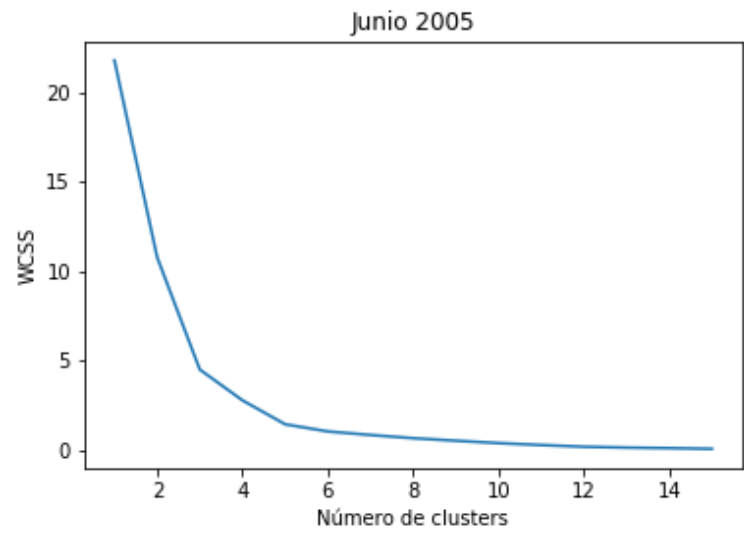
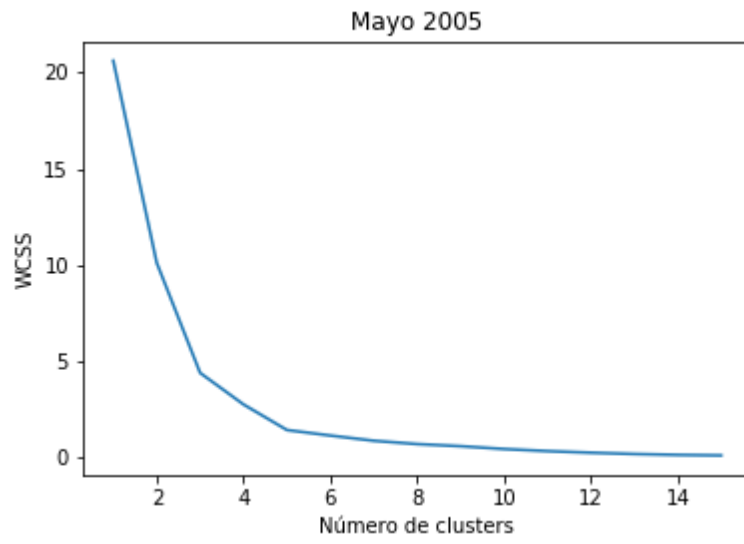
GRAFICAS OBTENIDAS AL EJECUTAR WCSS PARA LA BÚSQUEDA DE EL VALOR "K"

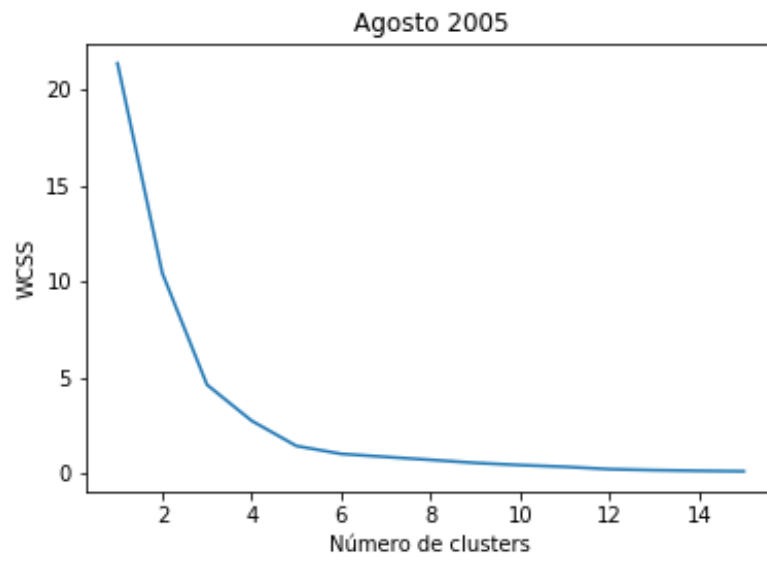
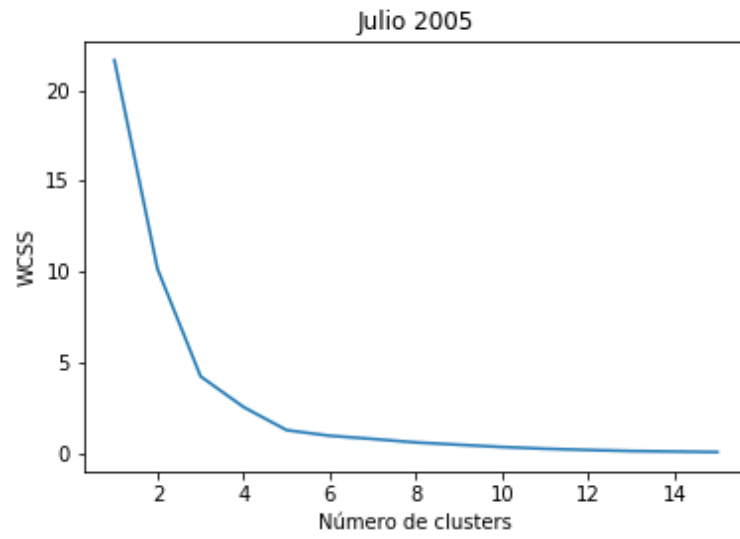


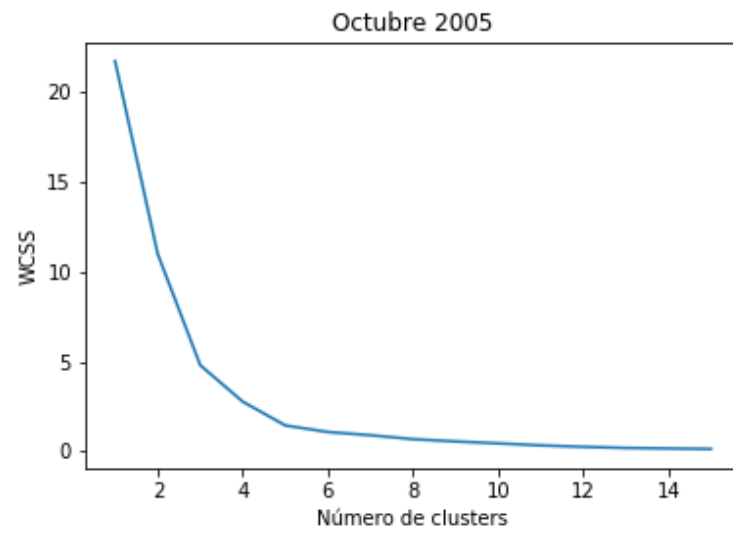
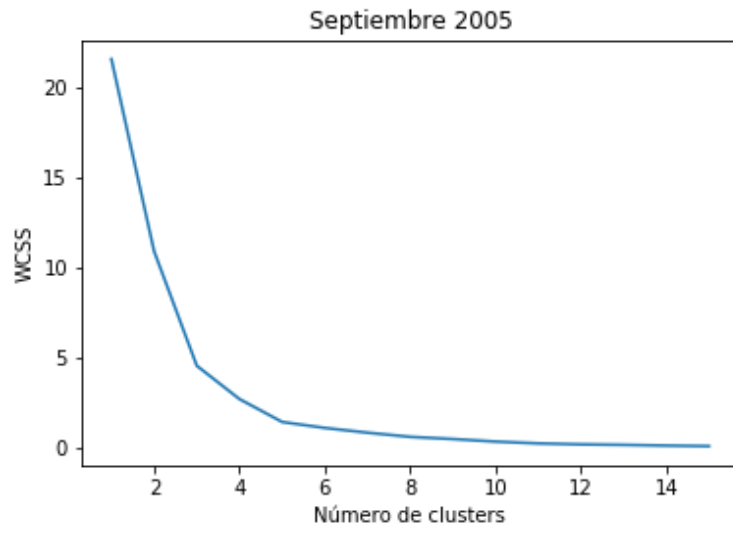


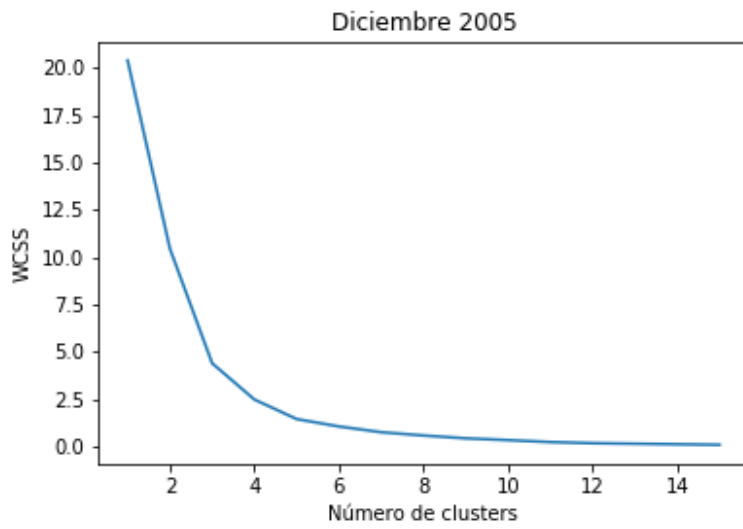
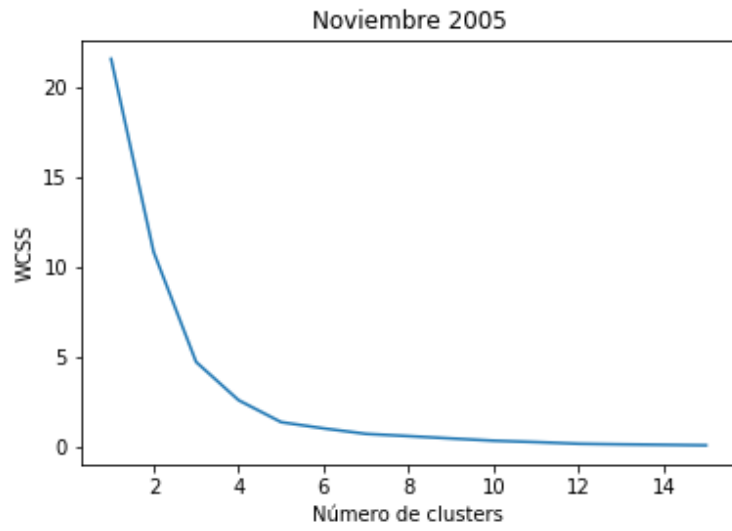












ANEXO 3

SALIDA OBTENIDA EN LA EJECUCION DE KMEANS

Index	Ago04	Sep04	Oct04	Nov04	Dic04	Ene05	Feb05	Mar05
0	3	1	2	2	3	3	4	1
1	3	1	2	2	3	1	3	2
2	3	4	2	2	3	3	4	1
3	3	3	4	3	0	3	4	1
4	0	3	4	3	4	2	2	4
5	1	2	1	1	2	4	1	3
6	4	0	0	3	4	2	2	0
7	0	3	0	0	4	2	2	0
8	0	3	0	0	4	2	2	0
9	2	0	3	1	2	4	1	3
10	4	0	0	0	2	4	2	0
11	0	3	4	3	0	0	0	4
12	2	0	0	0	4	2	2	0
13	4	0	0	0	2	4	1	0
14	4	0	0	0	2	4	1	3
15	4	0	3	0	2	4	1	3
16	0	3	4	3	0	0	0	4
17	2	0	0	0	4	2	2	3
18	2	0	3	1	2	4	1	3

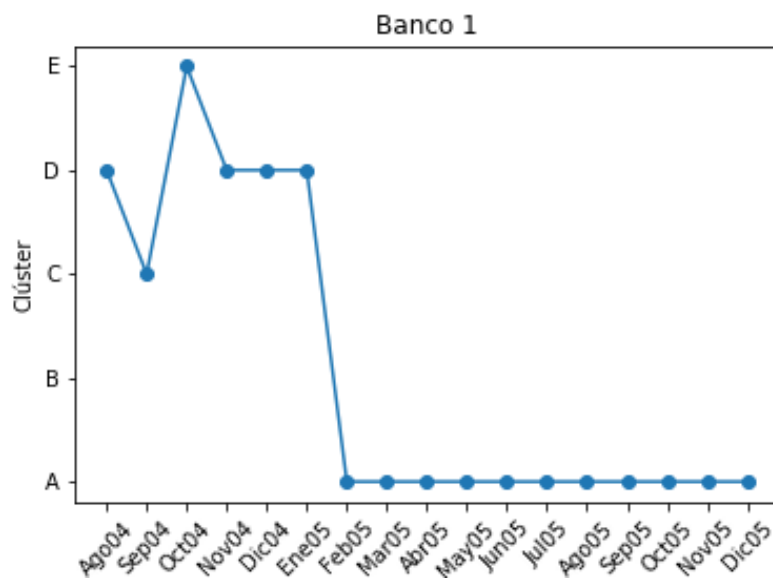
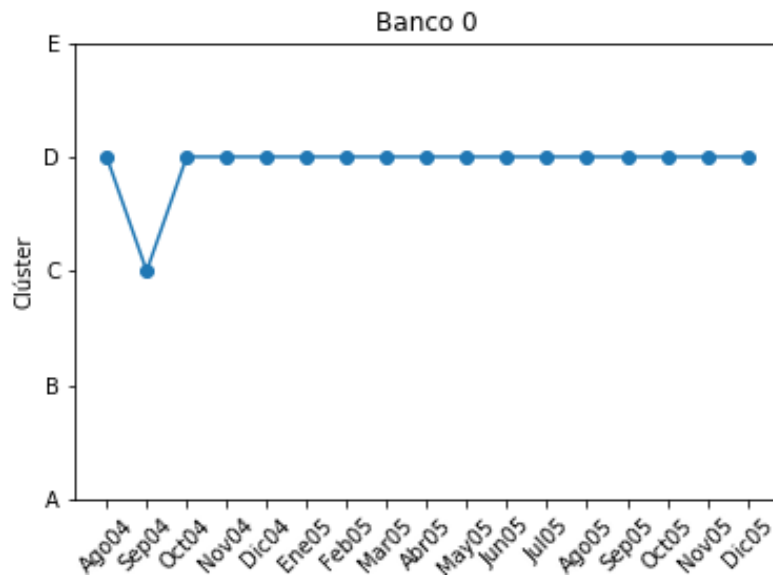
Index	Ago04	Sep04	Oct04	Nov04	Dic04	Ene05	Feb05	Mar05
19	1	3	4	3	4	2	2	0
20	2	0	3	0	2	4	1	3
21	0	1	4	3	4	2	2	0
22	2	0	3	1	2	4	1	3
23	1	2	3	1	2	4	1	3
24	2	0	3	1	1	1	3	2
25	0	3	0	0	4	2	2	3
26	1	2	1	4	1	1	3	2
27	4	0	0	0	4	4	1	3
28	1	2	3	1	2	4	1	3
29	4	0	3	0	2	4	1	3
30	1	2	1	4	1	1	3	2
31	1	2	1	4	1	1	3	2
32	1	0	0	0	2	4	1	3
33	4	0	0	0	2	4	1	3
34	2	0	3	1	2	4	1	3
35	0	3	4	3	0	0	0	0
36	2	0	3	1	2	4	1	3
37	1	2	1	4	2	4	1	3

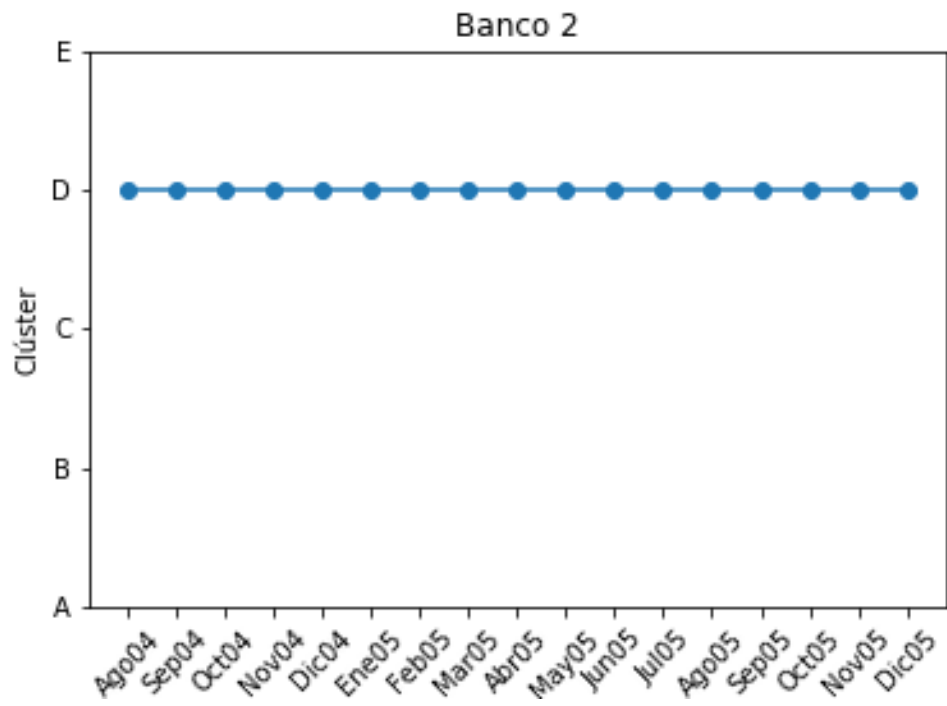
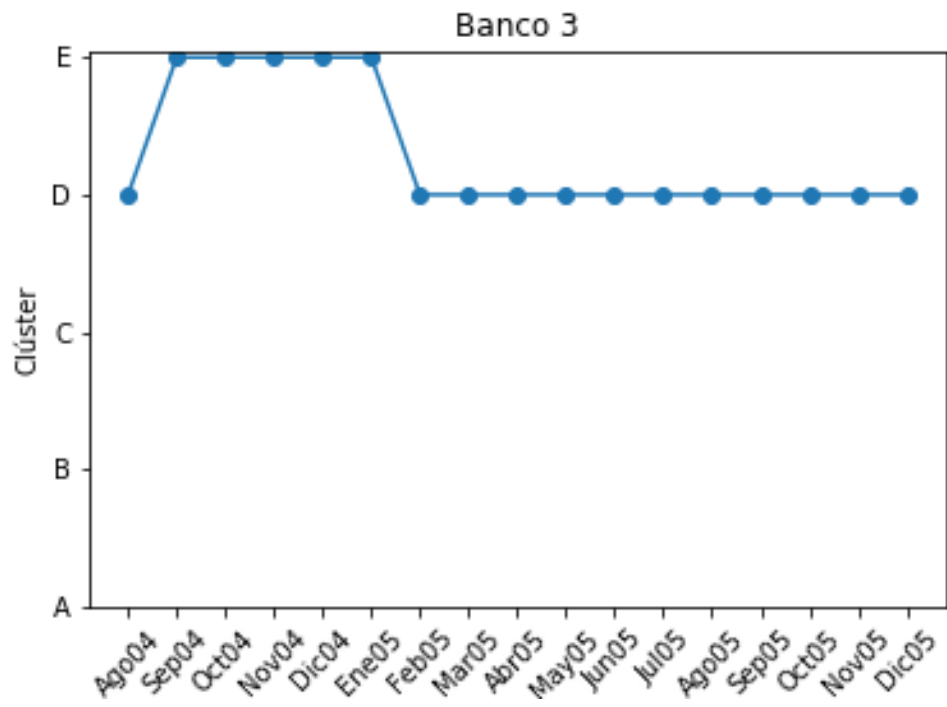
Index	Abr05	May05	Jun05	Jul05	Ago05	Sep05	Oct05	Nov05	Dic05
0	1	3	1	1	1	3	3	1	1
1	2	2	2	2	2	1	1	2	2
2	1	3	1	1	1	3	3	1	1
3	1	3	1	1	1	3	3	1	1
4	4	1	4	4	4	0	0	4	4
5	0	0	3	0	0	2	4	0	3
6	3	4	0	3	3	4	2	3	0
7	3	4	0	3	3	4	2	3	0
8	3	4	0	3	3	4	2	3	0
9	0	0	3	0	0	2	4	0	3
10	3	4	0	3	0	2	4	3	0
11	4	1	4	4	4	0	0	4	4
12	3	4	0	3	3	4	2	3	0
13	3	4	0	3	0	2	2	3	0
14	0	0	3	0	0	2	4	0	0
15	0	4	0	3	0	2	4	3	0
16	4	1	4	4	4	0	0	4	4
17	0	0	3	0	0	2	4	0	3
18	0	0	3	0	0	2	4	0	3

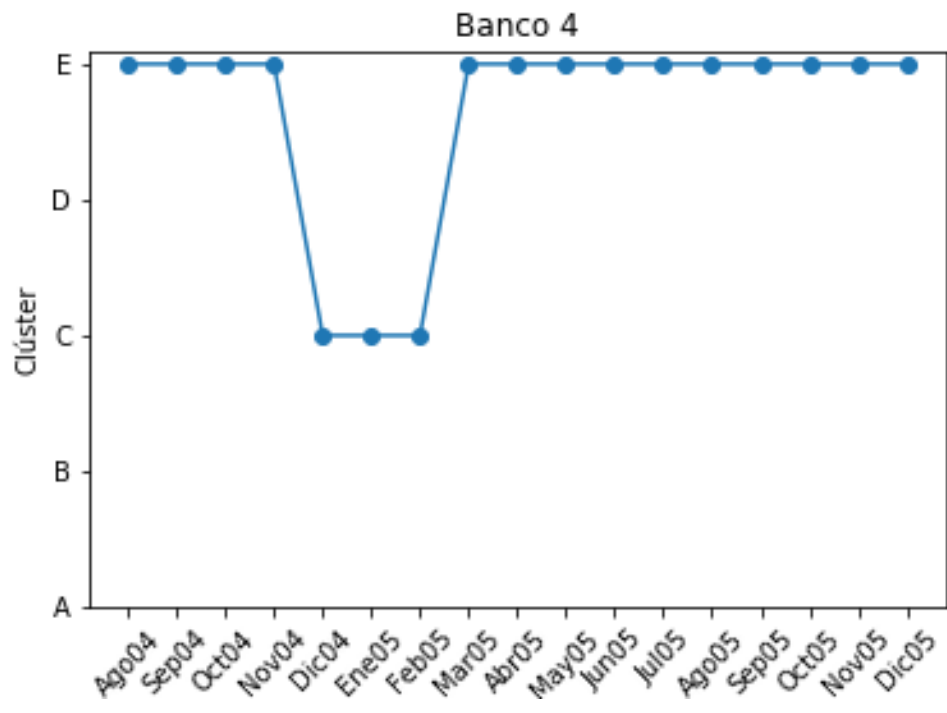
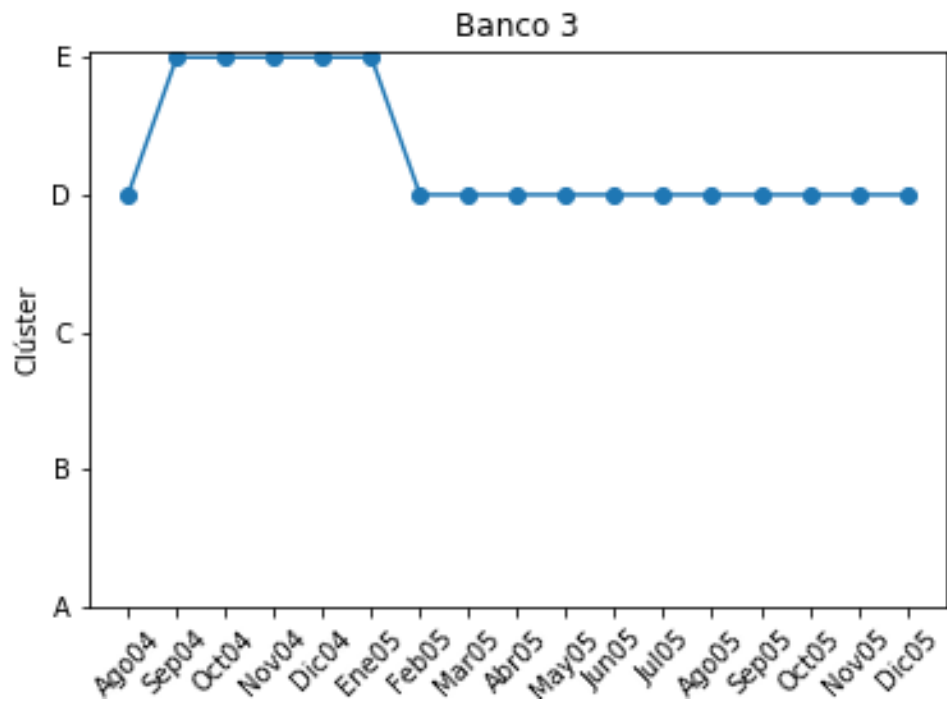
Index	Abr05	May05	Jun05	Jul05	Ago05	Sep05	Oct05	Nov05	Dic05
19	3	4	0	3	3	4	2	3	0
20	0	0	3	0	0	2	4	0	3
21	3	4	0	3	3	4	2	3	0
22	0	0	3	0	0	2	4	0	3
23	0	0	3	0	0	2	4	0	3
24	2	2	2	2	2	1	1	2	2
25	0	0	3	0	0	2	4	0	3
26	2	2	2	2	2	1	1	2	2
27	0	0	3	0	0	2	4	0	3
28	0	0	3	0	0	2	4	0	3
29	0	0	3	0	0	2	4	0	0
30	2	2	2	2	2	1	1	2	2
31	2	2	2	2	2	1	1	2	2
32	3	4	0	3	0	2	4	0	3
33	0	0	3	0	0	2	4	0	3
34	0	0	3	0	0	2	4	0	3
35	3	1	4	4	3	4	0	3	4
36	0	0	3	0	0	2	4	0	3
37	0	0	3	0	0	2	4	0	3

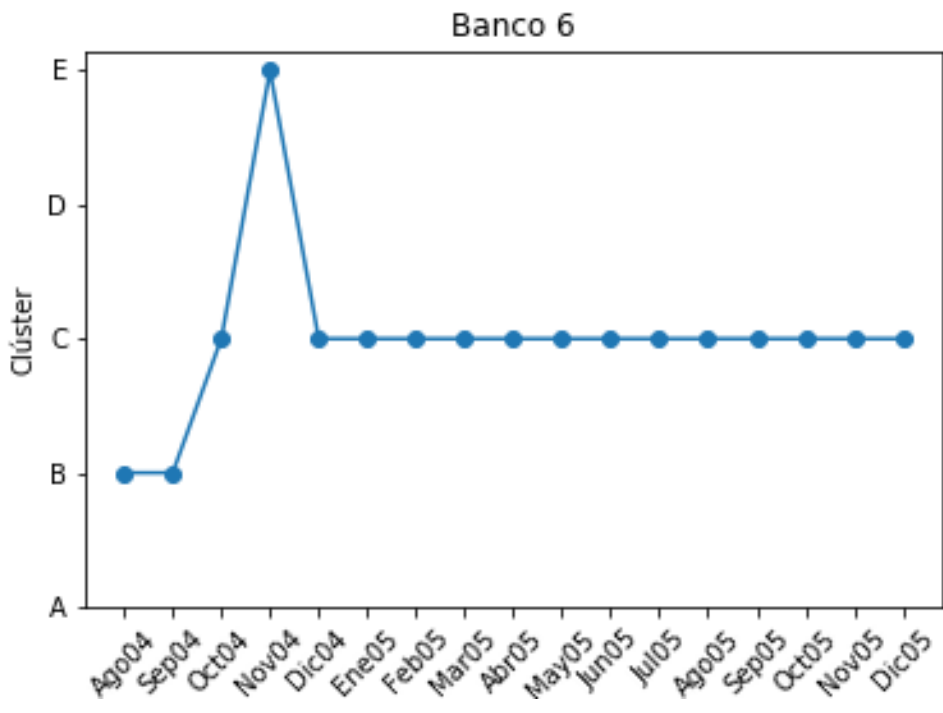
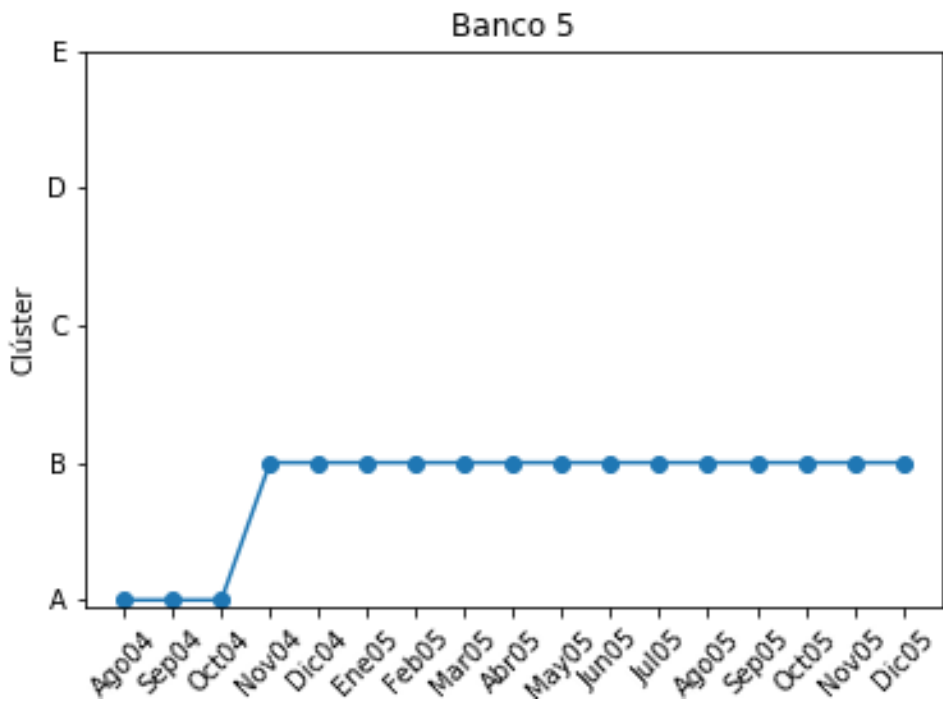
ANEXO 4

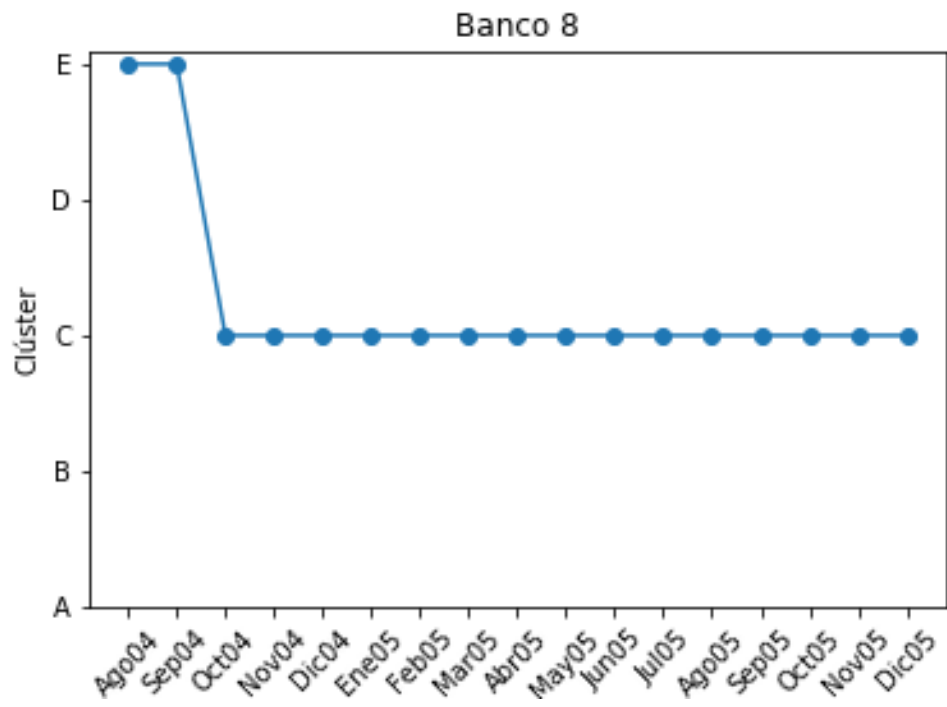
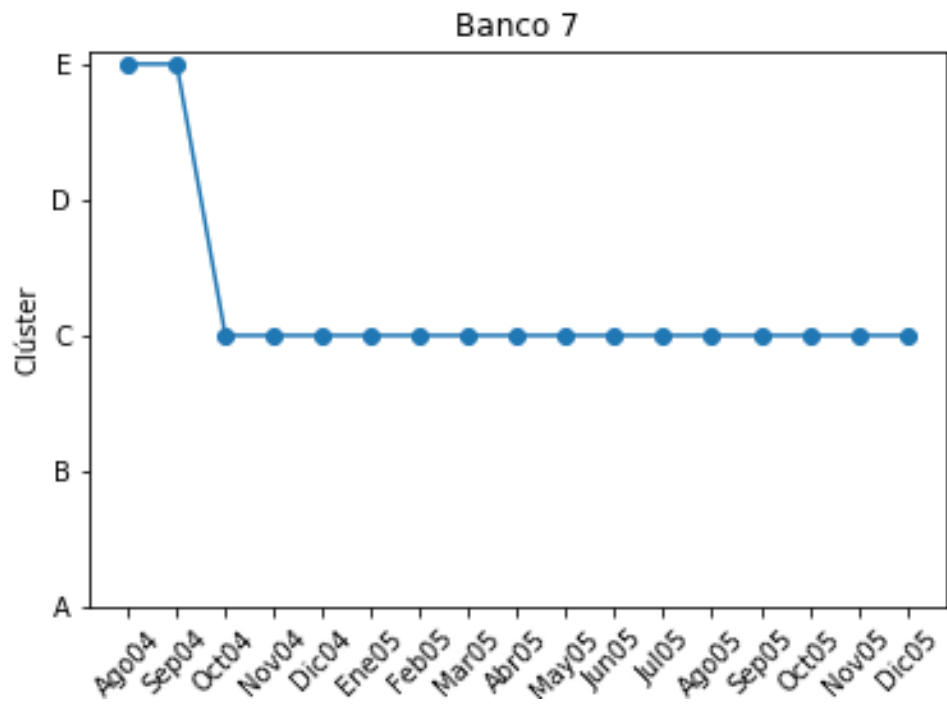
EVOLUCIÓN DE LOS BANCOS A TRAVÉS DEL TIEMPO EN LOS GRUPOS “A” “B” “C” “D” “E”

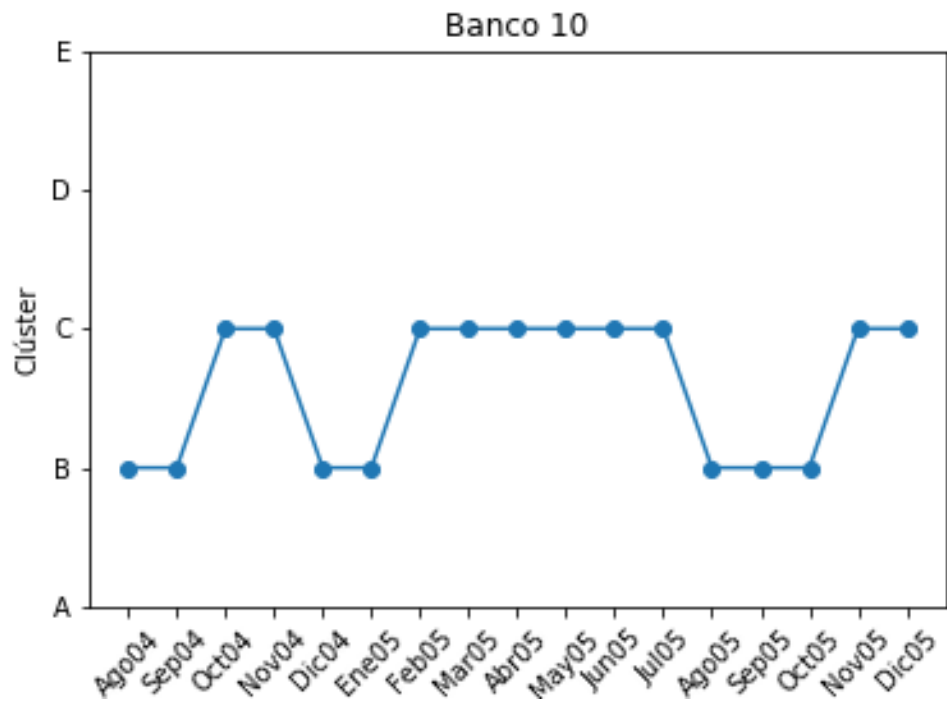
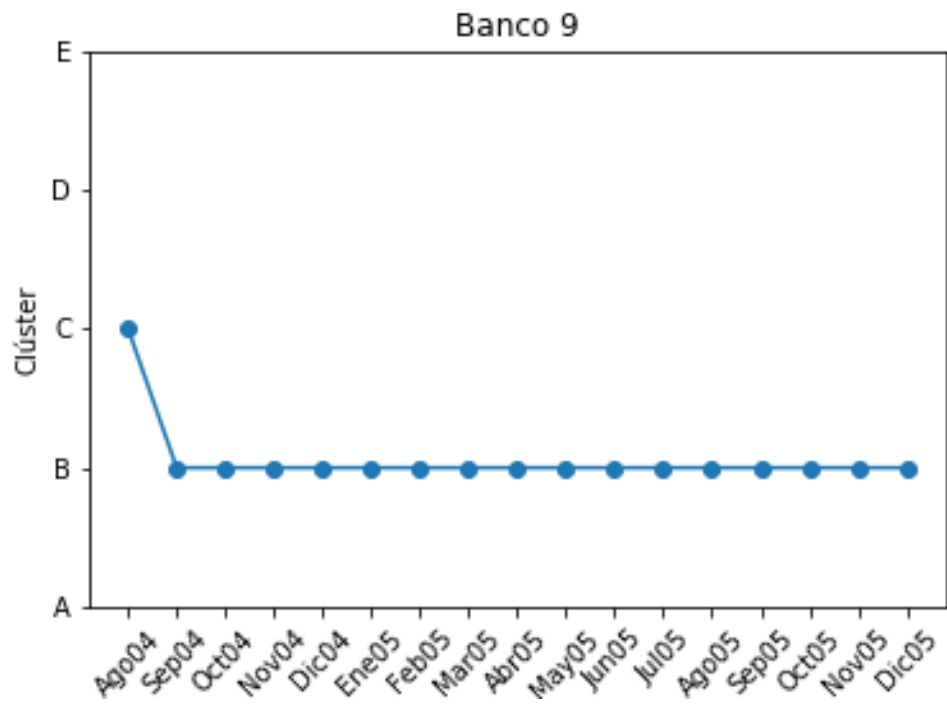


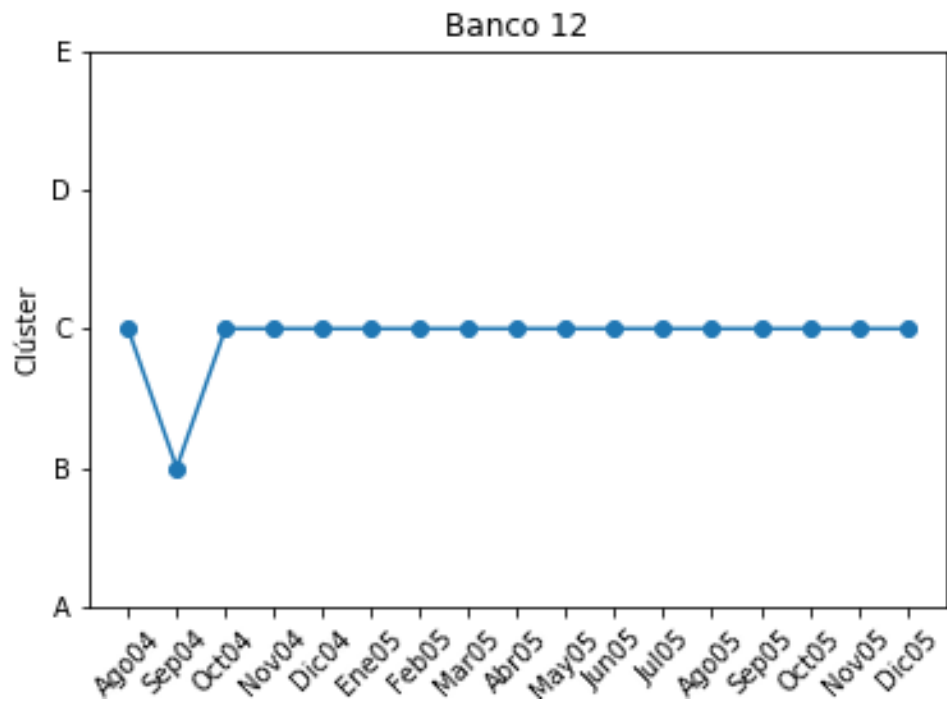
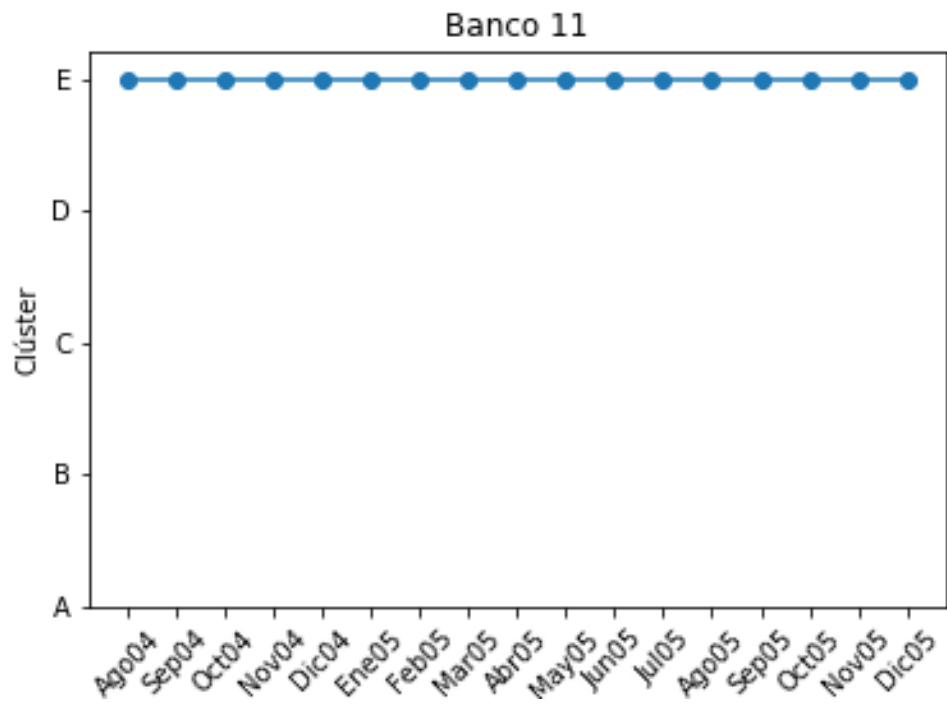


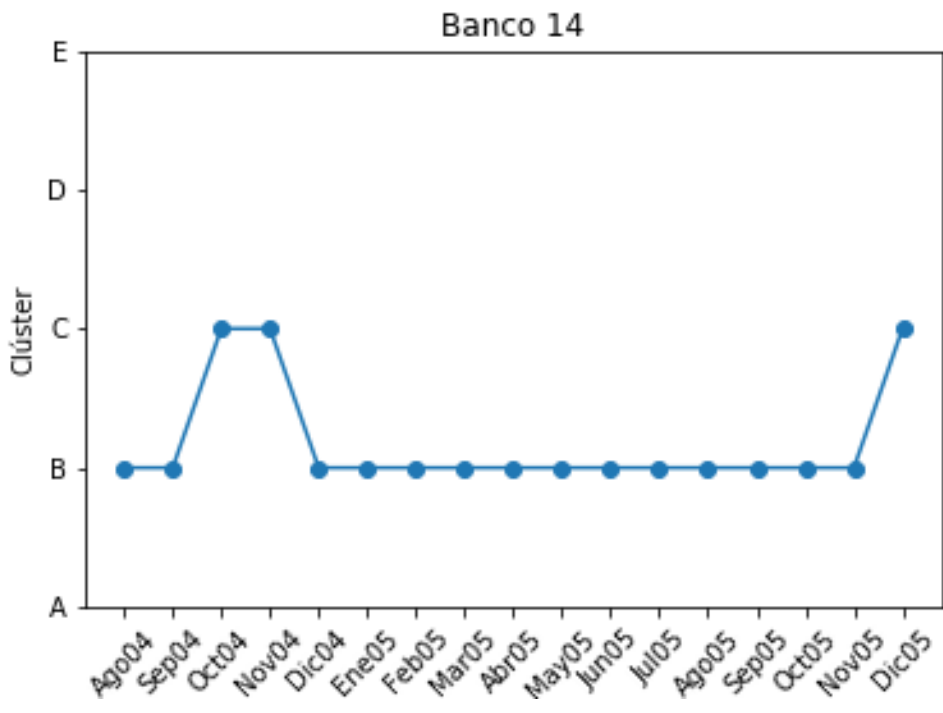
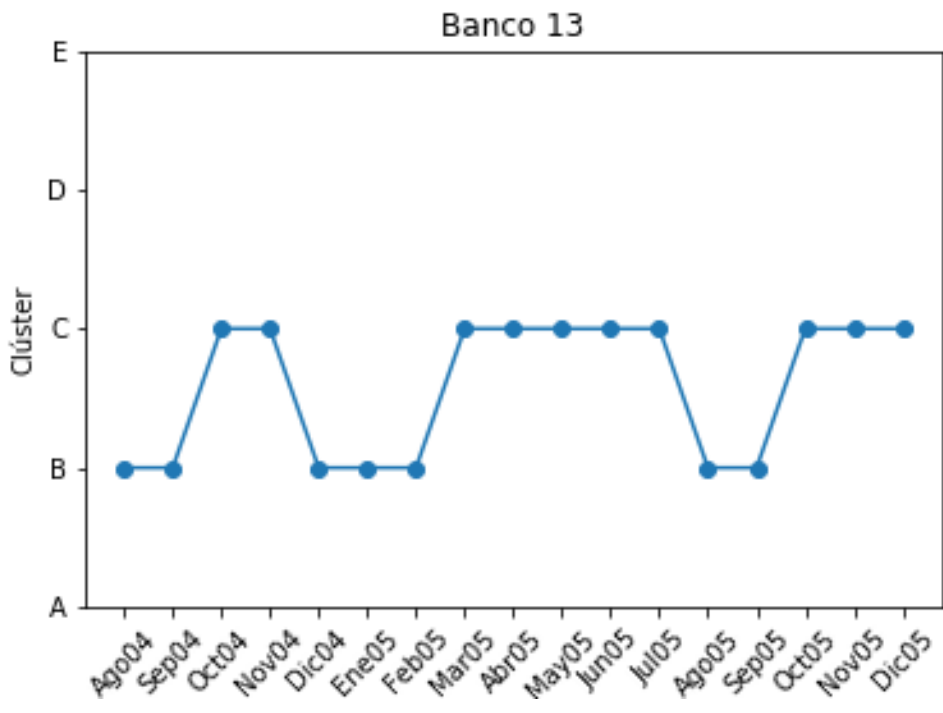


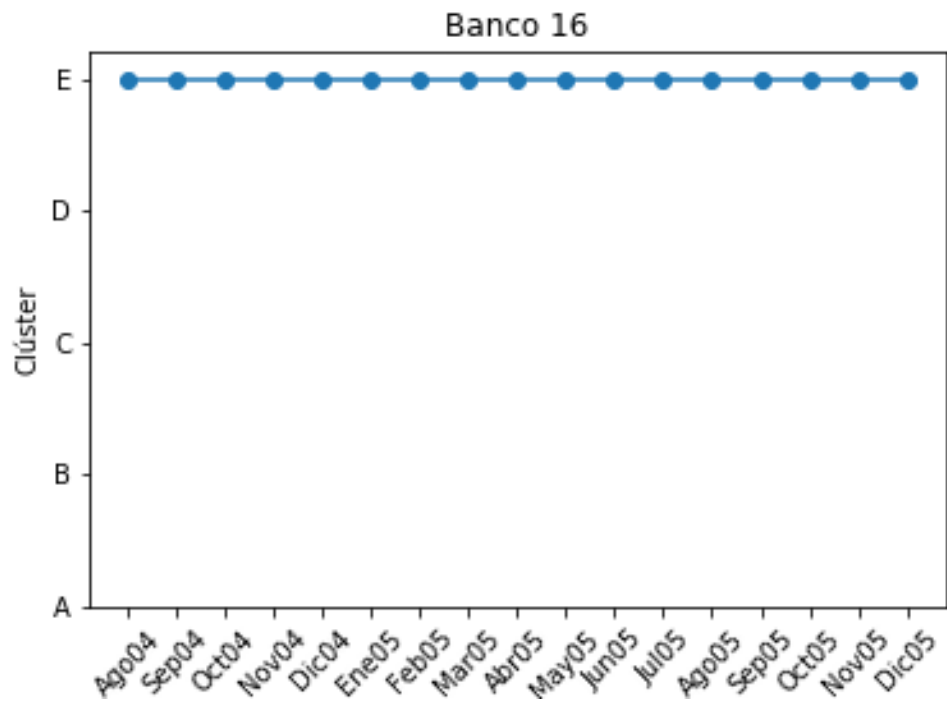
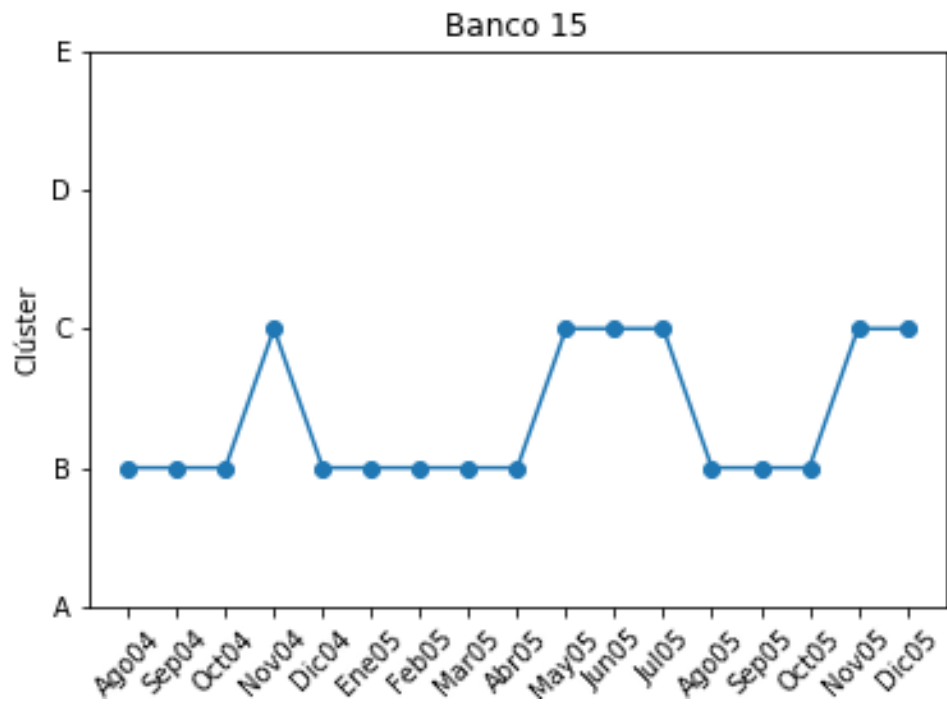


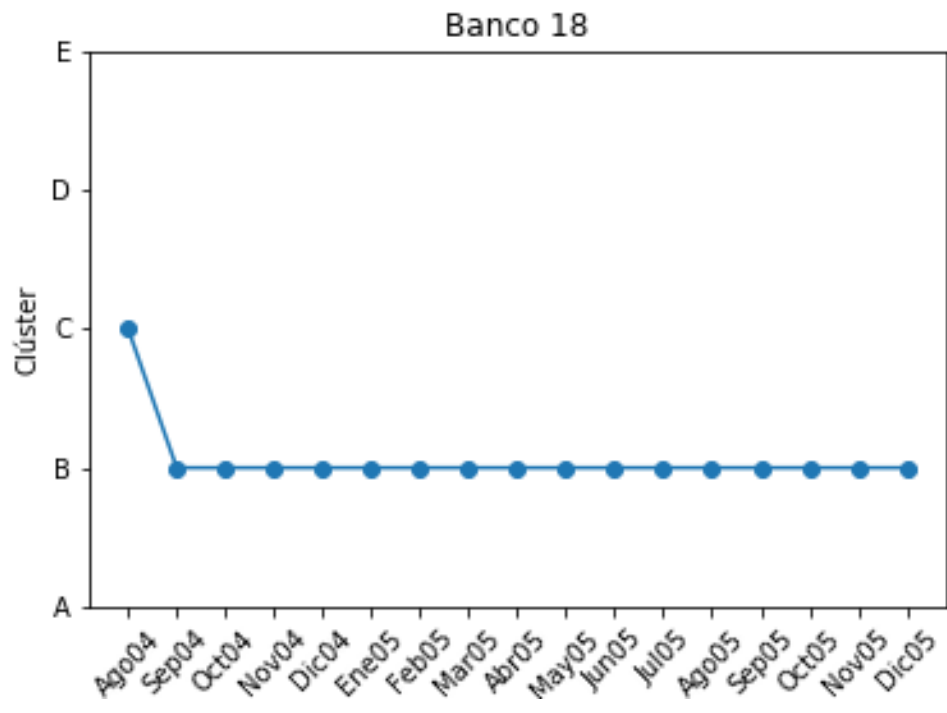
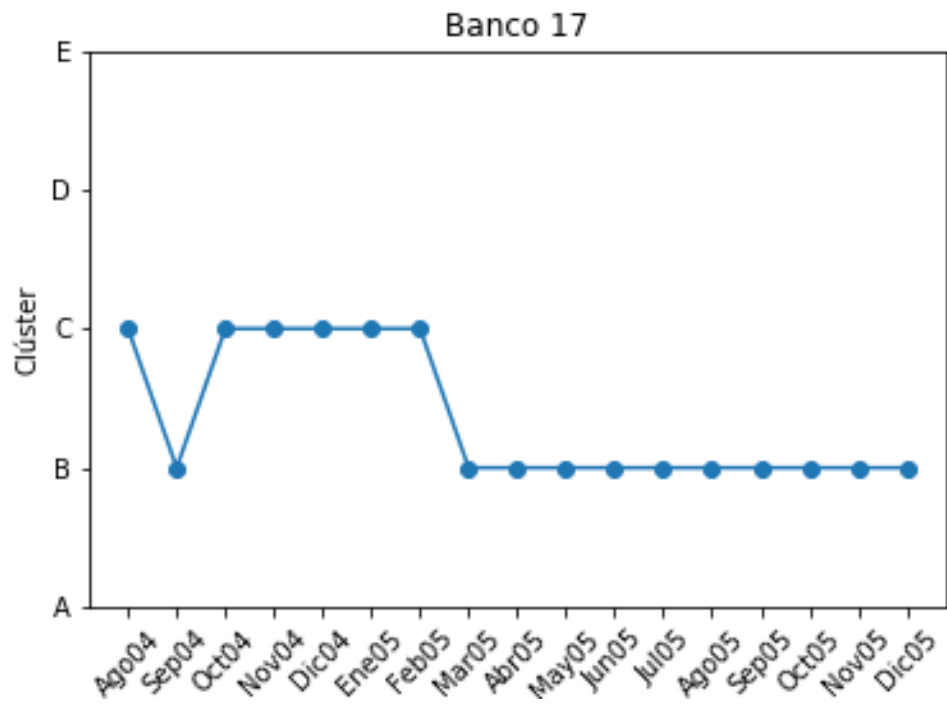


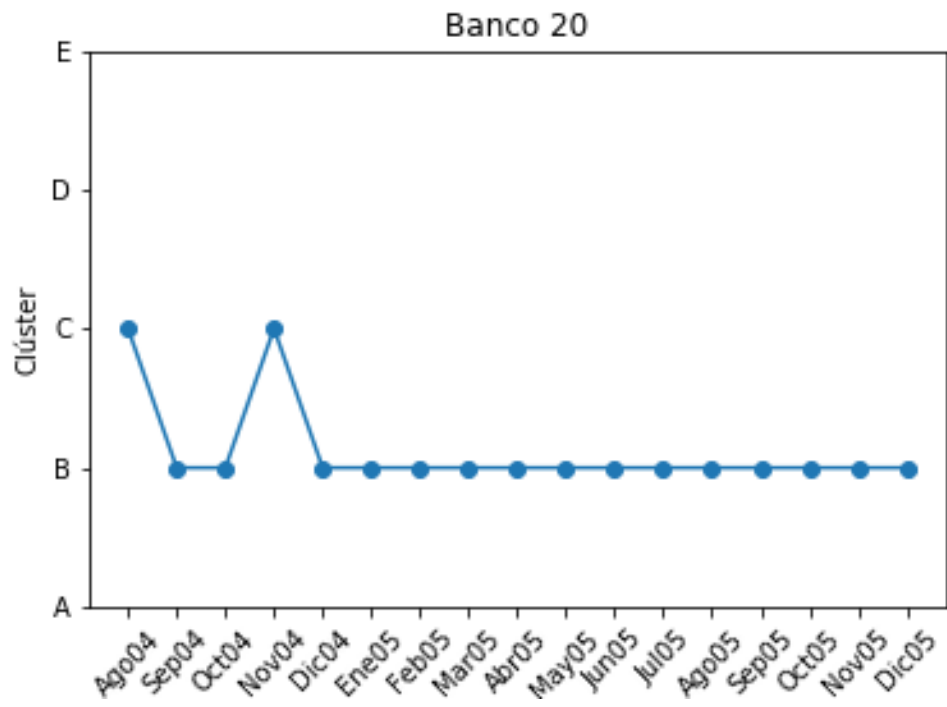
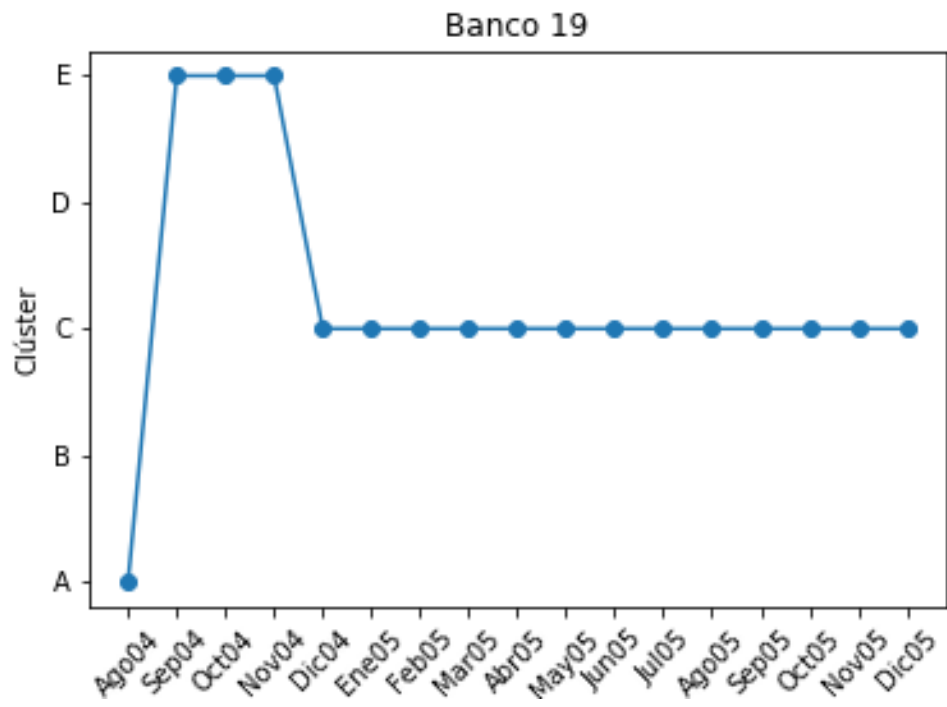


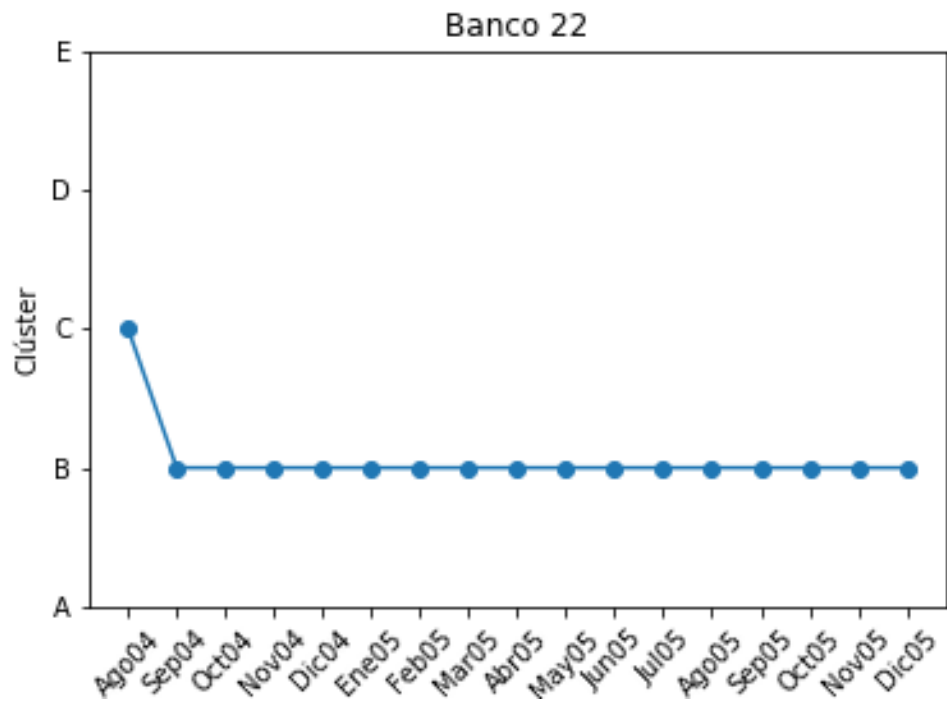
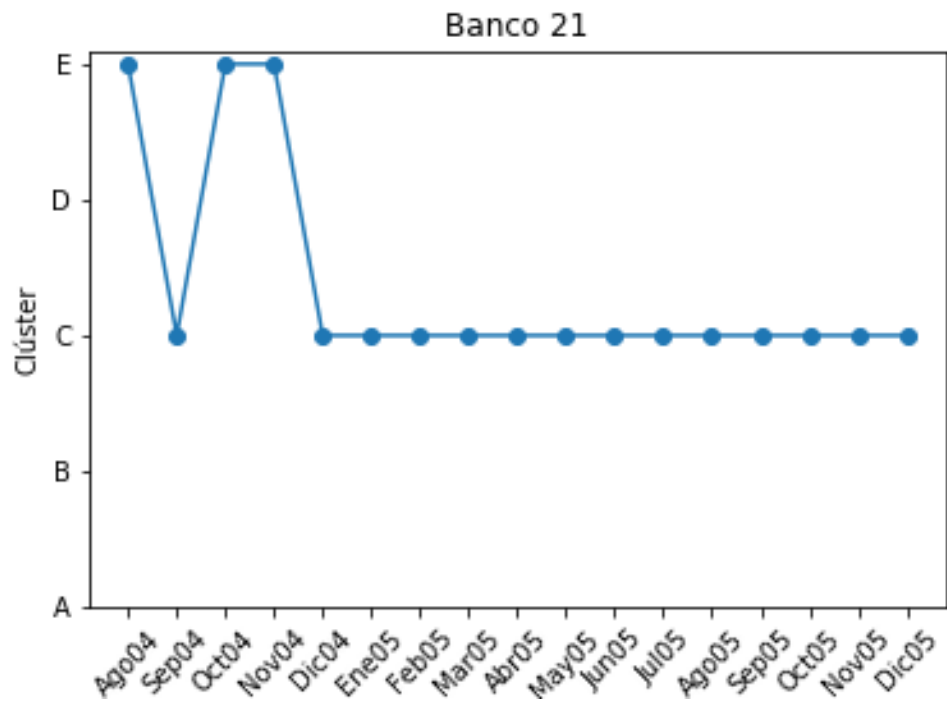


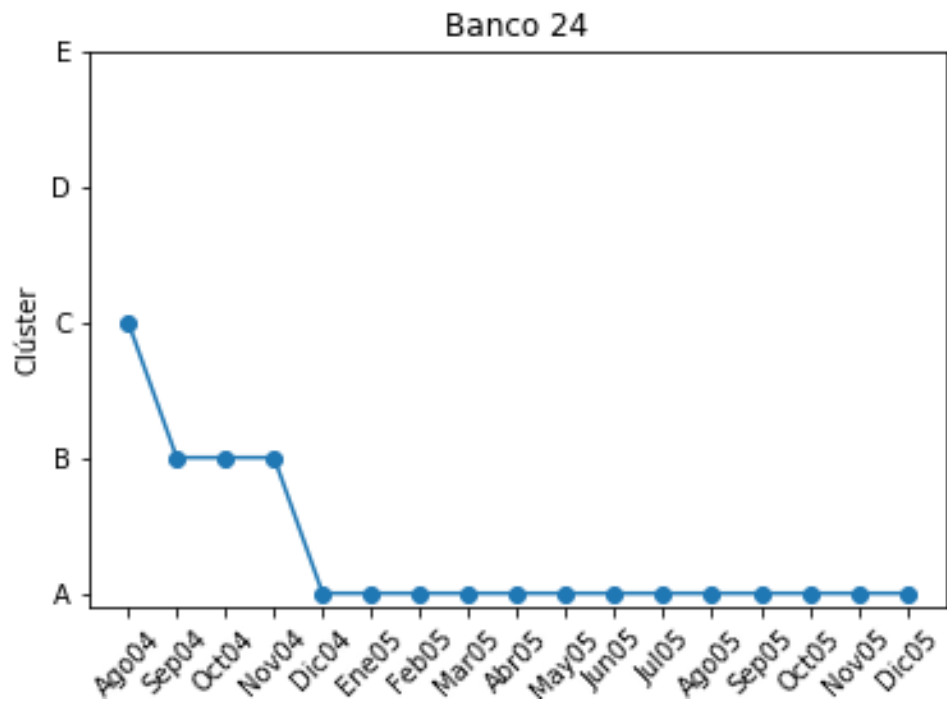
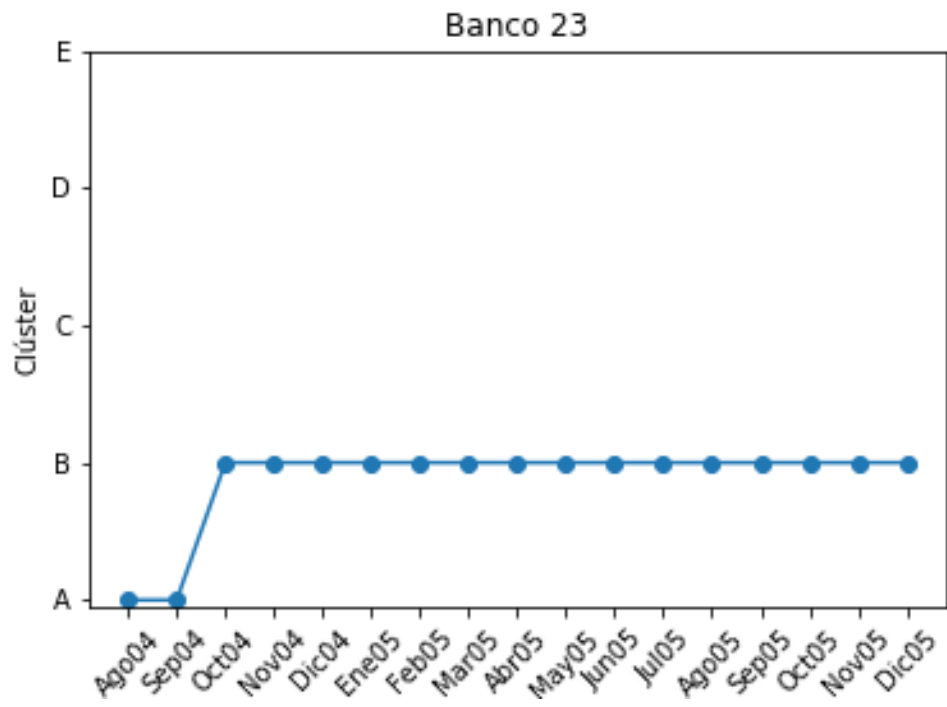


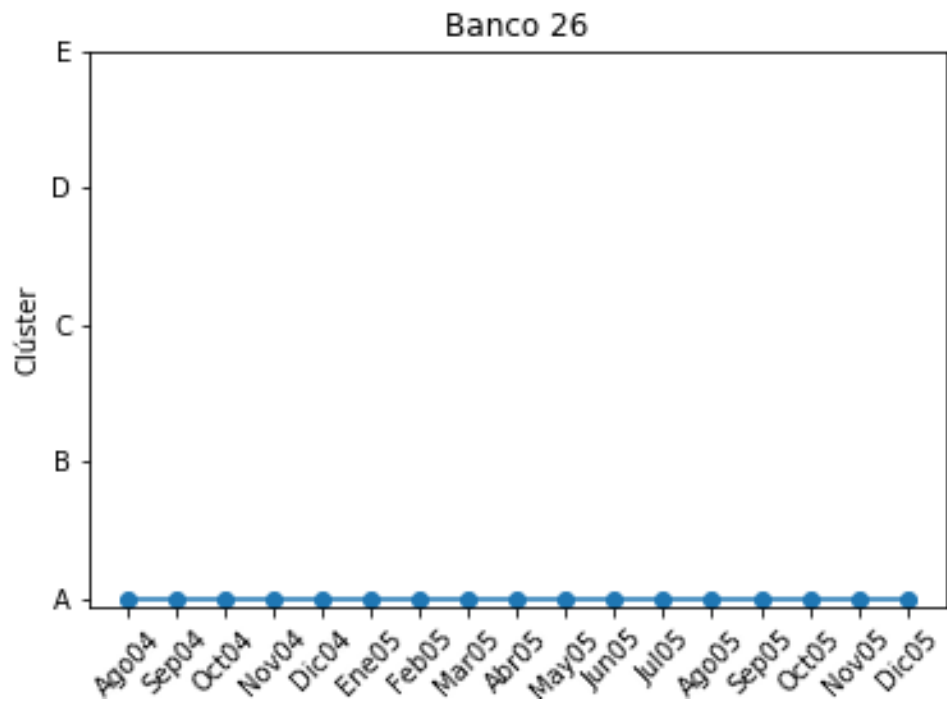
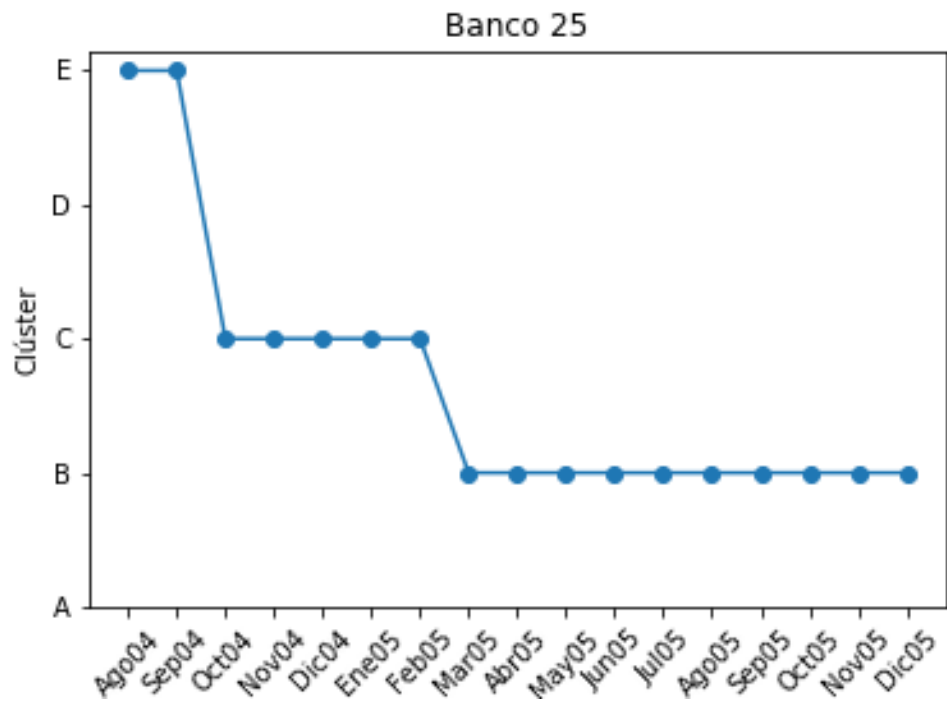


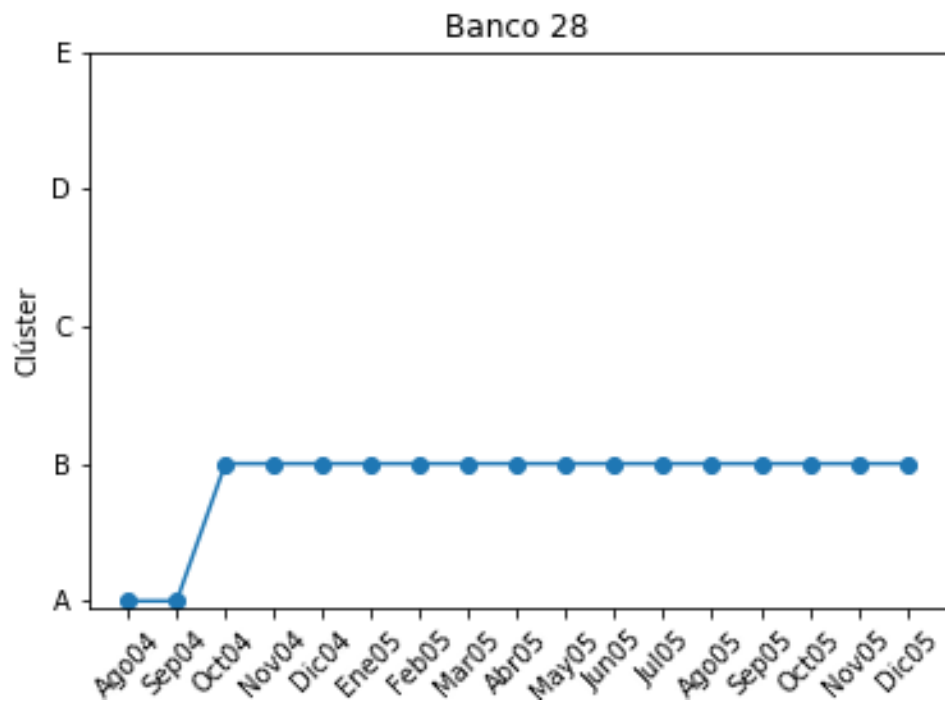
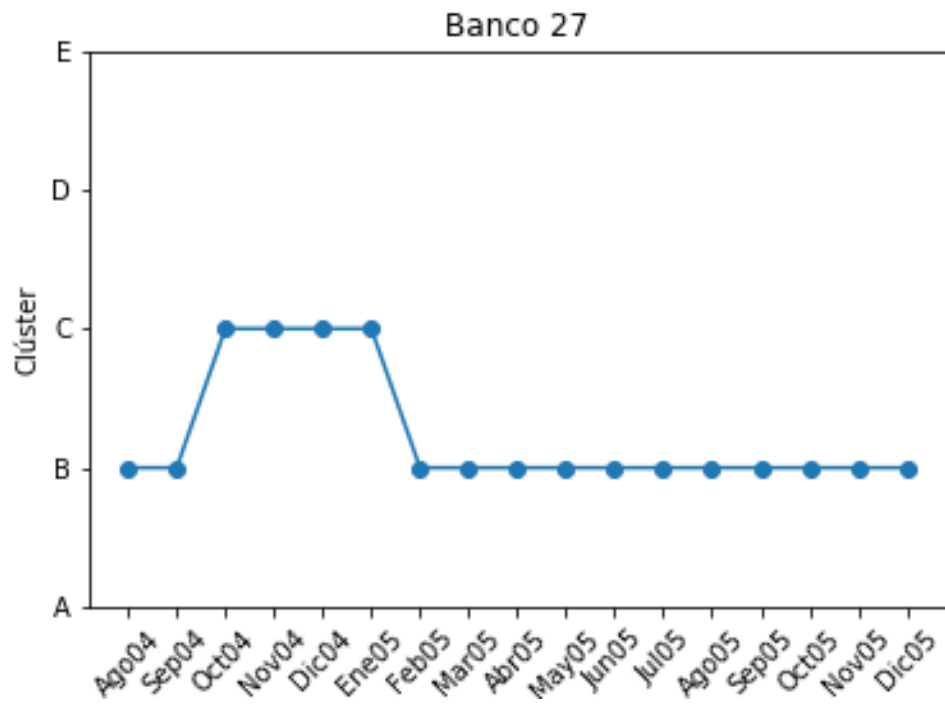


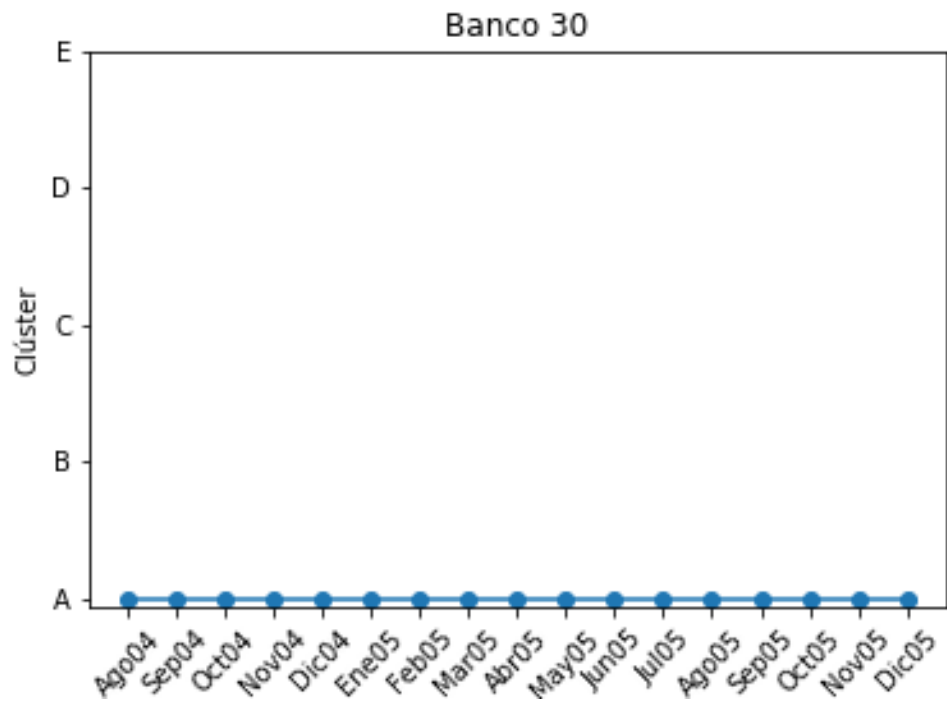
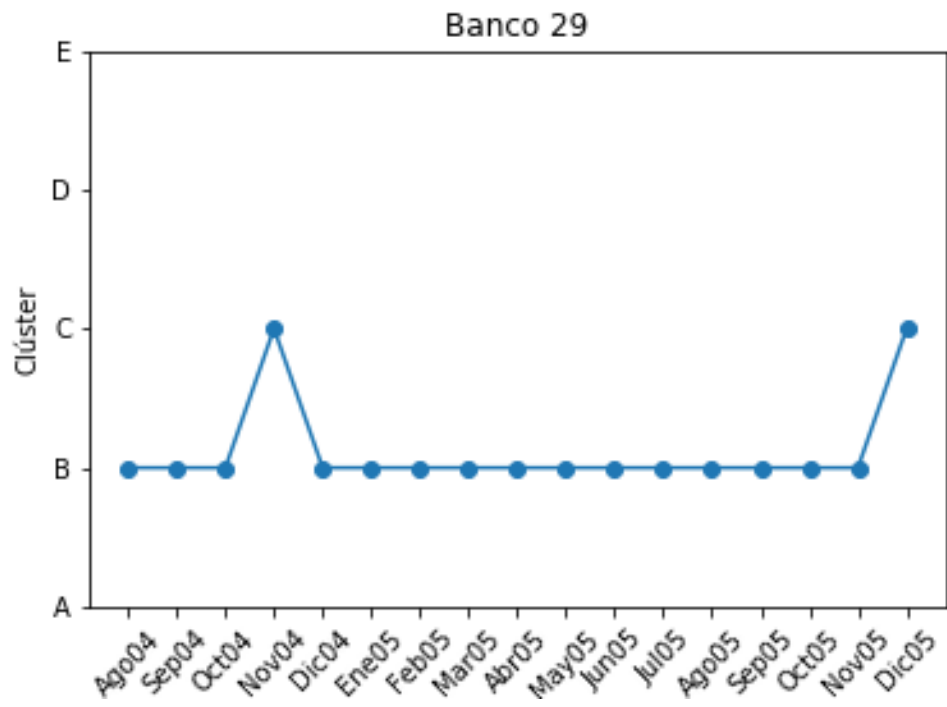


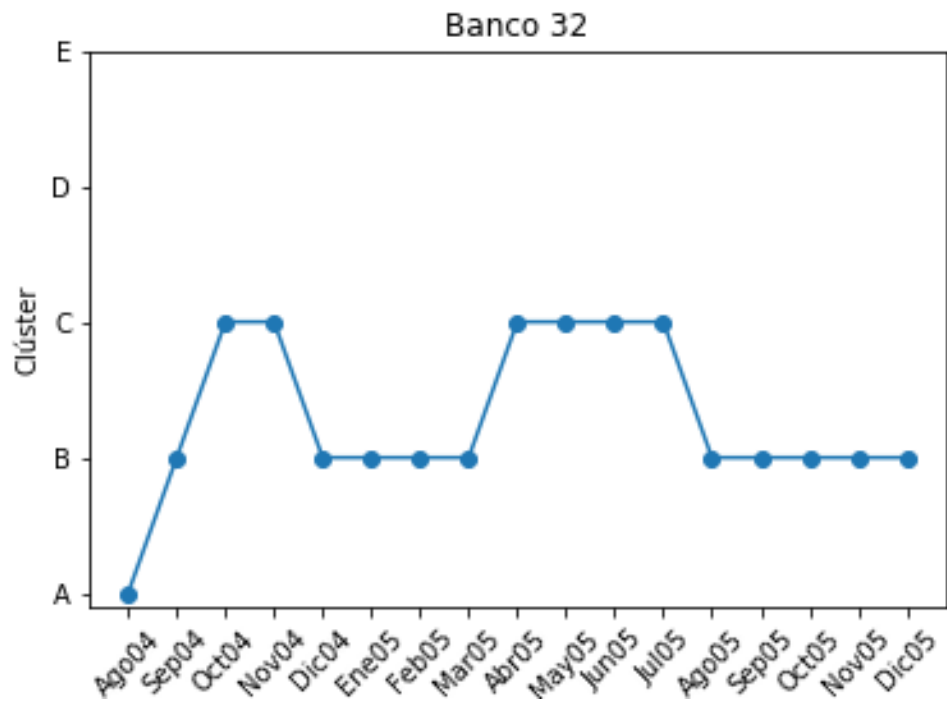
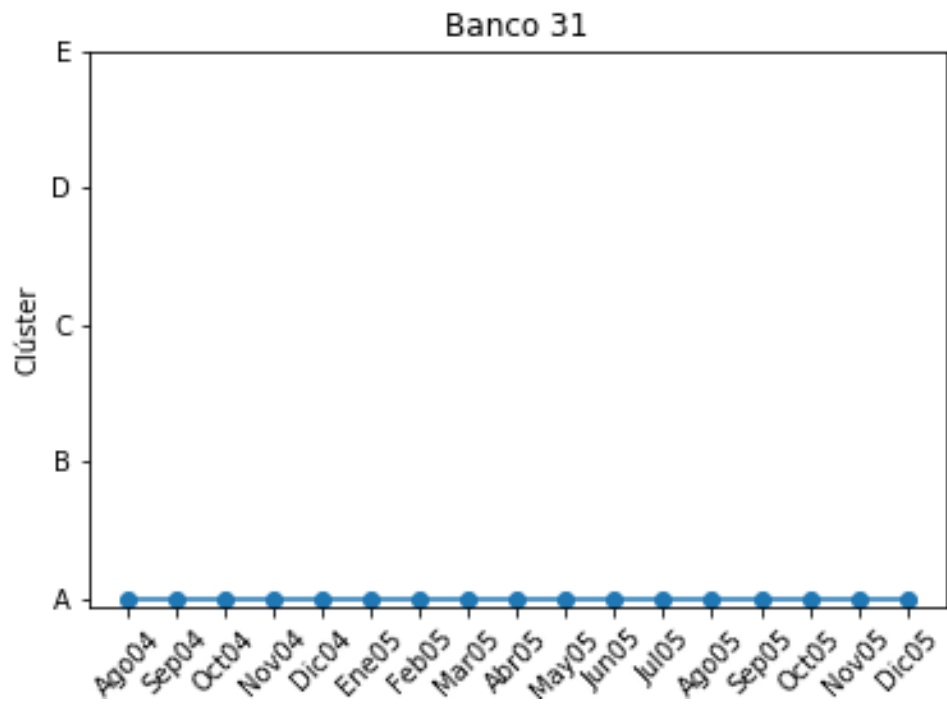


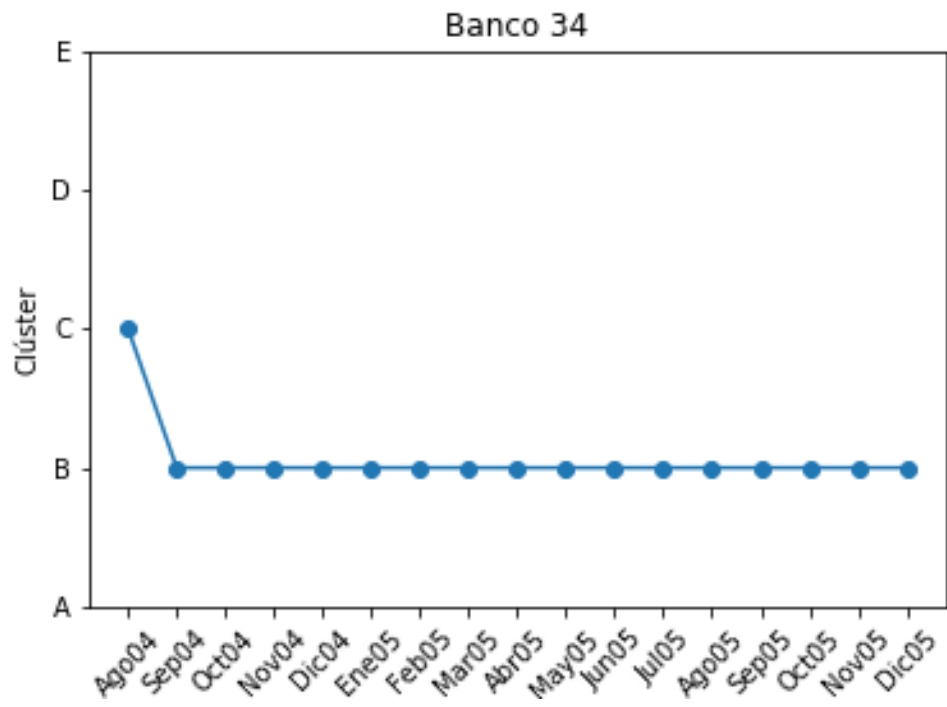
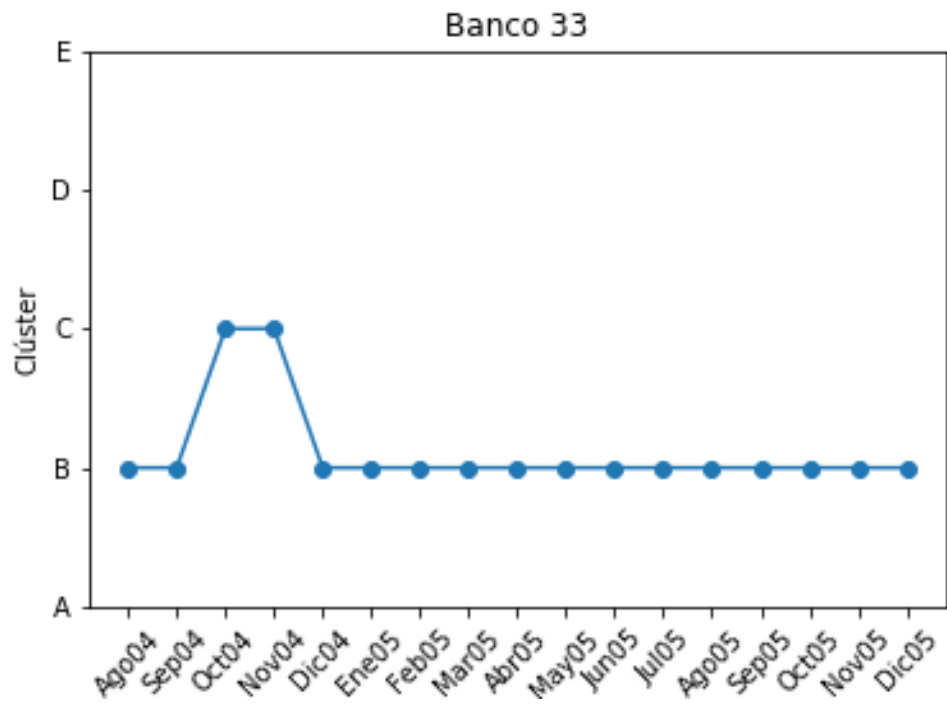


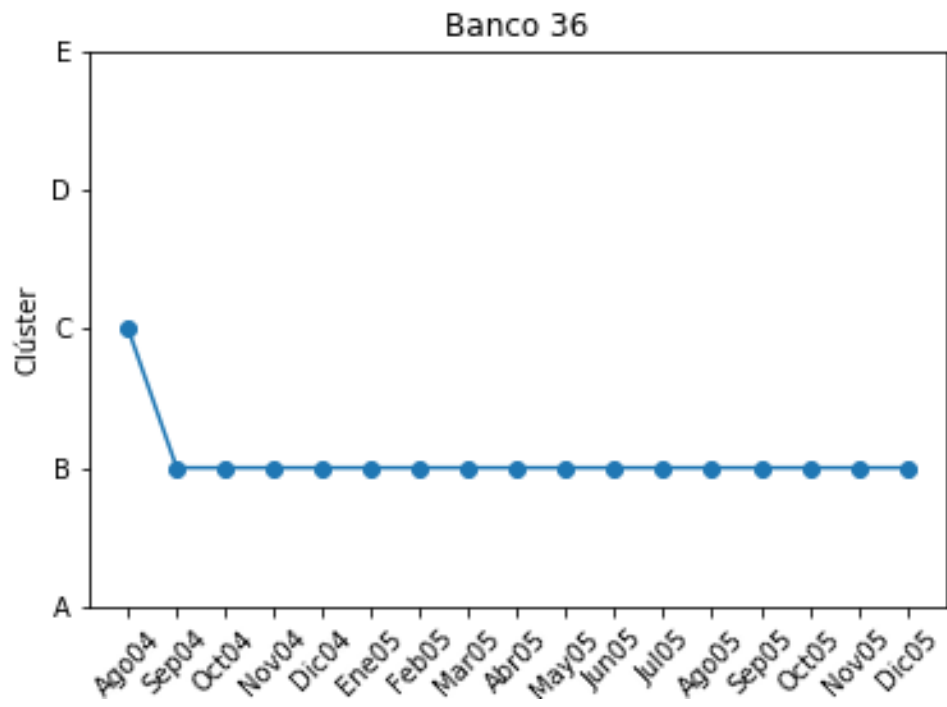
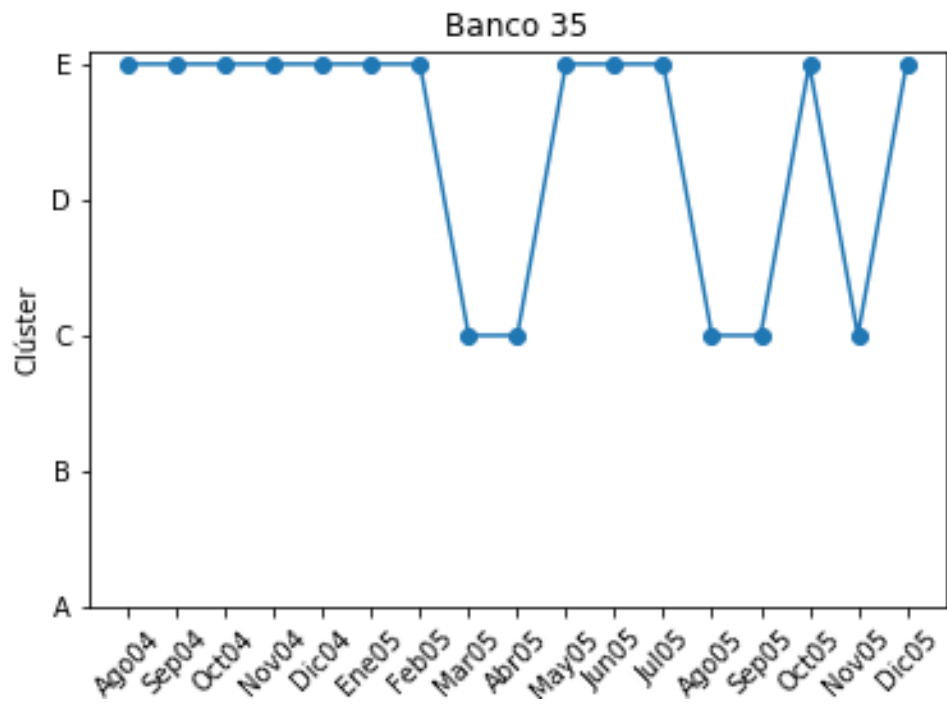


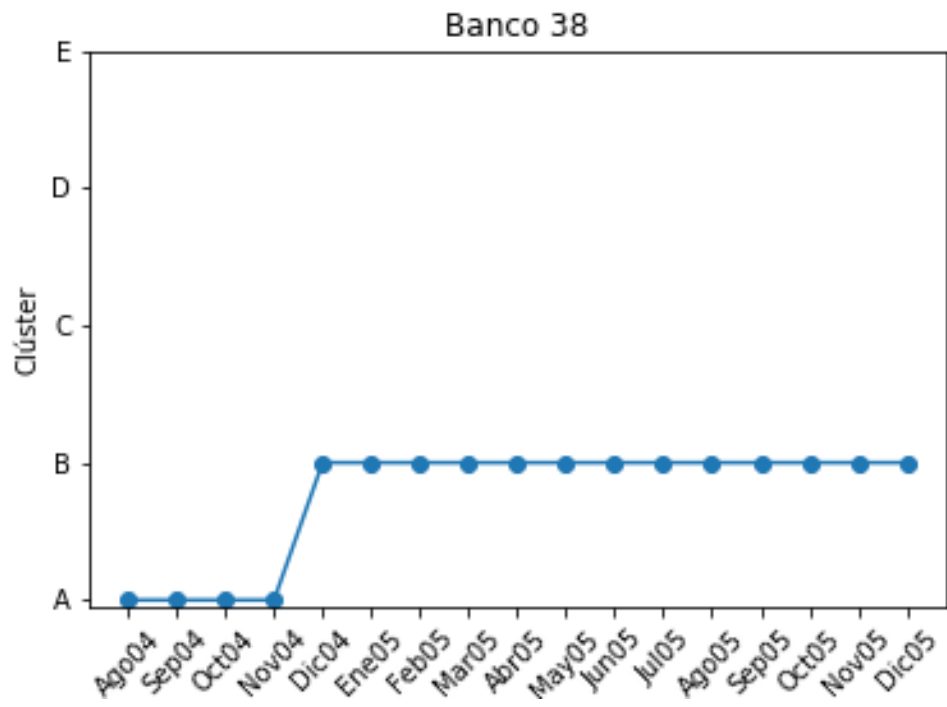
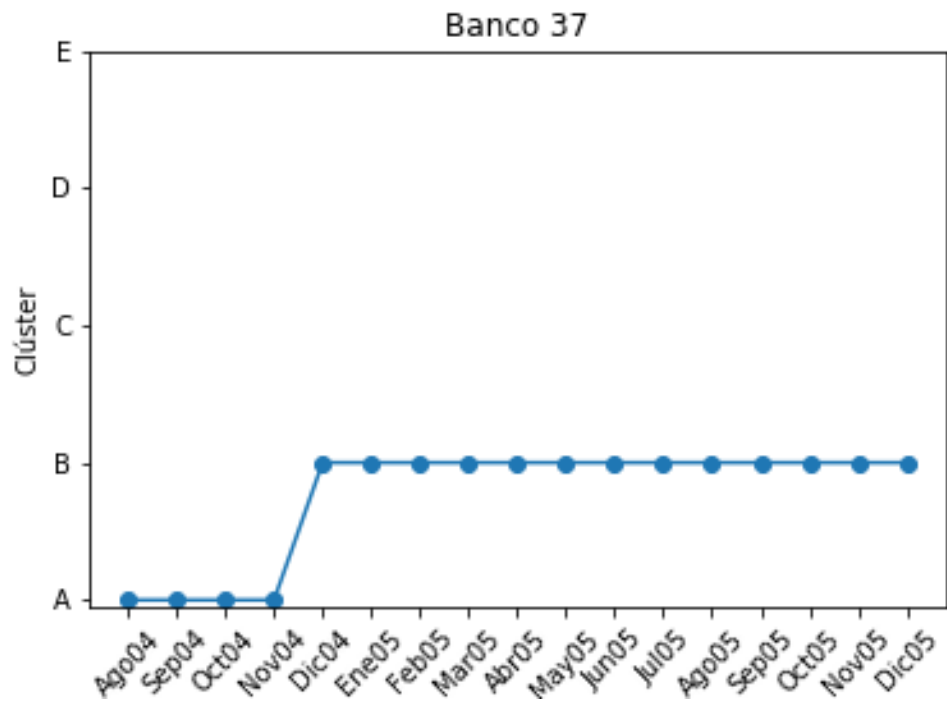








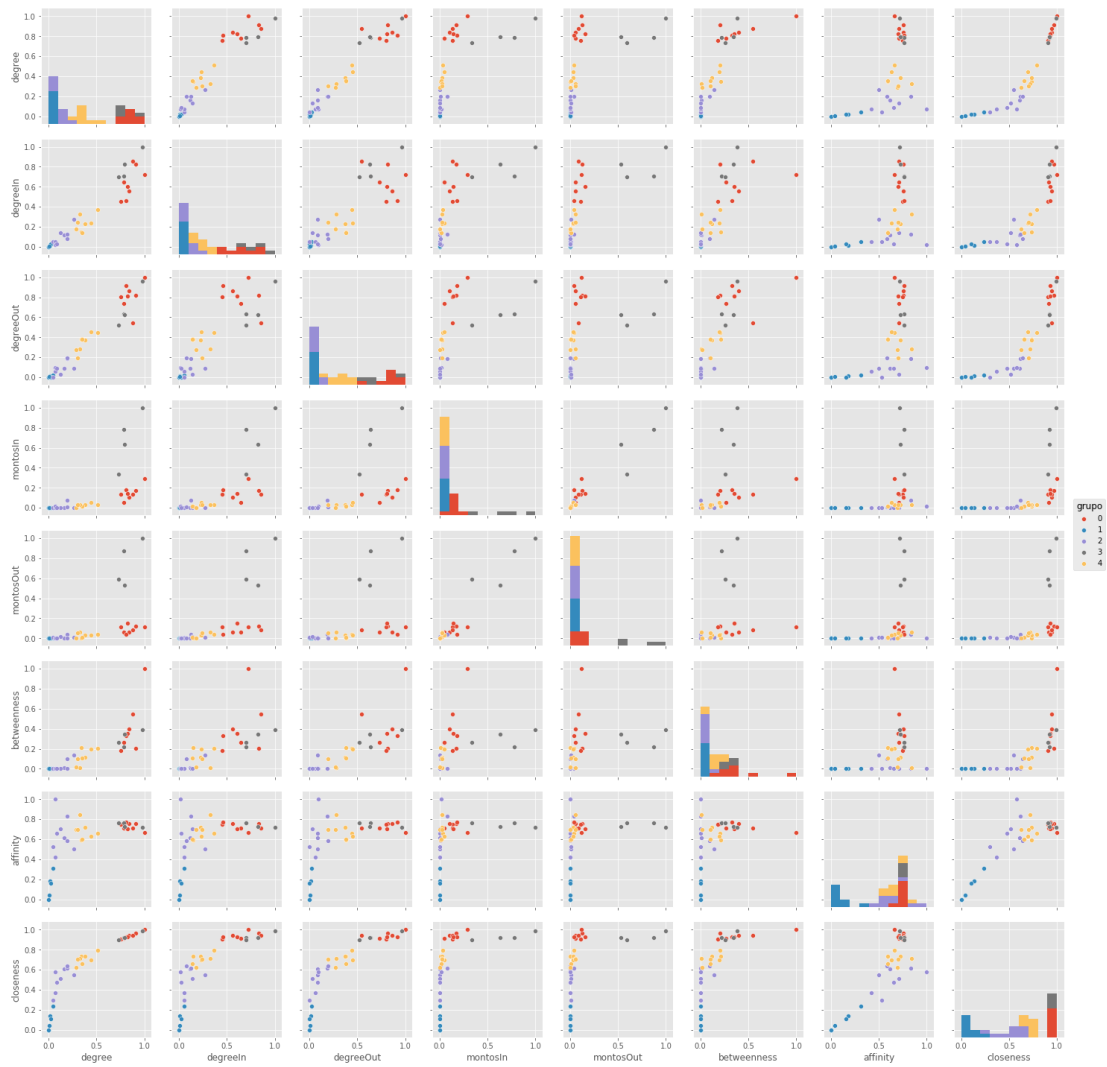




ANEXO 5

GRAFICA MULTIDIMENSIONAL POR MES

AGOSTO 2004



SEPTIEMBRE 2004



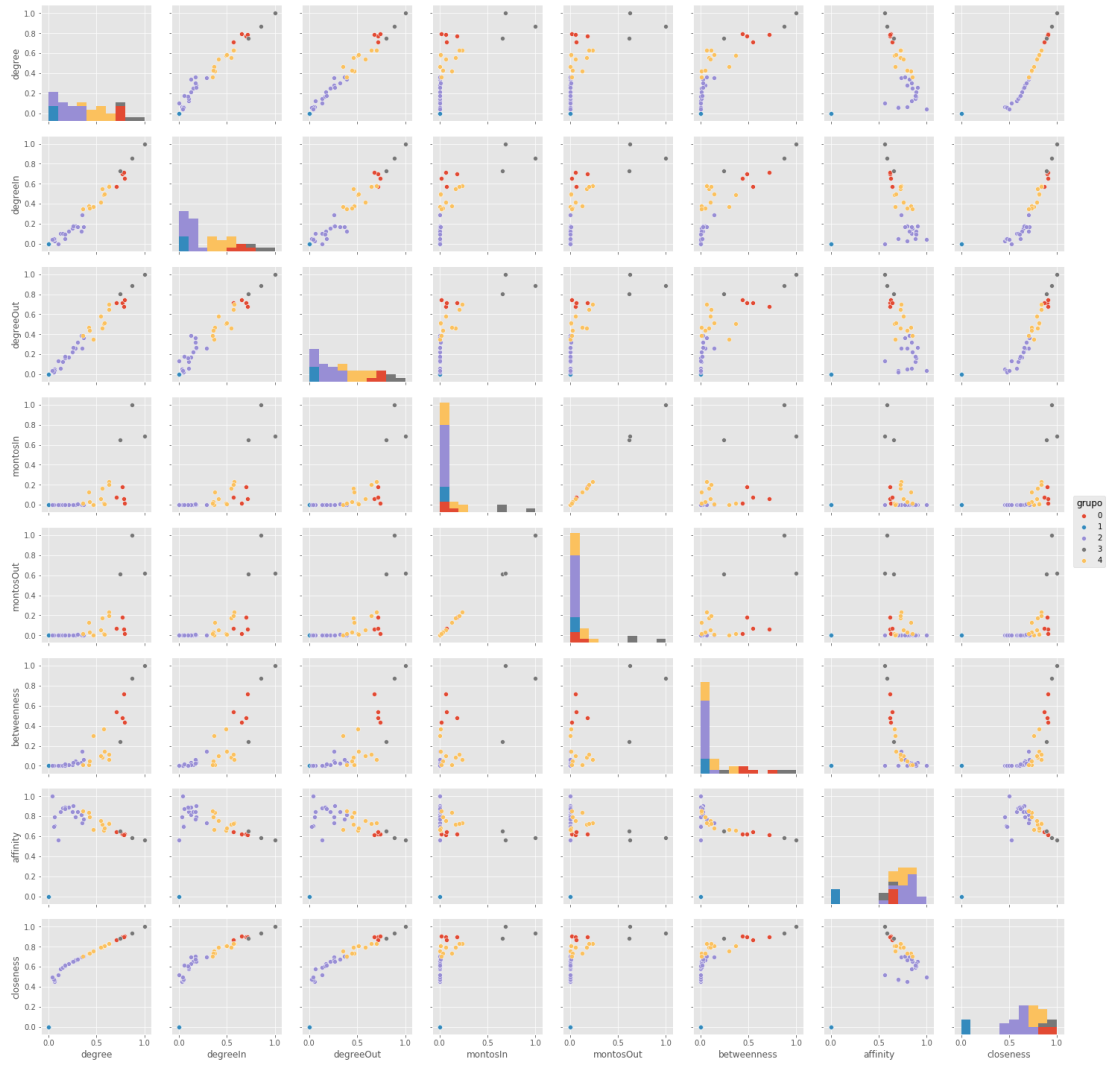
OCTUBRE 2004



NOVIEMBRE 2004



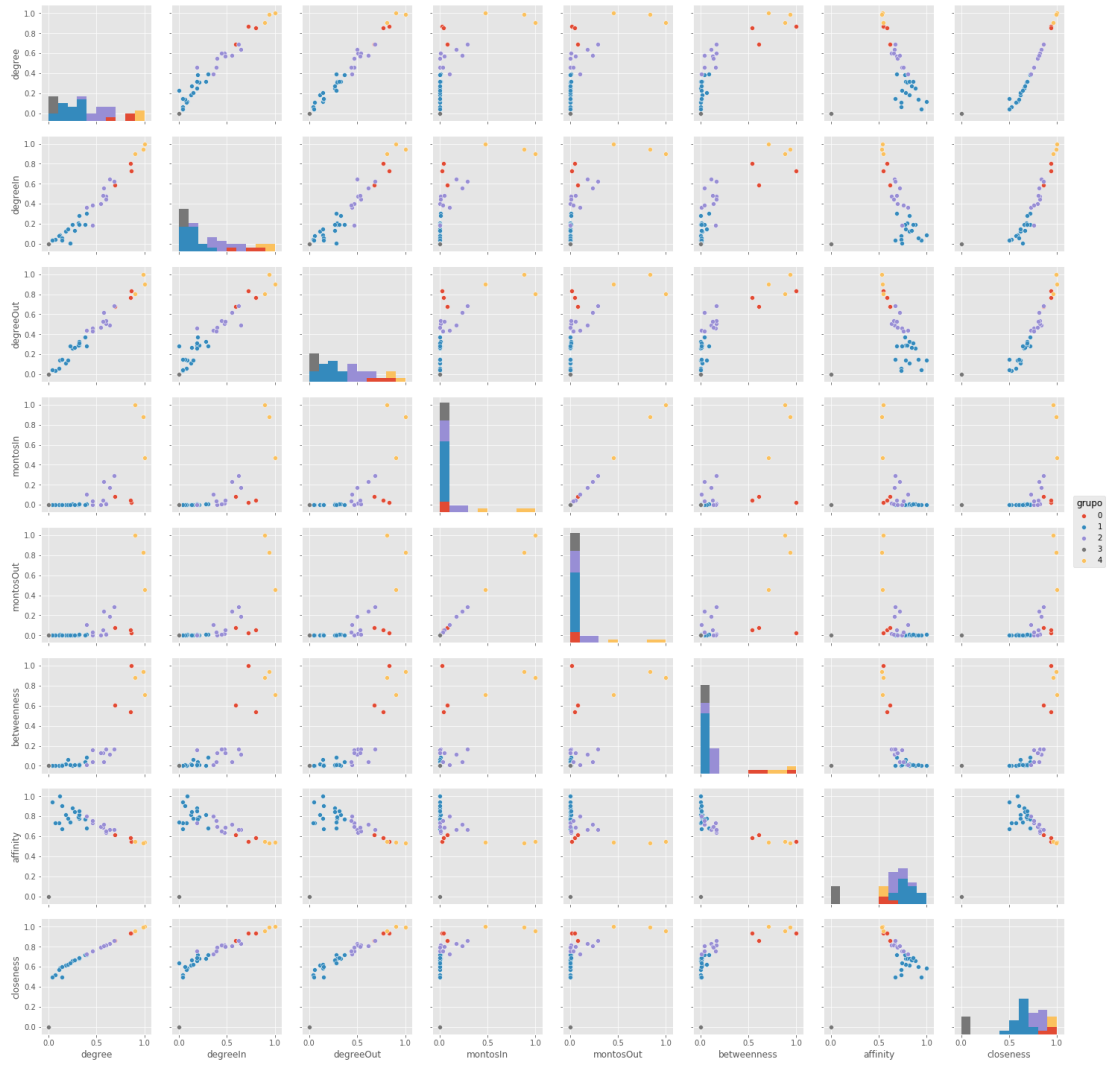
DICIEMBRE 2004



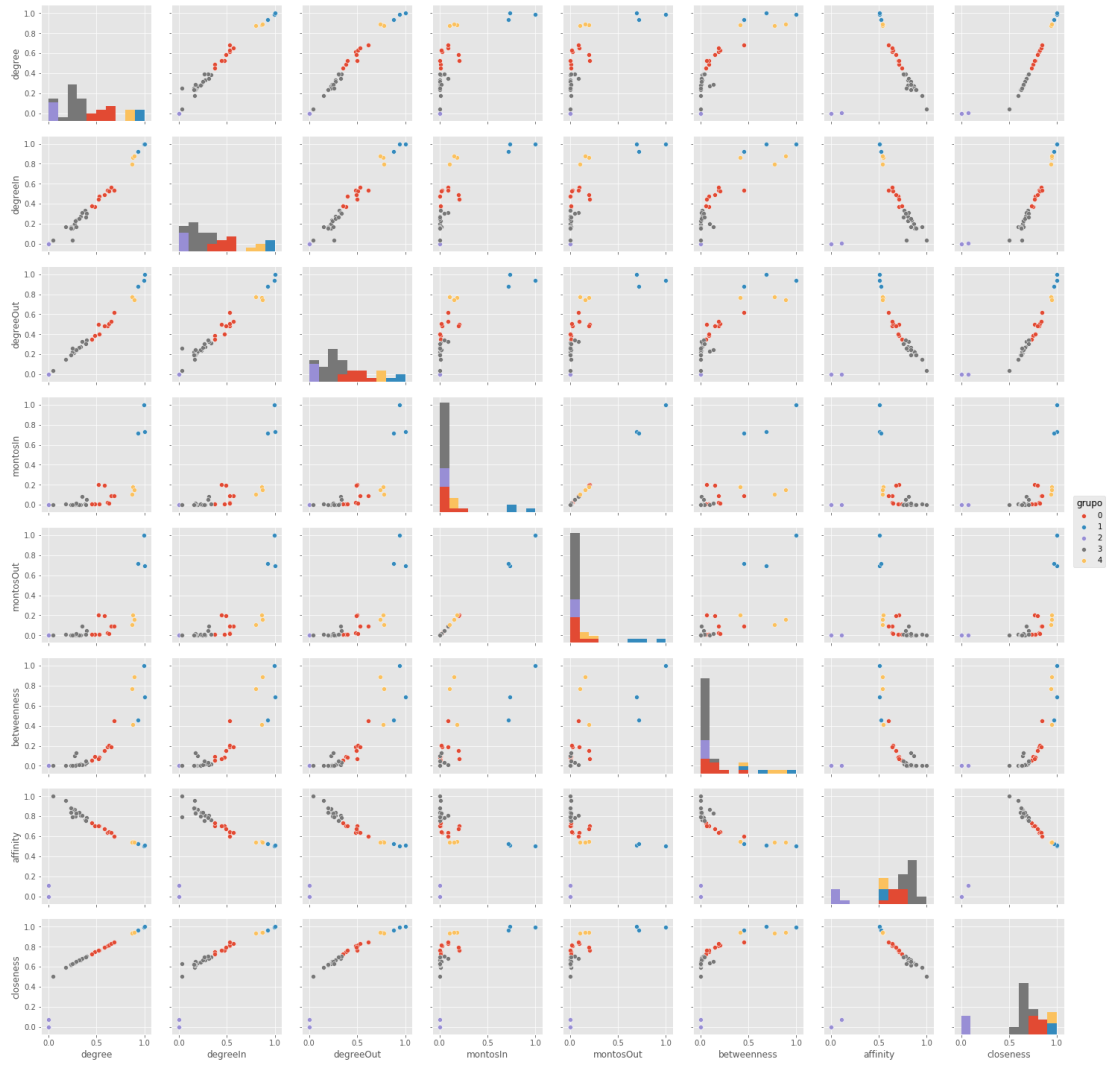
ENERO 2005



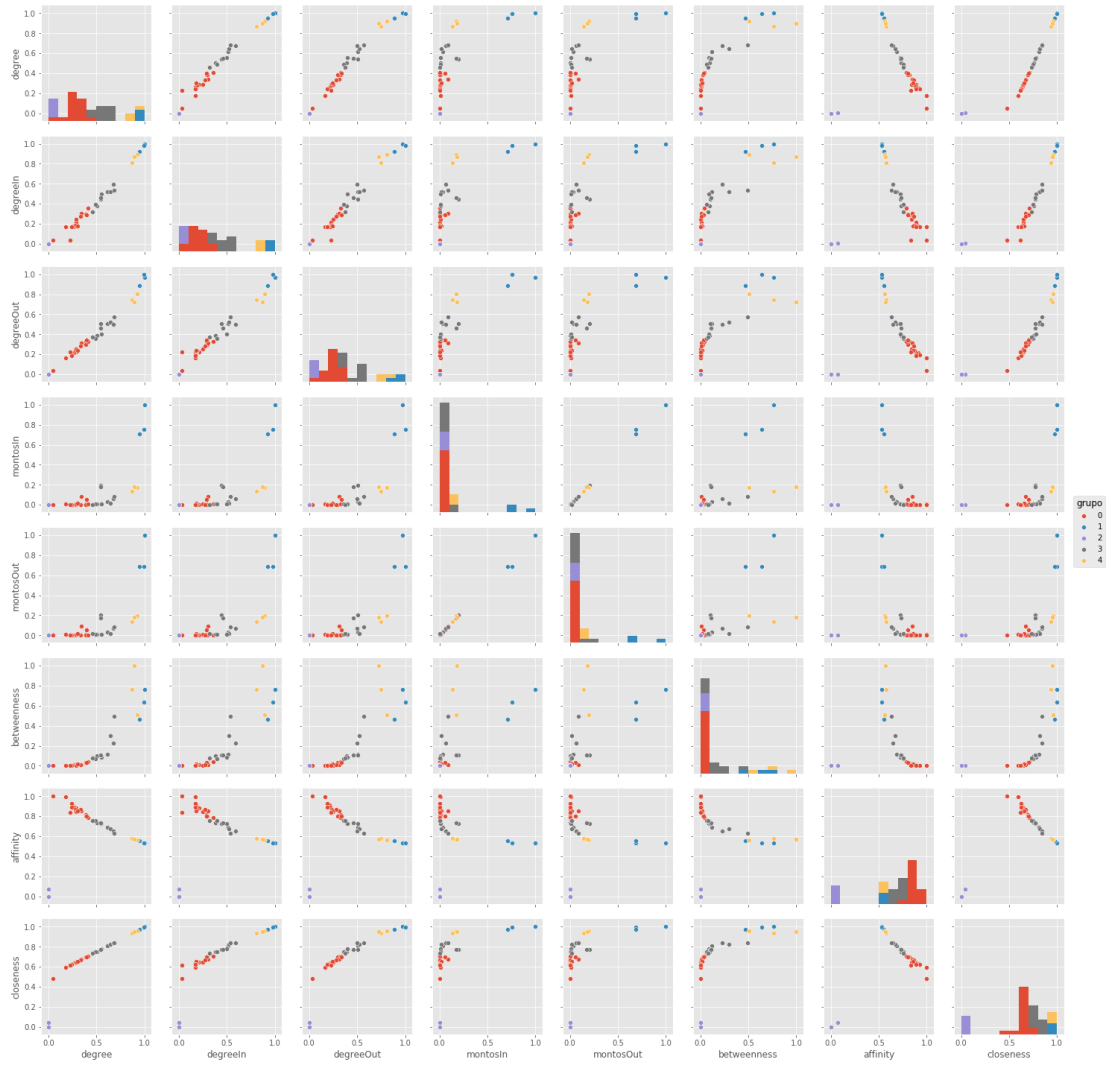
FEBRERO 2005



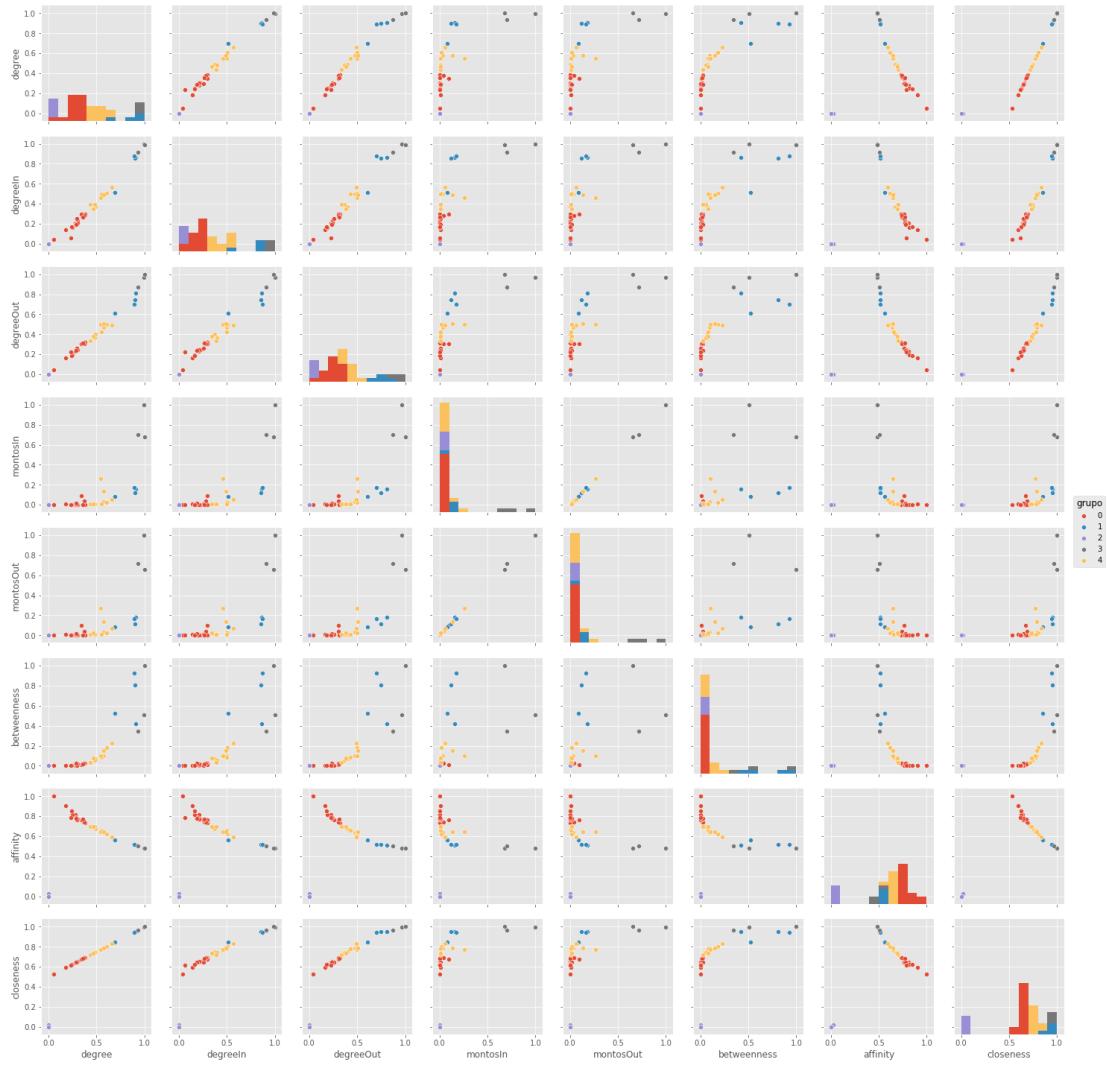
MARZO 2005



ABRIL 2005



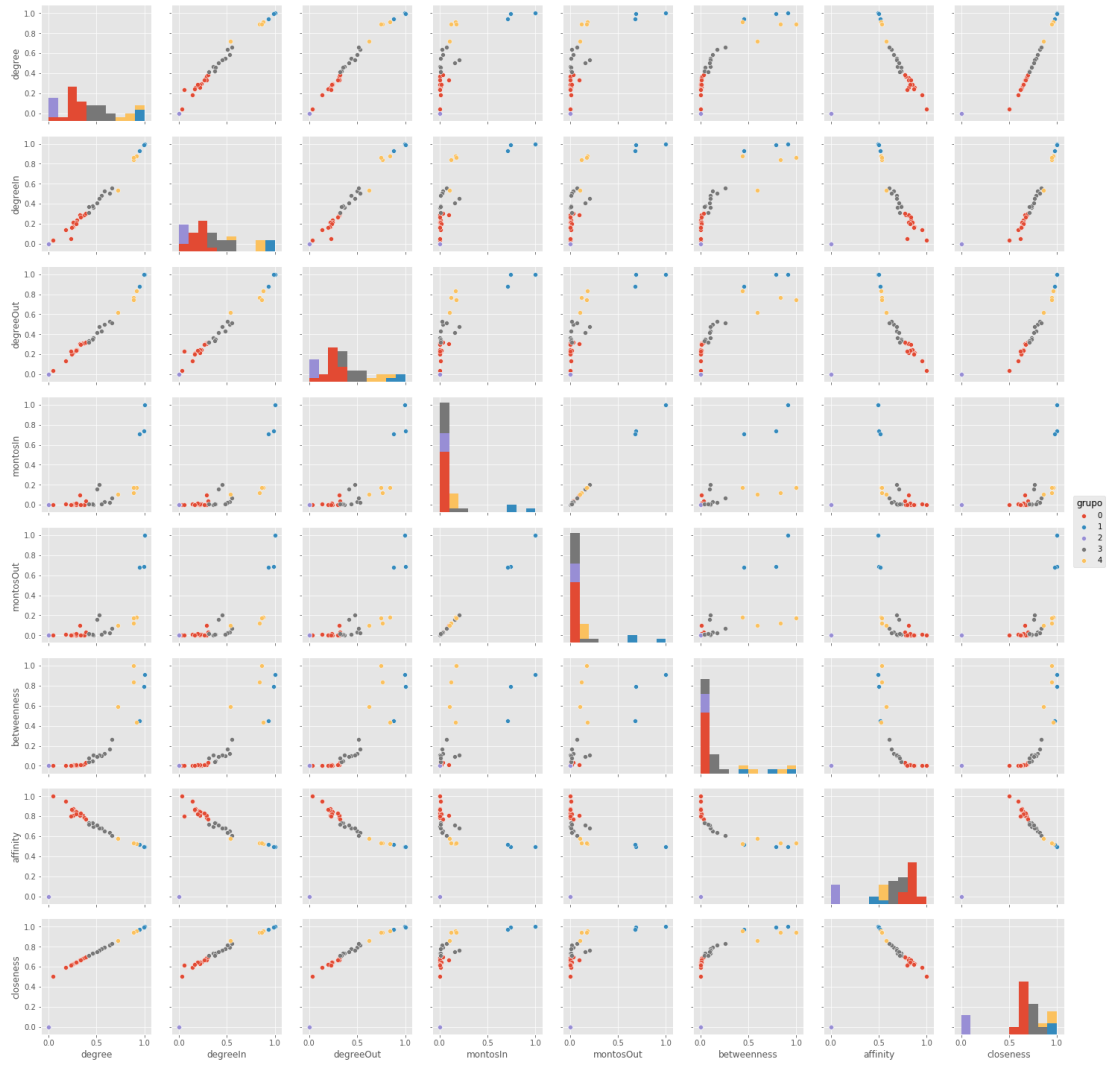
MAYO 2005



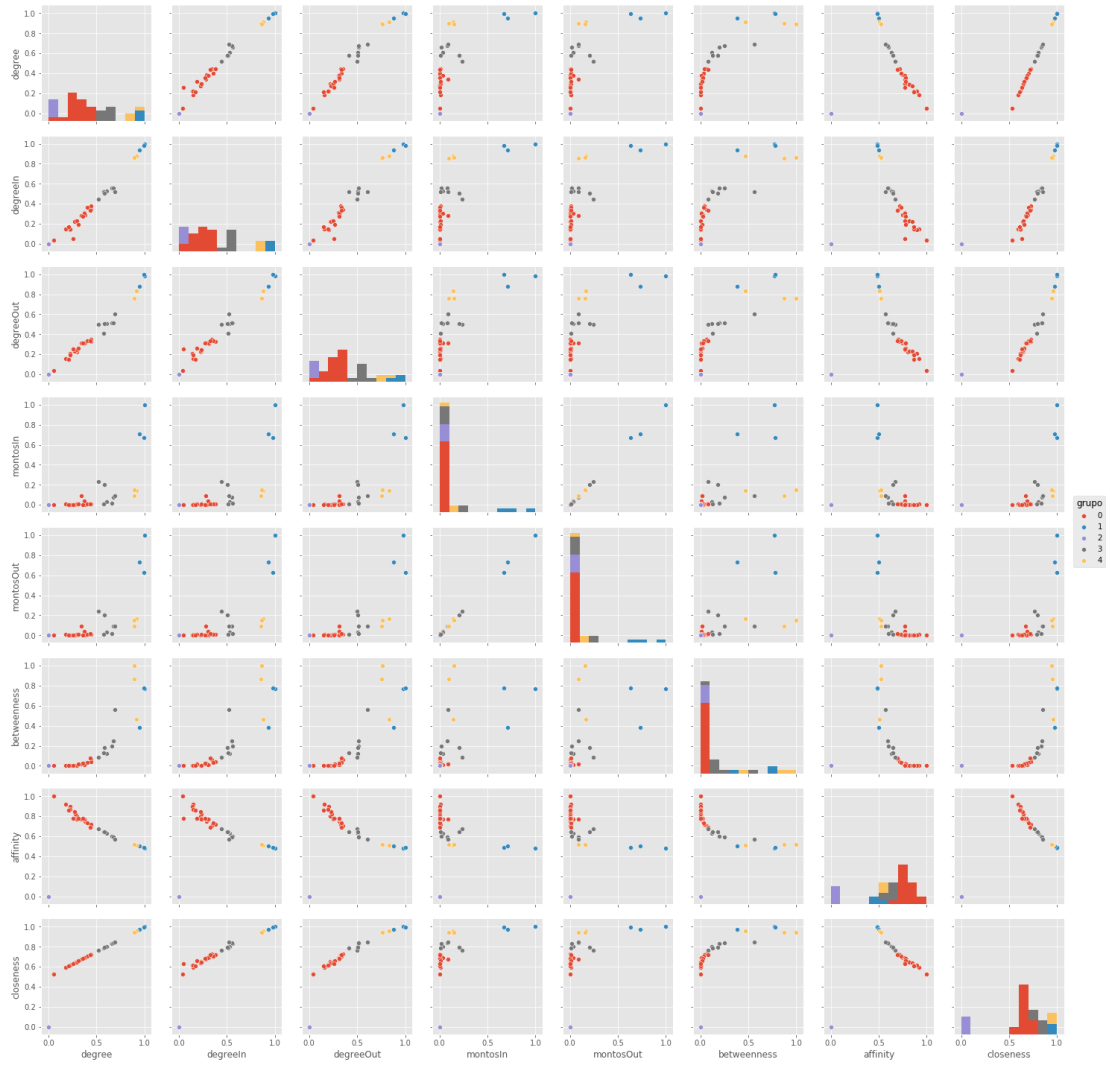
JUNIO 2005



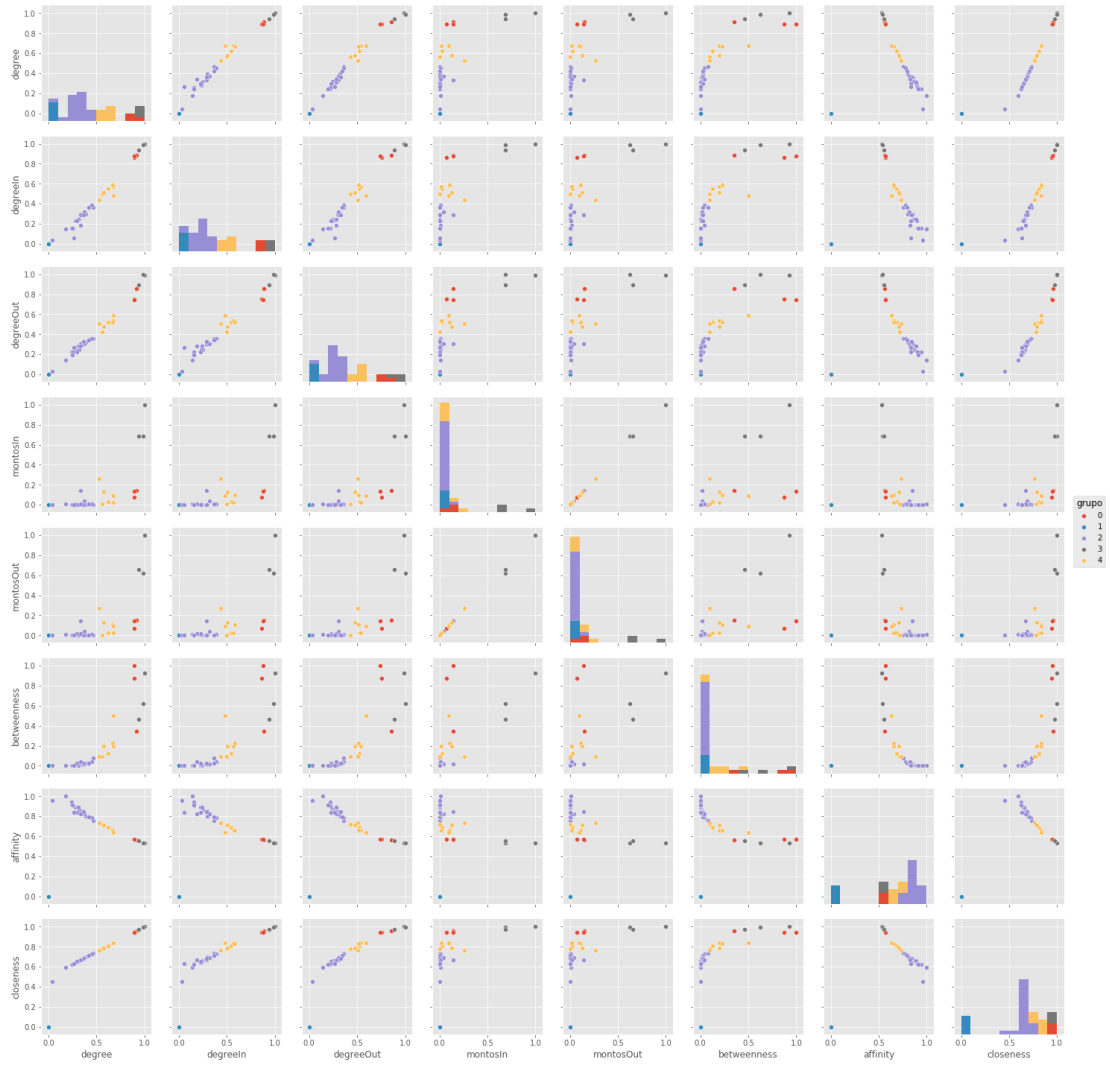
JULIO 2005



AGOSTO 2005



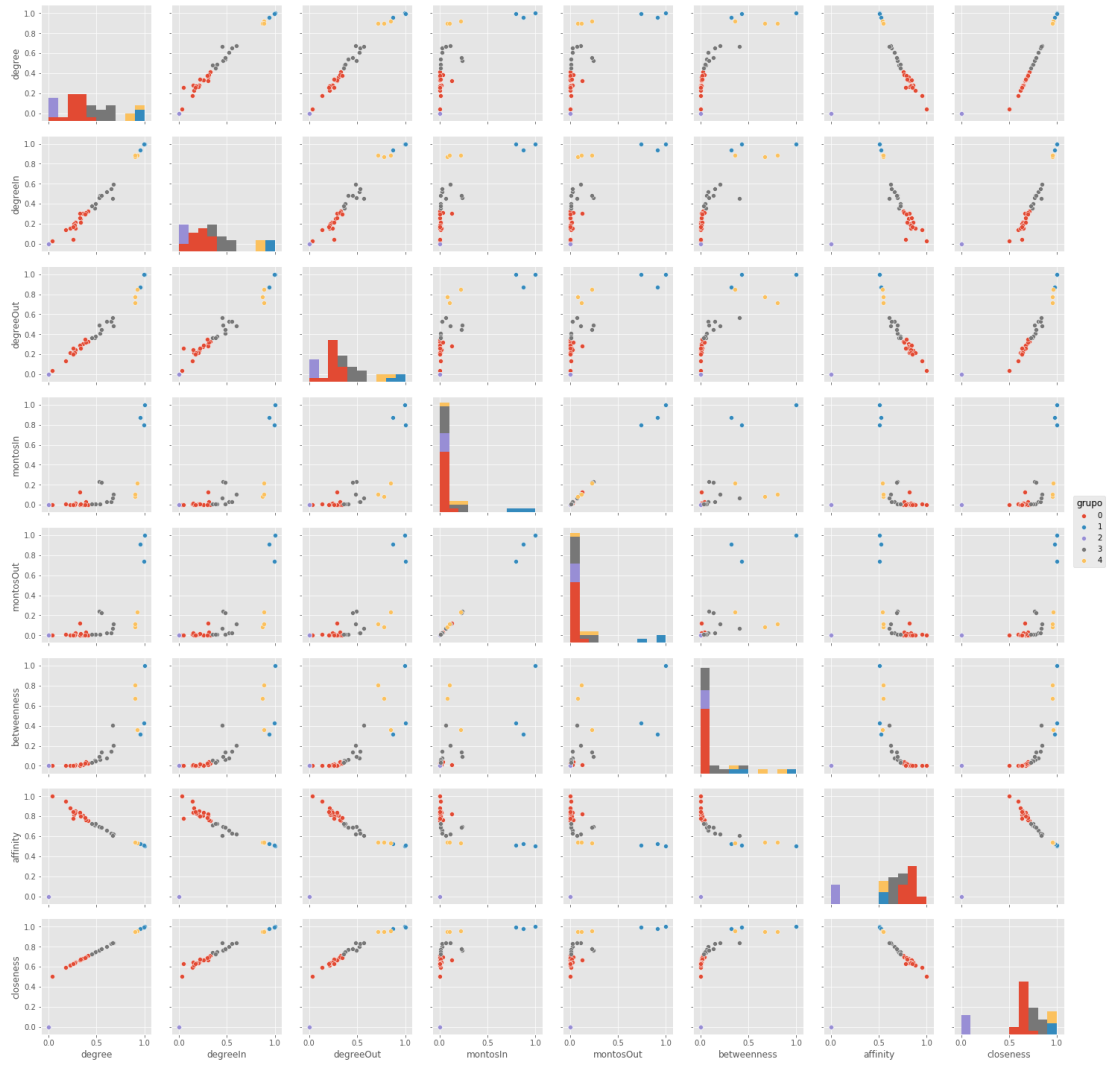
SEPTIEMBRE 2005



OCTUBRE 2005



NOVIEMBRE 2005



DICIEMBRE 2005



Bibliografía

Amoedo, D. (9 de diciembre de 2017). *Spyder, un potente entorno de desarrollo interactivo para Python*. [Entrada de Blog]. Recuperado de: <https://ubunlog.com/spyder-entorno-desarrollo-python/>

Banco de México. (2017). Obtenido de Divulgación: <http://www.banxico.org.mx/divulgacion/divulgacion.html>

Banco de México. (2017). Obtenido de Banco de México: <http://www.banxico.org.mx/home-inf.html>

Bravo-Benitez, B., Kabadjova, B. A., & Martinez-Jaramillo, S. (2014). *Centrality Measurement of the Mexican Large Value Payments System from the Perspective of Multiplex Networks, [42-44]*. Recuperado de: <https://www.researchgate.net/publication/263889322>

Cáceres Tello, J. (2017). *Universidad de Alcalá. Reconocimiento de patrones y el aprendizaje no supervisado*. Recuperado de: https://portal.uah.es/portal/page/portal/GP_EPD/PG-MA-PROF/OLD_PG-PROF-138886%202008-07-14%2010-07-31/TAB4348465/TAB4348469/TAB4348477/Articulo_Reconocimiento_Patrones_JesusCaceres_CISCI_06.pdf

Del Cerro Sánchez, T., & Novalbos Laina, P. (2005). *Minería de datos*. Recuperado de: <http://www.it.uc3m.es/jvillena/irc/practicas/13-14/06md.pdf>

Ester, M., Krieger, H.-P., Sander, J., & Xu, X. (1996). *Revista AAAI. A density-Based Algorithm for discovering clusters*. Recuperado de <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>

García Cambronero, C., & Gómez Moreno, I. (2017). UC3M. *Algoritmo de aprendizaje: KNN & KMeans [6-8]*. Recuperado de <http://www.it.uc3m.es/jvillena/irc/practicas/08-09/06.pdf>

Gallardo. (2017). *Métodos Jerárquicos de Análisis*. Recuperado de: <http://www.ugr.es/~gallardo/pdf/cluster-3.pdf>

González Bernal, J. A. (2017). Obtenido de https://ccc.inaoep.mx/~jagonzalez/AI/Sesion13_Data_Mining.pdf

González Duque, R. (2017). *Python para todos*. Recuperado de: http://www.utic.edu.py/citil/images/Manuales/Python_para_todos.pdf

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. elservier.

Ingeniesia Desarrollo Cloud, S. (2015). *¿Qué es el data mining? La definición de la minería de datos*. Recuperado de <https://clinic-cloud.com/blog/data-mining-que-es-definicion-mineria-de-datos/>

Intelligence, T. D. (2017). <http://www.iiia.csic.es>. Obtenido de <http://www.iiia.csic.es/udt/files/DataMining.pdf>

López Briega, R. (28 de Mayo 2014). *Python - Librerías esenciales para el análisis de datos*. [Entrada de Blog]. Recuperado de: <https://relopezbriega.github.io/blog/2014/05/28/python-librerias-esenciales-para-el-analisis-de-datos/>

López Domínguez, I. (25 de Junio 2017). *expansion.com*. Recuperado de: <http://www.expansion.com/diccionario-economico/riesgo-sistemico.html>

Morales, E., & Escalante, H. J. (2017). *Clustering*. Recuperado de: <https://ccc.inaoep.mx/~emorales/Cursos/NvoAprend/Acetatos/clustering.pdf>

Paniagua Arís, E., & López Ayuso, B. (2007). *La gestión tecnológica del conocimiento*. edit.um.

Pascual. (2017). *Algoritmos de agrupamiento*. Universidad de Oriente, Universidad Jaime I. Recuperado de: http://marmota.dlsi.uji.es/WebBIB/papers/2007/1_Pascual-MIA-2007.pdf

Sánchez-Montañés, M., Lago, L., & González, A. (2017). *Métodos Avanzados en aprendizaje Artificial: Teoría y aplicaciones a problemas de predicción*. Recuperado de: http://arantxa.ii.uam.es/~msanchez/docencia/maaa/transparencias/Intro_1112.pdf

Zúñiga, R. (2008). *Introducción al Uso de Matlab. Conceptos Generales*. Recuperado de: <http://www.geociencias.unam.mx/~bole/eboletin/Matlabintro0408.pdf>