



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS DE CONGLOMERADOS Y DE REDES
PARA LA IDENTIFICACIÓN DE MODELOS DE
NEGOCIOS
EN EL SISTEMA BANCARIO MEXICANO

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

PRESENTA:

LUIS ONÉSIMO LEONARDO ESCOBAR FARFÁN

ASESOR:

DR. SERAFIN MARTÍNEZ JARAMILLO



CIUDAD UNIVERSITARIA, CDMX, 2019



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

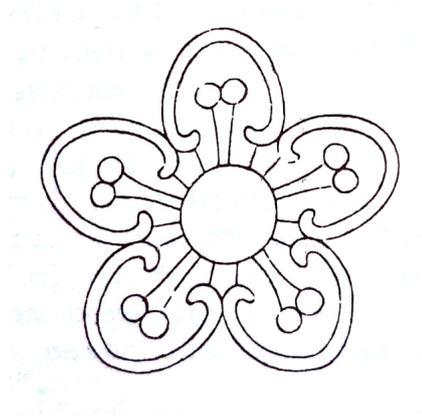
Datos del Jurado:

1. *Datos del alumno*
Escobar
Farfán
Luis Onésimo Leonardo
(044)(55)45254570
Universidad Nacional Autónoma de
México
Facultad de Ciencias
Actuaría
308202245
2. *Datos del tutor*
Dr.
Serafín
Martínez
Jaramillo
3. *Datos del sinodal 1*
Dra.
Claudia Orquídea
López
Soto
4. *Datos del sinodal 2*
Dra.
Bibiana
Obregón
Quintana
5. *Datos del sinodal 3*
Dr.
Antonio
Carrillo
Ledesma
6. *Datos del sinodal 4*
M. en C.
David Chaffrey
Moreno
Fernández
7. *Datos del trabajo escrito*
Análisis de conglomerados y de redes para
la identificación de modelos de negocios
en el sistema bancario mexicano
161 p
2019

*Ma camo tepalhuan tlaxitla tiyecan
totechco cah tomaquixtizque*

*No más tutela ni vivir despreciados
en nosotros hay fuerza para hacernos libres.*

Fidencio Villanueva.



*Para Ana, One, Pablo y Jonathan.
Sin su apoyo y ejemplo, nada es posible.*

A mi familia.

A Ingrid, Nelly, Emmanuel, David y Darío.

A Serafín.

A la UNAM.

Reconocimientos

Agradezco grandemente a la Mtra. **Jessica Maricarmen Serrano Bandala**, pues sin su apoyo e inspiración la presente tesis no hubiera sido posible.

También quiero agradecer al Banco de México, especialmente a la Dirección General de Estabilidad Financiera, por proporcionarme los datos necesarios para realizar éste trabajo.

Declaración

Declaro conocer el Código de Ética de la Universidad Nacional Autónoma de México, plasmado en la Legislación Universitaria. Con base en las definiciones de integridad y honestidad ahí especificadas, aseguro mediante mi firma al calce que el presente trabajo es original y enteramente de mi autoría. Todas las citas de, o referencias a, la obra de otros autores aparecen debida y adecuadamente señaladas, así como acreditadas mediante los recursos editoriales convencionales

Las opiniones expresadas en la presente tesis son mi responsabilidad y no representan la opinión del Banco de México.

LUIS ONÉSIMO LEONARDO ESCOBAR FARFÁN. CIUDAD UNIVERSITARIA,
CDMX, 2019

Índice general

Índice de figuras	xI
1. Introducción	1
1.1. Presentación	1
1.2. Motivación del trabajo	5
1.3. Objetivos del trabajo	7
1.4. Contribución del trabajo	7
1.5. Datos	8
1.6. Infraestructura de cómputo y software utilizado	9
1.7. Estructura de la tesis	9
2. Modelos de negocios	11
2.1. Antecedentes regulatorios	11
2.2. ¿Qué entendemos por “Modelos de negocio”?	13
2.3. ¿Por qué es importante monitorear los modelos de negocios que siguen los bancos?	14
2.4. Matemáticas para identificar modelos de negocios	15
3. Modelos de Conglomerados y de Reducción de Dimensión	19
3.1. Datos	20
3.2. Reducción de Dimensionalidad	23
3.3. Métodos de agrupamiento	27
3.4. K-Medias Reducido y K-Medias Factorial	33
3.5. Resultados	37
4. Árboles de decisión y Bosques aleatorios	49
4.1. Árboles de decisión	52
4.2. Bosques Aleatorios	57
4.3. Resultados	59
5. Redes Financieras	69
5.1. Datos	71
5.2. Características generales de las redes	71

ÍNDICE GENERAL

5.3. Análisis estructural de Redes	72
5.4. Principales modelos generativos de redes	82
5.5. Resultados	94
6. Modelos de Bloques Estocásticos	105
6.1. Breve introducción a la estadística Bayesiana	105
6.2. Modelos de Bloques Estocásticos	119
6.3. Resultados	123
7. Discusión de resultados y conclusiones	129
7.1. Discusión	129
7.2. Conclusiones	140
A. Apéndice A	143
A.1. Validación cruzada y Remuestreo (<i>Remuestreo</i>)	143
B. Apéndice B	147
B.1. Medidas de red	147
Bibliografía	155

Índice de figuras

3.1. Ejemplo de una gráfica de codo.	27
3.2. Evolución temporal de los criterios de ajuste del método FKM	39
3.3. Evolución temporal de los criterios de ajuste del método RKM	39
3.4. Visualizaciones de los ajustes con FKM para el nivel uno de desagregación para la última fecha de estudio.	40
3.5. Visualizaciones de los ajustes con FKM para el nivel dos de desagregación para la última fecha de estudio.	41
3.6. Visualizaciones de los ajustes con FKM para el nivel tres de desagregación para la última fecha de estudio.	41
3.7. Ajuste FKM N1 SinCol.	42
3.8. Ajuste FKM N2 SinCol.	43
3.9. Ajuste FKM N3 SinCol.	43
3.10. Visualizaciones de los ajustes con RKM para el nivel uno de desagregación para la última fecha de estudio.	44
3.11. Visualizaciones de los ajustes con RKM para el nivel dos de desagregación para la última fecha de estudio.	45
3.12. Visualizaciones de los ajustes con RKM para el nivel tres de desagregación para la última fecha de estudio.	45
3.13. Ajuste RKM N1 SinCol.	46
3.14. Ajuste RKM N2 SinCol.	47
3.15. Ajuste RKM N3 SinCol.	47
4.1. Ejemplo de overfitting.	51
4.2. Ejemplo de un árbol de regresión.	54
4.3. Ejemplo de un árbol de clasificación.	56
4.4. Variables más importantes de acuerdo al método de bosque aleatorio para llegar a la agrupación RKM N1 SinCol.	60
4.5. Variables más importantes de acuerdo al método de bosque aleatorio para llegar a la agrupación RKM N2 SinCol.	62
4.6. Variables más importantes de acuerdo al método de bosque aleatorio para llegar a la agrupación RKM N3 SinCol.	63

ÍNDICE DE FIGURAS

4.7. Distribución del porcentaje de la cuenta general de balance que representan las variables más importantes según el método de bosques aleatorios para el primer nivel del catálogo mínimo.	64
4.8. Distribución del porcentaje de la cuenta general de balance que representan las variables más importantes según el método de bosques aleatorios para el segundo nivel del catálogo mínimo.	66
4.9. Distribución del porcentaje de la cuenta general de balance que representan las variables más importantes según el método de bosques aleatorios para el tercer nivel del catálogo mínimo.	67
5.1. Visualizaciones de la red interbancaria Mexicana, cada círculo representa un banco mexicano, mientras que las ligas representan monto de dinero potencialmente perdido dado el estrés del sistema.	70
5.2. Comparación entre las colas de las distribuciones de grado obtenidas de redes aleatorias (Poisson) y redes libres de escala (Ley potencia)	87
5.3. Error incurrido por el modelo Core-Periphery de Craig y von Peter en la red interbancaria mexicana.	92
5.4. Diferencias de asignación al centro o a la periferia entre el modelo de Craig y von Peter y el método by Surprise para una muestra de bancos en México.	94
5.5. Promedio del sistema bancario de las principales medidas de red a lo largo del período de estudio (1).	96
5.6. Promedio del sistema bancario de las principales medidas de red a lo largo del período de estudio (2).	97
5.7. Centralidad por Cercanía y por DebtRank para las 10 instituciones con los valores más altos y más pequeños.	98
5.8. Centralidad por Fuerza y Grado para las 10 instituciones con los valores más altos y más pequeños.	99
5.9. Índice Herfindahl-Hirschman como prestamista y prestatario para las 10 instituciones con los valores más altos y más pequeños.	100
5.10. Centralidad por Interinidad y PageRank para las 10 instituciones con los valores más altos y más pequeños.	101
5.11. Promedio del sistema bancario de las principales medidas de red diferenciando por el centro y la periferia obtenidos con el método by Surprise (1).	102
5.12. Promedio del sistema bancario de las principales medidas de red diferenciando por el centro y la periferia obtenidos con el método by Surprise (2).	103
6.1. Distribución del grado para la red interbancaria Mexicana.	124
6.2. Gráfica aluvial de los movimientos entre los grupos encontrados con el Modelo de Bloques Estocásticos de las instituciones bancarias en México.	125
6.3. Importancia de las medidas de centralidad para llegar a la agrupación obtenida por SBM según el método de bosque aleatorio.	126

6.4. Distribución del grado y la fuerza del sistema bancario diferenciando por los grupos obtenidos con el método de bosque aleatorio.	127
6.5. Distribución de otras medidas de centralidad del sistema bancario diferenciando por los grupos obtenidos con el método de bosque aleatorio .	128
7.1. Número de conexiones entre los grupos obtenidos con el modelo de bloques estocásticos como porcentaje del total de ligas para la última red del periodo de estudio.	134
7.2. Comparación de los grupos asignados por SBM y por RKM a los bancos para el primer nivel del catálogo mínimo.	136
7.3. Comparación de los grupos asignados por SBM y por RKM a los bancos para el segundo nivel del catálogo mínimo.	137
7.4. Comparación de los grupos asignados por SBM y por RKM a los bancos para el tercer nivel del catálogo mínimo.	138
7.5. Valor en las cuentas más importantes del ajuste RKM N2 diferenciando las dos instituciones que son agrupadas en un grupo no existente para la fecha anterior del mismo ajuste	139

Introducción

1.1. Presentación

Los sistemas bancarios de todo el mundo se han vuelto más complejos con el paso del tiempo. Situaciones como la globalización, los cambios regulatorios y el crecimiento del número de instituciones han propiciado que los intermediarios financieros mundiales interactúen en nuevas “dimensiones” (mercados, países, jurisdicciones, productos), creando así relaciones con efectos no detectables por la regulación convencional, efectos con implicaciones *sistémicas*.

De manera general, definimos al *riesgo sistémico* como aquel que de materializarse generaría un “efecto dominó” que afectaría, de manera directa o indirecta, a una gran proporción de los participantes de un sistema, sin importar la naturaleza del mismo.

Dentro de todo sistema existen participantes “centrales” (veremos más adelante el concepto de centralidad), cuyo funcionamiento tiene efectos importantes sobre la actividad y la estabilidad del sistema en conjunto. Dichos participantes tienen el potencial de entorpecer la operatividad del resto de los participantes del sistema en caso de que dejara de realizar sus actividades con normalidad, o si detuviera por completo su participación en el sistema.

Por ejemplo, en una red de abasto eléctrico, existen plantas que dejarían sin electricidad a una gran porción de la red si dejaran de funcionar. En [Fornito et al. \(2016\)](#) señalan diversos estudios donde se ha demostrado que ciertas alteraciones estructurales en el cerebro tienen efectos en su funcionamiento y tienen una relación directa con enfermedades neurológicas.

En el contexto del presente trabajo, el sistema con el que trabajaremos es el sistema bancario Mexicano. En este, el riesgo sistémico se traduce en las afectaciones que tendría la quiebra o el estrés financiero de un banco en el resto de las instituciones bancarias del país, e incluso, sobre participantes de la economía real (personas morales, personas físicas, aunque estos agentes están fuera del alcance del presente trabajo), así como sobre otro tipo de intermediarios financieros (casas de bolsa, SOFOMEs¹, fondos de

¹Las Sociedades Financieras de Objeto Múltiple son un tipo de intermediario financiero con la

1. INTRODUCCIÓN

inversión, etcétera), provocando pérdidas significativas para la economía completa.

Después de la crisis financiera mundial que inició en 2008, surgieron preocupaciones sobre la efectividad de la regulación vigente que aplicaba a las instituciones financieras, especialmente a los bancos, que hasta la fecha siguen siendo los intermediarios financieros de mayor tamaño y de mayor importancia sistémica, dado el alcance de sus operaciones.

El enfoque de la mayoría de las pruebas de estrés y de las medidas regulatorias de la época fallaron notoriamente en la detección de situaciones de riesgo derivadas de la actividad habitual de un sistema financiero, es decir, derivadas de la manera en la que se conectan los bancos entre sí, y con otro tipo de instituciones financieras.

Como se descubrió durante el desarrollo de la crisis, fueron las exposiciones provocadas por instrumentos conocidos como *Collateralized Debt Obligation* (CDOs) y *Mortgage-Backed Securities* (MBS) las que desencadenaron los problemas que resultaron en la crisis. Aquí, entenderemos por *exposición* la pérdida potencial que enfrenta una institución bancaria dada su actividad de préstamos, operaciones con divisas, operaciones con títulos de deuda y operaciones de derivados, con cualquier otra institución financiera.

Gracias a la crisis, se volvió evidente la necesidad de desarrollar nuevos métodos capaces de identificar las situaciones de riesgo que no estaban siendo capturadas por el enfoque microprudencial tradicional, pues incluso en un sistema financiero fuerte, el estrés del participante más débil puede inducir un riesgo no despreciable en el sistema, y con la creciente complejidad de los sistemas financieros del mundo ha crecido también la dificultad para regularlos y monitorearlos.

Los esfuerzos principales para mejorar la regulación después de la crisis emanaron del Banco de Pagos Internacionales (*BIS*, por sus siglas en inglés), específicamente, del Comité de Basilea para la Supervisión Bancaria (*BCBS*), quienes propusieron a los bancos centrales miembros una serie de reformas a la regulación vigente, con la finalidad de aumentar la resistencia del sistema financiero a choques similares a los que provocaron la crisis del 2008. Dichas reformas estuvieron basadas, principalmente, en el incremento del tamaño y la calidad del capital regulatorio que tienen que conformar las instituciones financieras, así como en mejorar los mecanismos de compensación a los afectados por la quiebra de una institución financiera.

Adicionalmente, se crearon los conceptos de “Banco Global de Importancia Sistémica” (*Global systemically important bank*, o *G-SIB*), denominación que se otorgó a las instituciones bancarias con actividad multinacional, y cuyo estrés tendría repercusiones importantes para instituciones financieras (e incluso no financieras) de todo el mundo.

Análogamente, se creó el concepto de “Banco Doméstico de Importancia Sistémica” (*Domestic systemically important bank*, o *D-SIB*), con un significado similar al de *G-SIB*, pero en una escala nacional, o regional, sin tener alcance al sistema financiero mundial. La metodología para la definición de los *G-SIB*'s y de los *D-SIB*'s se puede

capacidad de ofrecer créditos, pero sin la posibilidad de realizar captación de recursos, como sí puede hacerlo un banco comercial.

consultar en (BIS, 2017); la lista de las instituciones sistémicas globales se encuentra en (FSB-BIS, 2017).

Cabe mencionar que la regulación convencional también sufrió cambios notorios: se endurecieron las medidas seguidas por los bancos para gestionar su riesgo, se impusieron límites más conservadores para la exposición de las instituciones a ciertos tipos de instrumentos, y se incrementaron los montos de capital que los bancos tienen que reservar para hacer frente a situaciones de estrés, siendo los G-SIBs los obligados a constituir las reservas de capital más grandes.

Dado el impacto que la crisis financiera de 2008, iniciada en Estados Unidos, tuvo para la economía mundial, principalmente en Estados Unidos y Europa, la identificación de los *G-SIBs* cobró un papel preponderante para los reguladores de todo el mundo, y fue en la teoría de redes que éstos encontraron una de las mejores herramientas para dicha tarea.

La crisis del 2008 tuvo severas consecuencias en la economía, especialmente de Estados Unidos y de Europa, donde algunos países (por ejemplo Grecia) siguen sufriendo las consecuencias, materializadas en una economía poco dinámica y poco crecimiento, lo que a su vez provoca desempleo y pobreza. Para un análisis de las causas de la crisis se puede consultar Krugman (2012). Una línea del tiempo detallada de los acontecimientos más importantes de la crisis se encuentra en Guillén (2015). Además, algunos datos de las consecuencias en EE.UU., y las medidas tomadas para mitigar el impacto se pueden encontrar en The Department of the Treasury, USA (2012).

Los bancos interactúan entre sí y con otro tipo de instituciones financieras y no financieras de muchas formas: toman préstamos de los que heredan una obligación de pago, emiten títulos de deuda con los que también adquiere obligaciones de pago, hacen operaciones de reportos², préstamos directos, compra-venta de títulos de deuda, etcétera. Es gracias a todas las actividades mencionadas que se forman las redes financieras. Los pasivos de una institución son los activos de otra y viceversa.

La complejidad del sistema que se forma con la dinámica descrita puede resultar abrumadora, pero con conceptos como la centralidad, y el análisis topológico de una red se ha logrado obtener conocimiento valioso de los procesos generadores de dichas redes, además de su organización estructural y el impacto que ésta tiene en los fenómenos que se desarrollan en la red.

Con lo anterior, se hizo evidente la necesidad de un marco regulatorio y de supervisión que no sólo tomara en cuenta la salud financiera de las instituciones de manera individual; tomó gran importancia el estudio de la salud del sistema financiero en su conjunto.

Fue gracias a las nuevas herramientas de redes y de otros ramos de las matemáticas que las causas de la crisis se hicieron evidentes, por lo que varias de éstas se integraron

²Las operaciones de reportos son un tipo de préstamo con garantía utilizado por las instituciones financieras para obtener liquidez, o para aprovechar el exceso de la misma. En un reporto, la institución reportadora (parte activa) compra títulos a cambio de un monto pactado, y se obliga a devolver dichos títulos por el mismo monto al que los compró más una tasa premio pactada entre las dos partes.

1. INTRODUCCIÓN

con gran éxito a las herramientas de trabajo de los reguladores de todo el mundo, pues proveen un marco teórico y práctico altamente interpretable, además de visual, que facilitan el análisis exploratorio del sistema en situaciones de estrés y en situaciones de calma, ayudando a la comparación de ambos escenarios, además de proveer métricas que ayudan a entender la estructura compleja del sistema.

Derivado del trabajo regulatorio internacional mencionado previamente, se desarrollaron diversos modelos de contagio financiero con redes, por ejemplo el trabajo de [Canedo and Martínez-Jaramillo \(2009\)](#), que permitieron medir la exposición total de los sistemas financieros de todo el mundo, o al menos en los países donde los datos permiten dicho estudio, pues son pocos los bancos centrales que recopilan de manera consistente datos desagregados de las operaciones de los bancos en sus respectivas jurisdicciones.

Los modelos de redes son altamente demandantes en términos computacionales y de datos, pues para poder hacer mediciones certeras del riesgo derivado de la actividad de todo un sistema bancario, es necesario contar con datos desagregados, incluso a nivel de transacción, que permitan conocer las posiciones de las instituciones contra cada una de sus contrapartes.

Por suerte, el Banco de México ha hecho un gran trabajo de recolección de datos regulatorios durante las últimas tres décadas (producto de la crisis financiera que golpeó al país en la década de los años 90). Se tiene información a nivel transacción de **todas las operaciones de todos los bancos** en diferentes mercados (derivados, divisas, préstamos directos, reportos, etcétera); con dichos datos es posible realizar un análisis de redes como el que se realizó en el presente trabajo.

Por otro lado, la información de los estados financieros de los bancos es más accesible, pues éstos están obligados a reportar los rubros principales con cierta periodicidad, principalmente si son públicos (es decir, si son bancos cuyos títulos de propiedad operan en una bolsa de valores). Dado que los estados financieros contienen información agregada de las actividades de cada institución, se necesita otro enfoque para extraer información valiosa de ellos.

El área de “Aprendizaje máquina” (*machine learning*) también ha aprovechado y fomentado el avance tecnológico. En la búsqueda de desarrollar sistemas computacionales capaces de “aprender”, se han propuesto técnicas sumamente flexibles que tienen como finalidad, entre otras cosas, detectar patrones, clasificar objetos, aproximar funciones; y es aquí donde encontramos herramientas de gran utilidad para extraer información de los datos de balance.

Muchos esfuerzos se han hecho recientemente para detectar estas nuevas dimensiones de interacción entre los bancos, pero ha sido gracias al desarrollo tecnológico que se han logrado avances significativos, pues ha permitido la aplicación de metodologías no necesariamente nuevas, pero que no eran factibles en el pasado debido a su complejidad computacional.

Utilizando datos regulatorios del Banco de México, fue posible realizar un estudio con la red interbancaria real de exposiciones y con los balances contables de los bancos, con la finalidad de obtener agrupaciones (conglomerados) que se puedan interpretar

como **modelos de negocio**, pues las diferencias en la manera en la que operan los bancos, así como los instrumentos con los que lo hacen, pueden ayudar a monitorear y regular con mayor precisión a las entidades.

1.2. Motivación del trabajo

Es en la teorías de redes y los métodos de clasificación dentro del aprendizaje máquina que encontramos herramientas de gran utilidad para el problema que buscamos resolver en el presente trabajo: encontrar grupos homogéneos de bancos en términos de actividad (con datos de balance) y de conectividad (con la red bancaria). Dentro de la literatura, dichos grupos se han definido como “modelos de negocios” (“*business models*”).

¿Por qué es importante identificar los modelos de negocios de los bancos en México? Como veremos más adelante, una preocupación en la literatura y para los reguladores, es que el exceso de normas que tienen que cumplir las instituciones genere a la larga un proceso de *homogeneización*, que lleve a los bancos a utilizar los mismos productos, en los mismos mercados, con mayor intensidad, creando así situaciones de riesgo derivadas de la alta correlación entre las inversiones de todas las instituciones; adicionalmente, a últimas fechas los directivos de los bancos en México han manifestado que una regulación homogénea para la banca puede propiciar problemas de competencia.³

Más aún, la manera en la que se organizan las instituciones puede tener una implicación directa sobre la influencia de un grupo de instituciones (incluso de una sola institución), que guarda una estrecha relación con los conceptos de “Demasiado Grande para Caer” (*Too-Big-to-Fail*, TBTF) y “Demasiado Interconectado para Caer” (*Too-Interconnected-To-Fail*, TITF), que son de las situaciones que más se ha intentado erradicar a partir de la crisis, pues tienen el potencial de generar incentivos para que las instituciones con dicha denominación tomen niveles de riesgo mayores a los “saludables”, bajo el supuesto de que, dada su importancia, los reguladores tomarían acciones para su rescate en caso de estrés.

Dentro de un sistema financiero, un alto grado de homogeneidad está asociado con una mayor fragilidad del mismo (León, 2017), por lo que, en el contexto del riesgo financiero sistémico, es importante conocer qué tan parecidas son las instituciones bancarias en términos de conectividad y operatividad. Un sistema altamente conectado y sincronizado puede propiciar la rápida propagación de un shock, cuya gravedad se vería potenciada en un sistema que además de estar altamente conectado, es sumamente homogéneo, es decir, en un sistema donde las instituciones bancarias operan en los mismos mercados y con el mismo conjunto de instrumentos.

También en León (2017) hay un muy buen repaso de trabajos donde se estudian las consecuencias de una alta homogeneidad en el sistema bancario, para mencionar sólo algunos:

³El Financiero: Regulación a la banca debe ser diferenciada.

1. INTRODUCCIÓN

- [Haldane and May \(2011\)](#) sostienen que en la búsqueda de diversificación, los balances financieros de los bancos, así como sus prácticas de manejo de riesgos adquieren grados preocupantes de homogeneidad.
- En [Goodhart and Wagner \(2012\)](#) mencionan que al obtener fondeo de las mismas fuentes el sistema financiero se vuelve susceptible a los choques que pueden recibir sus fondeadores.
- [Beale et al. \(2011\)](#) sugiere que la falta de diversidad entre los bancos puede deberse a un proceso de “*diversificación* uniforme”, en el que los bancos diversifican sus portafolios con el mismo conjunto de instrumentos. Lo anterior crea un efecto contradictorio, pues se supone que diversificarse tiene el propósito de disminuir el riesgo derivado de la correlación de los productos que se tienen, y por lo tanto se minimiza la probabilidad individual de quiebra, pero, al poseer portafolios con los mismos activos que el resto de las instituciones, aumenta la fragilidad del sistema en su conjunto.
- Aún más, en [Aymanns and Georg \(2015\)](#) establecen que al tener portafolios muy similares, las instituciones se vuelven susceptibles a sufrir estrés derivado de golpes de mercado a activos comunes para gran parte del sistema.

Son los modelos de negocios y los patrones de conectividad los que nos darán información sobre los instrumentos con los que operan las instituciones, y la intensidad con la que lo hacen, respectivamente. La información de balance (altamente desagregada) contiene, de manera implícita, las decisiones de inversión que toman los bancos: podemos saber el monto que se tiene en prestamos interbancarios sin garantía, con garantía, en títulos gubernamentales, privados, derivados, etcétera; por lo anterior, es que resultará de gran utilidad para identificar bancos con patrones de inversión similares. Si además resulta que tienen patrones de conectividad parecidos, lo que los colocaría en un mismo grupo en términos de conectividad, estaremos encontrando una situación de riesgo para el sistema entero.

Por último, resulta por demás sorprendente que, a pesar de que las técnicas que nos brinda el Aprendizaje Máquina han sido aplicadas con gran éxito en las actividades cotidianas de las instituciones bancarias en todo el mundo, sufran de una evidente ausencia en las actividades de monitoreo de las instituciones reguladoras mundiales. Dada la gran oferta de técnicas, y la flexibilidad que éstas brindan para solucionar una gran cantidad de problemas, es imprescindible que sean incorporadas a los procesos que se llevan a cabo día a día en las actividades de los reguladores financieros. En [Chakraborty and Joseph \(2017\)](#) aplican métodos de Aprendizaje Máquina a dos casos reales con interés regulatorio, sin embargo, la cantidad de técnicas cuya utilidad aún no ha sido explorada es muy grande, y los reguladores de todo el mundo deberían de redoblar esfuerzos en hacerlo.

1.3. Objetivos del trabajo

1.3.1. Objetivos generales

En pos de una metodología mucho más robusta, capaz de identificar y cuantificar el riesgo derivado de interacciones no triviales entre los bancos, aplicaremos métodos de teoría de redes (**Modelos de Bloques Estocásticos**, o SBM) y de *machine learning* (**Conglomerados**, o Clustering) para medir qué tan homogéneos son los bancos en México en términos de conectividad (SBM) y de actividad (Conglomerados).

1.3.2. Objetivos particulares

Cuantificar qué tan similares son las instituciones bancarias en el país, y qué factores favorecen dicha similitud (o disimilitud), pues un alto grado de homogeneidad conlleva una fragilidad peligrosa en un sistema financiero, como se vio en la crisis de 2008.

Adicionalmente, se espera encontrar evidencia de la existencia o ausencia de instituciones “Demasiado Grandes para Caer”, o “Demasiado Interconectadas para caer”, cuya presencia impone un alto riesgo ético sobre un sistema financiero.

1.4. Contribución del trabajo

Enumeramos las principales contribuciones de la presente tesis como sigue:

- Hasta donde conocemos, este es el primer trabajo en el que se analiza la estructura de agrupación de un sistema bancario en el mundo, utilizando dos métodos distintos en dos bases de datos distintas, y por un periodo de tiempo tan largo (enero de 2007 a junio de 2017). Como veremos en el siguiente capítulo, hay amplia literatura sobre modelos de negocios y su modelado, sin embargo, el presente trabajo es el único en el que se explora la agrupación de un sistema bancario en dos dimensiones distintas (actividad y conectividad), además del aspecto dinámico del comportamiento de dichas instituciones.
- Explicar cuales factores de los que intervienen en la actividad y conectividad de las instituciones bancarias en México contribuyen en la formación de grupos diferenciados, utilizando para ello una técnica más del aprendizaje máquina (Bosque Aleatorio).
- Proponer una metodología que pueda servir para hacer aún más robusta la supervisión de las instituciones financieras del país.
- A pesar de que el sistema bancario Mexicano es pequeño (alrededor de 50 bancos en la primera mitad de 2018) si lo comparamos con los sistemas bancarios

1. INTRODUCCIÓN

de otros países, la metodología que se propone en este trabajo es aplicable a un sistema de tamaño y complejidad arbitrario; aún más, los métodos propuestos podrían incluso arrojar resultados más robustos en sistemas bancarios sumamente heterogéneos, y son capaces de manejar sistemas financieros con otro tipo de intermediarios financieros sin necesidad de ajustar drásticamente los procedimientos computacionales requeridos.

- Verificar si el conjunto de datos tiene un papel importante en la determinación de una agrupación.
- Conocer el comportamiento de los bancos como unidad (resultados del método de conglomerados) y su comportamiento al incorporarse a un sistema bancario (resultados del método de Modelo de Bloques Estocásticos).
- Aportar más ejemplos de aplicaciones reales de las técnicas de aprendizaje máquina y de teoría de redes para la solución de problemas de interés regulatorio.

Éste es un primer esfuerzo para conciliar dos metodologías de agrupación distintas, teniendo cuidado de mantener un sentido económico y financiero de los grupos obtenidos, proveyendo así un marco de referencia que sirva de apoyo a las distintas entidades reguladoras del sistema financiero mexicano.

1.5. Datos

Los conjuntos de datos explotados provienen de dos repositorios institucionales del Banco de México. El primero incluye información altamente desagregada de datos de balance general de los bancos (lo que se conoce como el “catálogo mínimo” de la Comisión Nacional Bancaria y de Valores, o CNBV) ⁴, la cual es mucho más amplia que la versión pública que da a conocer la misma CNBV⁵.

En el segundo repositorio de información se incluye información sobre los montos agregados de la actividad de las instituciones bancarias en los mercados financieros nacionales e internacionales. Dichos montos representan las pérdidas potenciales que enfrentarían los bancos dada la quiebra de sus contrapartes, por lo que nos ayuda a medir la resiliencia del sistema, así como la conectividad del mismo, pues la información se encuentra con un alto nivel de detalle, a saber, es información transaccional para la que es posible identificar las partes.

Ambos conjuntos de datos contienen información confidencial de la actividad de los intermediarios bancarios del país, por lo que no es posible hacerla disponible, sin embargo, dada la flexibilidad de los métodos que presentaremos a lo largo del trabajo, es posible aplicar los métodos implementados a datos no necesariamente idénticos, siempre y cuando conserven el formato que requieren las rutinas escritas. En los capítulos siguientes se da mayor detalle sobre la composición de los conjuntos de datos.

⁴Documentación del Catálogo Mínimo

⁵Base Pública del Catálogo Mínimo

1.6. Infraestructura de cómputo y software utilizado

La parte computacional de la tesis fue realizada en distintas máquinas con dos distintos sistemas operativos: Windows 8 y Ubuntu Linux. Todas las máquinas utilizadas presentan procesadores multi núcleos que varían desde los 32 procesadores lógicos y 16 físicos, hasta los cuatro procesadores lógicos y dos físicos, lo que permitió la paralelización de las rutinas implementadas.

Por su parte, el software utilizado incluye Microsoft Excel 2013 para el pre procesamiento de los datos de balance general; R project ([R Core Team, 2018](#)), para el ajuste de los conglomerados y la obtención de las visualizaciones, y Python ([Rossum, 1995](#)) para la obtención de las comunidades usando los modelos de bloques estocásticos. De nuevo, en los subsecuentes capítulos se describe con detalle las bibliotecas específicas utilizadas para cada tarea desarrollada a lo largo de la tesis.

Las funciones y los códigos desarrollados para el presente trabajo se encuentran disponibles de manera pública en un repositorio en *GitLab*, mismo que se puede clonar localmente con la siguiente línea de comando:

```
git clone https://gitlab.com/lescobarfarfan/codigos_replicacion_tesis.git ,  
o al que se puede acceder a través de la siguiente página web: https://gitlab.com/lescobarfarfan/codigos\_replicacion\_tesis.
```

Los datos para los métodos de conglomerados que se encuentran en el repositorio son una versión pública de la Comisión Nacional Bancaria y de Valores. Finalmente, las redes en el repositorio son completamente simuladas y su estructura puede diferir significativamente de las redes originales utilizadas para esta tesis.

1.7. Estructura de la tesis

Para el resto de la tesis, en cada capítulo introduciremos el marco teórico de cada una de las herramientas que se utilizaron; también, incluiremos los resultados relevantes de la aplicación de los métodos descritos en el marco teórico de cada capítulo con la finalidad de hacer más fluido el contenido del trabajo, así como la discusión de los resultados, para tener una idea general más clara y amplia en el capítulo 7, donde concluimos la tesis.

Así, el resto de la presente tesis está estructurada de la siguiente forma: en el capítulo 2 introducimos brevemente el marco conceptual de los modelos de negocios, y los distintos tratamientos que se les han dado en términos matemáticos.

En el capítulo 3 describimos el análisis de componentes principales, los métodos de conglomerados, así como los métodos específicos de conglomerados que utilizamos para la identificación de los modelos de negocios.

El capítulo 4 contiene las definiciones de los árboles de decisión y los bosques aleatorios, herramientas que utilizamos para la medición de la importancia de las cuentas de balance y las medidas de red para la agrupación de instituciones bancarias.

1. INTRODUCCIÓN

Las principales herramientas utilizadas de teoría de redes en el análisis de redes financieras se presentarán en el capítulo 5.

En el capítulo 6, escribimos los conceptos básicos necesarios de la estadística bayesiana para entender los modelos de bloques estocásticos, así como las generalidades de los mismos.

Finalmente, en el capítulo 7 se incluye la presentación de la agrupación seleccionada, la discusión de los resultados y las conclusiones del trabajo.

Modelos de negocios

En este breve capítulo, argumentamos por qué es importante refinar las medidas regulatorias, complementando el enfoque puramente microprudencial con métodos que tomen en cuenta las diferencias significativas entre las instituciones, y que capturen de la mejor manera posible cómo las instituciones interactúan con los bancos similares, y con los bancos fundamentalmente diferentes.

2.1. Antecedentes regulatorios

En el análisis tradicional que los reguladores hacen hasta la fecha, el enfoque microprudencial ¹ sigue siendo preponderante, las pruebas de estrés que se realizan se enfocan en el efecto que un choque financiero en un mercado determinado tendría en la salud de una sola institución, medido a través del impacto de dicho choque en los principales indicadores financieros de una institución (RoA, RoE, Índice de Capitalización, Índice de morosidad, etc).

Sin embargo, dichos análisis pierden de vista que las actividades realizadas por una institución bancaria van más allá de convertir depósitos en préstamos, en la realidad, las operaciones de un banco son mucho más complejas. Dicha complejidad las vuelve susceptibles a recibir impactos negativos desde diversos mercados, nacional o internacionalmente. Esa susceptibilidad depende de las actividades realizadas y del grado en el que las instituciones están involucradas en dichas actividades/mercados.

Lo anterior no quiere decir que los análisis microprudenciales no tengan un papel importante en la regulación financiera, por el contrario, han sido utilizados (y refinados a lo largo del tiempo) ampliamente por los reguladores para realizar pruebas de estrés, que ponen a prueba, y finalmente demuestra la resistencia de un sistema bancario a choques económicos y financieros adversos. Más aún, uno de los esfuerzos que se

¹Para una discusión sobre las bondades y las debilidades de los enfoques microprudenciales y macroprudenciales, véase *Micro- versus Macro-prudential supervision* y *Macroprudential and Microprudential Policies: Toward Cohabitation*

han hecho en el Banco de México para no perder los efectos adversos producto de la conectividad del sistema bancario, es incluir las pérdidas por contagio en los escenarios de estrés que se simulan para los reportes de estabilidad financiero que se realizan en el banco central cada año².

Como se menciona en [Blundell-Wignall et al. \(2014\)](#), después de la crisis financiera mundial que inició en 2008, las autoridades reguladoras han tenido que mejorar las políticas regulatorias “sobre la marcha”, dichos esfuerzos han estado enfocados principalmente a mejorar los requerimientos de capital de las instituciones de acuerdo con los activos ponderados por riesgo (literalmente, imponer un costo a la deuda que se obtiene). Sin embargo, el esfuerzo sigue concentrado en el enfoque microprudencial.

Lo anterior no quiere decir que el monitoreo microprudenciales no es importante, pues ha probado ser bueno para la vigilancia de la salud de las instituciones; la salud individual de éstas es el primer paso para la estabilidad del sistema al que pertenecen, pues la fortaleza del sistema bancario está estrechamente relacionada a la salud individual de cada institución. Aún así, dada la creciente complejidad de los sistemas financieros, derivada de la globalización y la creciente cooperación entre instituciones de diferentes países y jurisdicciones, es necesario extender el entendimiento de los fenómenos complejos que ocurren en el sistema financiero.

En [Blundell-Wignall et al. \(2014\)](#) los autores dan un ejemplo de los primeros esfuerzos regulatorios para controlar los modelos de negocios que adoptan los bancos: la ley Glass-Steagall fue una respuesta de las autoridades reguladoras a la crisis que inició en 1929 en Estados Unidos; dentro de sus puntos principales se establecía la separación total entre la banca de depósito y la banca de inversión, con la finalidad de evitar una situación como la que desencadenó dicha crisis. Además, establecía la creación de un sistema bancario que diferenciara entre bancos nacionales, estatales y locales, así como la prohibición a dichas instituciones de participar en los consejos administrativos de empresas no financieras. Sin embargo, dicha Ley fue derogada a finales del siglo XX.

Actualmente, el principal conjunto de reglas que tienen que seguir las instituciones financieras, particularmente los bancos, lleva por nombre “Basilea”, desarrollado por el Comité de Basilea para la Supervisión Bancaria. Desde un principio, con la implementación de “*Basilea I*” en 1992, las reglas han estado encaminadas a lograr que los bancos mantengan niveles saludables de capital, que, en términos prácticos, es un ahorro que están obligados a mantener para enfrentar problemas financieros derivados de choques a sus actividades.

Posteriormente, en “*Basilea II*” se agrega un gran número de mejoras a la regulación existente, mejorando la calidad del capital obligatorio al obligar su constitución a partir del peso de los activos riesgosos.

Finalmente, las mejoras presentadas en *Basilea III* incluyen mejoras en los procesos internos de gestión de riesgos de las instituciones, aumento en la ponderación para algunos riesgos en la constitución del capital de riesgo, establecimiento de límites en el apalancamiento de las instituciones y un coeficiente para medir la cobertura contra

²Para mayores referencias se pueden consultar los reportes de estabilidad financiera del año 2017 y 2019: [Reporte de Estabilidad Financiera 2017](#), [Reporte de Estabilidad Financiera 2019](#)

riesgo de liquidez. Finalmente, y quizá el aspecto más importante para el presente trabajo, se determinaron montos extra de capital para hacer frente al estrés generado de la conectividad del sistema, en especial el derivado de las instituciones de importancia sistémica.

2.2. ¿Qué entendemos por “Modelos de negocio”?

De manera general, un “*modelo de negocio*” es el conjunto de actividades que realiza una empresa, banco, institución, o cualquier otra organización que opere con el fin de obtener un rendimiento por sus actividades. Dentro de un modelo de negocios se deben definir una serie de acciones y medios para lograr tener un negocio rentable: mercado objetivo (clientes potenciales), metas de crecimiento, planes de operatividad, etcétera. Un buen modelo de negocios es pieza clave para el éxito de un negocio, pues con él se establecen procedimientos y medidas que ayudan a supervisar el seguimiento de los objetivos planteados en un inicio.

En el caso específico de una institución financiera, con “*modelo de negocio*”, entendemos el conjunto de actividades que una institución financiera realiza en sus operaciones diarias, en las que se apoya para obtener fondos, proveer de liquidez al sistema, o maximizar sus ganancias. Además, un modelo de negocio va más allá de considerar únicamente las actividades que se realizan; comprende también la intensidad con la que llevan a cabo una actividad, el alcance geográfico, e incluso el mercado objetivo de sus operaciones.

Dentro de las instituciones bancarias, también hay diferencias entre los modelos de negocio que se siguen: hay bancos enfocados en el crédito al consumo, bancos de inversión, bancos regionales, bancos de nicho, etcétera. En México, tenemos bancos sumamente heterogéneos en cuanto al enfoque de sus operaciones. Por ejemplo, American Express es un banco enfocado en el crédito al consumo, aún más, los créditos que otorga están enfocados a personas con un nivel de solvencia alto. Por otro lado, hay bancos como Finterra (de nicho), enfocado en créditos agropecuarios.

De los ejemplos anteriores, podemos observar que las instituciones son susceptibles a diferentes choques dependiendo de las actividades que realizan en mayor medida, es decir, son susceptibles a diferentes choques dependiendo de su modelo de negocios.

La diferencia existente entre los modelos de negocios de las instituciones bancarias se debe a que éstas necesitan diferenciarse entre sí para mantenerse vigentes y competitivas; las instituciones bancarias están compitiendo constantemente en los mercados en los que operan, por lo que un buen modelo de negocios puede ser determinante para garantizar seguir siendo viables para sus inversores.

2.3. ¿Por qué es importante monitorear los modelos de negocios que siguen los bancos?

En la búsqueda de mantener su competitividad, las instituciones bancarias incurren en prácticas potencialmente peligrosas para su estabilidad, y por consiguiente, para la estabilidad del sistema completo. Como se mencionó anteriormente, la constitución del capital de cada institución se basa en los activos ponderados por riesgo. Sin embargo, dicha ponderación falla al tomar en consideración la estructura de activos, que a su vez depende de las actividades principales de cada institución, es decir, de su tipo de negocio.

Por ejemplo: un descenso fuerte en el precio de activos riesgosos (por ejemplo, acciones o productos financieros derivados) puede ocasionar serios problemas a instituciones que posean una cantidad considerable de los mismos, y cuya provisión de liquidez dependa fuertemente de ellos; los bancos enfocados a la inversión tendrán fuertes problemas, pues generalmente ocuparán gran parte de su capital para solventar las pérdidas derivadas de la devaluación de sus activos, por lo que serán percibidas como instituciones riesgosas y dejarán de percibir préstamos que en otra situación obtendrían sin mayores problemas, ocasionando una espiral de falta de pago a sus deudores con el potencial de provocar contagio.

Lo anterior es un ejemplo de por qué la regulación debe de tomar en consideración los diferentes tipos de bancos. El peso con el que ponderan sobre el capital los activos como los del ejemplo anterior debería de ser considerablemente mayor para las instituciones cuyas actividades las hagan depender fuertemente de ellos. Aún más, los pesos para la construcción de los activos ponderados por riesgo deberían de estar sujetos al modelo de negocio de cada institución.

Con Basilea II, los reguladores permitieron a las instituciones bancarias implementar sus propias metodologías para determinar los ponderadores para sus activos riesgosos, además de no penalizar la alta concentración de dichos activos (es decir, demasiado dinero invertido en unos cuantos activos), lo que llevó a las instituciones a apalancarse³ en exceso, y que derivó de manera inevitable en la crisis de 2008.

En Basilea III se atacan los problemas anteriores obligando a instituciones de importancia sistémica a aumentar el capital obligatorio, y forzando a las instituciones a mantenerse por debajo de un nivel de apalancamiento que las ponga en riesgo en caso de choques en sus operaciones.

Derivado de la creciente complejidad del sistema financiero, se han acuñado dos términos que se han vuelto sumamente populares para los reguladores de todo el mundo: “Demasiado-Grande-Para-Caer” (Too-Big-To-Fail, TBTF), y “Demasiado-Conectado-Para-Caer” (Too-Interconnected-To-Fail, TITF). Dichos conceptos nacen para evidenciar el riesgo ético en el que incurren las instituciones TBTF (o TITF) y sus contrapartes.

³El apalancamiento de un banco se define como la relación entre la deuda que se obtiene para el financiamiento de nuevas inversiones, y el capital propio; a mayor apalancamiento, más riesgo.

Dicho riesgo se desprende de la idea de que esas instituciones sin duda serían rescatadas por los gobiernos/reguladores, lo que a su vez genera una seguridad “artificial” al momento de exponerse a bancos clasificados como demasiado grandes/ conectados para caer, teniendo como consecuencia la subestimación de la magnitud de una exposición a un banco TBTF/TITF.

Entonces, resulta evidente que las reglas aplicables a las instituciones sistémicas deben de ser más amplias, y deberían de buscar evitar la proliferación del riesgo moral ya mencionado. Las reglas deberán tener en cuenta las diferencias entre los modelos de negocio que presenten las instituciones TBTF/TITF.

¿Cuál es el puente entre los modelos de negocios y las instituciones TBTF/TITF? En la literatura [Lucas et al. \(2017\)](#) y [Farnè and Vouldis \(2017\)](#) identifican diferentes tipos de modelos de negocios de instituciones bancarias en Europa utilizando algoritmos de *machine learning* (conglomerados) dentro de los clusters (grupos de bancos con modelos similares). En [Lucas et al. \(2017\)](#) identifican en un solo cluster a las instituciones que se tienen como G-SIB’s, por lo que incluso el conocimiento previo del sistema puede darnos una idea inicial del resultado de dichos algoritmos.

[Farnè and Vouldis \(2017\)](#) no describen similitudes entre los grupos que encuentran con su algoritmo y los grupos de bancos importantes definidos por los reguladores, sin embargo, una comparación entre los bancos catalogados como G-SIB’s por las autoridades y los grupos de bancos identificados en su estudio podría dar luz sobre la eficacia de su método para identificarlos; lamentablemente, debido a la confidencialidad de los datos que utilizan no es posible saber la identidad de los bancos pertenecientes a cada grupo.

Como complemento al análisis de modelos de negocios que se realizó en el presente trabajo, hacemos una breve descripción del sistema bancario mexicano utilizando teoría de redes, que además nos ayuda a obtener medidas del sistema y de las instituciones que lo conforman, mismas que sirven de apoyo para obtener evidencia de la importancia de las características de la red bancaria en la identificación de los modelos de negocios presentes en la misma, así como de la importancia individual de cada banco para el sistema (centralidad).

2.4. Matemáticas para identificar modelos de negocios

Existen otros trabajos en los que se han aplicado diferentes técnicas con el mismo objetivo de obtener grupos de bancos. [León \(2017\)](#) examina la homogeneidad entre las instituciones bancarias en Colombia utilizando un conjunto de datos similares al que se usa en el presente trabajo, sin embargo, aplica un método de *clustering aglomerativo*, mientras que nosotros usamos un método derivado del clásico K-Medias ([MacQueen, 1967](#)). En [León \(2017\)](#) analizan datos del estado financiero agrupado en tres categorías: portafolio de préstamos, inversiones y de fondeo. Encuentran evidencia de que los tres portafolios mencionados son muy similares en los dos bancos colombianos más grandes.

De lo anterior, resulta intuitivo pensar que mientras más parecidas sean las institu-

ciones financieras en términos de portafolios y actividades de inversión y fondeo, más propensas son a sufrir efectos negativos derivados del estrés de otra institución similar; si además se toma en consideración la gestión de riesgos, que tiene un grado significativo de homogeneidad derivado de la regulación que tienen que cumplir, las instituciones tomarán una serie de decisiones alineadas al resto de los bancos con problemas similares, creando así una espiral de problemas que pueden resultar altamente nocivos para todo el sistema financiero. Con lo anterior se hace evidente la importancia de trabajar en un marco regulatorio que tome en consideración el grado de homogeneidad entre instituciones en el sistema completo.

Respecto a la similitud en los portafolios de inversión, en [Poledna et al. \(2018\)](#) muestran que las pérdidas derivadas únicamente de la homogeneidad de los portafolios de los bancos en México no deberían pasarse por alto. Utilizando datos de transacciones entre las instituciones bancarias mexicanas, demuestran que al descartar el riesgo de contagio por “*overlapping portfolios*”⁴ se subestima el riesgo sistémico hasta en un 50 % en periodos de estrés generalizado.

El presente trabajo toma prestado mucho del análisis realizado por [Farnè and Vouldis \(2017\)](#), especialmente diseñado para optimizar el trabajo con “fat data”, es decir, datos donde el número de variables observadas para cada institución es mucho mayor que el número de instituciones. Utilizan el mismo procedimiento de clusterización derivado de K-Medias en datos regulatorios similares; agregan además un criterio para clasificar bancos con modelos de negocio “idiosincráticos”, es decir, modelos que difieren significativamente del resto de los grupos que encuentran en su estudio. Sin embargo, dejaremos de lado el procedimiento de identificación de modelos de negocios idiosincráticos.

Utilizan información regulatoria de 365 bancos que operan en la eurozona. Encuentran cinco grupos distintos de bancos: “de fondeo mayorista”, “comerciales tradicionales”, “tradicionales complejos” y “enfocados en valores”, así como un conjunto de bancos que siguen modelos de negocio únicos.

Los autores dan evidencia de que los grupos definidos por el algoritmo de clustering presentan distribuciones estadísticas distintas en sus indicadores de riesgo y de rendimiento.

- Los bancos *enfocados en valores* presentan los rendimientos más altos junto con las reservas de capital más grandes, así como un portafolio de activos relativamente riesgoso.
- Los bancos *de fondeo mayorista* también poseen portafolios relativamente riesgosos, con rendimientos altos relativos a capital.
- Los bancos *comerciales tradicionales* tienen en general los portafolios menos riesgosos, presentan además rendimientos superiores comparados con los bancos *comerciales complejos*, aunque por debajo de los otros dos grupos.

⁴La exposición por overlapping portfolios, o portafolios traslapados, se da a partir de la tenencia de títulos similares entre instituciones.

- Finalmente los bancos *comerciales complejos* presentan una combinación aparentemente no óptima de riesgo-rendimiento.

Sostienen que un proceso de selección de variables previo a la aplicación de un método de agrupamiento puede sesgar los resultados con preconcepciones de qué tienen que hacer y cómo se tienen que comportar los diferentes tipos de instituciones bancarias, por lo que el modelo que aplican es lo más agnóstico posible en lo que se refiere a datos y al procedimiento a seguir, en sus palabras, les interesa “*dejar que los datos hablen*”.

En el presente trabajo se comparte dicho punto de vista, pues como ya se mencionó, la interacción entre los bancos es mucho más compleja de lo que tradicionalmente se piensa, y se puede perder una gran cantidad de información en el proceso de selección de variables, información que puede ser de gran utilidad para desenmascarar la verdadera estructura de agrupación de las instituciones. Es por lo anterior que los métodos de aprendizaje no supervisados resultan sumamente útiles para nuestro fin.

En [Lucas et al. \(2017\)](#) aplican un método altamente complejo de agrupamiento basado en mezclas de distribuciones para identificar grupos similares de bancos en Europa. El desarrollo incorpora una dimensión temporal a la identificación de los modelos. La agrupación se hace con un conjunto de variables definidas por los autores, sobre las que hacen suposiciones de distribuciones estadísticas. Sin embargo, como ya se mencionó, en el presente trabajo preferimos mantener un enfoque agnóstico al respecto. El componente de tiempo es incorporado para describir las medidas tomadas por los bancos europeos después de la crisis, donde los reguladores se movieron a un entorno de tasas de interés sumamente bajas que hasta la fecha sigue vigente, aunque a últimas fechas se han dado pasos hacia la “normalización” de las tasas de interés en las economías más avanzadas.

Utilizando datos de 208 bancos operantes en Europa, identifican y caracterizan seis grupos distintos de bancos: “*grandes bancos internacionales*” (que incluyen varios bancos sistémicos), “*bancos mayoristas/corporativos*”, “*bancos comisionistas/de inversión*”, “*pequeños prestamistas diversificados*”, “*prestamistas minoristas locales*” y “*bancos cooperativos/mutualistas*”.

- Los *grandes bancos internacionales* son los más grandes en tamaño de activos; 60% de sus rendimientos provienen de activos con pago de interés; son los más apalancados; poseen grandes portafolios de derivados y de valores, en términos de tamaño y relativos a activos; finalmente, tienen importante actividad más allá de las fronteras de su país de residencia.
- Los *bancos mayoristas/corporativos* son los segundos en tamaño; prestan significativamente a corporaciones; comercian derivados y valores a nombre de sus clientes; obtienen poco o nulo fondeo de depósitos.
- Por su parte, los *bancos comisionistas* siguen en tamaño de activos; gran parte de sus ingresos provienen de comisiones y cuotas por actividad financiera; sólo la mitad de sus activos son préstamos, de los cuales todos son para clientes corporativos en lugar de clientes minoristas.

2. MODELOS DE NEGOCIOS

- Respecto a los *prestamistas minoristas locales*, ofrecen préstamos en igual proporción a clientes corporativos y minoristas; están bien capitalizados (apalancamiento pequeño); tienen depósitos sustanciales de corporaciones y clientes minoristas, así como una cartera de crédito diversificada entre clientes domésticos y no domésticos.
- Finalmente, los *prestamistas minoristas locales* y los *bancos cooperativos/mutualistas* son los más pequeños; son los más numerosos en la muestra utilizada, por lo que resultan importantes para la estabilidad financiera. La gran mayoría de sus activos son préstamos, los cuales ofrecen casi exclusivamente a prestatarios minoristas locales.

En comparación con el sistema bancario europeo, el mexicano es mucho más pequeño en términos de activos, y en número de instituciones, por lo que una clasificación tan amplia como la anterior parece improbable; más aún, resulta llamativo que, a pesar de tener una muestra de instituciones más grande, los resultados en [Farnè and Vouldis \(2017\)](#) no estén tan desagregados, ni son tan específicos en su caracterización como los obtenidos en [Lucas et al. \(2017\)](#), situación que podríamos interpretar como el efecto del sesgo en la selección de variables para la clasificación.

Modelos de Conglomerados y de Reducción de Dimensión

Un vez que hemos establecido la importancia de identificar los modelos de negocios de los bancos, resta la identificación de los mismos. Sacando provecho de las mejoras tecnológicas que se han dado recientemente, se han desarrollado diversas técnicas computacionales altamente eficientes que ayudan a encontrar estructura en grandes cantidades de datos.

En el presente trabajo, utilizaremos dos técnicas de “aprendizaje máquina” (“*machine learning*”) para intentar identificar grupos de bancos que podamos considerar modelos de negocios diferentes, además de un algoritmo que nos ayude a identificar las variables que están interviniendo significativamente en la identificación de dichos grupos. Para lograr lo anterior, usaremos dos algoritmos de análisis de conglomerados “mejorados” con un procedimiento de reducción de dimensión de datos (De Soete and Carroll (1994) y Vichi and Kiers (2001), para un repaso general de los métodos de conglomerados, se puede consultar Anderberg (1973)), y un procedimiento llamado “bosque aleatorio” (véase Breiman (2001)) (“*random forest*”) respectivamente.

Los algoritmos de aprendizaje máquina se dividen en dos subconjuntos principales: de aprendizaje supervisado y de aprendizaje no supervisado. El análisis de conglomerados es un algoritmo de aprendizaje no supervisado, mientras que los bosques aleatorios pertenecen al aprendizaje supervisado.

La diferencia entre el aprendizaje supervisado y el no supervisado radica en los datos que sirven como entrada para el algoritmo, mientras que en el aprendizaje supervisado es necesario proporcionar una *variable respuesta*; no es así en el aprendizaje no supervisado.

Por *variable respuesta* entendemos una variable cuyo comportamiento se pretende modelar y predecir a partir del resto de las variables que se tengan, por ejemplo, en una regresión probit (véase Finney (1952) para una presentación general), se busca la relevancia de ciertas variables en la clasificación de un objeto, por lo que es necesario proporcionar la clasificación previamente conocida. En cambio, en el aprendizaje no

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

supervisado no es necesario proveer al algoritmo de dicha clasificación, pues justo la finalidad de dichos procedimientos es clasificar objetos semejantes en grupos donde el resto de los objetos tengan características similares.

La identificación de los modelos de negocios a partir de métodos computacionales recientes es relativamente nueva, por ejemplos [Lucas et al. \(2017\)](#), [Farnè and Vouldis \(2017\)](#), ambos trabajos de investigación del Banco Central Europeo (ECB, por sus siglas en inglés), además de León [León \(2017\)](#), trabajo del Banco Central de Colombia.

El modelo que utilizamos para agrupar las instituciones que forman parte del sistema bancario mexicano está basado en el algoritmo desarrollado propuesto en [De Soete and Carroll \(1994\)](#) (Reduced K-Means), posteriormente, tomando en consideración las fallas del método anterior en [Vichi and Kiers \(2001\)](#) proponen el método de “Factorial K-Means” aplicado para encontrar modelos de negocios en el sistema bancario europeo en [Farnè and Vouldis \(2017\)](#). Más adelante discutiremos las ventajas y desventajas de cada modelo.

Después de que encontremos una clasificación de los bancos con el análisis de conglomerados, podemos utilizar los bosques aleatorios para conocer la importancia de cada variable en la clasificación de las instituciones, para finalmente hacer un análisis cualitativo de dichas variables, y así poder describir las características inherentes a cada grupo.

3.1. Datos

Para la identificación de los modelos de negocios en el sistema bancario mexicano, siguiendo a [Farnè and Vouldis \(2017\)](#), utilizamos datos desagregados del balance de cada institución, obtenidos del *Catálogo Mínimo*¹ de un repositorio institucional del Banco de México. Dicha información contiene el monto de las cuentas de Balance General de 56 bancos mexicanos abarcando desde enero de 2007 hasta junio de 2017.

El nivel de detalle es elevado, pues de los tres grandes rubros que forman un balance general (activos, pasivos y capital), tenemos más de 1000 distintas cuentas. Como se mencionó, se tienen datos desde enero de 2007 hasta junio de 2017, sin embargo, el modelo que utilizamos es “estático”, es decir, no toma en cuenta la dinámica que pueda haber en la agrupación de las instituciones de acuerdo con su modelo de negocio. Para sobreponerse a dicha limitante, se obtienen los grupos de bancos para diversos periodos, a saber, todos los trimestres de los años ya mencionados, y se obtendrá una gráfica aluvial² para verificar la estabilidad (o inestabilidad) de los grupos encontrados.

¹Los datos usados son confidenciales, más adelante proveemos la liga para consultar versiones públicas de dichos datos almacenadas por la Comisión Nacional Bancaria y de Valores (CNBV).

²Las gráficas aluviales son una herramienta de visualización de gran utilidad para representar los cambios en la estructura de una “red”, en este caso, los nodos de nuestros diagramas son los grupos a los que puede pertenecer cada banco, para consultar una serie de ejemplos, se puede visitar: <https://datavizproject.com/data-type/alluvial-diagram/>

Derivado del aspecto estático de nuestro modelo, el algoritmo que utilizamos (implementado en R (R Core Team, 2018), que se encuentra en la biblioteca “clustrd” (Markos et al., 2018a)) y que recibe como entrada una tabla de dimensión $n \times p$, con n el número de instituciones existentes en una fecha dada y p el número de variables observadas para cada institución. En nuestro caso, la base completa que se obtiene del repositorio de Banco de México contiene 859 variables (muchas de las cuentas existentes en el catálogo mínimo tienen valores en cero para todas las instituciones, por ello es que no se cuentan con las más de mil cuentas reportadas en el párrafo anterior), sin embargo, después de eliminar para cada fecha las cuentas para las que ninguna institución reportó montos distintos de cero, en promedio tenemos 400 variables. Además, también rastreamos el número de instituciones activas para cada fecha, comenzando con 32 bancos en marzo de 2007, para terminar con 48 instituciones activas en junio de 2017.

Dentro del análisis realizado en el presente estudio implementamos una rutina a las variables para reducir aún más el número de columnas presentes en nuestros datos. Dicha rutina tiene como finalidad seleccionar las variables más importantes en términos de su correlación con el resto de las variables, lo que se logra con la eliminación de variables “perfectamente” correlacionadas a pares (correlación > 0.95), seleccionando sólo una de ellas de acuerdo con su “importancia” con el resto de las variables. Dicho procedimiento puede consultarse en Farnè and Vouldis (2017). Para el presente trabajo, el nivel de correlación elegido para eliminar variables es de 0.75.

El número de cuentas que se eliminan de acuerdo con su nivel de correlación con el resto es mucho más variable que el número de variables que se eliminan al ser cero. Como veremos en la sección de resultados, lo anterior tiene efectos claros en los agrupamientos encontrados por los métodos que usamos. La aparición esporádica de montos distintos de cero en cuentas que normalmente son cero para todos los bancos provoca la creación de grupos, por lo general con una o dos instituciones, cuya única diferencia con los grupos encontrados anteriormente (o posteriormente) es esa única cuenta que es distinta de cero para una fecha específica.

El catálogo mínimo del que obtuvimos los datos está homologado con el catálogo de cuentas de la Comisión Nacional Bancaria y de Valores y puede consultarse en línea³. Ahí, podemos corroborar que los datos están formados de “grandes cuentas”, y de subcuentas que desagregan los rubros principales del balance.

Con la finalidad de obtener información sobre el nivel óptimo de desagregación de los datos necesario para nuestro algoritmo, ejecutamos los algoritmos de K-Medias Factorial y K-Medias Reducido (que detallamos adelante) por separado en tres conjuntos de datos. La forma de identificar los niveles de desagregación en el catálogo se desprende de las claves numéricas asignadas a cada rubro. Cada cuenta del balance general tiene asignada una clave única de 12 dígitos. El primer dígito corresponde a la identificación de las “grandes cuentas” (Activo = 1, Pasivo = 2, Capital = 4); en adelante, cada dos dígitos es un nuevo nivel de desagregación.

El primer conjunto contiene las variables en el primer nivel de desagregación, el

³En la página <https://datos.gob.mx/busca/dataset/serie-r01-catalogo-minimo>

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

segundo conjunto es el segundo nivel de desagregación, y análogamente el tercer nivel. Cabe destacar que el número de cuentas en cada conjunto, al igual que el número de instituciones, no es el mismo para todas las fechas.

Una de las discusiones iniciales respecto a las variables a incluir en el estudio fue relativa al uso de indicadores de rendimiento financiero y de capitalización, por ejemplo, el índice de capitalización (ICAP), el índice de morosidad (IMOR) y el ROE (*Return on equity*)⁴. Sin embargo, se decidió dejar fuera esas variables (y similares), la razón principal está relacionada con el marco conceptual que definimos para los modelos de negocios. Recordemos que definimos un modelo de negocios como el conjunto de actividades que realiza una entidad. Por otro lado, las variables de desempeño que se pensaban incluir no cargan información respecto a las actividades realizadas por una institución bancaria, más bien, contienen información de la efectividad con las que dichas actividades son realizadas. Lo anterior sirve para argumentar la exclusión de cualquier variable *calculada*, pues difieren fuertemente de las cuentas de balance que estamos utilizando tanto en interpretación como en su misma naturaleza.

Más aún, la incorporación de indicadores de desempeño al análisis presenta múltiples problemas en lo que a la manipulación de los datos se refiere. Dado el escalado que se le da a los datos, todas las cuentas terminan siendo un porcentaje, hasta aquí no hay mayor problema, pues los indicadores en su mayoría también se presentan en porcentaje, sin embargo, derivado de la evolución natural del sistema bancario hay instituciones que presentan indicadores sumamente alejados de los valores promedio del sistema; por ejemplo, las instituciones de reciente creación suelen tener índices de capitalización muy altos, pues dado que están comenzando a operar no tienen los activos ponderados por riesgo de un banco que tiene más tiempo operando.⁵ Dichas desviaciones pueden sesgar fuertemente los resultados del análisis de conglomerados.

Una de las posibilidades que se presentaron como solución al problema anterior fue la discretización de los indicadores de acuerdo a su nivel, sin embargo, al hacerlo surge de nuevo el problema de mezclar información fundamentalmente distinta, pues aún dejando que las categorías tengan un identificador numérico, éste volvería a estar fuera de escala.

Por último, hemos mencionado que el enfoque que se le quiere dar al análisis en el presente trabajo sigue la idea de “dejar que los datos hablen”, lográndolo al evitar la incorporación de información que pueda incluir un sesgo de parte del investigador, y se considera que asumir que la información de desempeño pueda influir en la estructura

4

- $ICAP = \text{Capital neto} / \text{Activos ponderados por riesgo}$
- $IMOR = \text{Cartera de crédito vencida} / \text{Cartera de crédito total}$
- $ROE = \text{Beneficio neto} / \text{Activos totales (incluyendo deuda)}$

⁵Para tener un ejemplo con nuestros datos, para cierta fecha hay un banco con un ICAP de 3400 %, mientras que el valor regulatorio actual está fijo en 15 %.

de agrupación de los bancos por modelos de negocios inyecta un sesgo indeseado en el análisis.

3.2. Reducción de Dimensionalidad

El algoritmo de conglomerados que utilizamos está pensado para trabajar con “*fat data*”, esto es, datos en los que el número de variables supera por mucho al número de objetos para los que se miden dichas variables.

Los datos de entrada de nuestro método de conglomerados es una matriz $\mathbf{A} \in \mathcal{M}_{n \times p}$, donde $\mathcal{M}_{n \times p}$ es el subconjunto de matrices de tamaño $n \times p$, donde n representa el número de observaciones (en nuestro caso, el número de bancos) y p representa el número de variables (cuentas de balance) para cada banco. Nuestros datos cumplen la condición para ser “*fat data*”, pues como ya vimos, el valor de p es significativamente mayor que el de n .

Al utilizar un algoritmo de conglomerados surge un problema cuando se aplica a datos de altas dimensiones: la visualización de los resultados. Es claro que cualquier espacio con más de tres dimensiones está por encima de nuestra capacidad de visualización, y es justo la visualización de las estructuras en el espacio lo que nos ayuda a discernir entre diferentes ajustes (sin ser un criterio absoluto para seleccionar una partición).

Dada la naturaleza de nuestros datos, es necesario darles un tratamiento previo para poder observar las estructuras resultantes de nuestro análisis, pues si bien un espacio de dimensión mayor a tres ya es imposible de observar, un espacio de dimensión mayor a 400 está completamente fuera de nuestro entendimiento.

Existen diferentes métodos que nos ayudan a reducir la complejidad de un conjunto de datos en dimensiones superiores a las que el ser humano puede visualizar. Dentro de las técnicas de reducción de dimensionalidad, a continuación definimos una de ellas, pues es el método que de manera indirecta utilizan los dos algoritmos que utilizamos para obtener una clusterización. Obtenemos las siguientes definiciones de [Johnson and Wichern \(2007\)](#).

3.2.1. Análisis de Componentes Principales

Al momento de realizar un análisis de componentes principales se tiene la intención de explicar la estructura de varianza-covarianza entre las p variables que componen nuestros datos a partir de un número (idealmente menor a p) de combinaciones lineales de dichas variables. El análisis de componentes principales tiene dos objetivos: reducción de dimensión de nuestros datos y brindar interpretabilidad. Además, como mencionan [Johnson and Wichern \(2007\)](#), el análisis de componentes principales es más una herramienta intermedia en un análisis más complejo que un resultado en sí; por ejemplo, en el “análisis tándem”, se aplican componentes principales como paso previo a un análisis de conglomerados.

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

Por otro lado, el método de conglomerados que aplicamos involucra reducción de dimensión, pero ésta se realiza de manera simultánea a la agrupación, profundizamos más en dicho método más adelante.

La interpretación geométrica del análisis de componentes principales es que dichas combinaciones lineales dan como resultado un nuevo conjunto de ejes en un espacio rotado y, dependiendo del número de componentes principales que se seleccionen, en un espacio de menor dimensión.

Los componentes principales se construyen de la siguiente forma:

Sea $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ un vector aleatorio con matriz de covarianzas Σ , con valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ (es decir, la matriz de covarianzas tiene que ser semi-definida positiva).

Los componentes principales son las combinaciones lineales no correlacionadas cuya varianza es lo más grande posible. Se construyen de la siguiente manera:

$$Y_1 = \mathbf{a}'_1 \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (3.1)$$

$$Y_2 = \mathbf{a}'_2 \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \quad (3.2)$$

\vdots

$$Y_p = \mathbf{a}'_p \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \quad (3.3)$$

$$Var(Y_i) = \mathbf{a}'_i \sum \mathbf{a}_i, i \in \{1, 2, \dots, p\} \quad (3.4)$$

$$Cov(Y_i, Y_k) = \mathbf{a}'_i \sum \mathbf{a}_k, i, k \in \{1, 2, \dots, p\}. \quad (3.5)$$

Escogemos los componentes principales de la siguiente forma:

1er componente = combinación lineal $\mathbf{a}'_1 \mathbf{X}$ que maximiza $Var(\mathbf{a}'_1 \mathbf{X})$, sujeto a $\mathbf{a}'_1 \mathbf{a}_1 = 1$.

2do componente = combinación lineal $\mathbf{a}'_2 \mathbf{X}$ que maximiza $Var(\mathbf{a}'_2 \mathbf{X})$, sujeto a $\mathbf{a}'_2 \mathbf{a}_2 = 1$ y así sucesivamente.

Dado que nuestra matriz es definida positiva, tenemos el siguiente resultado.

Resultado 1 Sea Σ la matriz de varianzas-covarianzas asociada al vector aleatorio $\mathbf{X}' = [X_1, X_2, \dots, X_p]$. Sean $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ los pares de valores y vectores propios asociados a Σ con $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Entonces, la i -ésima componente principal está dada por:

$$Y_i = \mathbf{e}'_i \mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p. \quad (3.6)$$

Así, se tiene que:

$$Var(Y_i) = \mathbf{e}'_i \sum \mathbf{e}_i, i \in \{1, 2, \dots, p\}. \quad (3.7)$$

$$Cov(Y_i, Y_k) = \mathbf{e}'_i \sum \mathbf{e}_k, i, k \in \{1, 2, \dots, p\}. \quad (3.8)$$

Lo anterior nos dice que es posible calcular las componentes principales de nuestra matriz a partir de los valores y los vectores propios de nuestros datos, siempre que la matriz derivada de los datos sea semidefinida positiva. Esto resulta muy conveniente, pues dado que el objetivo del análisis de componentes principales es explicar la mayor variabilidad posible de los datos con un número menor de variables, es importante saber cuánta variabilidad estamos capturando con la inclusión de cada componente principal. Tenemos el siguiente resultado.

Resultado 2 Sea $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ con matriz de varianzas-covarianzas Σ con pares de valores-vectores propios $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Sean $Y_1 = \mathbf{e}'_1 \mathbf{X}, Y_2 = \mathbf{e}'_2 \mathbf{X}, \dots, Y_p = \mathbf{e}'_p \mathbf{X}$ las componentes principales. Entonces:

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p Var(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p Var(Y_i). \quad (3.9)$$

Del resultado anterior obtenemos que la varianza total de nuestros datos se puede calcular como la suma de los valores propios de la matriz de varianzas-covarianzas. Esto nos ayuda a encontrar la varianza total explicada por cada componente. Sea PVT_k la proporción de la varianza total explicada por la componente k , entonces:

$$PVT_k = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i} \quad (3.10)$$

3.2.1.1. ¿Cómo seleccionar el número de componentes principales?

Hay muchas consideraciones que se deben tomar en cuenta para decidir el número de componentes principales que se conservarán para análisis posteriores; para el análisis hecho en el presente trabajo, consideramos en especial dos: la cantidad de varianza de los datos totales explicada por cada componente y la interpretabilidad de los resultados. La primer razón es importante, pues de poco sirve hacer el análisis de conglomerados con datos que explican poco o casi nada de la estructura real de los datos a partir de su variabilidad. La interpretabilidad de los resultados resulta sumamente importante por algo mencionado previamente: dada la naturaleza de nuestros análisis, obtener resultados que se puedan visualizar resulta de mucha ayuda para poder entender y dar una explicación de los grupos.

Dentro de las diferentes herramientas que se tienen para seleccionar un número óptimo de componentes principales, la “*gráfica de codo*” es la más útil e interpretable; como vimos, en una matriz definida positiva la proporción de la varianza total que explica cada componente está dada por los valores propios ordenados de mayor a menor de dicha matriz (en nuestro caso, la matriz de varianzas-covarianzas). Con los valores propios ordenados de mayor a menor, graficamos la posición de cada uno contra su valor (*ivs* λ_i) y obtenemos una gráfica como la mostrada en la Figura 3.1.

Vemos que se forma una curva con un “*codo*”, dicho codo marca el número de componentes “óptimo” para retener la mayor variabilidad posible de nuestros datos, pues dada la interpretación de la fórmula 3.2.1, la inclusión de las componentes correspondientes a los valores propios después del codo significan una mejora despreciable.

En muchos casos, las primeras componentes principales suelen explicar la mayor parte de la variabilidad de los datos originales.

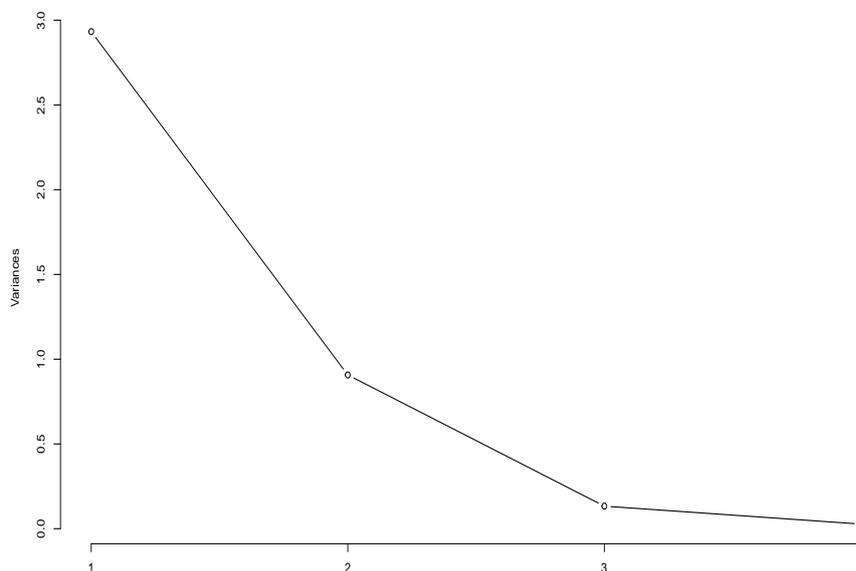


Figura 3.1: Ejemplo de una gráfica de codo.

El análisis de componentes principales está integrado con un método de agrupamiento, a saber, K-Medias (“*K-Means*” en inglés), que forma parte de los métodos de agrupamiento que describiremos a continuación.

3.3. Métodos de agrupamiento

Dentro del análisis exploratorio que se hace sobre un conjunto de datos, el agrupamiento de las observaciones en grupos homogéneos puede ayudarnos a obtener un mejor entendimiento de los mismos. Por ejemplo, en una gráfica de dispersión (cuando es posible obtener una, no en nuestro caso), nos ayuda a observar si hay “agrupación natural” en los datos, o si hay alguna observación que se aleje significativamente del resto. En un problema como el que intentamos resolver, los algoritmos de agrupamiento (“*clustering*”) son de gran utilidad, pues su objetivo es justamente obtener una clasificación de grupos de observaciones con características semejantes.

Los métodos de análisis de conglomerados (*Clustering*) pertenecen al subconjunto de técnicas de aprendizaje máquina (*machine learning*) sin supervisión (*unsupervised*). Lo anterior hace referencia a que previo al análisis no tenemos una variable respuesta cuyo comportamiento se quiere aproximar con el resto de los datos, por el contrario, el objetivo de aplicar dichas técnicas es obtener un valor que se puede entender como la variable respuesta de nuestros datos, o encontrar estructuras significativas en ellos.

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

En las técnicas de aprendizaje supervisado por el contrario, conocemos la clasificación de las observaciones que tenemos; aplicamos dichas técnicas para conocer qué variables tienen mayor importancia en la estructura de agrupación de nuestras observaciones, y, a partir de la medición de dichas variables a un conjunto de observaciones sin clasificar, podemos asignarlas al grupo con las características más cercanas.

Es claro que el significado de los grupos con observaciones “similares” depende de qué entendemos por *similar*. Por ejemplo, si queremos agrupar frutas de acuerdo con sus características observables, es muy probable que variables como el tamaño, la forma y el color resulten útiles para realizar dicha clasificación; sin embargo, si se quiere realizar una clasificación de acuerdo con su sabor, puede resultar desafiante definir u obtener variables relevantes para dicha clasificación.

En el problema que nos ocupa, se utilizan medidas de distancia como en la mayoría de la literatura. Como hemos mencionado, la dimensión de nuestros datos no nos permite obtener visualizaciones directas de los datos, es a través del análisis de componentes principales en combinación con uno de los métodos de agrupación más utilizados que obtenemos una visualización adecuada de nuestros datos.

Las siguientes definiciones y ejemplos fueron obtenidos de [Johnson and Wichern \(2007\)](#).

Existen varias medidas de distancia que se pueden utilizar en los métodos de agrupamiento, a continuación definimos algunas de las más utilizadas en el aprendizaje máquina, así como en otras áreas de las matemáticas.

Definición 1 Sean $\mathbf{x}' = [x_1, x_2, \dots, x_p]$ y $\mathbf{y}' = [y_1, y_2, \dots, y_p]$ dos vectores (observaciones) p -dimensionales. La distancia euclidiana entre ambos se define como:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})}. \quad (3.11)$$

Existen otras medidas de distancia ampliamente usadas, entre las más frecuentes encontramos la distancia estadística.

Definición 2 Sean \mathbf{x} y \mathbf{y} igual que en la definición anterior. Definimos la distancia estadística como:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \mathbf{A} (\mathbf{x} - \mathbf{y})}. \quad (3.12)$$

Donde $\mathbf{A} = \mathbf{S}^{-1}$, con \mathbf{S} la matriz de varianzas y covarianzas muestrales.

Otra medida de distancia es la de Minkowski.

Definición 3 Sean \mathbf{x} y \mathbf{y} igual que en la definición anterior. Definimos la distancia de Minkowski como:

$$d(\mathbf{x}, \mathbf{y}) = \left[\sum_{i=1}^p |x_i - y_i|^m \right]^{\frac{1}{m}}. \quad (3.13)$$

Donde p es el orden de la distancia, donde $p = 2$ es la distancia Euclidea.

Como mencionamos, el sentido que le damos a los grupos homogéneos que pretendemos encontrar con los algoritmos que aplicamos depende de lo que entendamos por “similitud”. En nuestro caso, aplicando la medida de distancia euclidiana ya definida, la similitud entre uno o más objetos está dada por la cercanía que éstos tengan entre sí. Mientras más pequeño sea el valor dado por la distancia euclidiana, más “parecidos” son los bancos en nuestros datos.

En teoría, podríamos obtener la agrupación deseada checando todas las combinaciones posibles del número de grupos y de la pertenencia de cada observación a dichos grupos, sin embargo, computacionalmente nuestro problema se volvería demasiado complejo, específicamente puede volverse un problema *NP-Difícil*⁶, y no podemos perder de vista que los recursos computacionales son siempre limitados.

Diversos algoritmos de agrupamiento se han desarrollado desde hace varias décadas, por ejemplo, el algoritmo K-Medias (que describimos en la sección 3.3.2) fue desarrollado por James McQueen en 1967 (MacQueen, 1967). Queda claro entonces que muchas técnicas exploratorias, como el agrupamiento, no son precisamente modernas, sin embargo, se han visto beneficiadas por los grandes avances que se han dado en los últimos años en el área computacional, lo que ha permitido aplicar dichos métodos de manera más eficiente y en conjuntos de datos mucho más grandes e interesantes.

Las técnicas de agrupamiento se dividen en dos grupos: agrupamiento jerárquico y agrupamiento no jerárquico. Veremos a continuación las diferencias principales entre los dos tipos.

3.3.1. Métodos de agrupamiento jerárquicos

Los métodos jerárquicos se basan en una sucesión finita de separaciones o de aglutinaciones de las observaciones en grupos distintos. Dichas separaciones/ aglutinaciones se dan tomando en cuenta los valores de similitud entre las observaciones. En los últimos, se inicia con N distintos grupos que se unen iterativamente hasta que las observaciones terminen aglomeradas en un solo grupo; por el contrario, en los primeros se inicia con un sólo grupo y se termina el algoritmo hasta que cada observación se encuentre sola en un grupo.

⁶Un problema NP-Difícil es aquel que no puede ser resuelto en un tiempo que crezca de forma polinomial.

De lo anterior se desprenden los llamados *métodos aglomerativos* y los *métodos disociativos*. Describiremos tres de los métodos jerárquicos aglomerativos más usados, llamados “*linkage methods*”, o métodos de vinculación.

3.3.1.1. Métodos de vinculación

En general, el procedimiento que siguen todos los métodos de vinculación es el mismo:

1. Iniciamos con N grupos distintos. De ellos, obtenemos una matriz $\mathbf{D} = \{d_{ik}\}$ de tamaño $N \times N$, donde cada entrada representa la distancia (o similitud) entre la observación i y la observación k
2. Buscamos la distancia (similitud) más pequeña entre cada par de grupos en \mathbf{D} . Y definimos la distancia entre los dos grupos más similares, A y B , como d_{AB}
3. Fusionamos los grupos A y B en un nuevo grupo (AB) y actualizamos la matriz \mathbf{D} en dos pasos: 1) eliminamos las columnas y los renglones pertenecientes a los grupos A y B y 2) agregamos la columna y el renglón con la distancia del cluster (AB) al resto de los grupos.
4. Repetimos los pasos 2 y 3 $N - 1$ veces hasta que todas las observaciones estén en un mismo grupo. Es importante no perder la pista de qué observaciones se unen en un mismo grupo y a qué distancia/similitud.

Tenemos tres métodos aglomerativos de vinculación. La diferencia entre los tres radica en la selección de la distancia entre un grupo recién formado y el resto de los grupos (paso 3).

3.3.1.1.1. Distancia mínima Como el nombre indica, la distancia entre un grupo recién formado, digamos (AB) , y un grupo W está dada por $d_{(AB)W} = \min\{d_{AW}, d_{BW}\}$

3.3.1.1.2. Distancia máxima La distancia entre un grupo recién formado y alguno de los grupos restantes está dada por $d_{(AB)W} = \max\{d_{AW}, d_{BW}\}$

3.3.1.1.3. Distancia promedio La distancia está dada por

$$d_{(AB)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(AB)}N_W}, \quad (3.14)$$

donde d_{ik} es la distancia entre la observación i en el cluster (AB) y la observación k en el cluster W ; $N_{(AB)}$ y N_W son el número de observaciones en el cluster (AB) y W respectivamente.

Los métodos de agrupación jerárquicos se pueden utilizar para obtener grupos no sólo de observaciones (objetos), también se pueden obtener grupos de variables en un conjunto de datos; aquí, la similitud utilizada es la correlación entre dichas variables y para lograrlo, simplemente se aplican los algoritmos descritos sobre la matriz de correlación entre las variables. De lo anterior, y en línea con el significado estadístico de la medida de correlación, las variables con correlación negativa cercana a uno serán consideradas como “lejanas” por el algoritmo, mientras que las variables con una correlación positiva cercana a uno serán consideradas “cercanas” (similares).

3.3.2. Métodos de agrupación no jerárquicos

El método de K-Medias (*K-Means Clustering*) nos ayuda a particionar un conjunto de observaciones en K distintos grupos mutuamente excluyentes. Una desventaja del método de K-Medias respecto a los métodos jerárquicos descritos en la sección anterior es que el usuario tiene que especificar el valor de K a priori, el cual no siempre es conocido.

La agrupación K-Medias está considerada dentro de los métodos no jerárquicos, pues sólo es posible agrupar observaciones, no variables, y como ya se mencionó, es necesario proveer un número de grupos antes de aplicarlo.

Los métodos no jerárquicos se pueden iniciar de dos formas distintas: con una partición dada de las observaciones en grupos o con K puntos iniciales que sirven como centroides de un grupo cada uno. Es importante mencionar que cualquier sesgo respecto a la estructura de agrupación de los datos debe ser evitado, por ello generalmente los K puntos iniciales deben ser elegidos de manera aleatoria, y se comienza a iterar a partir de esa inicialización.

En [MacQueen \(1967\)](#) se presenta por primera vez el algoritmo de K-Medias (*K-Means*). El principio básico de dicho método es que se inicia con un conjunto de K puntos iniciales seleccionados aleatoriamente, y posteriormente se asignan los puntos restantes al conglomerado representado por el punto inicial más cercano. De manera general, la distancia entre los puntos es calculada usando la distancia euclidiana descrita en las secciones anteriores. Complementamos la definición del algoritmo con [James et al. \(2014\)](#).

Sean C_1, \dots, C_K los conjuntos que contienen los índices de las observaciones que pertenecen a cada conglomerado, por ejemplo, si la observación i pertenece al conglomerado k entonces $i \in C_k$; dichos conjuntos deben de cumplir que:

1. $C_1 \cup C_2 \cup \dots \cup C_K = \{1, 2, \dots, n\}$,
2. $C_k \cap C_{k'} = \emptyset \forall k \neq k'$,

donde n es el número de objetos por agrupar y k es el número de grupos en los que se partirán las n observaciones.

El criterio de optimalidad para el algoritmo de K-Medias es la “Variación Intra-Grupos” (*within-cluster variation*), que se define como $W(C_k)$ para un cluster C_k y

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

representa qué tan diferentes son los objetos que están dentro de un mismo cluster entre ellos; recordemos que la medida de “similitud” en nuestro contexto es la distancia euclidiana, por lo que $W(C_k)$ representa qué tan distantes están entre sí los objetos que pertenecen a un mismo grupo, por lo tanto, el algoritmo de K-Medias consiste en resolver:

$$\underset{C_1, \dots, C_K}{\text{minimizar}} \left\{ \sum_{k=1}^K W(C_k) \right\}, \quad (3.15)$$

es decir, encontrar una partición C_1, \dots, C_K con la que se obtenga la variación intra-grupo más pequeña para **todos** los grupos.

Como ya mencioné, la medida de similitud utilizada es generalmente la distancia euclidiana al cuadrado, con lo que podemos definir $W(C_k)$ como:

$$W(C_k) = \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2, \quad (3.16)$$

donde $|C_k|$ es el número de observaciones en el k -ésimo cluster, así, la variación intra-grupo es la suma de la distancia entre cada par de puntos del conglomerado, normalizada con el total de observaciones en ese grupo. Por lo tanto, K-Medias se reduce a resolver:

$$\underset{C_1, \dots, C_K}{\text{minimizar}} \left\{ \sum_{k=1}^K \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right\}. \quad (3.17)$$

A partir de la ecuación anterior es posible definir una función de costo para el algoritmo de K-Medias como:

$$J = \sum_{k=1}^K \sum_{i=1}^n \|x_i^k - c_k\|^2, \quad (3.18)$$

donde x_i^k es un punto i asignado al grupo k y c_k es el centroide del grupo k . La función J nos da las distancias entre los objetos de cada uno de los K grupos.

Una vez más, nos encontramos con el hecho de que resolver la ecuación anterior probando todas las combinaciones posibles de particiones de tamaño K es computacionalmente ineficiente. Por lo anterior, se tiene un algoritmo con el que se logra encontrar mínimos locales:

1. Asignar los n puntos de manera aleatoria a uno de los K grupos.
2. Paso iterativo:
 - a) Para cada conglomerado se calcula el centroide, que es un vector p -dimensional en el que cada entrada es el promedio de los puntos en ese grupo.
 - b) Se asigna cada punto al conglomerado con el centroide más cercano.

El paso iterativo se repite hasta que las asignaciones dejen de cambiar.

Es importante mencionar que dado que el algoritmo sólo garantiza la convergencia a un **mínimo local**, los resultados pueden variar dependiendo de la agrupación inicial, por lo que es recomendado correr el algoritmo descrito varias veces para un mismo conjunto de datos, variando la inicialización de los grupos.

3.4. K-Medias Reducido y K-Medias Factorial

Los dos métodos que utilizamos para obtener conglomerados de instituciones bancarias a partir de información contable de balance fueron diseñados con base en el algoritmo básico de K-Medias, con la diferencia de que están contruidos para lograr una agrupación y **al mismo tiempo** una reducción en la dimensión de los datos, por lo que resultan sumamente convenientes para problemas similares al nuestro, es decir, problemas de detección de conglomerados en los que p es mucho más grande que n . Ambos métodos forman parte del enfoque de análisis de conglomerados en subespacios de las variables, y tienen como supuesto que los *centroides* de los grupos están localizados en un **subespacio** de las p variables. Al respecto, uno de los primeros métodos desarrollados fue el “análisis tándem”, procedimiento que obtiene conglomerados en dos pasos: primero se realiza un análisis de componentes principales y posteriormente se aplica análisis de conglomerados en el espacio generado por las componentes seleccionadas, (Arabic and Hubert, 1994), sin embargo, los dos métodos que describimos a continuación nacieron a partir de la preocupación de que el procedimiento realizado en el análisis tándem pudiera enmascarar la *verdadera* estructura de agrupación de los objetos a través de la construcción de nuevos espacios con el análisis de componentes principales.

En Timmerman et al. (2010) se hace un estudio exhaustivo con datos sintéticos y datos reales para comparar los dos métodos con la finalidad de descubrir en qué situaciones (con qué tipo de datos) es preferible un método sobre el otro. Las conclusiones que obtienen son:

- Demuestran de manera teórica que los dos métodos se ven afectados en diferente medida por la presencia de residuales en el subespacio donde se encuentran los conglomerados, así como en el subespacio complemento ortogonal al mismo.
- Para los dos métodos, la efectividad de la clasificación se ve disminuida con el aumento en el traslape de los conglomerados.
- Cuando la mayoría de las variables contiene información importante sobre la estructura de agrupación de los datos, o cuando se utilizan variables estandarizadas, el método de K-Medias Reducido tiene resultados buenos.
- Dependiendo de la estructura de los datos, los dos métodos son complementarios.

Utilizamos dicho trabajo para describir los dos métodos, así como la principal diferencia entre los dos. Adicionalmente, del mismo trabajo (Timmerman et al. (2010))

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

obtenemos la siguiente lista de definiciones para describir los algoritmos de K-Medias Factorial y K-Medias Reducido (*Factorial K-Means & Reduced K-Means*). Para conservar la notación original del trabajo usamos I en lugar de n y J en lugar de p :

1. I : número de observaciones, $i \in \{1, 2, \dots, I\}$.
2. J : número de variables, $j \in \{1, 2, \dots, J\}$.
3. C : número de conglomerados, $c \in \{1, 2, \dots, C\}$.
4. Q : número de componentes formados con las J variables, $q \in \{1, 2, \dots, Q\}$.
5. X : matriz de $I \times J$ con los datos.
6. U : matriz de pertenencia de tamaño $I \times C$, con $u_{ic} = 1$ si la observación i pertenece al cluster c y cero en caso contrario.
7. F : matriz de tamaño $C \times Q$ con las coordenadas de los centroides en el espacio reducido Q -dimensional.
8. A : matriz ortonormal ⁷ de tamaño $J \times Q$ con los valores de cada variable para formar los q componentes.
9. A^\perp matriz ortonormal de tamaño $J \times (J - Q)$ que cumple $A'A^\perp = 0$.
10. EA' matriz de tamaño $I \times Q$ de los residuales en el subespacio Q -dimensional.
11. $E^\perp A'^\perp$ matriz de $I \times (J - Q)$ complemento de los residuales.
12. T matriz ortonormal de rotación de $Q \times Q$.
13. $0_{I \times J}$ matriz de ceros de tamaño $I \times J$.
14. I_J identidad de tamaño $J \times J$.
15. $diag(c)$ matriz diagonal formada a partir del vector c .

3.4.1. K-Medias Reducido vs K-Medias Factorial

El método de K-Medias Reducido fue introducido por primera vez en [De Soete and Carroll \(1994\)](#), mientras que K-Medias Factorial se propone en [Vichi and Kiers \(2001\)](#). Ambos métodos tienen como objetivo obtener una agrupación de los objetos en la matriz X en grupos homogéneos y al mismo tiempo simplificar la estructura de los datos para facilitar la interpretación, e incluso la visualización de los mismos.

Ambos métodos tienen como primer objetivo descomponer la matriz de datos X en tres submatrices U , A y F . Estas tres matrices tienen un significado muy importante para la interpretación de los resultados que nos arrojen los modelos, pues de acuerdo

⁷Ortonormal implica que $A'A = I_Q$

con su definición dada en la lista anterior, U nos da la pertenencia de cada objeto a un grupo, así como el número de grupos encontrados; la matriz F nos da las coordenadas de los centroides de cada cluster, y la matriz A nos da los coeficientes de cada variable en cada una de las Q dimensiones que resulten de la reducción de la dimensión, por lo que nos dice qué tan relacionada está una variable con una dimensión en particular. Tomando el ejemplo que se da en [Timmerman et al. \(2010\)](#): Sean $I = 4$, $J = 5$, $C = 2$ y $Q = 2$, y las siguientes matrices:

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

$$F = \begin{pmatrix} 1.2 & 1.3 \\ -0.1 & 0.2 \end{pmatrix}$$

$$A = \begin{pmatrix} 0.7 & 0.0 \\ 0.7 & 0.0 \\ 0.0 & 0.7 \\ 0.0 & 0.7 \\ 0.0 & 0.0 \end{pmatrix}$$

La matriz U nos dice que los objetos 1 y 2 pertenecen a un grupo, mientras que los objetos 3 y 4 pertenecen a otro. Luego, las coordenadas del centroide del primer conglomerado son (1.2, 1.3) y las del centroide del segundo son (-0.1, 0.2). Finalmente, vemos que las variables 1 y 2 tiene una relación directa con la primera dimensión del subespacio encontrado por los métodos, mientras que las coordenadas 3 y 4 cuatro tienen relación con la segunda dimensión del subespacio, y por último, la variable 5 no tiene relación con ninguna de las dos dimensiones del subespacio, por lo que se le conoce como una variable de enmascaramiento (*masking variable*). Vemos que con la matriz A podemos incluso ver la relevancia que tienen las variables para describir la estructura de agrupación de los objetos, por lo que estos métodos nos ayudan además a realizar una selección de variables.

De lo anterior, ¿qué es lo que hace diferentes los métodos? Recordando el método de K-Medias clásico, tenemos una función de costo J , para la cual se tiene que encontrar el parámetro K que la minimice. De la misma forma, se definen dos funciones objetivo que se tienen que minimizar. Es importante mencionar que los puntos en estos métodos son cercanos en un sentido Euclidiano, es decir, existen en un espacio Euclidean, sin embargo, la métrica que se utiliza para medir la distancia entre los puntos son distintas, y corresponden a las últimas ecuaciones de las siguientes igualdades.

Para el método de K-Medias Factorial:

$$F_{FKM}(U, F, A) = \|XAA' - UFA'\|^2 = \|XA - UF\|^2 = tr(A'X'XA) - tr(A'X'UF). \quad (3.19)$$

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

Para el método de K-Medias Reducido:

$$F_{RKM}(U, F, A) = \|X - UFA'\|^2 = tr(X - UFA')(X - UFA'), \quad (3.20)$$

donde el operador $tr(M)$ es la traza de una matriz.

Es necesario notar que cuando el subespacio de los centroides tiene dimensión Q , las fórmulas anteriores son iguales a la función de costo del algoritmo de K-Medias clásico. Así, vemos que en K-Medias Factorial se busca minimizar la distancia entre los centroides en el subespacio proyectado por A (UF) y los puntos originales proyectados en el mismo subespacio (XA). Por otro lado, con K-Medias Reducido se busca minimizar las distancias entre los puntos en su espacio original (X) y los centroides en el subespacio definido por A (UFA').

La diferencia más clara entre los dos modelos es fácil de identificar en los residuales de cada uno. Dado que todos los modelos estadísticos están sujetos a tener una desviación de los datos que se quieren modelar, podemos reescribir de acuerdo con las fórmulas descritas arriba los dos modelos.

Para K-Medias Reducido (RKM):

$$X = UFA' + E_R. \quad (3.21)$$

Para K-Medias Factorial (FKM):

$$XAA' = UFA' + E_F. \quad (3.22)$$

Dado que uno de los modelos ofrece una solución al problema de agrupación en una proyección de los datos originales X en un subespacio ortogonal definido por A , es necesario considerar cómo la forma de los residuales del modelo de K-Medias Factorial marca la diferencia con el modelo RKM. De la fórmula 3.22 vemos que

$$E_F = XAA' - UFA', \quad (3.23)$$

es decir, E_F se encuentra en el subespacio definido por las columnas de A' , entonces podemos escribir $E_F = EA'$, donde E es una matriz en el mismo espacio que E_F . Lo anterior quiere decir que E_F contiene los residuales del modelo en el subespacio generado por A' respecto al espacio completo en el que se encuentra X . Entonces, tenemos que

$$XAA' = UFA' + E_F = UFA' + EA' = (UF + E)A', \quad (3.24)$$

además, sabemos que rotar un espacio no afecta su distancia euclidiana, por lo que, multiplicando por A obtenemos que $XA = UF + E$, es decir, lo que mencionamos arriba: *el modelo de K-Medias Factorial ofrece un resultado que vive en el subespacio definido por A* . Por lo que, para hacer comparables los dos modelos es necesario tomar en cuenta los residuales que viven en el subespacio complementario ortogonal al subespacio generado por A . De lo anterior, distinguimos dos tipos de residuales: residuales del

subespacio de A , y residuales del espacio ortogonal complementario a A , los cuales denotaremos como EA' y $E^\perp A^{\perp'}$ respectivamente.

Con los anterior, podemos definir completamente un modelo general para los dos métodos como:

$$X = UFA' + EA' + E^\perp A^{\perp'}. \quad (3.25)$$

El análisis para diferenciar y comparar la efectividad realizado en [Timmerman et al. \(2010\)](#) se desprende de la expresión anterior. Distinguen cuatro tipos de datos que se pueden utilizar en los dos métodos: datos sin residuales en el subespacio de A ni en el subespacio complementario de A , datos sin residuales en el subespacio de A , datos sin residuales en el subespacio complementario de A , y datos sin ninguno de los dos tipos de residuales.

Demuestran que los datos “perfectos” para el modelo de K-Medias Factorial son aquellos sin residuales en el subespacio complementario, y que los datos “perfectos” para el modelo de K-Medias Reducido son aquellos sin residuales de los dos tipos. Se realiza un análisis con datos sintéticos simulados para probar la efectividad de los dos métodos con cada tipo de datos.

Un resultado importante que los autores demuestran en el trabajo, es que el número máximo de dimensiones del subespacio en el que viven los residuales está acotado superiormente por el número de conglomerados, en específico, demuestran que $Q \leq \min(C, J)$. Por lo anterior, y como vemos más adelante, la implementación de los métodos utilizada en el presente trabajo tienen la condición anterior para los dos parámetros (número de dimensiones y de grupos).

3.5. Resultados

Los ajustes fueron hechos con el paquete “clustrd” ([Markos et al., 2018b](#)) del software estadístico [R Core Team \(2018\)](#). Dicho paquete contiene la implementación de los dos métodos de detección de conglomerados para datos de alta dimensionalidad que presentamos en la sección anterior. Se obtuvieron ajustes con los dos métodos (K-Medias Factorial o FKM y K-Medias Reducido o RKM), y se compararon sus métricas de bondad de ajuste además de las agrupaciones resultantes, pues como ya mencionamos, es importante mantener resultados coherentes en términos económicos.

Además, la comparación entre los resultados de ambos métodos puede ayudarnos a identificar el tipo de datos que son las cuentas de balance, diferenciando según los tipos de datos que se definieron como ideales para cada método.

Es común que en la práctica, siempre que se hace un análisis de reducción de dimensionalidad, los datos se normalicen⁸ con el objetivo de hacer **comparables** las cantidades que se utilizan. Por ejemplo, supongamos un conjunto de datos con las siguientes variables para un conjunto de países: PIB, PIB per capita y esperanza de vida. Si se hace cualquier análisis sobre los datos originales, las variables relacionadas

⁸Restar la media y dividir entre la desviación estándar de cada variable.

con el PIB dominarán los resultados, pues su magnitud es mucho más grande que la esperanza de vida, y cualquier método que utilicemos tendrá un sesgo fuerte hacia las cantidades más grandes, por lo anterior, si cada una de las variables anteriores se normaliza, las cantidades utilizadas estarán en un rango de valores similares.

Sin embargo, contrario al escalado clásico, el escalado que proponen los autores en [Farnè and Vouldis \(2017\)](#) consiste en hacer cada cuenta del balance general relativa a su categoría principal, es decir, dividir cada cuenta del activo entre el activo total, del pasivo entre el pasivo total, y del capital entre el capital total. Adicionalmente, los datos pueden ser centrados directamente con la función de ajuste de conglomerados, no obstante, en el presente trabajo omitiremos centrar los datos, pues solamente con el escalado descrito arriba logramos hacer comparables todas las variables utilizadas. En este caso, a pesar de que todas las cantidades tienen como unidad el peso mexicano, las cantidades reportadas por los bancos más grandes son mucho mayores, y el tamaño de las instituciones es algo que puede inducir un sesgo que queremos evitar, pues no refleja objetivamente un ordenamiento por grupos de acuerdo con el conjunto de actividades que se realizan con mayor intensidad.

3.5.1. Calidad del ajuste

Una de las salidas que nos proporciona la función empleada para los ajustes es un criterio de optimalidad de acuerdo con las funciones objetivo de cada método (ecuaciones [3.19](#) y [3.20](#)). Dado que la agrupación óptima es aquella que minimice dichas ecuaciones, para cada ajuste la función devuelve el valor de la respectiva función objetivo evaluada con la partición resultante.

Podemos observar los valores históricos de las funciones objetivo para cada uno de los métodos (FKM y RKM) en cada uno de los niveles de desagregación (véanse las gráficas [3.2](#) y [3.3](#)). Es fácil notar que la escala del criterio para los ajustes del método FKM es significativamente menor que la del criterio de los ajustes con RKM, lo que en principio sugiere que los ajustes obtenidos con el método de K-Medias Factorial es mejor.

No obstante, hay que remarcar el hecho de que los ajustes obtenidos con el método de K-Medias Reducido presentan una tendencia mucho más estable, lo que daría evidencia de que las agrupaciones obtenidas con dicho método son menos variables en el tiempo, lo que es una característica deseable para nuestro análisis.

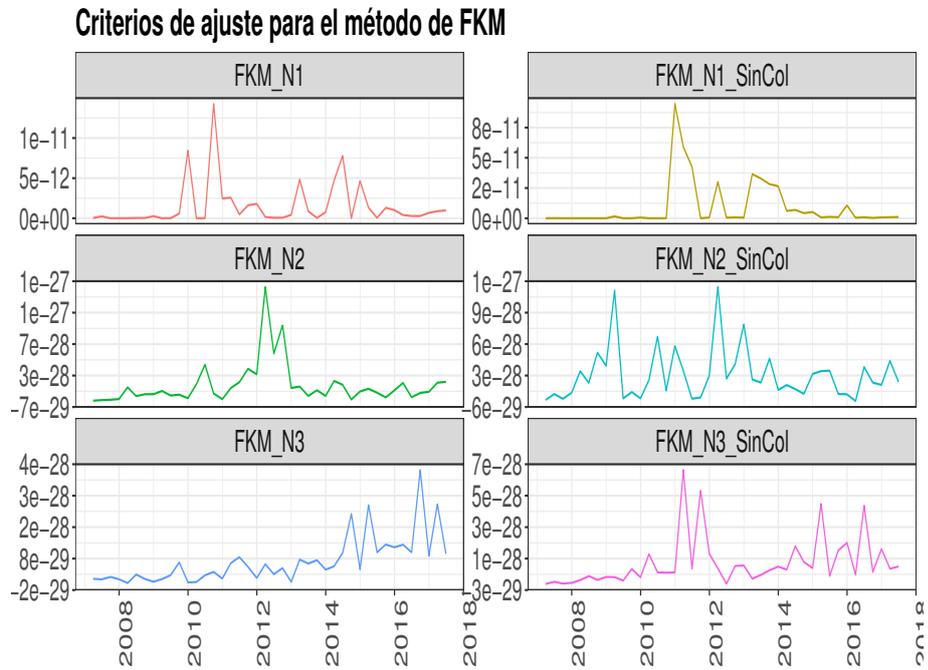


Figura 3.2: Evolución temporal de los criterios de ajuste del método FKM

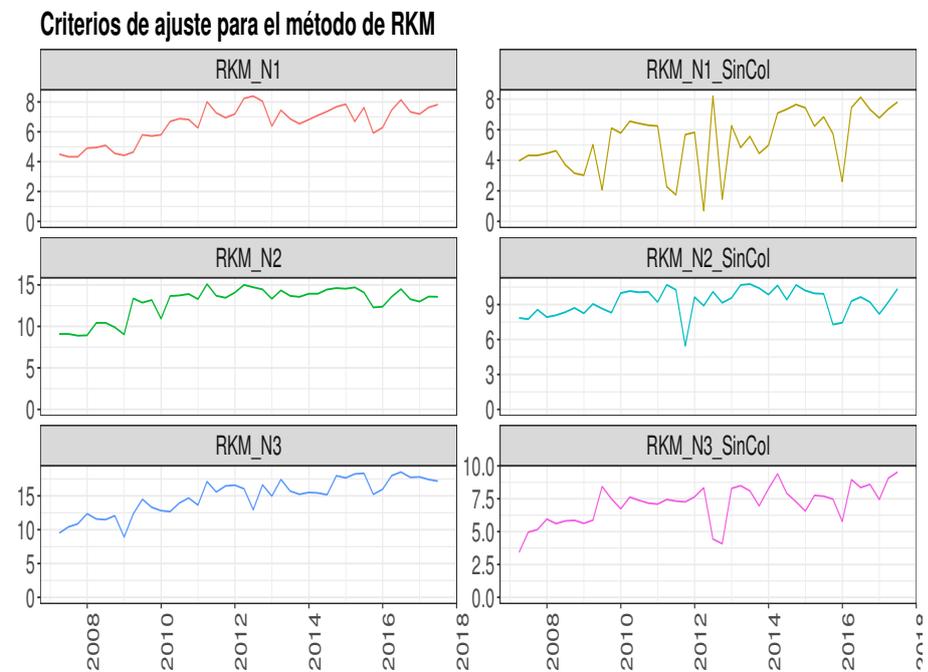


Figura 3.3: Evolución temporal de los criterios de ajuste del método RKM

Para diferenciar aún más la calidad de los ajustes hechos con ambos métodos, obtuvimos visualizaciones de las agrupaciones obtenidas con cada uno para la última fecha que considera nuestro estudio.

3.5.1.1. Ajustes con K-Medias Factorial

Observamos claramente que los grupos no están bien diferenciados en ninguno de los dos casos (véase 3.4⁹), más aún, los ajustes tienen un nivel de complejidad alto, pues el mejor sub espacio de ajuste es de tres dimensiones.

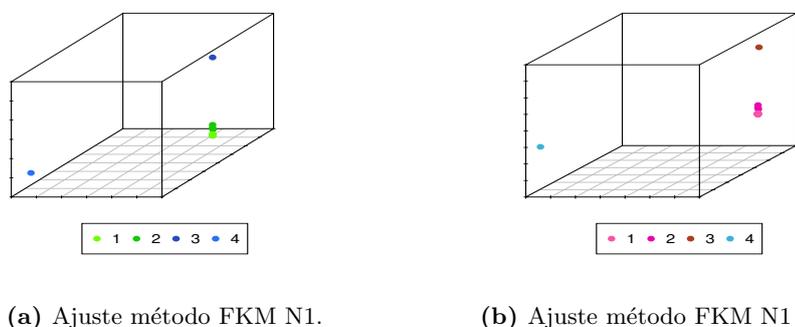


Figura 3.4: Visualizaciones de los ajustes con FKM para el nivel uno de desagregación para la última fecha de estudio.

Para poder verificar si el método FKM se ve beneficiado de la agregación de más información, podemos observar las mismas visualizaciones para el segundo nivel de desagregación (véase 3.5). Tomando en cuenta las variables altamente correlacionadas mantenemos un alto nivel de complejidad, sin embargo, al eliminar las variables redundantes logramos obtener un espacio óptimo de sólo dos dimensiones. Sin embargo, al igual que en las visualizaciones del nivel anterior, todas las instituciones de un mismo grupo están en un mismo punto en el espacio, lo que brinda evidencia de un posible sobreajuste de la agrupación resultante en ambos casos. Además, el número de agrupaciones encontradas (siete) puede complicar fuertemente la interpretabilidad de los grupos resultantes.

⁹SinCol se refiere a los ajustes para los que fueron eliminadas las variables altamente correlacionadas entre si de acuerdo con el procedimiento en [Farnè and Vouldis \(2017\)](#)

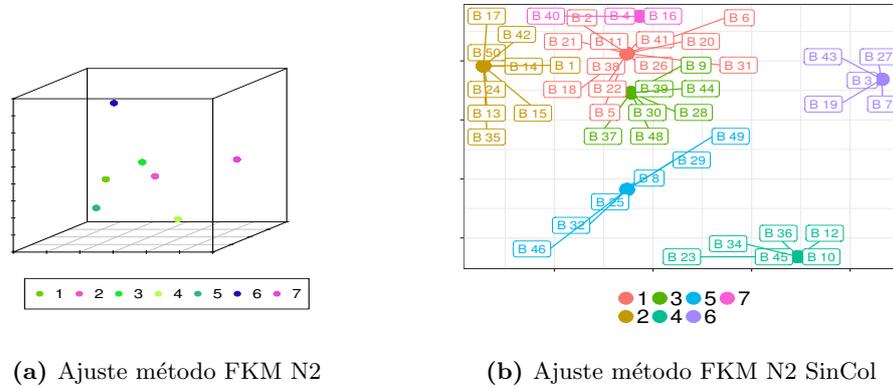


Figura 3.5: Visualizaciones de los ajustes con FKM para el nivel dos de desagregación para la última fecha de estudio.

Finalmente, para el tercer grado de desagregación observamos de nuevo que eliminar las cuentas altamente correlacionadas disminuye la complejidad, pues se obtiene un espacio de dos dimensiones (véase 3.6). Sin embargo, el número de agrupaciones sigue siendo alto, y todas las instituciones de un mismo grupo se encuentran en el mismo punto espacial una vez más.

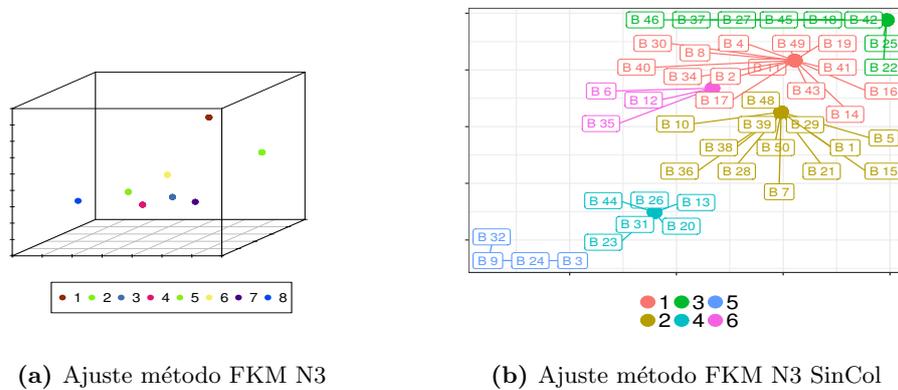


Figura 3.6: Visualizaciones de los ajustes con FKM para el nivel tres de desagregación para la última fecha de estudio.

El hecho de que la incorporación de más información (niveles de desagregación) provoque la formación de más agrupaciones es natural, pues es mucho mayor el número de cuentas de balance que se consideran, y es mucha más la información en la que las instituciones se pueden diferenciar entre sí. No obstante, estas fluctuaciones en el número de grupos pueden ser resultado de la incorporación de cuentas que no figuran normalmente en los balances de los bancos, por lo que la repentina aparición de un

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

monto distinto de cero en una cuenta que aparece con poca frecuencia puede provocar que las instituciones que presentan el monto sean asignadas a su propio grupo. Por lo anterior, es necesario obtener una visualización histórica de los movimientos del número de grupos, y los movimientos de los bancos dentro de dichos grupos. Por lo tanto, como ya mencionamos, utilizamos **gráficas aluviales** (*alluvial diagrams* o *parallel set plots*) (véase Rosvall and Bergstrom (2010)) para ver cómo cambia la agrupación de las instituciones a lo largo del tiempo.

Dado que los ajustes sin variables altamente correlacionadas logran disminuir el número de grupos y las dimensiones utilizadas para el ajuste, sólo agregaremos los diagramas aluviales correspondientes a los ajustes que no las toman en cuenta. En la siguiente gráfica hay resultados interesantes (véase 3.7). A pesar de que las fechas que se consideran como el inicio de la crisis son en el último cuarto de 2008, la incertidumbre mundial era alta desde inicios del año, y los efectos son claros en las agrupaciones encontradas por el algoritmo. Antes de enero de 2008 el número de agrupaciones encontradas era alto (un mínimo de 3 en septiembre 2007, hasta 8 grupos en marzo y junio del mismo año), en cambio, en marzo de 2008 el número baja a tres grupos, y se mantiene relativamente constante hasta 2011 (con excepción de septiembre de 2009). Posteriormente, hay un periodo de mucha variabilidad entre septiembre 2011 y diciembre 2012; después, el número de grupos se mantuvo entre 3 y 4 hasta el fin del periodo.

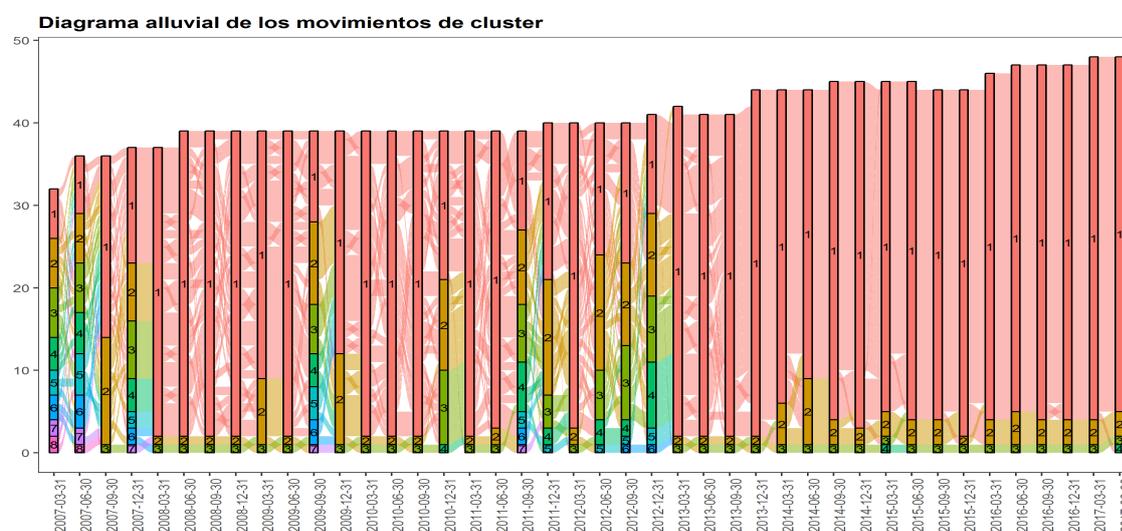


Figura 3.7: Ajuste FKM N1 SinCol.

Como mencionamos al momento de describir los ajustes de los niveles de agregación dos y tres, un gran número de grupos puede deberse a un sistema bancario con un conjunto de actividades altamente diversificado; así, la caída en el número de grupos encontrados es un reflejo más del fuerte cambio que sufrieron los sistemas financieros de todo el mundo: en general, las instituciones bancarias adoptaron una actitud más aversa al riesgo, por lo que resulta creíble pensar que el conjunto de actividades realizadas

disminuyó como consecuencia de ello.

Por otro lado, vemos un patrón similar al de la gráfica anterior, con la diferencia de que el número mínimo de grupos encontrados es mucho mayor (alrededor de 7 y 8 grupos a lo largo del periodo, véase 3.8). Aunque menos evidente, sigue siendo notable la disminución del número de grupos encontrados en los dos últimos trimestres del 2008.

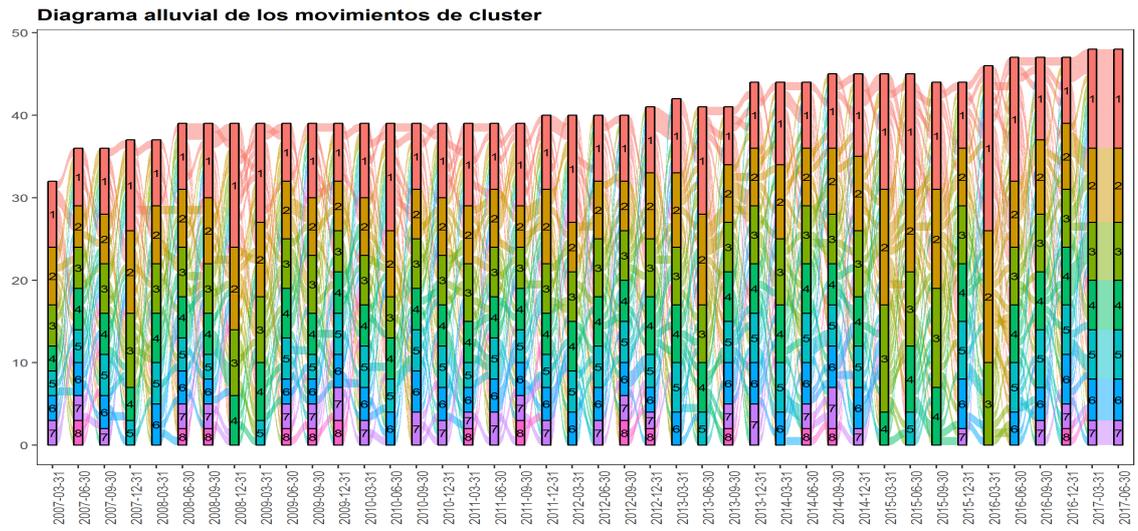


Figura 3.8: Ajuste FKM N2 SinCol.

Por último, vemos que las agrupaciones resultantes no son muy diferentes de las que se observan en la gráfica anterior para el tercer nivel de desagregación (véase 3.9).

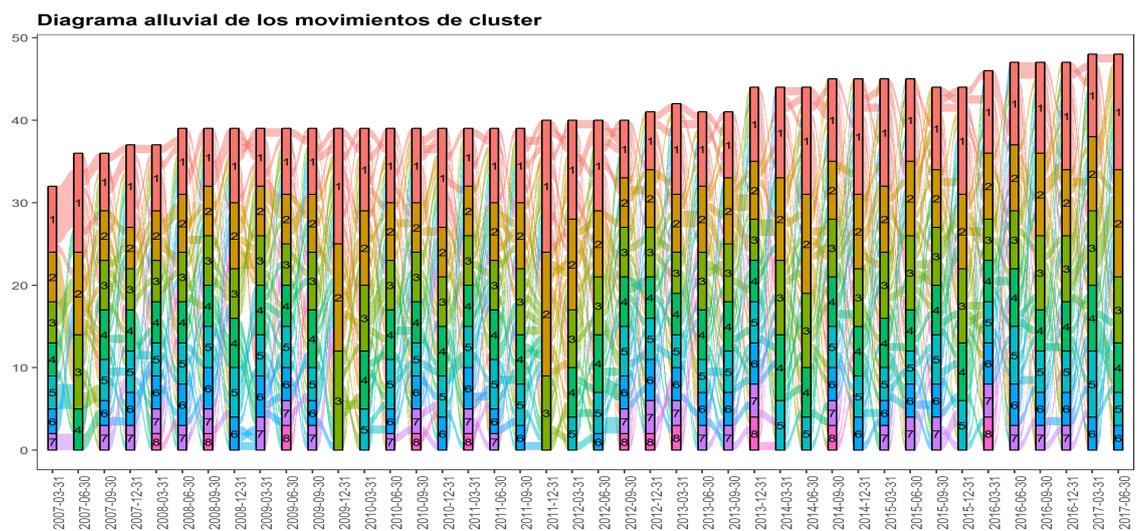


Figura 3.9: Ajuste FKM N3 SinCol.

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

En general, los ajustes del modelo **FKM** son poco estables, y son sumamente sensibles a la incorporación de nueva información, lo que, sumado a la variabilidad del criterio de optimalidad (ver 3.2), da evidencia de que el método no es el mejor para manipular los datos que estamos utilizando.

3.5.1.2. Ajustes con K-Medias Reducido

Análogo al análisis anterior, ahora revisamos los resultados obtenidos con el método K-Medias Reducido. Primero, observamos las agrupaciones para el primer nivel de desagregación de las cuentas de balance (ver 3.10). La reducción en la complejidad de los resultados es notable respecto a los ajustes con FKM del mismo nivel, pues no sólo los dos ajustes se encuentran en un espacio de dos dimensiones, también el número de agrupaciones es menor, y la dispersión de las instituciones en el espacio es mucho más realista, pues en los ajustes anteriores en los que un grupo es un solo punto se estaría asumiendo que todos los bancos en un mismo grupo son idénticos.

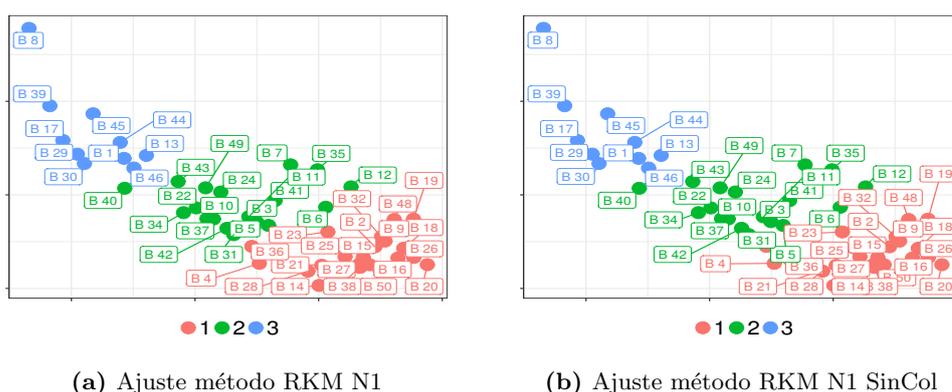


Figura 3.10: Visualizaciones de los ajustes con RKM para el nivel uno de desagregación para la última fecha de estudio.

En el segundo nivel de desagregación se observa que los ajustes tienen un patrón de dispersión similar al de las agrupaciones del nivel 1 (ver 3.11), con la diferencia de que se tiene un grupo más, tanto en el ajuste con variables colineales como en el que contiene todas las cuentas. También, el espacio en el que se encuentran los grupos está rotado respecto a los ajustes del nivel uno, pero en términos prácticos esto no hace ninguna diferencia.

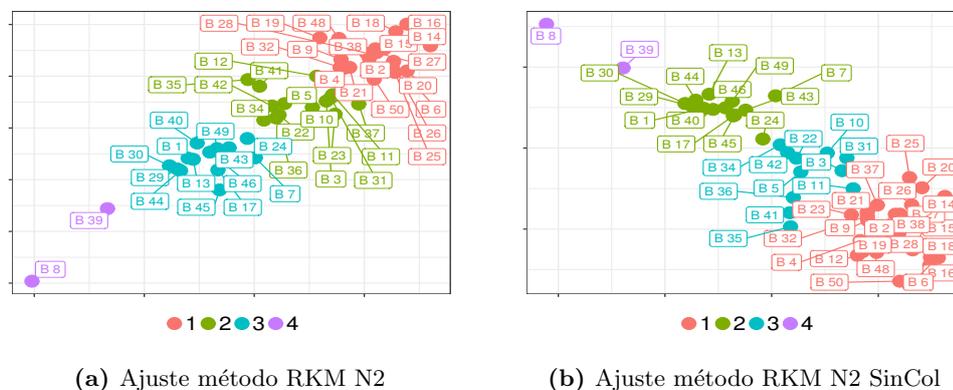


Figura 3.11: Visualizaciones de los ajustes con RKM para el nivel dos de desagregación para la última fecha de estudio.

Finalmente, llama la atención que en el último nivel de desagregación (véase 3.12) el método de K-Medias Reducido reduce de nuevo el número de agrupaciones encontradas, sin embargo, son los mismos bancos que en los ajustes con el nivel dos se “despegan” para crear su propio grupo los que se encuentran más separados del resto de los bancos, lo que nos da evidencia de consistencia en los ajustes.

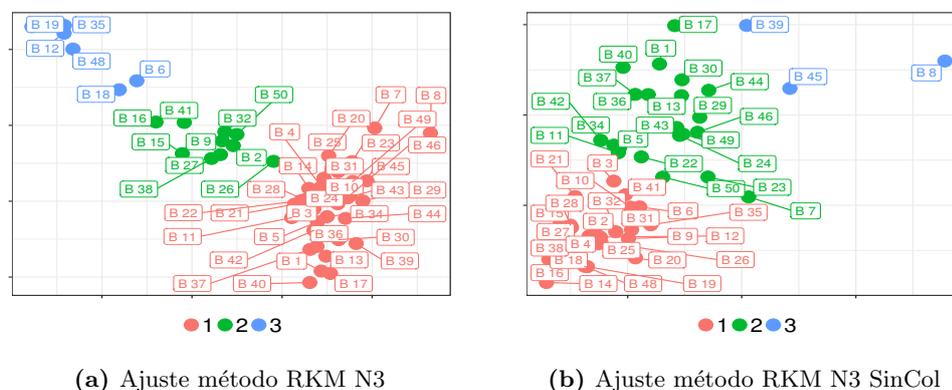


Figura 3.12: Visualizaciones de los ajustes con RKM para el nivel tres de desagregación para la última fecha de estudio.

Es muy importante que recordemos que todas las visualizaciones de las agrupaciones obtenidas con ambos métodos corresponden a **una sola fecha** de nuestro análisis, a saber, son resultados para junio de 2017. Por lo anterior, vemos también el comportamiento a lo largo del tiempo apoyándonos en los diagramas aluviales de los ajustes con el método RKM. Dado que todos los ajustes presentan agrupaciones relativamente

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

similares entre sí, nos enfocaremos en los ajustes que no consideran las variables colineales, pues además de que no consideran información redundante, presentan valores consistentemente más bajos en el criterio de optimalidad.

En el primer nivel notamos que el número de grupos es relativamente estable (ver 3.13), salvo junio 2009 y marzo 2012, en las que el método obtiene 7 y 6 grupos respectivamente. También es necesario resaltar la manera en la que el tercer grupo ha crecido con el paso del tiempo, lo que da evidencia de que hay un conjunto de instituciones que poco a poco han convergido a un modelo de negocios similar.

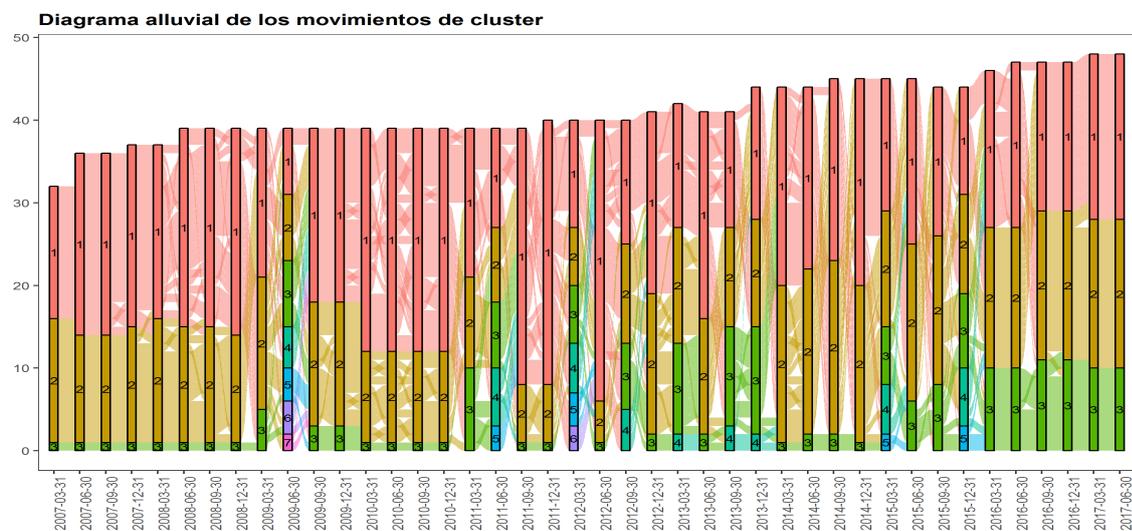


Figura 3.13: Ajuste RKM N1 SinCol.

El gráfico aluvial para el segundo nivel de desagregación (ver 3.14) muestra de igual forma un patrón casi constante en el número de agrupaciones encontradas por el método. Ahora son los ajustes correspondientes a septiembre de 2009 y de marzo de 2017 los que tienen saltos importantes en el número de grupos.

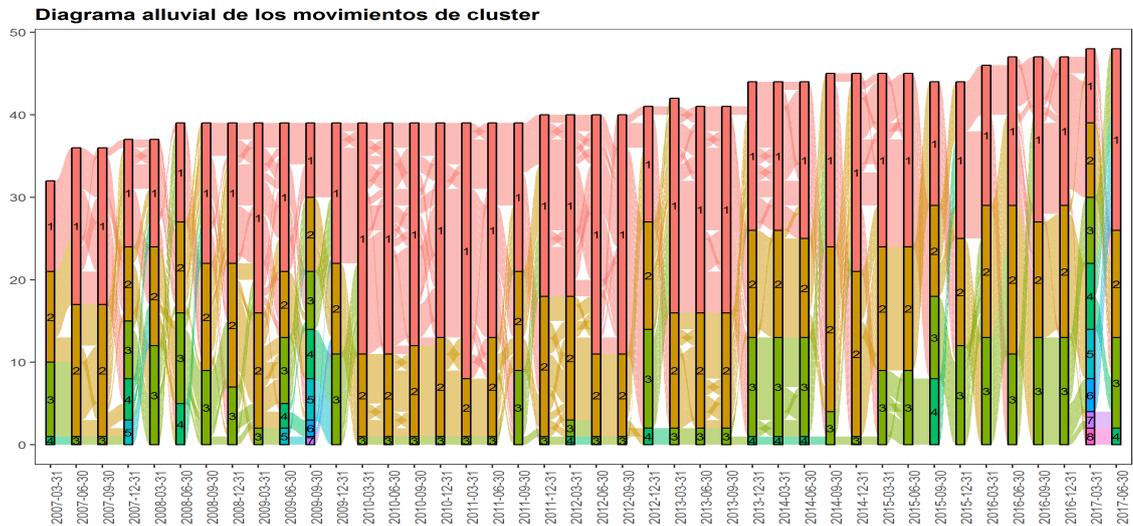


Figura 3.14: Ajuste RKM N2 SinCol.

Por último, en el diagrama alluvial del tercer nivel de agregación (ver 3.15) podemos observar movimientos fuertes en la estructura de agrupación únicamente para diciembre de 2007 y marzo de 2008, y el mismo comportamiento semi constante, con el número de grupos moviéndose entre 3 y 4 grupos.

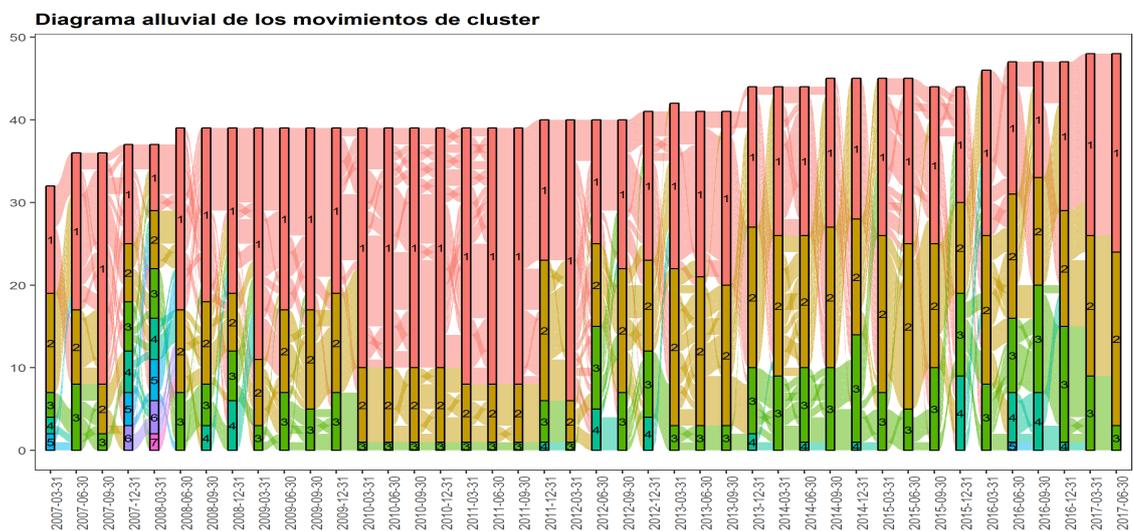


Figura 3.15: Ajuste RKM N3 SinCol.

Como vimos en todos los ajustes, hay un conjunto de fechas para las que en al menos uno de los ajustes hay movimientos significativos en el número de grupos encontrados. En la mayoría de esas fechas se dieron acontecimientos importantes en el mundo. Para

3. MODELOS DE CONGLOMERADOS Y DE REDUCCIÓN DE DIMENSIÓN

empezar, el 2008 está marcado por el inicio de la peor crisis financiera mundial en las últimas décadas, cuyo efecto es notable incluso desde finales del 2007 en muchos de los diagramas aluviales. También hay algunos ajustes con perturbaciones en septiembre de 2009, cuya causa muy probablemente sea la caída del PIB Mexicano por casi 8% en el segundo trimestre del mismo año. Así, tenemos evidencia de que los modelos de negocios de los bancos en México son sensibles al entorno económico y financiero en el que se desenvuelven.

Árboles de decisión y Bosques aleatorios

Con los métodos de conglomerados de la sección anterior somos capaces de asignar una institución en un grupo determinado de acuerdo con las características que comparte con el resto de las instituciones en el mismo grupo. Como mencionamos, las técnicas anteriores pertenecen al aprendizaje no supervisado (*unsupervised learning*).

También dentro del aprendizaje estadístico tenemos técnicas de aprendizaje supervisado (*supervised learning*). Dichas técnicas pueden ser utilizadas para diferentes tareas, aunque, primordialmente, nos sirven para hacer regresión (predicción/aproximación de una variable dependiente) y clasificación de objetos en un grupo homogéneo (predicción de una variable cualitativa). Las siguientes definiciones fueron obtenidas de [James et al. \(2014\)](#), complementando algunas ideas con [Hastie et al. \(2009\)](#).

A diferencia de los métodos no supervisados, en el aprendizaje supervisado se cuenta con una variable y_i (valor observado o clase observada, dependiendo del problema), que se intenta aproximar/predecir a partir de un conjunto de variables explicativas x_1, x_2, \dots, x_p .

En los modelos de aprendizaje supervisado, dada la presencia de la variable dependiente y , los datos se dividen en dos subconjuntos de observaciones: los datos de entrenamiento del modelo (*training data*) y los datos de prueba (*test data*). Generalmente, los datos de entrenamiento son la mayoría de los registros que se tengan, y los datos de prueba se utilizan para medir qué tan bueno es el modelo que se está obteniendo, pues se tiene un valor real observado de la variable dependiente que puede ser contrastado con el valor que nos arroje el modelo ajustado a los datos.

Sin embargo, hay ocasiones (nuestra aplicación cae en ese caso) en las que el tamaño de nuestro conjunto de datos no es suficiente para obtener un subconjunto de datos de prueba idóneo para obtener una comprobación creíble. Por lo anterior, existen métodos para validar un modelo que no implican partir el conjunto de datos en dos de manera directa, dentro de dichos métodos los más utilizados son la **Validación cruzada** (*Cross-Validation*, ver [Stone \(1974\)](#)) y el **Remuestreo (Bootstrap)** (véase [Efron \(1979\)](#)) los cuales se describen con mayor detalle en el apéndice A. Para obtener un modelo se tienen que realizar dos estimaciones diferentes: de los parámetros estructurales del mismo (por ejemplo, los coeficientes en un modelo de regresión, o los valores

de corte para partir el espacio de las variables predictivas en un modelo de árbol), así como parámetros de ajuste del modelo (por ejemplo, el grado en una regresión polinomial, o el número de ramas óptimo en un árbol). La validación cruzada y el remuestreo nos pueden ayudar a definir los mejores parámetros de ajuste de un modelo, pues nos permiten ajustar varios modelos en diferentes subconjuntos de datos y comparar los resultados de todos ellos.

Los modelos de aprendizaje supervisado son, además, sumamente útiles para medir e interpretar la importancia de las variables explicativas $\{x\}_1^n$, pues es posible obtener valores ligados a cada variable que nos ayudan a medir el impacto (positivo o negativo) que tiene cada una para la estimación de la variable dependiente y , además de la cantidad de información que aporta cada una a la predicción de ésta, es decir, su *significancia estadística*.

Los modelos de aprendizaje supervisado son sumamente flexibles, pero esto presenta un problema, pues se pueden volver tan complejos como se permita, lo que irremediablemente atenta contra la interpretabilidad del modelo, si ésta es prioridad del estudio un modelo con un alto grado de complejidad no es deseable, si, por el contrario, la exactitud de las predicciones es lo más importante, un modelo con mayor grado de complejidad será capaz de devolver resultados más apegados a la realidad. Con lo anterior, el investigador se enfrenta a un problema: hacer una concesión entre la interpretabilidad del modelo y la exactitud de las predicciones. A pesar de lo anterior, otro concepto importante en el aprendizaje estadístico, el “sobreajuste” (*overfitting*) nos dice que no siempre los modelos más complejos arrojarán los mejores resultados.

El sobreajuste se da cuando el modelo ajusta de manera casi perfecta *los datos de entrenamiento*, sin embargo, dicho modelo puede tener una varianza demasiado pequeña, lo que puede dejar fuera del rango de predicción observaciones para las que se desee aproximar su variable dependiente con el modelo ajustado. Por lo anterior, es deseable obtener modelos que tengan un buen rendimiento en los datos de prueba, no sólo en los datos de entrenamiento. Por ejemplo, cuando se está haciendo regresión (predicción de una variable continua), se busca minimizar el **Error Cuadrático Medio** (*Mean Squared Error*); lo definimos como:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2, \quad (4.1)$$

donde \hat{f} es el modelo ajustado con los datos de entrenamiento. Sin embargo, dado que predecir el valor de y para observaciones nuevas es importante, estamos más interesados en minimizar el error cuadrático medio **de los datos de prueba**. Con lo anterior, el sobreajuste (*overfitting*) se da cuando un modelo tiene un error cuadrático medio pequeño para los datos de entrenamiento, y un valor grande para los datos de prueba, es decir, cuando un modelo “aprende” el patrón exacto de los datos de entrenamiento.

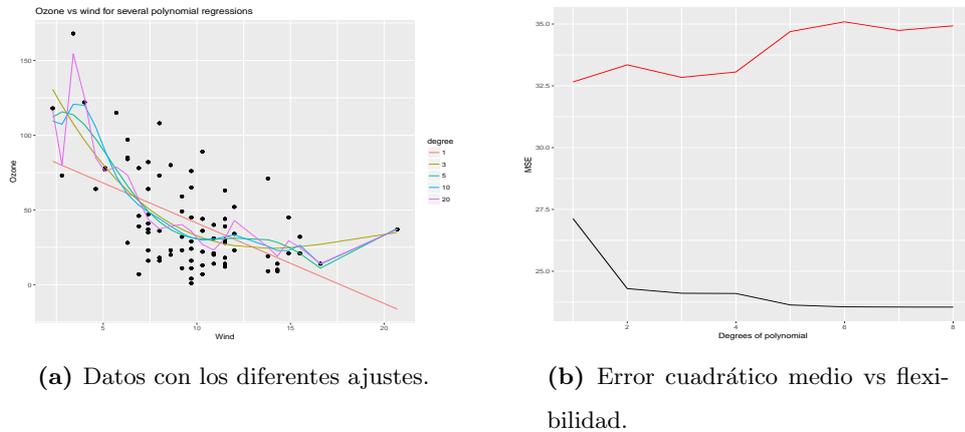
Finalmente, existe un resultado más que nos ayuda a medir la calidad de un modelo, además de ayudarnos a elegir el mejor ajuste de manera informada: la concesión entre varianza y sesgo (*Bias-Variance Trade-off*), el cual nos dice que el error cuadrático

medio de los datos de prueba siempre se puede descomponer en tres cantidades:

$$E(y_0 - \hat{f}(x_0))^2 = Var(\hat{f}(x_0)) + [Sesgo(\hat{f}(x_0))]^2 + Var(\varepsilon), \quad (4.2)$$

donde la varianza mide la sensibilidad de \hat{f} a cambios en los datos que recibe como entrada, es decir, qué tanto cambiarían las predicciones bajo un conjunto de datos distinto. Por otro lado, el sesgo se refiere al error inherente en el que se incurre al intentar aproximar un fenómeno complejo real con un modelo que lo simplifica. Mientras que los modelos más complejos son capaces de aproximar mejor un conjunto de datos (y por lo tanto, incurrir en un sesgo menor), son sumamente sensibles a cambios en el conjunto de datos (por lo que su varianza aumenta como consecuencia). Por otro lado, modelos con un bajo grado de complejidad manejan mejor cambios en los datos con lo que se alimenta, pues son mucho más restrictivos (menor varianza), sin embargo, su capacidad de aproximar un fenómeno real se ve disminuida como consecuencia (mayor sesgo).

Para ejemplificar lo anterior replicamos la gráfica presentada en ([EnhancedDataScience, 2017](#)) (ver gráfica 4.1). Se ajustan regresiones polinomiales con diferente grado (a mayor grado mayor flexibilidad del modelo) para medir la relación (queremos hacer inferencia) entre la velocidad del viento y la cantidad de ozono en el aire ¹.



(a) Datos con los diferentes ajustes.

(b) Error cuadrático medio vs flexibilidad.

Figura 4.1: Ejemplo de overfitting.

Tomando como variable dependiente y las asignaciones de grupo obtenidas con los métodos de la sección anterior, y como variables explicativas las cuentas de balance a las que aplicamos las técnicas de conglomerados, usamos una técnica de aprendizaje supervisado, *Bosque aleatorio (random forest)*, ver [Breiman \(2001\)](#), para medir la importancia de las variables explicativas en la clasificación de las instituciones, esto con

¹Vemos que conforme crece el grado del polinomio, más se ajusta el modelo a los datos (figura a), sin embargo, el error cuadrático medio (MSE) crece con el grado del polinomio (figura b, la línea roja corresponde al MSE de los datos de prueba, y la línea negra al MSE de los datos de entrenamiento).

la intención de adquirir intuición sobre qué variables son consideradas por el algoritmo como las más representativas de la estructura de agrupación de nuestros objetos (bancos).

4.1. Árboles de decisión

Los bloques básicos de construcción de los bosques aleatorios son los árboles de decisión (*decision trees*, ver [Breiman et al. \(1984\)](#)). Como vemos más adelante, los árboles de decisión tienen ciertas desventajas naturales, mismas que dieron pie al desarrollo de los bosques aleatorios.

Al igual que los modelos de regresión, podemos utilizar árboles de decisión para hacer predicciones de una variable continua, así como para clasificar nuestras observaciones en grupos con características similares. De manera general, los modelos basados en árboles tienen como objetivo segmentar el espacio generado por las variables explicativas en tantos subespacios como sea necesario de acuerdo con las reglas de decisión sobre dichas variables. Cuando tenemos como objetivo hacer una predicción, el resultado será el promedio de todas las observaciones en el subespacio al que pertenece la observación para la que se desea obtener un valor, y cuando se desea clasificar un objeto, la partición del espacio va dando reglas sobre las variables explicativas para poder clasificar una observación.

4.1.1. Árboles de regresión

Iniciamos con una matriz $A_{n \times p}$ y un vector y , donde cada entrada del vector y es la variable dependiente de cada vector $x_{1 \times p}$. El objetivo del árbol es partir el espacio generado por las variables predictivas X_1, X_2, \dots, X_p en J regiones excluyentes R_1, R_2, \dots, R_J , finalmente, para todas las observaciones que caigan en cada una de las J regiones se asigna el mismo valor (predicción): el promedio de la variable respuesta y de los objetos en cada región.

La mejor partición del espacio de variables predictoras lo da aquella que minimice la suma de los errores al cuadrado (*Residual Sum of Squares*, o RSS por sus siglas en inglés), dada por:

$$\sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2, \quad (4.3)$$

donde \hat{y}_{R_j} es el promedio de la variable respuesta en la región j .

Sin embargo, a medida que crece el número de variables independientes (p), la viabilidad de calcular todas las combinaciones posibles de regiones decrece, y recordemos que nuestros datos tienen una dimensión alta (p alrededor de 400 variables cada mes), por lo que se necesita otro enfoque para resolver el problema.

Derivado de lo anterior, se define el algoritmo de **Partición Binaria Recursiva** (*Recursive binary splitting*, ver [Breiman et al. \(1984\)](#)). Dicho algoritmo construye de “*arriba hacia abajo*” el árbol de manera “*codiciosa*” (*greedy*).

Se dice que la construcción es de “*arriba hacia abajo*” porque al inicio del algoritmo se tienen todas las observaciones en un mismo grupo, posteriormente se parte sucesivamente el espacio de variables predictoras de acuerdo a un valor de corte para cada variable independiente. Adicionalmente, se dice que es un algoritmo “*codicioso*” porque en cada iteración busca la partición que provoque la mayor disminución en la suma de los errores al cuadrado, en lugar de buscar todas las particiones posibles con todas las variables.

Para cada paso en el algoritmo, se selecciona la variable $X_j, j \in \{1, \dots, p\}$ y un valor de corte s , dicho valor cumple la propiedad de reducir lo máximo posible la suma de los errores al cuadrado. Así, el espacio de predictores queda partido en las regiones:

$$R_1(j, s) = \{X|X_j < s\} \text{ y } R_2(j, s) = \{X|X_j \geq s\} \quad (4.4)$$

Se seleccionan valores para j y s tales que minimicen la ecuación:

$$\sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2, \quad (4.5)$$

donde \hat{y}_{R_i} es el promedio de las observaciones que caen en la región R_i .

El procedimiento anterior se realiza sucesivamente hasta que dejemos de tener variables predictivas o hasta que se alcance algún criterio de paro definido para el algoritmo. En cada paso se divide la rama creada en el paso anterior, no el espacio entero. Finalmente, cuando se desea aproximar un valor (como es el caso en los árboles de regresión), a cada nueva observación se le asigna el promedio de la variable y de las observaciones de entrenamiento en la misma región.

Para ejemplificar la interpretabilidad de los árboles de decisión, recreamos la gráfica propuesta en [Kabacoff \(2017\)](#), en ella, es posible observar la visualización de un árbol de regresión, que aproxima el rendimiento de combustible de diferentes automóviles de acuerdo con cuatro variables: precio, país, confiabilidad y tipo de auto (ver Figura 4.2).

Podemos observar que las variables de confiabilidad y país de origen del auto no son consideradas en el mejor modelo ajustado de manera automática por el software R, además, siguiendo una de las ramas del árbol, los autos con un precio mayor a \$9,446 dólares, de tipo mediano-grande-camioneta, y en una partición más fina, de grande-camioneta, tienen el rendimiento más bajo de combustible (19 millas por galón), mientras que los autos con un precio menor a \$9,446 dólares tienen el rendimiento más alto, con 32 millas por galón.

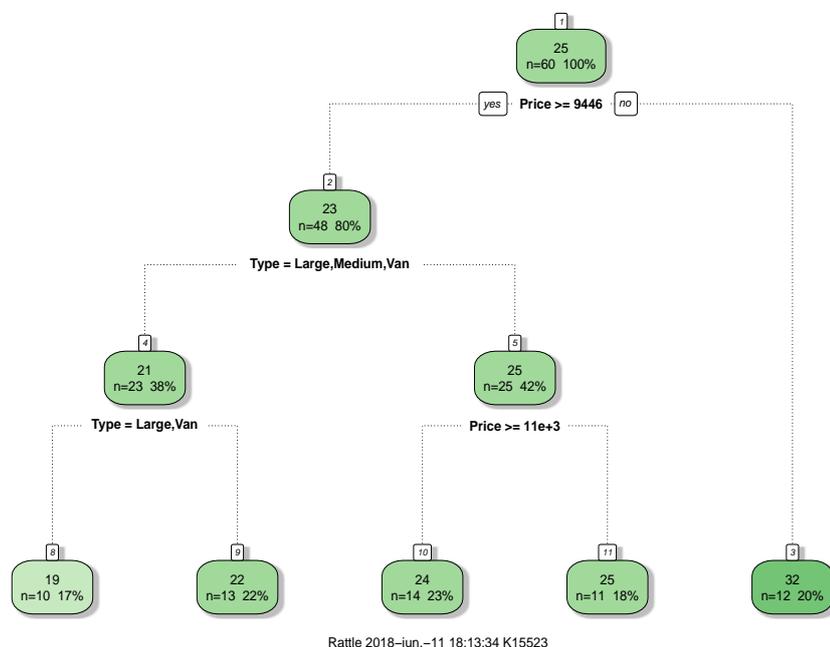


Figura 4.2: Ejemplo de un árbol de regresión.

4.1.2. Árboles de clasificación

Bajo los mismos supuestos sobre la estructura de los datos que se tienen, se pueden utilizar árboles de decisión con la finalidad de clasificar objetos de acuerdo con sus características (variables $X_1 \dots X_p$); la diferencia con los árboles de regresión es que la variable que se desea predecir no es continua, es una variable discreta donde cada valor que puede tomar representa un grupo diferente de observaciones.

Para la construcción de un árbol de clasificación también se utiliza la Partición Binaria Recursiva (Breiman et al. (1984)). Luego, cuando se hacen predicciones para nuevas observaciones con el modelo ajustado, cada nuevo objeto se asigna a la clase más común en la región en la que cae. Sin embargo, dado que la variable que se quiere predecir no es continua, no es posible utilizar la suma de errores cuadrados como criterio de optimalidad, pues la ecuación 4.3 no está definida para valores cualitativos de y ; en la práctica se tienen dos alternativas para la construcción del árbol: el *coeficiente GINI* (ecuación 4.6) y la *entropía* (ecuación 4.7).

Tomando \hat{p}_{mk} como la proporción de observaciones en la región m (ramas en el árbol) que pertenecen a la clase k (k es conocido a priori, pues es el número de clases observado en los datos), definimos al coeficiente GINI como:

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk}). \quad (4.6)$$

El coeficiente GINI también es conocido como la *pureza* de los nodos del árbol, pues mide qué tan homogéneos son éstos según las clases de las observaciones que se encuentren en ellos. Por ejemplo, si $\hat{p}_{mk} = 1$ para un nodo, quiere decir que todas las observaciones en el nodo m pertenecen a la clase k , por lo que $G = 0$. En general, si G es pequeño, quiere decir que la mayoría de las observaciones en m pertenecen a una misma clase.

Como alternativa al coeficiente de GINI se utiliza la entropía, que se define como:

$$D = - \sum_{k=1}^K \hat{p}_{mk} \log(\hat{p}_{mk}). \quad (4.7)$$

De manera similar al coeficiente GINI, si la entropía es pequeña, es señal de que la mayoría de las observaciones en la rama m pertenecen a una misma clase.

Dadas las dos medidas anteriores, el proceso de construcción del árbol es el mismo que para los árboles de regresión, con la diferencia de que en cada paso se busca la partición que presente el valor óptimo de pureza en la partición.

Análogo al ejemplo del árbol de regresión, podemos obtener un ejemplo de cómo se construye un árbol de decisión para clasificar pacientes con una deformación en la columna vertebral llamada cifosis después de someterse a una operación (gráfica 4.3), utilizando tres variables: edad, número de vértebras involucradas en la operación y la vértebra más alta (en la espalda).

Vemos que un paciente con una vértebra de inicio menor a “8.5”² tiene la deformación; siguiendo otra rama del árbol donde la deformación está presente, vemos que ésta se presenta en pacientes con vértebra de inicio menor a 14 y una edad entre 55 y 111 años. (Fuente: (Kabacoff, 2017)).

²Este número hace referencia a la vértebra más alta que fue manipulada en la operación, empezando de las vértebras más cercanas al coxis.

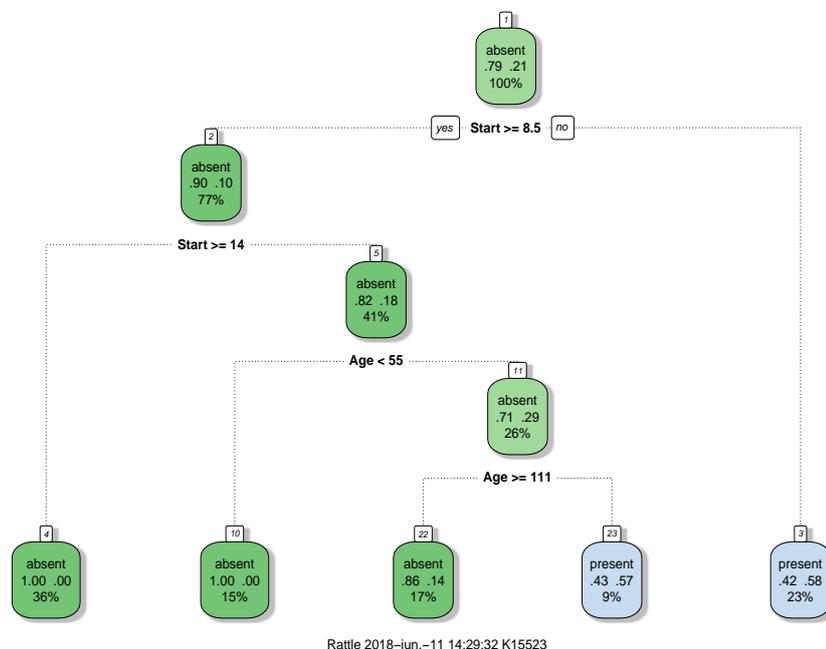


Figura 4.3: Ejemplo de un árbol de clasificación.

4.1.3. Poda de un árbol

Cuando se tiene un número grande de variables explicativas (es decir $p \gg 1$), es fácil que un árbol ajustado a los datos se vuelva sumamente complejo, lo que provoca, como ya vimos, que el modelo resultante presente sobreajuste, y por consiguiente, que las predicciones del modelo sean de baja calidad. Aquí se nos presenta el dilema del **intercambio sesgo-varianza** (*Bias-Variance Trade-off*, Geman et al. (1992)), donde perdemos credibilidad al simplificar demasiado la solución de un problema complejo o aumentamos la variabilidad con datos fuera del conjunto de entrenamiento.

La *poda* de un árbol de decisión es una técnica que se desarrolló para luchar contra estos problemas, suponiendo que un árbol con menos nodos finales (regiones R_i) presenta menor probabilidad de tener sobreajuste y es mucho más interpretable. Por lo anterior, la poda de un árbol consiste en iniciar con un árbol grande T_0 (puede incluso considerar las p variables) y cortar (*podar*) ramas del árbol para obtener subárboles.

La decisión de cómo podar el árbol no es aleatoria. Dado que calcular el MSE de prueba para todos los árboles posibles resultaría computacionalmente ineficiente, especialmente si el número de variables independientes p es muy grande (como es nuestro caso). Por lo anterior se define la “**Poda con costo en complejidad**” (*Cost complexity pruning*). Dado que un árbol demasiado grande tiene una complejidad muy alta,

se utiliza un parámetro de ajuste α . Para cada valor α hay un subárbol $T \subset T_0$ tal que

$$\sum_{m=1}^{|T|} \sum_{i: x_i \in R_m} (y_i - \hat{y}_{R_m})^2 + \alpha |T|, \quad (4.8)$$

donde $|T|$ es el número de nodos finales del árbol, R_m es el m -ésimo nodo final, y el resto de variables corresponden a las definiciones anteriores. Notamos que la fórmula para calcular el error promedio cuadrado está incluida, más un factor que depende de α y del tamaño del subárbol, por lo que, controlando el valor de α controlará a su vez el tamaño (y por lo tanto la complejidad) del árbol resultante. De esta forma se puede construir una serie de subárboles que dependerán de dicha variable. Por último, para seleccionar un valor adecuado de α se puede utilizar validación cruzada (*Cross-Validation*, Stone (1974)) o partiendo los datos y utilizando el conjunto de prueba.

4.2. Bosques Aleatorios

Como se mencionó al principio del capítulo, los árboles de decisión presentan algunos problemas asociados a su misma naturaleza, los cuales van muy ligados al principio del *intercambio sesgo-varianza* (Geman et al. (1992)). Dado que las particiones del espacio de variables dependientes no son lineales bajo los árboles de decisión, y que se pueden tomar tantas de ellas como se desee, los modelos resultantes pueden ser demasiado complejos (lo que nos lleva a un problema si se quiere hacer un inferencia con dicho modelo) y tener un problema de sobreajuste o de sesgo grave que harán que el modelo falle al momento de hacer nuevas predicciones.

Para sobreponerse a dichos problemas, se desarrollaron los **bosques aleatorios** a partir de los métodos de remuestreo descritos en el apéndice A, en particular, definimos la **Agregación Bootstrap** o **Bagging** (*Bootstrap aggregation*). Gracias a la teoría estadística, sabemos que la varianza del promedio de n observaciones Z_1, \dots, Z_n cada una con varianza σ^2 está dada por $\frac{\sigma^2}{n}$, es decir, promediar la varianza de las n observaciones *reduce* la varianza. Siguiendo el mismo principio y utilizando el remuestreo, podemos construir tantos conjuntos de entrenamiento como se quieran, construir un modelo diferente con cada uno y promediar los resultados de todos ellos (justo lo que hacen los bosques aleatorios). Es decir, si tenemos B conjuntos de entrenamiento diferentes, cada uno construido usando remuestreo, podemos obtener $\hat{f}^1(x), \hat{f}^2(x), \dots, \hat{f}^B(x)$ ajustes, con los cuales obtenemos el modelo

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x), \quad (4.9)$$

el cual tendrá una varianza menor a la varianza individual de los B modelos ajustados.

Dado lo anterior, la agregación bootstrap consiste en construir B conjuntos de entrenamiento usando remuestreo, a partir de los cuales construimos B árboles de decisión,

4. ÁRBOLES DE DECISIÓN Y BOSQUES ALEATORIOS

los cuales se promedian al final para obtener el modelo \hat{f}_{bag} definido. Es importante mencionar que cada uno de los B árboles de decisión se construye “profundamente” (tomando todas las variables del conjunto de predictores) y **no se podan**.

Lo anterior es la especificación para hacer predicciones de una variable Y cuantitativa, sin embargo, para resolver un problema de clasificación lo único que tenemos que hacer es clasificar una observación en la categoría que tenga mayoría dentro de los B árboles. Una característica más de los árboles de decisión que resulta conveniente es que incrementar el tamaño del parámetro B no nos lleva a tener sobreajuste, pues el error converge conforme B se hace grande.

Una vez que tenemos los modelos, sigue siendo necesario hacer una estimación de qué tan bueno es el modelo obtenido, para ello existe un resultado que demuestra que cada árbol construido con el procedimiento anterior utiliza dos tercios de las n observaciones, al tercio restante se le conoce como las observaciones “**fuera-de-la-bolsa**” (*OOB*), las cuales son usadas para calcular el error promedio. Dado que se tienen B árboles y cada uno utiliza alrededor de dos tercios de las observaciones, nos quedamos con alrededor de $\frac{B}{3}$ predicciones para cada observación, las cuales podemos promediar (cuando se está haciendo regresión) o contar y obtener la categoría con mayoría (en clasificación), con lo que nos quedamos con una sola predicción para cada observación, las cuales llamamos “predicciones fuera-de-la-bolsa”.

Cuando usamos agregación bootstrap para construir un modelo promedio utilizamos las p variables predictivas disponibles, las cuales a su vez pueden variar en importancia para la clasificación de las observaciones, por lo que todos los árboles tendrán un sesgo a favor de las más importantes, teniendo como consecuencia que todos los árboles sean muy parecidos entre sí, lo que a su vez puede llevarnos a tener predicciones altamente correlacionadas. Por lo anterior, se desarrollaron los **Bosques Aleatorios**; en éstos, los árboles de decisión no contienen el mismo conjunto de variables de predicción, pues para un conjunto de entrenamiento construido con remuestreo se eligen de manera aleatoria m variables independientes, donde $m < p$, así, se garantiza que los árboles serán distintos entre ellos y por consiguiente las predicciones tendrán una correlación más pequeña. En la literatura es común encontrar $m \approx \sqrt{p}$

Para efectos del presente trabajo, una de las características más importantes de los bosques aleatorios es que nos permiten obtener la **importancia** de las variables independientes para la predicción de la variable dependiente y . En cada uno de los B árboles del bosque tenemos la capacidad de obtener qué tanto mejora un árbol de acuerdo con qué tanto decrecen los errores al cuadrado (RSS) en los árboles de regresión, y qué tanto decrece el índice GINI para los árboles de clasificación. Se promedia dicho valor para los B árboles y se obtiene la importancia, mientras más grande sea ésta, más importante es una variable para la variable dependiente, pues mayor es la exactitud de las predicciones.

4.3. Resultados

Una vez que hemos obtenido una agrupación valiéndonos del aprendizaje no supervisado en el capítulo anterior, es necesario dar una explicación a los resultados, pues en nuestro problema es importante saber qué es lo que está provocando que las instituciones sean agrupadas de una u otra forma. Por lo anterior, obtuvimos la importancia de las cuentas de balance para la obtención de una agrupación dada con el método del Bosque Aleatorio.

Los bosques aleatorios fueron ajustados utilizando la biblioteca VSURF (Genuer et al. (2018)), que a su vez hace uso de la biblioteca randomForest (Liaw and Wiener (2002)). Dichas paqueterías nos permiten realizar múltiples ajustes y obtener así la importancia de las cuentas promediada a través de un número grande de simulaciones, mejorando así las estimaciones obtenidas, más aún, el paquete VSURF está diseñado específicamente para la selección de variables haciendo uso de la importancia que los bosques aleatorios asignan a las variables independientes de un conjunto de datos.

Para los resultados de los tres ajustes obtenidos con el método de RKM sobre los datos sin variables colineales, obtuvimos la evolución de la importancia de las cuentas de balance más importantes para la estructura de agrupación según el método de Bosque Aleatorio.

Como mencionamos, la biblioteca que estamos utilizando hace de manera automática una selección de variables, eliminando aquellas cuya importancia es despreciable, así para cada uno de los ajustes seleccionados en el capítulo anterior (los obtenidos con el método RKM quitando las variables colineales) obtuvimos la evolución de la importancia para las variables seleccionadas en cada fecha.

En las siguientes gráficas (ver 4.4 a 4.9) presentamos la medida utilizada para dar importancia a una variable: decaimiento promedio de la precisión, dicha medida nos dice qué tanto se movería (en sentido negativo) la precisión de un árbol de clasificación al remover una variable del mismo, por lo que mientras más alto es el valor en las gráficas más importante es la variable para lograr la estructura de clasificación encontrada por el método de conglomerados.

Por cuestiones de calidad en la visualización, utilizamos dos criterios para seleccionar las variables que se presentan en las gráficas que mostramos a continuación: la frecuencia con la que aparecen en las variables seleccionadas por la función (recordemos que tenemos 42 ajustes distintos, uno para cada fecha de nuestro periodo de estudio), y el promedio de la clasificación ³ a lo largo de todas las veces que aparece en las variables más importantes. Así, sólo graficamos las variables que aparecen para al menos el 80% de nuestras fechas, y que tienen el promedio de clasificación más bajo.

Para el primer nivel de granularidad en las cuentas, la gráfica 4.4 muestra que las variables más influyentes para las agrupaciones obtenidas son el capital contribuido y

³Aquí, utilizamos la palabra clasificación en el sentido de orden o posición dentro de un listado ordenado de acuerdo a un valor, en este caso, respecto a su importancia de acuerdo al método de bosque aleatorio.

4. ÁRBOLES DE DECISIÓN Y BOSQUES ALEATORIOS

el capital ganado; después, a pesar de tener valores mucho menores, la importancia de la cartera de crédito vigente, la captación tradicional y las inversiones en valores tienen saltos importantes. Teniendo en mente el conjunto de fechas con saltos importantes en los ajustes de grupos, llama la atención el salto en la importancia de la inversión en valores y la cartera de crédito vigente en la segunda mitad del 2009, lo que nos da evidencia de la fuerza con la que los bancos mexicanos ajustaron sus operaciones como respuesta a la crisis financiera y a la caída del PIB en el país.

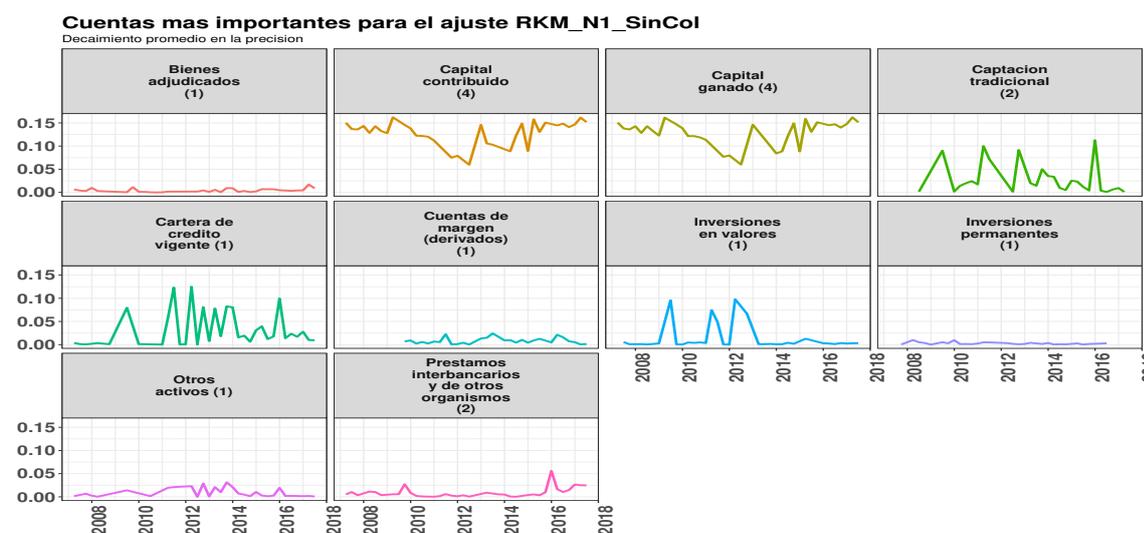


Figura 4.4: Variables más importantes de acuerdo al método de bosque aleatorio para llegar a la agrupación RKM N1 SinCol.

Hasta aquí, las cuentas que saltan en importancia sin duda pueden cargar información sobre los modelos de negocios, en particular la cartera de crédito, la inversión en valores y la captación tradicional, sin embargo, al ser cuentas muy generales, no aportan mayor intuición sobre las decisiones que toman los bancos. Pensemos por ejemplo en la cartera de crédito vigente, hay bancos en México que se especializan en crédito al consumo (como las tarjetas de crédito), otros que se especializan en crédito comercial para empresas, y otros que se especializan en créditos inmobiliarios, así, la cuenta “Cartera de crédito vigente” no nos da suficiente información para identificar un modelo de negocios.

Lo anterior es la razón principal por la que se decidió realizar ajustes con distintos niveles de granularidad de la información, así, presentamos la importancia del siguiente nivel de profundidad del catálogo mínimo (gráfica 4.5).

Primero, hay que notar que la mayoría de las series seleccionadas son subniveles de cuentas en la gráfica anterior, a saber:

1. Capital contribuido:
 - Incremento por actualización del capital social.

- Capital social.
2. Capital ganado:
 - Incremento por actualización de reservas de capital.
 - Incremento por actualización de resultados de ejercicios anteriores.
 - Reservas de capital.
 3. Captación tradicional:
 - Depósitos a plazo.
 4. Inversiones en valores:
 - Títulos para negociar.
 5. Inversiones permanentes:
 - Subsidiarias.
 6. Préstamos interbancarios y de otros organismos:
 - De largo plazo.

Lo anterior resulta significativo para nuestro estudio, pues es evidencia clara de que el método de conglomerados utilizado y la selección de variables con bosques aleatorios son consistentes con la estructura de los datos, lo que implica que el balance general de las instituciones bancarias contiene de manera inherente una estructura de agrupación entre instituciones; lo anterior se hace más claro si consideramos el caso en el que las variables más importantes para la selección en el segundo nivel no estuvieran relacionadas en lo absoluto con las elegidas en el primer nivel, situación que complicaría la interpretación y la justificación de los resultados obtenidos con ambos niveles.

Una vez más, las cuentas relacionadas al capital están presentes, y son además las que tienen los valores más altos para la métrica de importancia. Es importante mencionar que no hay ninguna cuenta relacionada con la cartera de crédito vigente, que sí está presente en el primer nivel, y tiene saltos significativos en su importancia.

4. ÁRBOLES DE DECISIÓN Y BOSQUES ALEATORIOS

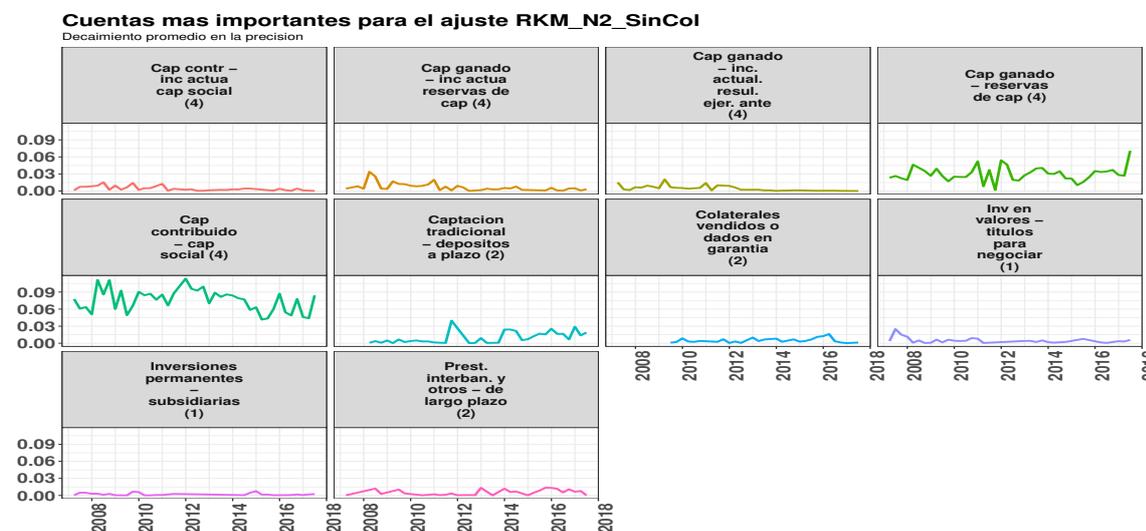


Figura 4.5: Variables más importantes de acuerdo al método de bosque aleatorio para llegar a la agrupación RKM N2 SinCol.

En el tercer grado de granularidad (gráfica 4.6) volvemos a tener cuentas en niveles subsecuentes a las cuentas en los dos niveles anteriores, y de nuevo son las variables de Capital las que dominan en importancia. Además, vuelve a aparecer una variable relacionada con la cartera de crédito vigente: créditos comerciales sin restricción. Por último, resulta relevante un cambio en las variables relacionadas a los valores, pues mientras que en los primeros dos niveles de granularidad las cuentas importantes relacionadas con valores son de la parte activa del balance, en el último nivel hay un claro cambio a la parte pasiva de las operaciones con valores.

Una cuenta del pasivo que tiene presencia en los tres niveles es la correspondiente a los préstamos interbancarios, sin embargo, las cuentas relacionadas a ésta para el segundo y tercer nivel difieren, pues en el segundo nivel tenemos préstamos de largo plazo, y el tercer nivel tenemos préstamos a corto plazo de instituciones de banca de desarrollo.

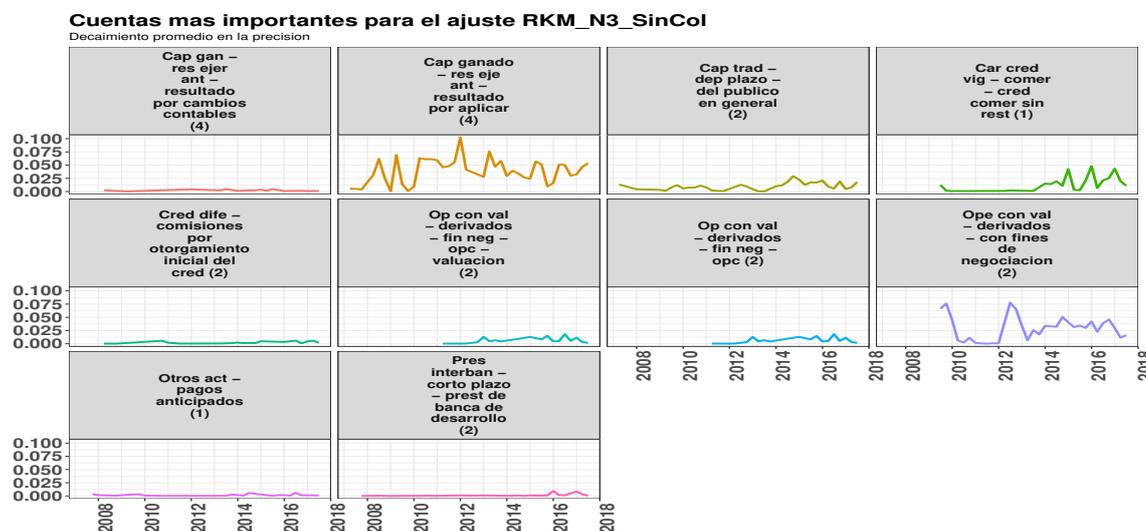


Figura 4.6: Variables más importantes de acuerdo al método de bosque aleatorio para llegar a la agrupación RKM N3 SinCol.

El cambio que se da entre cuentas de activo y de pasivo al pasar del segundo al tercer nivel demuestra que los resultados son sensibles a la incorporación de más información, por lo que es importante analizar hasta qué punto la granularidad de la información es benéfica para el análisis, y en cuál comienza a distorsionar los resultados.

4.3.1. Distribución de las cuentas más importantes por grupo

Con los diagramas aluviales presentados en el capítulo anterior y las gráficas de la evolución de la importancia de la sección anterior, queda claro que los modelos de negocios de los bancos mexicanos no son estáticos, más aún, son sensibles al ambiente económico en el que se desenvuelven. Por lo anterior, la tarea de obtener una agrupación resulta sumamente complicada al incluir el aspecto dinámico de la estructura de agrupación implícita.

Una de las consecuencias del aspecto dinámico de la agrupación es que el conjunto de variables importantes obtenidas con el bosque aleatorio también es propenso a cambiar, lo que complica aún más la identificación de conglomerados, por lo anterior, limitamos el análisis de las cuentas importantes a una fecha, específicamente la última fecha de nuestro periodo de estudio, junio 2017 para los análisis obtenidos con el método RKM quitando las variables altamente correlacionadas.

Ahora, observamos la distribución de las cuentas importantes seleccionadas por el bosque aleatorio para el primer nivel de desagregación (gráfica 4.7). Al ver la distribución de las cuentas de **Capital Contribuido** y de **Capital Ganado**, es claro por qué son seleccionadas como las más importantes, pues la distinción entre los tres grupos es evidente, siendo el grupo 1 el que tiene los valores más grandes (en valor absoluto).

4. ÁRBOLES DE DECISIÓN Y BOSQUES ALEATORIOS

La tercera y la cuarta cuenta tienen una interpretación directa y muy intuitiva para los modelos de negocios, pues vemos bien diferenciados (a pesar de tener traslapes importantes) los tres grupos y es claro que la cartera de crédito y la deuda interbancaria son actividades muy relacionadas con las decisiones tomadas para generar ganancias en una institución financiera.

Las inversiones en valores (séptima cuenta más importante) también reflejan fuertemente las decisiones de negocios de un banco. A pesar de que los tres grupos tienen distribuciones muy parecidas, es importante notar que por cada grupo hay instituciones que dan mayor peso a los valores dando como resultado la existencia de que los grupos se diferencian por lo pesado de sus colas.

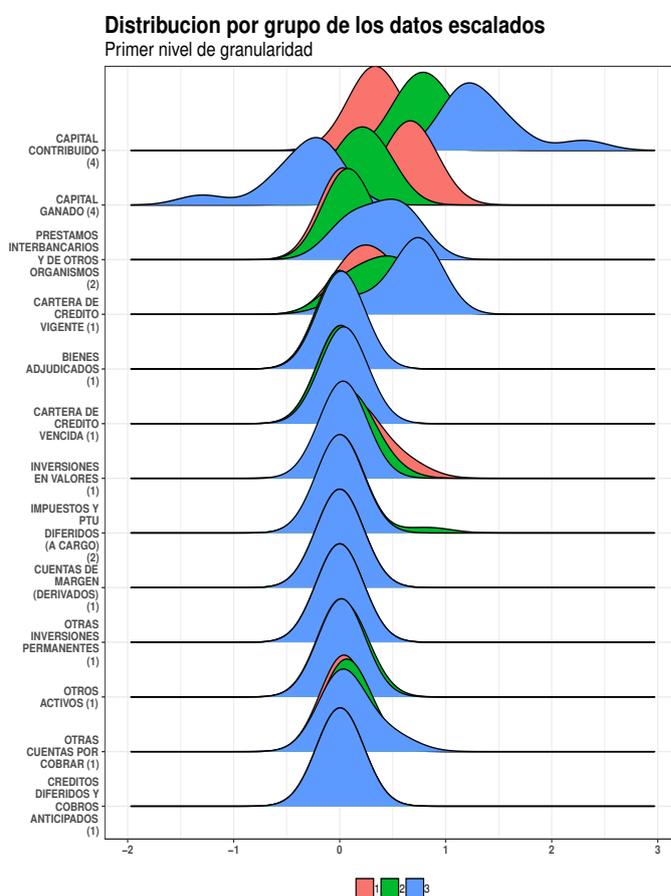


Figura 4.7: Distribución del porcentaje de la cuenta general de balance que representan las variables más importantes según el método de bosques aleatorios para el primer nivel del catálogo mínimo.

Las cuentas del segundo nivel de desagregación (gráfica 4.8) muestran que la distribución del capital social (dentro del capital contribuido del nivel anterior) es el que

resulta relevante para la obtención de la agrupación.

Respecto a las variables con una interpretación directa para los modelos de negocios, vemos que la captación tradicional es la tercera cuenta más importante, lo que resulta significativo, pues hay instituciones que tienen una clara preferencia entre los principales métodos de fondeo que existen. Específicamente, son los depósitos a plazo los que figuran en las cuentas más importantes.

La cuarta cuenta, operaciones de derivados con valores también es interesante, pues muestra al grupo 1 con una concentración mayor en dichas operaciones que los otros dos grupos. Por último, la séptima cuenta (cartera de créditos comerciales vigentes) complementa a la captación tradicional y las operaciones de derivados para poder decir que el grupo tres es, en general, el más equilibrado respecto a la conformación de su balance, mientras que los grupos uno y dos presentan distribuciones diferenciadas en cuentas significativas para hablar de modelos de negocios.

Más aún, las distribuciones en la captación tradicional, la cartera de crédito e incluso los préstamos interbancarios nos muestran que los bancos del grupo dos tienen un enfoque en el que dan mayor importancia a las actividades de captación de recursos y a la oferta de crédito. Por otro lado, viendo las operaciones con derivados y las inversiones en valores notamos que el grupo uno tiene un enfoque más fuerte a la inversión en mercados, con lo que podemos hacer una partición razonable de los bancos con un sentido de modelos de negocios.

4. ÁRBOLES DE DECISIÓN Y BOSQUES ALEATORIOS

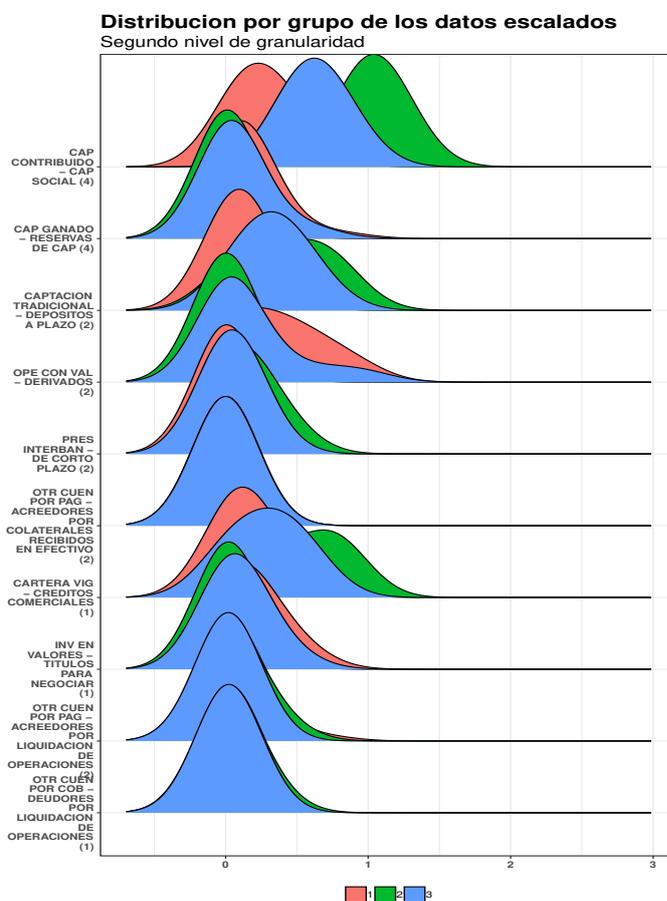


Figura 4.8: Distribución del porcentaje de la cuenta general de balance que representan las variables más importantes según el método de bosques aleatorios para el segundo nivel del catálogo mínimo.

Finalmente, vemos una vez más (gráfica 4.9) una cuenta del capital en el primer lugar de importancia: capital ganado - resultados del ejercicio anterior - resultado por aplicar. Sin embargo, las cuentas subsiguientes brindan de nuevo información clara para definir modelos de negocios.

La segunda y la tercer cuenta (depósitos a plazo del público en general y derivados con fines de negociación, respectivamente) confirman las hipótesis hechas con las cuentas del segundo nivel: mientras que el grupo uno tiene un balance mucho más homogéneo, los grupos dos y tres se diferencian claramente en la captación (en la que el grupo dos tiene una clara preferencia) y en las operaciones con derivados, en las que el grupo uno se diferencia claramente del resto de los bancos. Resulta interesante ver que en cuentas que corresponden a otros activos, como el equipo e inmobiliario, todos los bancos tienen porcentajes similares.

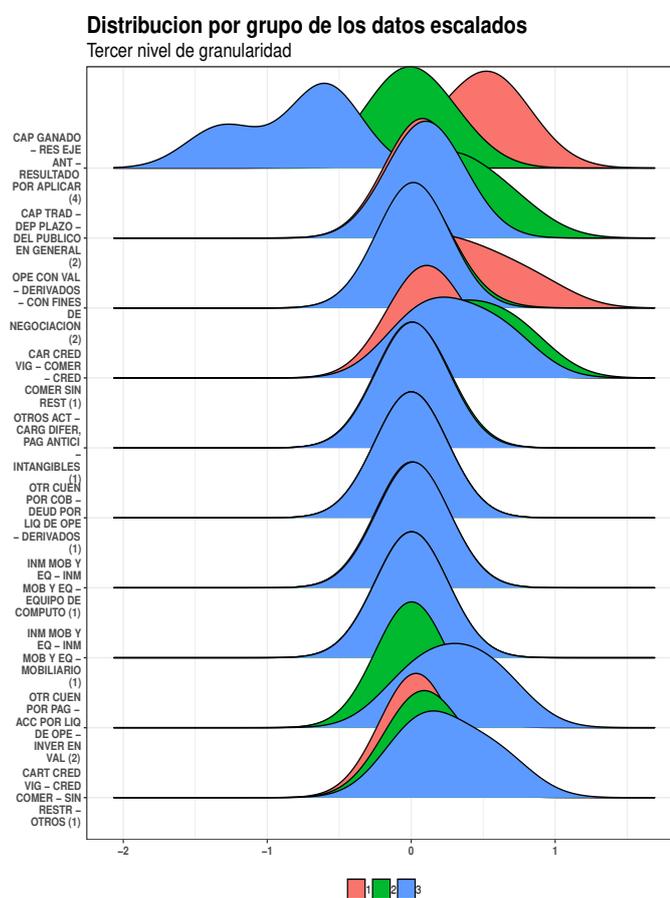


Figura 4.9: Distribución del porcentaje de la cuenta general de balance que representan las variables más importantes según el método de bosques aleatorios para el tercer nivel del catálogo mínimo.

Es claro que al movernos hacia niveles más desagregados de información, los resultados parecen más intuitivos respecto a los modelos de negocios, sin embargo, no debemos perder de vista la observación que hicimos en la sección donde se presentan las variables más importantes históricamente: al cambiar del segundo al tercer nivel, ya no son variables anidadas las que tenemos como más importantes, situación que sí pasa al cambiar del primer al segundo nivel.

Redes Financieras

La interacción de las instituciones bancarias en diferentes mercados y actividades da lugar a la formación de estructuras complejas, para las cuales las herramientas microprudenciales clásicas de análisis utilizadas por los bancos centrales mencionadas en la introducción quedan cortas en su capacidad de extracción de información valiosa. Dichas estructuras se presentan en una forma general: redes. Las redes pueden darse en diferentes mercados (tipos de transacciones financieras): divisas, derivados, compra-venta de títulos, reportos, préstamos directos (interbancario), etcétera. Todos los mercados tienen características diferentes entre sí, dependiendo, entre otras cosas, del número de instituciones participantes, la regulación existente, el nivel de desarrollo económico de la nación en la que operan. Por ejemplo, en [Fricke and Lux \(2015\)](#) encuentran una estructura Core-Periphery (que definimos más adelante) en el mercado interbancario electrónico e-MID.¹

De manera muy simple podemos pensar que las interacciones entre dos instituciones se da en sus estados financieros, pues algunos de los activos (“posesiones”) de una, son los pasivos (“deudas”) de la otra, es decir, hay un flujo de efectivo entre dos instituciones, dinero que una parte del acuerdo se compromete a pagar a la otra, siguiendo los términos pactados, y que constituye un **riesgo** para la institución prestamista de los fondos.

Las redes financieras han probado ser un método efectivo para visualizar y modelar las estructuras complejas derivadas de la interacción de los bancos, pues han permitido conocer con más detalle la verdadera forma en la que interactúan los participantes de un sistema financiero, no sólo las instituciones bancarias. Obtuvimos dos visualizaciones de la red interbancaria mexicana: una en la fecha inicial del estudio, y la otra en la fecha final (Figura 5.1), el color y tamaño de los nodos está relacionado con el número de conexiones de los mismos, mientras que el grosor y el color de las ligas dependen de su peso.

Es notorio que el sistema ha sufrido cambios con el paso del tiempo, en particular, es

¹En la literatura, el mercado interbancario es aquel en el que los bancos ofrecen y obtienen liquidez a través de préstamos sin garantía.

5. REDES FINANCIERAS

claro que se han agregado nuevas instituciones, lo que ha tenido como consecuencia una baja en la concentración de los flujos de dinero, como se puede observar en el cambio del color y el grosor de las ligas, así como en el color y el tamaño de los nodos. Sin embargo, una inspección visual no es suficiente para desenmascarar las características no evidentes de un sistema, por lo que la teoría de redes ha cobrado fuerza en el quehacer científico de las últimas dos décadas.

Para un análisis exhaustivo del uso que se le ha dado a la teoría de redes en un contexto económico-financiero, y su relación con el riesgo sistémico y la estabilidad financiera, el trabajo de [Battiston and Martinez-Jaramillo \(2018\)](#) contiene un recuento de las principales aplicaciones de la teoría de redes, así como de los resultados obtenidos hasta la fecha, además de las posibles rutas futuras de trabajo.

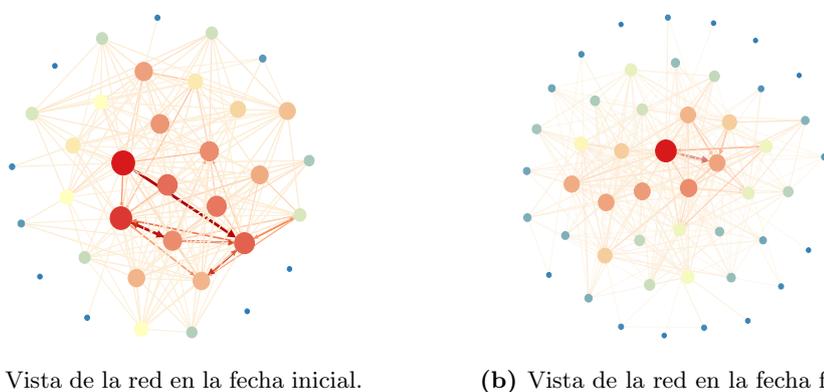


Figura 5.1: Visualizaciones de la red interbancaria Mexicana, cada círculo representa un banco mexicano, mientras que las ligas representan monto de dinero potencialmente perdido dado el estrés del sistema.

En un contexto de riesgo sistémico, para los reguladores es importante conocer a detalle a los principales jugadores del sistema bancario en términos de tamaño, de qué tanta labor de intermediación realizan en un mercado, y, en un contexto de redes, en términos de centralidad, pues son esos jugadores los que pueden desencadenar una serie de eventos con el potencial de poner en riesgo la estabilidad del sistema financiero.

Aún más, la teoría de redes ha permitido desenmascarar relaciones no triviales de instituciones financieras que a “simple vista” no representan un riesgo importante para el sistema, pero que dada su posición en la red sobresalen del resto de las instituciones gracias a la importancia que tienen para el buen funcionamiento del sistema.

El estudio de la red que forma el sistema bancario provee las herramientas necesarias para responder a preguntas importantes para el monitoreo del riesgo sistémico. A continuación, definimos los conceptos necesarios para entender el análisis de redes que aplicaremos al sistema bancario mexicano.

En el presente capítulo, usaremos de manera intercambiable los términos “sistema”,

“red” y “estructura”.

5.1. Datos

Los datos que utilizamos para realizar el análisis de red de nuestras instituciones provienen de una base de datos calculada y resguardada por Banco de México. Los datos pertenecen a un proyecto que se utiliza para la medición del **riesgo de contagio** en el sistema financiero Mexicano. Dichos datos representan las *obligaciones vigentes* entre cada par de instituciones financieras en el país (e incluso con instituciones financieras del exterior). Los resultados más actuales del monitoreo que se realiza con dicho proyecto se pueden consultar en [Banco de México \(2018\)](#).

Las obligaciones vigentes provienen de cuatro principales fuentes de información: operaciones con títulos de deuda, operaciones interbancarias, operaciones con derivados y operaciones con divisas, mercados para los cuales se cuenta con información a nivel transaccional para todos los bancos que operan (y han operado) en México.

A partir de dichas obligaciones bilaterales es que somos capaces de construir matrices de pesos (que definimos a continuación) que representan las redes de exposición interbancaria. Con los datos que se tienen, es posible construir matrices diarias desde 2005 hasta la fecha (y desde 2011 para el sistema financiero). Sin embargo, para el presente análisis nos limitaremos a matrices trimestrales para el periodo comprendido entre enero de 2007 y junio de 2017 (para tener congruencia con los datos utilizados en los métodos para obtención de conglomerados).

En las siguientes definiciones, los conceptos de ligas, arcos y conexiones se usaran para describir lo mismo, al igual que los conceptos de nodo, banco y arista representan lo mismo.

5.2. Características generales de las redes

Gran parte de la teoría de redes se desprende de los trabajos de los matemáticos *Paul Erdős* ([Wikipedia \(2019b\)](#)) y *Alfréd Rényi* ([Wikipedia \(2019a\)](#)), los cuales demostraron que es posible describir una red exhaustivamente con un conjunto pequeño de características inherentes a cada una. En [Barabási and Pósfai \(2016\)](#) se presenta una excelente introducción a la teoría moderna de redes, de donde obtenemos muchas de las siguientes definiciones, adicionalmente, otros materiales generales de apoyo para la presente tesis son [Newman \(2010\)](#) y [Bravo-Benítez \(2012\)](#).

5.3. Análisis estructural de Redes

5.3.1. Definiciones Generales

5.3.1.1. Gráfica no dirigida

Definimos una *gráfica no dirigida* como el conjunto

$$\mathcal{G} = [\mathcal{E}, \mathcal{V}], \quad (5.1)$$

donde \mathcal{V} es un conjunto finito distinto del vacío, y \mathcal{E} es el conjunto de pares no ordenados $\mathcal{E} = \{(i, j) | i, j \in \mathcal{V}\}$. A los elementos de \mathcal{V} se les llama vértices o nodos, y los elementos de \mathcal{E} son aristas, o conexiones que representan las relaciones existentes en una red. Es decir, una gráfica \mathcal{G} es un conjunto de nodos conectados con un conjunto de aristas.

5.3.1.2. Gráfica dirigida

Lo anterior se puede extender para definir una *gráfica dirigida* agregando un orden específico a los pares (i, j) , es decir, es el conjunto

$$\mathcal{D} = [\mathcal{N}, \mathcal{A}], \quad (5.2)$$

donde $\mathcal{N} \neq \emptyset$ finito, y $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$. A los elementos de \mathcal{N} se llaman nodos, y a los de \mathcal{A} arcos; la relación (i, j) , con $i, j \in \mathcal{N}$, implica que hay un arco con sentido de i a j , que a su vez no implica que haya una relación con sentido de j a i . Dicha relación puede interpretarse de muchas formas dependiendo del contexto; por ejemplo, en una red de computadoras la relación (i, j) implicaría que hay un flujo de información de la computadora i a la computadora j .

5.3.1.3. Red no dirigida

Las definiciones anteriores se extienden para definir nuestro objeto principal de estudio: las redes. En las redes tenemos un *peso* asignado a cada arista (arco) de nuestra gráfica, que denotamos con una función $w: \mathcal{E} \rightarrow \mathbb{R}$, es decir, el valor $w(i, j) = w_{ij}$ es el peso de la conexión entre los vértices i y j . Extendiendo el ejemplo anterior, el peso w_{ij} sería la cantidad de información que fluye entre los nodos i y j . Como notación, definimos una red no dirigida como la terna

$$\mathcal{R} = [\mathcal{V}, \mathcal{E}, w], \quad (5.3)$$

donde $[\mathcal{V}, \mathcal{E}]$ es una gráfica no dirigida como en 5.1

5.3.1.4. Red dirigida

Análogamente, definimos una *red dirigida* como la terna

$$\mathcal{R} = [\mathcal{N}, \mathcal{A}, w], \quad (5.4)$$

donde $[\mathcal{N}, \mathcal{A}]$ es una gráfica dirigida como en 5.2

5.3.1.5. Matriz de adyacencia

Hay dos formas principales de describir a una red matemáticamente, y que se utilizan para trabajar computacionalmente con ella. Una de ellas es a través de una “lista de arcos”, que es una lista ordenada de la forma (i, j) , por ejemplo, los pares $(1, 2)$, $(2, 3)$, $(3, 4)$, $(4, 1)$ nos definen una gráfica. Sin embargo, la forma más común para definir una red es con una *matriz de adyacencia* \mathbf{A} . La representación matricial resulta sumamente útil, pues es con operaciones matriciales que logramos obtener las medidas de red que definimos más adelante. Para una *gráfica no dirigida*, la matriz de adyacencia se define como

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{si } (i, j) \in \mathcal{E} \text{ o } (j, i) \in \mathcal{E}, \\ 0 & \text{en otro caso} \end{cases} \quad (5.5)$$

Dado que en una gráfica dirigida los arcos tienen un sentido específico entre dos nodos, la definición de la matriz de adyacencia puede ser dividida en dos partes, pues hay arcos que “entran” a un nodo y arcos que “salen” del él. Definimos las entradas de la *matriz de adyacencias exteriores* como:

$$a_{ij}^+ = \begin{cases} 1 & \text{si } (i, j) \in \mathcal{E}, \\ 0 & \text{e.o.c} \end{cases} \quad (5.6)$$

Análogamente, las entradas de la *matriz de adyacencias interiores* son:

$$a_{ij}^- = \begin{cases} 1 & \text{si } (j, i) \in \mathcal{E}, \\ 0 & \text{e.o.c} \end{cases} \quad (5.7)$$

En general, se cumple que $\mathbf{A}^+ = (\mathbf{A}^-)^T$

5.3.1.6. Vecinos de un nodo i

De la matriz de adyacencia, definimos a los vecinos de un nodo i . Para una gráfica no dirigida $\mathcal{G} = [\mathcal{V}, \mathcal{E}]$, el vértice $j \in \mathcal{V}$ es vecino del nodo $i \in \mathcal{V}$ si existe una arista que los conecte, por lo tanto, los vecinos del nodo i son el conjunto:

$$N(i) = \{j \in \mathcal{V} : a_{ij} = 1\}. \quad (5.8)$$

En una red dirigida, los *vecinos interiores* de i son aquellos nodos que tienen un arco dirigido hacia i , es decir, es el conjunto

$$N^-(i) = \{j \in \mathcal{V} : a_{ij}^- = 1\}. \quad (5.9)$$

Análogamente, los *vecinos exteriores* de i son aquellos nodos que reciben un arco dirigido de i . Es el conjunto:

$$N^+(i) = \{j \in \mathcal{V} : a_{ij}^+ = 1\}. \quad (5.10)$$

5.3.1.7. Matriz de Pesos

Con las matrices de adyacencia, definimos las *matrices de pesos* para una red dirigida. Dichas matrices son especialmente importantes, pues con ellas representaremos al sistema bancario mexicano. Definimos la *matriz de pesos exteriores* \mathbf{W}^+ a partir de una red dirigida $\mathcal{R} = [\mathcal{N}, \mathcal{A}, w]$ como:

$$\mathbf{W}_{ij}^+ = \begin{cases} w_{ij}^+ & \text{si } (i, j) \in \mathcal{A}, \\ 0 & \text{e.o.c} \end{cases} \quad (5.11)$$

Además, la *matriz de pesos interiores* \mathbf{W}^- se define como:

$$\mathbf{W}_{ij}^- = \begin{cases} w_{ij}^- & \text{si } (j, i) \in \mathcal{A}, \\ 0 & \text{e.o.c} \end{cases} \quad (5.12)$$

El sentido económico-financiero que le daremos en el presente trabajo a los pesos de los arcos en las redes dirigidas es que representan un flujo de dinero. La interpretación de dicho flujo depende de la matriz a la que pertenezca. Por ejemplo, en una red de exposiciones para calcular pérdidas por contagio, la entrada (i, j) de la matriz \mathbf{W}^+ representa la cantidad de dinero con la que la institución i está expuesta a la institución j , es decir, la cantidad de dinero que j le debe a i como resultado de las operaciones pasivas (de fondeo) de j con i ; para la matriz \mathbf{A}^- es la misma interpretación, pero en sentido contrario.

Por el contrario, en la red del Sistema de Pagos Electrónicos Interbancarios (SPEI), por ejemplo, la entrada (i, j) de la matriz \mathbf{W}^+ representa una transferencia de dinero del banco i al banco j .

Al igual que con las matrices de adyacencia, se cumple que $\mathbf{W}^+ = (\mathbf{W}^-)^T$. Además, podemos definir una matriz $\mathbf{W} = \mathbf{W}^+ + \mathbf{W}^-$, cuyas entradas representan el peso total de la conexión entre dos instituciones.

5.3.2. Medidas estructurales de red

Con las matrices definidas arriba podemos calcular las principales medidas estructurales que representan una red, dichas medidas nos ayudarán a “desenmarañar” las

características más importantes de nuestro sistema bancario, así como de los participantes del mismo.

De acuerdo con [Silva et al. \(2016\)](#), podemos dividir las medidas estructurales de red en tres categorías:

- **Medidas estrictamente locales:** Medidas que corresponden y describen únicamente al nodo al que pertenecen, no toman en cuenta información de los nodos vecinos ni de la totalidad de la red. Son útiles para identificar los nodos importantes en una red.
- **Medidas cuasi-locales:** También son medidas pertenecientes a cada nodo, sin embargo, sí toman en cuenta las características del vecindario de cada nodo para proveer información. Las medidas cuasi-locales nos ayudan a medir la importancia de un nodo *a partir* de la importancia de los nodos con los que está conectado, por lo que brindan información importante sobre los nodos de una red que se complementa muy bien con la información que dan las medidas estrictamente locales. Por ejemplo, hay que prestar especial atención a un nodo que, además de tener valores altos en sus medidas locales, también los tiene en sus medidas cuasi-locales.
- **Medidas globales:** Usan información de la red en su totalidad, por lo que ayudan a desenmascarar características de la red como sistema, más allá de sus componentes individuales (nodos).

Cada una de las medidas brindan información diferente sobre el sistema del que se obtienen, en el apéndice [B](#) se encuentran las definiciones formales, así como una breve descripción de la implicación que tienen en la red interbancaria con la que se trabaja en el presente trabajo.

5.3.3. Componentes de una red

Dado que uno de los objetivos de la teoría de redes es obtener información relevante de estructuras sumamente complejas, es conveniente caracterizar una red en distintos componentes de acuerdo con la conectividad de cada uno, con ciertas propiedades que permiten estudiarlos por separado, o al menos hacer más fácil la interpretación de las propiedades de cada uno.

Para entender el significado de los componentes de una red, es importante conocer primero los conceptos de *camino* y *ruta*. Un **camino** es una sucesión de vértices donde cada par consecutivo está conectado con una arista, es decir, es una descripción del orden en el que se tiene que recorrer un conjunto de nodos para llegar de un inicio a un final. Si las aristas tienen dirección, éstas se deben recorrer de acuerdo a ella, y se pueden recorrer en cualquier sentido en caso contrario. Por último, si dicha sucesión de nodos pasa una sola vez por cada uno (es decir, el camino no se intersecta consigo mismo), lo llamamos **ruta**.

El concepto de *componente* se desprende del hecho de que es posible que no haya un camino entre algún par de vértices. Más aún, es posible que una red R esté dividida en dos partes, digamos A y B , tal que es posible que ninguno de los vértices en A se conecte con ninguno de los vértices en B , pero sí es posible encontrar un camino entre todos los pares de vértices en cada conjunto.

En el caso anterior, se dice que R es una red **desconectada**, mientras que las subredes definidas por A y B , con sus respectivos conjuntos de ligas, son redes **conectadas**. De lo anterior, toda red conectada tiene un solo componente.

Para las redes dirigidas, es posible obtener subconjuntos más finos de nodos condicionando con la dirección de las ligas. Si omitimos la dirección, los componentes obtenidos son llamados **componentes débilmente conexos** (*weakly connected components*). Por el contrario, si la dirección de los ejes cobra relevancia, y es posible encontrar un camino *en ambos sentidos* para todo par de vértices, al conjunto que cumpla esas dos condiciones se le conoce como un **componente fuertemente conexo** (strongly connected component).

El concepto de componente se puede generalizar al de k -componente. Un k -componente es un subconjunto de vértices tal que cada par está conectado por al menos k caminos independientes de vértices.² La idea de un k -componente va de la mano con la noción de **robustez** de una red, es decir, de la resistencia de la red a la falla de alguno de sus vértices. Por ejemplo, en una red de distribución de bienes es deseable tener caminos independientes de vértices entre cada par de destinos, pues eso significa tener más de un camino para llegar de un punto a a un punto b .

Por otro lado, en [Dorogovtsev et al. \(2001\)](#) se propone una forma de particionar una red dirigida que retoma y extiende muchos de los conceptos ya definidos. Además, se dan los pasos para encontrar dichas particiones. Esencialmente podemos particionar una red de acuerdo con Dorogovtsev en dos partes, sin importar su tamaño:

1. El *componente desconexo*, que está formado por todos los nodos desconectados, o por pequeños subconjuntos de nodos que sólo se conectan entre ellos y con nadie más fuera de su conjunto.
2. El *componente débilmente conexo*, que, como ya mencionamos, es aquel en el que todo par de vértices está conectado por al menos un camino, sin considerar la dirección de los arcos que lo forman.

Al tomar en cuenta la dirección de los arcos, se puede particionar a su vez el componente débilmente conexo:

- El *componente fuertemente conexo*, en el que para cada par de vértices existen dos caminos que los conectan, uno en sentido contrario del otro.
- El *componente exterior*, al cual sólo se puede acceder desde el componente fuertemente conexo.

²Dos caminos son independientes de vértices cuando no comparten nodos más que el inicial y el final.

- El *componente interior*, el cual puede acceder únicamente al componente fuertemente conexo.
- Los *tallos*, los cuales están completamente desconectados del componente fuertemente conexo, y que además están desconectados entre sí.

5.3.4. Centralidad

Las redes financieras heredaron el concepto de *centralidad* de las redes sociales; nos ayudamos de algunas de las medidas de red ya definidas (principalmente de las medidas cuasi-locales) y definiremos otras para medir la centralidad de un nodo. Entendemos a la centralidad como un indicador de la importancia de un nodo en la red, dicha importancia tiene un significado particular en la red bancaria que nos ocupa, pues gracias a las llamadas *medidas de centralidad* podemos detectar a las instituciones que más importan para la salud y la estabilidad del sistema completo, lo que las convierte en herramientas de gran utilidad para los reguladores y supervisores encargados de mantener la resiliencia del sistema.

De [Martinez-Jaramillo et al. \(2014\)](#), reproducimos las características que [Henggeler-Müller \(2006\)](#) considera, vuelven *central* a una institución en una red financiera:

- Posee muchas ligas hacia otros miembros de la red (centralidad por grado).
- El tamaño de sus activos y sus pasivos en la red es muy grande (fuerza).
- Su quiebra puede transmitir contagio en pocos pasos (cercanía, DebtRank).
- Sus contrapartes también son *centrales* en la red (PageRank, Eigenvectores, DebtRank).
- Hay muchos caminos que pasan a través de ella (interinidad).

Ordenando las medidas de centralidad que definimos a continuación, podemos obtener la importancia de cada institución dentro de la red de acuerdo con cada indicador, sin embargo, como mencionan en [Martinez-Jaramillo et al. \(2014\)](#), ninguna es lo suficientemente robusta para tomarla como la medida “a considerar”. Proponen una combinación de todas las medidas en una combinación lineal, sin embargo, no haremos el mismo análisis en el presente trabajo. Del mismo trabajo tomamos las definiciones de las medidas de centralidad.

5.3.4.1. Centralidad por fuerza

Definimos la centralidad por fuerza de cada nodo simplemente como su fuerza misma (ver definición en apéndice [B](#)), es decir:

$$C_s(i) = s_i. \quad (5.13)$$

Es importante diferenciar la fuerza exterior de la fuerza interior para darle significado a la centralidad por fuerza, la centralidad por fuerza exterior nos da la importancia del nodo en la red como prestamista, y la interior como prestatario. Al mismo tiempo, el efecto de un nodo central por fuerza depende de su papel en la red.

5.3.4.2. Centralidad por grado

La centralidad por grado será mayor en aquellos nodos que tengan muchas más conexiones que el resto de los participantes de la red, pues los problemas originados en dicho nodo se propagarían a un mayor número de instituciones. Una vez más, la centralidad por grado es simplemente el grado de cada nodo:

$$C_d(i) = d_i. \quad (5.14)$$

El grado cumple un papel sumamente importante en la teoría de redes, pues a través de las conexiones de los nodos se pueden descubrir características importantes que permiten clasificar un sistema dentro de alguno de los modelos de redes conocidos (los cuales exponemos más adelante), o incluso detectar propiedades poco convencionales para una red.

Una de las herramientas más utilizadas es la **distribución del grado**, pues frecuentemente basta para caracterizar una red y diferenciarla de distintos modelos cuyas distribuciones de grado presentan comportamientos claramente distintos. En palabras de Albert-László Barabási en [Barabási and Pósfai \(2016\)](#) (página 115):

“Indeed, we will come to see that no matter what network property we are interested in, from communities to spreading processes, it must be inspected in the light of the network’s degree distribution”.

[“De hecho, veremos que sin importar la propiedad en la que estemos interesados, desde comunidades hasta procesos de difusión, debe de ser inspeccionada a través de la distribución del grado de la red.”]

5.3.4.3. Centralidad por interinidad

La centralidad por interinidad nos dice qué tanto está intermediando una institución (nodo) en las conexiones entre cualesquiera dos instituciones diferentes en la red. Para auxiliarnos en la definición de la centralidad por interinidad, sea $\sigma_{ij} = \sigma_{ji}$ el número total de caminos más cortos entre los nodos i y j , y sea $\sigma_{ij}(k)$ el número total de caminos más cortos entre i y j que pasan por el nodo k , entonces, la centralidad por interinidad del nodo k es:

$$C_I(k) = \sum_{i \neq k \neq j \in \mathcal{N}} \frac{\sigma_{ij}(k)}{\sigma_{ij}}. \quad (5.15)$$

Es posible calcular centralidad por interinidad exterior e interior, considerando dirección en los arcos.

5.3.4.4. Centralidad por cercanía

Un nodo es central por cercanía en la red si al sumar la inversa de las distancias de dicho nodo al resto de nodos es un valor grande, es decir, si está conectado por caminos cortos con el resto de los nodos. Se define como:

$$C_c(i) = \sum_{j \in \mathcal{N} \setminus \{i\}} \frac{1}{d_{ij}}. \quad (5.16)$$

Abusando de la notación, la distancia aquí no solo es la longitud de cualquier camino entre dos nodos, es la longitud del camino **más corto** entre los dos.

De [Newman \(2010\)](#) obtenemos la definición de la centralidad por *Eigenvectores*.

5.3.4.5. Centralidad por Eigenvectores

La centralidad por eigenvector extiende la centralidad por grado al tomar en cuenta a los vecinos de un nodo, asignando una ponderación distinta por cada vecino, dicha ponderación depende a su vez de la centralidad de cada vecino. Así, la idea detrás de la centralidad por eigenvectores es que la importancia de un nodo aumenta si sus vecinos son a su vez importantes.

En redes dirigidas existe un inconveniente que es importante mencionar. Dado que tenemos dos matrices de adyacencia (la interior y la exterior), tenemos también dos conjuntos de eigenvectores. En [Newman \(2010\)](#) se menciona que se suele usar la matriz interior para medir la centralidad de un nodo, y la razón se hace evidente con el siguiente ejemplo: una página de internet es útil si muchas páginas apuntan a ella como referencia, por otro lado, que ésta apunte a muchas otras no brinda mucha información sobre su importancia.

La centralidad por eigenvectores es tomada como caso general para construir otras medidas de centralidad, como PageRank y Katz, que no veremos aquí. Las variaciones que se han construido a partir de la centralidad por eigenvectores pretenden disminuir ciertos efectos adversos presentes en el cálculo de la misma.

5.3.5. Centralidad por DebtRank

Dentro de la teoría de redes aplicada al estudio de sistemas financieros, sin duda el *DebtRank* marcó un antecedente, pues es la primera medida de centralidad pensada y desarrollada específicamente para redes financieras. Más aún, al mismo tiempo de cumplir un papel de medida de centralidad, es también una medida del **riesgo sistémico** implícito en un sistema financiero. Dada su importancia, daremos mucho más detalle en la definición del DebtRank en este capítulo.

El DebtRank fue introducido en [Battiston et al. \(2012\)](#), y es presentado como una novedosa medida para cuantificar el riesgo sistémico, con la característica de que toma en cuenta efectos de “retroalimentación” (*feedback effects*) entre las instituciones que participan en el sistema financiero, y son esos efectos de “retroalimentación” (la

traducción es forzada, y va en un sentido negativo, pues es una retroalimentación a las pérdidas de las instituciones) los que hacen interesante a la medida.

En los primeros modelos de contagio financiero a partir de teoría de redes, (por ejemplo, los modelos presentados en [Canedo and Martínez-Jaramillo \(2009\)](#) y en [Gai and Kapadia \(2010\)](#)), una vez que una institución quiebra³ como consecuencia del proceso de contagio (*cascade default*⁴), se detiene su potencial de seguir contagiando a otras instituciones. Lo novedoso del DebtRank, como veremos en su definición, es la capacidad que se le da a las instituciones de seguir amplificando los choques (pérdidas) sufridos por otras instituciones conectadas a éstas en la red, haciéndolas a su vez más propensas a quebrar.

En [Battiston et al. \(2012\)](#) demuestran cómo el DebtRank captura de mejor forma los efectos potenciales de la quiebra de una institución (y así, asignando un grado de centralidad de acuerdo con su potencial de riesgo), y cómo otras medidas de centralidad aplicadas a redes financieras subestiman los efectos reales de una institución sistémica.

Para describir el algoritmo para calcular el DebtRank, es necesario considerar las siguientes definiciones, que reproducimos de [Battiston et al. \(2012\)](#):

- A es la matriz de pesos, donde A_{ij} representa la cantidad de dinero que i usa para fondear a j . Así, $A_i = \sum_l A_{il}$ es la cantidad total invertida por i en el sistema.
- E_i es el capital básico o neto de i .⁵
- γ es un umbral positivo tal que si $E_i \leq \gamma$, i entra en default.
- $W_{ij} = \min\{1, A_{ji}/E_j\}$ es el impacto de i en j . Si el impacto es 1, quiere decir que en caso de que i quiebre, también lo hará j de manera **directa**.
- $v_j = \frac{A_j}{\sum_l A_l}$ es el valor económico de j relativo al sistema.
- $I_i = \sum_j W_{ij}v_j$ es el impacto de i en sus vecinos directos.

Con lo anterior, queda claro que si i quiebra, la institución j tendrá una pérdida de A_{ji} , y también quebrará si $A_{ji} > E_j$. Hasta ahora, todas las definiciones han sido tomando en consideración únicamente los vecinos directos de una institución i . Pero como ya mencioné, esta medida también toma en cuenta el impacto *indirecto* de i en instituciones con las que no está conectada directamente.

³Por quiebra o default entendemos que una institución se vuelve incapaz de honrar sus obligaciones, provocando así pérdidas a sus acreedores.

⁴Un proceso de contagio *cascade default* es aquel en el que la quiebra de una institución provoca la quiebra de otra institución con una alta dependencia a ella, lo que a su vez provoca la caída de otra(s) institución(es) y así sucesivamente.

⁵El capital regulatorio de una institución bancaria es la cantidad de dinero que deben de acumular para hacer frente a sus obligaciones en situaciones de estrés.

Para tomar en cuenta los efectos indirectos de i en el sistema, se define la ecuación recursiva

$$I_i = \sum_j W_{ij}v_j + \beta \sum_j W_{ij}I_j, \quad (5.17)$$

vemos que el primer término es el impacto directo. El segundo término da cuenta del impacto subsecuente de i en los vecinos de sus vecinos. $\beta < 1$ es un término de “suavizado” del impacto de acuerdo con la longitud de los caminos que conectan a i con el resto de la red. El papel de β se hace más claro cuando escribimos la ecuación 5.17 en forma matricial:

$$I = Wv + \beta Wv = (Id - \beta W)^{-1}Wv = \sum_{k=0}^{\infty} (\beta^k W^k)Wv. \quad (5.18)$$

Un resultado importante en teoría de redes es que la entrada ij de la k -ésima potencia de la matriz de adyacencia es el número de caminos de distancia k que existen entre i y j . En una matriz de pesos, es la suma de los productos de los pesos a través de los caminos de longitud k . Así, vemos que el factor β^k provoca que el impacto de una institución se diluya en caminos más largos. Es importante notar que bajo la ecuación anterior es posible que una institución tenga un impacto indirecto en sí misma, cuestión que también tiene sentido en redes reales, pues i puede provocar indirectamente la quiebra de una institución k que a su vez tiene obligaciones con i .

El valor I obtenido con la ecuación anterior podría considerarse en sí una medida de centralidad, sin embargo presenta una inconsistencia: permite que una institución tenga impacto sobre sí misma más de una vez, lo que puede crear problemas. Para que una institución tenga impacto sobre sí misma, es necesario que en la red existan ciclos, sin embargo, a través de esos ciclos, en caso de que una institución i tenga impacto propio, se estaría contando más de una vez el impacto directo de i sobre el sistema al llegar de nuevo a sí misma.

Por lo anterior, los autores proponen un método para calcular el DebtRank definitivo. Definen dos variables de estado: $h_i \in [0, 1]$ y $s_i \in \{U, D, I\}$, donde D significa “estresada”, U “no estresada” e I “inactiva”. Adicionalmente, S_f es el conjunto de nodos estresados en el “tiempo” 1, es decir, los nodos con un impacto de algún otro nodo i mayor a uno. Finalmente, se suponen las siguientes condiciones iniciales: $h_i(1) = \psi \forall i \in S_f$, $h_i(1) = 0 \forall i \notin S_f$, $s_i = D \forall i \in S_f$, $s_i = U \forall i \notin S_f$. ψ es el nivel de estrés inicial, $\psi = 1$ significa default. Finalmente, se define el siguiente algoritmo de actualización para cada ronda t consecutiva:

$$h_i(t) = \min \left\{ 1, h_i(t-1) + \sum_j W_{ji}h_j(t-1) \right\}, \quad (5.19)$$

donde j corre sobre las instituciones estresadas en $t-1$. Además, la variable s se

actualiza de la siguiente forma:

$$s_i(t) = \begin{cases} D & \text{si } h_i(t) > 0; s_i(t-1) \neq I, \\ I & \text{si } s_i(t-1) = D, \\ s_i(t-1) & \text{e.o.c.} \end{cases} \quad (5.20)$$

El procedimiento anterior se repetirá hasta que todos los nodos se encuentren en el estado U o I . Vemos que en el paso de actualización de s_i un nodo que se encuentra en estrés pasará a estar inactivo en el tiempo siguiente, lo que evita el inconveniente de contar más de una vez el impacto de un banco. Finalmente, el **DebtRank** se define como:

$$R = \sum_j h_j(T)v_j - \sum_j h_j(1)v_j, \quad (5.21)$$

donde T es el tiempo en el que la dinámica se detiene una vez que se haya cumplido la condición.

Así, podemos interpretar al DebtRank como el porcentaje de activos del sistema bajo estrés dada la quiebra hipotética de un banco. Es importante recalcar que para obtener la centralidad por DebtRank de una sola institución i es necesario hacer $S_f = i$.

5.4. Principales modelos generativos de redes

Dentro del análisis de redes se han logrado identificar diferentes modelos estructurales que se pueden encontrar en diversos tipos de sistemas: financieros, biológicos, físicos o sociales. Dichas estructuras presentan características diferentes de acuerdo con la forma en la que sus componentes interactúan entre sí, y de acuerdo con la forma en la que ciertos procesos se desarrollan en ellas, por ejemplo, en una red social altamente conectada, un proceso de contagio tendrá un alcance (y por lo tanto un nivel de severidad) mucho mayor que en un sistema en el que sus participantes están mayormente aislados el uno del otro.

Las medidas estructurales y el concepto de centralidad definidos previamente son buenas herramientas para la identificación de ciertos comportamientos en una red, en particular, las medidas locales de red nos ayudan a descifrar comportamientos en una escala micro (nivel de nodo), y las medidas globales nos ayudan a identificar comportamientos a escala macro (sistema completo).

Sin embargo, en todo sistema visto como una red, existen comportamientos y formas de organización no triviales que no son posibles de capturar con las medidas estructurales de red. Es por esto que se definen las **mesoescalas** como las características de una red que no son perceptibles a simple vista.

En general, la estructura de una red (y por lo tanto, sus características globales) están determinadas por un proceso generador desconocido, por lo que uno de los objetivos principales de la teoría moderna de redes es desarrollar modelos que **reproduzcan** las características de una red de la mejor forma posible.

A continuación presentaremos brevemente algunos de los modelos de redes más utilizados y sus características principales.

5.4.1. Redes aleatorias

Existen varios ejemplos de estructuras en redes, siendo el más común el modelo de red aleatoria. Es ampliamente utilizado como modelo nulo para saber si una red bajo estudio presenta las propiedades de un sistema aleatorio o no, sin embargo, existen varios trabajos empíricos que han demostrado que las redes reales distan de tener un comportamiento aleatorio, más aún, las propiedades de la red más grande creada por el hombre (internet) motivaron la formulación de uno de los modelos más recientes de formación de redes: el modelo *Barabási-Albert* (véase [Barabási and Albert \(1999\)](#)), que es un modelo particular derivado de las redes aleatorias.

Dependiendo de la forma en la que se genera una red, se pueden distinguir dos tipos generales de redes aleatorias:

- Modelo $G(N, L)$: Se asumen N nodos y un número **fijo** de conexiones L . Para simular una red aleatoria con el modelo $G(N, L)$, en cada paso se selecciona un par de nodos de manera aleatoria y se conecta con algunas de las ligas disponibles. Dicho procedimiento se repite hasta que las L ligas sean usadas. Fue desarrollado y utilizado por Erdős y Rényi en varios trabajos: [Erdős and Rényi \(1959\)](#), [Erdős and Rényi \(1960\)](#).
- Modelo $G(N, p)$: En este modelo, cada par de nodos se conecta con una probabilidad p . Es ampliamente utilizado dada la facilidad con la que pueden derivarse de manera analítica muchas de sus características gracias a la inclusión de la idea de probabilidad. Fue introducido inicialmente por [Gilbert \(1959\)](#).

Tomando como base al modelo $G(N, p)$, las redes aleatorias tienen las siguientes características generales:

- El grado tiene una distribución binomial, que bajo ciertas condiciones puede ser aproximada con una distribución Poisson, es decir, si definimos $p(k)$ como la probabilidad de que al seleccionar un nodo de la red aleatoriamente, éste tenga grado k , entonces la distribución del grado es:

$$p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}, \quad (5.22)$$

y en su forma Poisson:

$$p(k) = \exp(-\hat{k}) \frac{\hat{k}^k}{k!}. \quad (5.23)$$

donde \hat{k} es el grado promedio de la red.

- Dado el grado promedio, pueden surgir diversos regímenes de conectividad: si el grado promedio $\hat{k} > 1$, la red tendrá un componente gigante conectado, es decir, la mayoría de los nodos estarán conectados entre sí a partir de un camino, con solo algunos nodos desconectados del resto de la red. Al punto $\hat{k} = 1$ se le conoce como *punto crítico* de conectividad.
- La distancia promedio de la red está dada por $\hat{d} \propto \frac{\ln(N)}{\ln(\hat{k})}$.
- La medida de agrupamiento promedio se puede calcular como $\hat{C} = \frac{\hat{k}}{N}$.

Para las redes reales se ha observado que, en general, el grado **no** sigue una distribución Binomial-Poisson, los regímenes de conectividad no están necesariamente asociados con el grado promedio y en su mayoría tienen un grado promedio mayor a 1, y el coeficiente de agrupamiento es mucho mayor que el que presentan las redes aleatorias. Por su parte, la forma de calcular la distancia promedio de una red aleatoria da una aproximación aceptable a la distancia promedio de redes observadas.

5.4.2. Redes libres de escala

Es bien sabido que los procesos que suceden de manera natural a nuestro alrededor difícilmente pueden ser descritos o simulados de manera exacta, a lo más, los modelos matemáticos que se han desarrollado sirven como (a veces muy buenas) aproximaciones. La teoría de redes no es la excepción.

Existen incontables materiales, varios de ellos citados en la presente tesis, en los que se ha demostrado con información empírica que las redes reales tienen comportamientos distintos a los que presentan las redes aleatorias. Uno de los ejemplos más importantes es [Albert et al. \(1999\)](#), donde los autores demuestran que la distribución del grado para el Internet está lejos de ser Binomial-Poisson.

Al observar la distribución del grado de la red formada por el Internet⁶, los autores notaron de inmediato una diferencia fundamental: mientras que las redes aleatorias, desde su construcción tienen prácticamente prohibida la existencia de nodos de gran conectividad, el Internet presentaba nodos con un grado de entrada y de salida sumamente altos.

Dichos nodos, de aquí en adelante llamados *hubs* o *centros de conectividad*, están presentes en la gran mayoría de las redes reales, y son el bloque principal sobre el que está construida la teoría de las redes libres de escala. A partir de sus experimentos, [Albert et al. \(1999\)](#) demostraron que la distribución del grado de Internet puede ser aproximada de mejor forma por una función

$$p_k \sim k^{-\gamma}. \quad (5.24)$$

⁶Podemos considerar dos redes de internet: la red física, formada por computadoras y servidores, y la red intangible, formada por todas las páginas de internet; la red considerada en el trabajo citado corresponde a la red de páginas.

La función anterior se conoce como una **distribución Power-Law, Ley Potencia** o **Distribución Pareto** (Pareto et al. (1964)), de la cual podemos obtener una característica sumamente útil para hacer una inspección inicial sobre qué tanto se parece una red real a una red aleatoria. Si aplicamos logaritmo a la función anterior, obtenemos

$$\ln(p_k) \sim \gamma \ln(k). \quad (5.25)$$

Lo anterior nos da una relación lineal entre el grado y su probabilidad, con γ representando la pendiente de la recta. Por lo anterior, una de las primeras inspecciones que se hacen para saber si una red es aleatoria o no, es graficar su distribución de grado en un plano log-log. Como vemos más adelante, las redes que utilizamos para obtener las comunidades por Modelo de Bloques Estocásticos presentan una distribución de grado más parecida a la de las redes aleatorias que a la de una red libre de escala. En general, *una red libre de escala es aquella cuya distribución de grado es una Ley Potencia*.

Existen dos formas de definir la distribución Ley Potencia para redes: la forma discreta y la forma continua. A continuación, derivamos las dos formas dadas las características que deben de cumplir como funciones de distribución.

5.4.2.1. Forma discreta

Sabemos que el grado de un nodo puede ser un número entero $k \in \{0, 1, 2, \dots\}$. De la ecuación 5.24 podemos definir que la probabilidad de que un nodo seleccionado al azar tenga grado k está dada por:

$$p_k = Ck^{-\gamma}, \quad (5.26)$$

donde, al ser p_k una distribución de probabilidad debe cumplir

$$\sum_{k=0}^{\infty} p_k = 1. \quad (5.27)$$

Así, obtenemos:

$$\sum_{k=0}^{\infty} p_k = \sum_{k=1}^{\infty} Ck^{-\gamma}, \quad (5.28)$$

$$1 = C \sum_{k=0}^{\infty} k^{-\gamma}, \quad (5.29)$$

entonces

$$C = \frac{1}{\sum_{k=0}^{\infty} k^{-\gamma}}, \quad (5.30)$$

donde $\sum_{k=0}^{\infty} k^{-\gamma} = \zeta(\gamma)$ es la función zeta de Riemann. Por lo tanto, la forma discreta de la distribución del grado para una red libre de escala es:

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}. \quad (5.31)$$

Dado que la función anterior tiene una discontinuidad en $k = 0$, si es necesario, se puede declarar p_0 como el porcentaje de nodos desconectados de nuestra red.

5.4.2.2. Forma continua

Si se supone que el grado de los nodos puede ser cualquier número real positivo, se puede derivar la distribución continua siguiendo los mismos pasos que con la distribución discreta a partir de la igualdad

$$p(k) = Ck^{-\gamma}, \quad (5.32)$$

tomando como condición

$$\int_{k_{min}}^{\infty} p(k)dk = 1. \quad (5.33)$$

Entonces

$$C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} dk}. \quad (5.34)$$

En [Barabási and Pósfai \(2016\)](#) hay una discusión de qué pasa con la distribución del grado de las redes variando el valor de γ . Sin embargo, en todas las redes que se analizan en el texto, se cumple que $2 < \gamma < 3$, así, $-\gamma + 1 < 0$, por lo que al concentrarnos en la integral de la ecuación anterior tenemos que

$$\int_{k_{min}}^{\infty} k^{-\gamma} dk = \frac{k^{-\gamma+1}}{-\gamma+1} \Big|_{k_{min}}^{\infty} = 0 - \frac{k_{min}^{-\gamma+1}}{-\gamma+1} = \frac{k_{min}^{-\gamma+1}}{\gamma-1}. \quad (5.35)$$

Entonces

$$C = \frac{1}{\frac{k_{min}^{-\gamma+1}}{\gamma-1}} = \frac{\gamma-1}{k_{min}^{-\gamma+1}} = (\gamma-1)k_{min}^{\gamma-1}. \quad (5.36)$$

Por lo tanto, la forma continua de la distribución de grado para una red libre de escala es:

$$p(k) = (\gamma-1)k_{min}^{\gamma-1}k^{-\gamma}. \quad (5.37)$$

De la ecuación anterior vemos que para $k = 0$ tenemos una indeterminación en $k^{-\gamma} = \frac{1}{k^\gamma}$, razón por la cual se utiliza k_{min} en la determinación de la constante de normalización C .

La forma continua tiene una interpretación distinta a la que se tiene con p_k . Pues $p(k)$ no es la probabilidad de que un nodo seleccionado aleatoriamente tenga grado k , más bien, si calculamos

$$\int_{k_1}^{k_2} p(k)dk, \quad (5.38)$$

obtendremos la probabilidad de que un nodo seleccionado aleatoriamente tenga grado k tal que $k_1 \leq k \leq k_2$.

Es importante mencionar que para **redes dirigidas**, la distribución del grado se puede ver por separado para el **grado interior** y el **grado exterior**, cumpliéndose que cada una sigue una Ley Potencia por separado siempre y cuando ambas medidas estén altamente correlacionadas. En [Newman \(2010\)](#) mencionan como ejemplo el internet, que presenta distribuciones Power-Law en el grado de entrada y en el de salida, mientras que en muchas de las redes de citas académicas sólo el grado interior presenta dicho comportamiento.

La diferencia primordial entre las redes aleatorias y las redes libres de escala es la existencia de los ya mencionados *hubs*, que son nodos con un número muy grande de conexiones, ya sea de salida o de entrada. En la Figura 5.2 (reproducida por el autor, de [Barabási and Pósfai \(2016\)](#), página 120, tomando los mismos parámetros para simular los datos necesarios para las gráficas) podemos observar claramente lo anterior. Mientras que en la gráfica del lado izquierdo tenemos las dos distribuciones en una escala lineal, en la gráfica de la derecha, con escala log-log la diferencia se hace evidente. La distribución Ley potencia permite la existencia de nodos con un grado mucho mayor, mientras que la distribución Poisson presenta una caída muy pronunciada después de cierto grado.

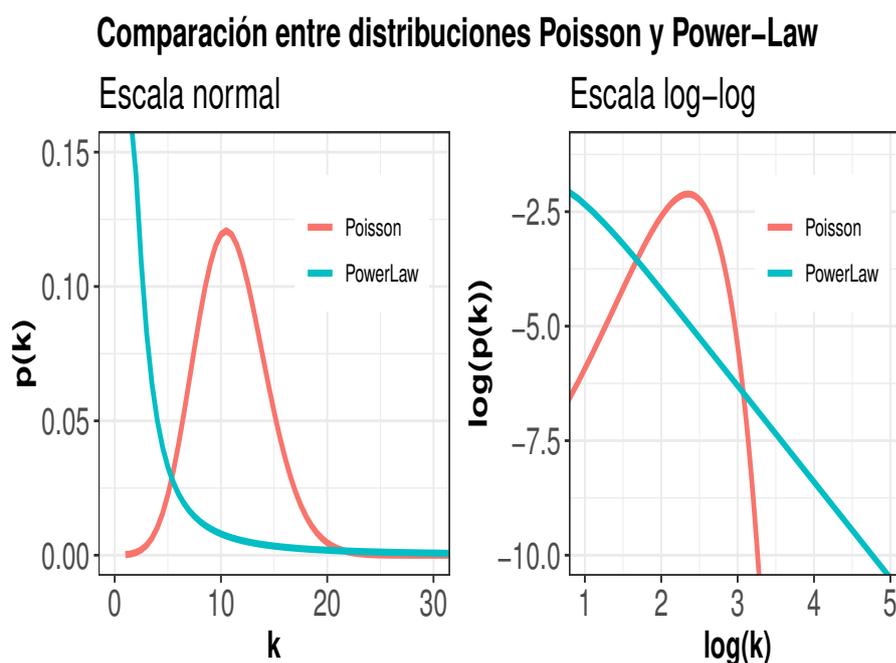


Figura 5.2: Comparación entre las colas de las distribuciones de grado obtenidas de redes aleatorias (Poisson) y redes libres de escala (Ley potencia)

El comportamiento anterior se ha encontrado en una gran diversidad de redes reales, desde sistemas biológicos, sociales, hasta redes tecnológicas, por lo que han tenido a bien dar la característica de **universalidad** a la propiedad de libertad de escala.

Otra de las características particulares de las redes libres de escala, y con una interpretación directa y sumamente útil, es que gracias a la existencia de nodos con un gran número de conexiones éstas resultan ser altamente vulnerables a los ataques dirigidos, mientras que muestran una gran resistencia a fallas aleatorias en un sistema.

En [Albert et al. \(2000\)](#) realizan un experimento para demostrar lo anterior. Tomando una red libre de escala, seleccionan nodos de manera aleatoria, y recalculan la longitud promedio de la red al remover del sistema el nodo seleccionado. Encontraron que la longitud promedio de la red aumentaba considerablemente al remover nodos altamente conectados, en comparación de los nodos con conectividad promedio, que no provocan grandes cambios en la conectividad de la red.

El resultado anterior resulta natural si consideramos que en términos absolutos, es mucho más probable seleccionar un nodo con pocas conexiones a un nodo altamente conectado. Sin embargo, si un sistema con un número limitado de *hubs* recibe un ataque dirigido a alguno de ellos, el número de nodos que pueden quedar desconectados entre sí es considerablemente alto.

5.4.3. Grado de los nodos en redes aleatorias versus redes libres de escala

Para descubrir por qué las redes aleatorias se diferencian de las redes libres de escala por la existencia de nodos altamente conectados, tenemos que dirigir la atención a la distribución del grado para los dos casos.

Por un lado, como ya mencionamos, la distribución del grado de las redes aleatorias puede aproximarse de buena forma con una distribución binomial, es decir:

$$p(k) = e^{(-\hat{k})} \frac{\hat{k}^k}{k!}. \quad (5.39)$$

Por la fórmula de Stirling, sabemos que

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \quad (5.40)$$

Con lo anterior, podemos reescribir [5.39](#) como:

$$p(k) = e^{(-\hat{k})} \left(\frac{\hat{k}^k}{\sqrt{2\pi n} \left(\frac{k}{e}\right)^k} \right). \quad (5.41)$$

Reordenando

$$p(k) = \frac{e^{-\hat{k}}}{\sqrt{2\pi k}} \left(\frac{\hat{k}}{k}\right)^k = \frac{e^{-\hat{k}}}{\sqrt{2\pi k}} \left(\frac{e\hat{k}}{k}\right)^k. \quad (5.42)$$

Dado que el grado promedio de la red \hat{k} es un número fijo, para valores $k > e\hat{k}$ el término $\left(\frac{e\hat{k}}{k}\right) < 1$, así, con $k \rightarrow \infty$ ambos términos tienden a cero, pues $\frac{1}{\sqrt{k}}$ también

decrece rápidamente conforme k crece. Es decir, con lo anterior es claro que la probabilidad de observar un nodo con un grado muy grande se aproxima rápidamente a cero a medida que el grado k crece en una red aleatoria.

Por su parte, para las redes libres de escala podemos hacer un análisis similar partiendo también de la distribución del grado. Podemos calcular los momentos de una distribución de probabilidad discreta como

$$E[X^n] = \sum_x x^n p_x \approx \int_{-\infty}^{\infty} x^n p(x) dx. \quad (5.43)$$

Con $n = 1$ obtenemos el valor esperado de la distribución, y con $n = 2$ obtenemos el segundo momento, que nos ayuda a calcular la varianza de la distribución.

En el caso de la distribución de grado de una red libre de escala, podemos encontrar el n -ésimo momento de la distribución (caso continuo) como

$$\hat{k}^n = \int_{k_{min}}^{\infty} k^n p(k) dk. \quad (5.44)$$

Sustituyendo la distribución del grado para el intervalo observado en nuestra red, tenemos

$$\hat{k}^n = \int_{k_{min}}^{k_{max}} k^n (\gamma - 1) k_{min}^{\gamma-1} k^{-\gamma} dk = (\gamma - 1) k_{min}^{\gamma-1} \int_{k_{min}}^{k_{max}} k^{-\gamma+n} dk. \quad (5.45)$$

Utilizando la solución de la integral en 5.35 y definiendo $C = (\gamma - 1) k_{min}^{\gamma-1}$

$$\hat{k}^n = C \frac{k^{n-\gamma+1}}{n-\gamma+1} \Big|_{k_{min}}^{k_{max}} = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}. \quad (5.46)$$

Una vez más, centramos nuestra atención en el exponente $n - \gamma + 1$. Para que la comparación que estamos haciendo con las redes aleatorias tenga sentido, necesitamos considerar el caso extremo en el que $k_{max} \rightarrow \infty$; si $n - \gamma + 1 \leq 0$ (es decir, si $n \leq \gamma - 1$) entonces en el límite, el momento converge. Sin embargo, para valores de n tales de $n - \gamma + 1 \geq 0$ (es decir, $n \geq \gamma - 1$), el límite se hace infinito.

Lo anterior se entiende claramente si se complementa con un resultado en [Barabási and Pósfai \(2016\)](#) previamente discutido: siendo que para todas las redes estudiadas se cumple $2 < \gamma < 3$, tenemos que para los momentos mayores o iguales a 2, todos serán infinitos, pues $n - \gamma + 1 > 0$. Es decir, si pensamos que el rango de valores que puede tomar el grado de un nodo está dado por $\hat{k} \pm \sigma_k$, con σ_k la desviación estándar, entonces para una red libre de escala el grado de un nodo es

$$k = \hat{k} \pm \infty. \quad (5.47)$$

Por lo tanto, es claro porqué las redes libres de escala presentan nodos altamente conectados.

5.4.4. Estructuras Centro-Periferia

La teoría de redes ha avanzado grandemente en los últimos 20 años, sin embargo, la presente tesis no pretende brindar una introducción exhaustiva al respecto, por lo que hemos centrado nuestra atención en los conceptos esenciales para nuestro trabajo.

Con lo anterior en mente, es necesario introducir un nuevo concepto de gran utilidad en teoría de redes: **asortatividad** (véase [Newman \(2002\)](#)) y su contraparte, **disasortatividad**. Se puede medir la asortatividad en una red respecto a muchas de las variables que se pueden obtener a nivel de nodo. Por ejemplo, respecto al grado, respecto a la fuerza. En una red social, los nodos pueden tener características cualitativas útiles para medir la asortatividad, por ejemplo, la edad.

Se dice que los nodos de una red se mezclan asortativamente, o tienen una asortatividad positiva (respecto a una característica dada) si los nodos con valores similares se conectan mayormente entre ellos, y en menor medida con nodos con valores distantes. Lo anterior resulta relevante para definir una de las estructuras de red más socorrida en las redes financieras: **Centro-Periferia** (*Core-Periphery*).

Una estructura Centro-Periferia es un caso especial derivado de la asortatividad de una red de acuerdo con el *grado* de los nodos. Según [Newman \(2010\)](#), cuando una red presenta asortatividad por grado, los nodos altamente conectados (entre sí) forman una subestructura compacta que se puede pensar como un “núcleo” (traducción directa del inglés) que ocupa una posición central para la conectividad de la red.

Por otro lado, los nodos débilmente conectados tienden a formar ligas mutuamente, generando así otra subestructura mucho menos densa, que se considera como una periferia posicionada alrededor de los nodos del centro. Luego, gracias a que es posible que existan ligas que conecten nodos en el centro con nodos en la periferia, una red toma la forma de Centro-Periferia, en el que los nodos del centro tienen un papel importante de intermediación, ya que al estar débilmente conectados los nodos en la periferia, es mucho más factible que cualesquiera dos nodos logren conectarse a través del centro.

En [Craig and von Peter \(2014\)](#) dan una interpretación económica a la estructura Centro-Periferia que resultó atractiva para el estudio de redes financieras. En dicho trabajo se presenta un modelo para particionar una red en un centro y una periferia y lo aplican en redes financieras reales construidas con datos de exposición interbancaria del Banco Central Alemán. Encuentran evidencia de que dichas redes tienen un comportamiento distinto al de una red aleatoria, y que el flujo de dinero tiende a concentrarse en un subconjunto de nodos que hacen una fuerte labor de intermediación, es decir, las redes presentan una división de Centro y Periferia.

El modelo arroja un vector para identificar qué nodos en una red pertenecen al centro y cuáles a la periferia. Lo anterior se logra haciendo una comparación entre una red Centro-Periferia “perfecta” y la red real bajo estudio. El modelo C-P base se toma como una matriz

$$M = \begin{bmatrix} CC & CP \\ PC & PP \end{bmatrix}$$

donde cada entrada representa los arcos dirigidos Centro-Centro, Centro-Periferia, Periferia-Centro y Periferia-Periferia. El procedimiento lleva cuenta de todas las desviaciones que la red real tiene respecto al modelo base, siendo estas desviaciones conexiones presentes entre secciones de la red que no deberían de existir para poder considerarse una red C-P ideal.

Como podemos observar (Figura 5.3), el error⁷ del modelo de Craig y von Peter es considerablemente elevado, llegando casi al 35% en el periodo de la crisis financiera, lo cual resulta interesante, pues sugiere que en periodos de crisis, las instituciones financieras que en condiciones normales funcionan como intermediarios importantes cambian su actividad de manera drástica.

Más aún, el error tiene una tendencia creciente, lo que brinda evidencia de que el sistema bancario mexicano ha cambiado sus patrones de conectividad con el paso del tiempo, y cada vez se aleja más de un modelo Centro-Periferia, que en principio no es bueno para aproximar la red en todo el periodo de estudio. La diferencia con los resultados obtenidos por los autores del modelo puede deberse, entre otras cosas, a que el sistema bancario mexicano es mucho más pequeño que el que consideran los autores, lo que provoca que los bancos menos centrales sean más propensos a formar conexiones entre sí, violando así los supuestos del modelo.

⁷El error del modelo se calcula como el total de ligas existentes que se desvían del modelo base dividido por el total de ligas existentes en la red observada.

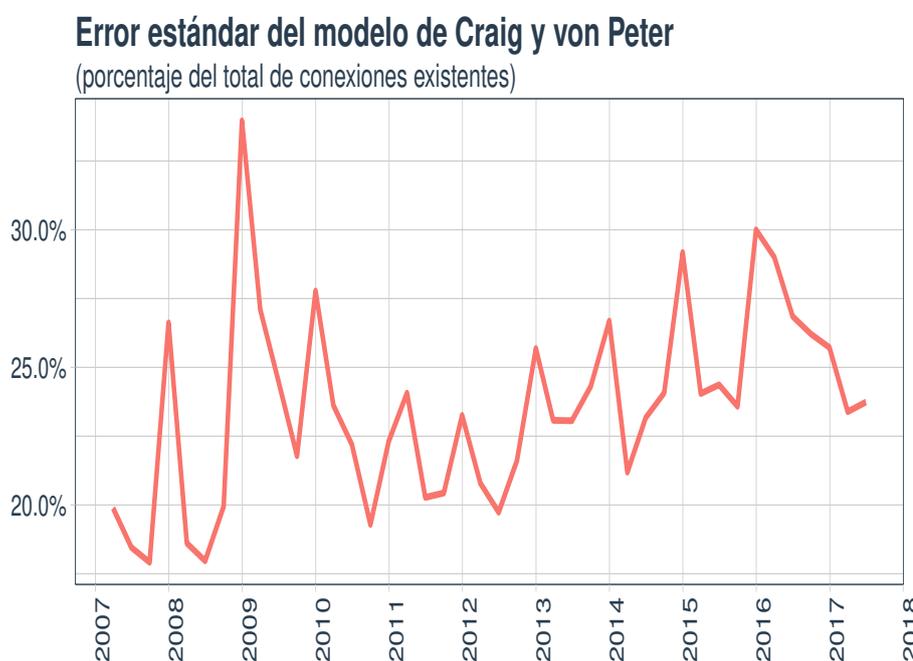


Figura 5.3: Error incurrido por el modelo Core-Periphery de Craig y von Peter en la red interbancaria mexicana.

En [van Lidth de Jeude et al. \(2018\)](#) se presenta un modelo diferente para detectar una estructura Centro-Periferia, con la gran ventaja de que gracias a su construcción es capaz de devolver una significancia estadística a la partición encontrada representada con un p-valor, lo cual resulta muy conveniente, pues como podremos comprobar en la sección de resultados, los promedios de las medidas de centralidad son muy similares diferenciando por las particiones obtenidas con ambos métodos. Sin embargo, como podemos observar en la siguiente tabla, la significancia del método se encuentra consistentemente por debajo de 0.05, salvo el valor máximo que empuja algunos cuantiles fuertemente, aún así tenemos la condición suficiente para considerar significativa la partición en todos los periodos salvo uno.

Cuantil	Valor
Min.	$2.300000000000000e - 153$
1st Qu.	$4.715000000000000e - 133$
Median	$1.245000000000000e - 127$
Mean	$2.38095238095238e + 00$
3rd Qu.	$8.243750000000000e - 117$
Max.	$1.000000000000000e + 02$

Llama la atención el valor máximo observado en la significancia de las particiones. El código original no cuenta con una excepción para el caso en el que el algoritmo no logre encontrar una partición, por lo que el 100 que observamos en la tabla fue un valor default que se agregó al código con el que se hicieron los ajustes para manejar errores en la ejecución. Las razones por las que no fue posible encontrar una partición para esa red no son claras, pues no es un día sin actividad en los mercados, no es menos densa que el resto de las fechas y el volumen total en el sistema no es mucho más bajo.

Dado lo anterior, es necesario comparar los resultados de los métodos de acuerdo con las asignaciones que obtiene para los bancos que forman nuestro sistema. En la Figura 5.4 vemos para una muestra de bancos que sí hay diferencias significativas entre los resultados de ambos métodos.

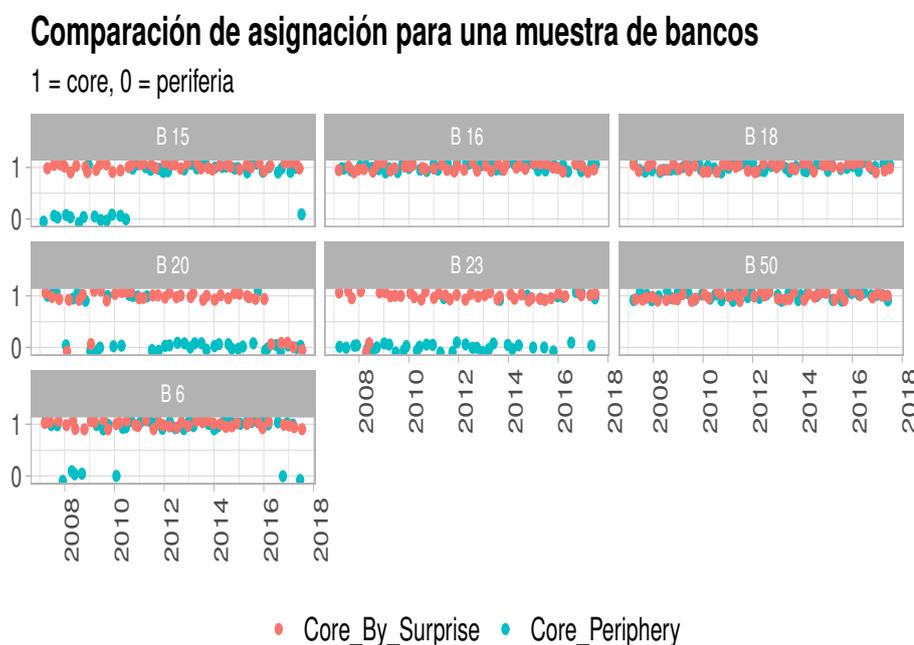


Figura 5.4: Diferencias de asignación al centro o a la periferia entre el modelo de Craig y von Peter y el método by Surprise para una muestra de bancos en México.

Hay que notar que a pesar de que ambos métodos difieren bastante en la asignación de algunas de las instituciones, hay otras para las que ambos métodos concuerdan a la perfección, en la figura anterior vemos que los bancos 16, 18 y 50 pertenecen al core con ambos métodos para todo el periodo de estudio, lo que da evidencia de que el fenómeno *Too-Big(Interconnected)-To-Fail* sigue vigente en el sistema bancario mexicano.

5.5. Resultados

Como ya mencionamos, el análisis de las medidas estructurales de una red es una de las herramientas básicas de la teoría de redes, pues nos permiten identificar participantes con importancia sistémica para una red. En el caso de las redes financieras, sin importar su tamaño, la identificación de dichas instituciones es de suma importancia, y ha sido integrado a las tareas de supervisión y regulación en todo el mundo.

Con la intención de brindar un poco más de información del sistema bancario mexicano, a continuación analizamos la evolución de un conjunto de medidas de red del sistema interbancario en nuestro país. Primero, presentamos gráficas correspondientes a los promedios de las medidas de red, y posteriormente vemos gráficas correspondientes a las instituciones bancarias más importantes en cada medida analizada.

5.5.1. Promedio del sistema

En general, observar el promedio de las medidas de red nos da una idea del comportamiento general del sistema bancario, principalmente, nos ayudan a detectar tendencias y fechas con comportamientos atípicos. En las Figuras 5.5 y 5.6 observamos la evolución del promedio para un amplio conjunto de medidas de red.

Primero, es importante notar la tendencia creciente que tienen algunas de las series. En el caso de la afinidad, cercanía, interinidad y la fuerza como prestamista y prestatario, la tendencia positiva puede ser consecuencia directa del crecimiento del sistema (principalmente en la fuerza y en el aumento del camino más corto promedio) y del aumento en la conectividad del sistema (cercanía e interinidad), es decir, las instituciones bancarias en el país han aumentado, en promedio, el número de instituciones con las que interactúan. El aumento coordinado de la longitud promedio de los caminos más cortos y de la centralidad por cercanía puede parecer contradictorio, sin embargo, al considerar el aumento en el número de instituciones en el sistema, y el que algunas instituciones cobran mayor importancia (y por ende se conectan con más contrapartes directamente), podemos explicar que tome más pasos llegar de un banco a otro, pero que el promedio de conexiones directas aumente. No obstante, la tendencia presenta un corte importante a partir de 2012, fecha para la que las series se mantienen casi constantes hasta el final del periodo.

Vemos que la longitud del camino más corto, en general se mueve de manera inversa al resto de las medidas mencionadas, siendo más visible el efecto en las fechas posteriores al inicio de la crisis financiera del 2008, en el que la longitud del camino más corto se incrementó considerablemente, provocando una desconexión generalizada en la red que puede observarse en las métricas comparadas con la longitud en el párrafo anterior.

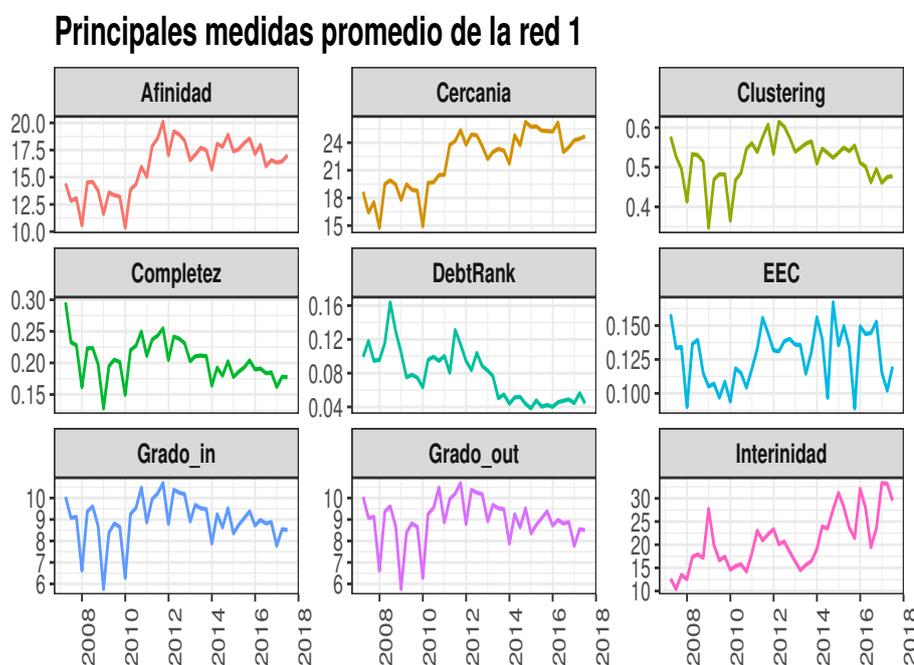


Figura 5.5: Promedio del sistema bancario de las principales medidas de red a lo largo del período de estudio (1).

Un par de series más con una tendencia creciente son las correspondientes al índice Herfindahl-Hirschman, aquí, los valores altos no son deseables, pues significan una mayor concentración de riesgo, así, una concentración de fondeo (HHI_B) y de proveedor (HHI_L) de entre 35 y 40 % resulta elevada. Pese a lo anterior, es una buena señal observar que el promedio del DebtRank haya experimentado una tendencia a la baja tan fuerte en los últimos años, para terminar en una serie constante alrededor de valores pequeños. Lo anterior sin duda es resultado del trabajo regulatorio realizado en los últimos años, específicamente el enfocado al mejoramiento del capital que tienen que formar los bancos.

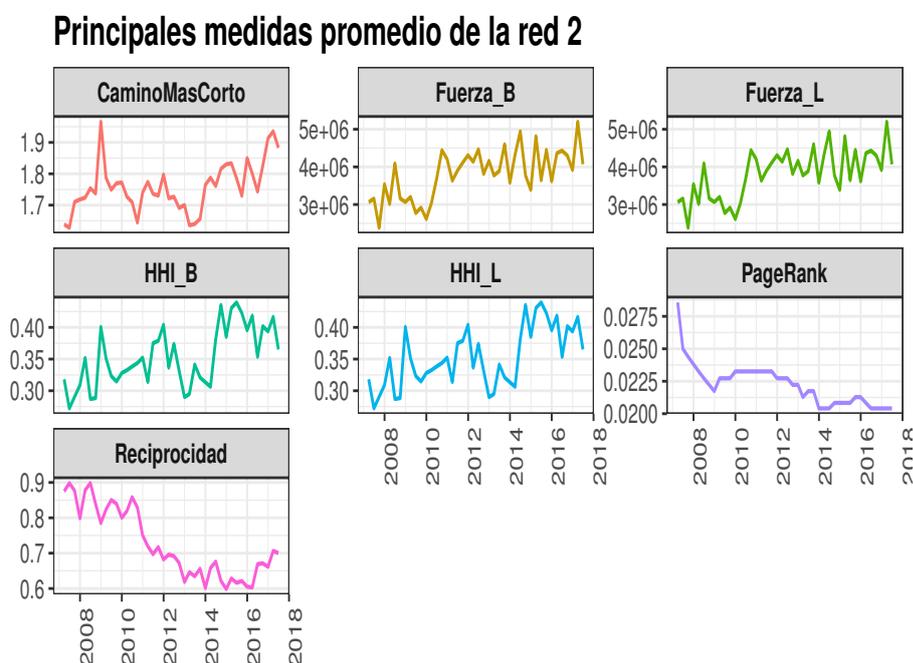


Figura 5.6: Promedio del sistema bancario de las principales medidas de red a lo largo del período de estudio (2).

5.5.2. Centralidad

Desagregando los resultados, podemos obtener visualizaciones de un subconjunto de las medidas de red para cada institución, es decir, podemos obtener series para visualizar la centralidad de cada institución para un conjunto de medidas con una interpretación clara para nuestras redes.⁸

Las siguientes gráficas (ver las Figuras 5.7 a 5.10) contienen la evolución de las medidas de centralidad para las 10 instituciones más centrales, así como para las 10 menos centrales, esto se hizo con la intención de facilitar la visualización (y la identificación) de las instituciones más y menos “importantes” de acuerdo a cada métrica. Lo primero que salta a la vista es la heterogeneidad de las series para algunas de las medidas, lo que habla en sí de que los bancos en México pueden ser diferenciados de acuerdo a su centralidad.

Por ejemplo, si observamos las series correspondientes a la cercanía y al DebtRank (Figura 5.7). Es claro que las instituciones más “cercanas” en la red tienen un comportamiento similar y creciente, cosa que se puede observar desde el promedio. Por el

⁸Las series de las gráficas por institución fueron suavizadas automáticamente con la función “geom_smooth” de la biblioteca Wickham (2016), con el objetivo de detectar con mayor facilidad la tendencia de cada serie.

5. REDES FINANCIERAS

contrario, las series son mucho más heterogéneas para las 10 instituciones menos centrales. En ambos casos, hay que notar que los análisis capturan la entrada y salida de instituciones a la red.

En lo que respecta a las gráficas C y D, es importante notar que un conjunto reducido de instituciones son consistentemente las más centrales por DebtRank, y con niveles bien diferenciados entre ellas, en particular, el banco “B 16” es consistentemente el más sistémico. Si vemos ahora las series para los bancos con el DebtRank más chico, hay un comportamiento bastante homogéneo a partir de 2014, más aún, como podemos observar, la fecha inicial de la gráfica D es 2012. Dado que la selección de los 10 bancos más y menos centrales se realizó promediando para cada institución a lo largo de todo el periodo, el hecho de que antes de 2012 no se hayan considerado quiere decir que son los bancos de más reciente creación los menos sistémicos.

Lo anterior sugiere que mientras más tiempo de actividad tenga un banco, mayor es la percepción positiva que da al resto del sistema, por lo que es más fácil para él interactuar con el resto de la red, aumentando así su conectividad y por lo tanto su “potencial” sistémico.

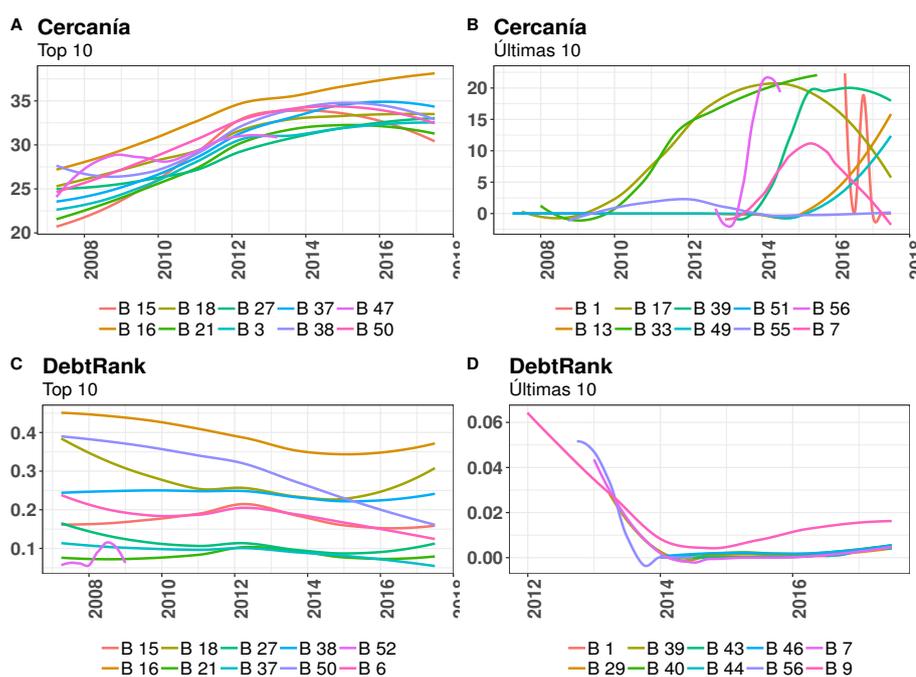


Figura 5.7: Centralidad por Cercanía y por DebtRank para las 10 instituciones con los valores más altos y más pequeños.

Para las medidas de Grado y Fuerza (Figura 5.8) es claro el alto nivel de correlación que tienen, lo que se hace evidente en las tendencias de las series, incluso al final del periodo cuando ambas medidas sufren una caída. Al final, como podemos ver en el

ápndice B, la fuerza es un grado con pesos. Sin embargo, no sorprende observar que la institución “B 16” es también la más conectada y la que más dinero mueve dentro de la red. Más aún, el conjunto de instituciones que aparecen en el top 10 de las más centrales por DebtRank, es casi el mismo conjunto que aparece en las más centrales por grado y fuerza, a saber, las instituciones 15, 16, 18, 27, 37, 38 y 50 aparecen en las tres medidas dentro de los 10 bancos más centrales.

Para las instituciones menos conectadas, hay una selección mucho más heterogénea, empezando por el conjunto de bancos que aparecen en las tres medidas (sólo los bancos 7 y 56 están presentes en las instituciones menos conectadas para las tres medidas), lo que apunta a una mayor “competencia” en término de conectividad para los bancos menos centrales.

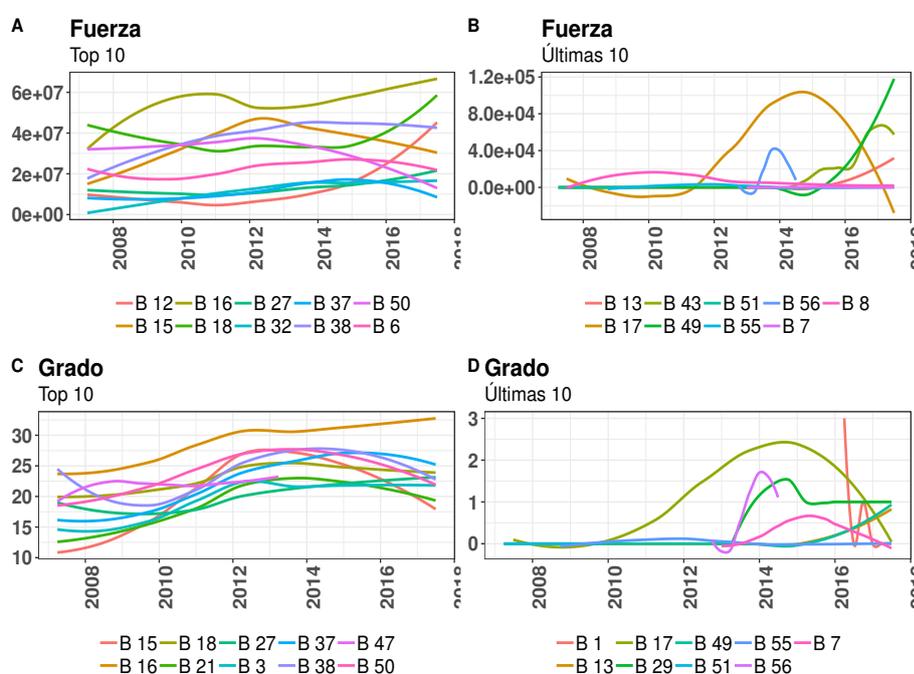


Figura 5.8: Centralidad por Fuerza y Grado para las 10 instituciones con los valores más altos y más pequeños.

Como ya mencionamos, la concentración es un concepto de gran importancia en cualquier análisis de riesgos, por lo anterior, observamos la evolución del índice Herfindahl-Hirschman para las instituciones más y menos concentradas como prestamista y como prestatario (Figura 5.9).⁹

⁹Como se menciona en el apéndice B, el índice de Herfindahl-Hirschman está acotado al intervalo de valores $[0, 1]$, sin embargo, en las gráficas podemos observar valores que están fuera del intervalo permitido hacia arriba y hacia abajo, esto es producto del suavizado aplicado a los datos.

5. REDES FINANCIERAS

Uno de los primeros resultados importantes respecto a la concentración es que la institución 16, a pesar de ser la que acarrea un mayor riesgo sistémico para la red, está entre las instituciones menos concentradas como prestamista, lo que quiere decir que son muchas las instituciones que obtienen flujos de efectivo de ella, aquí sería necesario verificar las exposiciones hacia 16, pues si la gran mayoría están cerca de una cantidad que mueva significativamente los índices de capitalización de los bancos expuestos a 16, dicha concentración es sin duda nociva, pues acarrea un efecto de potenciador para una hipotética cadena de contagio. Por el contrario, si todas las exposiciones a 16 están en un rango saludable, dicha dispersión tiene un efecto de diversificación.

Por lo general, para las instituciones con una centralidad más alta, especialmente en DebtRank, es deseable tener índices de concentración pequeños, ya que así se “diluyen” efectos potenciadores de estrés en la red, por lo anterior, es importante mencionar que ninguna de las instituciones presentes en el top 10 de centralidad por DebtRank se encuentran en ninguna de las dos gráficas de las instituciones más concentradas.

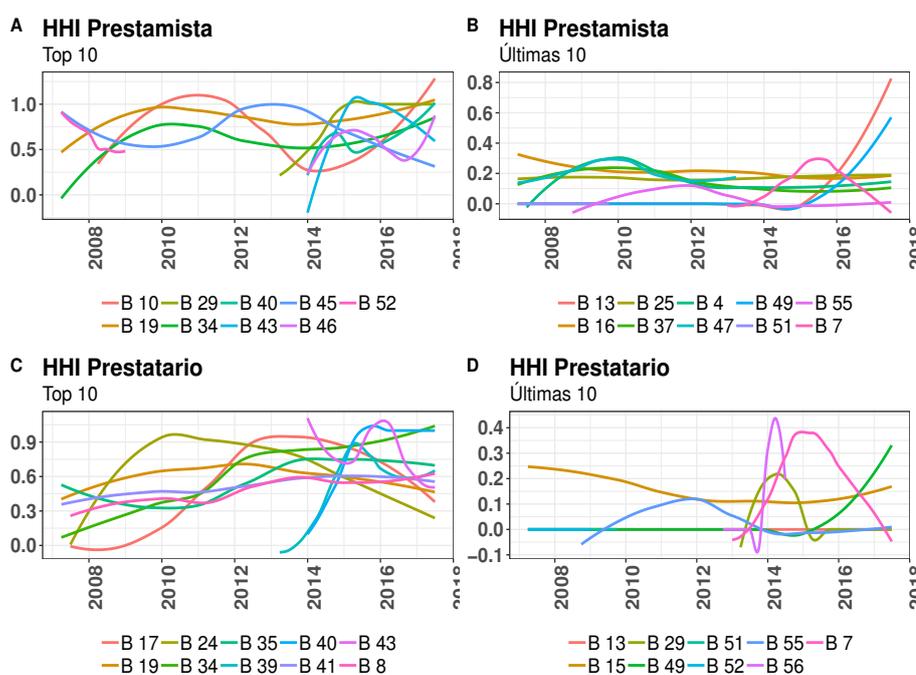


Figura 5.9: Índice Herfindahl-Hirschman como prestamista y prestatario para las 10 instituciones con los valores más altos y más pequeños.

Finalmente tenemos la centralidad por Interinidad y PageRank (Figura 5.10), donde vemos de vuelta a las instituciones más centrales en otras medidas (15, 16, 18, 27, 37, 38 y 50) en el top 10. Para los 10 menos centrales en interinidad, la medida es básicamente cero, lo que apunta al nulo trabajo de intermediación de las instituciones que vemos en dicha gráfica.

Como sabemos, el PageRank considera centrales a las instituciones que están conectadas a su vez con otras instituciones centrales, por lo tanto, se confirma que la interacción entre las instituciones más importantes en otras medidas tiene también un impacto en su centralidad por PageRank (ver Figura 5.10 C).

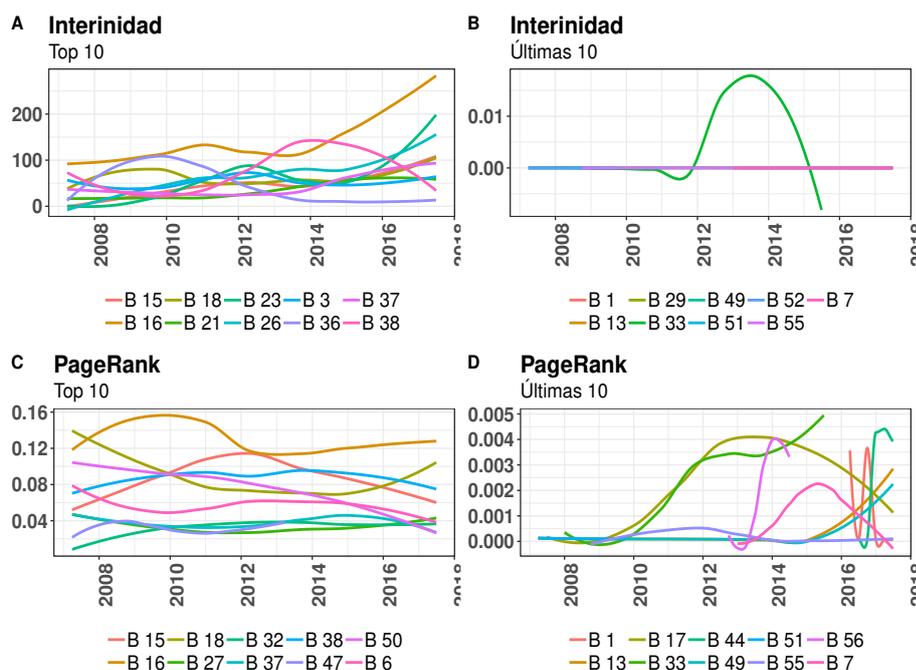


Figura 5.10: Centralidad por Interinidad y PageRank para las 10 instituciones con los valores más altos y más pequeños.

5.5.3. Medidas de red promedio por centro y periferia

En [van Lidth de Jeude et al. \(2018\)](#) presentan un procedimiento estadístico para la detección de estructuras Centro-Periferia, mismo que fue utilizado en nuestras redes. Los resultados que arrojó el algoritmo son significativos para las redes bancarias que nos ocupan, pues podemos observar claramente la diferencia entre los promedios de las principales medidas de red separando por centro y periferia (Figuras 5.11 y 5.12). En general, las instituciones pertenecientes al Centro (color azul) tienen una centralidad claramente mayor que las instituciones pertenecientes a la periferia (color rojo).

Las medidas promedio de la red diferenciando por centro o periferia nos permiten ver que en general, los bancos con la centralidad más alta tienden a estar en un mismo grupo, lo que nos confirma la teoría de que el Too-Big-Too-Fail y el Too-Interconnected-To-Fail siguen vigentes en el sistema financiero mexicano, por lo que el trabajo para disminuir sus efectos sigue siendo relevante, así, seguirá siendo un tema de interés en la agenda regulatoria de México.

Es necesario notar que en el caso particular del DebtRank, hay una clara distinción entre las dos series después del 2014, pues las instituciones del centro conllevan en promedio un **riesgo sistémico más alto** (Figura 5.11).

En el caso del Índice de Herfindahl-Hirschman tenemos un comportamiento interesante. Dada la interpretación del mismo, es deseable tener valores bajos, pues esto significa una exposición más diversificada entre las instituciones. Vemos que en 2008 e inicio de 2009, el índice era más alto para las instituciones del centro actuando como prestatarias; recordemos que en 2008 y 2009 se vivió la parte más álgida de la crisis financiera mundial, las instituciones sistémicas estaban concentrando su fondeo, y ésto se ve claramente reflejado en el estrés del sistema en el DebtRank (gráfica 5.5).

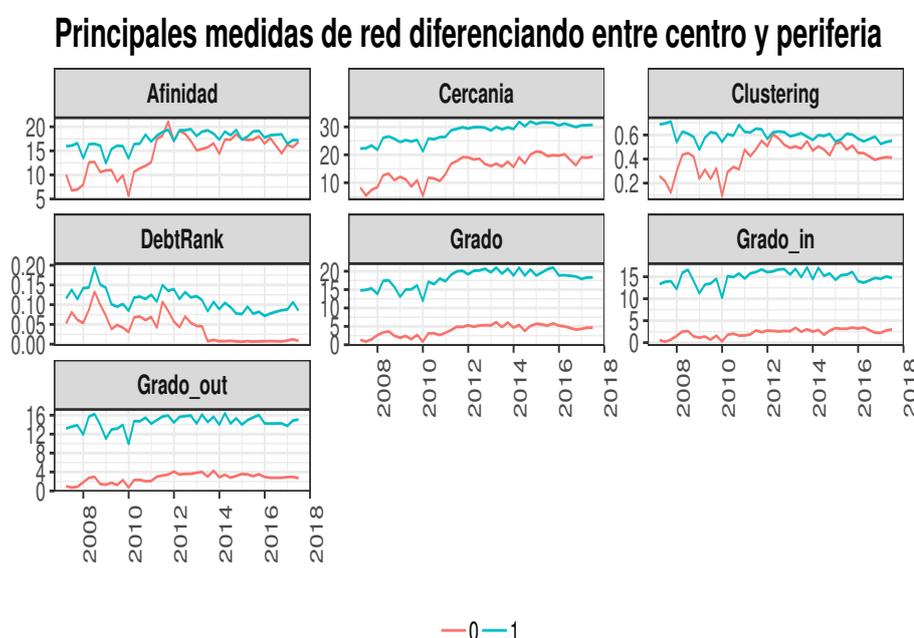


Figura 5.11: Promedio del sistema bancario de las principales medidas de red diferenciando por el centro y la periferia obtenidos con el método by Surprise (1).

En general, las instituciones de la periferia tienen un índice de Herfindahl-Hirschman (HH) más alto que las instituciones del centro, situación que está en línea con la hipótesis de [Craig and von Peter \(2014\)](#), es decir, que los bancos del centro tienden a hacer una labor de intermediación intensa, lo que se traduce en una mayor cantidad de conexiones con un número mayor de instituciones; dicha hipótesis se hace más fuerte si vemos las series de interinidad: los bancos del centro sirven como centros de conectividad en mucho mayor medida que los bancos de la periferia (Figura 5.11).

Al contrastar los resultados de las medidas de centralidad por institución con las medidas promedio por centro-periferia, queda claro que partir a los bancos en sólo

dos grupos no toma en cuenta la heterogeneidad del comportamiento de los bancos de manera individual, particularmente entre las instituciones que se encuentran fuera de las 10 más centrales, por lo anterior, la partición en un número mayor de grupos, como la realizada con los métodos de conglomerados y con los modelos de bloques estocásticos (que exponemos en el siguiente capítulo) resulta interesante, pues se benefician de una estructura rica en variaciones y significado.

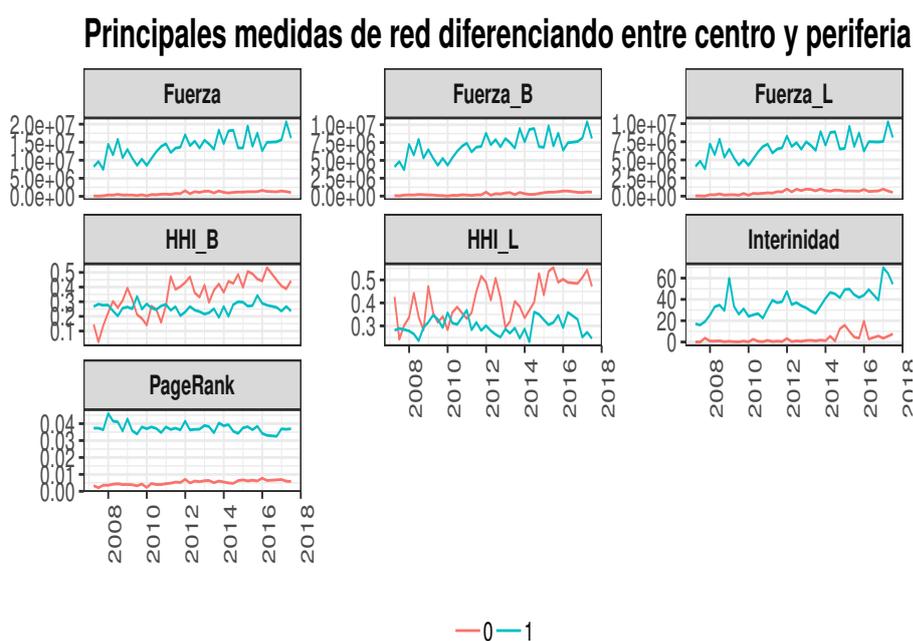


Figura 5.12: Promedio del sistema bancario de las principales medidas de red diferenciando por el centro y la periferia obtenidos con el método by Surprise (2).

Modelos de Bloques Estocásticos

En el capítulo anterior mencionamos que los mercados financieros son en esencia sistemas complejos que nacen de la interacción entre sus participantes. En el presente trabajo limitamos el conjunto de participantes a los bancos que han operado y siguen operando en México, un conjunto pequeño si lo comparamos con los sistemas bancarios de otras regiones, sin embargo es un conjunto de participantes heterogéneos en sus características, tanto en tamaño como en intensidad de la actividad. Es esa heterogeneidad la que nos permite hacer un análisis para desenmascarar estructuras de agrupación de nuestros bancos.

Hemos mencionado que en el presente trabajo tenemos dos objetivos: obtener una clasificación de los bancos en México de acuerdo con sus actividades utilizando información de balance, cuyos resultados presentamos en los capítulos 3 y 4, y otra de acuerdo con sus niveles de actividad en la red de exposiciones. Para identificar la clasificación de acuerdo con la actividad, utilizaremos una de las muchas herramientas para detección de comunidades que nos brinda la teoría de redes, a saber, los **Modelos de Bloques Estocásticos** o *SBM* (Nowicki and Snijders (2001)), por sus siglas en inglés, mismas que utilizaremos para referirnos al método en adelante.

Dado que los SBM tienen como finalidad la *detección de comunidades* en una red, mismas que representan bloques de nodos con características similares, los cuales ayudan a describir un sistema de forma más sencilla al permitirnos agregar propiedades generales de la red a nivel de bloque, resultan ser un método conveniente para el problema que nos ocupa en el presente trabajo. Utilizando las mismas redes analizadas en el capítulo anterior, aplicaremos el Modelo de Bloques Estocásticos para identificar grupos de nodos similares de acuerdo a su conectividad en la red.

6.1. Breve introducción a la estadística Bayesiana

Más adelante vemos que el método utilizado para la búsqueda de comunidades que aplicamos utiliza herramientas de la *estadística Bayesiana* (para un curso general, ver Gelman et al. (2014)). Por lo anterior, introducimos las nociones básicas necesarias

para entender cómo funciona el Modelo de Bloques Estocásticos.

6.1.1. Herramientas básicas: Probabilidad total, Teorema y Fórmula de Bayes

Como su nombre lo indica, la estadística bayesiana hace uso de uno de los resultados más importantes en la teoría de la probabilidad: la Fórmula de Bayes, por lo que es importante recordar brevemente la definición de la fórmula, y dos resultados previos necesarios para obtenerla.

Primero, recordemos el teorema de la **Probabilidad Total**: Sean A_1, A_2, \dots, A_n una partición del espacio muestral Ω , y sea $B \in \Omega$ para el cual se conocen las probabilidades $\mathbb{P}(B | A_i) \forall i \in \{1, 2, \dots, n\}$; el Teorema de la Probabilidad Total nos dice que:

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B | A_i) \mathbb{P}(A_i). \quad (6.1)$$

Por otro lado, para el **Teorema de Bayes** necesitamos A_1, A_2, \dots, A_n eventos aleatorios mutuamente excluyentes, tales que $\mathbb{P}(A_i) > 0 \forall i \in \{1, 2, \dots, n\}$; sea B otro evento aleatorio para el que se conocen todas las probabilidades $\mathbb{P}(B | A_i)$, entonces, el **Teorema de Bayes** nos dice que:

$$\mathbb{P}(A_i | B) = \frac{\mathbb{P}(B | A_i) \mathbb{P}(A_i)}{\mathbb{P}(B)}. \quad (6.2)$$

Finalmente, al sustituir la forma de calcular $\mathbb{P}(B)$ que nos brinda el Teorema de la Probabilidad Total en el Teorema de Bayes obtenemos la **Fórmula de Bayes**:

$$\mathbb{P}(A_i | B) = \frac{\mathbb{P}(B | A_i) \mathbb{P}(A_i)}{\sum_{k=1}^n \mathbb{P}(B | A_k) \mathbb{P}(A_k)}. \quad (6.3)$$

Más adelante vemos de manera explícita cómo es que la estadística bayesiana se vale de los resultados anteriores para hacer inferencia sobre la forma en la que se comporta un fenómeno aleatorio en una población.

6.1.2. Diferencias entre el enfoque frecuentista y el bayesiano

Es necesario entender cuáles son las diferencias entre la estadística clásica (también llamada frecuentista) (ver [Casella and Berger \(2001\)](#) para un curso introductorio a la estadística frecuentista) y la estadística Bayesiana, para ello, consideramos dos interpretaciones distintas que se le pueden dar a una probabilidad: la *frecuentista* y la *bayesiana*.

En [Murphy \(2012\)](#) encontramos un ejemplo que clarifica dichas interpretaciones: supongamos que realizamos el experimento aleatorio de lanzar una moneda justa. En

la interpretación *frecuentista* una probabilidad representa la frecuencia con la que ocurre un evento aleatorio en una serie de repeticiones del experimento. Por otro lado, la interpretación *Bayesiana* cuantifica la **incertidumbre** que tenemos respecto al resultado del experimento. Así, bajo el enfoque frecuentista, si la probabilidad de obtener “sol” al lanzar una moneda es 0.5, estamos diciendo que al repetir el experimento un gran número de veces esperamos que en la mitad el resultado sea “sol”, y la otra mitad sea “aguila”. Por el contrario, una probabilidad de 0.5 bajo el enfoque Bayesiano quiere decir que los eventos de obtener “sol” o “aguila” en el siguiente lanzamiento son equiprobables.

En el ejemplo anterior se encuentra implícitamente una más de las diferencias primordiales, pues mientras que la estadística frecuentista depende en gran medida de la existencia de muestras de tamaño considerable para hacer estimaciones, la estadística bayesiana nos permite traducir la escasez de información en una **distribución a priori**, que encapsula la ya mencionada incertidumbre que hay respecto a los parámetros que gobiernan un fenómeno aleatorio, más aún, la estadística bayesiana, haciendo uso de la Fórmula de Bayes, nos dice exactamente cómo debemos de ajustar la probabilidad de un evento aleatorio dada la incorporación de nueva información a nuestro conjunto original.

Los parámetros obtenidos con cualquiera de los dos enfoques se utilizan para hacer **inferencia** sobre un proceso aleatorio que ocurre en una población. Hacer inferencia se refiere a obtener evidencia para llegar a una conclusión, obtenida a partir de una **muestra** de observaciones de un fenómeno que se desea estudiar; estamos limitados a trabajar con una muestra para hacer inferencia estadística dado que obtener información de cada uno de los elementos de una población siempre resulta inviable por el alto costo y la gran complejidad de hacerlo, por lo que el enfoque bayesiano resulta conveniente gracias a la facilidad con la que nueva información se agrega al cálculo de una estimación.

En general, son dos tipos de estimaciones las que deseamos hacer con un análisis estadístico: estimaciones de valores *observables* y valores *no observables*. Los valores observables hacen referencia a cantidades que podemos encontrar como resultado de un evento aleatorio, por ejemplo, el estado de una institución financiera en periodos de crisis (por ejemplo en quiebra, saludable, bajo estrés, etcétera) o la altura de una persona seleccionada al azar en una población; las cantidades no observables son aquellas que gobiernan un proceso aleatorio, por ejemplo, los parámetros de una regresión lineal, o la media y la varianza de un proceso con distribución normal.

Los intervalos para un estimador también resultan primordiales en cualquier análisis que realicemos, pues son el primer paso para conocer la importancia de un parámetro en el modelaje de un proceso aleatorio, y en ellos encontramos una vez más una diferencia importante entre la estadística frecuentista y la bayesiana.

Un intervalo frecuentista (intervalo de confianza) tiene la característica de que, dados los supuestos de distribución hechos sobre los fenómenos bajo estudio, se puede calcular de manera analítica incluso antes de obtener datos para hacer inferencia, por lo que un intervalo frecuentista sólo es válido para un conjunto de inferencias que

contengan los mismos supuestos, y con un conjunto de datos obtenidos de la misma (replicable) forma, por lo que, como se menciona en Hoff (2009), da información sobre el posible valor de un parámetro *antes* de que los datos sean observados. Por otro lado, un intervalo bayesiano (de probabilidad), al estar calculado con parámetros calculados con un procedimiento que implica la incorporación de datos a un modelo, nos da información sobre el valor de un parámetro *después* de observar datos.

6.1.3. Conceptos principales

Ahora, presentaremos las herramientas necesarias para definir los conceptos principales de la estadística bayesiana. Primero, fijamos la notación que utilizaremos para identificar las partes primordiales de un estudio bayesiano.

- En general, en la literatura se utiliza la letra griega θ para denotar cantidades no observables, es decir, parámetros.
- Para identificar el conjunto de datos con el que se realiza el análisis (datos **observados**) utilizamos la letra y .
- Los datos observables que estimamos a partir de los valores obtenidos de θ los denotamos con \hat{y}

Para cualquier estudio, es necesario reunir información de un conjunto de objetos (una muestra), comúnmente, al número de objetos a los que se les mide la cantidad de interés es denotado como n , por lo que podemos acomodar los datos observados en un vector $y = (y_1, \dots, y_n)$, donde la entrada i contiene el valor observado para el i -ésimo objeto.

Adicionalmente, es común que a cada objeto en un estudio se le midan otras variables que pueden servir para explicar el valor observado en y_i , dichas variables pueden llamarse *covariables* o *variables explicativas*, y se denotan con x . Así, terminamos con un conjunto de observaciones $(x, y)_i$, que representan el resultado de la variable de interés y_i y las variables explicativas x_i para cada uno de los i objetos en nuestra muestra. Es importante notar que siempre es posible utilizar algunas de las variables en x como variable respuesta y , e incorporar la variable y original al conjunto de predictores.

Como se discutió anteriormente, la estadística bayesiana tiene el objetivo de integrar la incertidumbre que se tiene al modelaje y la estimación de cantidades de interés, es por esto que todo se traduce a un lenguaje probabilístico; así, partimos de los datos que conocemos para llegar a conclusiones sobre los datos que no conocemos, por lo que condicionamos nuestro análisis a los datos observados y , es decir, nos interesa llegar a tener expresiones para:

- $$p(\theta | y), \tag{6.4}$$

que representa la probabilidad de observar un valor para θ dada la información existente. Esta densidad es de gran importancia en el análisis que realizamos,

pues representa el conocimiento (o completa ignorancia) que tenemos respecto al parámetro gobernador de nuestro proceso aleatorio.

■

$$p(\hat{y} | y), \quad (6.5)$$

que es la probabilidad de tener un valor \hat{y} para un objeto del que no se tiene la medición, calculada a partir de la información que se tiene de otros objetos similares.

En términos prácticos, necesitamos una herramienta que nos permita hacer declaraciones conjuntas para θ y y , pues al final estamos asumiendo que θ gobierna el proceso a partir del cual se generan los valores en y . Podemos lograr lo anterior a partir de la forma que le podemos dar a una probabilidad conjunta utilizando probabilidad condicional. Recordemos que para dos eventos aleatorios A y B , la probabilidad condicional de B dado A es

$$p(B | A) = \frac{p(A, B)}{p(A)}, \quad (6.6)$$

donde $p(A, B)$ es la probabilidad conjunta, si reescribimos la expresión anterior, tenemos que

$$p(A, B) = p(B | A)p(A), \quad (6.7)$$

sustituyendo las cantidades de nuestro interés, tenemos que

$$p(\theta, y) = p(y | \theta)p(\theta), \quad (6.8)$$

de donde identificamos dos funciones de probabilidad importantes. A $p(\theta)$ se le conoce como la **distribución a priori**, y a $p(y | \theta)$ se le conoce como la distribución de muestreo, o la **distribución de los datos**.

Dado que conocemos los datos y , y a partir de ellos queremos actualizar nuestro conocimiento respecto al parámetro θ , tiene sentido que nos enfoquemos en la función de probabilidad

$$p(\theta | y) = \frac{p(\theta, y)}{p(y)}, \quad (6.9)$$

sustituyendo 6.8 en 6.9, obtenemos la función de distribución **posterior** de nuestro parámetro θ dados los datos observados:

$$p(\theta | y) = \frac{p(\theta)p(y | \theta)}{p(y)}, \quad (6.10)$$

donde, por el teorema de la Probabilidad Total sabemos que $p(y) = \sum_{\theta} p(\theta)p(y | \theta)$ (o $p(y) = \int_{\theta} p(\theta)p(y | \theta)d\theta$ si θ tiene una distribución continua).

Ya que tenemos una expresión para hacer inferencia sobre el parámetro θ , que es una cantidad no observable, de acuerdo a lo que definimos antes, ahora necesitamos una expresión para hacer inferencia sobre cantidades observables, es decir, para hacer

predicciones de nuevos elementos de nuestra población, para ello, definimos dos cosas, la primera, es la **distribución predictiva a priori**, que corresponde a la distribución marginal de nuestros datos y que escribimos como

$$p(y) = \int p(y, \theta) d\theta, \quad (6.11)$$

utilizando de nuevo la probabilidad condicional tenemos que

$$p(y) = \int p(y | \theta) p(\theta) d\theta. \quad (6.12)$$

Para obtener la **distribución predictiva posterior** necesitamos condicionar los valores nuevos \hat{y} a los existentes y , además, los valores que estimamos dependen directamente del parámetro θ , pues no olvidamos que es el valor que gobierna el proceso generador de las observaciones y , y por lo tanto de \hat{y} ; así, tenemos

$$p(\hat{y} | y) = \int p(\hat{y}, \theta | y) d\theta, \quad (6.13)$$

usando la definición de la probabilidad condicional

$$p(\hat{y} | y) = \int \frac{p(\hat{y}, \theta, y)}{p(y)} d\theta, \quad (6.14)$$

ahora, dado que la única información con la que contamos es y , y a partir de ésta podemos estimar valores (probabilidades) para θ , condicionamos la probabilidad conjunta de la ecuación anterior y escribimos

$$p(\hat{y} | y) = \int \frac{p(\hat{y} | \theta, y) p(\theta, y)}{p(y)} d\theta, \quad (6.15)$$

condicionamos de nuevo la probabilidad conjunta

$$\begin{aligned} p(\hat{y} | y) &= \int \frac{p(\hat{y} | \theta, y) p(\theta | y) p(y)}{p(y)} d\theta \\ &= \int p(\hat{y} | \theta, y) p(\theta | y) d\theta \\ &= \int p(\hat{y} | \theta) p(\theta | y) d\theta, \end{aligned} \quad (6.16)$$

donde el último paso se debe al supuesto de independencia entre \hat{y} y y dado un parámetro θ . La última expresión tiene un significado importante, pues representa a la probabilidad de observar un nuevo valor \hat{y} condicionando al parámetro θ , que a su vez está ponderada por la probabilidad de observar un valor para θ dados los datos observados y .

Los conceptos anteriores constituyen los bloques básicos sobre los que se construye la forma Bayesiana de hacer inferencia estadística. Es importante tener presente que

las distribuciones que seleccionemos para representar la distribución a priori de nuestro parámetro puede pertenecer a alguna de las distribuciones más comunes en la teoría de la probabilidad. Como mencionamos, debe de reflejar el conocimiento que el investigador tiene respecto a los parámetros del modelo, así, es común encontrar distribuciones *a priori* uniformes¹, pues al desconocer el fenómeno bajo estudio, considerar a los eventos como equiprobables resulta natural.

Ahora, para entender mejor la forma en la que la estadística Bayesiana funciona, reproducimos los pasos que se definen en [Kruschke \(2014\)](#) para realizar un análisis bayesiano sobre un conjunto de datos:

1. Identificar los datos más relevantes para responder una pregunta de interés. Es importante diferenciar entre posibles covariables y la variable respuesta.
2. Definimos un modelo que sea relevante para modelar la información, cuyos parámetros y características sean adecuadas para los datos recabados en el paso inicial.
3. Imponer una distribución a priori a los parámetros del modelo, que contenga la mayor cantidad de información posible sobre los mismos de acuerdo al conocimiento previo que se tenga del fenómeno a modelar.
4. Usamos la fórmula de Bayes (a través de la distribución posterior de los parámetros) para recalculamos el grado de incertidumbre de los parámetros utilizando la información existente.
5. Verificamos la exactitud del modelo ajustado con los parámetros estimados “replicando” los valores que se tienen con la especificación obtenida. Si el modelo devuelve datos demasiado alejados de los datos reales, es necesario repetir el proceso desde la selección de un modelo que describa el fenómeno aleatorio.

Con lo anterior, queda claro que la estadística bayesiana se ocupa de “reacomodar” la incertidumbre que tenemos respecto a un proceso en todas las posibilidades de resultados que tiene el fenómeno, tomando en consideración la incorporación de información nueva a nuestro análisis.

Por ejemplo, [Kruschke \(2014\)](#) da un excelente ejemplo para entender la idea detrás de la estadística bayesiana: supongamos que se tiene una lista de posibles sospechosos de un crimen. Inicialmente, dada la inexistencia de datos sobre el crimen, lo mejor que podemos hacer como investigadores es repartir la probabilidad de ser el criminal de manera uniforme en todos los sospechosos. Sin embargo, después de hacer un análisis preliminar llega a ser de nuestro conocimiento que el criminal es del sexo masculino, así, podemos eliminar a todas las mujeres de nuestra lista de sospechosos y repartir la incertidumbre entre los sospechosos que queden en la lista.

Siguiendo un procedimiento similar, podemos eliminar poco a poco la incertidumbre sobre quién es el criminal, hasta llegar al culpable agregando evidencia en su contra de

¹La distribución uniforme es un caso particular de la distribución beta, la cual resulta ser una de las más estudiadas y aplicadas en el enfoque bayesiano.

manera iterativa. Para ejemplificar la forma en la que la estadística bayesiana reasigna certeza / incertidumbre a un evento a partir de nueva información, de [Gelman et al. \(2014\)](#) reproducimos un ejemplo numérico relacionado con la *hemofilia*, enfermedad que afecta la coagulación en la sangre. A partir de estudios se ha demostrado que la presencia de dicha enfermedad está vinculada con el cromosoma X heredado de padres a hijos. Se sabe que los hombres tienen un cromosoma X y uno Y, mientras que las mujeres tienen ambos cromosomas X.

Calcularemos la probabilidad *posterior* de que una mujer tenga la enfermedad a partir de la siguiente información:

1. La mujer tiene un hermano con la enfermedad (es decir, la madre es portadora del gen).
2. Su padre no tiene la enfermedad, lo que le da una probabilidad de tener el gen de 50% (de los dos cromosomas X de su madre, puede heredar el portador de la enfermedad).

Con lo anterior, el estado de la mujer, definido como θ tiene dos posibilidades: $\theta = 1$, la mujer tiene el gen y $\theta = 0$, la mujer no tiene el gen. Así, podemos definir la distribución *a priori* como $p(\theta = 1) = p(\theta = 0) = \frac{1}{2}$.

Luego, la información que será utilizada para calcular la distribución *posterior* del estado de la mujer es el estado de sus hijos, en específico, supongamos que tiene dos hijos, y definimos y_i el estado del i -ésimo hijo, por lo que tenemos $y_i = 1$ si el i -ésimo hijo tiene la enfermedad, y $y_i = 0$ si no la tiene. Asumiendo que los hijos no son gemelos idénticos, sus estados son independientes entre sí.

Para actualizar la incertidumbre sobre el estado de la mujer con los estados de sus hijos utilizamos (como ya mencionamos) la regla de Bayes, por lo que necesitamos calcular lo siguiente:

$$\begin{aligned} p(y_1 = 0, y_2 = 0 \mid \theta = 1) &= (0.5)(0.5) = 0.25, \\ p(y_1 = 0, y_2 = 0 \mid \theta = 0) &= (1)(1) = 1. \end{aligned} \tag{6.17}$$

La primera ecuación sigue del hecho de que si la mujer es portadora del gen ($\theta = 1$), entonces cada uno de sus hijos tiene una probabilidad de heredar el gen de 0.5. La segunda ecuación se debe a que si la mujer no es portadora, entonces es un hecho que los hijos no pueden heredar el gen.

Con lo anterior, podemos actualizar nuestras creencias respecto al estado de la mujer utilizando la regla de Bayes calculando lo siguiente:

$$\begin{aligned} p(\theta = 1 \mid y_1 = 0, y_2 = 0) &= \frac{p(y \mid \theta = 1)p(\theta = 1)}{p(y_1 = 0, y_2 = 0 \mid \theta = 1)p(\theta = 1) + p(y_1 = 0, y_2 = 0 \mid \theta = 0)p(\theta = 0)}, \\ &= \frac{(0.25)(0.5)}{(0.25)(0.5) + (1)(0.5)}, \\ &= \frac{0.125}{0.625}, \\ &= 0.2 \end{aligned} \tag{6.18}$$

Como mencionamos, una de las características que hacen a la estadística Bayesiana sumamente útil es la facilidad con la que se puede incorporar nueva información para recalculer una distribución posterior. Por ejemplo, si se supone que la mujer tiene un tercer hijo que tampoco tienen la enfermedad, podemos utilizar la distribución calculada en el paso anterior para recalculer la probabilidad de que la mujer sea portadora del gen como sigue:

$$\begin{aligned}
 p(\theta = 1 \mid y_1, y_2, y_3) &= \frac{p(y_3 = 0 \mid \theta = 1)p(\theta = 1 \mid y_1, y_2)}{p(y_3 = 0 \mid \theta = 1)p(\theta = 1 \mid y_1, y_2) + p(y_3 = 0 \mid \theta = 0)p(\theta = 0 \mid y_1, y_2)}, \\
 &= \frac{p(y_3 = 0 \mid \theta = 1)p(\theta = 1 \mid y_1, y_2)}{p(y_3 = 0 \mid \theta = 1)p(\theta = 1 \mid y_1, y_2) + p(y_3 = 0 \mid \theta = 0)(1 - p(\theta = 1 \mid y_1, y_2))}, \\
 &= \frac{(0.5)(0.2)}{(0.5)(0.2) + (1)(0.8)}, \\
 &= 0.1111
 \end{aligned} \tag{6.19}$$

Que da el mismo resultado que calcular con la distribución *a priori* inicial, tomando en cuenta el resultado del tercer hijo:

$$\begin{aligned}
 p(y_1 = 0, y_2 = 0, y_3 = 0 \mid \theta = 1) &= (0.5)(0.5)(0.5) = 0.125, \\
 p(y_1 = 0, y_2 = 0, y_3 = 0 \mid \theta = 0) &= (1.0)(1.0)(1.0) = 1
 \end{aligned} \tag{6.20}$$

Lo que nos da la siguiente distribución *a posteriori*:

$$\begin{aligned}
 p(\theta = 1 \mid y) &= \frac{p(y \mid \theta = 1)p(\theta = 1)}{p(y \mid \theta = 1)p(\theta = 1) + p(y \mid \theta = 0)p(\theta = 0)} \\
 &= \frac{(0.125)(0.5)}{(0.125)(0.5) + (1)(0.5)}, \\
 &= \frac{0.0625}{0.5625}, \\
 &= 0.1111
 \end{aligned} \tag{6.21}$$

Finalmente, para ver más claramente la forma en la que la incertidumbre se mueve dependiendo de los datos, calculamos la probabilidad de que la mujer sea portadora si el tercer hijo sí tiene la enfermedad:

$$\begin{aligned}
 p(\theta = 1 \mid y_1, y_2, y_3) &= \frac{p(y_3 = 1 \mid \theta = 1)p(\theta = 1 \mid y_1, y_2)}{p(y_3 = 1 \mid \theta = 1)p(\theta = 1 \mid y_1, y_2) + p(y_3 = 1 \mid \theta = 0)(1 - p(\theta = 1 \mid y_1, y_2))}, \\
 &= \frac{(1.0)(0.2)}{(1)(0.2) + 0(0.8)}, \\
 &= 1
 \end{aligned} \tag{6.22}$$

6.1.4. Modelos no paramétricos

Las definiciones anteriores, junto con los ejemplos numéricos, forman y ejemplifican la base del análisis estadístico bayesiano; como pudimos ver, sigue siendo necesario

hacer supuestos de distribución para los fenómenos, los cuales son dependientes a un conjunto (generalmente pequeño) de parámetros sobre los cuales hacemos inferencia.

La utilidad del análisis bayesiano descrito hasta el momento resulta limitada cuando nos encontramos con problemas más complejos. Por ejemplo, retomemos el análisis de conglomerados con datos de balance desarrollada en el capítulo 3. El objetivo es obtener un número k de grupos similares de acuerdo a sus datos de balance, sin embargo, el gran número de variables que se tienen para cada institución puede propiciar una gran variedad de grupos bien diferenciados entre sí; sabemos que la mayoría de los métodos clásicos de conglomerados necesitan como parámetro el número de grupos que *suponemos* existen, no obstante, cuando nuestros datos tienen un alto grado de complejidad, es muy complicado (y aventurado incluso) hacer suposiciones sobre dicho número.

Si nuestra creencia inicial respecto al número de grupos es errónea, el resultado del análisis de conglomerados puede estar altamente sesgado, lo cual puede ser difícil de detectar sin hacer una inspección más profunda de los datos, por lo que tener supuestos realistas y adecuados sobre los parámetros de un modelo es de suma importancia.

Sabemos que uno de los objetivos de la inferencia estadística es lograr ajustar un modelo que tenga el mejor balance entre precisión, parsimonia y facilidad de implementación (Rodríguez and Müller, 2013), y los modelos no paramétricos (una discusión introductoria a los métodos no paramétricos se encuentra en MacEachern (2016)) buscan moverse tomando en cuenta esas tres dimensiones. Así, la posibilidad de relajar los supuestos sobre los parámetros de un modelo con el que se pretende modelar un conjunto de datos resulta ser algo deseable, y es justo esto lo que permiten los modelos no paramétricos bayesianos.

Los modelos no paramétricos fueron desarrollados para eliminar la dependencia que tienen los resultados de los supuestos hechos sobre los parámetros de un modelo, lo que se logra a partir de la definición de modelos con un número “*infinito*” de parámetros. Es claro que lo anterior no quiere decir que tendremos modelos con una infinidad de parámetros, lo cual no solo iría en contra del principio de parsimonia que se tiene en estadística, sino que resulta infactible computacionalmente; más bien, en los métodos bayesianos no paramétricos se permite una gran flexibilidad sobre las suposiciones *a priori* que hacemos sobre los parámetros de un modelo, esperando que después de agregar información al modelo se logren dibujar supuestos más realistas sobre el número y el comportamiento real de los parámetros.

Una de las problemáticas que surgen con los modelos bayesianos está relacionada con la selección del modelo *a priori* para la distribución de los parámetros, ¿cómo saber si se está seleccionando un modelo adecuado, o cómo saber si los resultados de un ajuste son realistas?

En MacEachern (2016) definen lo anterior en términos de **consistencia**. En estadística frecuentista, la consistencia de un modelo se refiere a que, mientras más grande es una muestra aleatoria, el estimador calculado *convergerá* al valor “real” del parámetro (si el modelo supuesto es correcto). Por otro lado, en los métodos bayesianos la consistencia se traduce en vecindarios arbitrariamente pequeños, es decir, la distribu-

ción posterior de un parámetro se concentra alrededor del valor verdadero, lo que se traduce en que la probabilidad de que la distribución posterior se acerque al valor verdadero será uno para muestras suficientemente grandes.

Así, los modelos bayesianos paramétricos tienen una gran limitante: si los supuestos a priori de la distribución de un parámetro **no contienen** el valor “real” del mismo, es imposible que la distribución posterior se concentre alrededor del valor correcto. Así, queda aún más clara la necesidad de los modelos no paramétricos, que tienen como objetivo principal deshacerse de la dependencia de los supuestos a priori.

Dado lo anterior, una de las motivaciones principales para desarrollar la teoría bayesiana no paramétrica fue encontrar distribuciones a priori que tuvieran un soporte mucho más amplio sobre el dominio del (de los) parámetro(s) a estimar y, en caso de ser posible, con soporte sobre **todo** el dominio del parámetro.

Para aclarar lo anterior, el autor (MacEachern, 2016) pone como ejemplo una sucesión de lanzamientos de moneda que resulta en un promedio de 0.7². Si se supone que el valor real del parámetro θ es 0.5, o incluso que se encuentra en el intervalo $[0.4, 0.6]$, entonces las distribuciones posteriores (tomando subconjuntos acumulativos de la muestra) se concentrarán alrededor de 0.7. Por lo tanto, además de la dependencia a la selección de una distribución a priori, las estimaciones tradicionales pueden fallar si la muestra está sesgada.

Así, como ya mencionamos, los métodos bayesianos no paramétricos toman como distribuciones a priori funciones con un espacio paramétrico “infinito”, entendiéndose con esto que el número de parámetros de la distribución a priori crece junto con el tamaño de la muestra. Lo anterior requiere que se definan medidas de probabilidad para múltiples funciones de distribución (es decir, medidas que asignen probabilidades a funciones de distribución completas), contrario al enfoque paramétrico que asigna probabilidades (o que acumula probabilidad) a puntos cercanos al valor real de un parámetro. Dichas medidas son llamadas **Medidas Aleatorias de Probabilidad** o *Random Probability Measures* RPM por sus siglas en inglés) (Ferguson (1973) describe en qué consisten dichas medidas, y propone varias características que deben de cumplirse).

Ahora, presentamos las herramientas utilizadas por los modelos de bloques estocásticos.

6.1.4.1. Distribución Beta

Como mencionamos previamente, la distribución a priori en un análisis bayesiano representa (o contiene) la información que se conoce con anticipación sobre los parámetros de un fenómeno aleatorio, y la distribución Beta resulta útil para codificar dicha información dada su flexibilidad y el hecho de que su dominio está en el intervalo $[0, 1]$, lo que resulta conveniente para acumular probabilidad en puntos de su dominio que pueden ser interpretados a su vez como probabilidades.

Una variable aleatoria X se distribuye Beta con parámetros a y b si su función de

²La asignación de 1 o 0 a los dos resultados del volado no es importante para el ejemplo.

distribución es de la forma:

$$Beta(x | a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}, \quad (6.23)$$

donde $B(a, b)$ es la función Beta, que definimos como:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (6.24)$$

donde a su vez definimos la función Gamma como:

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du. \quad (6.25)$$

Para que la función de distribución beta exista, es necesario tener $a, b > 0$. Al tomar $a = b = 1$ recuperamos la distribución uniforme, pues:

$$\begin{aligned} Beta(x | a = 1, b = 1) &= \frac{1}{B(1, 1)} x^{1-1} (1-x)^{1-1}, \\ &= \frac{1}{B(1, 1)} x^0 (1-x)^0, \\ &= 1 \end{aligned} \quad (6.26)$$

La distribución uniforme resulta conveniente en el caso en el que la información previa sobre un fenómeno es casi nula, pues la mejor suposición que se puede hacer es que todos los posibles valores que toma la distribución tienen la misma probabilidad.

Por último, la distribución beta tiene media y varianza con la siguiente forma:

$$media = \frac{a}{a+b}, \quad (6.27) \quad varianza = \frac{ab}{(a+b)^2(a+b+1)}. \quad (6.28)$$

6.1.4.2. Distribución de Dirichlet

La distribución de Dirichlet es ampliamente usada en los métodos bayesianos no paramétricos, pues resulta ser una generalización multidimensional de la distribución Beta, que como vimos, su flexibilidad ayuda a hacer mejores suposiciones para la distribución a priori del modelo.

Al igual que la distribución Beta, la distribución Dirichlet tiene soporte en el “rango” $[0, 1]$, solo que en este caso, el soporte resulta ser el “simplex de probabilidad”³:

$$S_K = \{\mathbf{x} : 0 \leq x_k \leq 1, \sum_{k=1}^K x_k = 1\}. \quad (6.29)$$

³Esto es, todos los puntos en un hiperplano tal que la suma de sus entradas es uno.

La función de distribución Dirichlet tiene la siguiente forma:

$$Dir(x | \alpha) = \frac{1}{BG(\alpha)} \prod_{k=1}^K x_k^{\alpha_k - 1} \mathbb{1}(x \in S_k), \quad (6.30)$$

donde $BG(\alpha)$ es la generalización de la función Beta para K variables, definida como:

$$BG(\alpha) = \frac{\prod_{k=1}^K \Gamma(\alpha_k)}{\Gamma(\alpha_0)}, \quad (6.31)$$

donde $\alpha_0 = \sum_{k=1}^K \alpha_k$ y $\alpha_k \geq 0 \forall k \in \{1, 2, \dots, k\}$.

Para tener una imagen más clara del papel que juegan los parámetros de la distribución en su comportamiento, resulta útil reescribirla como:

$$Dir(x | \alpha g_{0_1}, \dots, \alpha g_{0_K}) = \frac{\Gamma(\alpha)}{\prod_{k=1}^K \Gamma(\alpha g_{0_k})} \prod_{k=1}^K x_k^{\alpha g_{0_k} - 1} \mathbb{1}(x \in S_k), \quad (6.32)$$

donde definimos $g_{0_i} = \frac{\alpha_i}{\sum_i \alpha_i}$. Así, es claro que α_{0_i} sirve como valor a priori para el i -ésimo parámetro de una distribución multinomial, y α sirve para controlar qué tanto se alejan otros puntos del vector dado por $g_0 = (g_{0_1}, \dots, g_{0_K})$.

Notamos que al tomar $K = 2$, obtenemos la distribución $Beta(\alpha_1, \alpha_2)$.

Esta distribución corresponde a la distribución a priori *conjugada*⁴ de los parámetros de una distribución multinomial. Tiene además las siguientes características:

$$\mathbb{E}[x_k] = \frac{\alpha_k}{\alpha_0}, \quad (6.33) \quad Var[\alpha_k] = \frac{\alpha_k(\alpha_0 - \alpha_k)}{\alpha_0^2(\alpha_0 + 1)}. \quad (6.34)$$

Como recordamos, la distribución Beta toma valores entre 0 y 1, por lo que resulta útil como función a priori, pues al incorporar información para el cálculo de la distribución posterior, ésta “concentra” la masa de la distribución en el rango de valores que con mayor probabilidad contiene al valor real del parámetro que se quiere estimar, por lo que se puede pensar a la distribución beta como una “distribución sobre distribuciones”, pues se ajustará a la distribución real del parámetro a estimar; análogamente, la distribución de Dirichlet se ajusta a distribuciones más generales, propiedad que se hace más evidente al ver que, por ejemplo, la distribución marginal de cada x_k en $x \in S_k$ sigue una distribución $Beta(\alpha_k, \alpha - \alpha_k)$.

6.1.5. Muestreo de Gibbs

Una de las últimas herramientas que se utilizan en el ajuste del Modelo de Bloques Estocásticos es el Muestreo de Gibbs (mencionado por primera vez en [Geman and Geman \(1984\)](#)), por lo que daremos la intuición general de dicho algoritmo para entender

⁴En estadística Bayesiana, una distribución “ F ” es conjugada de una distribución “ G ” si “ F ”, como distribución a priori de los parámetros de “ G ”, conserva su función de distribución (salvo cambios en el valor de sus parámetros) al pasar a la distribución posterior.

de mejor forma cómo se encuentran las comunidades en la redes que nos ocupan. El muestreo de Gibbs es un caso particular de los métodos conocidos como *Markov Chain Monte Carlo* (MCMC, [Robert and Casella \(2011\)](#) hacen un recuento de los inicios y las distintas etapas en el desarrollo de los métodos MCMC.), que en nuestro contexto buscan estimar valores de parámetros desconocidos a partir de la convergencia de una cadena de Markov a su distribución de equilibrio.

La utilidad del muestreo de Gibbs en el problema que nos ocupa es la flexibilidad que nos da para obtener una muestra de la distribución conjunta de dos o más variables aleatorias (es decir, se busca obtener una distribución multivariada) a partir del supuesto de que se conocen las distribuciones marginales (condicionales) de cada variable aleatoria respecto al resto. Nos apoyamos en [Rossi et al. \(2012\)](#) para las siguientes definiciones.

Sabemos que una cadena de Markov nos ayuda a construir una sucesión de variables aleatorias $\theta_1, \theta_2, \dots, \theta_r, \dots$, partiendo de un valor inicial θ_0 . La manera en que “recorremos” la cadena está determinada por las probabilidades de transición p_{ij} , acomodadas en una matriz estocástica⁵ P , donde la entrada p_{ij} representa la probabilidad de que la matriz cambie del estado θ_i al estado θ_j en la siguiente unidad de tiempo.

Lo anterior quiere decir que la cadena evoluciona de acuerdo a las distribuciones condicionales $\theta_{r+1} | \theta_r \sim F(\theta_r)$. En este caso, la función F está determinada por la matriz de transición P , y con ésta se pueden simular observaciones de una cadena de Markov partiendo de un punto (estado) inicial θ_0 . Con los supuestos anteriores, el muestreo de Gibbs se lleva a cabo con los siguientes pasos iterativos:

1. Iniciamos el algoritmo con un punto inicial θ_0 .
2. Obtenemos $\theta_1 \sim F(\theta_0)$.
3. Reemplazamos θ_1 en $F(\theta_0)$ y reemplazamos R veces.

Bajo ciertas condiciones sobre F , al tomar $R \rightarrow \infty$ la cadena de Markov converge a su distribución estacionaria, y así obtenemos una muestra de la distribución F , distribución de una cadena de Markov.

Como mencionamos arriba, el muestreo de Gibbs nos ayuda a obtener muestras de variables multivariadas, así, para hacer más claro el algoritmo del muestreo de Gibbs, nos enfocaremos en el caso de dos variables aleatorias, es decir, en un vector aleatorio (X, Y) , a cuya distribución le llamaremos $\pi(X, Y)$.

Para obtener los puntos bidimensionales de interés (x_i, y_i) damos por hecho que conocemos las distribuciones condicionales $\pi_{X|Y}$ y $\pi_{Y|X}$, y procedemos de la siguiente forma:

1. Iniciamos la cadena con un punto x_0 .
2. Para cada “tiempo” de nuestra simulación t obtenemos:

⁵Una matriz es estocástica si sus entradas cumplen $0 \leq p_{ij} \leq 1$, y la suma por renglón (columna) suma uno. Si la suma por columna y por renglón es uno, la matriz es doblemente estocástica.

$$a) y_t = \pi_{Y|X}(y | x_{t-1}),$$

$$b) x_t = \pi_{X|Y}(x | y_t).$$

3. Repetimos el número de veces necesarias.

Del procedimiento anterior, formamos las duplas (x_t, y_t) , donde cada par representa un punto obtenido de la distribución conjunta $\pi(X, Y)$.

Como veremos, para los modelos de bloques estocásticos se tienen que resolver problemas de inferencia mucho más complejos que involucran todas las herramientas definidas hasta esta sección, y se resuelven utilizando muestreo de Gibbs para obtener valores de distribuciones con un conjunto grande de parámetros.

6.2. Modelos de Bloques Estocásticos

Una vez que tenemos las herramientas necesarias para describir el procedimiento general de los modelos de bloque estocásticos (como los presentan inicialmente en [Nowicki and Snijders \(2001\)](#)), pasamos a definir las generalidades de dicho método.

Existen múltiples métodos para la detección de comunidades, donde la mayoría basan sus resultados en una noción de *distancia*, la cual puede ser sumamente arbitraria⁶, y puede fallar al momento de capturar el patrón de conectividad de un sistema complejo. Lo anterior provoca que muchos de los modelos de detección de comunidades en redes sean propensos a sufrir de sobreajuste, pues dadas sus limitaciones para considerar elementos importantes de una red (por ejemplo, características de los nodos, o pesos de las conexiones) pueden terminar afectados por fluctuaciones aleatorias en la conectividad de un sistema.

Por lo anterior, se creó una nueva corriente de métodos para la detección de comunidades, en ésta, decidieron atacar el problema generado por las fluctuaciones aleatorias (naturales en todo sistema real) con teoría de probabilidad, proponiendo modelos que traducen la idea de equivalencia estructural a un marco probabilístico. Aquí, la *equivalencia estructural* se refiere a la similitud que comparten los participantes de un sistema en términos de conectividad, lo que provoca que el algoritmo los identifique dentro de un mismo grupo.

Todos los métodos que utilizan la probabilidad para la detección de comunidades en una red utilizan como supuesto la existencia de un conjunto de n nodos, para los cuales es posible observar una variable y_{ij} llamada la “relación de i a j ”.

El enfoque más exitoso para modelar las relaciones existentes en una red ha sido suponer que dichas relaciones pueden ser modeladas condicionando su probabilidad de existencia a un vector \mathbf{x} de atributos de los nodos, en nuestro caso, la entrada x_i ,

⁶En el capítulo 3 se introducen los métodos generales para identificar conglomerados, y se brinda una discusión un poco más detallada de cómo el concepto de distancia puede tener efectos no deseados en los resultados de un análisis.

$i \in \{1, 2, \dots, n\}$ representa la comunidad a la que pertenece el nodo i . Así, es posible modelar la relación

$$\mathbf{y} = (y_{ij})_{1 \leq i \neq j \leq n}, \quad (6.35)$$

condicionado al vector $\mathbf{x} = (x_1, \dots, x_n)$, donde x_i pertenece al conjunto $\mathcal{C} = \{1, 2, \dots, c\}$, que contiene todas las clases (grupos) posibles a las que pueden pertenecer nuestros nodos, lo anterior es equivalente a pensar que la probabilidad de que dos nodos i y j formen una conexión en una red depende completamente de la comunidad a la que pertenecen. En el caso en el que las clases en \mathcal{C} son conocidas, y se sabe a cuál pertenece cada nodo, el enfoque anterior se conoce como *modelaje de bloques a priori*.

En [Holland et al. \(1983\)](#) los autores generalizan el marco teórico anterior para gráficas dirigidas, asumiendo que las relaciones y_{ij} y y_{ji} no son necesariamente iguales, y además son independientes, condicionando de igual forma al vector de clasificación \mathbf{x} . A dicho modelo se le conoce como modelaje de bloques dependiente-a-pares (pair-dependant block-modeling), en éste, se dice que dos nodos son *equivalentes estocásticamente* si su probabilidad de conexión es igual con nodos de su mismo grupo, y con nodos de otro grupo, lo que se cumple si pertenecen al mismo grupo.

En [Snijders and Nowicki \(1997\)](#) se propone un método de modelaje a posteriori para gráficas no dirigidas, el cual sólo permite la existencia de dos clases, es decir $\mathcal{C} = \{0, 1\}$ ⁷. Sin embargo, dadas las limitaciones que imponen los supuestos anteriores, en [Nowicki and Snijders \(2001\)](#) se generaliza el mismo marco teórico para considerar gráficas dirigidas y con un número arbitrario de comunidades, en el cual nos basamos para las siguientes definiciones. Los autores mencionan que el método es básicamente una mezcla de distribuciones, ya que la comunidad a la que pertenece cada nodo es el resultado de una variable aleatoria X_i .

Como ya mencionamos, el objetivo de la extensión que dan en [Nowicki and Snijders \(2001\)](#) al trabajo seminal en [Snijders and Nowicki \(1997\)](#) es el de eliminar varios supuestos que limitan al modelo. Dado que el trabajo posterior cubre gráficas dirigidas (para las cuales no es necesario que $y_{ij} = y_{ji}$), la “unidad” de trabajo es el par (y_{ij}, y_{ji}) .

En nuestro caso, no es posible el análisis a priori, pues no conocemos la estructura de agrupación de nuestros bancos, por esto, es necesario un enfoque que encuentre dicha agrupación *a posteriori* utilizando únicamente las relaciones \mathbf{y} , las cuales conocemos gracias a las redes de exposición definidas en el capítulo anterior. Dado que en nuestro caso sólo existen relaciones entre los bancos de acuerdo con las matrices de exposición, y dado que el método utilizado trabaja con matrices dirigidas sin pesos, simplificaremos la notación original utilizada por [Nowicki and Snijders \(2001\)](#).

De manera resumida, el modelo hace los siguientes supuestos:

- N denota el conjunto de todos los pares de nodos (i, j) , $i, j \in \{1, \dots, n\}$.
- Cada nodo i pertenece a una clase k , $k \in \{1, \dots, c\}$.

⁷El uso de 0 y 1 para identificar los grupos es arbitrario, pues en el trabajo se menciona que la etiqueta de identificación de las comunidades es irrelevante para la estimación de las mismas.

- Sólo observamos la matriz de relaciones \mathbf{Y} y queremos estimar el vector de grupos \mathbf{X} .

Lo primero que se define es la probabilidad conjunta del grupo de cada uno de los vértices, que son resultado de variables aleatorias X_i . Así, tenemos que:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \theta_1^{m_1} \dots \theta_c^{m_c}, \quad (6.36)$$

donde m_i representa el número de nodos pertenecientes a la clase i , y θ_i representa la probabilidad de que un nodo pertenezca a la clase i .

Como ya mencionamos, uno de los supuestos fuertes es que la existencia de una conexión entre dos nodos depende de la clase a la que pertenecen, así, dado el vector $\mathbf{X} = \mathbf{x}$ tenemos que:

$$P(\mathbf{Y}_{ij} = 1 \mid \mathbf{X} = \mathbf{x}) = \eta(x_i, x_j), \quad (6.37)$$

donde $\eta(k, h)$ es la probabilidad de conexión que depende de la agrupación de i y j , y que cumple:

$$\sum_{k, h \in \mathcal{C}} \eta(k, h) = 1. \quad (6.38)$$

Más aún, la probabilidad $\eta(k, h)$ se puede partir en $\eta(k, k)$ y $\eta(k, h)$, que representan la probabilidad de conexión entre dos nodos del mismo grupo, y entre nodos de grupos distintos, respectivamente. Así, tomando en cuenta la partición anterior podemos escribir:

$$P(y \mid x, \theta, \eta) = \left(\prod_{1 \leq k < h \leq c} (\eta(k, h))^{e(k, h)} \right) \times \left(\prod_{k=1}^c (\eta(k, k))^{e(k, k)} \right), \quad (6.39)$$

donde $e(k, h)$ es el número de relaciones existentes entre los bloques k y h , que se calcula como:

$$e(k, h) = \sum_{(i, j) \in N} \mathbb{1}\{y_{ij} = 1\} \mathbb{1}\{x_i = k\} \mathbb{1}\{x_j = h\}. \quad (6.40)$$

Así, el modelo que nos interesa está dado por la distribución conjunta de las relaciones y de la pertenencia de los nodos a un grupos, es decir, la distribución conjunta de (\mathbf{Y}, \mathbf{X}) , que está dada por:

$$\begin{aligned} P(y, x \mid \theta, \eta) &= \theta_1^{m_1} \dots \theta_c^{m_c}, \\ &\times \left(\prod_{1 \leq k < h \leq c} (\eta(k, h))^{e(k, h)} \right), \\ &\times \left(\prod_{k=1}^c (\eta(k, k))^{e(k, k)} \right). \end{aligned} \quad (6.41)$$

Los tres componentes de la ecuación anterior representan, respectivamente: la distribución del grupo de los nodos, la de las conexiones entre nodos de clases distintas y

la de las conexiones entre nodos del mismo grupo. Los parámetros son estimados utilizando métodos Bayesianos, y para cada uno de los componentes descritos se supone que la distribución a priori es el producto de distribuciones Dirichlet independientes.

Dado que sólo la estructura de \mathbf{Y} es conocida, reescribimos la probabilidad anterior como

$$P(y | \theta, \eta) = \sum_{x \in \mathbb{C}^n} P(y, x | \theta, \eta), \quad (6.42)$$

es decir, sumamos sobre todas las posibles combinaciones de grupos para los n nodos de nuestra red. Para aplicar la ecuación anterior es necesario estimar los parámetros θ y η utilizando la distribución posterior

$$f(\theta, \eta | y) = \sum_x f(\theta, \eta, x | y), \quad (6.43)$$

con los cuales se puede estimar el vector de clases \mathbf{x} utilizando la densidad posterior predictiva

$$P(x | y) = \int f(\theta, \eta, x | y) d\eta d\theta. \quad (6.44)$$

El muestreo de Gibbs es aplicado para obtener la probabilidad condicional

$$f(\theta, \eta, x | y). \quad (6.45)$$

a partir de la cual hacemos inferencia para las dos ecuaciones anteriores. Para mayores detalles de cómo es aplicado dicho método para encontrar las comunidades en las redes de interés, en [Nowicki and Snijders \(2001\)](#) presentan de manera detallada el procedimiento a seguir.

6.2.1. Problemas del Modelo

Como mencionamos, la distribución del grado es una de las características más importantes de una red. En general, las redes reales son significativamente grandes, incluso las financieras (el tamaño de las redes que estamos utilizando es pequeño, pues por construcción sólo estamos considerando las instituciones bancarias del país), lo que provoca que en general existan nodos con un grado por encima del promedio, dando pie a la existencia de una distribución del grado mucho más variable, y con colas más pesadas.

Relacionado con lo anterior, el modelo de detección de comunidades presentado resulta sensible a una distribución del grado sumamente heterogénea, pues se ha demostrado que la inferencia hecha con el modelo presenta sesgos en favor de nodos altamente conectados. Por lo anterior, [Karrer and Newman \(2011\)](#) presentan una extensión del modelo de [Nowicki and Snijders \(2001\)](#), en el que el grado de los nodos es incluido en las distribuciones sobre las que se aplican los métodos de inferencia bayesianos que ya discutimos.

Adicionalmente, nuestro modelo no considera los pesos de las conexiones entre las instituciones bancarias. El peso cobra gran importancia en el contexto de nuestras

redes, ya que puede además ayudar a revelar dependencia o patrones no triviales de conectividad con la capacidad de aumentar el riesgo sistémico si, por ejemplo, el peso de las ligas fuera el porcentaje del capital que representa una conexión para una institución que está enviando dinero (o lo que es lo mismo, una institución que se está exponiendo).

6.2.2. Estimación de SBM con la librería Graph-Tool de Python

En el presente trabajo se utilizó la paquetería para análisis de gráficas (redes) *Graph-Tool* (Peixoto, 2014). En Peixoto (2017) se presenta un excelente repaso de los principios de los modelos de bloques estocásticos, y se explica a detalle la metodología que utiliza el paquete de Python ya mencionado para encontrar las comunidades que forman una red.

6.3. Resultados

6.3.1. Distribución del grado

Por el problema descrito en la sección anterior, obtuvimos la visualización de la distribución del grado interior y exterior, y la comparamos con las respectivas distribuciones Ley potencia y Poisson.⁸

La forma en la que obtuvimos los datos para la comparación de las distribuciones sigue la misma lógica con la que obtuvimos la gráfica de comparación entre la distribución Poisson y la distribución Power Law (ver 5.2). De acuerdo con la distribución de grado observada⁹, obtuvimos dos parámetros: el promedio del grado que se utiliza como parámetro λ para simular valores Poisson, y el parámetro α , para simular observaciones de una distribución Ley potencia¹⁰. Como podemos observar, la distribución empírica contiene nodos con grado mayor a los que se obtienen con la correspondiente distribución Poisson, sin embargo, se queda muy corta en comparación a los nodos más conectados obtenidos con la correspondiente distribución Ley potencia.

⁸Fórmulas 5.23, 5.31 y 5.37

⁹Es importante mencionar que la gráfica se obtuvo utilizando las distribuciones de todo el periodo de estudio como una sola.

¹⁰Para la estimación de α y la simulación de las observaciones de la distribución Ley potencia se utilizó la biblioteca `powerLaw` (Gillespie (2015))

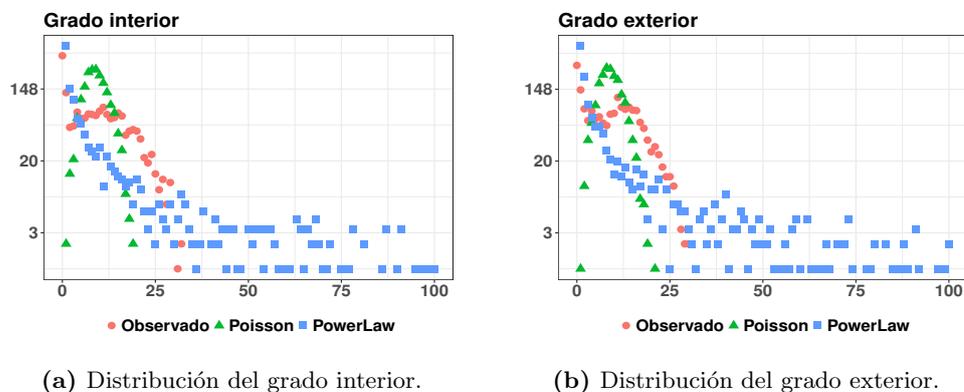


Figura 6.1: Distribución del grado para la red interbancaria Mexicana.

Por lo anterior, se decidió utilizar Modelos de Bloques Estocásticos sin corrección de grado, pues la distribución empírica de nuestra red no presenta el nivel de heterogeneidad necesario para la corrección por grado. Más aún, los modelos con corrección por grado fueron ajustados, pero el efecto de tener una distribución más cercana a la aleatoria que a la ley potencia es claro, pues en todos los ajustes (es decir, para todas las fechas), todas las instituciones son asignadas a un único grupo.

6.3.2. Movimientos de las agrupaciones en el tiempo

Para visualizar la estabilidad de las agrupaciones obtenidas con el modelo de bloques estocásticos utilizamos de nuevo un diagrama aluvial similar a los que presentamos en los resultados del capítulo 3. En el caso de los SBM, para cada fecha sólo tenemos un ajuste, por lo que tenemos una sola visualización (Figura 6.2).

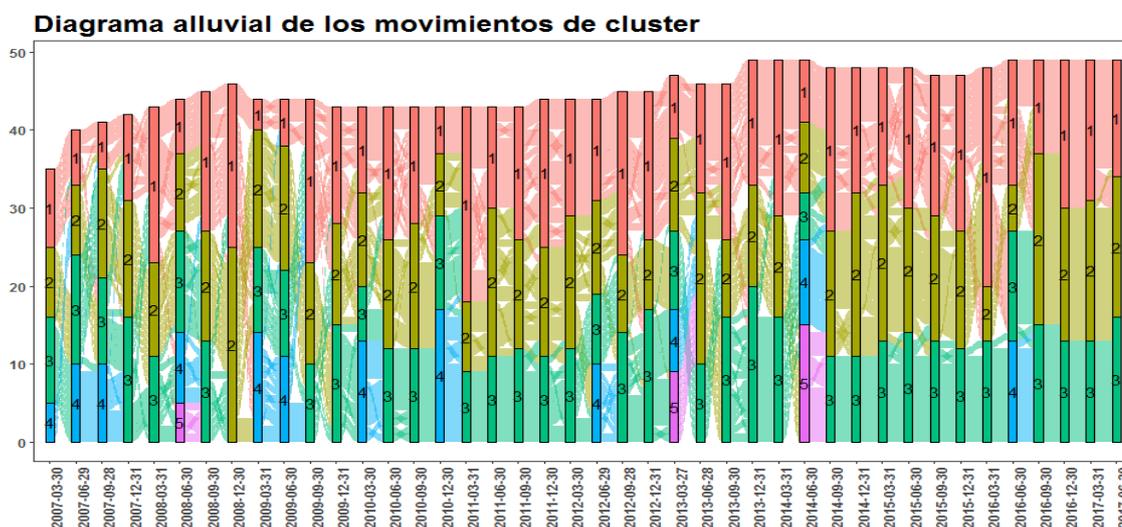


Figura 6.2: Gráfica aluvial de los movimientos entre los grupos encontrados con el Modelo de Bloques Estocásticos de las instituciones bancarias en México.

Al inspeccionar la figura anterior, vemos que los movimientos de los conglomerados son ligeramente más frecuentes que los movimientos en las agrupaciones de los ajustes con FKM, excepto al final del periodo, donde los grupos son bastante constantes desde septiembre de 2016 a junio de 2017. Sin embargo, hay que notar que el número de grupos se mantiene estable a lo largo del periodo de estudio.

Es importante notar que en las fechas cercanas a la crisis financiera se observa el que podría ser un efecto contrario al que se tiene en los grupos por RKM ¹¹, esto es, el número de grupos encontrados disminuye; si trasladamos el significado de los grupos a los encontrados con el Modelo de Bloques Estocásticos, tenemos evidencia de que las crisis financieras no sólo afectan la estructura general de la red interbancaria (Martinez-Jaramillo et al. (2014)), si no que también afectan la estructura de agrupación que siguen los bancos, materializándose como aversión al riesgo, lo que provoca que los bancos disminuyan significativamente su actividad en la red, generando así un sistema menos denso, que podemos observar en diciembre de 2008, donde pasamos de tener 5 grupos justo antes del punto más álgido de la crisis, a tener sólo dos grupos.

6.3.3. Distribución de las medidas de centralidad por grupo

Análogo al análisis de la distribución de las cuentas de balance por grupo, hacemos una breve exploración del comportamiento de la centralidad de los bancos de acuerdo a

¹¹Vimos que alrededor de las fechas con mayor volatilidad e incertidumbre los bancos tienden a separarse en un mayor número de grupos, lo que sugiere que las instituciones bancarias recurren a una mayor diversificación en épocas de crisis.

su agrupación. A pesar de que no tenemos una gran cantidad de medidas de centralidad para que sea necesario seleccionar las más importantes para la agrupación, también obtuvimos la importancia de las medidas con un bosque aleatorio para complementar el análisis de qué influye en la agrupación obtenida por el modelo de bloques estocásticos.

Podemos ver la importancia de las medidas de centralidad (Figura 6.3). Dada la construcción del método, no es sorprendente que el grado sea la medida más influyente para llegar a la agrupación obtenida. La cercanía es la segunda medida más importante, lo que complementa de buena forma al grado, pues los patrones de conectividad que buscan los modelos de bloques estocásticos toman en cuenta la conectividad intra e inter grupo.

Resulta interesante que la fuerza de los nodos no figure dentro de las medidas más importantes, pues sabemos que en la red que estamos utilizando (exposición) el peso de las ligas puede ser un determinante importante para la propagación de contagio; una mejora necesaria al análisis aquí propuesto sería tomar en cuenta el peso de las ligas en los SBM.

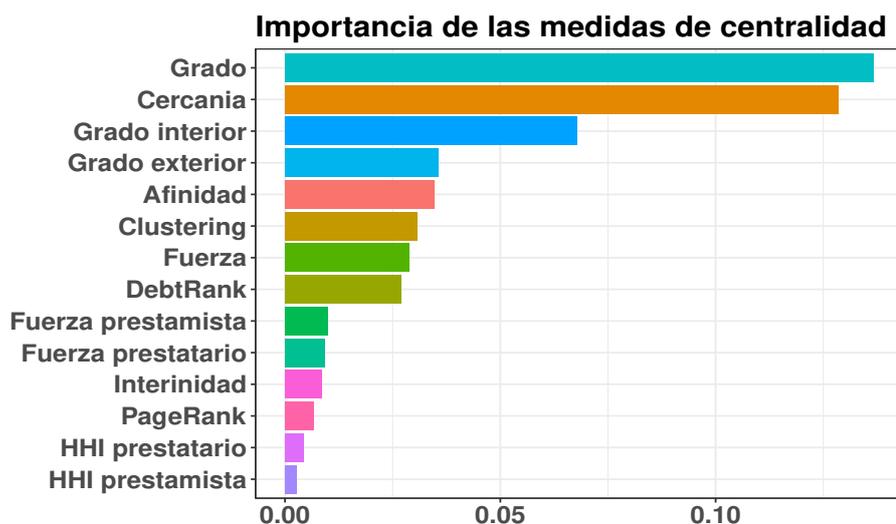


Figura 6.3: Importancia de las medidas de centralidad para llegar a la agrupación obtenida por SBM según el método de bosque aleatorio.

Al igual que en la distribución de las cuentas de balance, complementando la figura anterior con la información de la distribución por grupo de la fuerza y el grado (Figura 6.4), es claro porqué el grado es la medida más importante para la obtención de la agrupación dada, pues la distribución de dicha medida está claramente diferenciada para cada grupo, siendo el clúster 1 el que contiene a los nodos más conectados, y el grupo 3 el que contiene los nodos con menor cantidad de conexiones; más aún, el grupo 1 también es el que se encuentra más “cerca” del resto de la red.

Resulta relevante también la afinidad de los nodos. Como podemos observar, el tercer grupo tiene una distribución altamente heterogénea, evidenciando así que los

bancos pertenecientes a dicho grupo forman conexiones con nodos de todos los niveles de conectividad. Además, resulta significativo que el segundo grupo sea el que tiene la afinidad más alta, pues esto da la señal de que el grupo 3 forma relaciones con nodos más diversos en términos de conectividad, evidenciando así que el modelo de Centro-Periferia de [Craig and von Peter \(2014\)](#) se queda corto respecto a las redes financieras reales, pues el supuesto de que los nodos de la periferia no pueden formar conexiones entre sí no es fiel a la realidad, como se demuestra en las dos figuras mencionadas arriba, hipótesis que se hace más fuerte al ver la tendencia creciente del error del modelo (Figura 5.3). A pesar de que algunas hipótesis son soportadas por figuras del capítulo anterior, el paradigma de tener una red partida en sólo dos grupos se ve rebasado al analizar a profundidad las características de la red interbancaria, por lo que una generalización del modelo Centro-Periferia parece realista.

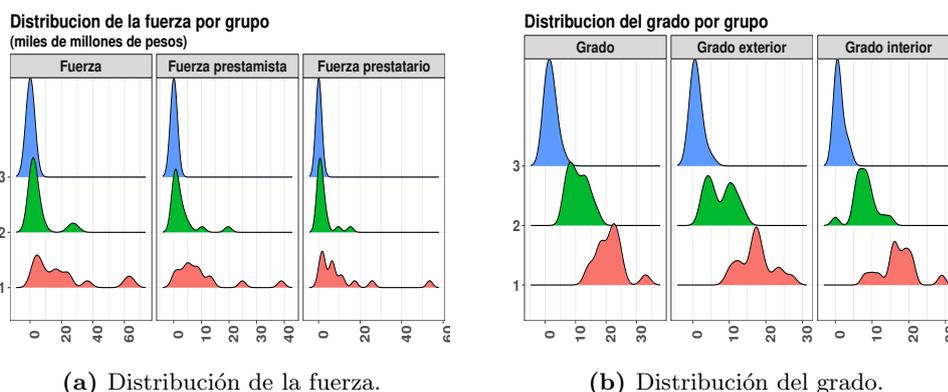


Figura 6.4: Distribución del grado y la fuerza del sistema bancario diferenciando por los grupos obtenidos con el método de bosque aleatorio.

Para el resto de las medidas de centralidad (Figura 6.5) el DebtRank resulta interesante, pues a pesar de que la media de la distribución por grupos no es tan distante, es claro que hay muchos valores extremos en el grupo 1, es decir, la mayoría de los bancos sistémicos en México se encuentran en el mismo grupo, lo que implica que tienen patrones de conectividad similares entre sí, y con los otros dos grupos. Una vez más, la sincronización en la conectividad de los bancos sistémicos puede potenciar los efectos de un choque, por lo que la regulación diferenciada resulta de nuevo relevante.

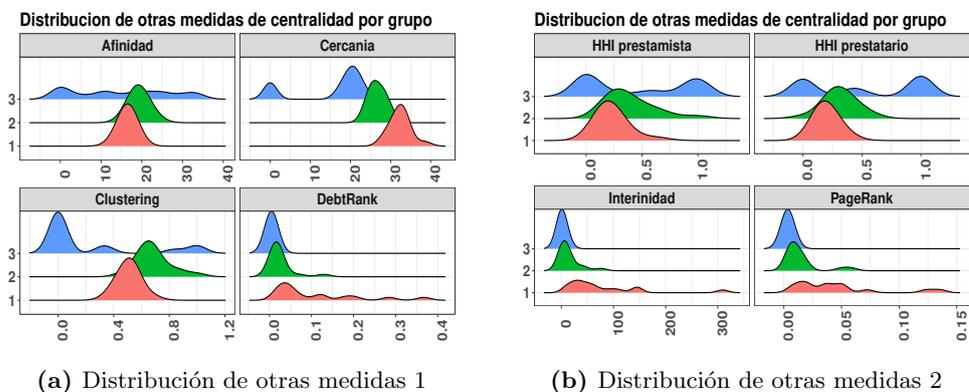


Figura 6.5: Distribución de otras medidas de centralidad del sistema bancario diferenciando por los grupos obtenidos con el método de bosque aleatorio

Con las distribuciones presentadas en este capítulo complementando las gráficas de la centralidad de los bancos del capítulo anterior (Figuras 5.7 a 5.10), queda claro que la agrupación de los bancos en la red interbancaria también está sujeta a determinantes dinámicos, y al ambiente económico-financiero, lo que hace difícil proponer una agrupación que capture el aspecto dinámico de la evolución de la red, y proponer un método para lograr una clasificación dinámica está fuera del alcance del presente trabajo. Así, nos limitamos a hacer un análisis estático de la agrupación obtenida por ambos métodos (RKM y SBM) en la última fecha del periodo de estudio, mismo que realizamos como parte de las conclusiones en el siguiente capítulo.

Discusión de resultados y conclusiones

7.1. Discusión

A lo largo de los capítulos anteriores hemos presentando los resultados de los análisis que se llevaron a cabo, sin embargo, no hemos sido concluyentes respecto a la mejor agrupación para los bancos Mexicanos de acuerdo con los resultados de los métodos aplicados a nuestros datos. Lo anterior se debe a dos resultados importantes de nuestro análisis:

1. A pesar de que los modelos de negocios son relativamente estables (como podemos ver en las gráficas aluviales 3.7 a 3.9, 3.13 a 3.15 y 6.2) y son eventos coyunturales los que propician cambios fuertes, los modelos de negocios son dinámicos, y son propensos a cambiar de acuerdo al entorno económico-financiero en el que se desenvuelven las instituciones bancarias de un país.
2. Los modelos de negocios son sensibles al conjunto de datos utilizado, mientras más específica es la información, más propensa es la existencia de un número mayor de grupos. Más aún, el conjunto de datos puede hacer necesaria la implementación de un método distinto, como vimos en el capítulo 3 respecto a los métodos FKM y RKM.
3. Los métodos de conglomerados y de redes utilizados para identificar grupos contienen de forma inherente un componente aleatorio con el potencial de mover los resultados, por lo que es importante obtener un gran número de simulaciones.

Así, en el presente capítulo caracterizamos de manera más clara el mejor grupo para cada uno de los dos métodos usados: análisis de conglomerados y modelos de bloques estocásticos; también hacemos un análisis comparativo de las diferencias entre los grupos encontrados por ambos métodos, con la intención de verificar si los bancos cambian la forma en la que se comportan al tomar en cuenta su interconexión con el resto de las instituciones.

7.1.1. Agrupación con RKM

En los capítulos 3 y 4 vimos los resultados de los ajustes de conglomerados, y la importancia de las cuentas de balance para la estructura de agrupación misma. Ahora, dadas las observaciones hechas en dichos capítulos, seleccionaremos la agrupación más adecuada. Los parámetros que tomaremos en cuenta para seleccionar la mejor partición son las métricas de desempeño (de acuerdo al criterio de K-Medias) arrojadas por la función de RKM y el nivel de diferenciación de acuerdo a la distribución de las cuentas más importantes. Recordemos que dada la complejidad que agrega el querer hacer un análisis dinámico, nos enfocaremos en el ajuste de la última fecha, es decir, junio de 2017.

En el capítulo 3 vimos que los valores del criterio de K-Medias devuelto por la función usada para los ajustes del método FKM son significativamente más pequeños que los del método RKM (Figuras 3.2 y 3.3), sin embargo, al verificar la distribución espacial de los bancos (figuras 3.4 al 3.6) el método identifica a todos los bancos de un grupo en el mismo punto, haciéndolos idénticos en el sub espacio generado por el método, lo cual no es factible, pues a pesar de que las instituciones puedan tener comportamientos parecidos en su balance, es sumamente complicado que sean idénticas.

Por lo anterior, los ajustes obtenidos con el método RKM son preferidos, más aún, a pesar de que los ajustes sin columnas muy correlacionadas tienen un poco más de variabilidad en su criterio de ajuste, tienen valores consistentemente más bajos que los ajustes con todas las variables.

Adicional a lo anterior, observamos que la dispersión de las instituciones en el sub espacio generado por el método es mucho más realista en términos de distancia entre los balances de los bancos (Figuras 3.10 a 3.12).

Si ahora consideramos el análisis de las cuentas más importantes arrojadas por el método de Bosque Aleatorio realizado en el capítulo 4, como ya vimos, las cuentas de capitalización de los bancos juegan un papel importante para la agrupación, incluso históricamente (Figuras 4.4 a 4.6). Por lo anterior, serán el resto de las cuentas las que determinen con mayor certeza los grupos. Sin embargo, es importante notar que los bancos del grupo uno tienen los valores más altos en las cuentas importantes de capitalización, si se toman en cuenta las observaciones hechas en los capítulos 1 y 2 (sobre las modificaciones al capital regulatorio a todo el sistema, en particular a los bancos de importancia sistémica), esto puede ser el resultado de las nuevas regulaciones respecto a los cambios en las ponderaciones de los activos en riesgo, así como de la importancia de los bancos para el sistema.

Viendo los conjuntos de variables relevantes para cada nivel de granularidad es notable que en todos están incluidas variables que, dado su significado, resultan importantes para la determinación de las características de un modelo de negocios, es por esto que incluso la decisión de tomar un nivel de granularidad puede ajustarse al objetivo que se tenga, así, si se desea que la agrupación esté determinada por la mayor cantidad de información posible, tomamos el tercer nivel de granularidad de los datos, y el primer nivel si, por el contrario, queremos que los grupos estén determinados únicamente por

cuentas generales.

Así, tomando en cuenta la estabilidad tanto del criterio de ajuste como de las agrupaciones (Figuras 3.3 y 3.13 respectivamente), especialmente para las últimas fechas del periodo de estudio, **el ajuste seleccionado es el correspondiente al primer nivel de granularidad.**

Otro criterio para justificar la selección de dicho ajuste corresponde a la importancia histórica de las cuentas de balance (Figura 4.4), donde los valores son más altos que en las siguientes dos Figuras (4.5 y 4.6), es decir, en general las variables tienen un nivel de importancia mayor, y la importancia de cuentas con un significado directo para modelos de negocios tiene saltos importantes a lo largo del tiempo, por ejemplo, la importancia de la captación tradicional, de la cartera de crédito y de las inversiones en valores.

Lo anterior se puede explicar si tomamos en cuenta que para los niveles dos y tres son más las cuentas existentes, por lo que la importancia se divide entre un mayor número de variables, sin embargo, en cualquier modelo estadístico siempre resulta deseable mantener la complejidad de un modelo al mínimo, por lo que lo anterior no es algo necesariamente negativo.

Para describir los grupos del ajuste seleccionado, nos concentramos en la distribución de las cuentas más importantes según el bosque aleatorio (Figura 4.7). Como ya mencionamos, las cuentas de capital son las más relevantes para la agrupación, así, el capital contribuido representa el porcentaje más grande del capital contable para los bancos del grupo 3, seguidos por los bancos del grupo 2 y el grupo 1 con los porcentajes más pequeños.

Respecto al capital ganado, resulta relevante que las instituciones del grupo 1 tienen valores positivos, lo que quiere decir que son bancos con resultados positivos consistentemente, los bancos del grupo 2 tienen valores positivos, pero dicha cuenta es menos importante para el capital total. Por otro lado, los bancos del grupo 3 tienen valores negativos, lo que podría tener dos significados:

1. Son bancos con resultados poco favorables, lo que los obliga a reportar valores negativos en la contribución al capital derivada de pérdidas.
2. Son bancos con pérdidas registradas producto de, por ejemplo, un aumento en los gastos de operación derivado de un proceso de inversión.

La tercera y la cuarta cuenta más importante dejan ver que los bancos del grupo 1 dan una mayor importancia a los préstamos interbancarios y a los créditos (en general) para la conformación de sus activos. Le siguen los bancos del grupo 2 y los bancos del grupo 3, que tienen los valores más pequeños para estas dos cuentas.

Por último, complementando lo anterior, vemos que los bancos del grupo 1 tienen una composición de activos donde la inversión en valores tiene una importancia mayor en la cola de la distribución, es decir, el grupo 1 contiene ciertos bancos que le dan un peso grande a la inversión en valores. Para los bancos del grupo 2 también resultan importantes las inversiones en valores, aunque con una cola menos pesada, y las instituciones del grupo 3 tienen el menor valor para dicha cuenta.

7. DISCUSIÓN DE RESULTADOS Y CONCLUSIONES

El resto de las cuentas tienen una distribución muy similar para todos los grupos. Recordemos que estamos utilizando el primer nivel de cuentas, por lo que el número no es demasiado grande. Por el contrario, la información de la distribución de las cuentas más importantes de los siguientes niveles del catálogo mínimo (Figuras 4.8 y 4.9) puede resultar complementaria, pues como mencionamos en el capítulo 4, las cuentas contenidas en dichas visualizaciones resultan ser subcuentas de las categorías más generales de la Figura 4.7.

De todo lo anterior y apoyándonos en la Figura 4.7, podemos concluir que los bancos en México pueden ser agrupados de acuerdo a su capitalización, tanto la parte aportada por los dueños / accionistas de la institución (**capital contribuido**), como por la parte que se obtiene de la actividad de la institución, ya sean pérdidas o ganancias (**capital ganado**). El último punto resulta relevante, pues como se mencionó en el capítulo 3, la inclusión de información referente al desempeño de las instituciones se evito a toda costa, sin embargo, la inclusión del capital ganado y sus subcuentas a los conceptos más importantes puede funcionar como intermediario de los efectos del desempeño de una institución para su agrupación.

Adicional a lo anterior, la importancia que tienen los **préstamos interbancarios** para los pasivos de las instituciones también juega un papel importante en la agrupación, específicamente, es el tercer grupo el que se diferencia claramente de los otros dos, pues los bancos de dicho conglomerado tienen una actividad importante de fondeo proveniente a su vez de otros bancos.

Por su parte, la **cartera de crédito** también tiene un papel importante en la clasificación de los bancos Mexicanos. Viendo la distribución de esta cuenta, es claro que en general la tercera agrupación le da una mayor importancia a los créditos como modelo de negocio, seguida de la segunda agrupación y es la agrupación 1 la que le da la menor importancia (lo que se refleja como un menor porcentaje de los activos totales).

Finalmente, las **inversiones en valores** como porcentaje del activo ayudan a diferenciar a los bancos. A pesar de que la mayor parte de la masa de la distribución no es tan diferente entre los tres grupos, queda claro que en las colas los grupos sí son diferentes. Algunos bancos del grupo 1 le dan mayor importancia a dicha actividad respecto al resto de las instituciones, seguidas por instituciones del grupo 2 y al final el grupo 3.

Es claro que la descripción anterior está limitada por la cantidad de información contenida en las cuentas incluidas en el primer nivel, pero como ya mencionamos, el objetivo del presente trabajo es presentar un primer esbozo de una metodología de clasificación de los bancos en México. Ya discutimos que la incorporación de más información del balance tiene un efecto importante en la clasificación, por lo que es importante definir qué tan específico se quiere hacer el análisis, y los efectos potenciales de hacerlo.

Adicional a lo anterior, a pesar de que la inclusión financiera en México ha tenido avances importantes en los últimos años¹, aún hay grandes oportunidades para el sector

¹Los principales resultados del año 2018 pueden ser consultados en <https://www.gob.mx/cnbv/>

bancario en México, pues si se considera que sólo el 68 % de los adultos en México contaba con al menos un producto financiero al cierre de 2018, es claro que la variedad de productos que aún se pueden desarrollar y ofrecer en México es amplia, lo que tendría beneficios claros para la inclusión financiera en el país, lo que a su vez ayudaría a crear un sistema bancario más heterogéneo y menos concentrado en los productos bancarios clásicos.

7.1.2. Agrupación con SBM

Para obtener una agrupación a partir de los patrones de conectividad de los bancos en la red de exposiciones interbancarias la tarea fue más sencilla, pues no hubo necesidad de realizar múltiples ajustes, simplemente se usó la herramienta de detección de modelos de bloques estocásticos. Así, la tarea de seleccionar un modelo sobre otro no es necesaria. Más aún, en la sección de resultados del capítulo 6 se hizo una breve caracterización de los grupos de acuerdo a la centralidad en la red de cada banco, mismo que complementaremos a continuación.

Como mencionamos anteriormente, la distribución del grado de nuestra red se queda muy corta con la distribución de grado de una red libre de escala, por lo que se ajustaron modelos de bloques estocásticos sin corrección de grado. Esto tiene un efecto notable en los determinantes de la clasificación obtenida; como podemos observar en la Figura 6.3, el grado total de los nodos es la medida más influyente para llegar a la clasificación obtenida, sin embargo, al hacer el mismo ajuste con corrección de grado todos los bancos son clasificados bajo el mismo grupo, lo que da evidencia de la importancia que toma la ponderación por magnitud del grado para obtener los ajustes en redes con una mayor heterogeneidad.

Por otro lado, si nos enfocamos en las Figuras 6.4 y 6.5 notamos que los bancos del grupo 1 son, en general, los más centrales en la red, lo que se hace claro al ver la distribución de medidas como el grado, la fuerza, el DebtRank, la interinidad y el PageRank. Lo anterior se hace evidente, particularmente con el DebtRank y el PageRank, que tienen un sentido directo de importancia, el primero de importancia económica/financiera sistémica y el segundo de importancia estructural. En segundo lugar, se encuentran los bancos del grupo 2 y los bancos del grupo 3 son los que tienen los valores más pequeños de las medidas mencionadas.

Una vez más es visible el fenómeno que identificamos en el capítulo 5, en el que las instituciones sistémicas tienen valores consistentemente más bajos para indicadores de riesgo, como el índice de Herfindahl-Hirschman que podemos observar en la Figura 6.5 *b*, lo que es deseable para la salud del sistema, y que da cuenta de cómo las nuevas regulaciones tienen un efecto positivo en la resiliencia del sistema bancario.

En la Figura 7.1 vemos que las conexiones están repartidas entre los grupos de manera casi uniforme, lo que da más evidencia para alegar que, a pesar de los resultados en el capítulo 5, el modelo de Centro-Periferia no es el mejor para la red interbancaria

[articulos/resultados-de-la-encuesta-nacional-de-inclusion-financiera-enif2018?idiom=es](#)

7. DISCUSIÓN DE RESULTADOS Y CONCLUSIONES

Mexicana, pues no hay un grupo dominante en términos de conectividad. Más aún, a pesar de que es el grupo 1 el que tiene las instituciones más conectadas y con mayor importancia sistémica, no es el que tiene la mayor conectividad en número de conexiones, tanto intragrupo como intergrupo, lo que va en contra del supuesto de conectividad dominante del centro, más aún, el tercer grupo, que contiene las instituciones menos centrales en la red concentra un porcentaje de conexiones intragrupo más grande que el grupo 1, respecto al total de conexiones existentes en la red.

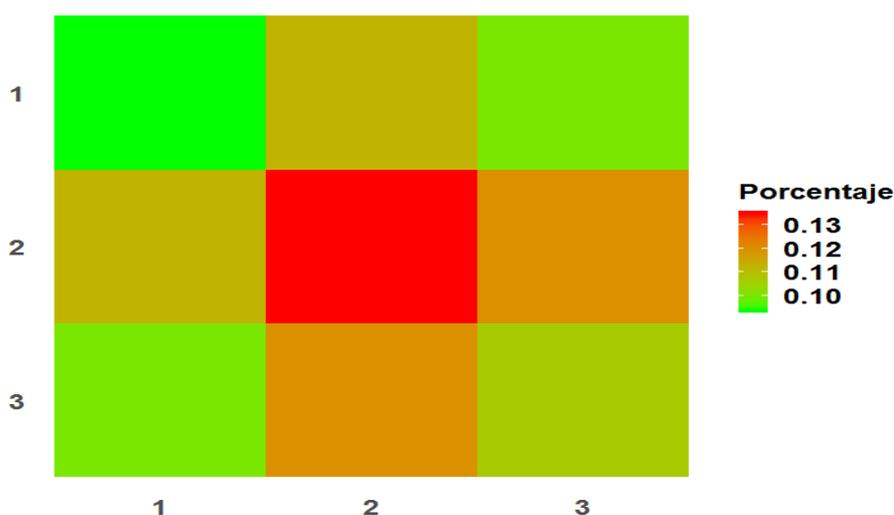


Figura 7.1: Número de conexiones entre los grupos obtenidos con el modelo de bloques estocásticos como porcentaje del total de ligas para la última red del periodo de estudio.

En conclusión, las instituciones con la importancia sistémica más alta y al mismo tiempo más influyentes en la red tienden a tener patrones de conectividad similares, esto es, recordando el capítulo 6, tienen las mismas probabilidades de conectarse entre sí, como con el resto de la red. Por su parte, los bancos del segundo grupo mantienen una fuerte actividad con instituciones de su mismo grupo, al igual que las instituciones del grupo 3. Es importante recordar que hay una propiedad dinámica de los patrones de conectividad que no estamos capturando, pues estamos analizando diferentes momentos de la red.

7.1.3. Comparación de ajustes con RKM contra SBM

Uno de los objetivos del presente trabajo era encontrar evidencia de si la conectividad en la red interbancaria modifica la manera en la que los bancos se agrupan, para ello, se obtuvieron agrupaciones con dos métodos distintos. Los grupos resultantes de ambos métodos están presentes en las Figuras 7.2, 7.3 y 7.4. A pesar de que sólo el ajuste RKM N1 fue seleccionado, presentamos los tres niveles de profundidad de los

datos para ilustrar que, sin importar la información con la que se alimente al método de conglomerados, la agrupación de los bancos de acuerdo a su nivel de conectividad siempre cambia.

Al principio hicimos incapié en la importancia de realizar un monitoreo diferenciado para las instituciones bancarias, y hay evidencia de que, en efecto, el comportamiento de los bancos como elementos de un sistema se ve influenciado por su interacción con el resto de los participantes de dicho sistema (Figuras 7.2, 7.3 y 7.4).

Por ejemplo, el grupo 2 de los ajustes con modelos de bloques estocásticos es constantemente nutrido por instituciones de todos los grupos encontrados por RKM (con excepción del ajuste con datos del segundo nivel de profundidad) lo que habla de la influencia que tiene la interacción que tienen las instituciones entre sí sobre la estructura de agrupación del sistema.

Más aún, con los bancos con los que podría resultar natural que estén en un mismo grupo (por ejemplo, bancos de inversión o bancos de importancia sistémica), existe una discrepancia entre los dos métodos, como se puede observar en las Figuras 7.2, 7.3 y 7.4, pues hay instituciones que constantemente cambian del segundo grupo obtenido con RKM al primer grupo obtenido con SBM.

7. DISCUSIÓN DE RESULTADOS Y CONCLUSIONES

Comparación de grupos obtenidos con SBM vs RKM

Primer nivel de granularidad

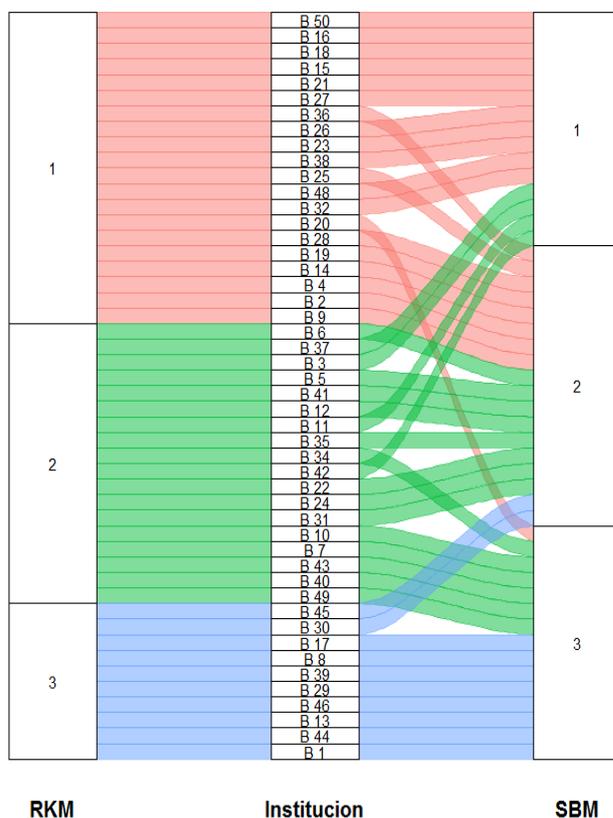


Figura 7.2: Comparación de los grupos asignados por SBM y por RKM a los bancos para el primer nivel del catálogo mínimo.

En el segundo nivel de agregación obtenemos un grupo más con el método de RKM, sin embargo, como podemos constatar en la Figura 7.3, son dos instituciones que a través de los tres niveles se encuentran en el mismo grupo obtenido con SBM; la razón, como lo explicamos anteriormente, es la sensibilidad a la incorporación de nuevos datos, así como la sensibilidad a valores extremos. Podemos ver la posición de los bancos en las mismas cuentas que resultan importantes para el ajuste RKM N2 (Figura 7.5), donde se hace evidente que los valores extremos en Capital Social (4) son los que provocan la formación del cuarto grupo, pues las dos instituciones en dicho conjunto no resultan significativamente distintos del resto para las otras cuentas en la Figura.

Comparacion de grupos obtenidos con SBM vs RKM

Segundo nivel de granularidad

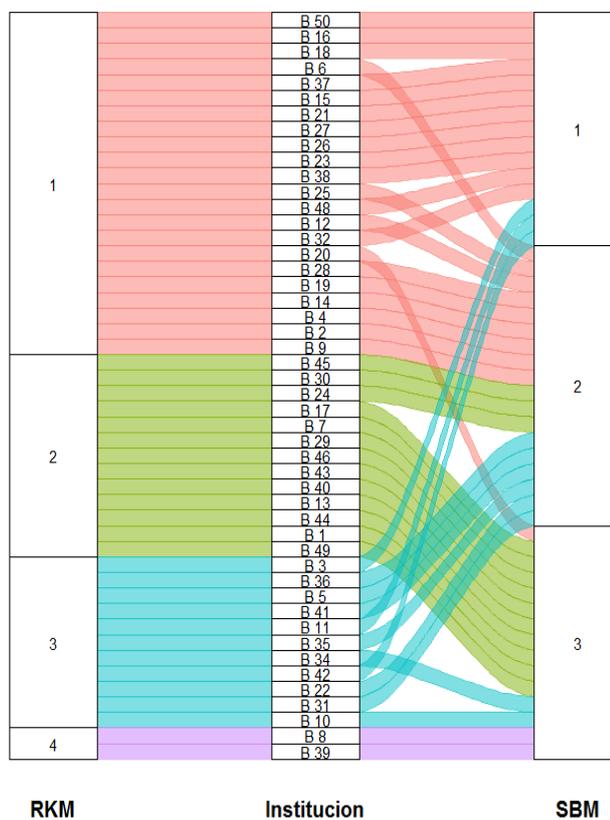


Figura 7.3: Comparación de los grupos asignados por SBM y por RKM a los bancos para el segundo nivel del catálogo mínimo.

7. DISCUSIÓN DE RESULTADOS Y CONCLUSIONES

Comparacion de grupos obtenidos con SBM vs RKM

Tercer nivel de granularidad

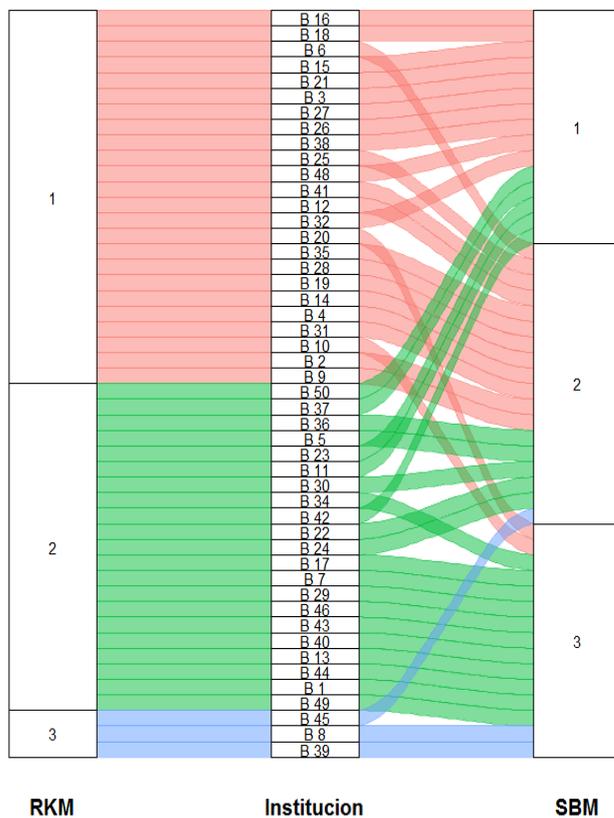


Figura 7.4: Comparación de los grupos asignados por SBM y por RKM a los bancos para el tercer nivel del catálogo mínimo.

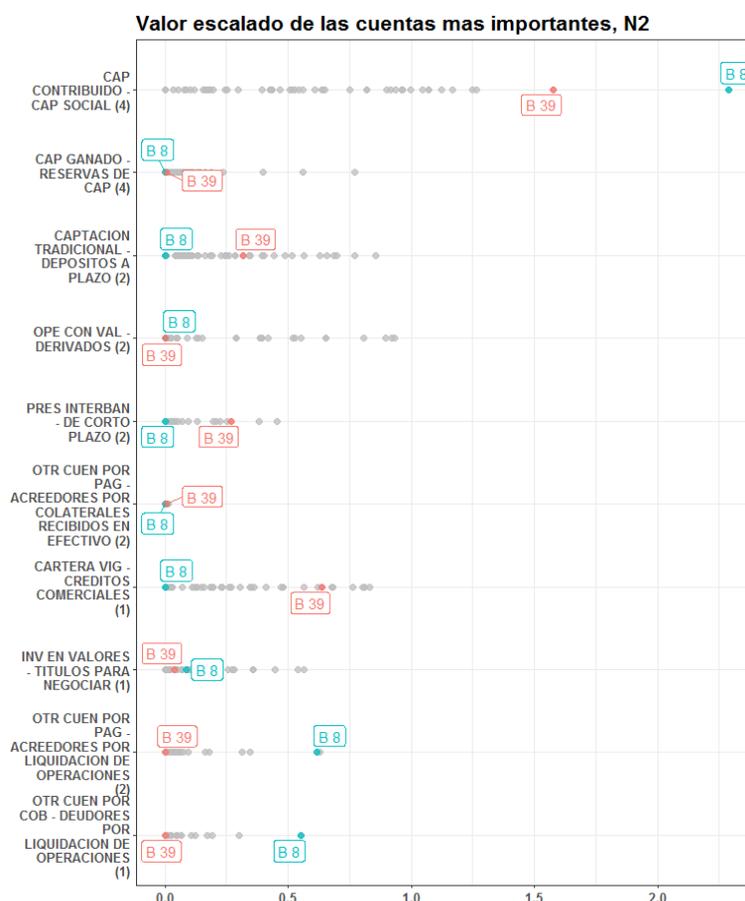


Figura 7.5: Valor en las cuentas más importantes del ajuste RKM N2 diferenciando las dos instituciones que son agrupadas en un grupo no existente para la fecha anterior del mismo ajuste

De lo anterior, queda claro que la agrupación de los bancos es propensa a cambiar al considerarlos como parte de un sistema, en lugar de analizarlos de manera individual. Esto tiene implicaciones importantes para hacer más fuertes los argumentos a favor de la consideración por agrupaciones del sistema bancario mexicano pues, por ejemplo, puede existir el caso de una institución altamente concentrada en un tipo específico de actividad, y al mismo tiempo tener una posición central en la red interbancaria, y como sabemos, la concentración combinada con importancia sistémica en la red tiene el potencial de provocar severos desequilibrios en el buen comportamiento de los bancos, y por consiguiente, del sistema financiero en su totalidad.

7.2. Conclusiones

Considerando todos los resultados obtenidos a lo largo de la tesis, las conclusiones generales del trabajo pueden ser resumidas como sigue:

- Los modelos de negocios de los bancos se ajustan de acuerdo al ambiente económico y financiero en el que se desenvuelven.
- El modelo de conglomerados utilizado es sensible a la incorporación de nueva información, así como a los valores extremos en las cuentas de balance para las que siempre tienen valores distintos de cero.
- Para los bancos mexicanos, las cuentas de capital parecen tener un impacto consistente a lo largo del periodo de estudio sobre la estructura de agrupación del sistema. Dicho resultado es interesante si consideramos que el sistema bancario Mexicano ha mantenido niveles de capitalización altos y consistentes aún con la introducción de las nuevas regulaciones, por lo que resulta complicado pensar que se deba a un tema de ajustes no coordinados por parte de las instituciones.
- Adicionalmente, las Figuras 4.4 a 4.6 nos dejan ver que las carteras de crédito y las operaciones con otros instrumentos financieros (en especial los títulos y los derivados) cobran alta importancia en momentos específicos del tiempo, en los que la incertidumbre global puede ser utilizada para explicar dichos movimientos en la composición de los balances bancarios.
- La inclusión de nueva información, o la utilización de niveles del catálogo mínimo más profundos, puede ayudar a la interpretabilidad de los resultados, de nuevo, podemos usar las Figuras 4.4 a 4.6 para notar que los conjuntos de variables seleccionadas como importantes están contenidos en las cuentas del nivel anterior de desagregación.
- Las agrupaciones son propensas a cambiar cuando se agrega la interacción que tienen las instituciones en el sistema interbancario, provocando incluso que bancos con “poca” importancia sean incluidos por los modelos de bloques estocásticos en los grupos donde se encuentran las instituciones con mayor importancia sistémica.
- Los resultados obtenidos con ambos métodos de agrupamiento brindan evidencia de que las instituciones bancarias en México no se agrupan de manera trivial, por lo que es complicado resumir sus patrones de conectividad así como las actividades que realizan, tomando en cuenta sólo información individual, es importante tomar en cuenta la interconectividad del sistema bancario y financiero.
- Dado el tamaño y la poca heterogeneidad del sistema bancario Mexicano, no fue necesario el uso de la corrección de grado en los modelos de bloques estocásticos, lo que provoca que el grado sea reconocido como la variable de red más importante para la obtención de la agrupación, sin embargo, es interesante ver otras medidas en el top de importancia, pues dan evidencia de un sistema altamente conectado.

- La bancarización en México es un tema que deja ver una gran oportunidad para que las instituciones bancarias amplíen la variedad de productos que ofrecen, impulsando la composición de un sistema bancario más heterogéneo y menos concentrado en los mismos productos y con las mismas contrapartes, impulsando la salud y la resistencia del sistema financiero.

7.2.1. Trabajo futuro

A lo largo de la tesis quedó claro que como primer acercamiento al problema de la identificación de modelos de negocios, la metodología propuesta puede ser mejorada tomando en consideración las situaciones que se ha demostrado tienen un impacto en los resultados obtenidos. Por lo anterior, podemos mencionar algunos puntos que se pueden implementar como trabajo futuro, con la finalidad de hacer más robusto el procedimiento propuesto:

- Incluir el aspecto dinámico de los modelos de negocios. Como mencionamos, en [Lucas et al. \(2017\)](#) utilizan un método que toma en cuenta el aspecto dinámico, sin embargo, es aplicado a un conjunto significativamente más pequeño que el nuestro, lo que puede provocar un deterioro importante en la viabilidad de la implementación computacional.
- Incluir el peso de las ligas en los ajustes con Modelos de Bloques Estocásticos. Hay métodos que lo hacen, sin embargo los ajustes tendrían que hacerse tomando en consideración que los supuestos hechos sobre la distribución del peso de las conexiones puede ser restrictiva, o alejada de la realidad.
- Aplicar los mismos dos métodos de obtención de agrupaciones en un sistema bancario (o incluso financiero) más grande y más heterogéneo. En el caso de los datos de balance, los resultados serían interesantes al aplicar el método de conglomerados a un sistema bancario internacional, lo que nos ayudaría a detectar si un componente geográfico tiene un efecto importante en los modelos de negocios de los bancos, es decir, si las condiciones nacionales de cada banco afectan la toma de decisiones, o si por el contrario, es más fuerte el impacto de las condiciones internacionales.

Por su parte, la aplicación del Modelo de Bloques Estocásticos a un sistema interfinanciero (incluyendo otro tipo de intermediarios, como fondos de inversión, casas de bolsa o aseguradoras, por ejemplo) nos permitiría verificar si los distintos tipos de intermediarios tiene patrones de conectividad muy distintos, o si forman grupos homogéneos en términos de tipo.

- Dentro de la literatura de redes, el enfoque “multicapa” ha cobrado mucha fuerza en los últimos años dado el poder que tiene para desenmascarar aún más características no triviales de los sistemas complejos que representan. La representación del sistema financiero como red multicapa, donde cada capa es un mercado, podría

7. DISCUSIÓN DE RESULTADOS Y CONCLUSIONES

dar información adicional para el descubrimiento de grupos con instituciones similares, los cuales a su vez pueden ser comparables con los resultados obtenidos con otro conjunto de datos y otro método.

Apéndice A

A.1. Validación cruzada y Remuestreo (*Remuestreo*)

A.1.1. Validación cruzada

La validación cruzada y el remuestreo pertenecen a los **métodos de remuestreo** (*resampling methods*, una referencia clásica es Fisher (1935)), los cuales, como su nombre lo dice, extraen subconjuntos de observaciones de los datos totales de manera aleatoria e iterativa, con la finalidad de realizar pruebas de validación de los modelos obtenidos con cada uno de los subconjuntos extraídos. Dado que los conjuntos de datos no tienen porqué estar balanceados, o no tienen porque ser suficientemente grandes para poder partirlos sin perder información, la validación cruzada nos ayuda a obtener información de un modelo ajustado.

Existen diferentes tipos de validación cruzada, pero como ya se mencionó, el principio es el mismo para todos: obtener numerosas particiones de tamaño dos del conjunto de datos, utilizar una parte de cada partición para entrenar un modelo y la otra para probar su eficacia y promediar los errores de cada ajuste; las diferentes formas de validación cruzada difieren principalmente en el tamaño de las particiones que se obtienen.

A.1.1.1. Validación Cruzada Deja-Uno-Fuera (LOOCV)

Como su nombre indica, la validación cruzada “Deja-Uno-Fuera” (*Leave-One-Out Cross-Validation*) es parecida al método clásico de partir en un subconjunto de entrenamiento y en uno de prueba los datos, sin embargo, es un proceso iterativo en el que se selecciona sólo una de las observaciones, misma que se deja fuera del conjunto de entrenamiento, con el que se entrena el modelo, y posteriormente se obtiene el error cuadrático medio MSE evaluando el ajuste en cada una de las observaciones.

Dado que se tienen n observaciones, se obtienen n mediciones diferentes del error cuadrático medio $MSE_1, MSE_2, \dots, MSE_n$ las cuales se promedian para obtener la

estimación final del método “Deja-Uno-Fuera” de la siguiente forma:

$$CV_{(n)} = \frac{1}{n} \sum_{i=1}^n MSE_i. \quad (\text{A.1})$$

Las dos ventajas principales del método LOOCV es que, a diferencia del enfoque clásico de particionar los datos en dos partes, al tomar la mayoría de las observaciones en cada iteración ($n-1$ de n) el sesgo se ve disminuido significativamente. Además, dado que las iteraciones son iguales sin importar cuántas veces se aplique el procedimiento el resultado seguirá siendo el mismo y por lo consiguiente no existe varianza, pues no importa el orden en el que se seleccione la observación para obtener el MSE.

A.1.1.2. Validación cruzada k-partes

La validación cruzada “ k -partes” (*k-fold Cross-Validation*) es similar al método LOOCV, con la diferencia de que aquí el conjunto de datos es dividido en k grupos de tamaño aproximadamente igual, y en cada iteración uno solo de esos grupos de observaciones es utilizado como conjunto de prueba de los ajustes y los restantes $k-1$ grupos se usan como datos de entrenamiento. Para cada iteración se obtiene el error promedio cuadrado MSE_1 , y así sucesivamente hasta tener las mediciones $MSE_1, MSE_2, \dots, MSE_k$, con los que se calcula la estimación total como:

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^k MSE_i. \quad (\text{A.2})$$

Es claro que la validación cruzada en k -partes es la generalización de LOOCV. La ventaja directa es que es mucho menos intensiva computacionalmente, pues no es necesario ajustar y comprobar un modelo n veces. Adicionalmente, como mencionamos en el capítulo de Bosques Aleatorios, tener un modelo demasiado simple (o complejo) aplicado a muy pocos datos puede tener un efecto fuerte en la relación Varianza-Sesgo. En [James et al. \(2014\)](#) se menciona que empíricamente se ha demostrado que la validación cruzada “ k -partes” con $k=5$ o $k=10$ tiene errores de prueba sin exceso de sesgo ni exceso de varianza.

A.1.1.3. Validación Cruzada en problemas de clasificación

Las definiciones anteriores trabajan bajo el supuesto de que la variable dependiente y es continua, es decir, es un problema de regresión, sin embargo, la validación cruzada también se puede aplicar para problemas de clasificación. En lugar de utilizar el error cuadrático medio, utilizamos como criterio de optimalidad el número de clasificaciones incorrectas bajo el modelo ajustado. Así, la estimación del error de prueba bajo la validación “deja-uno-fuera” está dada por:

$$CV_{(n)} = \frac{1}{n} \sum_{i=1}^n Err_i, \quad (\text{A.3})$$

donde $Err_i = I(y_i \neq \hat{y}_i)$. El error se define de la misma forma para la validación k -partes.

A.1.2. Remuestreo

Al igual que la validación cruzada, el **remuestreo** (Efron (1979)) es un método de remuestreo del conjunto de datos original que nos ayuda a obtener información extra sobre la efectividad de un modelo ajustado, que de otra forma resulta demasiado complicado obtener, específicamente, nos ayuda a obtener información sobre la variabilidad (sensibilidad) de un modelo a la incorporación de datos no observados previamente vista como la desviación estándar del estimador del valor que queremos aproximar.

La diferencia principal con los métodos de validación cruzada es que en el remuestreo se seleccionan n observaciones del conjunto original de datos (es decir, se genera un conjunto de datos del mismo tamaño que el original) **con reemplazo**. Es decir, una observación puede estar representada más de una vez en un conjunto generado por el remuestreo. Nombramos a cada conjunto como Z^{*i} , $i \in 1, \dots, B$, para un valor de B suficientemente grande para reducir la varianza de la aproximación. Con cada conjunto Z^{*i} se genera una nueva estimación del valor que se está aproximando, y^{*i} , a partir de los cuales podemos calcular el error estándar de las estimaciones con Remuestreo como:

$$SE_B(\hat{\alpha}) = \sqrt{\frac{1}{B-1} \sum_{r=1}^B \left(\hat{\alpha}^{*r} - \frac{1}{B} \sum_{r'=1}^B \hat{\alpha}^{*r'} \right)^2}. \quad (\text{A.4})$$

Apéndice B

B.1. Medidas de red

Clasificamos todas las medidas que describimos a continuación de acuerdo con las tres categorías mencionadas en el capítulo 5.

B.1.1. Número de nodos (*Medida global*)

El número de nodos es simplemente la cardinalidad del conjunto \mathcal{N} , que denotaremos con $n = |\mathcal{N}|$.

B.1.2. Número de arcos (*Medida global*)

El número de de arcos es la cantidad de conexiones entre todos los pares de nodos presentes en una red, lo denotamos como $m = |\mathcal{A}|$.

B.1.3. Distancia (*Medida cuasi-local*)

En una gráfica, definimos la distancia entre dos vértices $i, j \in \mathcal{V}$ como el mínimo número de arcos que se tienen que recorrer para llegar de un nodo al otro. Se denota como d_{ij} . Si dicha secuencia de arcos no existe, denotamos la distancia como $d_{ij} = \infty$.

B.1.4. Grado (*Medida estrictamente local*)

En una red no dirigida, el grado de un nodo $i \in \mathcal{V}$ es el número de arcos que tienen a i como origen en alguno de sus extremos, se calcula a partir de la matriz de adyacencias como:

$$d_i = \sum_{j \in N(i)} a_{ij} = |N(i)|. \quad (\text{B.1})$$

En una red financiera, el grado de un nodo (institución) representa el número de contrapartes con las que dicho nodo realiza transacciones, cuya interpretación depende del mercado en el que se den. En una red dirigida, definimos a partir de las matrices de adyacencia interior y exterior el grado interior y el grado exterior como sigue.

B.1.4.1. Grado exterior (*Medida estrictamente local*)

El grado exterior de i es el número de arcos que dirigidos que tiene su origen en él. Se define como:

$$d_i^+ = \sum_{j \in N^+(i)} a_{ij}^+ = |N^+(i)|. \quad (\text{B.2})$$

El grado exterior se entiende como el número de instituciones hacia las que la institución i tiene un flujo dirigido dinero, es decir, el número de instituciones a los que i está expuesta.

B.1.4.2. Grado interior (*Medida estrictamente local*)

Análogamente, definimos el grado interior de i como:

$$d_i^- = \sum_{j \in N^-(i)} a_{ij}^- = |N^-(i)|. \quad (\text{B.3})$$

Al contario del grado exterior, el grado interior es el número de instituciones que tienen un flujo de efectivo dirigido hacia i , es decir, son las instituciones que están expuestas a i .

B.1.5. Nodos desconectados (*Medida global*)

De las definiciones anteriores, decimos que un nodo está desconectado si su grado es igual a cero. En una red dirigida, el grado exterior no tiene porqué ser igual al interior, el sentido es el siguiente: una institución puede decidir sólo prestar dinero (grado exterior mayor a cero), y no tomar préstamos de nadie mas (grado interior cero), y viceversa.

B.1.6. Fuerza (*Medida estrictamente local*)

En una red dirigida con pesos asociados a las conexiones entre nodos, definimos a la fuerza como la suma total de los pesos de dichas conexiones. Se define como:

$$s_i = \sum_{j \in N(i)} w_{ij}. \quad (\text{B.4})$$

Dado que tenemos dirección en las conexiones, a partir de la matriz de pesos definimos la fuerza interior y exterior como sigue:

B.1.6.1. Fuerza exterior (*Medida estrictamente local*)

La *fuerza exterior* de un nodo i es:

$$s_i^+ = \sum_{j \in N^+(i)} w_{ij}^+. \quad (\text{B.5})$$

En nuestra red bancaria, representa el flujo total de efectivo que la institución i provee al sistema.

B.1.6.2. Fuerza interior (*Medida estrictamente local*)

Análogamente, la *fuerza interior* de un nodo i se define como:

$$s_i^- = \sum_{j \in N^-(i)} w_{ij}^-. \quad (\text{B.6})$$

Representa el flujo total de efectivo que la institución i está tomando del sistema bancario.

B.1.7. Flujo Neto (*Medida estrictamente local y global*)

El *flujo neto* de un nodo i es una medida que nos ayuda a detectar el papel que juega un nodo en una red. Dicha medida se puede obtener para un par de instituciones, así como a nivel sistema. El flujo neto entre un par de instituciones se define como:

$$f_{ij} = w_{ij}^+ - w_{ij}^-. \quad (\text{B.7})$$

Representa la cantidad neta de dinero entre las instituciones i y j . Si $w_{ij}^+ > w_{ij}^-$, significa que i tiene una relación de pretamista neto con j ; si se tiene el caso contrario, significa que i tiene una relación de prestatario neto con j .

A nivel sistema, el flujo neto de una institución i se define como:

$$f_i = \sum_{j \in N(i)} f_{ij} = s_i^+ - s_i^-. \quad (\text{B.8})$$

Si el flujo neto de i contra todo el sistema es positivo, quiere decir que i es un pretamista neto para el sistema, es decir, provee de liquidez al sistema. Por el contrario, si el flujo es negativo, quiere decir que i es un sumidero de liquidez, es decir, absorbe mayores recursos del sistema de los que aporta al mismo.

B.1.8. Volumen (*Medida global*)

El volumen se calcula como la suma de los pesos de todos los arcos presentes en una red, es decir:

$$Vol = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} w_{ij}^+ = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} w_{ij}^-. \quad (\text{B.9})$$

Representa el volumen total de efectivo que se movió en el sistema bancario en el día al que pertenece la matriz de pesos correspondiente. Puede ser utilizada para medir la intensidad de las transacciones realizadas en todo el sistema.

B.1.9. Reciprocidad (*Medida global*)

El índice de reciprocidad sólo es válido en una red dirigida, pues representa la proporción de arcos que tienen un arco en dirección opuesta en toda la red. Toma valores entre cero y uno, donde cero significa que no existe ningún arco en sentido contrario a alguno de los arcos existentes, y uno significa que para todos los arcos de la red existe un arco en sentido contrario. Lo definimos como.

$$r = \frac{\sum_{i \in \mathcal{N}} \sum_{j \in N(i)} a_{ij}^+ a_{ji}^+}{\sum_{i \in \mathcal{N}} \sum_{j \in N(i)} a_{ij}^+}. \quad (\text{B.10})$$

B.1.10. Afinidad (*Medida cuasi-local*)

La afinidad de un nodo i nos ayuda a descubrir si éste tiende a formar conexiones con nodos de grado similar, o por el contrario, si tiende a conectarse con nodos de grado menor o mayor. De lo anterior se desprenden dos fenómenos importantes, el primero de ellos es llamado *assortative mixing*, que se da cuando los nodos de una red forman conexiones con nodos de grado similar al propio. Al fenómeno contrario, que se da cuando los nodos de la red forman conexiones con nodos de grado diferentes (nodos de grado alto con nodos de grado pequeño y vice versa) se le conoce como *diassortative mixing*. Se calcula a partir de la red no dirigida asociada a la red dirigida, pues solo nos interesan las conexiones de cada nodo, no la dirección de cada arco. Se define como:

$$a_i = \frac{1}{d_i} \sum_{j \in N(i)} d_j. \quad (\text{B.11})$$

El fenómeno de *diassortative mixing* es común en muchas redes financieras que presentan una estructura llamada *Core-Periphery*.

B.1.11. Índice de Herfindahl-Hirschman (*Medida cuasi-local*)

El índice de Herfindahl-Hirschman es común en las finanzas y la economía; es un indicador que permite evaluar el nivel de concentración en, por ejemplo, la cartera de crédito de un banco, un portafolio de inversión, etcétera. En nuestro caso, el índice nos indica el nivel de concentración en las contrapartes de una institución, es decir, nos dice si una institución tiene la mayoría (o el total) de sus flujos de efectivo con una sola contraparte o varias.

Al igual que en la mayoría de las medidas descritas hasta ahora, para una red dirigida podemos definir un índice exterior y uno interior.

B.1.11.1. Índice de Herfindahl-Hirschman exterior (*Medida cuasi-local*)

Se define como:

$$HHI^+(i) = \sum_{j \in N(i)} \left(\frac{w_{ij}^+}{\sum_j w_{ij}^+} \right)^2. \quad (\text{B.12})$$

El índice anterior nos ayuda a descubrir qué tan concentradas están los flujos de efectivo exteriores de una institución, si $HHI^+(i)$ es cercano a uno, quiere decir que la gran mayoría de los préstamos que está realizando i se están dirigiendo a muy pocas contrapartes, por el contrario, si es cercano a cero, quiere decir que la institución tiene repartidos los préstamos que realiza entre un número mayor de instituciones; más aún, los montos mismos están repartidos de una manera más homogénea.

B.1.11.2. Índice de Herfindahl-Hirschman interior (*Medida cuasi-local*)

Se define como:

$$HHI^-(i) = \sum_{j \in N(i)} \left(\frac{w_{ij}^-}{\sum_j w_{ij}^-} \right)^2. \quad (\text{B.13})$$

Si el índice interior tiene valores cercanos a uno, quiere decir que los préstamos que está recibiendo la institución i provienen de pocas contrapartes; por el contrario, y de manera análoga al índice exterior, si es cercano a cero, quiere decir que los préstamos se reciben de un mayor número de instituciones y de manera más homogénea.

El índice de Herfindahl-Hirschman resulta importante para conocer la sensibilidad de las instituciones a un proceso de contagio, pues, si tiene un índice muy alto (i.e. da/recibe préstamos a/de muy pocas contrapartes), la quiebra de hasta una sola de esas contrapartes puede poner en serios problemas a una institución muy concentrada. Desde el punto de vista regulatorio, es deseable que las exposiciones/ responsabilidades de las instituciones bancarias estén repartidas de manera homogénea entre un número grande de instituciones (el qué tan grande dependerá de muchos factores, desde la salud de las mismas, como su calificación, o su importancia para el sistema completo).

B.1.12. Coeficiente de agrupamiento (*Medida cuasi-local*)

El coeficiente de agrupamiento nos dice qué tanto los nodos tienden a formar grupos altamente conectados entre sí, con este índice podemos medir que tan densas son las conexiones entre los vecinos de un nodo. El coeficiente indica qué tan probable es que dos vecinos de un nodo sean también vecinos entre ellos. Para el nodo i , el coeficiente de agrupamiento se define como:

$$c_i = \frac{t_i}{d_i(d_i - 1)/2}, \quad (\text{B.14})$$

donde t_i es el número de triángulos a los que pertenece i . Un triángulo es una subestructura en una red en la que, en tres “pasos” se llega del nodo i a él mismo.

En [Silva et al. \(2016\)](#) utilizan además dos medidas extendidas del coeficiente de agrupamiento con interpretaciones interesantes.

B.1.12.1. Coeficiente de agrupamiento ponderado de un nodo (*Medida cuasi-local*)

Se define como:

$$cc_i = \frac{1}{s_i(d_i - 1)} \sum_{(i,j) \in \mathcal{N}^2} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{ik} a_{jk}, \quad (\text{B.15})$$

donde s_i, d_i son la fuerza y el grado, respectivamente, y w y a son las entradas de las matrices de pesos y de adyacencia respectivamente. Según [Silva et al. \(2016\)](#), mientras más grande sea cc_i , más fácil es reemplazar los flujos de efectivo de i .

Para una red dirigida, podemos extender el coeficiente para prestamistas (exterior) y prestatarios (interior). El *coeficiente de agrupamiento ponderado exterior* se define como:

$$cc_i^+ = \frac{1}{s_i^+(d_i^+ - 1)} \sum_{(i,j) \in \mathcal{N}^+} \frac{w_{ij}^+ + w_{ik}^+}{2} a_{ij}^+ a_{ik}^+ a_{jk}^+. \quad (\text{B.16})$$

El *coeficiente de agrupamiento ponderado interior* (*Medida cuasi-local*) se define como:

$$cc_i^- = \frac{1}{s_i^-(d_i^- - 1)} \sum_{(i,j) \in \mathcal{N}^-} \frac{w_{ij}^- + w_{ik}^-}{2} a_{ij}^- a_{ik}^- a_{jk}^-. \quad (\text{B.17})$$

Dichos coeficientes representan qué tan sustituible es una institución para invertir (exterior) o para conseguir recursos (interior). Si el coeficiente para una institución es pequeño, quiere decir que dicha institución es importante para su vecindario, pues la falta de la misma reduce considerablemente las instituciones de las que se puede conseguir o a las que se les puede prestar recursos.

B.1.12.2. Coeficiente de agrupamiento ponderado de la red (*Medida global*)

El coeficiente ponderado se puede extender para ser una medida global como sigue:

$$CC = \frac{1}{n} \sum_{i \in \mathcal{N}} cc_i, \quad (\text{B.18})$$

donde n es el número de nodos en la red. Es decir, el *coeficiente de agrupamiento ponderado de la red* es el promedio de los coeficientes de cada nodo. Un coeficiente grande (cercano a uno), quiere decir que la gran mayoría de las instituciones de la red son sustituibles, es decir, todos los bancos de la red tienen diversas opciones para llevar a cabo sus actividades de solicitud y otorgamiento de préstamos.

B.1.13. Índice de completitud (*Medida global*)

El *índice de completitud* nos dice qué tan cercana está la red a tener todos los arcos posibles, es decir, qué tan cercana está a que todos los nodos estén conectados entre sí de manera directa. Si tenemos n nodos y m arcos en una red dirigida, definimos el índice como:

$$p = \frac{m}{n(n-1)}. \quad (\text{B.19})$$

Si el índice es igual a uno, quiere decir que todas las instituciones realizan transacciones con el resto de la red; caso contrario si el índice es cero.

Ahora, obtenemos la definición de centralidad por *eigenvectores* de Newman (2010).

B.1.14. Eigenvectores

La construcción de la centralidad por eigenvectores es iterativa. Se inicia asumiendo un valor arbitrario, y se actualiza tomando en cuenta la centralidad de cada vecino. Así, supongamos que la centralidad por eigenvectores inicia como $x_i = 1 \forall i$; posteriormente, actualizamos a una centralidad por eigenvector x'_i , que definimos como la suma de las centralidades de los vecinos de i :

$$x'_i = \sum_j A_{ij} x_j, \quad (\text{B.20})$$

donde A es la matriz de adyacencia, podemos escribir la ecuación de forma matricial como $\mathbf{x}' = \mathbf{A}\mathbf{x}$. Repitiendo el procedimiento t veces, obtenemos el valor

$$\mathbf{x}(t) = \mathbf{A}^t \mathbf{x}(0). \quad (\text{B.21})$$

La medida recibe el nombre debido a que es posible escribir $\mathbf{x}(0)$ como una combinación lineal de los eigenvectores \mathbf{v}_i de la matriz de adyacencias:

$$\mathbf{x}(0) = \sum_i c_i \mathbf{v}_i, \quad (\text{B.22})$$

para valores apropiados de c_i . Así, podemos reescribir B.21 como:

$$\mathbf{x}(t) = \mathbf{A}^t \sum_i c_i \mathbf{v}_i = \sum_i c_i k_i^t \mathbf{v}_i = k_1^t \sum_i c_i \left[\frac{k_i}{k_1} \right]^t \mathbf{v}_i, \quad (\text{B.23})$$

donde k_i son los eigenvalores de A , k_1 es el más grande. Como vimos en el capítulo 3, una de las propiedades de los eigenvalores es que $\frac{k_i}{k_1} < 1 \forall i$, así, los términos de la derecha en B.23 decrecen rápidamente, por lo que al tomar el límite nos quedamos con

$$x(t) \rightarrow c_1 k_1^t \mathbf{v}_1 \text{ si } t \rightarrow \infty, \quad (\text{B.24})$$

es decir, el vector límite de centralidades es proporcional al primer eigenvector de la matriz de adyacencias. Finalmente, la centralidad x_i es proporcional a la suma de las centralidades de los vecinos de i :

$$x_i = \frac{1}{k_i} \sum_j A_{ij} x_j, \quad (\text{B.25})$$

de lo anterior, podemos ver que la centralidad por eigenvectores tiene la propiedad de que puede ser grande por dos razones: si i tiene muchos vecinos, o si i tiene vecinos importantes.

Bibliografía

- Albert, R., Jeong, H., and Barabási, A.-L. (1999). Diameter of the world-wide web. *Nature*, 401. [84](#)
- Albert, R., Jeong, H., and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, 406. [88](#)
- Anderberg, M. R. (1973). *Cluster Analysis for Applications*. Academic Press. [19](#)
- Arabie, P. and Hubert, L. (1994). *Cluster analysis in marketing research*, pages 160–190. John Wiley & Sons. [33](#)
- Aymanns, C. and Georg, C.-P. (2015). Contagious synchronization and endogenous network formation in financial networks. *Journal of Banking And Finance*, 50:273–285. [6](#)
- Banco de México (2018). Reporte sobre el Sistema Financiero. <http://www.banxico.org.mx/publicaciones-y-prensa/reportes-sobre-el-sistema-financiero/%7B0B4861B7-2810-7FE9-88FE-2427C182512A%7D.pdf>. [71](#)
- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512. [83](#)
- Barabási, A.-L. and Pósfai, M. (2016). *Network science*. Cambridge University Press, Cambridge. [71](#), [78](#), [86](#), [87](#), [89](#)
- Battiston, S. and Martinez-Jaramillo, S. (2018). Financial networks and stress testing: Challenges and new research avenues for systemic risk analysis and financial stability implications. *Journal of Financial Stability*, 35:6 – 16. Network models, stress testing and other tools for financial stability monitoring and macroprudential policy design and implementation. [70](#)
- Battiston, S., Puliga, M., Kaushik, R., Tasca, P., and Caldarelli, G. (2012). Debrank: Too central to fail? financial networks, the fed and systemic risk. *Scientific Reports*, 2(541). [79](#), [80](#)

BIBLIOGRAFÍA

- Beale, N., Rand, D. G., Battey, H., Croxson, K., May, R. M., and Nowak, M. A. (2011). Individual versus systemic risk and the regulators 's dilemma. *Proceedings of the National Academy of Sciences*, 108(31):12647–12652. 6
- BIS (2017). Metodología para definición de g-sibs. <https://www.bis.org/publ/bcbs255.pdf>. 3
- Blundell-Wignall, A., Atkinson, P., and Roulet, C. (2014). *Bank business models and the Basel system: Complexity and Interconnectedness*, page 51–74. Anthem Press. 12
- Bravo-Benítez, B. (2012). Análisis empírico de la red de exposiciones interbancarias en méxico y del sistema de pagos electrónicos interbancarios (spei) y su conexión con el riesgo sistémico. 71
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32. 19, 51
- Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA. 52, 53, 54
- Canedo, J. M. D. and Martínez-Jaramillo, S. (2009). A network model of systemic risk: stress testing the banking system. *Intelligent Systems in Accounting, Finance and Management*, 16(1-2):87–110. 4, 80
- Casella, G. and Berger, R. (2001). *Statistical Inference*. Duxbury Resource Center. 106
- Chakraborty, C. and Joseph, A. (2017). Machine learning at central banks. Bank of England working papers 674, Bank of England. 6
- Craig, B. and von Peter, G. (2014). Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23(3):322 – 347. 90, 102, 127
- De Soete, G. and Carroll, J. D. (1994). K-means clustering in a low-dimensional euclidean space. In Diday, E., Lechevallier, Y., Schader, M., Bertrand, P., and Burtschy, B., editors, *New Approaches in Classification and Data Analysis*, pages 212–219, Berlin, Heidelberg. Springer Berlin Heidelberg. 19, 20, 34
- Dorogovtsev, S. N., Mendes, J. F. F., and Samukhin, A. N. (2001). Giant strongly connected component of directed networks. *Phys. Rev. E*, 64(2):025101. 76
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26. 49, 145
- EnhancedDataScience (2017). Machine learning explained: Overfitting. <http://enhancedatascience.com/2017/06/29/machine-learning-explained-overfitting/>. 51
- Erdős, P. and Rényi, A. (1959). On random graphs i. *Publicationes Mathematicae (Debrecen)*, 6:290–297. 83

-
- Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. In *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, pages 17–61. 83
- Farnè, M. and Vouldis, A. (2017). Business models of the banks in the euro area. <https://ideas.repec.org/p/ecb/ecbwps/20172070.html>. 15, 16, 18, 20, 21, 38, 40
- Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1(2):209–230. 115
- Finney, D. (1952). Probit analysis. *Journal of the American Pharmaceutical Association*, 41(11):627–627. 19
- Fisher, R. (1935). *The design of experiments. 1935*. Oliver and Boyd, Edinburgh. 143
- Fornito, A., Zalesky, A., and Bullmore, E. T. (2016). *Fundamentals of brain network analysis*. Elsevier/Academic Press Amsterdam, Boston. 1
- Fricke, D. and Lux, T. (2015). Core–periphery structure in the overnight money market: Evidence from the e-mid trading platform. *Computational Economics*, 45(3):359–395. 69
- FSB-BIS (2017). Lista de g-sibs en 2017. www.fsb.org/wp-content/uploads/P211117-1.pdf. 3
- Gai, P. and Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2120):2401–2423. 80
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd ed. edition. 105, 112
- Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58. 56, 57
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741. 117
- Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2018). *VSURF: Variable Selection Using Random Forests*. R package version 1.0.4. 59
- Gilbert, E. N. (1959). Random graphs. *Ann. Math. Statist.*, 30(4):1141–1144. 83
- Gillespie, C. S. (2015). Fitting heavy tailed distributions: The poweRlaw package. *Journal of Statistical Software*, 64(2):1–16. 123

BIBLIOGRAFÍA

- Goodhart, C. and Wagner, W. (2012). Regulators should encourage more diversity in the financial system. <https://voxeu.org/article/regulators-should-encourage-more-diversity-financial-system>. 6
- Guillén, M. F. (2015). The global economic & financial crisis: A timeline. https://lauder.wharton.upenn.edu/wp-content/uploads/2015/06/Chronology_Economic_Financial_Crisis.pdf. Online. 3
- Haldane, A. and May, R. (2011). Systemic risk in banking ecosystems. *Nature*, 469:351–355. 6
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in Statistics. Springer, 2 edition. 49
- Henggeler-Müller, J. (2006). *The Potential for contagion in the Swiss interbank market*. PhD thesis, Universität Basel. 77
- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer Publishing Company, Incorporated, 1st edition. 108
- Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2):109 – 137. 120
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer New York. 31, 49, 144
- Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis (6th Edition)*. Pearson, 6 edition. 23, 28
- Kabacoff, R. (2017). Tree-based models. <https://www.statmethods.net/advstats/cart.html>. 53, 55
- Karrer, B. and Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107. 122
- Krugman, P. (2012). Economics in the crisis. <https://krugman.blogs.nytimes.com/2012/03/05/economics-in-the-crisis/>. Online. 3
- Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Elsevier Science. 111
- León, C. (2017). Banks in colombia: how homogeneous are they? Working Paper Series 1022, Banco de la República. 5, 15, 20
- Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18–22. 59

-
- Lucas, A., Schaumburg, J., and Schwaab, B. (2017). Bank business models at zero interest rates. Working Paper Series 2084, European Central Bank. [15](#), [17](#), [18](#), [20](#), [141](#)
- MacEachern, S. N. (2016). Nonparametric bayesian methods: a gentle introduction and overview. *Communications for Statistical Applications and Methods*, 23:445–466. [114](#), [115](#)
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif. University of California Press. [15](#), [29](#), [31](#)
- Markos, A., D’Enza, A. I., and van de Velden, M. (2018a). *clustrd: Methods for Joint Dimension Reduction and Clustering*. R package version 1.2.2. [21](#)
- Markos, A., D’Enza, A. I., and van de Velden, M. (2018b). *clustrd: Methods for Joint Dimension Reduction and Clustering*. R package version 1.2.2. [37](#)
- Martinez-Jaramillo, S., Alexandrova-Kabadjova, B., Bravo-Benitez, B., and Solórzano-Margain, J. P. (2014). An empirical study of the mexican banking system’s network and its implications for systemic risk. *Journal of Economic Dynamics and Control*, 40:242 – 265. [77](#), [125](#)
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press. [106](#)
- Newman, M. (2010). *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA. [71](#), [79](#), [87](#), [90](#), [153](#)
- Newman, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701. [90](#)
- Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087. [105](#), [119](#), [120](#), [122](#)
- Pareto, V., Busino, G., and Bousquet, G.-H. (1964). *Cours d’économie politique*. Droz. [85](#)
- Peixoto, T. P. (2014). The graph-tool python library. *figshare*. [123](#)
- Peixoto, T. P. (2017). Bayesian stochastic blockmodeling. Workingpaper, arXiv. 42 pages, 16 figures, Chapter in “Advances in Network Clustering and Blockmodeling”, edited by P. Doreian, V. Batagelj, A. Ferligoj, (Wiley, New York, 2018 [forthcoming]). [123](#)

- Poledna, S., Martínez-Jaramillo, S., Caccioli, F., and Thurner, S. (2018). Quantification of systemic risk from overlapping portfolios in the financial system. Papers 1802.00311, arXiv.org. [16](#)
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. [9](#), [21](#), [37](#)
- Robert, C. and Casella, G. (2011). A short history of markov chain monte carlo: Subjective recollections from incomplete data. *Statistical Science*, 26(1):102–115. [118](#)
- Rodriguez, A. and Müller, P. (2013). Nonparametric bayesian inference. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 9:i–110. [114](#)
- Rossi, P., Allenby, G., and McCulloch, R. (2012). *Bayesian Statistics and Marketing*. Wiley Series in Probability and Statistics. Wiley. [118](#)
- Rossum, G. (1995). Python reference manual. Technical report, CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands, The Netherlands. [9](#)
- Rosvall, M. and Bergstrom, C. T. (2010). Mapping change in large networks. *PLOS ONE*, 5(1):1–7. [42](#)
- Silva, T. C., de Souza, S. R. S., and Tabak, B. M. (2016). Network structure analysis of the brazilian interbank market. *Emerging Markets Review*, 26:130 – 152. [75](#), [152](#)
- Snijders, T. A. and Nowicki, K. (1997). Estimation and prediction for stochastic block-models for graphs with latent block structure. *Journal of Classification*, 14(1):75–100. [120](#)
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):111–147. [49](#), [57](#)
- The Department of the Treasury, USA (2012). The financial crisis response in charts. https://www.treasury.gov/resource-center/data-chart-center/Documents/20120413_FinancialCrisisResponse.pdf. Online. [3](#)
- Timmerman, M. E., Ceulemans, E., Kiers, H. A., and Vichi, M. (2010). Factorial and reduced k-means reconsidered. *Computational Statistics And Data Analysis*, 54(7):1858 – 1871. [33](#), [35](#), [37](#)
- van Lidth de Jeude, J., Caldarelli, G., and Squartini, T. (2018). Detecting core-periphery structures by surprise. *CoRR*, abs/1810.04717. [92](#), [101](#)
- Vichi, M. and Kiers, H. A. (2001). Factorial k-means analysis for two-way data. *Computational Statistics And Data Analysis*, 37(1):49 – 64. [19](#), [20](#), [34](#)

- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. 97
- Wikipedia (2019a). Alfréd Rényi — Wikipedia, the free encyclopedia. <http://es.wikipedia.org/w/index.php?title=Alfr%C3%A9d%20R%C3%A9nyi&oldid=104591623>. [Online; accessed 25-June-2019]. 71
- Wikipedia (2019b). Paul Erdős — Wikipedia, the free encyclopedia. <http://es.wikipedia.org/w/index.php?title=Paul%20Erd%C5%91s&oldid=107945575>. [Online; accessed 25-June-2019]. 71