



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**Aplicación del método de visibilidad en redes para
el pronóstico de Indicadores Económicos**

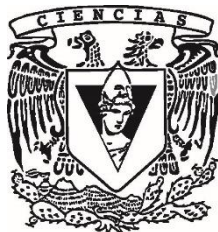
T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A:

RICARDO HERRERA MARTÍNEZ



**TUTORA
DRA. BIBANA OBREGÓN QUINTANA**

**CIUDAD UNIVERSITARIA, CD. MX.
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.

1. Datos del alumno
Herrera
Martínez
Ricardo

2. Datos del autor
Dra.
Bibiana
Obregón
Quintana

3. Datos del sinodal 1
Dr.
Yuri
Salazar
Flores

4. Datos del sinodal 2
Dr.
Lev
Guzmán
Vargas

5. Datos del sinodal 3
Act.
Victor Manuel
Solís
Nájera

6. Datos del sinodal 4
Dr.
Luis Alberto
Quezada
Téllez

7. Datos del trabajo escrito
Aplicación del método de visibilidad en redes para el pronóstico de Indicadores
Económicos
100 p.
2019

“ALGDGADU”.

AUTOR

FIRMA

DEDICATORIA

A mis padres, que sin su amor JAMÁS hubiera llegado tan lejos. Los amo infinitamente. Gracias por darme la oportunidad de vivir a su lado. Gracias por haber sido ustedes y permitirme ser yo.

A mi hermana, por ser la mejor cómplice. Amor, orgullo y admiración sin límite. Eres lo mejor que me dio la vida.

A mi novia, Carla Alondra Alonso Peniche, porque me enseñó que el latido del corazón es más veloz estando con ella. Te amo infinitamente.

A mis amigos, porque caminar con ellos de la mano en la oscuridad es mejor que caminar solo en la luz. Súper S., P., YMCA, EMA, EY.

A Grupo ARFil, por haberme enseñado lo que no se muestra en la Universidad, amor a lo que se hace.

Agradecimientos

Agradezco infinitamente el apoyo que me ha brindado cada una de las personas con las que me he encontrado en la vida.

Agradezco a la Dra. Bibiana Obregón Quintana por su dedicación y esfuerzo. Le agradezco el enseñarme a confiar en mis conocimientos. Sin ella, este trabajo nunca hubiera salido adelante.

Agradezco a mi tío Fernando Enrique Loyola Ahedo, a Silvia González Mesa y a Javier Pérez Tapia por el apoyo incondicional en mi trayectoria profesional.

Agradezco a mis HH:. por haberme enseñado el valor de ser siempre un aprendiz en esta vida. Por ellos y por todos.

Agradezco a Diego Isaías Prietto Luna y a Miguel Ángel Espinosa Ochoa porque sin ellos mi vida no hubiera sido igual.

Agradezco a José Morán, Danae Valdés y Shamanta Morelos porque sin ellos la universidad no hubiera sido igual.

Agradezco a mis amigos de siempre: Yolanda Morales, Aldo Ruíz, Daniel Díaz, Arturo Enriquez, Cristina y Fernanda González, Diego Ociel y Humberto Sandoval, Rafael Padilla y Christian González. En fin, cuantos nombres no podría meter a la lista.

Agradezco a la H.H. Universidad Nacional Autónoma de México y en especial a la Facultad de Ciencias por haberme brindado los conocimientos necesarios para enfrentarme al futuro. Un orgullo ser parte de esta universidad. **¡Goya!**

Aplicación del método de visibilidad en redes para el pronóstico de Indicadores Económicos

”Deja que los perros ladren Sancho amigo, es señal que vamos pasando.”

Don Quijote de la Mancha

Índice general

Índice de figuras	IX
Índice de cuadros	XIII
1. Antecedentes	3
1.1. Producto Interno Bruto de México	4
1.2. Ocupación en México	6
1.3. Inflación	7
2. Marco Teórico	9
2.1. Historia	9
2.2. Conceptos Básicos de la Teoría de Redes	13
2.3. Modelos de Redes y Conceptos Avanzados	17
2.4. Estadística y Series de Tiempo	21
2.4.1. Series de Tiempo [12]	21
2.4.2. Modelos de Pronóstico	23
3. Método de la Gráfica de Visibilidad y Método de Pronóstico	27
3.1. Gráfica de la Visibilidad	27
3.1.1. Gráfica de la Visibilidad Vertical	30
3.1.2. Gráfica de la Visibilidad Horizontal	31
3.2. Método de Pronóstico (Propuesta)	33
4. Datos, Metodología y Resultados	41
4.1. Datos	41

4.2.	Estadística Descriptiva de las Series	43
4.2.1.	Estadística Descriptiva de PIB	43
4.2.2.	Estadística Descriptiva de Ocupación (%)	44
4.2.3.	Estadística Descriptiva de Inflación Acumulada	46
4.3.	Generación de la gráfica de visibilidad vertical de las series	47
4.3.1.	Visibilidad y Red del PIB	47
4.3.2.	Visibilidad y Red de Ocupación	52
4.3.3.	Visibilidad y Red de Inflación Acumulada	55
4.4.	Resultados de la Aplicación del Método de Pronóstico	58
4.4.1.	Pronóstico del PIB	58
4.4.2.	Pronóstico de Ocupación (%)	59
4.4.3.	Pronóstico de Inflación Acumulada	60
4.5.	Resultados: Comparación de Pronósticos	62
4.5.1.	Comparación de PIB	62
4.5.2.	Comparación de Ocupación (%)	63
4.5.3.	Comparación de Inflación	65
	Conclusiones	71
	Bibliografía	73
	Anexos	77

Índice de figuras

1.1. PIB México de marzo de 2007 a septiembre de 2016.	5
1.2. Ocupación en México de enero de 2010 a marzo de 2017. . .	6
1.3. Inflación acumulada de enero de 2014 a marzo 2017.	8
2.1. Derecha puentes de Königsberg (Kaliningrado, Rusia). Iz- quierda representación de la red.	10
2.2. Recableado para obtener un modelo de mundo pequeño. . .	12
2.3. Tipos de redes.	13
2.4. Red conexas.	14
2.5. Distribución de grado.	15
2.6. Matriz de adyacencia.	16
2.7. Red aleatoria con probabilidad p	18
2.8. Distribución de Poisson.	18
2.9. Red de mundo pequeño.	19
2.10. Red de mundo pequeño mediante la interpolación de redes.	19
2.11. Red de libre escala.	20
2.12. Distribución exponencial.	20
2.13. Serie de tiempo del PIB.	22
3.1. Ejemplo de una serie de tiempo de 20 datos con su respectiva red derivada del algoritmo de la gráfica de visibilidad [6]. . .	28

3.2.	La gráfica de visibilidad de una serie de tiempo es invariante ante transformaciones. (a) Serie de tiempo original mapeada mediante el método de la gráfica de visibilidad. (b) Traslación de la serie. (c) Escalamiento vertical. (d) Escalamiento horizontal. (e) Cambio de tendencia lineal en la serie. Como se puede observar en la gráfica inferior de esta figura, la gráfica de visibilidad permanece invariante [6].	29
3.3.	Método de visibilidad vertical.	31
3.4.	Método de visibilidad horizontal.	32
3.5.	Arcos atrasados del nodo 4 (en verde) y arcos adelantados del nodo 4 (en rojo).	35
3.6.	Ejemplo de los vectores E y E_1	38
4.1.	Serie de tiempo de Producto Interno Bruto	44
4.2.	Serie de tiempo de Ocupación en México	45
4.3.	Serie de tiempo de Inflación Acumulada	47
4.4.	Gráfica de visibilidad del PIB	48
4.5.	Red del PIB	48
4.6.	Red del PIB disconexa (nodo 4).	50
4.7.	Red del PIB disconexa (nodo 2).	50
4.8.	Distribución de grado de la red del PIB.	51
4.9.	Gráfica de visibilidad de Ocupación.	52
4.10.	Red de Ocupación.	52
4.11.	Red de la ocupación en México, disconexa nodo 36.	54
4.12.	Distribución de grado de Ocupación en México.	55
4.13.	Gráfica de visibilidad de inflación acumulada.	55
4.14.	Red de inflación acumulada.	56
4.15.	Distribución de grado de la inflación acumulada.	57
4.16.	Pronóstico del PIB y_{t+1} con intervalos de confianza al 95.5%.	58
4.17.	Pronóstico del PIB con intervalos de confianza (5 periodos).	59
4.18.	Pronóstico del primer periodo y visibilidad de ocupación.	59
4.19.	Pronóstico de ocupación con intervalos de confianza (5 periodos).	60
4.20.	Pronóstico e intervalo de confianza inflación acumulada.	61
4.21.	Pronóstico de periodos de inflación acumulada.	61

4.22. Izquierda: Pronóstico e intervalo de confianza del método propuesto. Derecha: Pronóstico e intervalo de confianza del modelo ARIMA.	63
4.23. Izquierda: Pronóstico e intervalo de confianza del modelo propuesto. Derecha: Pronóstico e intervalo de confianza del modelo ARIMA.	65
4.24. Izquierda: Pronóstico e intervalo de confianza del modelo propuesto. Derecha: Pronóstico e intervalo de confianza del modelo ARIMA.	66
4.25. Backtesting: PIB	67
4.26. Backtesting: Ocupación en México	68
4.27. Backtesting: Inflación Acumulada en México	69

Índice de cuadros

4.1. Estadísticos Descriptivos del PIB	43
4.2. Estadísticos Descriptivos de la Ocupación en México	45
4.3. Estadísticos Descriptivos de la Inflación Acumulada	46
4.4. Métricas de la red del PIB.	49
4.5. Características de la red de la ocupación en México.	53
4.6. Características de la red de la inflación acumulada.	56
4.7. Valores de pronóstico método propuesto y modelo ARIMA.	62
4.8. Diferencias de pronóstico método propuesto y modelo ARIMA.	62
4.9. Valores de pronóstico método propuesto y modelo ARIMA.	64
4.10. Diferencias de pronóstico método propuesto y modelo ARIMA.	64
4.11. Valores de pronóstico método propuesto y modelo ARIMA.	65
4.12. Diferencias de pronóstico método propuesto y modelo ARIMA.	66

Introducción

Si observamos nuestro alrededor podemos darnos cuenta de una gran cantidad de sistemas que pueden modelarse como redes. Ejemplos de estas son las redes de transporte, drenaje, la infraestructura eléctrica de las ciudades, entre otras. Además, podemos también observarlas en fenómenos más complejos como las redes de narcotráfico, el contagio de enfermedades y en la interacción de los países en la economía global.

Estas representaciones se utilizan de forma general en diversos problemas, como en los procesos de inicio a fin de un producto, la distribución, la planificación de proyectos (como redes de tiempo), la localización de instalaciones (como sucursales para reducir costos e incrementar productividad), la administración de recursos y la planificación financiera. En fin, casi cualquier sistema dinámico podría ser representado mediante una red.

Se han desarrollado herramientas y metodologías que ayudan a comprender sistemas que se pueden modelar como redes y, aunque no todos los problemas que derivan de ellos, pueden ser resueltos de forma eficiente por su complejidad. La aparición de algunos algoritmos de optimización han tenido efectos positivos para la modelación informática de estos fenómenos, apoyados en las estructuras de datos y la manipulación eficiente de sus relaciones.

Si se retoma la idea de la modelación de sistemas complejos como la interacción de las economías globales, también se pueden modelar como redes los diversos indicadores económicos para analizar el comportamiento financiero de un país (creciente, estable o decreciente), mediante el uso de las series de tiempo. La presente tesis utiliza como insumo principal algunas de las series de indicadores económicos más importantes de México. Este trabajo se basó en el artículo *Statistical validation of financial time series via visibility graph*, de Matteo Seraffino *et al* [21], en donde se menciona el uso del método de la gráfica de visibilidad aplicada a series de tiempo financieras y económicas, para determinar la probabilidad de una crisis financiera.

Así, el *objetivo* central de esta tesis es presentar un método de pronósti-

co de series de tiempo utilizando la gráfica de visibilidad, para observar patrones que se encuentran en las relaciones existentes entre los datos a través del tiempo. También, permitirá observar si los datos estimados mediante este método son consistentes con los que resultarían de un modelo estadístico tradicional. Es importante resaltar que, es un método no estadístico que muestra una nueva forma de pronosticar series económicas.

La forma de lograr un análisis adecuado, será con base en la modelación de distintas series de tiempo de algunos de los indicadores económicos, para estudiarlos con algoritmos comúnmente usados en la teoría de redes. Así, se encontrará una forma distinta de analizar las series de tiempo en comparación con métodos tradicionales de estadística.

También, se compararán los resultados obtenidos contra pronósticos generados con modelos estadísticos como los Autoregresivos Integrados de Medias Móviles (descripción en el capítulo 1), con el fin de observar si los pronósticos tienen diferencias significativas.

Ahora bien, esta tesis está organizada como sigue, en el capítulo 1 se mencionan los antecedentes de los indicadores económicos y las series financieras. En el segundo capítulo se presenta un resumen histórico de la teoría de redes y se introducen conceptos básicos, tanto de la teoría de redes como de la estadística. En el capítulo 3 se presenta de manera formal el método de la gráfica de visibilidad y el método de pronóstico propuesto mediante el uso de herramientas de la teoría de redes. Por último, en el capítulo 4 se muestran las fuentes de información de donde se obtuvieron las series económicas, la metodología de la gráfica de la visibilidad y los resultados de la aplicación del método de pronóstico propuesto. Como anexos se agregan los datos y el código, en R.

Capítulo 1

Antecedentes

El estudio y análisis de indicadores financieros y económicos es fundamental para la salud de la economía de un país, pues basados en estos factores, los agentes del mercado (las empresas, accionistas y gobiernos) toman decisiones que tienen impacto directo dentro de la sociedad. La economía se define como [18]:

[...el estudio de la manera en que las sociedades utilizan recursos escasos para producir mercancías valiosas y distribuirlas entre los distintos individuos...] ,

y tiene diversas herramientas que permiten estudiar a fondo cada una de sus partes, para entender los fenómenos que repercuten de forma directa e indirecta en la vida de cada uno de los integrantes del sistema económico. Existen diversas áreas de investigación en la economía como, por ejemplo, la economía política, microeconomía, macroeconomía, economía internacional y la economía financiera, entre otras. En este capítulo se introduce a la economía financiera como objeto de estudio utilizando uno de sus componentes principales, las series de tiempo.

La economía financiera es, la rama de la economía que analiza la forma en que inversionistas racionales deben invertir sus fondos para alcanzar sus objetivos de la mejor manera posible. La economía financiera especialmente

se centra en la medida y la valoración del riesgo y la asignación temporal de recursos monetarios a opciones de inversión y financiamiento en entornos inciertos y volátiles. Por esto, las series financieras e indicadores económicos son generalmente modelados como series de tiempo, es decir, como realizaciones de un proceso estocástico a tiempo discreto, y por esto se utilizan herramientas estadísticas para predecir fluctuaciones futuras [18].

Es por lo anterior que, uno de los principales fenómenos de interés en esta área es el pronóstico de valores financieros y económicos, con el fin de prever o tomar ventaja de eventos que afecten de forma directa o indirecta los bienes y servicios en la economía.

Existe gran diversidad de indicadores financieros que permiten observar la salud económica de un país. Dentro de los principales se encuentran el producto interno bruto, la balanza comercial (exportaciones e importaciones), el tipo de cambio, la inflación, el desempleo, entre otras.

1.1. Producto Interno Bruto de México

La medida más clara de la producción total de una economía es el Producto Interno Bruto (PIB). El PIB es la medida del valor de mercado de todos los bienes y servicios finales (tales como producción petrolera, automóviles, conciertos, servicios de consultoría, entre otros), producidos en un país durante un año. Este indicador es la suma del valor monetario del consumo (C), la inversión bruta (I), el gasto público (G) y las exportaciones netas (X) dentro de un país en un año cualquiera [18]:

$$PIB = C + I + G + X.$$

El PIB se utiliza principalmente para medir el desempeño global de una economía (ver figura 1.1). Este indicador tiene diversos factores que afectan de forma directa o indirecta las fluctuaciones de la serie de tiempo del indicador, como por ejemplo las tasas de interés, el tipo de cambio, la balanza comercial y la política monetaria.

Tanto en el método de pronóstico propuesto (ver sección 3.2) como el método estadístico ($ARIMA(q,d,p)$), obsérvese el capítulo 2 para más detalles), se utiliza para fines de pronóstico el principio econométrico *Ceteris*

Paribus, indicando que las variables que pudiesen afectar a la serie permanecen constantes y sólo se utilizará la información numérica de la serie (suponiendo que toda la información adicional está integrada en estos datos, ver figura 1.1).

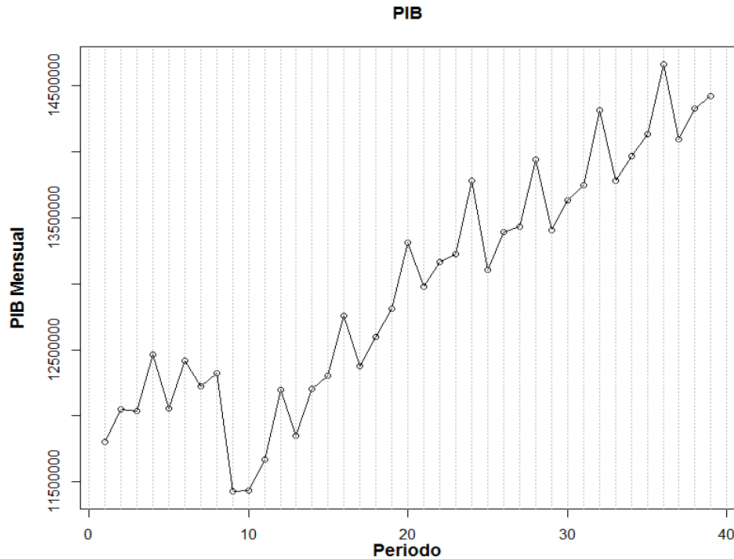


Figura 1.1: PIB México de marzo de 2007 a septiembre de 2016.

Fuente: Elaboración propia con datos de <https://www.inegi.org.mx/sistemas/bie/>

Puesto que este indicador permite observar si la economía creció, o no, en un año, los gobiernos buscan pronosticar esta serie con el fin de desarrollar modelos de presupuesto más acertados y así, optimizar los recursos que se invertirán en cada una de sus políticas públicas [18].

Debido a que este indicador es importante para la toma de decisiones, se han realizado diversos estudios estadísticos con el fin de comprender su comportamiento a través del tiempo y estimar de mejor forma su movimiento futuro. Es por esto que se utiliza la serie del PIB para el estudio.

1.2. Ocupación en México

En términos prácticos, la Ocupación es el porcentaje de población de la Población Económicamente Activa (PEA, que está compuesta por todos los habitantes en edad laboral que o bien trabaja en un empleo remunerado (población ocupada) o bien se halla en plena búsqueda de empleo), que tienen un trabajo subordinado o que ejercen una actividad por cuenta propia. Este indicador es importante dentro de la economía en tanto que muestra el bienestar laboral de un país y, derivado de este, el bienestar social [18].

En la figura 1.2 se muestra un histórico del porcentaje de la población ocupada en México de enero de 2010 a marzo de 2017.

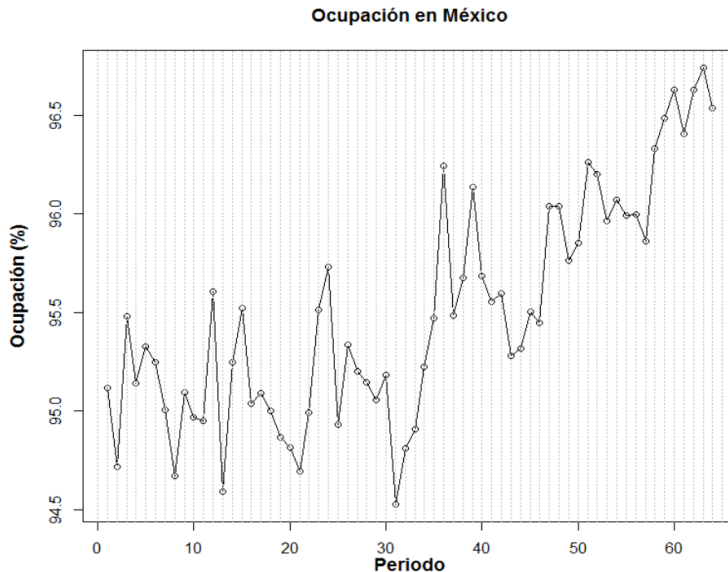


Figura 1.2: Ocupación en México de enero de 2010 a marzo de 2017.

Fuente: Elaboración propia con datos de <https://www.inegi.org.mx/sistemas/bie/>

Se decidió utilizar este indicador debido a que cumple con que es un indicador importante para la economía de cualquier país y también, con

que no tiene un comportamiento tan homogéneo como la serie del PIB. Lo anterior, para analizar si existen diferencias importantes en la aplicación del método de pronóstico propuesto.

1.3. Inflación

El concepto de inflación se entiende como el aumento porcentual generalizado de bienes y servicios durante un periodo de tiempo determinado [18]. Este porcentaje se calcula como se menciona a continuación:

$$\pi_t = 100 * \frac{P_t - P_{t-1}}{P_{t-1}}.$$

En donde,

π_t es la inflación del año t ,

P_t son los precios de bienes o servicios en el año t ,

P_{t-1} son los precios de bienes o servicios en el año $t - 1$.

Existen dos fenómenos en donde este indicador llega a sus niveles extremos.

1. *Deflación*: Hay una deflación cuando los precios bajan (lo que significa que la tasa de inflación es negativa). Una deflación puede indicar un incremento en el desempleo.
2. *Hiperinflación*: Existe una hiperinflación cuando se observa un alza en el nivel de precios de mil a un millón por ciento anual. Una hiperinflación puede entenderse como una inestabilidad financiera del país.

La estabilidad de precios es importante porque, para que un sistema de mercado funcione, es necesario que los precios reflejen adecuadamente la información sobre la escasez relativa de los bienes. Además, este indicador tiene un efecto en la producción de bienes y servicios dentro de una economía y es por esto que la correlación con la dinámica económica nacional es positiva [18].

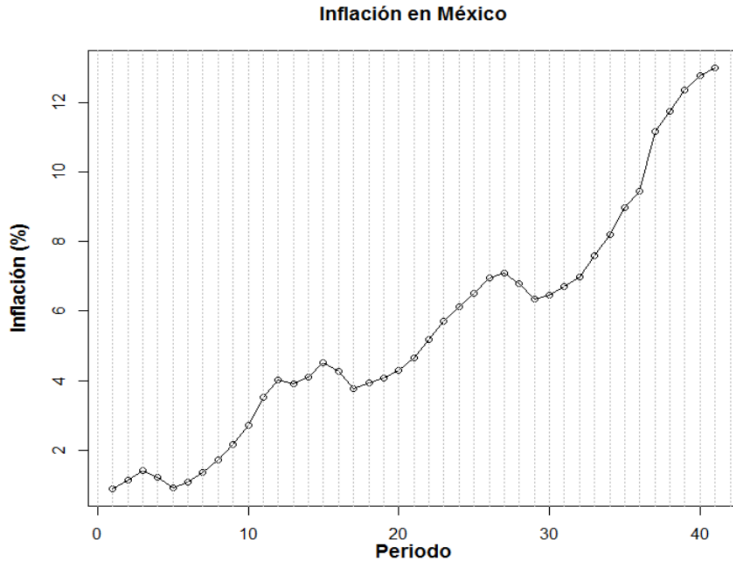


Figura 1.3: Inflación acumulada de enero de 2014 a marzo 2017.

Fuente: Elaboración propia con datos de <https://www.inegi.org.mx/sistemas/bie/>

En la figura 1.3 se puede observar el comportamiento de la inflación acumulada en México desde enero 2014 a marzo 2017.

Expertos en la materia concuerdan en que un modelo económico estable deberá tener crecimientos inflacionarios suaves, que modelen de forma eficiente el comportamiento de oferta y demanda de productos [18].

Capítulo 2

Marco Teórico

En este capítulo se presentan algunos acontecimientos históricos relevantes de la teoría de redes, personajes importantes que han aportado elementos fundamentales en esta teoría y también, se definen conceptos básicos de redes y de estadística que se utilizarán más adelante.

2.1. Historia

Las redes modelan sistemas reales con características complejas y, generalmente, dinámicas para el estudio de sus componentes.

Aunque no se considera que la teoría de redes sea una rama de la teoría de gráficas, ambas comparten sus fundamentos topológicos iniciados por Leonhard Euler. Uno de los trabajos del matemático Leonhard Euler en 1736, conocido como el Problema de los Siete Puentes de Königsberg, Prusia (hoy Kaliningrado, Rusia), es considerado el primer resultado concreto de la teoría de gráficas.

Este problema radica en que la ciudad de Königsberg estaba dividida por el río Pregel en cuatro partes conectadas por siete puentes. El objetivo de Euler era encontrar un recorrido utilizando cada uno de los puentes una sola vez y que se visitaran las cuatro partes. El matemático, modeló el problema representando las partes de tierra como nodos y los puentes

como arcos creando así la primera aproximación a las figuras topológicas llamadas redes (ver figura 2.1) [10].

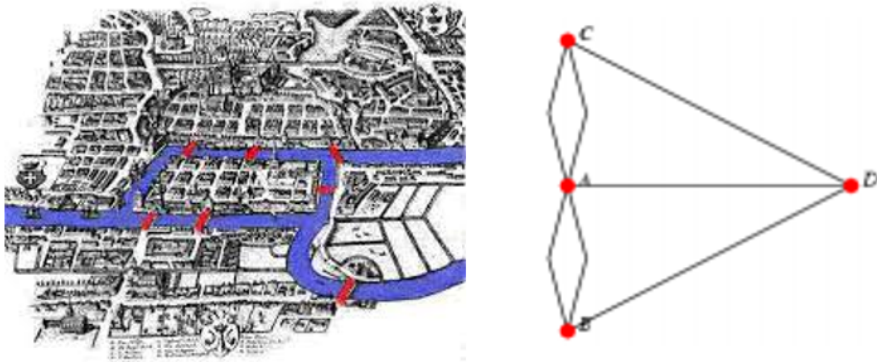


Figura 2.1: Derecha puentes de Königsberg (Kaliningrado, Rusia). Izquierda representación de la red.

A partir de estos estudios se comenzaron a modelar fenómenos con base en esta topología y, con ello se desarrolló la teoría que daría pie al estudio de las redes.

Frigyes Karinthy en la década de los 30's del siglo XX, escribió un libro titulado *Everything is Different* como una conglomeración de cuentos cortos. Uno de ellos lo nombra *Chain* (Cadena) y en este argumentaba que, como Julio Verne establece en *Viaje alrededor del mundo en 80 días*, el mundo se estaba haciendo pequeño en aspectos comunicativos y sociales. En este cuento se propone que cualquier persona, por la interconexión de sus conocidos, se encuentra a cinco personas de separación de cualquier otra en el mundo. Como ejemplo, menciona la posibilidad de conectar a un ganador del Premio Nobel con él mismo, por medio de cinco personas de separación y lo mismo podría cumplirse con un trabajador de una fábrica. Lo anterior, muestra que no importando el estatus social que se tenga siempre se podrá llegar, en promedio, por medio de cinco personas a cualquier otra [10].

Más tarde, en 1967, el psicólogo Stanley Milgram demostró de forma experimental que, como Frigyes Karinty había predicho, con un pequeño error, en promedio se tienen a seis personas de distancia entre uno y cualquier persona del mundo. Actualmente, conocemos a este efecto como *seis grados de separación*. Stanley Milgram desarrolló un experimento en la Universidad de Harvard, en donde envió 296 cartas a personas en Kansas y Nebrasaka, ambos estados pertenecientes a los Estados Unidos de América. Cada una de las personas elegidas por Milgram, debían reenviar la carta a un destinatario en Boston o a alguien a quien ellos consideraran tenía posibilidad de conocer un destinatario específico en aquella ciudad. El experimento terminó con 64 cartas en Boston, dando el resultado de que en promedio se llegó con 6 personas de un estado a otro. Si regresamos a 1950, Paul Erdős y Alfréd Rényi ambos matemáticos húngaros, propusieron a las redes aleatorias como espacios topológicos, en donde cada nodo es enlazado a otro con una probabilidad prefijada p , en cada tiempo establecido, así, si el tiempo crece, la red tendrá a la mayor parte de los nodos interconectados formando un *componente gigante*. Además, los matemáticos húngaros presentaron por primera vez modelos de redes como objetos estocásticos más que como deterministas. Fue el inicio de las redes aleatorias o complejas, dándole nacimiento formal a la teoría de redes. Derivado de esto, se iniciaron estudios matemáticos para la descripción de redes estocásticas que pueden observarse, por ejemplo, en los fenómenos como influencia y propagación de enfermedades [10].

A partir de las aportaciones de Paul Erdős y Alfréd Rényi, se comenzaron a modelar sistemas dinámicos como series estocásticas para analizar su comportamiento. Se utilizan principalmente para observar cambios de tendencia, ciclicidad y estacionalidad, además se ha estudiado el componente aleatorio con el fin de utilizarlo para la estimación de proyecciones.

En el año de 1998, Duncan Watts y Steven Strogatz propusieron el modelo de *mundo pequeño*. Este tipo de redes tienen la particularidad de que tienen un coeficiente de clustering alto y una distancia media pequeña (definiciones en la sección 2.2). El algoritmo propuesto por Watts y Strogatz comienza con una red en la que existen N nodos en forma de anillo, en esta red cada nodo está conectado a sus primeros k vecinos. Luego se realiza un recableado, esto es, se desconectan ciertos arcos y se vuelven a reconectar

a otros nodos con una probabilidad prefijada p (ver figura 2.2).

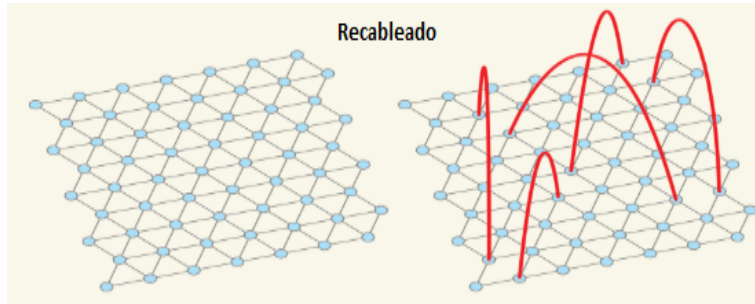


Figura 2.2: Recableado para obtener un modelo de mundo pequeño.

En 1999, Albert-Lászlo Barabási y Réka Albert comenzaron a observar en redes reales, que existen nodos con mayor conexión que otros, dando inicio al estudio de redes de libre escala [10].

Las redes de libre escala son estudiadas en diversas áreas. En mercadotecnia se buscan a los nodos con mayor número de conexiones para tener un impacto mayor en la publicidad dirigida. También se utilizan en sistemas de distribución, para hacer eficientes las rutas que debe tomar cierto producto desde algún nodo de distribución hasta el nodo de venta. En el sector financiero también se pueden observar este tipo de redes. Si una serie de tiempo es mapeada a esta topología y muestra nodos con gran cantidad de conexiones, además de que tienen similar número de arcos, se puede inferir que existe ciclicidad en la serie. Lo anterior debido a que se presentaría una conexión similar en los puntos máximos y/o mínimos y se traduciría en grados altos (definición en la sección 2.2.) [2].

Actualmente, el gran avance computacional ha ayudado a desarrollar modelos más robustos, que buscan modelar redes complejas de forma más eficiente y a realizar cálculos iterativos que serían imposibles generarlos sin esta herramienta.

Cabe señalar que el sistema económico mundial, está compuesto por entidades que intercambian activos en tiempo real mediante sistemas informáticos. Las cámaras de compensación son las encargadas de la transferencia de los bienes de un actor con otro. Si se mapeara el sistema económi-

co mundial tendería a formar una red de libre escala, ya que las cámaras de compensación serían nodos con gran cantidad de enlaces. Uno de los grandes retos es proteger a estas entidades de ataques cibernéticos de gran alcance, para evitar pérdidas de información y bienes dentro del sistema financiero. También se han trabajado redes en el mercado de derivados, más específico con los instrumentos Over The Counter (OTC). Debido a que no se cuenta con información del trading¹ de estos instrumentos, se utilizan las redes para observar el comportamiento de los precios de los instrumentos listados e inferir el precio de los OTC [2].

2.2. Conceptos Básicos de la Teoría de Redes

Una red es un conjunto de nodos unidos por arcos. La notación para describir una red es $G = \{N, A\}$, donde N es el conjunto de nodos, y A es el conjunto de arcos. Estas redes pueden tener o no dirección en sus arcos formando redes *dirigidas* y *no dirigidas* (ver figura 2.3).

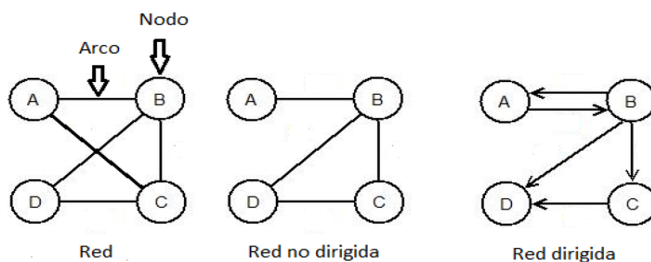


Figura 2.3: Tipos de redes.

Las redes, como sistemas complejos, se han estudiado principalmente desde un punto de vista estructural basado en el comportamiento de sus relaciones. Puesto que en el método de pronóstico propuesto, está basado en el estudio y análisis de las conexiones de una red, es necesario definir conceptos que nos den más profundidad en este tema. Así, definimos *camino*

¹Definición de trading : es negociar y/o especular en los mercados financieros con el objetivo de generar rentabilidades en el tiempo. A diferencia de una “inversión”, el trading es más de corto plazo; semanas, días, horas e incluso minutos [18].

como una sucesión de arcos adyacentes y diremos que una red está *conectada* si existe un camino entre cada par de nodos. A esta red la llamaremos *conexa* y llamaremos *conectividad* a esta característica de la red.

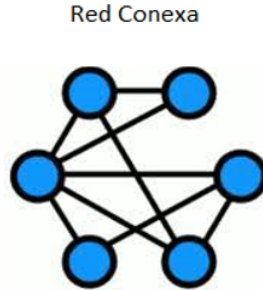


Figura 2.4: Red conexa.

A una red que no es conexa la llamaremos *disconexa* y esta puede dividirse en lo que denotaremos como *componentes*, cada uno de los cuales están internamente conectados. En una red no dirigida y disconexa, los componentes serán disjuntos, es decir, no existirá arco alguno entre ellos. En cambio, en una red dirigida, puede existir un arco entre sus componentes pero únicamente con una dirección, por lo que podrá entrar del componente A al componente B de la red pero no podrá regresar. Además, la conectividad puede estar asociada al número de arcos de la red, ya que, entre más arcos existan en la red es más probable encontrar una ruta de un nodo i a un nodo j [3].

A la cantidad de arcos que conectan al nodo i , se le llama *grado* y a los nodos conectados, a este nodo, se les conoce como *nodos vecinos* [9].

Se denota *ruta* a una secuencia alternada y ordenada de nodos y arcos del nodo i al nodo j , en donde todos los nodos son distintos. La *ruta más corta* (ruta geodésica) de un nodo i a un nodo j , será aquella de todas las rutas que contenga menor cantidad de arcos. La *distancia media* será el promedio aritmético de las distancias entre todos los pares de nodos (i, j) de la red.

Por otro lado, la *centralidad intermedia* es una métrica que nos ayuda a determinar la importancia de un nodo. Así, el nodo que tendrá mayor centralidad será aquel por donde pasen una mayor cantidad de rutas más cortas de cualquier nodo i a cualquier nodo j , este nodo será importante para conectar subgrupos y tendrá relevancia en las conexiones de información de forma eficiente [9].

$$C_i = \sum_{j,k \neq i} \frac{g_{j,k}(i)}{g_{j,k}}, \quad (2.1)$$

donde,

$C(i)$ es la centralidad intermedia del i –ésimo nodo,

$g_{j,k}(i)$ son las rutas más cortas de nodo j a k , que pase por el nodo i ,

$g_{j,k}$ ruta más corta de nodo j a k .

Una de las formas más sencillas de estudiar una red es mediante el comportamiento del grado de los nodos, esto es, la conectividad. Si se hace una gráfica de frecuencia del número de nodos por cada grado se tendría una *distribución de grado*, que sería entendido igualmente como la distribución de probabilidad de un grado en la red. La distribución de grado en una red se representa habitualmente como $P(k)$, y es definida como la fracción de nodos en la red con un cierto grado k (ver figura 2.5) [8].

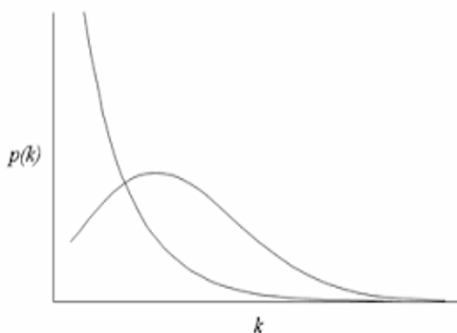


Figura 2.5: Distribución de grado.

La *densidad* se define como el grado de conectividad total de la red

vista como porcentaje, i.e., vista como las relaciones existentes dentro de la red entre relaciones posibles [9].

$$Densidad = \frac{\text{Conexiones Reales}}{\text{Conexiones Potenciales}}, \quad (2.2)$$

donde,

$$\text{Conexiones Potenciales} = \frac{N * (N - 1)}{2}, \quad (2.3)$$

donde N es el número de nodos dentro de la red.

La *matriz de adyacencia* es la forma en que se representa de forma matemática una red, es una matriz cuadrada $N \times N$ que muestra la relación que tienen los nodos de una red, en donde las entradas $a_{i,j} = 1$, representan la existencia de un arco entre el nodo i y el nodo j , y $a_{i,j} = 0$ si no existe arco entre ellos (ver figura 2.6). Una característica de la matriz de adyacencia de una red no dirigida es que es simétrica por construcción, lo cual facilita la manipulación de los datos ya que es posible trabajar solo con el triángulo inferior o el triángulo superior de la matriz.

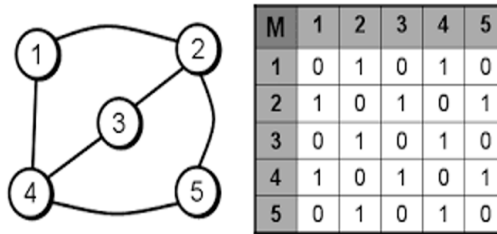


Figura 2.6: Matriz de adyacencia.

El *coeficiente de clustering* (coeficiente de aglomeración o propiedad de transitividad) se refiere a la tendencia de la red de formar triángulos entre los nodos y se puede obtener de manera local o global.

Si se supone un nodo i con k_i vecinos, se puede decir que a lo más existirán $k_i * (k_i - 1)/2$ enlaces entre ellos, con la restricción de que todos

los nodos estén conectados entre sí (creando una *red completa*), entonces se define al coeficiente de clustering local del nodo i , como la proporción de los enlaces que existen (E_i) y la cantidad máxima de enlaces que podrían existir, así [2]:

$$C_i = \frac{E_i}{\frac{k_i*(k_i-1)}{2}}, \quad (2.4)$$

donde,

C_i es el coeficiente de clustering local en el i – *ésimo* nodo, E_i son los enlaces presentes en la red, k_i es el grado del nodo i .

Con el comportamiento del clustering y la distribución de grado, se pueden clasificar las redes en diversas estructuras con características distintas [2]. A continuación, se presentan los diversos modelos de redes.

2.3. Modelos de Redes y Conceptos Avanzados

Redes aleatorias (Modelo ER) [2]

En la década de 1950, Paul Erdős y Alfred Rényi definen una red aleatoria $G = \{N, p\}$ como una red no dirigida con N nodos, donde cada par de nodos está conectado aleatoriamente con una probabilidad prefijada p (un ejemplo de esta puede verse en la figura 2.7). Si se obtienen N nodos y se comienzan a conectar con dicha probabilidad, existirá un momento en que se forme un componente gigante de la red. Lo anterior fue la aportación más importante de Paul Erdős y Alfred Rényi.

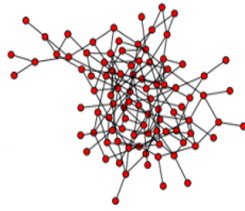


Figura 2.7: Red aleatoria con probabilidad p .

Este tipo de redes tiende a una distribución de grado binomial y cuando N tiende a infinito, se comporta como una distribución de grado tipo Poisson (ver figura 2.8).

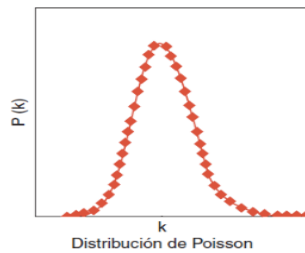


Figura 2.8: Distribución de Poisson.

Estas redes suelen tener un coeficiente de clustering bajo y una distancia media baja ($\varphi \sim \log(N)$).

Red de Mundo Pequeño (Modelo WS) [11]

En 1998, Duncan Watts y Steven Strogatz proponen el modelo de mundo pequeño. Lo más importante de este modelo es que junta la propiedad de mundo pequeño (distancia media baja) y alto coeficiente de clustering. Este tipo de red cumple con que si se tienen los nodos i y j , y no están conectados directamente, pueden unirse mediante una ruta relativamente pequeña (figura 2.9).

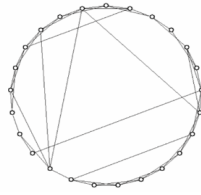


Figura 2.9: Red de mundo pequeño.

Estas redes, al igual que las redes aleatorias, tienen una distribución de grado de tipo Poisson. Es importante mencionar que las redes de mundo pequeño son la interpolación de una red regular (todos los nodos de la red tienen el mismo grado) y una red aleatoria (ver figura 2.10). Lo anterior se observa mediante una reconexión de enlaces de una red regular.

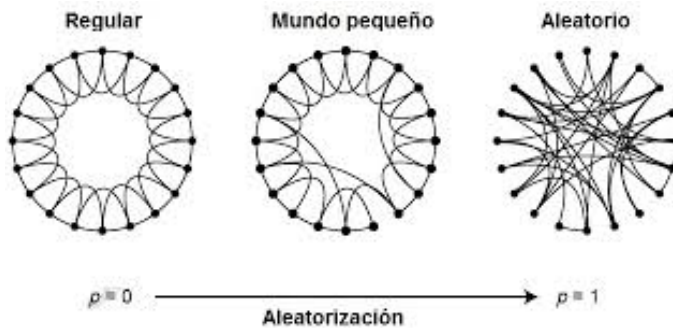


Figura 2.10: Red de mundo pequeño mediante la interpolación de redes.

Su distancia media crece proporcionalmente al logaritmo del número de nodos, esto es $\varphi \sim \log(N)$ y a nivel local presentan un clustering alto.

Red de Libre Escala (Modelo BA) [11]

Las redes de libre escala fueron propuestas por Albert-Lászlo Barabási y Réka Albert y se caracterizan por que pocos nodos tienen alto grado y muchos nodos tienen pocos enlaces (ejemplo en la figura 2.11). Existe una preferencia de enlace que hace que la probabilidad de conectar a un nuevo

nodo en la red, tiende a ser mayor si se conecta a un nodo con alto grado y menor a los nodos con menor grado.

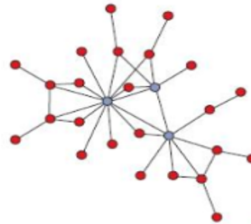


Figura 2.11: Red de libre escala.

Estas redes muestran una distribución de grado exponencial (ver figura 2.12). Una característica importante de estas redes es que tienen nodos cuyo grado es, en promedio, superior a otros nodos y son llamados *hubs*.

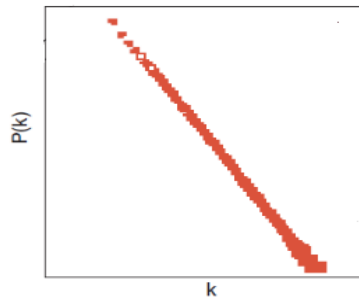


Figura 2.12: Distribución exponencial.

Los resultados numéricos muestran que, con el mismo número de nodos y el mismo grado promedio, el coeficiente de clustering es mucho mayor en comparación con una red aleatoria y su distancia media es $\varphi \sim \log(\log(N))$. Esto implica que la existencia de pocos nodos con grado alto desempeña un papel clave en conectar a los nodos entre sí.

Una vez obtenidas las redes para el análisis, se utilizó un modelo de pronóstico basado en los patrones generados por los arcos y cuyos resultados

se compararon con modelos de pronóstico estadísticos. A continuación, se mencionan los conceptos básicos de la estadística utilizados para generar el pronóstico con el que se hará la comparación del modelo basado en redes.

2.4. Estadística y Series de Tiempo

La estadística es tan antigua como la escritura misma. Los antiguos la utilizaban para poder medir impactos demográficos, sociales, contables y hasta para el comercio. Se define estadística como [16],

[...la ciencia que tiene por objeto aplicar las leyes de la cantidad a los hechos sociales para medir su intensidad, deducir las leyes que los rigen y hacer su proyección próxima...]

Así, la estadística será una herramienta para el análisis de hechos y sucesos con el fin de medirlos y pronosticarlos. Existen varias ramas de la estadística como la teoría de muestras, estadística descriptiva, estadística inferencial, análisis multivariado, entre otras. En este trabajo se utilizó la estadística inferencial como herramienta de análisis y comparación de resultados con el modelo de pronóstico propuesto. En estadística inferencial hay un tema que interesa para el desarrollo de esta tesis, las series de tiempo. Este tema consiste en encontrar patrones de una serie de datos o poblaciones y poder predecir movimientos o comportamientos futuros.

2.4.1. Series de Tiempo [12]

Por serie de tiempo nos referimos a datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares (diario, semanal, semestral, anual, entre otros).

El término serie de tiempo se aplica, por ejemplo, a datos registrados en forma periódica que muestran las ventas anuales totales de almacenes, el valor trimestral total de contratos de construcción otorgados, el valor del PIB (ver figura 2.13), entre otros.

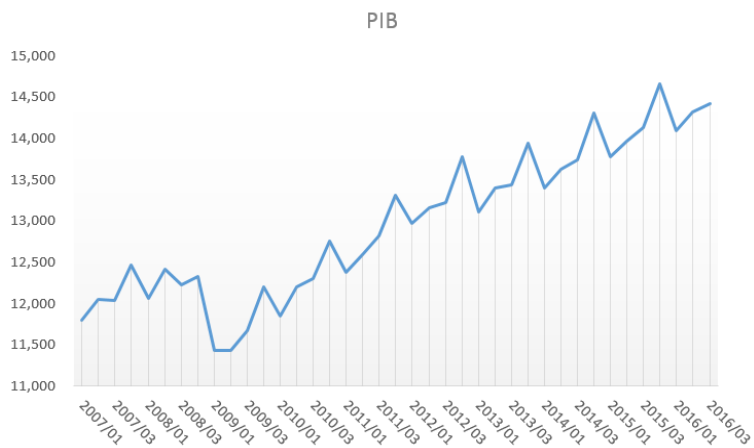


Figura 2.13: Serie de tiempo del PIB.

Fuente: Elaboración propia con datos de <https://www.inegi.org.mx/sistemas/bie/>

Las series de tiempo pueden descomponerse en los siguientes componentes [14]:

- a. *Tendencia (T)*: Es una componente de la serie temporal que refleja su evolución a largo plazo. Puede ser de naturaleza estacionaria o constante (se representa con una recta paralela al eje de las x), de naturaleza lineal, de naturaleza parabólica, de naturaleza exponencial, etc.
- b. *Variación cíclica (C)*: Es una componente de la serie que recoge oscilaciones periódicas de amplitud superior a un año. Estas oscilaciones periódicas no son regulares y se presentan, generalmente, en los fenómenos económicos cuando se presentan de forma alternativa etapas de prosperidad o de depresión.
- c. *Variación estacional (E)*: Es una componente de la serie que recoge oscilaciones que se producen alrededor de la tendencia, de forma repetitiva y en períodos iguales o inferiores a un año.
- d. *Variación aleatoria o accidental (A)*: Es una componente de la serie que recoge movimientos provocados por factores imprevisibles (un pedido

inesperado a una empresa, una huelga, una ola de calor, etc). También reciben el nombre de variaciones irregulares, residuales o erráticas.

Y, se dice que una serie es un *modelo aditivo* cuando,

$$y_t = T + C + E + A,$$

un *modelo multiplicativo* cuando,

$$y_t = T * C * E * A,$$

y un *modelo mixto* cuando,

$$y_t = T * C * E + A.$$

Una serie puede ser analizada mediante el estudio aislado de cada uno de sus componentes y, mediante herramientas estadísticas, pronosticar valores futuros.

2.4.2. Modelos de Pronóstico

Se define *pronóstico* como el proceso de estimación en situaciones de incertidumbre. Existen diversos modelos estadísticos para pronosticar series de tiempo, en particular series económicas.

En este trabajo nos enfocaremos en un modelo autoregresivo integrado de medias móviles (*ARIMA(p, d, q)*, por su acrónimo en inglés *autoregressive integrated moving average*) donde los parámetros p , d y q , son números enteros no negativos.

A continuación, se muestran los modelos individuales (autoregresivos y de medias móviles) que integran al modelo *ARIMA(p, d, q)*. El objetivo (ver Introducción) no es analizar estos modelos de forma profunda, por lo que únicamente se menciona su forma matemática.

a. Modelo autoregresivo $AR(p)$

El modelo autoregresivo de orden p , supone que el valor de la serie de tiempo y_t , depende de los valores anteriores y se puede expresar como una combinación de $y_{(t-1)}, y_{(t-2)}, \dots, y_{(t-p)}$ [12]. Así, un modelo autoregresivo se puede modelar de la siguiente forma:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \epsilon_t,$$

donde,

c es una constante,

ϕ_1, \dots, ϕ_p son los parámetros del modelo,

ϵ_t es ruido blanco ².

b. **Modelo de medias móviles $MA(q)$**

El modelo de Medias Móviles de orden q , se fundamenta en la idea de que lo que sucede en el periodo $t - 1$, permanece parcialmente en el periodo t . Así, dice que se puede expresar el momento t , como ruido blanco y una combinación de ruidos blancos previos [13].

De esta manera, se pueden modelar como:

$$y_t = \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i},$$

donde,

$\theta_1, \dots, \theta_p$ son los parámetros del modelo,

ϵ_t es ruido blanco.

²Un ruido blanco es un caso simple de los procesos estocásticos, donde los valores son independientes e idénticamente distribuidos a lo largo del tiempo con media cero e igual varianza, lo denotaremos por ϵ_t

c. **Modelos autoregresivos integrados de media móvil $ARIMA(p, d, q)$**

Los modelos de series de tiempo analizados hasta ahora se basan en el supuesto de estacionariedad, esto es, la media y la varianza para una serie de tiempo son constantes en el tiempo y la covarianza es invariante en el tiempo. Pero se sabe que muchas series de tiempo y en especial las series económicas no son estacionarias, porque pueden ir cambiando de nivel en el tiempo o sencillamente, la varianza no es constante en el tiempo, a este tipo de proceso se les considera procesos integrados. Para hacer estacionarias las series, se realizan diferencias con la misma serie d veces y luego se aplica a esta nueva serie un modelo autoregresivo de medias móviles. Se dice que la serie original es $ARIMA(p, d, q)$, es decir, una serie de tiempo autoregresiva integrada de media móvil. Donde p denota el número de términos autoregresivos, d , el número de veces que la serie debe ser diferenciada para hacerla estacionaria y q , el número de términos de la media móvil [12].

La expresión algebraica de un modelo $ARIMA(p, d, q)$ es:

$$y_t^d = c + \sum_{i=1}^p \phi_i y_{t-i}^d + \sum_{i=1}^q \theta_i \epsilon_{t-i}^d + \epsilon_t^d,$$

donde,

y_t^d es la serie de las diferencias de orden d ,

ϵ_t^d es un proceso de ruido blanco,

c , ϕ_i y θ_i son los parámetros del modelo.

Con estas definiciones se puede continuar con el análisis de las series de tiempo, con la utilización de herramientas de la teoría de redes. A continuación, se presenta el método de la gráfica de visibilidad y el método de pronóstico propuesto.

Capítulo 3

Método de la Gráfica de Visibilidad y Método de Pronóstico

Para la modelación de la serie en red, a continuación, se presenta la metodología de la gráfica de visibilidad. El método consiste en modelar una serie de tiempo como una red, para explorar sus propiedades mediante algoritmos usados en teoría de redes. Una vez obtenidas las redes, se propone un modelo que permita encontrar patrones dentro de la red y, basados en estos, pronosticar valores futuros.

3.1. Gráfica de la Visibilidad

El método de la gráfica de visibilidad fue propuesto por Lucas Lacasa *et al.* en el artículo *From time series to complex networks: The visibility graph* [6] en 2008. En este artículo se establece una metodología para mapear series de tiempo en redes y se mencionan las propiedades que heredan de estas.

En general, el método de la gráfica de visibilidad transforma una serie de tiempo en una red mediante reglas de visión, uniendo a cada punto de la serie con aquellos a los que pueda "ver" desde el punto más alto [6] (como los mostrados en la figura 3.1). Actualmente, el método de la gráfica

de visibilidad se ha utilizado para modelar series de tiempo univariadas y multivariadas, mostrando excelentes resultados [5].

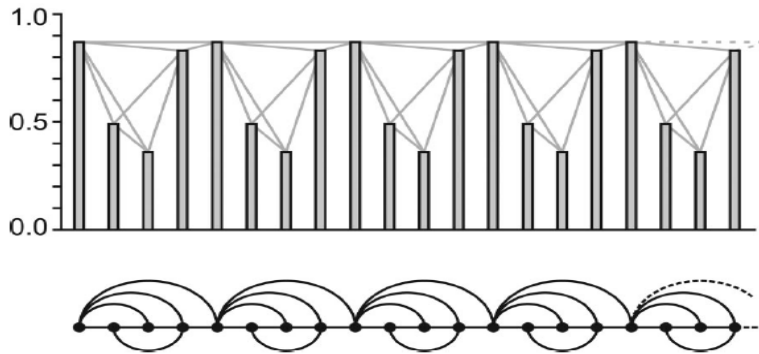


Figura 3.1: Ejemplo de una serie de tiempo de 20 datos con su respectiva red derivada del algoritmo de la gráfica de visibilidad [6].

De forma general, la gráfica de visibilidad cumple con las siguientes propiedades elementales [6]:

1. Es *conexa*, cada nodo tendrá al menos un arco con sus respectivos vecinos.
2. *Reversible*, es posible retornar a la serie original desde la red con la matriz de adyacencia. También es importante mencionar que la gráfica de visibilidad pierde cierta información con respecto a la serie de tiempo, partiendo del hecho de que la estructura de la red está determinada por la matriz de adyacencia (de forma binaria. Si existe conexión 1, en otro caso 0). Por ejemplo, la serie $T = \dots 3, 1, 3, 1, 3, \dots$ y la serie $T = \dots 2, 1, 2, 1, 2, \dots$ tendrán la misma gráfica de visibilidad aunque los datos de la serie sean distintos. Para corregir esto bastará asignar pesos en los arcos, modificando la matriz de adyacencia que tendrá estos valores en las entradas respectivas. Lo anterior generará la reversibilidad de la red a la serie original.
3. *Capacidad de crear una red dirigida y no dirigida*. Debido a la construcción de la gráfica de visibilidad, los nodos se unen de forma que la red es

no dirigida. Sin embargo, se puede realizar una red dirigida mediante el uso del grado interno y externo de cada nodo. Se define el *grado interno* (k_{in}) como la cantidad de nodos que tienen visibilidad con el nodo i , y el *grado externo* (k_{out}) como la cantidad de nodos que puede visibilizar el nodo i , y con esto se obtiene una red dirigida.

4. *Invariante* ante transformaciones como escalamientos en cualquiera de los ejes y ante traslaciones verticales u horizontales (como se muestra en la figura 3.2).

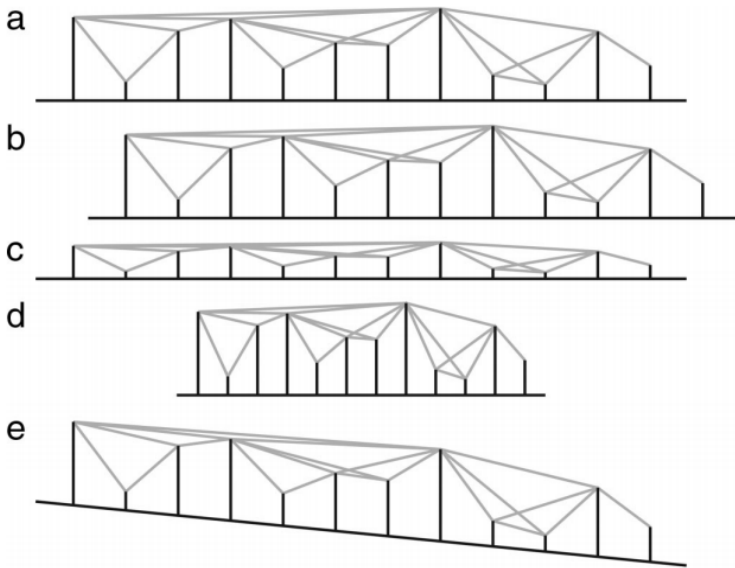


Figura 3.2: La gráfica de visibilidad de una serie de tiempo es invariante ante transformaciones. (a) Serie de tiempo original mapeada mediante el método de la gráfica de visibilidad. (b) Traslación de la serie. (c) Escalamiento vertical. (d) Escalamiento horizontal. (e) Cambio de tendencia lineal en la serie. Como se puede observar en la gráfica inferior de esta figura, la gráfica de visibilidad permanece invariante [6].

A partir del método de la gráfica de visibilidad se han realizado diversos estudios en diferentes áreas. Por ejemplo, se han analizado textos mediante

la transformación de sus palabras en redes [22], se ha estudiado el comportamiento de burbujas financieras para el pronóstico de crisis económicas [21], se ha modelado el flujo de tráfico automovilístico dentro de las ciudades [24], se han realizado estudios de la secuencia de movimientos sísmicos [23], entre otros.

El mapeo de series de tiempo en redes mediante el método de la gráfica de visibilidad puede realizarse mediante reglas definidas *a priori*. Estas reglas introducen los dos modelos de la gráfica de visibilidad: el *método de visibilidad vertical* (que se utilizará para la modelación de las series en este trabajo) y el *método de visibilidad horizontal*.

A continuación, se presenta el modelo matemático (tanto la metodología de la gráfica de visibilidad horizontal como de la vertical), que explica el algoritmo de la gráfica de visibilidad desarrollado a partir del artículo *Horizontal visibility graphs: Exact results for random time series* [5].

3.1.1. Gráfica de la Visibilidad Vertical

El método de la gráfica de la visibilidad vertical, busca transformar la serie $\{X_i\}_{i=1, \dots, n}$, con el criterio de que, si representamos los valores de dicha serie en un gráfico de barras, se une cada barra (punto en la serie) con los puntos que puedan verse desde la parte superior de la barra sin que ninguna otra se interponga.

Es decir, sean dos puntos (T_a, Y_a) y (T_b, Y_b) , tendrán visibilidad, y por lo tanto serán dos nodos conectados, si existe (T_c, Y_c) , con $a < c < b$ cumple con,

$$Y_c < Y_b + (Y_a - Y_b) \frac{(T_b - T_c)}{(T_b - T_a)}. \quad (3.1)$$

Con la desigualdad anterior se genera la figura 3.3 mostrada a continuación.

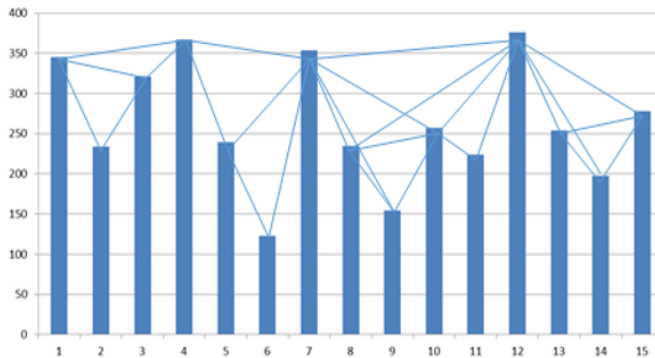


Figura 3.3: Método de visibilidad vertical.

Como lo demostraron Lucas Lacasa *et al*, la red generada por este método hereda propiedades de la serie de tiempo y sus estudios muestran información no trivial de la serie. Además, utilizando este método se puede demostrar que series periódicas, aleatorias y fractales se mapean en redes aleatorias, exponenciales y de libre escala, respectivamente [6].

3.1.2. Gráfica de la Visibilidad Horizontal

Este método, hace referencia a que si se define una serie de tiempo $\{X_i\}_{i=1, \dots, n}$, representados en una gráfica de barras, dos nodos i y j , de la gráfica estarán conectados si uno puede dibujar una línea horizontal de X_i a X_j , que no interseque ninguna altura intermedia de datos entre X_i y X_j , por lo que el nodo i y j , estarán conectados si el siguiente criterio geométrico se cumple (ver figura 3.4),

$$X_i, X_j > X_n,$$

donde $i < n < j$.

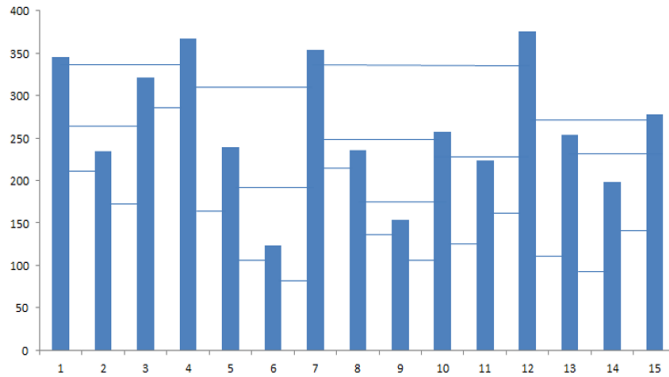


Figura 3.4: Método de visibilidad horizontal.

En primera instancia, se puede observar que la red que se genera del método de la gráfica de visibilidad horizontal es una subred, de la que se generaría con el método de la gráfica vertical con la misma serie de tiempo. Esto debido a que las conexiones generadas por el método horizontal son un subconjunto de las conexiones generadas por el método vertical [5].

Debido a la diferencia en la construcción de la red por los métodos antes expuestos, se pueden encontrar diferencias importantes en sus principales indicadores. Principalmente, en cuanto a la densidad y al clustering [5].

Otra diferencia es que el método de la gráfica de visibilidad horizontal tiene un costo computacional menor [5], que el del caso de la gráfica de visibilidad vertical, esto ayuda al análisis que requiere información en tiempo real, por ejemplo, el mercado de capitales.

Además en el caso general, el método horizontal nos generará una red con menor visibilidad que con el método vertical, esto se debe principalmente al funcionamiento matemático detrás de la construcción de dicha red. Si bien esto no provoca un impacto en las características cualitativas de las redes, hablando en términos cuantitativos, las redes generadas por el método de visibilidad horizontal tendrán por lo general menos conexiones [5].

La siguiente sección presenta el método de pronóstico propuesto.

3.2. Método de Pronóstico (Propuesta)

En esta sección se presenta un resumen del artículo de Matteo Seraffino *et al.* [21], en donde utiliza el método de la gráfica de visibilidad para series de tiempo. También, se muestran definiciones de conceptos que permitirán comprender la metodología del pronóstico y los cálculos a seguir para su estimación.

Es importante recordar que la metodología del pronóstico propuesto, se basó en la investigación de Matteo Seraffino y colaboradores para el análisis de series de tiempo, utilizando el método de la gráfica de la visibilidad.

Los autores del artículo *Statistical validation of financial time series via visibility graph* observan el comportamiento de los arcos en la red para analizar el crecimiento exponencial y después, la ruptura de las burbujas financieras, llevándolos a inferir posibles crisis a partir de este comportamiento. Lo anterior, se realiza mediante la observación de la relación de los nodos de la red generada a partir del algoritmo de la gráfica de visibilidad y se compara con la probabilidad de ocurrencia de dicha conexión utilizando modelos de tipo ARCH (Modelos autoregresivos con heterocedasticidad condicional) y GARCH (ARCH generalizados, ver ANEXO C), simulando la serie con su respectiva media y volatilidad [21].

Los modelos ARCH y GARCH son comúnmente utilizados en el modelado de series de tiempo financieras debido a que, presentan agrupaciones de volatilidad variables en el tiempo, es decir, periodos de oscilaciones entremezclados con periodos de relativa calma. Además cuentan con la propiedad de que tienen aleatoriedad máxima ayudando a modelar este tipo de series [12]. La desventaja de este tipo de familia es que no logran captar los crecimientos súper-exponenciales de las series que terminan por ser evidencia significativa de crisis, es ahí en donde se ve reflejado un cambio importante en la conectividad de la gráfica de visibilidad, mostrando un comportamiento inusual. Así, se muestra que las conexiones obtenidas mediante la gráfica de visibilidad crecen en correspondencia con periodos de crisis financiera [21].

A diferencia del artículo de Matteo Seraffino, se presenta un modelo de *pronóstico* de series de tiempo de indicadores financieros, en donde se utiliza la idea de Matteo Seraffino y sus colaboradores, en cuanto al análisis del comportamiento de los arcos de la red derivada del algoritmo de la gráfica

de visibilidad.

El método de pronóstico se basa en la conectividad (ver capítulo 2.2). Este, requiere del análisis de las conexiones existentes dentro de la red. Es por esto que, a continuación se definen conceptos basados en conectividad para seguir con el desarrollo de la metodología propuesta.

Definiciones

- a. *Arcos atrasados (AT)*: Serán los arcos que conecten a un nodo i , con un nodo j , tal que $i > j$. Definimos al vector AT , como aquel que contiene la cantidad de arcos atrasados de cada nodo de la red, en donde cada entrada AT_i , es el número de arcos atrasados que tiene el nodo i , así,

$$AT = \{AT_1, AT_2, \dots, AT_n\}.$$

Cada una de las entradas de AT se puede calcular si se utiliza la matriz de adyacencia mediante la suma de las entradas $a_{i,j}$, como se muestra a continuación:

$$AT_k = \sum_{i < j}^{j=k} a_{i,j} = \sum_{j < i}^{i=k} a_{i,j}, \quad (3.2)$$

con $k = 1, \dots, n$.

- b. *Arcos adelantados (AD)*: Serán los arcos que conecten a un nodo i , con un nodo j , tal que $i < j$. Definimos al vector AD , como aquel que contiene la cantidad de arcos adelantados de cada nodo de la red, en donde cada entrada AD_i , es el número de arcos adelantados que tiene el nodo i , así,

$$AD = \{AD_1, AD_2, \dots, AD_n\}.$$

Las entradas de AD se calculan con la matriz de adyacencia de la siguiente forma:

$$AD_k = \sum_{i>j}^{j=k} a_{i,j} = \sum_{j>i}^{i=k} a_{i,j}, \quad (3.3)$$

con $k = 1, \dots, n$.

Si, por ejemplo se requiere encontrar AT_4 y AD_4 , de la red de la figura 3.5, entonces se deberán sumar los valores dentro de alguno de los recuadros verdes y sumar los valores dentro de alguno de los recuadros rojos, respectivamente. Así, $AT_4 = 3$ y $AD_4 = 2$.

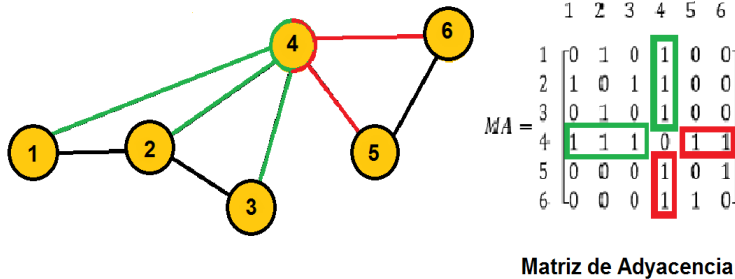


Figura 3.5: Arcos atrasados del nodo 4 (en verde) y arcos adelantados del nodo 4 (en rojo).

Para la estimación del pronóstico se deben seguir los siguientes pasos:

1. *Cálculo de factor ϕ_i*

El factor ϕ_i , ayudará a decidir el patrón que se buscará dentro de la red. Para realizar el cálculo del factor ϕ_i , con $i = 1, \dots, n$, se comienza con un vector que se llamará V_i (i indica el tamaño del vector) y contendrá valores de AT , de forma retrospectiva. Esto es, cuando $i = 1$, se tiene que $V_1 = \{AT_n\}$. Después, de forma iterativa se actualizará mediante un proceso de retroceso, como el presentado a continuación.

Con $i = 2$, entonces $V_2 = \{AT_{n-1}, AT_n\}$,
 con $i = 3$, entonces $V_3 = \{AT_{n-2}, AT_{n-1}, AT_n\}$,
 \dots ,
 con $i = n$, entonces $V_n = AT$.

Para cada V_i , se deben contabilizar las veces que este vector se repite en AT . Además, con cada V_i se debe calcular el factor ϕ_i , utilizando una ponderación por la cantidad de nodos con el fin de obtener el mejor parámetro, ya que, se busca maximizar el mayor número de veces que se repite el patrón encontrado basado en la mayor cantidad de nodos posibles. Esto es,

$$\phi_i = \frac{i * (R_i - 1)}{k},$$

donde,

R_i , es el número de veces que se repite V_i dentro de AT ,

k , es el número total de nodos de la serie AT .

Este procedimiento se realiza hasta que $R_i = 1$, ya que indica que no se tienen más patrones repetidos en la serie.

Cuando se finalice el proceso iterativo, se obtendrá el $\max\{\phi_i\}$ y se denotará como ϕ_{max} . El vector V_i , que se utilizó para obtener ϕ_{max} será el que se requiere en el siguiente paso.

2. Casos de obtención de nodos pivote

Los *nodos pivote* son aquellos nodos de los cuales, mediante su valor y el valor inmediato posterior en la serie de tiempo, se estimará el y_{t+1} .

Para aplicar la metodología de pronóstico propuesto, se requiere un análisis del comportamiento de los arcos de la red. Es importante mencionar los dos casos en los que puede derivar el análisis:

- *Caso I*: cuando se obtiene un patrón en los arcos de los nodos de la red. Cuando $R_i = 1$ cuando $i = 1$.

- *Caso II*: cuando no se obtiene un patrón directo en los arcos de los nodos de la red.

En el caso I se utiliza el *método de pronóstico simple*. En el caso II se realiza un ajuste en el método de pronóstico simple derivando en el *método de pronóstico modificado*.

Por lo anterior, la diferencia de la aplicación del método radica en cómo se obtienen los nodos pivote. Así, el método de pronóstico toma dos vertientes y es por eso que se presentan el método de pronóstico simple y el método de pronóstico modificado.

a. *Método de Pronóstico Simple*

Este método se aplica en el caso I, mencionado en casos de aplicación de la metodología de pronóstico (sección 3.2). Esto es cuando $R_i \neq 1$, con $i = 1$. Lo anterior indica que existe un patrón dentro de AT , que se utilizará para estimar y_{t+1} .

En este caso, los nodos pivote serán aquellos que se localizan en la i -ésima posición de cada V_i , que se encontró en AT . Y se definen el vector de los nodos pivote E , y el vector del nodo consecutivo del nodo pivote E_1 , como sigue,

$$E = (P_1, P_2, \dots, P_{R_{i-1}}) \text{ y } E_1 = (P_2, P_3, \dots, P_{i+1}, \dots, P_{R_{i-1}+1}),$$

donde,

P_i será la posición de cada nodo pivote dentro del vector AT .

Para ejemplificar lo anterior, se muestra la figura 3.6.

$$\begin{array}{cccccccccccc}
 & & & \boxed{3} & \boxed{4} & & & \boxed{7} & \boxed{8} & & & & & \\
 & & & \uparrow & & & & \uparrow & & & & & & \\
 AT = \{ & 0, & \boxed{2}, & \boxed{3}, & 4, & 1, & \boxed{2}, & \boxed{3}, & 5, & 6, & 7, & \boxed{2}, & \boxed{3} \} \\
 \\
 V_2 = \{ & 2, & 3 \}
 \end{array}$$

$$\text{Entonces, } \boxed{E} = \{3, 7\} \text{ y } \boxed{E_1} = \{4, 8\}$$

Figura 3.6: Ejemplo de los vectores E y E_1 .

b. *Método de Pronóstico Modificado*

Este método se aplica en el caso II, mencionado en casos de aplicación de la metodología de pronóstico (sección 3.2). Esto es, el caso en el que $R_i = 1$, cuando $i = 1$. Lo anterior indica que no existe un patrón, dentro de AT , que nos permita usar el método de pronóstico simple. Para encontrar los nodos pivote para este caso, primero, se necesita determinar un número representativo de datos para usarlos en la estimación. Se utilizó la siguiente ecuación para la obtención de este número:

$$c = \sqrt{N},$$

en donde,

c es el número de muestra que se tomará para la estimación de y_{t+1} ,
 N es el número de periodos en la serie.

Debido a que se busca encontrar los c nodos cuyo grado sean los más próximos a AT_n , se utilizan los c datos en AT , cuya diferencia con respecto a AT_n , sean las menores. Esto es, que se obtendrán los c menores datos de p , donde,

$$p = \{|AT_n - AT_i|\},$$

con,

p es el vector de las diferencias,

AT_n es el último valor de la serie de AT ,

AT_i es el valor i –ésimo de la serie de AT .

Estos c nodos son los que se usan para la estimación del pronóstico y, al igual que en el método anterior, el vector de la posición de estos nodos se denotará por E y utilizaremos también el vector E_1 .

3. Estimación de y_{t+1}

Para la estimación de y_{t+1} , se obtienen los valores originales de la serie del indicador económico de cada una de las entradas en E y E_1 , y se denotará por VSO_{E_k} y VSO_{E1_k} , respectivamente. Con esto, se realiza el siguiente cálculo,

$$y_{t+1} = y_t + \frac{1}{R_i - 1} \sum_{k=1}^{R_i-1} VSO_{E1_k} - VSO_{E_k},$$

en donde,

y_{t+1} , es el valor estimado,

y_t , es el último valor de la serie original,

R_i , es el número de veces que se repite V_i , dentro de AT ,

VSO_{E_k} , es el valor de la serie original que corresponde al k –ésimo valor del vector E ,

VSO_{E1_k} , es el valor de la serie original que corresponde al k –ésimo valor del vector E_1 .

4. Estimación de los Intervalos de Confianza

La estimación de los intervalos de confianza se generan mediante la siguiente ecuación.

$$Confianza(95\%) = (\tau_1, \tau_2) = (y_{t+1} - 2\sigma, y_{t+1} + 2\sigma),$$

donde,

σ es la desviación estándar de las diferencias de VSO_{E1_k} con VSO_{E_k} .

El método de pronóstico propuesto (de ambos casos) permitió estimar n valores futuros mediante un proceso iterativo de cálculo. En el siguiente capítulo se presenta estadística descriptiva de las series, la modelación de la red utilizando el algoritmo de la gráfica de la visibilidad, el pronóstico de las series mediante el método propuesto y la comparación de resultados del pronóstico con modelos convencionales de series de tiempo, en específico un modelo ($ARIMA(p, d, q)$).

Capítulo 4

Datos, Metodología y Resultados

En este capítulo, en primer lugar, se presentan las fuentes de donde se obtuvieron los datos para la aplicación de la metodología. Después, se muestran los estadísticos descriptivos de las series de tiempo elegidas. A continuación, se desarrolla la aplicación de la gráfica de visibilidad vertical y se presenta el método de pronóstico propuesto junto con la comparación de los resultados obtenidos y los resultantes de un modelo $ARIMA(q, d, p)$.

Para los cálculos requeridos se utilizaron las paqueterías de *networks*, *igraph* y *stata* del software R (código en ANEXO B). Para el pronóstico de las series se utilizó una hoja de cálculo por la simplicidad de su uso.

4.1. Datos

Se utilizaron tres series de tiempo obtenidas del Banco de Indicadores Económicos (BIE), de la página oficial de Instituto Nacional de Estadística y Geografía (INEGI, 2017) ¹. Estas series se mencionan a continuación.

- a. *Producto Interno Bruto*: La periodicidad de este indicador económico es trimestral. Se obtuvieron los datos del primer trimestre de 2007 (primero

¹Fuente: <http://www.inegi.org.mx/sistemas/bie/>

de marzo) hasta el tercer trimestre de 2016 (primero de septiembre). Con un total de 39 periodos.

- b. *Ocupación en México*: Este dato lo estima el INEGI de forma mensual. Se obtuvieron de enero de 2010 a marzo de 2017 con un total de 62 periodos.
- c. *Inflación*: Este indicador lo estima INEGI de forma mensual. Se obtuvieron los datos comprendidos de los periodos de enero de 2014 a marzo 2017, con un total de 39 periodos.

Por simplicidad, y para evitar pérdida de información dentro de las series, no se le da tratamiento especial a los datos. Esto es, se tomaron los datos brutos directos de la página de INEGI (sin transformaciones adicionales).

Estas series se seleccionaron debido a que tienen ciertas propiedades que nos permiten observar la consistencia del método de pronóstico propuesto.

En primer lugar, se buscó una serie de tiempo regular, buscando que tuviera patrones altamente observables (mostrando patrones a través del tiempo). Por ello se eligió la serie del PIB. Este indicador muestra regularidad en el tiempo debido a los ciclos económicos del país, con puntos de recesión y puntos de expansión económica regulares explicados por la dinámica económica nacional [12].

En segundo lugar, se buscó una serie de tiempo con datos que tuvieran un comportamiento más aleatorio. Se eligió la ocupación de México debido a que este indicador fluctúa sin que, de forma visual, se muestre un patrón a través del tiempo.

Por último, se buscó una serie que tuviera características como las presentadas en el trabajo de Matteo Seraffino *et al.* (series que muestran una tendencia positiva pronunciada, esto es, generalmente se observa un aumento en el grado de los nodos sucesivos a través del tiempo) y mostrar los efectos en el pronóstico en dicha serie. Por tal motivo, se decidió tomar como serie de estudio la inflación acumulada de México.

4.2. Estadística Descriptiva de las Series

En general, el objetivo de la estadística descriptiva es el brindar métodos (tanto gráficos como cuantitativos), que nos permitan resumir y extraer información de un conjunto de observaciones tomadas de una variable en un periodo de tiempo. Se utiliza principalmente en estudios transversales [13].

4.2.1. Estadística Descriptiva de PIB

En la tabla 4.1 se presenta un análisis descriptivo de la serie del producto interno bruto de México, desde el primer trimestre de 2007 hasta el tercer trimestre de 2016.

Tabla 4.1: Estadísticos Descriptivos del PIB

Descriptivos	Valor
Valor Mínimo	11,427,077
Primer Cuantil	12,281,921
Mediana	13,195,706
Media	13,181,800
Tercer Cuantil	13,999,048
Valor Máximo	15,318,586
Varianza	$1.10e^{12}$

Con los estadísticos descriptivos presentados en la tabla anterior se puede mencionar que, dado que la media y la mediana de la serie son similares, la distribución de los datos del PIB tiende a ser simétrica [12].

Como puede observarse en la figura 4.1, esta serie tiene un comportamiento regular a partir del onceavo dato, en donde se muestran patrones repetidos. Es por esto que se esperaría que el pronóstico replicara estos patrones.

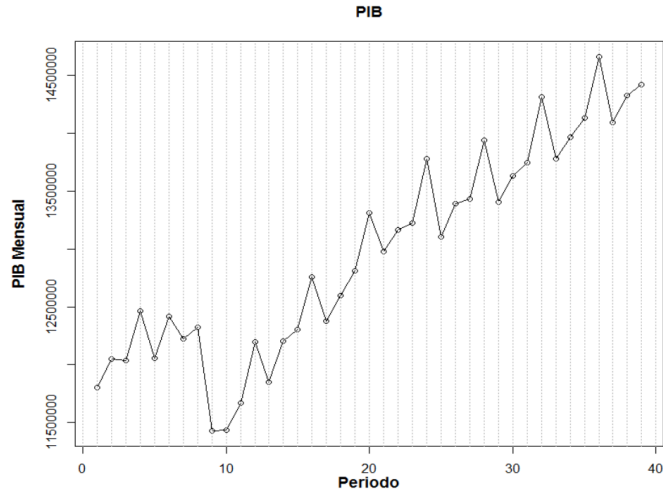


Figura 4.1: Serie de tiempo de Producto Interno Bruto

Si se observan los datos del primer trimestre de 2007 al tercer trimestre de 2010, se puede apreciar que también existía un comportamiento similar (produciendo un patrón) a los posteriores. Donde se genera este cambio de tendencia es en el primer trimestre del año 2009, justo en la crisis económica más reciente (*Sub-prime*). Este cambio de comportamiento es precisamente el que analiza Matteo Seraffino, encontrando un cambio significativo de tendencia en la serie, provocando un cambio en los grados de los nodos de la red. Con esto, Matteo Seraffino determina este comportamiento como un indicador de crisis financiera. Luego, la serie vuelve a estabilizarse mostrando nuevamente el comportamiento uniforme debido a la dinámica económica nacional, que también se estabiliza. Además, se puede ver que es una serie creciente (con tendencia positiva), regular y sin datos atípicos aparentes.

4.2.2. Estadística Descriptiva de Ocupación (%)

A continuación, se presenta un análisis descriptivo de la serie de la ocupación en México desde enero de 2010 a marzo de 2017.

Tabla 4.2: Estadísticos Descriptivos de la Ocupación en México

Descriptivos	Valor
Valor Mínimo	94.53
Primer Cuantil	95.04
Mediana	95.39
Media	95.46
Tercer Cuantil	95.86
Valor Máximo	96.63
Varianza	0.29145

De forma similar que la serie anterior, la media y la mediana son cercanas, sugiriendo que la distribución de los datos tiende a ser simétrica. En contraste con la serie anterior, la figura 4.2 no muestra regularidad a través del tiempo como se observan en la figura 4.1.

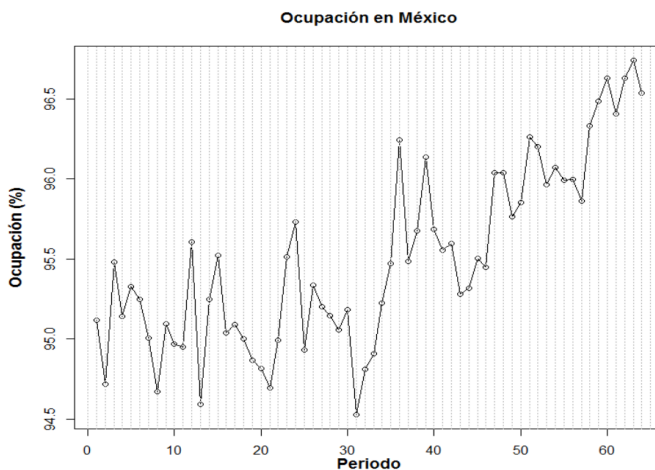


Figura 4.2: Serie de tiempo de Ocupación en México

Inmediatamente después del periodo 30 de la serie (junio de 2014), se puede observar un punto *mínimo*. Si se analiza este periodo, se infiere que debido a que existe una correlación positiva de este indicador con la productividad del país [18], se evidencian señales de una desaceleración o

estancamiento económico.

Para comprender algunos de los factores que provocaron este efecto, se menciona en una nota de Grupo MILENIO en estas fechas.

[...]el tímido crecimiento estadounidense y chino, la falta de acciones decididas por parte de la actual administración para estimular la economía y, encima, medidas como la pésima reforma fiscal que entró en vigor este año, han debilitado al mercado laboral de México que aún no consigue volver a los niveles previos a la crisis de 2008. La tasa de crecimiento económico y creación de empleos ha sido insuficiente para apuntalar el mercado interno y, por si fuera poco, México ha caído, en un solo año, del lugar 32 al 41 en el Índice global de competitividad que año tras año publica el Instituto Internacional para el Desarrollo Administrativo (IMD, por sus siglas en inglés)[...] [25].

También, en la figura 4.4 se observa que la ocupación en México incrementó en los últimos periodos indicando un crecimiento económico sostenido. Además, no hay patrones observables en la serie.

4.2.3. Estadística Descriptiva de Inflación Acumulada

A continuación, se presentan los estadísticos descriptivos básicos de la serie de la inflación acumulada en México de enero 2014 a marzo 2017.

Tabla 4.3: Estadísticos Descriptivos de la Inflación Acumulada

Descriptivos	Valor
Valor Mínimo	0.8941
Primer Cuantil	3.1204
Mediana	4.5202
Media	5.1285
Tercer Cuantil	6.8641
Valor Máximo	12.3382
Varianza	9.2888

Esta serie muestra un comportamiento creciente pronunciado debido a la acumulación de los datos. Además, la media y la mediana no son similares por lo que la distribución de este indicador no tiende a ser simétrica.

Además, se observa un comportamiento cíclico que puede explicarse debido a la dinámica de la economía nacional y puede observarse en la figura 4.3 (como se explicó en el capítulo 1).

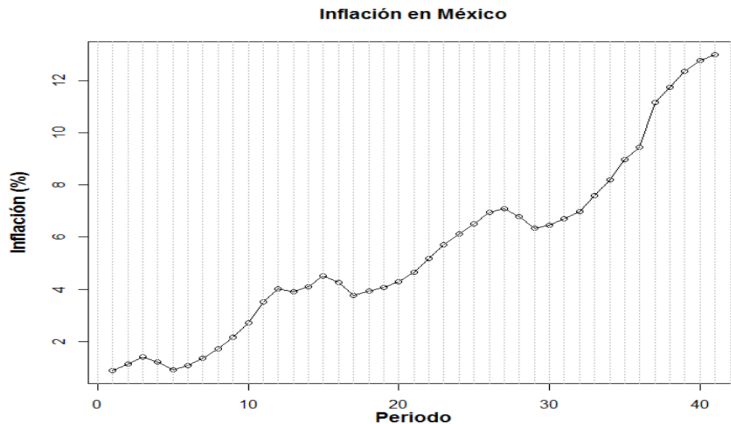


Figura 4.3: Serie de tiempo de Inflación Acumulada

4.3. Generación de la gráfica de visibilidad vertical de las series

En esta sección se aplica el método de la gráfica de visibilidad vertical a las redes seleccionadas para este análisis y se presentan algunos indicadores utilizados en la teoría de redes.

4.3.1. Visibilidad y Red del PIB

A continuación, se presentan los resultados obtenidos de la aplicación del algoritmo de la gráfica de visibilidad para modelar la serie de tiempo del PIB. Una vez obtenida la gráfica de visibilidad se modeló la red correspondiente.

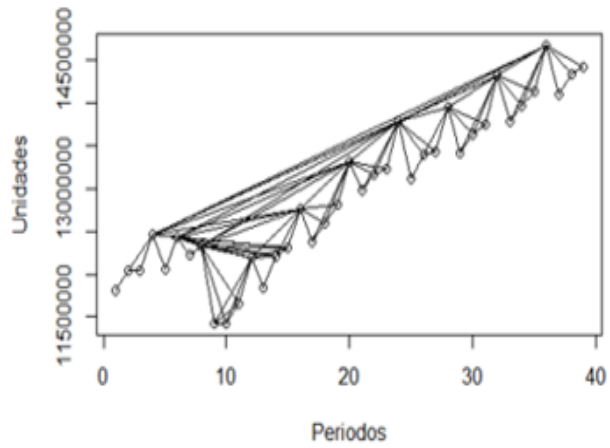


Figura 4.4: Gráfica de visibilidad del PIB



Figura 4.5: Red del PIB

En la figura 4.4 se observa la formación de patrones cada 4 periodos, mismos que a la hora de ser modelados a una red (figura 4.5) genera subredes de 4 nodos.

De esta red se calcularon algunas métricas de la teoría de redes para conocer su estructura.

Tabla 4.4: Métricas de la red del PIB.

Métrica	Valor
Nodo con Mayor Centralidad Intermedia	<i>Nodo24</i>
Coefficiente de Clustering	0.45
Densidad de la red	0.12955

Si se analiza el nodo con mayor centralidad intermedia (24), se puede observar que es un dato que sobresale de la serie, indicando un máximo local. Esto es, que los nodos con centralidad intermedia alta son resultado de observaciones que se encuentran en las colas de la distribución de los datos de la serie [8].

Debido a que este dato tuvo un crecimiento medio mayor que el resto, se analizó la fecha en que se dio este comportamiento, resultando ser el cuarto trimestre de 2012 que corresponde al nodo 24 de la red. La justificación del incremento en este periodo es debido a que como México es un gran exportador de automóviles a nivel mundial y, según datos de la Asociación Mexicana de la Industria Automotriz (AMIA), ese año se obtuvo la mayor venta histórica de automóviles hacia América Latina con 102.5 mil vehículos que ocasionó un crecimiento del PIB [26], que se observa en el periodo 24.

Una vez visto que la métrica de centralidad intermedia en la red mapeada nos indica cambios importantes dentro de la serie de tiempo, se analizaron los nodos que tuvieran una alta centralidad intermedia, estos son los nodos 2 y 4. Así, si se eliminan los nodos 2 y 4 de la red, se convierte en una red disconexa (ver figura 4.5 y 4.6).

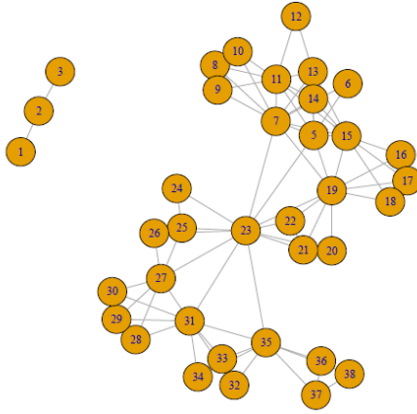


Figura 4.6: Red del PIB desconexa (nodo 4).

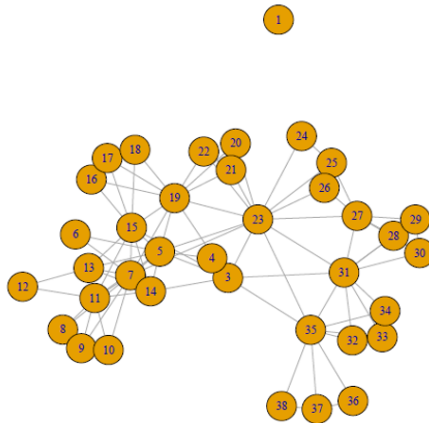


Figura 4.7: Red del PIB desconexa (nodo 2).

Los nodos 2 y 4 presentan un comportamiento atípico. Estos datos pertenecen al año 2007. Según la Comisión Económica para América Latina y del Caribe (CEPAL), México mostró una tasa de crecimiento económico de 3.3% en 2007. La Secretaría de Hacienda y Crédito Público (SHCP)

explicó que creció de forma importante impulsada por el transporte y las comunicaciones, llevando dicho indicador a cifras superiores a las esperadas [20].

No obstante lo anterior, se muestran fluctuaciones importantes para los periodos comprendidos de 2008 a 2010 (datos dentro de la serie del 5 al 12), explicados por la crisis *sub-prime* que generó menor demanda de los Estados Unidos de América, provocando inestabilidad financiera interna [20].

Para determinar la topología de la red que se produce si se utiliza el algoritmo del método de visibilidad vertical, a continuación, se presenta la distribución de grado de la red (figura 4.8).

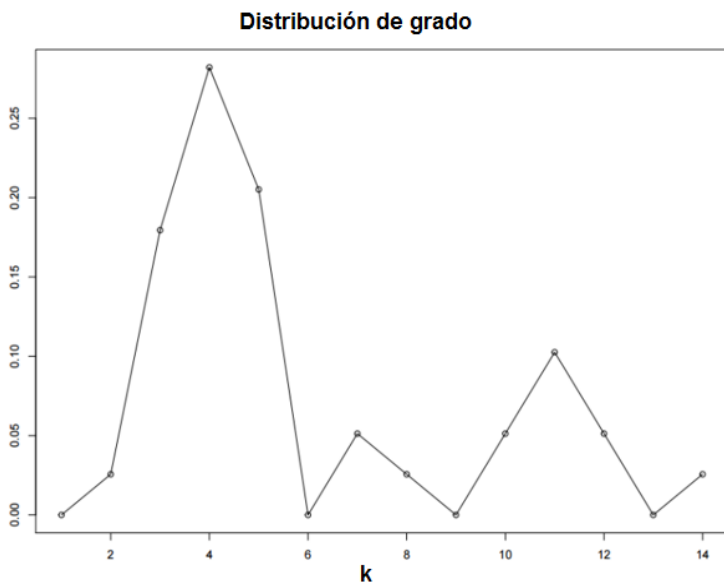


Figura 4.8: Distribución de grado de la red del PIB.

La distribución de grado de la red del Producto Interno Bruto se comporta de forma similar a una distribución Poisson $P(k) \sim e^{(-z(z^k/k!))}$ por lo que se puede hablar de que la red tiende a ser una red aleatoria. Lo anterior no puede asegurarse debido a la poca cantidad de datos de la red.

4.3.2. Visibilidad y Red de Ocupación

A continuación, se presentan los resultados obtenidos de la aplicación del algoritmo de la gráfica de visibilidad para la serie de tiempo de la ocupación en México.

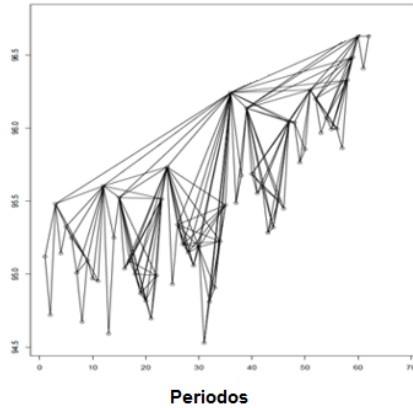


Figura 4.9: Gráfica de visibilidad de Ocupación.

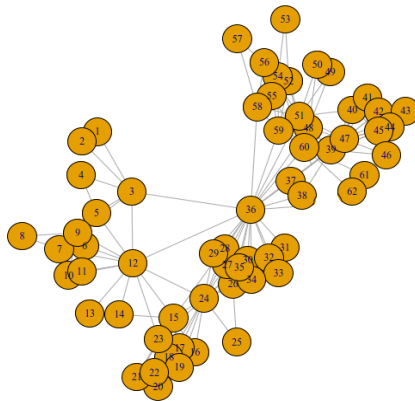


Figura 4.10: Red de Ocupación.

Se observa que esta serie tiene tendencia positiva y muestra una volatilidad superior a la serie del PIB.

En cuanto a la red de la ocupación en México, se observa que el nodo 36 (diciembre de 2014) tiene especial importancia. Este dato corresponde a un máximo local y divide la visibilidad de la red. A continuación, se presentan algunos indicadores importantes para el análisis y la comparación de las redes.

Tabla 4.5: Características de la red de la ocupación en México.

Métricas	Valor
Nodo con mayor Centralidad Intermedia	<i>Nodo36</i>
Coefficiente de Clustering	0.6157
Densidad de la red	0.10047

Si se analiza el nodo con mayor centralidad intermedia (36) se observa que es un dato en la serie que presenta un incremento significativo con respecto a los demás datos, teniendo un crecimiento medio mayor que el resto.

Debido al comportamiento del nodo 36, se analizó la fecha en la que se genera este para buscar una explicación conceptual.

La fecha que representa el nodo 36 es el periodo de diciembre de 2014, en donde según los resultados trimestrales de la Encuesta Nacional de Ocupación y Empleo, presentada por INEGI, en el periodo indica que el 96.24 % de la Población Económicamente Activa (PEA) estaba ocupada, indicando un máximo de la serie en el histórico.

No obstante, se hace mención de que dentro de estos encuestados se manifestaron casos (0.29 %), que declaraban tener la necesidad y la disponibilidad de tiempo para trabajar más horas, por lo que, aunque son del conjunto de los subocupados, se les contabilizó en este rubro y con ello se incrementó el número de personas ocupadas en este periodo [26].

Debido al crecimiento anormal en este dato de la serie, el efecto en la red es el incremento de la centralidad intermedia en este nodo. Por lo que se constató que si se elimina este nodo, la red queda disconexa como

se muestra en la figura 4.11. Esto indica que el nodo representa un dato máximo o mínimo local dentro de la serie.



Figura 4.11: Red de la ocupación en México, desconexa nodo 36.

Como anteriormente se comentó, la distribución de grado nos indica el tipo de red que se está trabajando y en el caso particular de esta red, indica que cae de forma exponencial (ver figura 4.12).

Además, se muestra una tendencia a ser una red de mundo pequeño, y lo anterior es congruente con el valor de su coeficiente de clustering (0.6157) y a que la diferencia del grado de los hubs con los demás nodos no varía significativamente (como lo haría en una red de libre escala).

Para confirmar la topología de la red se deberán tener más datos ya que la información estudiada es reducida.

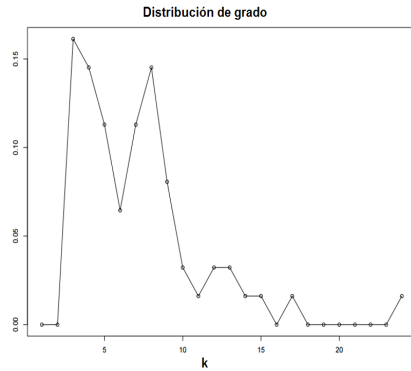


Figura 4.12: Distribución de grado de Ocupación en México.

4.3.3. Visibilidad y Red de Inflación Acumulada

De la misma forma que en las series anteriores, se generó la siguiente gráfica de visibilidad y su respectiva red.

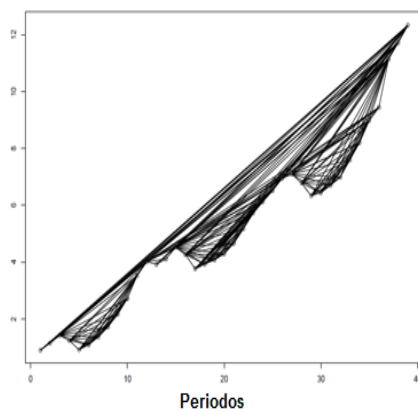


Figura 4.13: Gráfica de visibilidad de inflación acumulada.

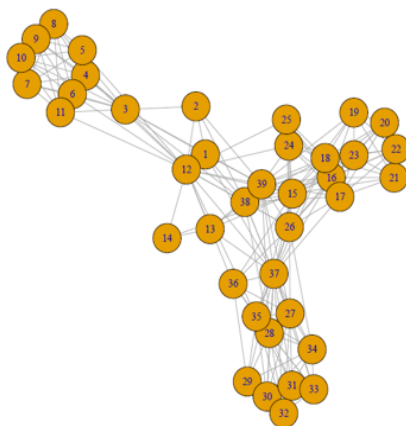


Figura 4.14: Red de inflación acumulada.

Como se aprecia en la figura 4.13, se puede observar que esta red es más densa que las anteriores, debido a que como tiene un comportamiento creciente con tendencia positiva pronunciada, el efecto de visibilidad aumenta. De la red anterior se calcularon métricas para poder conocer más a fondo la estructura de la red y se presentan en la tabla 4.6.

Tabla 4.6: Características de la red de la inflación acumulada.

Métricas	Valor
Nodo con mayor centralidad intermedia	<i>Nodo12</i>
Coefficiente de clustering	0.6376
Densidad de la red	0.268556

Esta red tiene coeficiente de clustering elevado debido al alto grado de conexión entre nodos. Además, si se eliminara el nodo con mayor centralidad intermedia (12) no se desconectaría la red. A continuación, se presenta la distribución de grado con el objetivo de ver la tendencia de la topología de esta red.

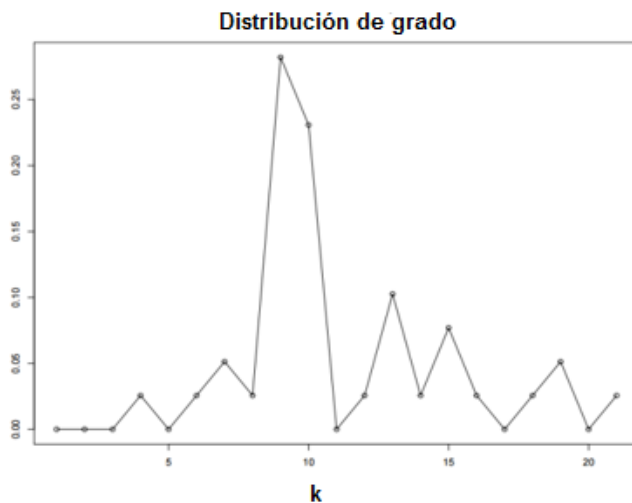


Figura 4.15: Distribución de grado de la inflación acumulada.

Esta red tiene una distribución de grado Poisson $P(k) \sim e^{-z}(z^k/k!)$ mostrando que es una red de mundo pequeño o aleatoria. Esto puede explicarse debido a la alta densidad y a que las conexiones de los nodos más recientes se relacionan directamente con los nodos más antiguos provocando un efecto de mundo pequeño dentro de la red.

4.4. Resultados de la Aplicación del Método de Pronóstico

En esta sección se presentan los resultados obtenidos utilizando el método propuesto (ver sección 3.2).

4.4.1. Pronóstico del PIB

En la figura 4.16 se muestra la gráfica del PIB con el valor estimado y_{t+1} (punto rojo) y su respectivo intervalo de confianza (puntos azules). Este pronóstico fue estimado utilizando el método de pronóstico simple (ver sección 3.2).

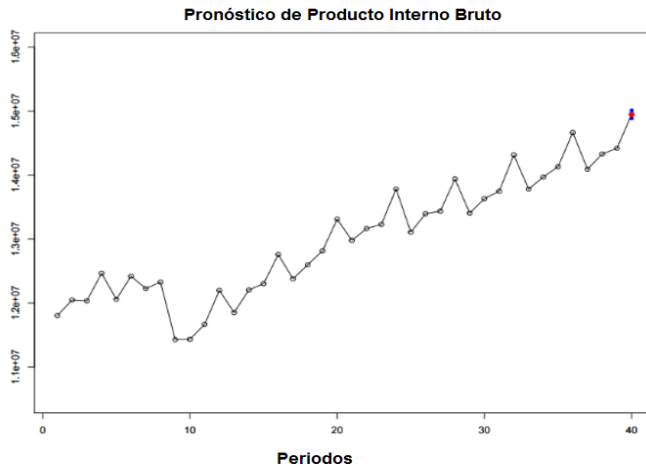


Figura 4.16: Pronóstico del PIB y_{t+1} con intervalos de confianza al 95.5 %.

Este método permite estimar n periodos posteriores de forma iterativa. Cabe señalar que los intervalos de confianza tendrán mayor amplitud conforme se estime un n más grande, y_{t+n} con $n = 1, 2, \dots$. En esta tesis, se utilizó $n = 5$. Así, se obtuvo la gráfica de la figura 4.17, en donde los puntos rojos son los valores estimados y los azules los respectivos intervalos de confianza.

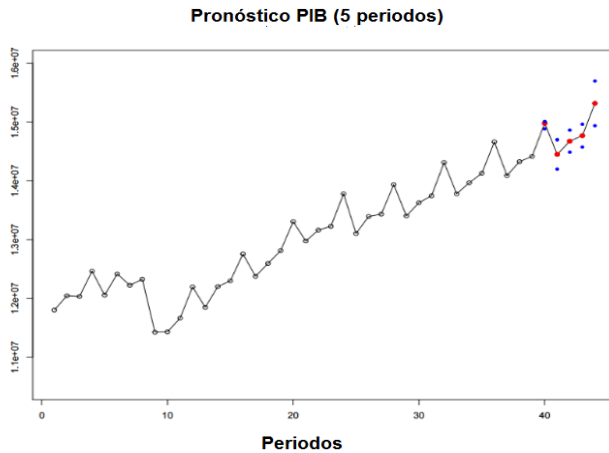


Figura 4.17: Pronóstico del PIB con intervalos de confianza (5 periodos).

4.4.2. Pronóstico de Ocupación (%)

En el caso de la serie de Ocupación en México se utilizó el método de pronóstico simple. Los resultados se muestran en la figura 4.18.

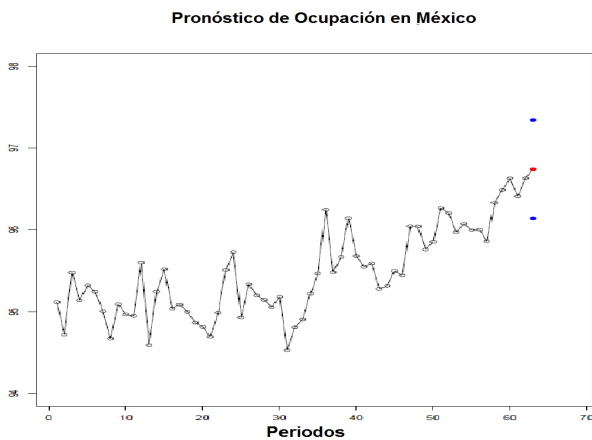


Figura 4.18: Pronóstico del primer periodo y visibilidad de ocupación.

La figura 4.19 muestra el pronóstico iterando el algoritmo para la estimación de 5 periodos. De igual forma que en el pronóstico del PIB, los puntos en color rojo son los datos estimados y los puntos azules son los respectivos intervalos de confianza.

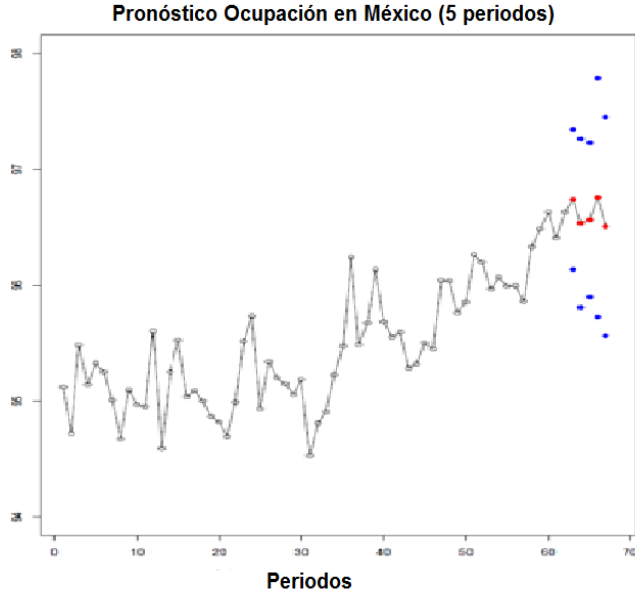


Figura 4.19: Pronóstico de ocupación con intervalos de confianza (5 periodos).

4.4.3. Pronóstico de Inflación Acumulada

Para el pronóstico de esta serie se utilizó el método de pronóstico modificado. Los resultados del y_{t+1} (punto azul) estimado se muestran en la figura 4.20 con su intervalo de confianza (puntos rojos).

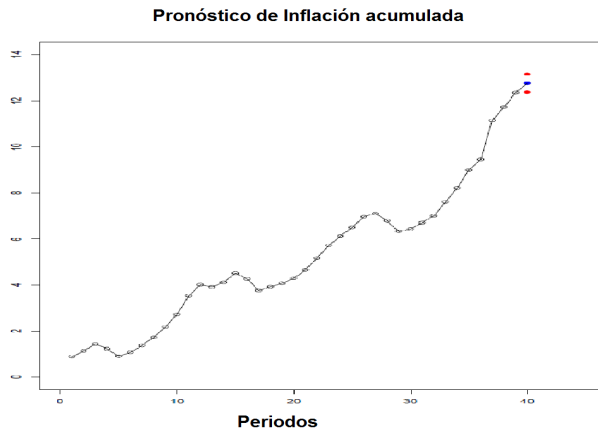


Figura 4.20: Pronóstico e intervalo de confianza inflación acumulada.

Mediante este método se generaron, de forma iterativa, los siguientes 5 periodos. Los resultados se muestran en la figura 4.21. Así, los puntos azules son los datos pronosticados y los datos rojos los respectivos intervalos de confianza.

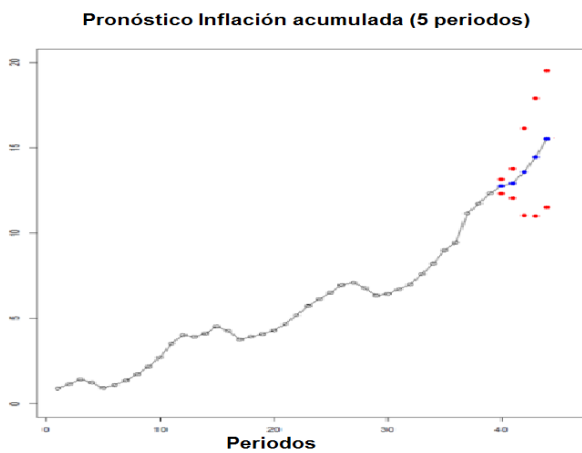


Figura 4.21: Pronóstico de periodos de inflación acumulada.

4.5. Resultados: Comparación de Pronósticos

Para ver si los datos obtenidos mediante el método propuesto son fiables, se desarrolló una comparación con un modelo $ARIMA(p, d, q)$. Cabe señalar que, para el ajuste mediante el modelo $ARIMA(p, d, q)$, se utilizó la función $auto.arima(x)$ de R , esta función optimiza los parámetros p, d y q para obtener la mejor estimación.

4.5.1. Comparación de PIB

La modelación de esta serie económica mediante el modelo estadístico arroja un modelo $ARIMA(2, 0, 2)$.

En la tabla 4.7 se muestran los valores obtenidos por el método de pronóstico de la presente tesis y los obtenidos mediante el $ARIMA(2, 0, 2)$.

Tabla 4.7: Valores de pronóstico método propuesto y modelo ARIMA.

Periodo	Modelo Redes	Conf.-95 %	Conf.+95 %	ARIMA	Conf.-95 %	Conf.+95 %
40	14, 948, 361	14, 885, 429	15, 011, 292	14, 907, 171	14, 645, 155	15, 169, 188
41	14, 264, 000	14, 013, 769	14, 514, 205	14, 377, 793	13, 949, 078	14, 806, 507
42	14, 486, 539	14, 302, 631	14, 676, 447	14, 585, 553	14, 018, 973	15, 152, 133
43	14, 583, 496	14, 389, 923	14, 777, 071	14, 655, 829	14, 015, 816	15, 295, 843
44	15, 132, 004	14, 752, 117	15, 511, 891	15, 123, 760	14, 434, 194	15, 813, 326

Diferencia PIB

A continuación, se presentan las diferencias numéricas y porcentuales de ambos modelos (tabla 4.8).

Tabla 4.8: Diferencias de pronóstico método propuesto y modelo ARIMA.

Periodo	Diferencia	Dif. %	Dif.Conf.-95 %	Dif%.Conf.-95 %	Dif.Conf.+95 %	Dif.%.Conf.+95 %
40	-133, 926	-0.76 %	240, 275	1.64 %	-157, 896	-1.04 %
41	-113, 792	-0.79 %	64, 718	0.46 %	-292, 302	-1.97 %
42	-96, 014	-0.66 %	283, 658	2.02 %	-475, 686	-3.14 %
43	-72, 332	-0.49 %	374, 107	2.67 %	-518, 772	-3.39 %
44	8, 244	0.05 %	317, 923	2.20 %	-301, 435	-1.91 %

Se observa que las diferencias arrojadas por el pronóstico estadístico y el pronóstico propuesto no difieren de forma significativa en esta serie. En primer lugar, se puede explicar debido a que se preserva la tendencia de

la serie en las estimaciones mediante ambos métodos. En segundo lugar, por ser una serie regular, tanto el modelo $ARIMA(2, 0, 2)$ como el modelo propuesto tenderán a replicar los *patrones* que se observa en los periodos anteriores.

La figura 4.22 muestra los datos pronosticados y los intervalos de confianza obtenidos mediante cada uno de los métodos.

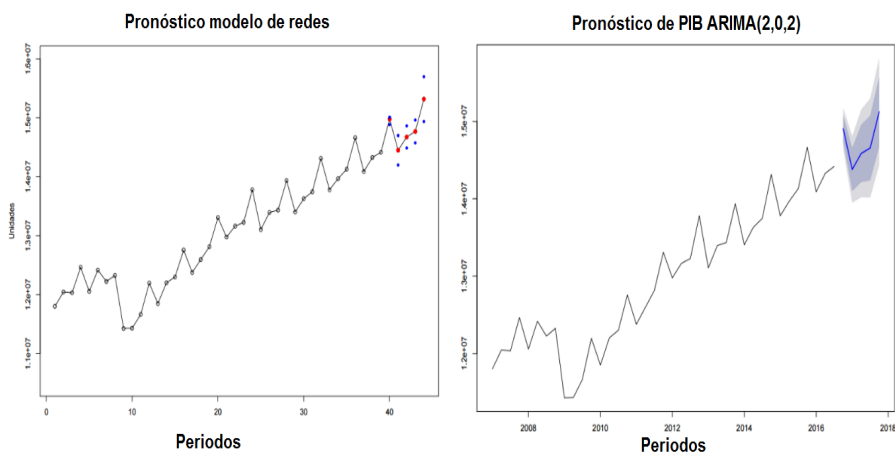


Figura 4.22: Izquierda: Pronóstico e intervalo de confianza del método propuesto. Derecha: Pronóstico e intervalo de confianza del modelo ARIMA.

4.5.2. Comparación de Ocupación (%)

La modelación de esta serie económica mediante el modelo estadístico arroja un modelo $ARIMA(2, 1, 0)$.

En la tabla 4.9, se muestran los valores obtenidos por el método de pronóstico y los obtenidos mediante el $ARIMA(2, 1, 0)$.

Tabla 4.9: Valores de pronóstico método propuesto y modelo ARIMA.

Periodo	Modelo Redes	Conf.-95 %	Conf.+95 %	ARIMA	Conf.-95 %	Conf.+95 %
63	96.7394	96.1360	97.3429	96.7824	96.2728	97.2921
64	96.5352	95.8086	97.2617	96.7632	96.2316	97.2949
65	96.5633	95.8983	97.2283	96.6351	96.0500	97.2203
66	96.7575	95.7288	97.7862	96.6810	96.0276	97.3344
67	96.5058	95.5633	97.4484	96.6407	96.9483	97.6700

Diferencias Ocupación (%)

La tabla 4.10, muestra las diferencias numéricas y porcentuales de los datos del modelo propuesto con el modelo generado por el $ARIMA(2, 1, 0)$.

Tabla 4.10: Diferencias de pronóstico método propuesto y modelo ARIMA.

Periodo	Diferencia	Dif. %	Dif.Conf.-95 %	Dif%.Conf.-95 %	Dif.Conf.+95 %	Dif.%Conf.+95 %
63	0.04308	-0.044 %	1.0701	1.112 %	1.156	-1.18 %
64	0.228	-0.236 %	1.0300	1.070 %	1.486	-1.52 %
65	0.0719	-0.074 %	1.1782	1.227 %	1.322	-1.36 %
66	0.0764	-0.079 %	1.7585	1.831 %	1.605	-1.65 %
67	0.1348	-0.140 %	1.5000	1.563 %	2.106	-2.15 %

Al igual que en la comparación del pronóstico en la serie del PIB podemos observar que las diferencias con respecto al método $ARIMA(2, 1, 0)$ no son significativos. Tanto los datos pronosticados como los intervalos de confianza son consistentes con el método propuesto.

Además, los periodos pronosticados (y también los intervalos de confianza) en ambos métodos cambian en el mismo sentido (+,-) pero con una ligera variación en la magnitud del movimiento.

La siguiente figura (4.23) muestra los pronósticos y los respectivos intervalos de confianza estimados por ambos métodos.

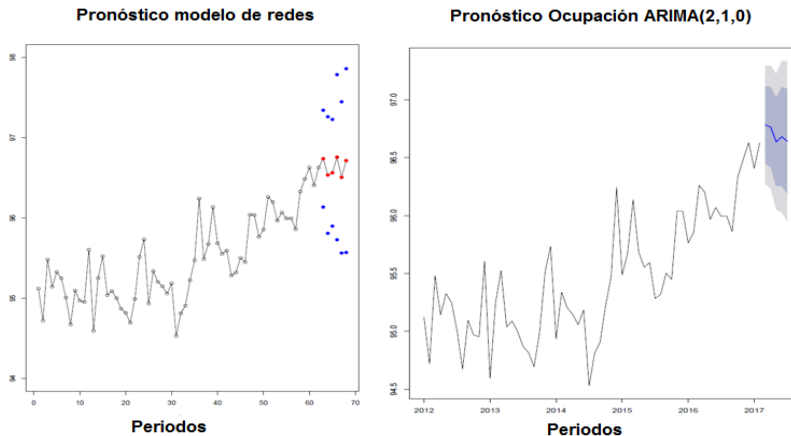


Figura 4.23: Izquierda: Pronóstico e intervalo de confianza del modelo propuesto. Derecha: Pronóstico e intervalo de confianza del modelo ARIMA.

4.5.3. Comparación de Inflación

La modelación de esta series económica mediante el modelo estadístico arroja un modelo $ARIMA(0, 2, 1)$.

En la siguiente figura se muestran los valores obtenidos por el método de pronóstico y los obtenidos mediante el $ARIMA(0, 2, 1)$.

Tabla 4.11: Valores de pronóstico método propuesto y modelo ARIMA.

Periodo	Modelo Redes	Conf.-95 %	Conf.+95 %	ARIMA	Conf.-95 %	Conf.+95 %
40	12.7528	12.3182	13.147	12.3872	11.783	12.991
41	12.9907	12.2794	13.702	12.3071	11.345	13.261
42	13.6630	11.114	16.212	12.7832	11.470	14.096
43	14.5273	11.079	17.974	13.4097	11.735	15.083
44	15.5968	11.591	19.602	14.0567	12.007	16.107

Diferencias Inflación

Es importante mencionar que como se utilizó el *método de pronóstico modificado* debido a que no se encontró un patrón al cuál seguir sino que se tomaron los valores menos distantes, la estimación de las diferencias de ambos modelos podrán ser significativamente mayores a las anteriores.

La tabla 4.12 muestra las diferencias numéricas y porcentuales de los datos del modelo propuesto con el modelo generado por el $ARIMA(2, 0, 1)$.

Tabla 4.12: Diferencias de pronóstico método propuesto y modelo ARIMA.

Periodo	Diferencia	Dif. %	Dif.Conf.-95 %	Dif%.Conf.-95 %	Dif.Conf.+95 %	Dif.%.Conf.+95 %
40	0.365	2.95 %	1.364	11.57 %	0.672	-5.18 %
41	0.603	4.90 %	2.425	21.38 %	1.218	-9.185 %
42	0.799	6.25 %	4.661	40.64 %	3.062	-21.72 %
43	1.037	7.73 %	6.159	52.48 %	4.083	-27.07 %
44	1.46	10.38 %	7.515	62.58 %	4.596	-28.53 %

Se observa que si se comparan las diferencias resultantes de los modelos anteriores, los datos pronosticados de esta serie son significativamente mayores.

La figura 4.24 muestra los valores estimados por ambos métodos de forma visual.

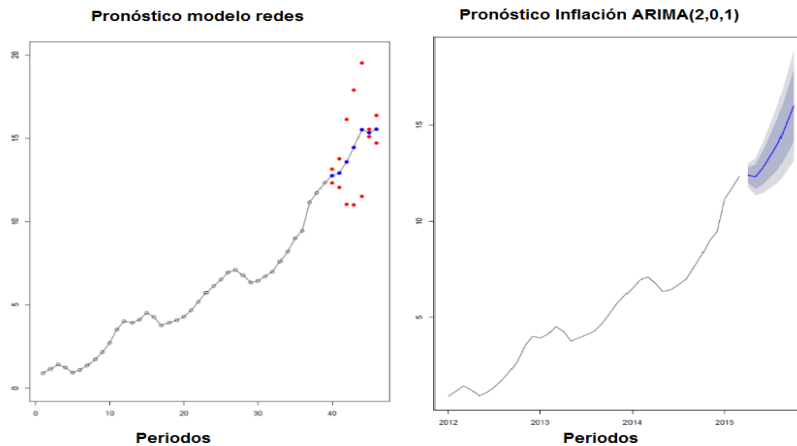


Figura 4.24: Izquierda: Pronóstico e intervalo de confianza del modelo propuesto. Derecha: Pronóstico e intervalo de confianza del modelo ARIMA.

Backtesting

Como forma de comprobación de estabilidad del método propuesto, y gracias a que las series de tiempo son de periodos anteriores, se realizó un *backtesting* para observar la estabilidad del método a través del tiempo.

Lo anterior, se logró mediante la comparación de los datos reales (obtenidos del INEGI) de los periodos pronosticados, los datos del modelo *ARIMA* y el método propuesto.

En la figura 4.25 se observa el comportamiento real del Producto Interno Bruto, los resultados obtenidos del pronóstico propuesto y el modelo *ARIMA*. En esta, se observa que ambos métodos pronostican de forma consistente la serie de tiempo, sin embargo, los datos obtenidos del modelo estadístico son más cercanos.

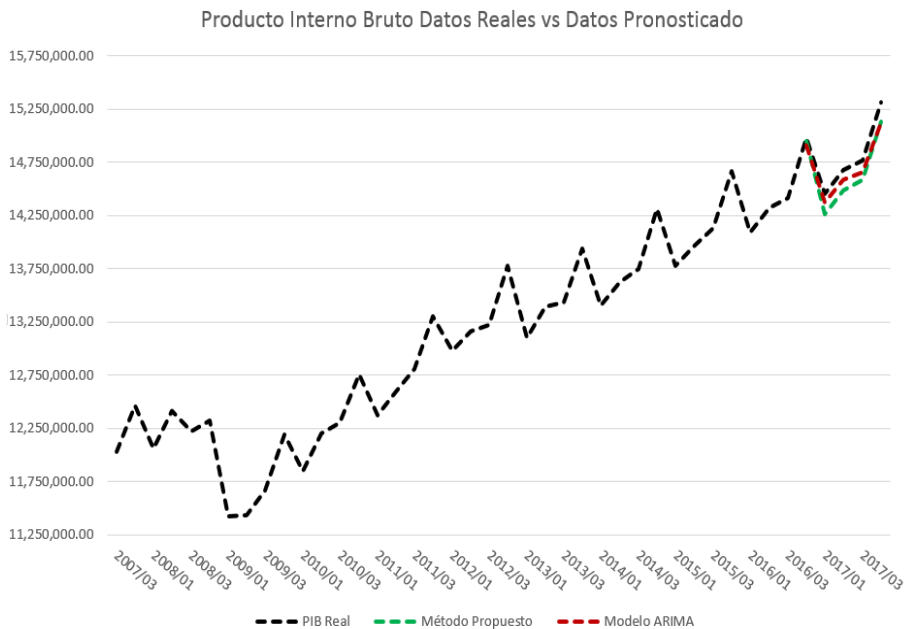


Figura 4.25: Backtesting: PIB

Si observamos la figura 4.16 (*backtesting* de la Ocupación en México) podemos notar que, dado que la serie es más volátil, los pronósticos, tanto

del *ARIMA* como del método propuesto tienen diferencias con respecto a los datos reales observados. También, se observa que los datos pronosticados por el método propuesto son más cercanos.

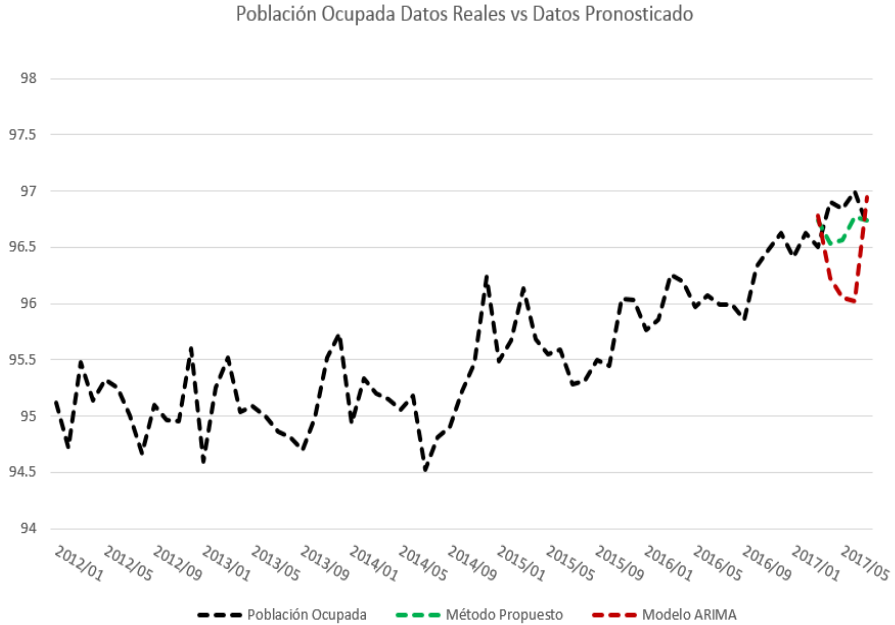


Figura 4.26: Backtesting: Ocupación en México

Ahora bien, si se analiza la figura 4.27, se observa que el método propuesto está significativamente por encima del dato real observado. En cuanto al modelo *ARIMA*, el resultado obtenido es bastante más acertado. Lo anterior, se explica por que la proyección de este indicador fue realizado mediante el método de pronóstico modificado y, como se mencionó anteriormente, los datos proyectados resultarán más volátiles.

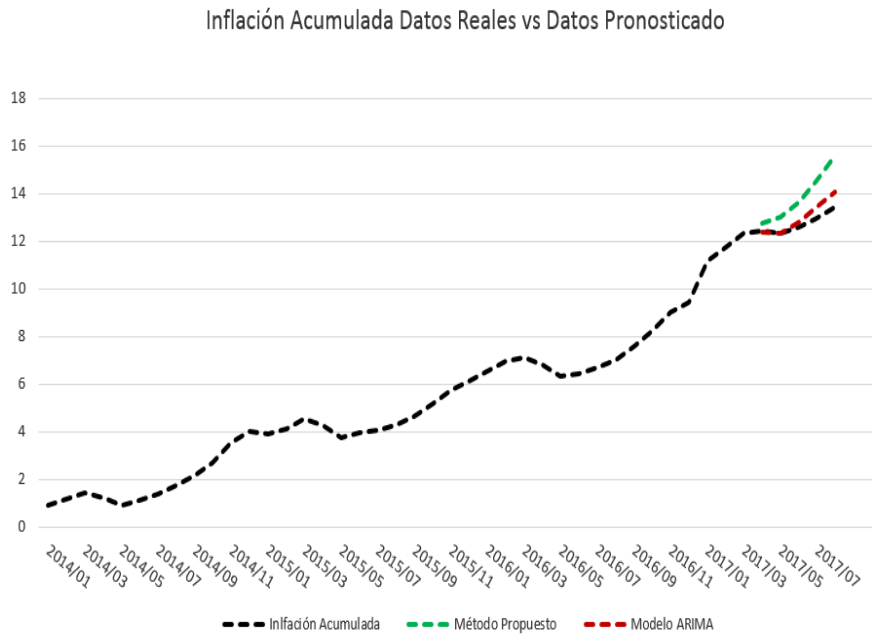


Figura 4.27: Backtesting: Inflación Acumulada en México

Para finalizar, en la siguiente sección se mostrarán las conclusiones y recomendaciones de esta tesis.

Conclusiones

El objetivo de esta tesis era proponer un método de *pronóstico* de series de tiempo utilizando la gráfica de visibilidad, para observar patrones que se encuentran en las relaciones existentes entre los datos a través del tiempo, esto es, a partir de su conectividad. Se buscó que el método propuesto tuviera congruencia numérica y es por esto que se compararon los resultados con un modelo estadístico tradicional.

Este objetivo se logró mediante la modelación de una serie de tiempo con el método de la gráfica de visibilidad vertical y el análisis de métricas usadas en teoría de redes. Los datos resultantes fueron satisfactorios comparándolos con el modelo estadístico seleccionado. En general, el comportamiento de los datos pronosticados fue adecuado.

También, se observó que para que el método de pronóstico propuesto estime de forma más consistente el valor y_{t+n} , la serie debe ser regular. Lo anterior, como en los modelos estadísticos, se basa en que el comportamiento de los datos es más predecible basado en su comportamiento pasado.

Derivado de que las series utilizadas tienen características diferentes, y que el método muestra consistencia en las tres, se puede decir que para el uso de este algoritmo podrán utilizarse series de tiempo que no requieren características específicas. Lo anterior, con el supuesto de que el resultado tenderá a ser más acertado en series regulares. En cuanto a series no regulares se observará que el pronóstico observará un mayor índice de error debido a la variabilidad de los datos y a las fluctuaciones dentro de la serie.

Es importante decir que, el método de pronóstico modificado muestra resultados estimados más volátiles que los resultados estimados mediante el método de pronóstico simple. Esto se justifica debido a que el primero utiliza nodos cuyo comportamiento es parecido a y_t , mientras que el se-

gundo utiliza nodos cuya característica principal es que están dentro de los patrones encontrados dándole mayor estabilidad al pronóstico.

Sin embargo, este método tiene ventajas significativas sobre los estadísticos. Por ejemplo, la modelación de la red es de bajo costo computacional gracias a la matriz de adyacencia, a pesar de la cantidad de datos que se pudieran tener.

Otra de las ventajas es que la serie no debe descomponerse para el estudio, ya que, con la técnica empleada, la tendencia, estacionalidad, ciclicidad y la aleatoriedad viene implícita en la modelación por la visibilidad de la red.

Ahora bien, como trabajo futuro y mejoras del método, se mencionan las siguientes recomendaciones.

- a. Se recomienda complementar la idea del pronóstico utilizando cálculos más sofisticados que diferencias aritméticas de los valores de los nodos pivotes.
- b. Se sugiere que el pronóstico se estime con una base de datos más grandes (Ley de grandes números) para observar estabilidad en las redes. Por lo que, se deberá usar información con más profundidad histórica.
- c. Debido a que se observó que el método modificado da señales de inestabilidad en los datos proyectados, aconseja realizar métodos distintos para el pronóstico de series de tiempo volátiles, en donde los patrones no se encuentren de forma sencilla. Este método buscó encontrar los patrones mediante el comportamiento de los arcos, sin embargo, podrían realizarse estudios posteriores en donde se utilice otra característica de la red.
- d. Como trabajo futuro se pretende realizar la demostración del método de forma inductiva y, mediante características específicas de las series de tiempo, establecer los criterios necesarios para los diversos tipos de series.

Los resultados obtenidos en este estudio servirán de guía para la proyección de diversas series de tiempo, con el fin de la mejora de los métodos convencionales y la apertura al estudio de distintas áreas de oportunidad que tenemos como comunidad científica.

Bibliografía

- [1] BRAND, E. y GÓMEZ, H., (2006), *Análisis de redes sociales como metodología de investigación*. Antioquía, Colombia.
- [2] DE'AGOSTINO, G., (2014), «Networks of Networks: The Last Frontier of Complexity», Roma, Italia, *Springer*, **311-322**.
- [3] ESTRADA, E. y PHILIP, A., (2015), «A first course in network theory», *Oxford university press*, **21-23**.
- [4] LACASA, L., LUQUE, B., LUQUE, J. y NUNO, J., (2009), «The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion», *EPL (Europhysics Letters)*, **86**.
- [5] LUQUE, B., LACASA, L., BALLESTEROS, F. y LUQUE, J. , (2009), «Horizontal visibility graph: Exact results for random time series», *The American Physical Society*, **1-2**.
- [6] LUQUE, B., LACASA, L., BALLESTEROS, F. y NUÑO, J. , (2007), «From time series to complex networks: The visibility graph», *PNAS*, **105(13),1-2**.
- [7] NUÑEZ, A. M., LACASA, L., GOMEZ, J. P. y LUQUE, B. , (2012), «Visibility algorithms: A short review. In New Frontiers in Graph Theory», *InTech*, **pags.121-128**.
- [8] BARABÁSI, (2002), «Linked: The New Science of Networks», *Perseus Publishing* ,Cambridge.
- [9] KOLACZYK, E. y CSÁRDI, G. , (2014), «Statistical Analysis of Networks with R», *Springer*, New York, USA.

- [10] MILGRAM, S., (1967), «The Small World Problem», *Psychology Today*, **2:60-68**.
- [11] WATTS, D. y STROGATZ, S., (1998), «Collective Dynamics of “small-world” networks», *Nature* 393, **pags.440-442** .
- [12] PEÑA, D., TIAO, G y TSAY, R., (2005), «A Course in Time Series Analysis», *Editorial John Wiley*.
- [13] PEÑA, D., (2005), «Análisis de series temporales», *Alianza Editorial*.
- [14] RUIZ, D., (2004), «Manual de Estadística», *Eudmednet*, ISBN 84-688-6153-7.
- [15] NOVALES, A., (2013), «Modelos ARCH univariantes y multivariantes», *Departamento de Economía Cuantitativa*, Universidad Complutense de Madrid.
- [16] RUÍZ, D., (2004), *Manual de Estadística*, Capítulo 1, Sevilla, España.
- [17] BROCKWELL, J. y DAVIS, R., (1996), «Introduction to Time Series and Forecasting», *Editorial Springer-Verlag*.
- [18] SAMUELSON, P. y NORDHAUS, W., (2010), «MACROECONOMÍA con aplicaciones en Latinoamérica», *Editorial Mc Graw Hill*.
- [19] SALDAÑA, J., (2018), «Modelos de propagación de enfermedades», Recuperado de <https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/matematicas-del-planeta-tierra-585/modelos-de-propagacion-de-enfermedades-11457>, 2013. Fecha de extracción de artículo 23/12/2018.
- [20] CNN, (2007), «PIB de México crece 3.3% en 2007», Recuperado de <https://expansion.mx/economia/2008/02/19/pib-de-mexico-crece-3-8-hacienda>, Febrero 2008. Fecha de extracción de artículo 15/01/2019.
- [21] SERAFINO, M. y et al., (2017), *Statistical validation of financial time series via visibility graph*, Departamento di Fisica, Univeritá Sapienza.

- [22] HERNÁNDEZ, C., BASURTO, R., OBREGÓN, B. y Guzmán, L., (2017), «Evaluating a Irregularity of Natural Languages».
- [23] TELESCA, L. y LOVATO, M., (2012), «Analysis of seismic sequences by using the method of visibility graph».
- [24] TANG, J. y et al, (2016), «Exploring dynamic property of traffic flow time series in multi-states based on complex networks: Phase space reconstruction versus visibility graph», *Elsevier*.
- [25] MILENIO, (2019), «La economía mexicana en 2014: ¿recesión o estancamiento?», Recuperado de <https://www.milenio.com/opinion/adriana-gonzalez-carrillo/argumentos-debate/la-economia-mexicana-en-2014-recesion-o-estancamiento> , Abril 2014. Fecha de extracción de artículo 02/04/2019.
- [26] AUTO, (2019), «LA INDUSTRIA AUTOMOTRIZ MEXICANA: SITUACIÓN ACTUAL, RETOS Y OPORTUNIDADES», Recuperado de <https://www.promexico.mx/documentos/biblioteca/industria-automotriz-mexicana.pdf>, 2016. Fecha de extracción de artículo 02/04/2019.

Anexos

ANEXO A

Código para modelar la serie mediante el método de la gráfica de visibilidad vertical.

```
#Lectura de datos
```

```
data<-read.csv("C:/Users/UM125RB/Desktop/Tesis/pruebaRpronfin.csv",sep=",")  
data<-read.csv("C:/Users/UM125RB/Desktop/Tesis/inflacion.csv",sep=",")  
data<-read.csv("C:/Users/UM125RB/Desktop/Tesis/ocu.csv",sep=",")
```

```
#Método de gráfica de visibilidad
```

```
s<-ts(data)  
plot(datab,xlab="Periodos",ylab="Unidades", main="PIB")#, ylim=c(0,20))  
n2 <- cbind(datab, nuevaCol=c(1:length(datab)))  
n3<-n2[,c(2,1)]  
lines(datab)  
ady<-matrix(0,nrow=length(datab),ncol=length(datab))
```


#Gráfica de visibilidad

```

for(i in 1:(length(n3)/2-1)){
  for(j in i:(length(n3)/2)){
    r<-c(0)
    for(k in (i+1):j){
      ta<-n3[i,1]
      ya<-n3[i,2]
      tb<-n3[j,1]
      yb<-n3[j,2]
      tc<-n3[k,1]
      yc<-n3[k,2]
      if((yc<(yb+(ya-yb)*((tb-tc)/(tb-ta)))) & (tb-ta)>0 & (tb-tc)>0){
        r<-r+1
      }
      if(r==(j-i-1)){
        segments(ta,ya,tb,yb)
        ady[i,j]<-1
      }
    }
  }
}
}

```

#Métricas de teoría de redes

```
library(sand)
x<-graph.adjacency(ady, mode=c("undirected"))
plot(x)
plot(degree.distribution(x), main="Distribución de Grado")
lines(degree.distribution(x))
graph.density(x)
transitivity(x)#clustering global
transitivity(x, type="average")#clustering promedio
mean(betweenness(x))#promedio de centralidad intermedia
plot(degree.distribution(x),type="b")
```

#Estadística

```
library("forecast")
ar<-plot(forecast(auto.arima(s),h=8))
ar1<-forecast.Arima(ar,h=8)
plot(ar, main="Pronóstico de Inflación ARIMA(2,0,1)")
plot(forecast(auto.arima(s),h=6))
```

ANEXO B

Series utilizadas.

Periodo	Índice de Ocupación en México
2012/01	95.1203
2012/02	94.7199
2012/03	95.4815
2012/04	95.1420
2012/05	95.3263
2012/06	95.2487
2012/07	95.0080
2012/08	94.6722
2012/09	95.0948
2012/10	94.9695
2012/11	94.9524
2012/12	95.6046
2013/01	94.5921
2013/02	95.2495
2013/03	95.5243
2013/04	95.0380
2013/05	95.0890
2013/06	95.0005
2013/07	94.8684
2013/08	94.8169
2013/09	94.6949
2013/10	94.9902
2013/11	95.5154
2013/12	95.7334
2014/01	94.9315
2014/02	95.3384
2014/03	95.2039
2014/04	95.1481
2014/05	95.0578
2014/06	95.1849
2014/07	94.5296
2014/08	94.8111
2014/09	94.9078
2014/10	95.2245
2014/11	95.4723
2014/12	96.2426
2015/01	95.4869
2015/02	95.6748
2015/03	96.1370
2015/04	95.6863
2015/05	95.5539
2015/06	95.5948
2015/07	95.2826
2015/08	95.3195
2015/09	95.5024
2015/10	95.4488
2015/11	96.0404
2015/12	96.0373
2016/01	95.7636
2016/02	95.8541
2016/03	96.2623
2016/04	96.2010
2016/05	95.9662
2016/06	96.0697
2016/07	95.9929
2016/08	95.9962
2016/09	95.8616
2016/10	96.3291
2016/11	96.4851
2016/12	96.6290
2017/01	96.4075
2017/02	96.6287
2017/03	96.7394
2017/04	96.5352

Periodo	Índice de Inflación en México
2014/01	0.89
2014/02	1.15
2014/03	1.42
2014/04	1.23
2014/05	0.92
2014/06	1.09
2014/07	1.36
2014/08	1.72
2014/09	2.16
2014/10	2.72
2014/11	3.52
2014/12	4.01
2015/01	3.92
2015/02	4.11
2015/03	4.52
2015/04	4.26
2015/05	3.76
2015/06	3.93
2015/07	4.08
2015/08	4.29
2015/09	4.66
2015/10	5.18
2015/11	5.72
2015/12	6.13
2016/01	6.51
2016/02	6.95
2016/03	7.10
2016/04	6.78
2016/05	6.33
2016/06	6.44
2016/07	6.70
2016/08	6.99
2016/09	7.60
2016/10	8.20
2016/11	8.99
2016/12	9.45
2017/01	11.15
2017/02	11.72
2017/03	12.34
2017/04	12.75
2017/05	12.99

Periodo	Producto Interno Bruto
2007/03	12,034,418.23
2007/04	12,465,488.29
2008/01	12,057,692.47
2008/02	12,418,219.69
2008/03	12,225,313.53
2008/04	12,326,228.19
2009/01	11,427,077.18
2009/02	11,432,356.88
2009/03	11,666,167.14
2009/04	12,197,396.22
2010/01	11,849,842.59
2010/02	12,203,103.44
2010/03	12,300,789.51
2010/04	12,756,899.78
2011/01	12,377,227.54
2011/02	12,597,108.19
2011/03	12,814,047.84
2011/04	13,308,587.31
2012/01	12,977,606.77
2012/02	13,164,228.90
2012/03	13,227,182.52
2012/04	13,781,117.83
2013/01	13,105,689.93
2013/02	13,394,784.88
2013/03	13,434,700.27
2013/04	13,937,843.79
2014/01	13,404,008.04
2014/02	13,630,323.07
2014/03	13,745,854.21
2014/04	14,313,237.33
2015/01	13,779,933.04
2015/02	13,968,414.79
2015/03	14,130,079.20
2015/04	14,663,623.31
2016/01	14,090,948.78
2016/02	14,328,128.33
2016/03	14,417,851.52
2016/04	14,979,826.70
2017/01	14,450,582.31
2017/02	14,676,120.91
2017/03	14,770,078.51
2017/04	15,318,585.71

ANEXO C

Modelos ARCH y GARCH [15]

ARCH(q)

Los modelos ARCH(q) aparecen en los años de 1980 con el objeto de recoger los episodios de agrupamiento temporal de volatilidad que suele observarse en las series de rentabilidad de casi todo mercado financiero. En realidad, hay precursores más antiguos [Bachelier (1900) y Mandelbrot(1963,1967)], trabajos en los que comenzó a caracterizarse las propiedades estadísticas de los precios de activos financieros.

Un modelo tipo ARCH(q) tiene la forma:

$$y_t = \epsilon_t * h_t,$$

en donde,

$$h_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2,$$

con $\alpha_0 > 0$, $\alpha_i \geq 0$, $\sum_{i=1}^q \alpha_i \leq 1$ y ϵ_t un ruido blanco.

GARCH(p, q)

En muchos casos, la especificación ARCH que recoge la estructura de autocorrelación en varianza precisa de un elevado número de retardos. Para evitar que el alto número de coeficientes en términos autoregresivos, generalmente bastante relacionados, produzca una importante pérdida de precisión en su estimación, se ha propuesto una parametrización alternativa, restringida, dependiente de un número reducido de parámetros. El modelo GARCH(p, q) (Generalized ARCH) de Bollerslev (1986) es de la forma:

$$y_t = \epsilon_t * h_t,$$

en donde,

$$h_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}^2,$$

con $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$, $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ y ϵ_t un ruido blanco.