

*Formulación de una metodología que combina herramientas
del Análisis de Redes y los Modelos Basados en Agentes.*

EDWIN JAVIER CUELLAR CAICEDO
ESTADÍSTICO



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
MAYO DE 2.016

*Formulación de una metodología que combina herramientas
del Análisis de Redes y los Modelos Basados en Agentes.*

EDWIN JAVIER CUELLAR CAICEDO
ESTADÍSTICO

TESIS PRESENTADA PARA OPTAR AL TÍTULO DE
MAGÍSTER EN CIENCIAS - ESTADÍSTICA

DIRECTOR
B. PIEDAD URDINOLA, PH.D.
PROFESORA ASOCIADA - DEPARTAMENTO DE ESTADÍSTICA

LÍNEA DE INVESTIGACIÓN
COMPUTACIÓN ESTADÍSTICA - DEMOGRAFÍA - MODELAMIENTO ESTADÍSTICO - INFERENCIA
BAYESIANA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
MAYO DE 2.016

Título en español

Formulación de una metodología que combina herramientas del Análisis de Redes y los Modelos Basados en Agentes.

Title in English

Methodological proposal using tools of Networks Analysis and Agent-Based Models.

Resumen: Los Modelos Basados en Agentes (MBA) es un método de micro-simulación que permite evaluar el comportamiento de un sistema de variables a través de la simulación de individuos, sus estados, reglas y comportamientos. A pesar de que los MBA no son una metodología estadística, son varias las etapas en las cuales la estadística puede jugar un rol importante y de las cuales puede obtener una retroalimentación para abordar problemas que involucran sistemas complejos. Los retos más importantes en el desarrollo de los MBA pueden clasificarse en dos procesos: validación e inferencia.

El proceso de validación de los MBA está estructurado en 4 etapas: validación de entrada, análisis de sensibilidad, calibración y validación de salida. Herramientas del diseño de experimentos, algoritmos genéticos y procesos estocásticos son útiles para establecer el soporte teórico de este proceso y para la validación empírica del MBA. Por otro lado, el proceso de inferencia requiere herramientas estadísticas más complejas como Emuladores Gaussianos y Computación Bayesiana Aproximada.

Muchas aplicaciones requieren considerar relaciones sociales entre los agentes. El Análisis de Redes es una metodología que estudia el comportamiento de un arreglo de individuos conectados, enfocándose en su formación, crecimiento, difusión y aprendizaje. El modelamiento estadístico, la estimación y el diseño de experimentos son temas de la estadística que han estado relacionados con avances recientes del Análisis de Redes.

Esta tesis combina herramientas de los MBA y el Análisis de Redes para proponer una metodología que permite estudiar el comportamiento a nivel micro y macro de diversas aplicaciones. La propuesta metodológica combina fenómenos de redes con MBA: la formación y el crecimiento se relacionan con la arquitectura de agentes, mientras que la difusión y el aprendizaje son usados para definir las reglas y el ambiente del MBA. Esta sinergia provee una herramienta para evaluar posibles escenarios de política pública en los cuales el aspecto social es importante para su desarrollo y puesta en práctica.

Para ilustrar la propuesta, se presenta una aplicación para estudiar diferentes escenarios de políticas públicas para la reducción de embarazos no deseados, a partir de información de la Encuesta Nacional de Demografía y Salud 2005 y 2010.

Abstract: Agent Based Models (ABM) is a micro-simulation method that allows assessing the behavior of a system of variables through the simulation of individuals, their states, rules and environment. Although, ABM is not a statistical methodology, there are many stages in which Statistics can play an important role to improve the methodology and which Statistics can get feedback to try complex systems. The most important challenges, in the development of ABM, could be classified in two processes: validation and inference.

The ABM validation process is structured in four stages: input validation, sensitivity analysis, calibration and output validation. Design of experiments, genetic algorithms and stochastic processes are useful statistical tools that could help to establish the theoretical support for this process and the empirical validation for the model. On the other hand, ABM inference process requires complex statistical tools, such as Gaussian Emulators and Approximate Bayesian Computation.

Many applications require including social relationships between agents. Network Analysis is a methodology that studies the behavior of an arrangement of linked individuals regarding its formation, growth, diffusion and learning. Statistical modelling, estimation and design of experiments are statistical topics that have been related with the recent advance of Network Analysis.

This document combines tools of ABM and Network Analysis to provide a methodological proposal that allows studying the micro and macro level behavior of different applications. The methodological proposal combines networks phenomena with ABM: network formation and growth are linked with agents' architecture while network diffusion and learning are used to define the rules and the environment of ABM. This synergy provides a tool to evaluate possible scenarios of public policies in which social aspects are important for its development.

To illustrate this proposal I use the information of The Demographic and Health Surveys in Colombia to study the simulated results under different alternatives of public policies oriented to decrease unintended pregnancy.

Palabras clave: Modelos Basados en Agentes (MBA), Análisis de Redes, micro-simulación, demografía, embarazo no deseado, escenarios de política pública, toma de decisiones.

Keywords: Agents-Based Models (ABM), Network Analysis, micro-simulation, Demography, unintended pregnancy, Public Policy Scenarios, Decision Making.

Índice general

Índice general	II
Índice de tablas	V
Índice de figuras	VI
OBJETIVOS	VII
INTRODUCCIÓN	VIII
ANTECEDENTES	X
1. ANÁLISIS DE REDES	1
1.1. Conceptos y definiciones	2
1.1.1. Red, enlaces y nodos	2
1.1.2. Grafos y representación	2
1.1.3. Características de una red	2
1.2. Muestreo y estimación en redes	4
1.3. Modelamiento para redes	5
1.3.1. Formación de una red	5
1.3.2. Crecimiento de una red	7
1.3.3. Difusión a través de una red	8
1.3.3.1. Modelo de difusión SIR	9
1.3.3.2. Modelo de difusión SIS	9
1.3.4. El aprendizaje dentro de una red	10
1.3.4.1. Modelo de DeGroot	10
1.3.4.2. Modelo de aprendizaje Bayesiano	11
1.4. Inferencia estadística en redes	12

1.5. Modelamiento y predicción para redes conocidas	12
2. MODELOS BASADOS EN AGENTES (MBA)	14
2.1. ¿Qué son los MBA?	14
2.2. Un enfoque de procesos estocásticos para los MBA	16
2.3. Componentes de los MBA y características de los agentes	16
2.3.1. Arquitectura de los agentes	19
2.3.2. Calibración y validación de los MBA	20
3. PROPUESTA METODOLÓGICA Y DESARROLLO ESTADÍSTICO	23
3.1. ¿Por qué el Análisis de Redes y los MBA?	23
3.2. Generalidades de la metodología propuesta	24
3.2.1. Esquema	25
3.2.2. Fenómenos de estudio	33
3.3. Inferencia estadística a partir de los MBA	34
3.3.1. Emuladores Gaussianos	34
3.3.2. Computación Bayesiana Aproximada	35
3.3.3. Implementación de la metodología	37
4. APLICACIÓN	38
4.1. Planteamiento y entendimiento del problema	38
4.2. Datos preliminares	40
4.3. Una mirada desde distintos enfoques	42
4.3.1. Análisis de redes	42
4.3.2. MBA	42
4.3.3. MBA y Análisis de redes	43
5. RESULTADOS Y ANÁLISIS	45
5.1. Datos empíricos y planteamiento de relaciones existentes entre las variables	45
5.2. Planteamiento del modelo	49
5.2.1. Definición de reglas para la simulación	49
5.3. Implementación de la metodología	54
5.4. Alternativas de políticas públicas	61
6. CONCLUSIONES	65
GLOSARIO	67

Bibliografía

68

Índice de tablas

5.1. Distribución por grupo de edad y estado conyugal.	47
5.2. Distribución por estado conyugal y actividad sexual.	48
5.3. Distribución por uso actual y uso pasado de métodos anticonceptivos	48
5.4. Distribución relativa por tipo de método anticonceptivo usado actualmente y uso pasado.	48
5.5. Análisis exploratorio de potenciales determinantes de la ocurrencia de embarazos. . .	50
5.6. Análisis exploratorio de potenciales determinantes de la ocurrencia de embarazos no deseados o inoportunos.	51
5.7. Razón de chances para los diferentes niveles de las variables explicativas incluidas en el análisis.	51
5.8. Distribución de problemas de fertilidad por grupo de edad.	53
5.9. Matriz de transición entre métodos anticonceptivos.	55
5.10. Matriz de transición entre estados conyugales.	55
5.11. Matriz de transición entre métodos anticonceptivos excluyendo mujeres que hayan discontinuado el uso por factores ajenos a los métodos.	55

Índice de figuras

1.	La presión social y el matrimonio	XI
1.1.	Ejemplo de red. Familia Medicci.	3
2.1.	Proceso de los MBA.	17
2.2.	Esquema BDI para la arquitectura de agentes.	20
2.3.	Esquema BOID para la arquitectura de agentes.	20
3.1.	Caja de herramientas de diseño de experimentos.	28
4.1.	Efectividad de los métodos anticonceptivos.	39
5.1.	Componentes de la Necesidad Insatisfecha de Planificación Familiar	46
5.2.	Umbral para regresión logística.	52
5.3.	Esquema general de la aplicación	53
5.4.	Análisis de sensibilidad. Escenario 1	56
5.5.	Análisis de sensibilidad. Escenario 1	57
5.6.	Análisis de sensibilidad. Escenario 1	57
5.7.	Análisis de sensibilidad. Escenario 1	58
5.8.	Análisis de sensibilidad. Escenario 2	58
5.9.	Análisis de sensibilidad. Escenario 2	59
5.10.	Análisis de calibración - Algoritmo genético. Escenario 2	60
5.11.	Resultados descriptivos. Escenario 2 - Alternativa 1	62
5.12.	Resultados descriptivos. Escenario 2 - Alternativa 2	62
5.13.	Resultados descriptivos. Escenario 2 - Alternativa 3	63
5.14.	Resultados descriptivos. Escenario 2 - Alternativa 4	63
5.15.	Resultados descriptivos. Escenario 2 - Alternativa 4	64

OBJETIVOS

Objetivo general

Diseñar una metodología que combine herramientas del análisis de redes y los Modelos Basados en Agentes (MBA) para poder abordar de manera simultánea la dependencia entre individuos en un sistema en donde interactúan múltiples variables.

Objetivos específicos

- Identificar las ventajas y desventajas teóricas y computacionales de cada una de las dos metodologías a combinar: análisis de redes y MBA.
- Evaluar la potencia de la metodología desarrollada en cuanto a su formulación teórica y su implementación computacional.
- Plantear el uso de herramientas estadísticas a nivel teórico y práctico con el fin de potenciar su uso y lograr mejoras en su desarrollo.
- Aplicar la metodología desarrollada a un estudio de caso particular de un diseño y formulación de política pública.

INTRODUCCIÓN

“We build too many walls and not enough bridges”. ISAAC NEWTON.

El origen y el desarrollo de la estadística han estado marcados por el vínculo directo con diferentes ciencias y disciplinas. En esta interacción se han producido múltiples desarrollos teóricos tras la búsqueda de soluciones a problemas prácticos. Muestra de ello son los desarrollos en análisis de regresión y su relación con la psicología y los progresos en diseño de experimentos y su vínculo con la agricultura. Estos ejemplos son una pequeña evidencia de cómo diferentes ciencias y disciplinas se nutren de los desarrollos estadísticos y la estadística se nutre de las complejidades que surgen en el desarrollo de las diferentes ciencias y disciplinas.

La estadística ha hecho parte fundamental del desarrollo científico, en la medida en que provee herramientas a las ciencias y disciplinas para la formulación de metodologías que buscan la solución de problemas. Existen dos metodologías que se destacan porque son utilizadas en una gran diversidad de temas, sus desarrollos recientes, su proyección y la retroalimentación que la estadística puede obtener de ellas para estudiar fenómenos complejos: el análisis de redes y los Modelos Basados en Agentes (MBA) ([24]). El análisis de redes es un enfoque metodológico que permite tomar en cuenta la estructura de dependencia entre observaciones, entendiéndolas como redes en las cuales los individuos interactúan. En el análisis de redes no importan los individuos como tal sino las relaciones que se establecen entre ellos ([49]). Este enfoque es relevante para la estadística, debido a que permite abordar fenómenos en donde no puede suponerse independencia entre los individuos. Además, existen diversas herramientas metodológicas y de modelación que permiten profundizar el análisis de redes desde una perspectiva estadística ([56]). Por su parte, los MBA es otro enfoque metodológico que permite estudiar la interacción de variables ([25]) dentro de un sistema complejo a partir de la simulación de agentes, en conjunto con sus comportamientos y reglas. Se entiende por simulación “el proceso de construir un modelo de un sistema que contiene un problema y conducir experimentos con el modelo en un computador con una propuesta específica de experimentación para resolver el problema” ([5]). Esta metodología se sustenta en herramientas de teoría de juegos, simulación computacional y sistemas multiagentes. Esta metodología resulta de interés para la estadística porque conjuga unos ingredientes atractivos: simulación computacional y modelamiento para el estudio de sistemas complejos. Como se presentó anteriormente, el análisis de redes se enfoca en las relaciones entre los individuos, mientras los MBA se orientan al estudio de los sistemas complejos de variables. Lo anterior permite pensar en su complementariedad ([24]).

En demografía existen trabajos que usan estas dos metodologías y que han abordado temas de nupcialidad ([13]), mortalidad ([27]), migración ([16]) y fertilidad ([80]), entre otros. Recientemente estas metodologías se han venido abordando en trabajos relacionados con una enorme

variedad de ciencias y campos de estudio, incluyendo finanzas ([60, 81, 15]), economía ([3, 46]) y estadística ([45, 47]).

Lo que se busca con el presente trabajo puede resumirse en dos grandes propósitos: en primer lugar presentar las ventajas y desventajas de cada metodología, sus usos y limitaciones, sus desarrollos teóricos y prácticos, y sus retos y avances; en segundo lugar se busca formular una metodología que, combinando herramientas del análisis de redes y los MBA, sirva como herramienta para la solución de problemas en donde sea necesario abordar simultáneamente estructuras de dependencia entre las variables e interacción entre los individuos a través de simulación computacional. Esta vinculación entre las dos metodologías ya ha sido planteada en trabajos como [24], [6], sin embargo no existe un planteamiento teórico-práctico que permita combinar ambas metodologías en un solo enfoque. Para potencializar el uso y desarrollo de la propuesta metodológica se plantea, a nivel práctico y teórico, la implementación de herramientas estadísticas que reúnen elementos de procesos estocásticos, diseño de experimentos e inferencia bayesiana, entre otros. Adicionalmente, se propone aplicar la metodología desarrollada en un estudio de caso particular de una formulación y diseño de política pública relacionado con la ocurrencia de embarazos inoportunos o no deseados.

El presente documento se estructura de la siguiente forma: a continuación se presentan algunos antecedentes relacionados con la presente propuesta; el primer capítulo presenta definiciones y herramientas necesarias para un buen entendimiento de la metodología del análisis de redes, mientras que en el segundo se sigue una estructura similar para los MBA. En el tercero se definen los lineamientos de la formulación metodológica que se busca producir en la presente tesis y el desarrollo estadístico de dicha propuesta; en el capítulo cuatro se explica la aplicación relacionada con una formulación de una política pública sobre la disminución de embarazos no deseados o inoportunos. El capítulo cinco incluye los resultados de la aplicación. En el sexto se presentan las conclusiones del presente trabajo y, finalmente, el capítulo siete incluye una propuesta de trabajo futuro.

ANTECEDENTES

La presentación de antecedentes se dividirá en dos partes: en primer lugar, se describirán de manera breve algunas aplicaciones del análisis de redes y los MBA y en segundo lugar, se presentará un panorama de las alternativas de solución más conocidas y utilizadas en el diseño y formulación de políticas públicas. Esta última sección es pertinente debido a que la metodología propuesta es aplicable en la evaluación de escenarios de política pública, lo que permite, vía simulación, aproximarse a un estudio de los posibles resultados de los diversos escenarios y escoger entre ellos el que mayor potencial tenga para impactar la variable de interés.

I. Análisis de redes y MBA

Las aplicaciones con análisis de redes y MBA son relativamente recientes pero abarcan una enorme diversidad de temas. Ambos enfoques metodológicos han tenido aplicaciones en áreas como la sociología, física, finanzas, economía, demografía, epidemiología, entre otras. A continuación se presentan algunos ejemplos que evidencian la diversidad de temas tratados con estas metodologías:

- **La propagación dinámica de la felicidad:** La felicidad ha sido por mucho tiempo objeto de estudio de muchos investigadores, en especial, con el fin de entender los principales factores que la determinan. En [32] se considera la felicidad como un fenómeno que presenta un contagio entre personas que comparten vínculos. Este estudio utiliza datos longitudinales, en donde se cuenta con información de las relaciones entre individuos, clasificándolas dependiendo de si estas corresponden a relaciones familiares, de amistad, matrimonio, vecindad o trabajo.

Entre los resultados relevantes se destaca la existencia de aglomerados o *clusters*, por su denominación en inglés, de felicidad e infelicidad. Los modelos longitudinales sugieren que los *clusters* de felicidad resultan de la difusión y no de una tendencia de las personas a relacionarse con individuos similares. Estos efectos no se evidencian en las relaciones entre compañeros de trabajo. Adicionalmente, se encuentra que los efectos se pierden a medida que pasa el tiempo y aumenta la separación geográfica.

- **La difusión de la obesidad en una red social:** Una creciente preocupación, en términos de salud pública, es el aumento de casos de obesidad, por configurarse en un factor de riesgo de enfermedades no transmisibles como las enfermedades cardiovasculares, diabetes, trastornos del aparato locomotor y algunos cánceres¹. En [17] se presenta un ejercicio en donde

¹Organización Mundial de la Salud. (<http://www.who.int/mediacentre/factsheets/fs311/es/>)

se utiliza análisis de redes como metodología para evidenciar que detrás de este aumento podría establecerse un contagio social. Utilizando los mismos datos que para el estudio sobre la difusión de la felicidad, se evidencia que diferentes vínculos o interacciones con personas obesas aumentan las probabilidades de presentar obesidad.

En este caso son las relaciones de tipo familiar y marital las que más efecto tienen sobre la difusión de la obesidad. Además, se encuentra que los efectos son diferenciales por sexo. El estudio promueve que la obesidad sea tratada a través de una intervención de salud pública y clínica, considerando a los individuos como sociales.

- **La presión social y el matrimonio:** La edad en la que se contrae matrimonio resulta para muchas personas razón de preocupación. Una aplicación relacionada con este tema es la desarrollada por [13], a través de un MBA en donde el principal interés está en hallar el riesgo de casarse (entendiendo riesgo como la tasa instantánea de ocurrencia de un evento). Para este modelo se determinó que el riesgo de contraer matrimonio depende de la disponibilidad de parejas y está mediado por la voluntad de casarse (la cual depende de la existencia de otras personas apropiadas dentro de la red social). El siguiente gráfico describe el bosquejo del sistema utilizado para el MBA:

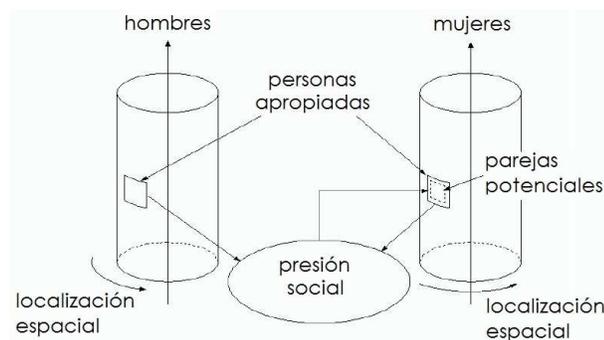


FIGURA 1. La presión social y el matrimonio. Tomada de [13]

La representación en cilindros resulta ser la más adecuada, según los autores, para mostrar la interacción entre diferentes factores: la altura corresponde a la edad, mientras que la ubicación sobre el correspondiente círculo (anillo) proporciona una distancia de similitud entre las personas, medido a través de variables como la edad, el nivel de educación, etc. Por un lado, las denominadas *personas apropiadas* hacen relación a las personas que hacen parte de la red social de un individuo, con las cuales podría establecer una relación. Este conjunto de personas se representa a través de una superficie de una vecindad bidimensional, es decir son personas similares y de edades cercanas. Por su parte, la presión social se cuantifica como una función de la proporción de personas casadas dentro de la red de personas apropiadas.

La implementación del MBA utiliza un total de 800 simulaciones y el sistema planteado anteriormente. Para inicializar algunos de los valores del modelo se utilizan datos del censo poblacional de 1995 de Estados Unidos. La primera conclusión es que al comparar la función de riesgo simulada de contraer matrimonio con la función de riesgo real, estas resultan bastante similares, lo que muestra que el modelo planteado reproduce la situación estudiada. Una segunda conclusión es que esta función de riesgo es robusta a cambios en los parámetros que determinan el tamaño de la red de personas apropiadas y la presión so-

cial. Sin embargo, se evidencia que esos cambios repercuten en la proporción de personas que contraen matrimonio.

- **El aprendizaje visto desde los MBA:** La idea en el trabajo realizado por [54] es entender de qué manera se comporta el aprendizaje dentro de un ambiente económico. Lo que subyace en esta inquietud es la teoría básica económica de optimizar dadas las restricciones, bajo la hipótesis de que se puede generar aprendizaje acerca de cómo optimizar, de acuerdo a la experiencia de otros.

Para ilustrar la metodología, el autor plantea tres escenarios en donde los MBA son útiles para estudiar este tipo de aprendizaje. En uno de los escenarios se estudia el aprendizaje de los agentes en los mercados de pescado. A través de un MBA se analiza de qué manera las relaciones entre vendedores y compradores influyen en el aprendizaje acerca de qué precios pagar y con quién negociar. En este caso, la metodología de los MBA es utilizada para estudiar la organización que emerge y el proceso de aprendizaje que se produce.

En este trabajo se desarrollará una aplicación de la metodología propuesta en un caso de estudio particular de diseño y formulación de una política pública relacionada con la ocurrencia de embarazos no deseados e inoportunos. Por esta razón, a continuación se presentan los antecedentes relacionados con las técnicas más comúnmente utilizadas en la generación de políticas públicas. La metodología propuesta busca contribuir en este tema proporcionando una alternativa para la evaluación ex-ante de escenarios de política pública bajo el paradigma de potenciar los resultados de la intervención.

II. ¿Qué son las políticas públicas?

Se conoce por política pública al conjunto de soluciones que plantea el Estado y las instituciones gubernamentales con el fin de dar respuesta a los requerimientos de los individuos que hacen parte de la sociedad en temas relacionados con problemas demográficos, de salud pública, alimentación, educación, empleo, entre otros. Las políticas públicas son un aspecto inherente de una rama de la economía denominada política económica y encierran toma de decisiones basadas en disciplinas como la sociología, el derecho, las relaciones internacionales, la historia, la geografía y la ciencia política, y cuyo fin común es establecer vínculos entre la economía y la política tanto a nivel teórico como práctico. Como señala [57], “la política puede ser analizada como la búsqueda de establecer políticas públicas sobre determinados temas o de influir en ellas”.

El desarrollo de las políticas públicas podría estructurarse en tres etapas básicas: el diseño y la formulación, la evaluación y la reformulación. Tanto la primera como la tercera etapa demandan la toma de decisiones importantes y para esto resultan de mucho interés diversas herramientas que ayuden a soportar tales decisiones. La relación entre individuos y la dependencia entre variables son dos realidades que no pueden desconocerse cuando se diseñan y se formulan intervenciones para individuos que coexisten en una sociedad y que se ven influenciados por múltiples variables del entorno. En relación con lo anterior, una herramienta potente para la toma de decisiones requiere que se puedan tomar en cuenta estos factores para la toma de decisiones.

III. ¿Cómo se formulan las políticas públicas?

En la revisión bibliográfica sobre el proceso de toma de decisiones sobre políticas públicas se identifican dos posturas: un enfoque político y un enfoque técnico. Dentro del planteamiento técnico, se reconoce que no hay un consenso sobre cuáles metodologías son pertinente para la formulación técnica de políticas públicas basada en evidencias. Existen múltiples metodologías que

contribuyen a la toma de decisiones en términos de diseño y formulación de políticas públicas. Dos de los enfoques utilizados, que son representativos y base de muchos otros, son los análisis de tipo costo-efectividad y costo-beneficio y las metodologías de toma de decisiones basada en múltiples criterios. Estas dos técnicas tienen planteamientos comunes con otras de las metodologías identificadas en la literatura y tienen una desventaja común y es que no consideran la realidad dinámica de los fenómenos en estudio.

- (a) **Costo-efectividad y costo-beneficio para la toma de decisiones:** Los análisis de costo-efectividad y costo-beneficio son dos técnicas que permiten evaluar los posibles resultados o impactos de una intervención en relación a los costos. En el caso del análisis costo-beneficio se llevan a una unidad común los beneficios y los costos y se realiza la comparación entre ambos valores. El análisis costo-beneficio “consiste en establecer un marco para evaluar si en un momento determinado en el tiempo, el costo de una medida específica es mayor que los beneficios derivados de la misma.”² En el análisis costo-efectividad se desea comparar la efectividad de una intervención en relación a los costos. Una de las utilidades de esta metodología “consiste en comparar las intervenciones que producen similares beneficios esperados con el objeto de seleccionar la de menor costo dentro de los límites de una línea de corte. Se aplica en los casos en los que no es posible efectuar una cuantificación adecuada de los beneficios en términos monetarios.”³

Son muchas las aplicaciones que se encuentran con estos dos tipos de análisis para la toma de decisiones. En Colombia algunos de estos ejemplos son: el análisis costo-beneficio de la Guerra contra las drogas ilícitas [67], la comparación costo-beneficio de los programas nutricionales de Familias en Acción y Hogares comunitarios [38] y el análisis costo - efectividad de dos intervenciones para el control de la enfermedad de malaria en Buenaventura [36].

- (b) **Tomando decisiones basados en múltiples criterios:** MCDM, por sus siglas en inglés (*Multi-Criteria Decision Making*), es una metodología relativamente reciente que soporta la toma de decisiones. En este caso, se entiende que para la toma de decisiones no solamente influyen los costos, sino otros factores relevantes. Por ejemplo la compra de una lavadora depende de aspectos como el precio, la capacidad, el tamaño y el color. En el ámbito económico, una decisión puede depender de la estabilidad financiera, el plazo de la inversión, la rentabilidad, etc. De igual forma, cuando se toman decisiones alrededor de una intervención por parte del Estado o de un ente gubernamental interesan muchos aspectos distintos a los costos, por ejemplo, el número de empleos generados, la mayor posibilidad de llegar a población vulnerable, la posibilidad de ejecutar de manera transversal la intervención en diferentes regiones o áreas y el tiempo de ejecución.

Como se presenta en [92], un análisis de MCDM puede representarse a través de una matriz de decisión Y de dimensiones $m \times n$, en donde Y_{ij} es el j -ésimo valor del atributo en la i -ésima alternativa. En este caso se entiende como atributo a una propiedad, cualidad o característica de una variable que cambia entre una y otra alternativa. Cada una de estas variables es lo que en este caso se denomina *criterio*.

El objetivo de esta metodología es proveer herramientas para la toma de decisiones, a través de la selección de una alternativa, conforme a criterios que usualmente son conflictivos. Dife-

²Leal, J. (2010) Análisis de Costo-Beneficio de regulaciones ambientales. Curso Internacional “Planificación y gestión sostenible de los recursos ambientales y naturales” CEPAL. Disponible en http://www.cepal.org/ilpes/noticias/paginas/7/40547/LEALVIERNES_2_ANALISIS_COSTO_BENEFICIO_REVISADO.pdf

³Glosario de Inversión Pública. Ministerio de Economía y Finanzas - Perú (http://www.mef.gob.pe/index.php?option=com_glossary&letter=A&id=11&Itemid=100284)

rentes métodos cualitativos y cuantitativos han sido propuestos con el objetivo de seleccionar una o varias alternativas que cumplen simultáneamente con unos criterios más favorables (Ver [92]).

(c) **Los MBA para el diseño de políticas públicas**

El uso de los MBA en el diseño de políticas públicas es un avance con respecto a las metodologías usualmente aplicadas en este contexto. Esta alternativa es más potente en la medida que permite estudiar el comportamiento de las variables que intervienen en la toma de decisiones a través de la simulación de individuos en conjunto con sus reglas, comportamientos y decisiones. Esto permite aproximarse de una mejor forma a las condiciones reales del fenómeno que se estudia, junto a su naturaleza dinámica, y hacerlo a través de los individuos, su relación con el ambiente y su interacción con los otros individuos.

Según el planteamiento de [12], una simulación de un MBA corresponde a un posible escenario del sistema bajo estudio y en esa medida diversas simulaciones sirven para el estudio de diversos escenarios, lo que ayuda a responder preguntas de política pública.

En [90] se presenta una aplicación de este enfoque, cuyo objetivo es encontrar el mejor escenario bajo el cual los habitantes de bajos ingresos de la ciudad de Búfalo, Nueva York, accedan a alimentación saludable.

Los dos componentes más importantes dentro del modelo son la población de hogares en condiciones de pobreza en la ciudad de Búfalo y los vendedores de alimentos que comercian frutas y vegetales. El consumo de estos dos tipos de alimentos se define como alimentación saludable. La distancia entre los hogares y los sitios de venta resulta relevante, ya que estudios anteriores han demostrado que a menor distancia, mayor probabilidad de acceso. Se diferencia entre supermercados, minimercados, plazas de mercado y mercados móviles. Tomando en cuenta algunos datos demográficos de la ciudad de Búfalo, el conocimiento sobre la ubicación de los sitios de venta, las probabilidades de encontrar alimentos saludables en cada uno de estos tipos de establecimientos y alguna información sobre hábitos de consumo, se implementa el modelo.

Se evalúan en total 4 escenarios: (a) integrar las plazas de mercado, (b) mejorar la frecuencia de compra, (c) aumentar el porcentaje de minimercados que venden alimentos saludables y (d) fortalecer la venta de mercados móviles. El uso de esta metodología permite dar un soporte a la evaluación de cada uno de los escenarios, y estudiar, en conjunto con las consideraciones de viabilidad de cada escenario un planteamiento de Política Pública. El primer escenario se considera un planteamiento base y se encuentra que los otros tres ayudan a mejorar el consumo de alimentos saludables.

Durante la introducción y los antecedentes se ha presentado un panorama general de las dos metodologías y algunas de las aplicaciones de cada una de ellas. En los siguientes capítulos se abordarán con mayor profundidad las ventajas y desventajas y las bases teóricas y prácticas de cada una de estas metodologías. Como se mencionó en la introducción, los análisis de redes tienen una mayor aplicabilidad en el estudio de los individuos, mientras los MBA la tienen en los sistemas complejos de variables, sin embargo, no debe perderse de vista que ambos métodos se basan en el estudio de los individuos, en el primer caso denominados nodos y en el segundo agentes. En el análisis de redes el énfasis del estudio se encuentra en la interacción de los individuos, lo que implica el estudio de fenómenos como el crecimiento, el

aprendizaje y la difusión. Por otro lado, los MBA estudian la interacción de variables dentro de un sistema a partir de los agentes, sus comportamientos, reglas y decisiones. El presente trabajo se fundamenta en estas dos metodologías, sus aplicaciones, ventajas y desventajas para producir una nueva metodología. Los siguientes dos capítulos presentan con mayor detalle herramientas importantes de cada una de las metodologías base.

ANÁLISIS DE REDES

La relación entre la estadística y el análisis de redes tiene varios ejes de trabajo ([56]): (a) en primer lugar, el tema de muestreo y estimación; (b) en segundo lugar, el modelamiento; (c) en tercer lugar, los temas relacionados con inferencia estadística en redes y finalmente, (d) el modelamiento y predicción para redes conocidas. Estos temas serán abordados a lo largo de este capítulo, haciendo énfasis principalmente en el modelamiento. Además, se presentan de manera general algunas herramientas teóricas y prácticas que permiten hacer uso del análisis de redes y se proveen algunas definiciones que permitirán un mejor entendimiento de los temas abordados.

El análisis de redes es una denominación común a diferentes temas trabajados en áreas tan diversos como la ingeniería de sistemas, la ingeniería eléctrica, las matemáticas, la economía y la sociología. La definición que se toma para este trabajo corresponde más exactamente a *Análisis de Redes Sociales* y corresponde a un enfoque metodológico que reúne herramientas de campos de estudio como la teoría de grafos, modelación estocástica y teoría sociológica. El análisis de redes, en este contexto, estudia los enlaces que se establecen entre nodos. Los nodos pueden corresponder a personas, empresas, lugares, organizaciones, etc., mientras que los enlaces corresponden a las relaciones que pueden establecerse entre ellos, como amistades, vínculos comerciales, conexiones viales y cooperaciones, respectivamente. Este enfoque metodológico permite tomar en cuenta la estructura de dependencia entre observaciones, entendiéndolas como redes en las cuales los actores interactúan entre ellos.

Como se presentó en los antecedentes, los análisis de redes se han implementado en diferentes disciplinas permitiendo estudiar la estructura de una red, su caracterización y descripción, y, con ayuda de otras metodologías, realizar avances en torno a análisis estadísticos espaciales y longitudinales. A propósito de la importancia de esta metodología, en [49] se menciona que “las redes sociales permean nuestras vidas sociales y económicas. Ellas juegan un rol central en la transmisión de información acerca de oportunidades de trabajo, y son críticas para el comercio de muchos bienes y servicios. Ellas son la base de la provisión de seguros en países desarrollados. Las redes sociales son también importantes en determinar cómo se difunden las enfermedades, cuales productos compramos, qué idiomas hablamos, cómo votamos, así como decidimos si nos volvemos criminales o no, cuánta educación obtenemos, y nuestra probabilidad de éxito profesional.”

1.1. Conceptos y definiciones

Los conceptos, definiciones y notación que se presentan a continuación corresponden a una combinación de los contenidos encontrados en [49] y [22].

1.1.1. Red, enlaces y nodos

Se parte de un conjunto de nodos $N = \{1, \dots, n\}$ entre los cuales se establecen enlaces que representan relaciones. Se denomina red al conjunto de nodos con sus correspondientes enlaces. Entre los nodos pueden o no existir relaciones y estas pueden ser de diferentes tipos, como se explicará más adelante.

1.1.2. Grafos y representación

Los grafos son representaciones de relaciones binarias que se establecen entre vértices, denominados nodos, y que se representan a través de líneas, también llamadas aristas o enlaces. Los grafos son el objeto de estudio de un campo de las matemáticas conocido como teoría de grafos. Los grafos permiten representar gráficamente las conexiones entre los individuos de la red. Se denotará (N, g) al grafo consistente en un conjunto de nodos $N = \{1, \dots, n\}$ y una matriz g de dimensión $n \times n$, en donde g_{ij} corresponde a la relación existente entre el nodo i y el j , siendo $|N| = n$ el número de nodos. Esta matriz es denominada matriz de adyacencia. Una red se denomina *dirigida* si las relaciones pueden establecerse de manera unilateral, de forma que es posible que $g_{ij} \neq g_{ji}$, mientras que la red se denomina *no dirigida* si las relaciones se establecen de manera recíproca, de forma que $g_{ij} = g_{ji}$ para todo i, j , haciendo g una matriz simétrica.

Las relaciones dentro de una red pueden ser ponderadas directamente proporcional a la intensidad del vínculo. En caso contrario los valores de g_{ij} serán 1 en caso que haya relación y 0 en caso que no exista. Se denomina *ruta* entre i y j a una secuencia de nodos que comienza en i y termina en j , tales que, cada pareja consecutiva está conectada y cada nodo es único en la secuencia. Por su parte, se denomina *caminata* entre i y j a una secuencia de nodos que comienza en i y termina en j , tales que, cada pareja consecutiva está conectada. Un *ciclo* es una caminata que comienza y termina en el mismo nodo. Se define la *distancia* entre dos nodos como la longitud de la caminata más corta que hay entre ellos. Adicionalmente se denomina *diámetro* de la red a la distancia más grande que hay entre cualquier par de nodos.

Una red se dice *conectada* si todo par de nodos está conectado por una caminata. Se llama *vecindad* de un nodo i a todos los nodos con los que se encuentra conectados.

1.1.3. Características de una red

Existen diferentes características que interesa estudiar en una red y que se presentan a continuación:

La distribución de los grados: Se denomina *grado* de un nodo al número de enlaces que tiene con otros nodos de la red, lo cual equivale a la cardinalidad de su vecindad. La distribución de los grados corresponde a la descripción de las frecuencias relativas de los nodos que tienen diferentes grados. Esta es una característica clave para el análisis de redes, ya que permite conocer

las probabilidades asociadas a que un nodo seleccionado al azar tenga un número determinado de enlaces. Una de las tipologías en las cuales se clasifican las redes corresponde a la distribución de grados. Algunas de las tipologías mas conocidas son Poisson, exponencial y libre de escala. ([21])

Partidismo o cliquishness: Este concepto hace referencia a la formación de grupos cerrados o exclusivos (*partidos*). Un partido o *clique* se define como el máximo subconjunto $S \subset N$, tal que todos los nodos de S están conectados. Un partido debe tener al menos tres nodos para sea considerado como tal. Una medida de partidismo es contar el número y el tamaño de los partidos.

Agrupamiento o clustering: Este concepto alude a la tendencia de los nodos a formar grupos o *clusters*. La medida de agrupamiento global se define como:

$$Cl(g) = \frac{\sum_i \#\{jk \in g | k \neq j, j \in N_i(g), k \in N_i(g)\}}{\sum_i \#\{jk | k \neq j, j \in N_i(g), k \in N_i(g)\}} = \frac{\sum_{i,j \neq i, k \neq j, k \neq i} g_{ij}g_{ik}g_{jk}}{\sum_{i,j \neq i, k \neq j, k \neq i} g_{ij}g_{ik}} \quad (1.1)$$

La anterior medida se basa en tripletas de nodos. La tripeleta se denomina cerrada si los tres nodos están conectados. La medida de agrupamiento global es la razón entre el número de tripeletas no conectadas sobre el número de tripeletas conectadas. Por su parte, la medida de agrupamiento individual para el nodo i , que cuantifica esta misma relación para cada nodo, está dada por:

$$Cl_i(g) = \frac{\#\{jk \in g | k \neq j, j \in N_i(g), k \in N_i(g)\}}{\#\{jk | k \neq j, j \in N_i(g), k \in N_i(g)\}} = \frac{\sum_{j \neq i, k \neq j, k \neq i} g_{ij}g_{ik}g_{jk}}{\sum_{j \neq i, k \neq j, k \neq i} g_{ij}g_{ik}} \quad (1.2)$$

El coeficiente de agrupamiento promedio se define como:

$$Cl^{Prom}(g) = \sum_i Cl_i(g)/n \quad (1.3)$$

El siguiente gráfico corresponde a un ejemplo presentado por [49], en donde se presenta la estructura relacional de las familias más importantes del Renacimiento en Florencia. Los enlaces en este caso representan vínculos matrimoniales entre miembros de dos familias. Se toman los vínculos matrimoniales ya que para la época esto repercutía en las relaciones políticas y comerciales entre dos familias.

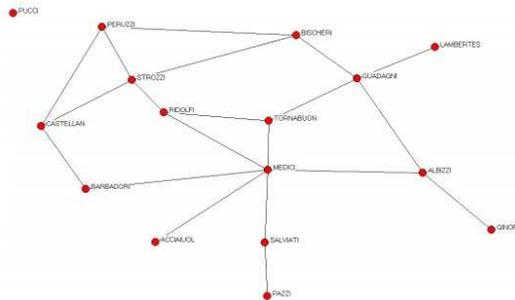


FIGURA 1.1. Ejemplo de red. Familia Medici. Tomada de [49]

En primer lugar, es importante resaltar que el diámetro de la red es 5, lo cual significa que para establecer vínculos entre dos familias existen máximo 5 vínculos entre ellas. En segundo lugar, la familia Pucci no está conectada a la red. En tercer lugar, un cálculo de los grados de cada una

de las familias de la red permite observar que la familia Medici es la que presenta un mayor número de enlaces, lo cual estaría asociado con más vínculos familiares, comerciales y políticos. Sin embargo, esta no es la única característica favorable para esta familia: su nivel de centralidad es el más alto, lo cual se evidencia en que la mayoría de rutas entre familias involucran a la familia Medici. Lo anterior no es coincidencia, esta familia fue una de las más influyentes y poderosas de Europa: aportó tres Papas, numerosos dirigentes y miembros de la familia real en Francia e Inglaterra. El anterior ejemplo es una muestra de la forma en cómo las redes pueden contribuir al análisis y entendimiento de las relaciones que se construyen y las implicaciones que se derivan de ello.

1.2. Muestreo y estimación en redes

Supóngase que se tiene un grafo (N, g) consistente en un conjunto de n nodos $N = \{1, \dots, n\}$ y una matriz g en donde g_{ij} corresponde a la relación existente entre el nodo i y el j . Los temas de estimación y muestreo surgen cuando no es posible observar toda la red y es necesario seleccionar una muestra dada por (N^*, g^*) , a partir de la cual se desea hacer inferencia estadística de la red poblacional. Existen algunos desarrollos estadísticos recientes que plantean muestreos probabilísticos en redes ([56]):

- **Muestreo de subgrafos inducido e incidente:** Según se presenta en ([56]), el muestreo de subgrafos inducido consiste en la selección sin reemplazamiento de una muestra aleatoria de nodos N^* y se observan los enlaces entre los nodos i, j , tal que $i, j \in N$ e $\{i, j\} \in g$. En este caso las probabilidades de inclusión para los nodos y los enlaces están dadas por:

$$\pi_i = \frac{n^*}{n} \text{ y } \pi_{\{i,j\}} = \frac{n^*(n^* - 1)}{n(n - 1)} \quad (1.4)$$

Por su parte, en el muestreo de subgrafos incidente se seleccionan aleatoriamente n^* enlaces del conjunto de enlaces de g , conformando una red muestreada g^* y luego se observan todos los nodos que conforman dichos enlaces. Las probabilidades de inclusión para los enlaces están dadas por $\pi_{\{i,j\}} = \frac{n^*}{N_e}$, donde N_e es el número de enlaces en la red. Por su parte las probabilidades de inclusión para los nodos dependen de si uno o más enlaces son muestreados y se definen como:

$$\pi_i = \begin{cases} 1 - \frac{\binom{N_e - d_i}{n^*}}{\binom{N_e}{n^*}} & \text{si } n^* \leq N_e - d_i \\ 1 & \text{si } n^* > N_e - d_i \end{cases} \quad (1.5)$$

- **Muestreo en estrella y bola de nieve:** Siguiendo la propuesta de [56], en el muestreo en estrella, un conjunto inicial de nodos N_0^* de tamaño n^* es seleccionado. Luego son observados todos los enlaces en los cuales alguno de los nodos i hayan sido seleccionados en la muestra N_0^* . Asumiendo un muestreo aleatorio simple, las probabilidades de inclusión de los enlaces están dadas por:

$$\pi_{\{i,j\}} = 1 - \frac{\binom{n-2}{n^*}}{\binom{n}{n^*}} \quad (1.6)$$

En cuanto a las probabilidades de inclusión de los nodos, si el nodo fue seleccionado en el conjunto N_0^* , corresponden a $\pi_i = \frac{n}{n^*}$. Por su parte, para los nodos seleccionados después de observar los enlaces de los nodos en N_0^* las probabilidades de inclusión están determinadas por una función de las vecindades de los nodos inicialmente seleccionados ([56]).

En cuanto al muestreo en bola de nieve, este puede entenderse como una extensión del muestreo en estrella, en la cual se parte de una muestra de n^* nodos iniciales N_0^* , se observan los vecinos inmediatos y se itera este procedimiento. El problema de este tipo de muestreo es que las probabilidades de inclusión suelen volverse intratables después de algunos pasos en la iteración.

- **Muestreo por rastreo de enlaces:** Este tipo de muestreo agrupa un conjunto de técnicas de muestreo que se caracterizan porque se seleccionan un conjunto de nodos y se rastrean algunos de los enlaces rastreando nodos adicionales. Según se presenta en [56], para este tipo de muestreo las probabilidades de inclusión están determinadas de acuerdo a la estrategia de rastreo definida en el diseño muestral.

Una vez implementado alguno de los diseños muestrales anteriormente expuestos, existen formulas que permiten estimar algunas de las medidas de las redes presentadas en la primera sección de este capítulo. No se hará énfasis en estos resultados debido a que no se relacionan con los objetivos del trabajo, sin embargo estas estimaciones en su mayoría se basan en los estimadores de Horvitz-Thompson y pueden ser consultadas en [56].

1.3. Modelamiento para redes

En esta sección se presentan algunas herramientas que permiten estudiar los comportamientos de las redes a partir de la construcción y uso de modelos. Existen diferentes fenómenos que se presentan en una red como la formación, el crecimiento, la difusión y el aprendizaje que son de vital importancia en las diferentes aplicaciones. En el análisis de redes es importante entender que algunos de los fenómenos se estudian a partir de modelos estáticos o dinámicos. Tal como se plantea en [49], se entenderá por estático aquel modelo que considere todos los nodos presentes en un mismo momento de tiempo, mientras que se entenderá por dinámico al modelo en donde se considera la introducción de nuevos nodos en el tiempo. Esta sección presenta los principios básicos que permiten entender estos fenómenos de las redes y su modelamiento. No todos los fenómenos ocurren de manera simultánea y, dependiendo de la aplicación, el fenómeno determinante podría ser solamente uno de los mencionados.

1.3.1. Formación de una red

La formación de una red es el fenómeno subyacente que permite el uso de esta metodología: si entre los nodos de interés no se establecen ningún tipo de relaciones, simplemente no es posible llevar a cabo un análisis de redes. Por esta razón, resulta clara la necesidad de establecer modelos de cómo y por qué se forma la red que se está estudiando. La formación de la red en un fenómeno que se estudia a partir de modelos estáticos.

- **Modelo de Erdős-Renyi:** La idea más intuitiva y básica que podría pensarse para modelar la formación de una red es considerar que la creación de los enlaces se produce bajo un proceso

completamente aleatorio. El modelo de Erdos-Renyi fue uno de los primeros modelos de formación y uno de los más sencillos¹. Su planteamiento se produjo dentro de la teoría de grafos (Ver [26]) y corresponde a un modelo para generar grafos aleatorios. Estableciendo la analogía con el análisis de redes, se supone que se tiene un conjunto fijo de n nodos entre los cuales se establecen enlaces de manera independiente con una probabilidad p . De esta forma, se tiene un modelo binomial de formación de enlaces, de manera tal que la probabilidad de tener una red con m enlaces en n nodos es:

$$p^m(1-p)^{\binom{n-1}{2}-m} \quad (1.7)$$

Cada nodo tiene $n-1$ potenciales vecinos con los cuales puede establecer $n-1$ enlaces con una probabilidad p . El grado esperado de un nodo es $(n-1)p$. Este valor es equivalente al grado promedio de la red.

Para n grande y p pequeño, puede aproximarse el modelo binomial con una distribución Poisson, razón por la cual a estos modelos de formación se les conoce más comúnmente como modelos Poisson.

Bajo este modelo, el cardinal del conjunto de nodos está dado por: $|E| = \binom{n}{2}$. La verosimilitud de la probabilidad de enlace p está dada por:

$$L(p; g) = (1-p)^{|E|} \left(\frac{p}{1-p} \right)^{\sum_{(i,j) \in E} x_{i,j}} \quad (1.8)$$

Si se denomina $t = \sum_{(i,j) \in E} x_{(i,j)}$, un estimador de máxima verosimilitud de la probabilidad de enlace p está dada por:

$$\hat{p} = \frac{t}{|E|} \quad (1.9)$$

Según se señala en [49] la importancia de este modelo radica en que la ley de los grandes números permite garantizar que este modelo de formación de enlaces tiene muchas cosas en común con un modelo donde una red tiene un número esperado de enlaces. Este planteamiento permite estudiar propiedades particulares de las redes, en particular cuando su tamaño es suficientemente grande.

- **Modelo de Watts-Strogatz:** Uno de los fenómenos más interesantes de estudio dentro de las redes sociales es el denominado *small world phenomenon*, el cual es un planteamiento que señala, coloquialmente, que a pesar de la existencia de redes grandes, la dimensión del diámetro es pequeña, mientras que el coeficiente de agrupamiento es mucho más grande que el de una red con formación aleatoria (Ver [21]). Una de las teorías fuertemente relacionadas con este fenómeno es la teoría de los seis grados de separación, inicialmente planteada por Frigyes Karinthy. Esta teoría “se refiere a la idea que toda persona está en promedio aproximadamente a seis pasos de cualquier otra persona en el mundo, a través de una cadena de afirmaciones “un amigo de un amigo” que puede ser hecha, para conectar dos personas cualesquiera, en seis pasos o menos”², lo que puede traducirse en que existe un *small world*

¹<http://mokslasplius.lt/rizikos-fizika/en/erdos-renyi-model> Consultado por última vez el 27 de julio de 2014.

²https://www.princeton.edu/~achaney/tmve/wiki100k/docs/Six_degrees_of_separation.html Consultado por última vez el 25 de julio de 2014

phenomenon entre los seres humanos. La base teórica detrás de este planteamiento es el modelo de Watts-Strogatz ([89], [21]). Para el planteamiento de este modelo se parte de una red con n elementos y un número k de enlaces para cada nodo. Con el fin de garantizar que no haya riesgo de una red no-conectada, el valor k debe cumplir la siguiente desigualdad: $1 \ll \ln(n) \ll k \ll n$, en donde $k \ll n$ garantiza que la red este conectada. Adicionalmente se plantea un procedimiento de reconfiguración en el cual cada enlace es renovado con una probabilidad p . En [89] se plantea que para valores intermedios de p el grafo es una *small-world network*, es decir que tiene un coeficiente de agrupamiento alto y un diámetro pequeño.

1.3.2. Crecimiento de una red

El crecimiento de una red corresponde al fenómeno determinado por la aparición de nuevos nodos y el establecimiento de enlaces en la red, razón por la cual la formación es un fenómeno que se estudia a partir de modelos dinámicos. Los fenómenos de crecimiento de redes se producen constantemente en el mundo real: la aparición de una nueva compañía, un nuevo integrante en una institución o un nuevo habitante de un barrio o municipio.

- **Modelo de crecimiento exponencial:** De manera análoga a los modelos de formación aleatoria, este modelo es el planteamiento más intuitivo y simple que puede utilizarse para estudiar el crecimiento en las redes. Según se presenta en [21], en cada incremento del tiempo se tiene un nuevo nodo, el cual se relaciona de manera aleatoria con m de los nodos existentes. Como se presenta en [49], se tiene que el grado esperado para un nodo i en el tiempo t que nació en el tiempo $m \leq s \leq t$ está dado por:

$$m + \frac{m}{s+1} + \frac{m}{s+2} + \cdots + \frac{m}{t} \quad (1.10)$$

o lo que equivale a:

$$m \left(1 + \frac{1}{s+1} + \frac{1}{s+2} + \frac{1}{t} \right) \quad (1.11)$$

Para t grande, esto es aproximadamente:

$$m \left(1 + \log \left(\frac{t}{s} \right) \right) \quad (1.12)$$

Realizando algunos cálculos matemáticos, para t grande, se deduce que la fracción de nodos con grados esperados menores que un valor d (donde $d < m(1 + \log(\frac{t}{m}))$) está dada por:

$$F_t(d) = 1 - e^{-\frac{d-m}{m}} \quad (1.13)$$

De la ecuación anterior, se tiene que el número de enlaces que espera ganar un nodo, que comienza con m enlaces, tiene distribución exponencial con valor esperado m .

- **Modelo Barabási - Albert:** Este modelo se enmarca dentro de lo que se conoce como redes libres de escala (*scale-free networks*), las cuales corresponden a redes cuya distribución de grados sigue una ley de potencias (Ver [49] y [21]). La importancia de este planteamiento

es que incorpora el proceso de Conexión Preferencial (*Preferential Attachment*) en la modelación del crecimiento de la red. Según se presenta en [7], este proceso no se basa en el supuesto de establecimiento de relaciones de forma aleatoria, lo cual se aproxima de una mejor manera al comportamiento de las redes sociales. Para este modelo se asume que los nuevos nodos prefieren formar enlaces con los nodos más conectados. En esta medida, la probabilidad de que un nodo i obtenga un nuevo nodo en el tiempo t , está dada por:

$$m \frac{d_i(t)}{\sum_{j=1}^t d_j(t)} \quad (1.14)$$

En la anterior expresión, m es el número de enlaces que se forman con la aparición de un nuevo nodo, $d_i(t)$ es el grado del nodo i en el tiempo t , mientras que $\sum_{j=1}^t d_j(t)$ corresponde al grado general de todos los nodos existentes en el tiempo t . Lo que indica esta expresión es que los nuevos enlaces son mayormente probables con aquellos nodos que tienen un grado relativamente más grande y por lo tanto existe una *Conexión Preferencial* orientada a nodos que podrían traer mayores beneficios de acuerdo a su posición relativa en la red.

1.3.3. Difusión a través de una red

Difundir es “propagar o divulgar conocimientos, noticias, actitudes, costumbres, modas, etc.”³ El fenómeno de difusión es de vital importancia para el estudio de redes en donde la estructura determina el comportamiento de diversas situaciones. Ejemplos de ello son el contagio de enfermedades, las tendencias de la moda o la tecnología, la popularidad de productos (*marketing viral*), la divulgación de información, la participación o impacto de un programa, etc.

- **El Modelo Bass:** Este modelo fue propuesto por Frank Bass [8] inicialmente, de manera muy breve, como un “*modelo de imitación*”, sin embargo, unos años más tarde se presentan formalmente junto a su derivación matemática y una evidencia empírica aplicada a los modelos de consumo, como se presenta en [9].

El principio del modelo Bass es⁴:

$$\frac{f(t)}{1 - F(t)} = p + \frac{q}{M}[A(t)] \quad (1.15)$$

La anterior ecuación corresponde a que: “la porción del mercado potencial que adopta en el tiempo t , dado que ellos no la habían adoptado, es igual a una función lineal de los adoptantes previos.” En este contexto se entiende por adopción la primera compra de un producto o servicio o la primera vez que se usa una innovación. Por su parte, M corresponde al mercado potencial (el último número de adoptantes), mientras que p y q corresponden a los coeficientes de innovación e imitación, respectivamente. El modelo Bass asume que M es constante. El término $A(t)$ se conoce como la función acumulativa de adoptantes. Se denomina a p el coeficiente de innovación, debido a que parte de la contribución de nuevos adoptantes no depende de los anteriores, sino a la influencia de factores externos al sistema que se estudia. Para el análisis de redes, los adoptantes pueden corresponder a nodos contagiados, informados, impactados por algún programa, etc.

³Diccionario de la Real Academia de la Lengua Española (<http://lema.rae.es/drae/>)

⁴Bass’s Basement Research Institute (<http://bassbasement.org/BassModel/>) Consultado por última vez el 17 de agosto de 2014.

El modelo de Bass más allá de plantear patrones de difusión, no es directamente aplicable al análisis de redes, debido a que no considera la manera en cómo están relacionados los nodos. La importancia de este modelo para el análisis de redes, como se presenta en [49], radica en que aporta herramientas importantes que han permitido derivar modelos de difusión en redes. Dos de los aspectos más relevantes son: incorpora los fenómenos de innovación e imitación, y además, plantea la posibilidad de adopción a partir de factores externos a la red. Sin embargo, este modelo tiene algunas carencias que han dado lugar a modelos más elaborados y enriquecidos.

1.3.3.1. Modelo de difusión SIR

Este modelo fue inicialmente planteado por Kernack y McKendrick y sus siglas corresponden a “*Susceptible, Infectado, Removido*” (SIR). En este modelo se considera que una vez el nodo supera el nivel de infectado (pasa a removido) este queda completamente inmune, y no vuelve a ser contagiado, queda susceptible, o muere.

Existen diferentes versiones del modelo SIR que incorporan algunas variaciones. En general, siguiendo la formulación que se presenta en [49], para este modelo se supone que se tiene una red cuya distribución de grados está dada por $P(d)$. Se asume que el proceso de infección es tal que la probabilidad de que un nodo infecte a algún nodo relacionado, antes que el nodo infectado sea removido, corresponde a un valor tr , que se conoce como probabilidad de transmisión.

Este modelo ha migrado de la epidemiología y resulta de mucha utilidad en el estudio de la propagación de enfermedades como el sarampión, la hepatitis A y la varicela, las cuales generan una inmunidad activa permanente después de haber contraído la enfermedad. Incluso este modelo ha sido utilizado recientemente en el estudio del contagio del chicungunya⁵.

1.3.3.2. Modelo de difusión SIS

En este modelo se considera que un nodo puede tener alguno de dos estados: susceptible o infectado. En este sentido, los nodos pueden recuperarse y volver a ser susceptibles. Las siglas de este modelo corresponden justamente a “*Susceptible, Infectado, Susceptible*”.

Así como en el modelo SIR, este corresponde a un modelo general que tiene diversas variaciones. Siguiendo la formulación de [49], el planteamiento general parte de los grados de un nodo, los cuales pueden expresarse como d_i , que corresponden a d_i interacciones. Estos valores son importantes puesto que las probabilidades de infección están relacionadas con el número de conexiones del nodo. Se denota por $P(d)$ al número de nodos que tienen grado d . Se denomina $\rho(d)$ a la fracción de individuos de grado d que se encuentran infectados en un momento dado. La probabilidad de que una relación se establezca con un individuos infectado corresponde entonces a:

$$\theta = \frac{\sum P(d)\rho(d)d}{\langle d \rangle} \quad (1.16)$$

⁵Yakob, L. & Clements, A.C.A. (2013). *A Mathematical Model of Chikungunya Dynamics and Control: The Major Epidemic on Réunion Island*. PLoS ONE 8(3): e57448.

En este caso se usa la notación $\langle d \rangle$ para denotar el valor esperado de la distribución de grados. Lo anterior corresponde a la idea natural de que un nodo tiene mayor probabilidad de ser infectado en la medida que más enlaces con nodos infectados tenga.

Este modelo es aplicable a la mayoría de las enfermedades contagiosas. Además resulta de mayor utilidad en situaciones que involucran temas como difusión de información, impacto y consumo de productos.

1.3.4. El aprendizaje dentro de una red

Como se ha venido presentando a lo largo de este capítulo, las redes juegan un papel central en diversos fenómenos sociales y dentro de las redes ocurren fenómenos que determinan los comportamientos y decisiones de los individuos. El aprendizaje es otro de los fenómenos que se presenta en la red. De acuerdo con [49], el aprendizaje es un fenómeno que tiene un rol central en contextos tan diversos como las recomendaciones al momento de ver una película, las opiniones sobre un político y la adquisición de habilidades dentro de una empresa. De acuerdo con [39], el aprendizaje puede definirse como el uso de sus experiencias pasadas y las experiencias de los otros para tomar decisiones en el presente. Algunos ejemplos sobre este fenómeno que son presentados en [39] son: la escogencia del consumidor, la innovación de un medicamento, las prácticas agrícolas, la adopción de una nueva tecnología de información, la escogencia de una segunda lengua en el proceso de aprendizaje y el cumplimiento de las normas sociales. El fenómeno del aprendizaje es inherente a la metodología que se desarrolla en la presente tesis, ya que en los MBA, el aprendizaje determinado por un red puede contribuir a un mejor entendimiento del comportamiento de los sistemas complejos. Algunos trabajos han construido un puente entre el aprendizaje en los MBA y la teoría de juegos ([54]). Por su parte, en múltiples trabajos se aborda el tema del aprendizaje en redes desde el enfoque del modelamiento ([19], [59], [49]). En [44] se presenta un completo desarrollo en torno a aprendizaje estadístico. A continuación se presentan dos de los modelos más utilizados en el estudio del aprendizaje dentro de las redes:

1.3.4.1. Modelo de DeGroot

El modelo de DeGroot ([19]) fue uno de los primeros en aproximarse al fenómeno del aprendizaje. Para proponer el modelo el autor plantea la siguiente situación: considere un grupo de individuos quienes deben actuar juntos como un equipo o un comité, y suponga que cada uno de ellos tiene su propia opinión. El autor propone un modelo en el cual se describe cómo el grupo puede lograr ponerse de acuerdo, combinando sus opiniones individuales. El primer paso en este modelo consiste en que el individuo i revisa su propia opinión ponderando las opiniones de los demás individuos del grupo a partir de su información, experiencia, opinión y juicios. En este sentido, el individuo i aprende de los demás. DeGroot plantea que la opinión del individuo en el segundo momento puede representarse como una combinación lineal de las primeras opiniones de los demás individuos y su primera opinión con coeficientes $p_{i1}, p_{i2}, \dots, p_{ik}$. Estos coeficientes representan el peso de la opinión de los otros sobre la opinión del individuo i . El anterior proceso es iterado hasta encontrar convergencia. Cada una de estas iteraciones se conoce como regla de ajuste.

El modelo de DeGroot resulta ser uno de los más sencillos y naturales para el entendimiento de los procesos de aprendizaje. La extensión al análisis de redes resulta de igual

forma sencilla debido a que solamente basta con considerar que la opinión de cada individuo está relacionada con las opiniones de sus vecinos. Tal y como se presenta en [49] este planteamiento ha sido desarrollado ampliamente y existen diversas variaciones que incluyen herramientas más potentes en la modelación, permitiendo incorporar características como la adopción de reglas de ajuste con variaciones en el tiempo y las circunstancias ([20]).

1.3.4.2. Modelo de aprendizaje Bayesiano

El aprendizaje observacional es la denominación que recibe un proceso en el cual, tal y como se plantea en [59], se estudia el aprendizaje de los individuos a partir de la tendencia de estos a imitar el comportamiento de los otros. Bajo este enfoque, se entiende que un individuo adopta alguna decisión o comportamiento cuando es capaz de observar las decisiones de los otros antes de tomar su decisión: se asume que los individuos pueden observar las decisiones de los otros y el resultado que genera. De acuerdo con [49], esto podría entenderse como una situación en donde hay un número k de posibles alternativas. En particular, supóngase que hay 2 posibilidades A y B, y con cada una de ellas se obtiene una ganancia con ciertas probabilidades. Lo que cada individuo buscaría es maximizar su ganancia. Para ello, necesita tener información de las probabilidades asociadas, sin embargo, como se está pensando que el individuo tiene la posibilidad de conocer las decisiones de los otros y sus resultados, esto será relativamente sencillo. El verdadero inconveniente tras este enfoque es que puede llegar a ser bastante complicado en la práctica. Por ejemplo, considérese que un individuo observa que uno de sus vecinos en la red opta por la opción B, lo cual a su vez estaría indicando que los vecinos de ese individuo han optado por esa opción y han obtenido buenos resultados. Esto hace que el estudio del aprendizaje presente retos no triviales ([34]). Una alternativa más aceptada y sencilla de implementar fue propuesta por [4]. En este caso los individuos solo procesan la información de las decisiones y resultados de los vecinos pero ignoran cualquier información indirecta de las secuencias de decisiones de ellos.

El modelo de aprendizaje propuesto por [4] se basa en un conjunto finito Θ de posibles estados del mundo, un conjunto de acciones X y Y un espacio de respuestas. Si el estado del mundo es $\theta \in \Theta$ y un nodo escoge una acción $x \in X$, él observa la respuesta y con densidad condicional $\phi(y; x; \theta)$ y obtiene una recompensa $r(x, y)$. Los supuestos de este modelo son:

- (I) Y es un espacio no-vacío con métrica separable. La distribución de las respuestas condicionales en x and θ puede ser representado por una función de densidad $\phi(\cdot; x; \theta)$ con respecto a una medida Γ definida en los subconjunto de Borel de Y .
- (II) Para cada $x \in X$, $r(x; \cdot)$ es acotada y medible en Y .

Los nodos no conocen el verdadero estado del mundo, y ellos entran con una creencia a priori en el conjunto de creencias sobre el estado de la naturaleza, denotado $\mathcal{D}(\Theta)$, tal que:

$$\mathcal{D}(\Theta) = \{\mu = \{\mu(\theta)\}_{\theta \in \Theta} \mid \text{para todo } \theta \in \Theta, \mu(\theta) \geq 0 \text{ y } \sum_{\theta \in \Theta} \mu(\theta) = 1\} \quad (1.17)$$

El parámetro μ representa la probabilidad que le da el nodo a que el parámetro verdadero sea θ . Dada esta creencia μ , la utilidad esperada de un nodo en un período es una función que depende de esa creencia y de la acción que toma $u(x, \mu)$ y está dada por:

$$u(x, \mu) = \sum_{\theta \in \Theta} \mu(\theta) \int_Y r(x, y) \phi(y; x; \theta) d\Gamma(y) \quad (1.18)$$

Sea $G : \mathcal{D}(\Theta) \rightarrow X$ la función de optimización de un período, la cual equivale a:

$$G(\mu) = \{x \in X | u(x, \mu) \geq u(x', \mu) \text{ para todo } x' \in X\}, \mu \in \mathcal{D}(\Theta) \quad (1.19)$$

Sea δ_θ el punto de masa de creencia en el estado θ ; entonces $G(\delta_\theta)$ denota el conjunto de acciones ex-post.

1.4. Inferencia estadística en redes

Los temas de inferencia estadística en redes hacen relación al uso de la información de una red y de algunas variables de interés de los nodos y/o los enlaces para estudiar el comportamiento de uno o más parámetros de interés. Tal y como sucede en inferencia estadística clásica, en lo que se piensa es que se tiene una muestra aleatoria (independiente e idénticamente distribuida) de una variable aleatoria X , dada por X_1, \dots, X_n , a partir de la cual puede establecerse un estadístico T , definido como una función de la muestra aleatoria, $T(X_1, \dots, X_n)$ y que sirve para estimar el parámetro o parámetros de interés, denotado θ . En el caso de las redes, siguiendo el planteamiento de [56], se supone que se tiene información de los nodos y/o los enlaces de una muestra de redes g de una colección de potenciales redes \mathcal{G} a partir de la cual se usan herramientas de modelamiento e inferencia estadística para contribuir al estudio de características como:

1. Predicción de enlaces a través del uso de métodos de scoring y métodos de clasificación probabilística basados en regresión logística.
2. Estudio de asociación y correlación en redes. Por ejemplo el estudio de los datos de microarreglo en la regulación génica y el estudio de la correlación entre niveles de expresión génica.

Como se mencionó en la introducción, el énfasis de este trabajo es el tema de modelamiento, razón por la cual no se abordan con mayor detalle los temas de inferencia estadística en redes. Más información puede encontrarse en [56].

1.5. Modelamiento y predicción para redes conocidas

Según se presenta en [56], existen una serie de herramientas enfocadas al estudio del comportamiento de redes conocidas, para llevar a cabo procesos de modelamiento y predicción. En este conjunto de herramientas se destacan por ejemplo los modelos gaussianos, modelos gravitacionales y modelos de Poisson, el uso de estadística no paramétrica y campos aleatorios de Markov ([56]).

En particular, los campos aleatorios de Markov son un tipo de procesos estocásticos que son una generalización de las cadenas de Markov, en la cual se pasa de un conjunto de índices de tiempo a índices de espacio, es decir, de un conjunto de índices unidimensional a uno bidimensional. De acuerdo con [56], para la generalización en el caso de redes, considérese una red (N, g) , y $\mathbf{X} = (X_1, \dots, X_n)^T$ una colección de variables aleatorias discretas definidas sobre N . Se dice que \mathbf{X} es un campo aleatorio de Markov si:

$$P(X_i = x_i | \mathbf{X}_{(-i)} = \mathbf{x}_{(-i)}) = P(X_i = x_i | \mathbf{X}_{N_i} = \mathbf{x}_{N_i}), \quad (1.20)$$

donde $\mathbf{X}_{(-i)}$ es el vector $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{N_v})$ y $\mathbf{X}_{\mathcal{N}_i}$ es el vector de todas las X_j , para $i \in \mathcal{N}_i$, siendo \mathcal{N}_i la vecindad del nodo i .

Los modelos basados en campos aleatorios de Markov permiten estudiar los comportamientos de variables de interés cuando puede suponerse que la probabilidad de que esa variable de interés asuma un estado, para un individuo i , solamente depende del comportamiento de sus vecinos en la variable de interés. Este supuesto es común cuando se hacen análisis de redes.

Las herramientas estadísticas presentadas a lo largo de este capítulo son un insumo que permite continuar con el estudio del análisis de redes desde una perspectiva estadística y pueden abordarse con mayor profundidad en trabajos de investigación futuros. El enfoque de este capítulo se orientó principalmente al tema de modelamiento, ya que este es retomado en el capítulo 3 cuando se plantea la combinación propuesta entre el análisis de redes y los MBA. El propósito de este capítulo, además de presentar las herramientas de la metodología que se abordarán más adelante, fue poner en evidencia los estrechos vínculos entre la estadística y el desarrollo teórico del análisis de redes.

MODELOS BASADOS EN AGENTES (MBA)

La relación entre los MBA y la estadística es muy estrecha y puede ser encontrada en todas las etapas de los MBA. En [78] se plantean 5 ejes temáticos que demandan herramientas estadísticas: en primer lugar, al contar con datos empíricos, pueden utilizarse métodos estadísticos para derivar patrones, ajustar modelos de regresión y construir y parametrizar las reglas; en segundo lugar, las reglas de los MBA dependen de métodos estadísticos aplicados durante la simulación; en tercer lugar, cuando el modelo es formulado e implementado, algunos parámetros son desconocidos y deben utilizarse métodos de modelamiento inverso; en cuarto lugar, al implementarse el MBA, es necesario usar análisis de sensibilidad e incertidumbre para entender mejor el sistema; finalmente, la estadística es muy potente en el momento de analizar los resultados de los modelos a través de herramientas descriptivas e inferenciales. Por otro lado, según se plantea en [45], los MBA resultan de especial interés para la estadística, ya que los datos de entrada y salida de los MBA pueden usarse como datos de estudio de relaciones complejas. Adicionalmente, bajo un enfoque bayesiano, se pueden plantear escenarios para llevar a cabo procesos de inferencia estadística. Estos temas serán abordados a lo largo de este capítulo y el capítulo 3, haciendo énfasis principalmente en los temas de validación, calibración, modelamiento e inferencia estadística. Adicionalmente, se presentan de manera general algunos de los componentes y herramientas de la metodología y se proveen algunas definiciones que permitirán un mejor entendimiento de los temas abordados.

De acuerdo con [68], los MBA han mejorado en calidad y confiabilidad en los años recientes debido a que han venido siendo tratados desde el enfoque del Modelamiento Basado en Patrones (Pattern-Oriented Modelling - POM). El Modelamiento Basado en Patrones se basa en la idea de desarrollar modelos que reproduzcan los diferentes patrones observados en el sistema y estudiar qué tanto el esquema computacional reproduce esas características. Este enfoque es el que ha permitido la implementación de herramientas estadísticas para las diferentes etapas de los MBA.

2.1. ¿Qué son los MBA?

Desde el punto de vista estadístico, de acuerdo con [73] los MBA, dependiendo su formulación, pueden llegar a ser vistos como modelos no-lineales complejos con los cuales, resueltas las dificultades relacionadas con el modelamiento y la parametrización, pueden aprenderse lecciones relacionadas con cambios en los parámetros. Dichos cambios pueden ser consecuencia de un

desarrollo inesperado (análisis de escenarios) o acciones propuestas implementadas (análisis de políticas).

Los MBA es una metodología que tiene sus orígenes en el planteamiento de los físicos y matemáticos Jhon von Neumann y Stanislaw Ulam sobre el autómata celular (*Cellular automaton*) en el año de 1940 ([51]). El autómata celular es un arreglo de células, en donde cada una de ellas tiene un conjunto finito de posibles estados. El arreglo comienza en un tiempo $t = 0$, en el cual las células tienen un estado inicial. Los estados van cambiando en el tiempo basados en la propia historia y la de la vecindad. El objetivo de esta formulación era estudiar fenómenos físicos dinámicos a partir de unidades discretas. A pesar de las bases teóricas que se establecen en la propuesta de von Neumann y Ulam, los MBA logran su auge desde la década de 1990 debido a los requerimientos computacionales de esta metodología. Esta metodología tiene estrechas relaciones en su formulación y desarrollo con el aprendizaje de máquinas, en particular, con los temas de inteligencia artificial y modelos adaptativos ([72]).

Un agente, en este contexto, es una entidad computacional a la cual se le pueden programar un conjunto de características relacionadas con heterogeneidad, autonomía, percepción, aprendizaje, memoria, etc. Según se presenta en [58], los MBA es un enfoque metodológico que hace parte de los métodos de microsimulación y que permite estudiar la interacción de variables dentro de un sistema complejo e investigar la interacción a dos escalas diferentes: la estructura a nivel macro y a nivel micro. Se entiende por sistema complejo aquel en donde existe una múltiple dependencia entre las variables de estudio y en donde los comportamientos a nivel macro no solamente corresponden a un agregado de los comportamientos micro sino que se basa en la capacidad de aprendizaje a nivel individual. Esta metodología se enmarca dentro de aquellas que utilizan herramientas computacionales para resolver, a través de simulación, diferentes problemas. Dentro de estos enfoques, los que presentan mayor reconocimiento y desarrollo son los Sistemas Multiagentes, la Inteligencia Artificial y la Programación Orientada a Objetos. Según se presenta en [91], los MBA se diferencian de otros métodos, debido a que la simulación de sistemas complejos se basa en el nivel micro, nivel en el cual las decisiones son modeladas como agentes autónomos inmersos en un ambiente dinámico. A diferencia de otros métodos de microsimulación, en los MBA se facilita la simulación de sistemas complejos debido a la descomposición del proceso desde un enfoque real que comienza por los individuos, enfocándose en el modelamiento de sus comportamientos; los patrones, estructuras y comportamientos del sistema emergen de la interacción entre agentes.

De acuerdo con [87], existen varias razones por las cuales, al menos en demografía, los métodos de microsimulación pueden ser preferidos sobre las técnicas de proyección a nivel macro, entre ellas: (1) Las microsimulaciones permiten tratar más fácilmente con poblaciones donde el espacio-tiempo es grande, o donde covariables, continuas como el ingreso, están envueltas; (2) Las microsimulaciones permiten incluir variables de salida más compuestas; (3) Las microsimulaciones permiten tratar con relaciones entre individuos. Por otro lado, en términos de [6], la adopción de sistemas complejos permiten el modelamiento de estructuras económicas desde una perspectiva ascendente, entendida como resultado de la interacción repetida de los agentes, sin ignorar las consecuencias de sus estructuras en el comportamiento individual, el surgimiento de patrones y el bienestar social. Por su parte, según [86], la metodología de los MBA es la más apropiada para modelar sistemas adaptativos complejos, además de ser uno de los enfoques más novedosos, es más general y potente. La principal diferencia entre los MBA y otros métodos es que en los MBA la simulación se hace programando de manera explícita las reglas que determinan el comportamiento micro mientras que el comportamiento a nivel macro emerge de la simulación de los agentes.

Según se expone en [41], los MBA son especialmente útiles cuando alguno de los siguientes aspectos de los individuos es importante para explicar el comportamiento de un sistema: la heterogeneidad entre individuos, las interacciones locales y el comportamiento adaptativo basado en toma de decisiones. Este comportamiento adaptativo puede ser interpretado desde la perspectiva de la inteligencia artificial, esto es, los agentes aprenden del estado del sistema y de sus resultados previos y se adaptan para tomar decisiones futuras ([72]).

2.2. Un enfoque de procesos estocásticos para los MBA

De acuerdo con [40], los MBA son sistemas recursivos, lo que significa que el comportamiento en un momento $t + 1$ puede ser explicado como función del comportamiento del sistema en el momento t . En particular, cuando se tiene una variable de salida discreta, según se presenta en [73], [40], [6], los MBA pueden ser vistos como Cadenas de Markov en donde el estado del sistema en el tiempo t está dado por la colección de los microestados en el tiempo t , esto es, $\mathbf{X}_t \equiv \{x_{it}\}$, $i = 1, \dots, N$, $t = 1, \dots, T$, con:

$$x_{i,t+1} = \mathbf{f}_i(\mathbf{X}_t, \Xi_t, \theta) \quad (2.1)$$

donde \mathbf{f}_i es una función que toma valores en \mathcal{R}^k , $\Xi_t \doteq \{\xi_{it}\}$ es un vector de elementos estocásticos y $\theta \in \Theta$ es un vector de parámetros.

Como representación del nivel macro, un conjunto de estadísticas agregadas $\mathbf{y}_t \equiv \{y_{kt}\}$, $k = 1, \dots, K$ puede ser definido como una función de \mathbf{X}_t , de forma que:

$$\mathbf{y}_{t+1} = g(\mathbf{X}_t, \Xi_t, \theta) \quad (2.2)$$

Como consecuencia de las propiedades de las cadenas de Markov, se tiene que:

$$\mathbf{y}_{t+1} = \mathbf{g}^t(\mathbf{X}_0, s, \theta) \quad (2.3)$$

donde s representa los valores iniciales de la microsimulación.

De acuerdo con [73] los análisis relacionados con los resultados de los MBA deben orientarse a submuestras de la simulación que tengan propiedades de estacionariedad y que en esta característica resulten cuantitativamente similares a los datos reales. En caso de no encontrar estacionariedad en la submuestra de interés, se puede optar por usar transformaciones.

2.3. Componentes de los MBA y características de los agentes

Dependiendo la aplicación en la cual se estén utilizando los MBA, el planteamiento del problema conduce a un desarrollo diferente del proceso de modelación y simulación. Sin embargo, existen algunos elementos y características comunes que se presentan a continuación. En primer lugar, el proceso se puede bosquejar de la siguiente manera según se presenta en [35]:

Como se observa en el anterior esquema, son tres los componentes que definen el proceso: los datos empíricos, el modelo y la simulación. Con base en unos datos empíricos existentes, se propone un modelo y se realiza la simulación, la cual se contrasta con información empírica exis-

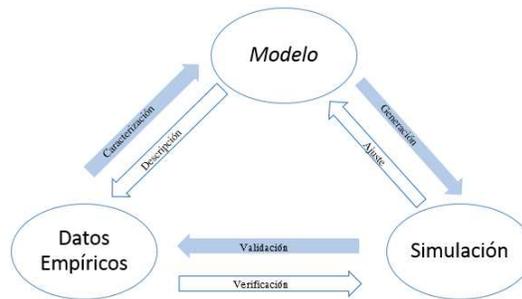


FIGURA 2.1. Proceso de los MBA. Tomada de [35]

tentes para observar si el modelo propuesto es factible o debe volverse a plantear su formulación. En este esquema también es claro que cada uno de estos componentes desempeña dos funciones dentro del proceso:

- **Los datos empíricos** sirven para (a) caracterizar el modelo, en la medida en que se utiliza información previa para inicializar los parámetros del mismo, y (b) para verificar que los resultados de la simulación sean factibles.
- **El modelo** sirve para (a) describir la estructura que se manifiesta en los datos empíricos, con ayuda de estudios previos que hayan analizado la relación existente entre diferentes factores de interés, y (b) ajustar los resultados de la simulación.
- **La simulación** sirve para (a) generar datos que correspondan a la estructura del modelo, para así experimentar con ellos, y (b) verificar que el comportamiento obedezca a la información empírica.

Por su parte, en [25] se enumeran un serie de componentes de los MBA, que se complementan con lo propuesto en [86]. A continuación se describen brevemente estos elementos:

- I. Agentes: Los agentes son objetos computacionales que pueden representar individuos, familias, organizaciones, etc., y que corresponden a las unidades de interés dentro del sistema que se estudia. Los agentes pueden tener las siguientes características:
 - **Autónomos:** Los agentes son autónomos en el sentido que no hay decisiones centrales que determinen sus comportamientos.
 - **Heterogéneos:** Los agentes son heterogéneos, ya que pueden tener diferentes atributos, reglas y comportamientos.
 - **Percepción:** Los agentes pueden percibir su entorno, incluyendo otros agentes en su vecindad.
 - **Desempeño:** Tienen un conjunto de comportamientos que son capaces de llevar a cabo, tales como el movimiento, la comunicación con otros agentes y la interacción con el ambiente.
 - **Memoria:** Los agentes tienen memoria y en ella recuerdan sus estados previos y acciones.
 - **Principios:** Los agentes tienen un conjunto de reglas, heurísticas o estratégicas que determinan, dada su presente situación y su historia, que harán.

- II. El estado: Se denomina estado a toda la información asociada a los agentes en un momento dado. En este sentido el estado incluye los valores que toman los agentes en las diferentes variables que definen el sistema. Esta información podría ser pública para los otros agentes, o por el contrario podría no ser observable por ellos.
- III. Las reglas: En un MBA, según se presenta en [86], este concepto corresponde más exactamente a “reglas de decisión mecánica o funciones de transformación” y hacen referencia a los patrones que determinan la forma en cómo ocurren transformaciones en los estados y cómo las acciones de los agentes están causadas por ellos. Un ejemplo clásico de regla en este contexto es que los agentes busquen maximizar sus utilidades y minimizar los costos.
- IV. El ambiente: El ambiente es el lugar en donde los agentes se ubican, contiene toda la información externa utilizada en los procesos de toma de decisiones y proporciona una estructura o espacio en donde los agentes interactúan. El ambiente contiene todo, incluyendo a los otros agentes. Los agentes pueden afectar el ambiente y pueden ser afectados por este.

Es importante resaltar la importancia que puede jugar el análisis de redes como complemento de esta metodología: el estado y las reglas están fuertemente influenciados por la forma en cómo se dan las relaciones entre los agentes y en esta medida implementar herramientas de redes a este contexto puede ayudar a modelar las interacciones sociales entre los agentes.

Un MBA puede definirse como un programa computacional que crea un mundo de agentes que interactúan con otros agentes y con el ambiente. Lo que se busca en últimas es tratar de reproducir un fenómeno del mundo real para estudiar como responde frente a variaciones en alguno o varios de sus componentes. Pero, ¿cómo interactúan con los otros individuos y con el ambiente? ¿cómo toman las decisiones y cómo pueden incorporarse las características de autonomía, heterogeneidad, percepción, desempeño, memoria, principios? ¿cómo puede implementarse parte de la complejidad de los procesos relacionados con los individuos, familias, organizaciones, etc. a través de un programa computacional? Estas inquietudes son la fuente de preguntas de investigación en esta metodología y a continuación se presentarán algunas herramientas importantes para tratar de dar respuesta a estos y otros aspectos.

Siguiendo el planteamiento de [58], el modelo y la simulación de un MBA pueden llevarse a un planteamiento basado en un conjunto de ecuaciones. Para ello, se asume que en cada tiempo t un individuo i , $i \in 1, \dots, n$ está bien descrito por una variable de estado $x_{i,t} \in \mathbb{R}^k$, cuya evolución puede especificarse a través de la ecuación:

$$x_{i,t+1} = f_i(x_{i,t}, x_{-i,t}; \alpha_i), \quad (2.4)$$

en donde se asume que las reglas del MBA para el individuo i pueden ser especificadas a nivel de los agentes a través de la forma funcional $f_i(\cdot)$ y el conjunto de parámetros α_i y con base en los estados de los otros x_{-i} individuos. Dada esta especificación del comportamiento a nivel individual, interesa estudiar el comportamiento a nivel agregado, el cual se representa a través de una estadística Y definida sobre todos los agentes:

$$Y_t = s(x_{1,t}, \dots, x_{n,t}) \quad (2.5)$$

Usando las dos ecuaciones anteriores para cada t puede especificarse una ecuación que puede ser resuelta de manera iterativa tal y como se presenta en el siguiente sistema:

$$\begin{aligned}
Y_0 &= s(x_{1,0}, \dots, x_{n,0}), \\
Y_1 &= s(x_{1,1}, \dots, x_{n,1}) \\
&= s(f_1(x_{1,0}, x_{-1,0}; \alpha_1), \dots, f_n(x_{n,0}, x_{-n,0}; \alpha_n)) \\
&= g_1(x_{1,0}, \dots, x_{n,0}; \alpha_1, \dots, \alpha_n) \\
&\vdots \\
Y_t &= g_t(x_{1,0}, \dots, x_{n,0}; \alpha_1, \dots, \alpha_n)
\end{aligned}$$

Las anteriores ecuaciones permiten que se formalice un MBA en el contexto de los modelos dinámicos a nivel micro, lo que facilita que puedan ser implementadas algunas herramientas aplicadas en otros contextos de modelación.

2.3.1. Arquitectura de los agentes

Los agentes tienen asociados un estado y unas reglas que determinan sus acciones. El conjunto de estos tres elementos puede ser de una diversa complejidad según el MBA que se esté planteando para estudiar el sistema de interés. Se define como arquitectura de los agentes a las especificaciones vinculadas con los agentes y sus acciones.

Según se plantea en [25], el estándar en la arquitectura de los agentes de los MBA es el esquema de Creencia-Deseo-Intención o BDI, por sus siglas en inglés (Belief-Desire-Intention), el cual corresponde a una concepción psicológica de que las acciones están determinadas por creencias, deseos e intenciones. Existen otros esquemas de arquitectura de agentes como el *Adaptive Control of Thought - Rational (ACT-R)*, el *Connectionist Learning with Adaptive Rule Induction On-line (CLARION)*, redes neuronales y Soar, sin embargo, según lo afirman los autores, el esquema BDI resulta relativamente más simple frente a estos esquemas, pero es altamente cognitivo y resulta de mucha utilidad en múltiples aplicaciones en donde puede adoptarse y ser implementado. A su vez, se señala una desventaja que es común a todos los esquemas de arquitectura de los agentes y es relacionada con que los agentes son átomos sociales que no tienen preestablecidas interacciones sociales. Estos autores presentan el gráfico 2.2, el cual resume el esquema BDI:

Como un intento de solución a la desventaja señalada se ha desarrollado el esquema de Creencia-Obligación-Intención-Deseo o BOID, por sus siglas en inglés (Belief-Obligation-Intention-Desire)^{1 2}. El gráfico 2.3 resume el esquema BOID:

Las obligaciones son entendidas como reglas que restringen el comportamiento de los agentes en caso de conflictos sociales. Bajo este esquema suelen presentarse conflictos entre los deseos y las obligaciones. Sin embargo, este esquema permite incorporar comportamiento normativo en el marco de los agentes. Una de las principales críticas a este esquema es que es un enfoque estático y no considera cambios en el comportamiento normativo; la otra crítica, que es común a los esquemas de arquitectura de agentes mencionados es que diferentes aspectos de los individuos pueden escaparse a la modelación computacional.

¹Castelfranchi, C., Dignum, F., Jonker, C. M., & Treur, J. (1999). Deliberate normative agents: Principles and architecture. In N. Jennings & Y. Lesperance (Eds.), *Intelligent agents VI. Volume proceedings of the sixth international workshop on agent theories, architectures, and languages ATAL 99*, Orlando (pp. 364 – 378).

²Dignum, F., Morley, D., Sonnenberg, E., & Cavedon, L. (2000). Towards socially sophisticated BDI agents. In E. Dufree (Ed.), *Proceedings of the international conference on multi-agent systems, ICMAS 00*, Boston (pp. 111?118).

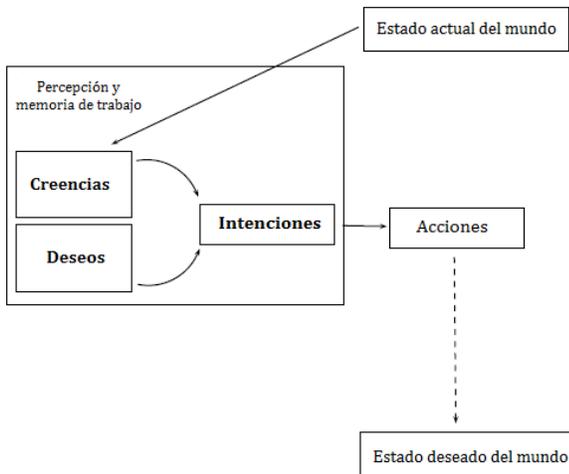


FIGURA 2.2. Esquema BDI para la arquitectura de agentes. Tomada de [25]

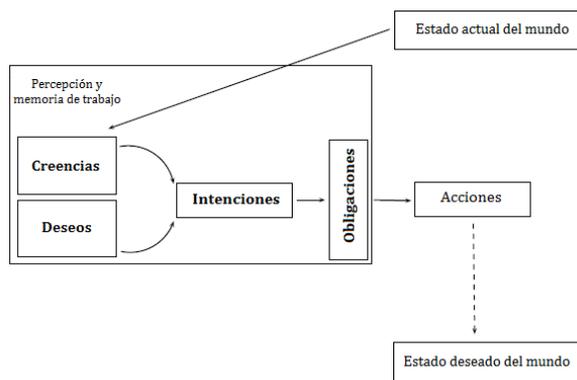


FIGURA 2.3. Esquema BOID para la arquitectura de agentes. Tomada de [25]

2.3.2. Calibración y validación de los MBA

Según se plantea en [64] la validación es el proceso que provee información necesaria para evaluar si el modelo provee resultados válidos y robustos y qué tanto se aproxima al sistema del mundo real y satisface los objetivos de el modelo desarrollado. En [45] se sugiere que la validación es un componente crucial dentro del desarrollo de los MBA en la medida que este proceso asegura que la simulación sea una representación razonable del sistema que se estudia y que a su vez esta logre reproducir los comportamiento del sistema a gran escala. A la par de la validación del modelo, el proceso de calibración busca “identificar un rango de valores para los parámetros del modelo y afinarlos en el ajuste a datos reales. Esto es llevado a cabo tratando el modelo general como una caja negra y usando métodos eficientes de optimización para encontrar el ajuste óptimo de los parámetros” [64].

Siguiendo el planteamiento de [64], existen varios tipos de validación como la empírica, la estadística, la conceptual, la interna, la operacional, la externa, la estructural y la de procesos. Estos tipos de validación son clasificados en tres grupos: *Validación replicativa*, *Validación predictiva* y *Validación estructural*. Este último será el tipo de validación utilizada, haciendo énfasis en la validación estadística, dado que la validación estructural evalúa si el modelo reproduce el sistema

observado y adicionalmente, sí este refleja la forma en la cual el sistema real opera para producir este comportamiento.

Siguiendo a los autores, quienes citan a Carley³ y Klügl⁴, son 4 los pasos que definen este tipo de validación, los cuales se presentan a continuación. La validación es uno de los aspectos que presentan mayor desarrollo en la actualidad. En [79] se presenta una interesante recopilación de las metodologías más recientes, las cuales incluyen herramientas del diseño de experimentos, métodos de optimización, algoritmos genéticos y métodos bayesianos.

- Validación de entrada: Este paso corresponde a una validación de proceso y reúne las siguientes etapas: (a) Una revisión general del comportamiento del sistema y de los agentes. (b) Monitorear la dinámica de un agente particular durante la ejecución del modelo. (c) Establecer que los valores de salida estén en un rango de valores reales y que las tendencias sean consistentes entre las diferentes simulaciones.
- Análisis de sensibilidad: En este paso se busca evaluar cuál es el efecto que tiene la alteración de los valores de las variables y de la estructura del modelo sobre las salidas generales del modelo. Los parámetros que no tengan un efecto significativo se remueven del modelo con el fin de hacerlo más parsimonioso. Para llevar a cabo el análisis de sensibilidad, la revisión bibliográfica conduce a una interesante alternativa que usa herramientas del diseño de experimentos para evaluar de manera simultánea el efecto de los parámetros en el modelo ([55, 43]).
- Calibración: En esta instancia se identifica el rango de valores posibles para los parámetros y se ajusta el modelo con datos reales. La calibración del modelo involucra además un proceso de optimización para encontrar el conjunto final de parámetros del modelo. Según se presenta en [66] los métodos de escalamiento y los algoritmos heurísticos son los más utilizados y difundidos para llevar a cabo la calibración en los MBA. En particular, los algoritmos genéticos y la optimización de Pareto son métodos que los autores identifican como los más apropiados para llevar a cabo la validación en esta etapa.
- Validación de salida: Este paso consiste en comparar los datos de la predicción del modelo contra los datos reales a través de herramientas gráficas y estadísticas. En [64] se presentan dos metodologías para llevar a cabo esta etapa de validación: por un lado, la técnica ROC (*Relative Operating Characteristic*, por sus siglas en inglés) compara los valores predichos con una clasificación determinada por un umbral o punto de corte de los valores de la variable que se analiza, obteniendo así 4 posibles alternativas: verdadero positivo, falso positivo, verdadero negativo y falso negativo. La validación de salida se lleva a cabo con ayuda de una curva, denominada curva de ROC, que provee una medida de la habilidad del modelo para clasificar correctamente o predecir valores de las variables del sistema. Por otro lado, se presenta la técnica de bondad de ajuste de resolución múltiple o *Multiple Resolution Goodness of Fit (MRG)*, la cual es aplicable especialmente en modelos con salidas que involucren patrones de espacio o tiempo.

Estos cuatro pasos son secuenciales y en este sentido debe haber un resultado satisfactorio en cada uno de ellos para pasar al siguiente. En el capítulo 3 se profundizará en este aspecto de

³Carley, K. M. (1996). *Validating computational models* (Working paper). Pittsburgh: Carnegie Mellon University.

⁴Klügl, F. (2008). *A validation methodology for agent-based simulations*. In *Proceedings of the 2008 ACM Symposium on Applied Computing - Advances in Computer Simulation*. ACM. pp 39-43. New York, USA.

validación y calibración y se hará énfasis en los temas relacionados con modelamiento e inferencia estadística a partir de MBA.

PROPUESTA METODOLÓGICA Y DESARROLLO ESTADÍSTICO

En los dos capítulos anteriores se abordaron temas importantes para el entendimiento de los enfoques metodológicos del análisis de redes y los MBA y se presentó su relación con la estadística. Como se ha mencionado, el análisis de redes permite tomar en cuenta la estructura de dependencia entre individuos, entendiéndola como una red en donde diferentes fenómenos están parcialmente determinados por la vecindad de los individuos y, en general, por las relaciones que se establecen entre ellos. Por su parte, los MBA son especialmente útiles para estudiar la interacción de variables dentro de un sistema complejo a partir de la simulación de agentes, en conjunto con sus comportamientos y reglas. Con este breve resumen de las dos metodologías, puede pensarse de manera directa en la posibilidad de complementarlas para formular una metodología más potente que permita incorporar de manera simultánea aspectos de las dos técnicas originales.

Como se mencionó en la introducción, el análisis de redes y los MBA son dos metodologías que tienen su origen y su mayor desarrollo en campos de las humanidades y de la sociología, sin embargo se han venido utilizando en trabajos relacionados con finanzas [60, 81, 15], economía [3, 46] y estadística [45, 47]. En este capítulo se presenta con mayor profundidad los temas de modelamiento e inferencia estadística de los MBA y se desarrolla la propuesta metodológica que combina las dos técnicas abordadas en los capítulos 1 y 2, junto con su desarrollo estadístico.

3.1. ¿Por qué el Análisis de Redes y los MBA?

El análisis de redes ayuda a abordar fenómenos en donde el paradigma de la independencia entre individuos no puede suponerse. Esto permite estudiar la forma en cómo las relaciones entre individuos afectan temas como la difusión de enfermedades, la compra de bienes, servicios y seguros, los idiomas que se estudian, la forma de votar, los comportamientos criminales, la probabilidad de éxito profesional, entre otros temas en donde la interacción entre individuos parece jugar un rol muy importante.

Por su parte, según [45], los MBA son una alternativa cuando se quiere entender y predecir el comportamiento y los procedimientos de un sistema complejo pero los individuos de interés no

pueden ser observados de manera lo suficientemente frecuente para obtener toda la información necesaria.

Basado en [88] los siguientes son aspectos que pueden incluirse dentro de los MBA y que puntualizan el porqué esta es una metodología apropiada para abordar el estudio de algunos temas:

- Se producen relaciones a nivel de agentes. Existe una interacción con otros agentes de manera dinámica a través del ambiente del modelo, de acuerdo a un conjunto de algoritmos que son establecidos de forma teórica o empírica.
- Pueden incorporarse herramientas de retroalimentación y umbrales críticos a través de especificaciones no lineales de comportamientos de los agentes, los cuales pueden generar cambios en unidades sociales más grandes. En este contexto, se entiende por umbral crítico el hecho de que hayan cambios en el comportamiento a nivel agregado cuando cierto porcentaje de agentes hayan cambiado de estado. Además, la retroalimentación puede operar a través de efectos composicionales, redes sociales, relaciones espaciales o el ambiente e incluir fenómenos de aprendizaje y adaptación de los agentes.
- Se impone la solución computacional, lo que permite que hayan especificaciones altamente no lineales y una mayor complejidad en las ecuaciones subyacentes. Sin embargo, también pueden incorporarse componentes estocásticos.

Los MBA son actualmente usados en estudios que involucran campos de la ciencia muy diversos, mientras que existe un creciente interés por la consolidación de sus bases teóricas y su uso en la investigación científica de alto nivel. Ejemplo de esto es que el Max Planck Institute for Demographic Research se planteó durante el año 2015 un importante objetivo de ayudar a difundir innovaciones importantes relacionados con MBA para mejorar las bases teóricas de esta metodología¹ y para ello definió tres pilares sobre los cuales soportar el trabajo futuro: (a) teoría de probabilidad, especialmente teoría de procesos estocásticos; (b) teoría de decisión / teoría de selección / teoría de acción; y (c) teoría de redes sociales. En este último aspecto existe una coincidencia con la propuesta metodológica que se desarrolla durante la presente tesis.

Hay dos aspectos que son de vital importancia dentro del desarrollo, mejora y expansión de los MBA: el proceso de validación y la inferencia. En estos dos aspectos la estadística tiene un rol muy importante. A continuación se presentan las generalidades de la metodología propuesta y simultáneamente se plantea el uso de herramientas estadísticas para su desarrollo.

3.2. Generalidades de la metodología propuesta

Los fenómenos del mundo real suelen corresponder a sistemas complejos en donde múltiples variables intervienen y tienen influencia sobre algunos resultados de interés. Partiendo de esto, la técnica base para esta formulación metodológica será los MBA. Una particularidad de esta propuesta será la definición de agentes estocásticos, para los cuales sus atributos son variables aleatorias, cuya distribución empírica se deriva de lo encontrado en datos auxiliares de la población. Sin embargo, los agentes simulados en las aplicaciones abordadas con estas técnicas suelen

¹http://www.demogr.mpg.de/en/education_career/international_advanced_studies_in_demography_3279/upcoming_courses_3762/a_stochastic_process_approach_to_agent_based_modeling_and_simulation_4103/default.htm

carecer de estructuras de base que permitan introducir interacciones sociales en su planteamiento ([25]). En este punto es donde el análisis de redes empieza a jugar un papel importante en dos aspectos básicos:

- **La estructura:** Un MBA resulta ser un programa computacional que crea un mundo de agentes autónomos y heterogéneos en el cual cada agente interactúa con otros agentes y con el ambiente ([25]). Sin embargo, la manera en cómo suelen abordarse esas interacciones depende de la aplicación y de esquemas de construcción de agentes. El planteamiento de una estructura en redes en donde los individuos se relacionan con otros, a los cuales se le llaman vecinos, y esos vecinos son justamente quienes se involucran en sus decisiones, opiniones, aprendizaje, difusión de información, etc., se presenta como una alternativa interesante para esquematizar las interacciones entre agentes.
- **Modelamiento del comportamiento de la red:** Tal y como se presentó en el capítulo 1, existen diferentes fenómenos que ocurren en las redes que ayudan a estudiar acontecimientos sociales muy útiles cuando se desea estudiar un sistema complejo. La idea de la metodología es proveer las herramientas para modelar la formación, el crecimiento, la difusión y el aprendizaje, según las necesidades del MBA. Esto ayudará a incorporar herramientas dinámicas en los MBA y aproximarse, de una mejor manera, al estudio del efecto de las variables de un sistema complejo sobre los resultados de interés.

3.2.1. Esquema

Los MBA parten de una necesidad de estudiar sistemas en donde interactúan múltiples variables. Dadas esas variables, interesa entender de qué manera variaciones en sus valores afectan los resultados de interés. Para esto, lo que se hace es simular escenarios dinámicos en donde los individuos, aquí llamados agentes, interactúan entre ellos y con el ambiente para así estudiar los resultados en las variables. En últimas, lo que se busca es tratar de reproducir un fenómeno del mundo real a través de un programa computacional.

Con esto en mente, para la nueva propuesta metodológica se plantea el siguiente esquema:

1. **Entendimiento del problema:** Este es el primer paso y uno de los más importantes. El entendimiento del problema y de las relaciones que podrían establecerse entre las variables es trascendental para los pasos que siguen. En este punto es muy importante definir qué variables podrían estar influenciado los resultados de interés. Para esto, deben considerarse dos aspectos: por un lado, que las variables expliquen suficientemente cómo trabaja el sistema y por otro lado, que no sean más de las necesarias, para así evitar que las relaciones pierdan sentido y la explicación del sistema sea compleja y poco entendible. Esta característica es conocido como parsimonia, y ese es un objetivo de vital importancia en el planteamiento de un MBA.
2. **Búsqueda de datos empíricos y planteamiento de las relaciones existentes entre las variables:** Una vez seleccionadas las variables que responden al entendimiento del problema, es necesario soportar las posibles relaciones que se establecen entre ellas y cuantificar algunos parámetros de interés. Para dar respuesta a esto, debe hacerse uso de estudios previos y/o datos empíricos. Este paso podría provocar decisiones acerca de agregar o eliminar variables.

3. **Planteamiento del modelo:** Teniendo claras las variables y las posibles relaciones entre ellas, se plantea el modelo que soportará la simulación del sistema. En este deben incluirse los datos empíricos y/o las relaciones encontradas en estudios previos. El modelo corresponde a las relaciones determinísticas que pueden establecerse entre las variables del sistema que se quiere estudiar.
4. **Simulación de los agentes y su ambiente:** Esta es la parte que corresponde realmente a la metodología propuesta. Los pasos anteriores son comunes a otros métodos pero la simulación de los agentes y su ambiente es lo que define la propuesta metodológica. Como la arquitectura de agentes incluirá herramientas del análisis de redes, debe considerarse dentro del desarrollo de esta etapa el modelamiento de la formación, el crecimiento, la difusión y/o el aprendizaje, y la incorporación de estos procesos a la arquitectura de redes y al planteamiento del MBA. Esta etapa de la metodología se desarrollará con mayor detalle en el capítulo 4.
5. **Validación y calibración:** Como la metodología aquí presentada se basa en microsimulaciones para el estudio de sistemas, resulta de mucha importancia validar y calibrar los resultados. Según lo presentado en el capítulo 2, son cuatro los pasos básicos que definen la validación estructural: Validación de entrada, análisis de sensibilidad, calibración y validación de salida. Los cuatro pasos de la validación no constituyen en sí una última etapa de la metodología. Para ser más exactos, cada uno de los pasos acompaña las 4 etapas propuestas anteriormente, esto es: la validación de entrada se relaciona con el entendimiento del problema y el primer planteamiento del modelo; el análisis de sensibilidad está conectado con la búsqueda de datos empíricos y el planteamiento de las relaciones existentes entre las variables conforme a las evidencias que se tienen del sistema; la calibración se encuentra vinculada con el planteamiento del problema y la validación de salida se asocia con la simulación de los agentes y su ambiente. Estos procedimientos dan un margen importante de trabajo para la presente tesis, en la medida que, tal y como se presentó en el capítulo 2, existen avances en diferentes aspectos relacionados con la validación y la calibración que tienen una relación directa con la estadística. Recientes trabajos ([79]) destacan la importancia que tiene la validación dentro de los MBA, a pesar de destacar su uso relativamente bajo. Por su parte, en [45] se asegura que la validación estadística puede incrementar la credibilidad de los MBA y puede llevarse a cabo en conjunto con validación predictiva o validación cruzada con otros modelos.

Para entrar en los temas relacionados con validación y calibración, se introduce el concepto de meta-modelo para hacer referencia al modelo estadístico subyacente del modelo computacional complejo ([29]). A continuación se presenta con mayor detalle cada una de los pasos de la validación y la calibración y la propuesta para llevarlos a cabo dentro de la metodología propuesta:

- **Validación de entrada:** Esta es la etapa más sencilla en el proceso de validación. En el capítulo 2 se presentaron las tres etapas que definen este paso, esto es, una revisión general del comportamiento del sistema y de los agentes; monitorear la dinámica de un agente particular durante la ejecución del modelo y establecer que los valores de salida estén en un rango de valores reales y que las tendencias sean consistentes entre las diferentes simulaciones. Adicionalmente, por tratarse de una combinación entre los MBA y el análisis de redes, se debe incluir dentro de este paso una subetapa de la revisión general del sistema y es la revisión de la arquitectura de los agentes. En particular, es necesario que el modelo que define la formación de la red esté operando

de manera adecuada. Por su parte, para evaluar la consistencia entre las tendencias de las simulaciones, en [64] se establece que esta etapa se puede efectuar a través de comparaciones de medias y de varianzas.

Para esta metodología se propone utilizar algunas medidas de las redes como la distribución de grados, la centralidad, el agrupamiento y la distancia para hacer comparaciones. Adicionalmente, dado que se abordarán algunos de los desarrollos vía simulaciones de Monte Carlo de Cadenas de Markov, se incluirán comparaciones basadas en las densidades de los datos muestreados, los autocorrelogramas y el comportamiento de las cadenas de Markov.

- **Análisis de sensibilidad:** Esta etapa está relacionada principalmente con la formulación del modelo y la existencia de efectos de las variables que se incluyen en la simulación. Con el análisis de sensibilidad el objetivo principal es buscar un modelo parsimonioso y en el que las variables de entrada expliquen de manera acertada el sistema que se busca estudiar. Para realizar el análisis de sensibilidad, en [64, 79, 55] se propone el uso de herramientas de diseño de experimentos, aunque otros autores sugieren explorar otras alternativas como el aprendizaje de máquinas². Otras técnicas de análisis de sensibilidad pueden derivarse desde el enfoque de los meta-modelos, sin embargo estos métodos serán explicados más adelante

Según Montgomery³, se entiende por diseño de experimentos al enfoque para planear y llevar a cabo un experimento, el cual se entiende como “una prueba o serie de pruebas en las que se hacen cambios deliberados en las variables de entrada de un proceso o sistema para observar e identificar las razones de los cambios que pudieran observarse en la respuesta de salida”. En particular, las variables de entrada se denominan factores. Cada factor puede ser una variable cualitativa o cuantitativa y puede tener dos o más niveles. Se conoce como tratamiento a los niveles de los factores o a combinación de ellos. En el diseño de experimentos se estudian diferentes configuraciones de los factores y se observan una o más variables de respuesta y se estudia, con ayuda de un modelo, si los factores son significativos para explicar el comportamiento de la variable de respuesta. En un MBA los factores son las variables de entrada (*input variables*) y las variables respuestas son las salidas de la simulación (*output variables*). Para llevar a cabo el análisis de sensibilidad, en [29] se propone el uso de tres métodos alternativos y complementarios: (a) diagramas de dispersión y correlaciones; (b) coeficientes de regresión estandarizados y (c) análisis de varianza. En cuanto a esta última técnica, dependiendo de las características de los factores involucrados en la simulación se debe desarrollar un análisis de acuerdo al tipo de diseño experimental con el cual se encuentre relacionado. Debido a la alta complejidad que puede involucrar la combinación de un número grande de factores, en [55] se presentan algunas alternativas para llevar a cabo muestreos.

El gráfico 3.1 ilustra los diferentes tipos de diseños de experimentos que pueden ser utilizados para llevar a cabo el análisis de sensibilidad.

La metodología usual para estudiar el efecto de los factores en el diseño se basa en la descomposición de la varianza y en la evaluación de la significancia de los parámetros asociados a los factores en la explicación de la variable respuesta. En [55] se reco-

²Miner, D. (2010). A framework for predicting and controlling system-level properties of Agent-Based Models. PhD Dissertation. University of Maryland. Baltimore County.

³Montgomery, D. (2004). *Diseño y análisis de experimentos (Versión autorizada en español de Design and Analysis of experiments)*. Editorial Limusa. México D.F.

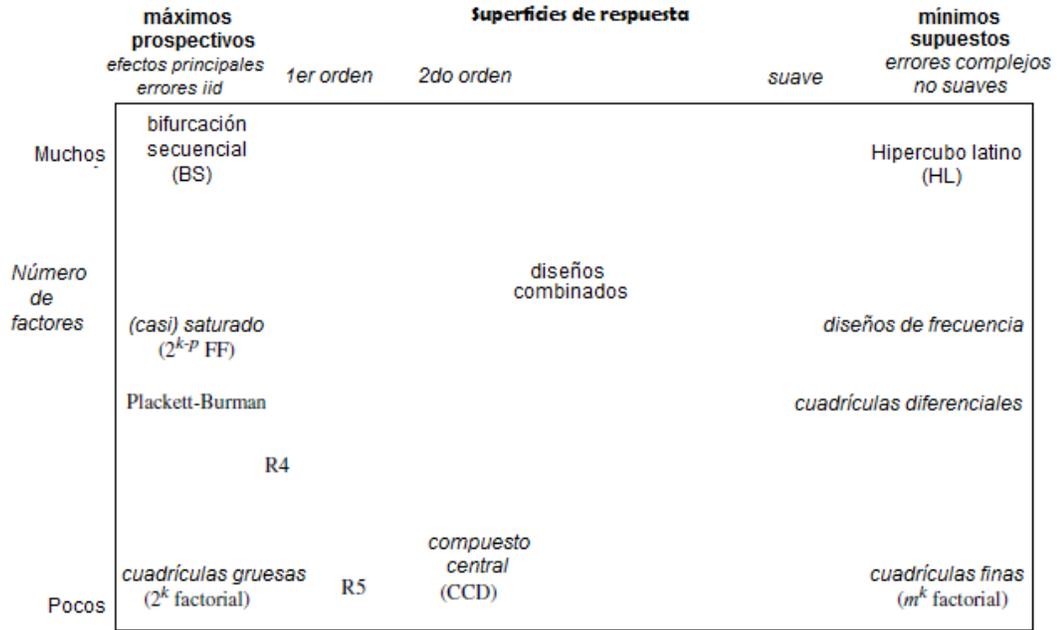


FIGURA 3.1. Caja de herramientas de diseño de experimentos. Tomada de [55].

mienda no hablar de significancia en el enfoque de la simulación pero se destaca la relevancia de estos análisis, previo a la definición de la simulación final. Para ello, se formula un modelo de regresión entre la variable respuesta y las variables incluidas en el diseño de la experimentación y se formulan los test estadísticos para evaluar si las variables de entrada tienen un efecto importante sobre las variables de salida. Por ejemplo, en el caso del diseño factorial general de 3 factores, esta especificación corresponde a:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl} \quad (3.1)$$

donde, $i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$; $k = 1, 2, \dots, c$ y $l = 1, 2, \dots, n$; siendo a el número de factores en el factor A, b el número de factores en B, c el número de individuos en C y n corresponde al número de individuos. La anterior especificación del modelo incluye todas las posibles interacciones entre los factores.

Una vez realizado el análisis de varianza, los estadísticos para probar la significancia de los parámetros quedan definidos a partir de la suma de cuadrados y los grados de libertad. Bajo la hipótesis nula, los estadísticos tienen una distribución F de Fisher bajo la cual se calculan los valores críticos y se aceptan o rechazan las hipótesis nulas, las cuales están definidas para cada factor como:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0 \text{ vs. } H_a : \text{al menos una } \tau_i \neq 0 \quad (3.2)$$

Dada esta breve explicación, llevar a cabo el análisis de sensibilidad en un MBA puede corresponder a evaluar, a través de un planteamiento análogo a un diseño de experimentos, qué variables deberían entrar inicialmente a la simulación usando como punto de partida cuales de ellas ayudan a explicar el comportamiento de la variable respuesta. Sin embargo, diseños de experimentos como los expuestos anteriormente, suelen asumir que la respuesta de un individuo al tratamiento asignado no es afectado por

otros individuos, lo que se conoce como el supuesto de no interferencia⁴ o SUTVA (*Stable Unit Treatment Assumption*)⁵.

Recientemente ha habido un interés creciente por abordar fenómenos en donde existe una interacción entre los individuos (también conocido como efecto de pares o interacciones sociales) en el desarrollo de diseños de experimentos. Entre 2013 y 2015 se han producido algunos avances que han evaluado el efecto de la interacción de redes en el diseño de experimentos encontrándose, por ejemplo, que el sesgo en las estimaciones es mayor en la medida en que las interacciones existentes entre los individuos tenga una mayor influencia sobre la respuesta al tratamiento. En esta medida, el efecto de la interacción entre los individuos puede llegar a ser desde ignorable hasta sustancial (para más información pueden consultarse los trabajos de [2], [62] y [23]).

Para el desarrollo de la presente propuesta metodológica los avances mencionados en términos de diseño de experimentos bajo interacción de los individuos son de vital importancia. Recordemos que en esta propuesta se introducen herramientas del análisis de redes en el desarrollo de los MBA y por lo tanto considerar la interacción dentro del análisis de sensibilidad ayuda a disminuir sesgos y errores. En [23] se proponen 4 pasos cuando se planea un diseño de experimentos bajo la presencia de interacción entre los individuos.

- Inicialización: Corresponde a lo que ocurre antes del experimento. Esto incluye la formación de la red y los procesos que definen el comportamiento a priori del experimento. Dentro de esta etapa es importante haber definido la matriz de adyacencia g . Adicionalmente, se identifica una matriz de características \mathbf{X} que pueden o no estar relacionadas con la estructura de la red.
- Asignación de tratamientos: En [23] se considera un tratamiento binario (tratamiento-control) denotado por Z_i , en donde el tratamiento de la unidad i es una variable aleatoria Bernoulli con $Z_i \sim \text{Bernoulli}(q)$. Se implementa un diseño denominado aleatorización por clusters del grafo modificado. La aleatorización por clusters del grafo ([82]) consiste en asignar tratamientos a clusters que son una partición de la red. Para ello los nodos de la red conforman una partición con N_C clusters C_1, C_2, \dots, C_{N_C} y se define una función $C(\cdot) : \{1, \dots, N\} \rightarrow \{1, \dots, N_C\}$ en donde $C(i)$ se refiere al cluster que contiene al nodo i . La conformación de estos clúster no tienen en cuenta las relaciones existentes entre los nodos. Los tratamientos son asignados a nivel de cada cluster C_j , esto es, C_j es asignado a un tratamiento $W_j \sim \text{Bernoulli}(q)$, de forma que el tratamiento asignado a los nodos corresponde al asignado a su clúster, esto es, $Z_i = W_{C_i}$. Esta alternativa tiene la dificultad de violar el requerimiento de que todas las unidades tengan una probabilidad positiva de ser asignada a todas las condiciones del tratamiento. La modificación propuesta por [23] consiste en introducir aleatorización a nivel de los nodos de forma que los clusters son asignados predominantemente a uno de los tratamientos pero algunos de los nodos reciben el otro tratamiento. Para ello

⁴Cox, D. (1958). *Planing of Experiments*. Wiley.

⁵Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. Vol. 66. pp. 668-701.

se establece que:

$$\begin{aligned} W_i &\sim \text{Bernoulli}(q_{C(i)}) \\ Q_i &\sim \text{Bernoulli}(\eta) \\ Z_i &= Q_i W_{C(i)} + (1 - Q_i)(1 - W_{C(i)}) \end{aligned}$$

Las Q_i son variables independientes conmutativas que hacen que Z_i sea asignada a $W_{C(i)}$ con una probabilidad η .

- Generación de resultados y observación: Se considera que la respuesta de un nodo está dada por una función de la asignación del tratamiento $Z \in \mathbb{Z}^N$ y un componente estocástico independiente $U \in \mathbb{U}^N$ para un vector de respuesta $Y \in \mathbb{Y}^N$, esto es:

$$f(\cdot) : \mathbb{Z}^N \times \mathbb{U}^N \rightarrow \mathbb{Y}^N \quad (3.3)$$

Dentro del desarrollo presentado en [62], el supuesto SUTVA se denomina ITR por sus siglas en inglés *Individualistic Treatment Response*, caso en el cual se asume que no hay interacción y por lo tanto la respuesta del nodo i estaría dada por una función de su propio tratamiento asignado:

$$f_i(\cdot) : \mathbb{Z} \times \mathbb{U}^N \rightarrow \mathbb{Y} \quad (3.4)$$

Una generalización del ITR propuesta por [62] se denomina CTR (*constant treatment response*) y permite incluir efectos asociados a las asignaciones globales de los tratamientos. Por su parte, en [2] se desarrolla de manera preliminar la teoría necesaria para realizar análisis bajo interacción de los individuos y se introduce el supuesto NTR (*neighborhood treatment response*) asumiendo que la respuesta de un nodo depende de su asignación al tratamiento y de la asignación de su vecindad.

Por otro lado, es importante definir el modelo de exposición que permite definir la función $f_i(\cdot)$ cuando hay interacción entre los individuos. En [23] se considera un modelo dinámico con pasos en tiempo discreto en el cual la respuesta del nodo i en el tiempo t , denotado por $Y_{i,t}$ está dado por una función h que depende de la asignación propia del tratamiento, la asignación y los comportamientos previos de la vecindad:

$$h_{i,t}(\cdot) : \mathbb{Z} \times \mathbb{Y}^{k_i+1} \times \mathbb{U}^N \rightarrow \mathbb{Y} \quad (3.5)$$

La anterior es la ecuación estructural del modelo. En este caso k_i representa el grado del nodo i . La función h y la red (N, g) determinan la función $f_i(\cdot)$

- Análisis y estimación: En [23], como se considera que existe un esquema de tratamiento-control, el análisis y la estimación se basan en el efecto promedio del tratamiento τ . Los autores presentan la derivación del estimador de τ bajo el supuesto NTR.

El diseño de experimentos en redes es un tema que hasta ahora produce sus primeros avances y en el momento no se encuentra desarrollada la teoría alrededor de diseños que consideren esquemas de tratamiento más allá del esquema de tratamiento-control. Avanzar en este propósito se escapa de los objetivos de la

presente tesis pero puede ser un tema de investigación que se aborde en próximos trabajos.

- **Calibración:** El algoritmo genético es una de las alternativas utilizadas para llevar a cabo el proceso de calibración dentro de los MBA ([28], [74]). Como se presenta en [31], el algoritmo genético es un tipo de algoritmo de evolución que se basa en fenómenos de la naturaleza como la selección, la combinación y la variación o mutación para resolver problemas de optimización complejos. La idea básica detrás de esta técnica es que cada individuo tiene asociado un conjunto de propiedades representadas a través de genes. Usualmente se comienza con una población de individuos, también llamada generación, y se desarrolla un algoritmo de evolución basado en los fenómenos naturales definidos y unas restricciones. En cada iteración del algoritmo se seleccionan unos individuos que pasan a la siguiente generación, así como algunos de ellos se combinan para heredar parte de sus características, mutan o varían y se calcula una función de aptitud que corresponde a la evaluación del problema de calibración. El algoritmo llega a su final después de haber logrado un número máximo de generaciones o haber alcanzado un nivel de satisfacción en la función de aptitud. A continuación se describen los diferentes operadores del algoritmo genético, según se presenta en [1]:

- I. **Selección:** En la naturaleza se da una selección natural de individuos que está mediada por el chance de sobrevivencia y la adaptación del individuo al ambiente. En el Algoritmo Genético la selección se basa en la evaluación de la función de aptitud. Dependiendo de si se trata de un problema de maximización o minimización, los individuos con valores más grandes o más pequeños en la función de aptitud, respectivamente, son los que tienen mayor probabilidad de sobrevivir.
- II. **Combinación:** En este proceso de reproducción, dos o más individuos combinan sus características (genes) para formar un nuevo individuo. Existen diferentes métodos para definir la herencia de los descendientes. El uso de cada método depende de las características del ejercicio que se esté desarrollando.
- III. **Mutación:** En la naturaleza se habla de mutación cuando el ADN cambia debido a errores durante el proceso de combinación de los genes de los padres o exposición a factores del ambiente. En el caso de los Algoritmos Genéticos este proceso se refiere a cambios aleatorios introducidos en el proceso.

El algoritmo genético tiene un importante número de ventajas ([77, 52]) sobre otros métodos de calibración. Entre las ventajas se destacan: (a) la capacidad de resolver diversos problemas de calibración basado en el enfoque de genes; (b) la capacidad de manejar problemas con múltiples soluciones; (c) la posibilidad de resolver problemas multidimensionales, no diferenciales, no continuos e incluso no paramétricos. Además se destaca su menor complejidad y flexibilidad en diferentes entornos de programación. Su principal desventaja es que este algoritmo no garantiza óptimos globales.

Según se presenta en [76] el algoritmo genético en el contexto de la calibración de los MBA parte de unos individuos que en este caso corresponden a una configuración específica de los parámetros del modelo. La generación 0 está compuesta por individuos escogidos aleatoriamente dentro de las posibles configuraciones de parámetros, esto debido a que el número posible de configuraciones aumenta de manera exponencial en la medida que aumenta el número de parámetros y puede volverse un problema intratable. En los siguientes pasos se evalúa la función de aptitud. En el contexto de los MBA resulta bastante difícil definir una función de aptitud dado que no es claro que función

de error se desea minimizar. Como una alternativa de solución a este problema, en [76] se proponen diferentes tipos de medidas como las medidas intra-agentes, medidas de grupos de agentes, medidas basadas en redes, medidas temporales, medidas de parámetros cruzados y medidas basadas en patrones.

Siguiendo a ([76]), las medidas basadas en redes son usadas en escenarios donde el MBA busca reproducir el comportamiento de una red. Para ello, se hace uso de las medidas definidas en el capítulo 1. Por su lado, en las medidas basadas en patrones lo que se hace es definir un patrón de referencia y definir una función de pérdida o de error en relación a ese patrón. Habiendo definido esas funciones de aptitud, lo que sigue en el algoritmo genético es calcular esa función para cada configuración de los parámetros y seguir el procedimiento canónico del algoritmo genético hasta haber logrado un número máximo de generaciones o haber alcanzado un nivel de satisfacción en las funciones de aptitud. Según se presenta en [66], típicamente cuando se tienen múltiples objetivos en un proceso de optimización lo que se hace es definir una función ad-hoc que vuelve unidimensional el problema en la función de aptitud.

Una alternativa que se presenta en [66] es la Optimización de Pareto, la cual es un método heurístico que retorna un conjunto de posibles soluciones al problema de calibración y que según se plantea en [48] constituye una modificación del algoritmo genético para manejar múltiples objetivos. En esta medida, la optimización de Pareto parece ser la alternativa apropiada para buscar calibrar el MBA basado en medidas de redes y de patrones.

En [48] se afirma que lo que se busca con las alternativas de optimización multi-objetivo es proveer soluciones no-dominadas, lo que significa que ninguna de las funciones objetivo por separado puede ser mejorada sin que esto implique que se degrade alguna de las otras funciones objetivos. En el caso de la optimización de Pareto las soluciones no-dominadas se ubican en una superficie conocida como la frontera óptima de Pareto, en la cual el algoritmo busca obtener una muestra de soluciones. Para ello lo que se plantea es buscar soluciones locales que sean no-dominadas ([48]).

En [79] se propone usar una alternativa denominada calibración categórica. Para esta estrategia de calibración se define un rango de valores aceptables para los parámetros que se desean calibrar, llamados criterios. Luego se define una *función de costo*, que depende de las diferencias estandarizadas de los valores simulados y los valores medios de los rangos aceptables para los criterios de la calibración. Por otro lado, cuando se tienen réplicas de datos observacionales y un modelo estocástico, en [79] se propone el uso de la medida *POMDEV - Pattern Oriented Modelling Deviance*, la cual se basa en la divergencia de Kullback-Leibler y teoría de verosimilitud. La divergencia de Kullback-Leibler es uno de los principios básicos del criterio de información de Akaike (AIC), ya que el objetivo es minimizar la divergencia I , la cual es evaluada entre la distribución teórica de x , descrita por la función de densidad $f(x)$, y la distribución empírica definida por el modelo, descrita por la función $g(x)$. De acuerdo con [68], $f(x)$ puede ser vista como influenciada por el conjunto completo de parámetros θ^* y expresada como $f(x|\theta^*)$, mientras que $g(x)$ está definida de acuerdo a un conjunto de parámetros considerados en el modelo, expresada como $g(x|\theta)$. La divergencia Kullback-Leibler se define como:

$$\begin{aligned}
I(f(x|\theta^*); g(x|\theta)) &= \int_{-\infty}^{\infty} f(x|\theta^*) \log \frac{f(x|\theta^*)}{g(x|\theta)} d(x) \\
&= \int_{-\infty}^{\infty} f(x|\theta^*) \log f(x|\theta^*) d(x) - \int_{-\infty}^{\infty} f(x|\theta^*) \log g(x|\theta) d(x) \\
&= E_f[\log f(x|\theta^*)] - E_f[\log g(x|\theta)]
\end{aligned} \tag{3.6}$$

Esta medida es un insumo para el denominado Criterio de Información para el Modelamiento Orientado a Patrones (*POMIC - Pattern Oriented Modelling Information Criterion*, por sus siglas en inglés). Según se presenta en [68], el *POMIC* se deriva partiendo de la misma lógica detrás del Criterio de Información de Akaike (*AIC*), sin embargo se tiene la dificultad de que en los MBA se carece de una función de verosimilitud. En el caso de los MBA se tiene que el proceso de simulación influencia el comportamiento de la variable x . La distribución de la variable respuesta es expresada a través de una función de densidad $g(x|\theta)$. Por su parte, la distribución del proceso de simulación, denominada distribución muestral, puede ser descrita como una función $b(x|\theta^f)$, en donde θ^f son los patrones que influyen la variable respuesta y los correspondientes valores de los parámetros. En el enfoque de los autores, $b(x|\theta^f)$ es tomada como un indicador de la función de densidad hipotética $f(x|\theta^*)$ en donde θ^* corresponde a todos los patrones que influyen la distribución teórica. Se plantea una estimación kernel de la densidad y, dada esta estimación, se calcula un indicador de deviance denominado *POMDEV (Pattern-Oriented Modeling Deviance)*, dado por:

$$POMDEV = -2 \sum_{i=1}^n \log(g(x_i|\theta)) \tag{3.7}$$

Siguiendo la propuesta presentada en [68], el *POMIC* es derivado después de combinar el muestreador Metropolis con el *POMDEV*. Esta propuesta parte de dos procedimientos: (a) la media del *POMDEV* del modelo es computado sobre la distribución posterior de los parámetros; (b) Se usa las distribuciones posteriores para escoger las estimaciones de los parámetros para la deviance del modelo.

- **Validación de salida:** Para hacer validación de salida puede partirse de dos escenarios: por un lado, del planteamiento estocástico de los MBA y por otro, de la aproximación a través de Emuladores Gaussianos y Computación Bayesiana Aproximada para acercarse al desarrollo de inferencia estadística a partir de los MBA. Estos temas son abordados en la siguiente sección.

3.2.2. Fenómenos de estudio

- (a) La formación: El fenómeno de formación de redes será transversal a las diferentes aplicaciones en las cuales pueda ser utilizada la metodología propuesta, ya que esta es una característica que se incorpora para mejorar los MBA. La formación de la red es utilizada para dar respuesta a las críticas sobre la estructura atómica expuesta en [25] en relación a la arquitectura de agentes. Dependiendo las características del sistema estudiado, pueden utilizarse alguno de los dos modelos de formación presentados en el capítulo 1 o pueden estudiarse otras alternativas dependiendo la aplicación que se esté abordando. Sin embargo, cuando se trabaje con relaciones entre personas, el modelo de Watts-Strogatz representa una mejor aproximación a la formación de las redes.

- (b) El crecimiento: El estudio del crecimiento busca modelar la forma en como se relacionan los agentes con otros nuevos agentes que aparecen en el tiempo. Como es de esperarse, la creación de nuevas relaciones, generalmente, no se produce de forma aleatoria sino que está mediada por otros factores. Los modelos que mejor representan esta situación son conocidos como modelos de Conexión Preferencial (como se presentó en el capítulo 1) y serán estos la base para la nueva metodología en la medida que son más cercanos a las condiciones en las que normalmente ocurren estos fenómenos. Una alternativa que puede ser abordada en próximos trabajos es la propuesta de modelos de Conexión Preferencial basados, por ejemplo, en los valores que toman los agentes en las variables de entrada. Es importante aclarar que no es necesario que hayan nuevos individuos en la población de estudio para usar modelos de crecimiento, ya que estos se relacionan es con la aparición de nuevos enlaces.
- (c) La difusión: El fenómeno de difusión es de vital importancia en estudios que involucren temas epidemiológicos, *marketing viral*, divulgación de información, entre otros. Los dos modelos más utilizados y explicativos, según se presentó en el capítulo 1, son los modelos de difusión tipo SIR y SIS. El uso de alguno de ellos depende de qué ocurre después de que la difusión alcanza algún individuo: si éste queda susceptible nuevamente o no a la difusión determinará que tipo de modelo usar.
- (d) El aprendizaje: Con el fin de incluir el fenómeno de aprendizaje dentro de la metodología propuesta, se decide adoptar el modelo de aprendizaje bayesiano. Para la metodología aquí propuesta, extender el fenómeno de aprendizaje para los agentes resulta bastante natural puesto que los agentes son tratados como nodos y entre ellos se establecen relaciones. Este modelo de aprendizaje permite incorporar la experiencia de la vecindad en torno a la toma de decisiones, lo cual se aproxima de una mejor manera a muchos fenómenos en donde los individuos tienen una naturaleza social. Como se mencionó en el capítulo 1, bajo este enfoque se entiende que un individuo adopta alguna decisión o comportamiento cuando es capaz de observar las decisiones de los otros antes de tomar su decisión. Además cada una de las decisiones determina algún tipo de ganancia con ciertas probabilidades. Lo que cada individuo busca es maximizar su ganancia.

3.3. Inferencia estadística a partir de los MBA

A pesar de la escasa literatura en el tema, se han encontrado resultados muy valiosos en la medida que abren una gran ventana que permite desarrollar inferencia estadística a partir de simulación computacional basada en agentes. Los trabajos más importantes en este tema han sido desarrollados por los estadísticos David Banks y Daniel Heard de la Universidad de Duke. Este último durante su tesis doctoral de 2014 [45] presenta el uso de dos herramientas estadísticas para la inferencia estadística a partir de MBA: los Emuladores Gaussianos y la Computación Bayesiana Aproximada (*Approximate Bayesian Computation (ABC)*). A continuación se presenta cada una de estas técnicas con mayor detalle.

3.3.1. Emuladores Gaussianos

Según se presenta en [45, 10], un emulador es un proceso estocástico que sirve como representación de un modelo computacional, también llamado simulador. Esta metodología se basa en los hallazgos de O'Hagan en 1978 ([42]), quién describe cómo puede representarse una función des-

conocida a través de un proceso Gaussiano. En este caso el simulador es visto como una función matemática desconocida.

De acuerdo a la definición presentada en [14], un proceso estocástico es una familia de variables aleatorias $(X_t)_{t \in T}$ indexadas en un conjunto llamado T , definidas todas ellas sobre un mismo espacio de probabilidad (Ω, \mathcal{J}, P) y con valores en un espacio medible (S, \mathcal{R}) . S se llama conjunto de estados. Por su parte, un proceso Gaussiano es un proceso estocástico tal que para cualquier subconjunto $F \subset T$, el vector aleatorio $(X_t)_{t \in F}$ tiene distribución Gaussiana. Esto es equivalente a que cualquier combinación lineal de variables aleatorias del proceso tiene distribución Gaussiana.

Siguiendo el planteamiento de [45], el objetivo del uso de Emuladores Gaussianos en el contexto de los MBA es modelar el comportamiento del sistema y las observaciones basado en unos ajustes de calibración \mathbf{t} que no son conocidos para las n observaciones. Los ajustes de calibración sirven como *input* en el modelo y afectan el *output* y pueden incluir, entre otras cosas, las reglas que determinan el comportamiento de los agentes. En este caso θ representa los valores óptimos de los ajustes.

Con base en la descomposición aditiva propuesta por [53], sea $\eta(\mathbf{x}, \mathbf{t})$ representa el *output* del MBA bajo las condiciones \mathbf{x} y los valores de calibración \mathbf{t} , los datos observados $\mathbf{y} = (y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))^T$ pueden ser modelados estadísticamente como:

$$y(\mathbf{x}_i) = \eta(\mathbf{x}_i, \theta) + \delta(\mathbf{x}_i) + \epsilon(\mathbf{x}_i) \quad (3.8)$$

donde $\delta(\mathbf{x}_i)$ es un término estocástico para cuantificar las discrepancias sistemáticas entre el MBA $\eta(\mathbf{x}_i, \theta)$ y el proceso físico $y(\mathbf{x}_i)$. Dada la complejidad natural de los MBA, la función que representa el *output*, η , es desconocida. Siguiendo lo propuesto en [42], se utiliza un modelo de proceso Gaussiano para $\eta(\mathbf{x}, \mathbf{t})$. En particular, de acuerdo a planteamiento anteriores de Hidgon y Sacks, el autor propone un proceso Gaussiano con función de media $\mu(\mathbf{x}, \mathbf{t})$ y una covarianza con función *exponential power*.

En [45] se desarrolla esta metodología a partir de un proceso Gaussiano arbolado, en el cual el espacio es particionado en diferentes conjuntos en los cuales los procesos Gaussianos estacionarios son ajustados dentro de cada conjunto.

Un ejemplo que utiliza emuladores gaussianos se presenta en [12]. En esta aplicación se extiende el MBA presentado en los antecedentes sobre la presión social y el matrimonio avanzando en la exploración de sus propiedades estadísticas. En este caso se usan procesos gaussianos para construir los emuladores.

3.3.2. Computación Bayesiana Aproximada

Este es otro enfoque que permite desarrollar inferencia Bayesiana a partir de MBA. El objetivo de la Computación Bayesiana Aproximada es hacer inferencia acerca de un parámetro (o conjunto de parámetros) θ a partir de los datos observados \mathbf{x}_0 cuando la función de verosimilitud es intratable. Este método usa la comparación de los datos simulados y los datos observados como criterio para hacer inferencia. De acuerdo con [45], el algoritmo más básico de la Computación Bayesiana Aproximada se basa en el muestreo por rechazo y se compone de los siguientes 4 pasos:

1. Muestrear un valor θ' de la distribución a priori, $\pi(\theta)$
2. Generar un conjunto de datos x' a partir de $f(\cdot|\theta')$
3. Medir la distancia entre los datos generados y los datos observados, $\rho(x', x_0)$. Si $\rho(x', x_0) \leq \epsilon$, acepte θ' , de lo contrario rechace este valor.
4. Retornar al paso 1.

Un inconveniente de este algoritmo es que si la distribución a priori es muy diferente de la distribución a posteriori, entonces la tasa de aceptación será baja. La diferencia con el muestreo por rechazo usual es que en el método de Computación Bayesiana Aproximada suele utilizarse una estadística \mathbf{S} , en cambio de los datos completos. Según lo señala el autor, idealmente debería utilizarse una estadística suficiente dado que simplifica los cálculos computacionales y además, por propiedades, se tiene que $\pi(\theta|x) = \pi(\theta|\mathbf{S})$. Sin embargo, dado que la función de verosimilitud no está disponible, no es fácil definir esta estadística.

Una alternativa al muestreo por rechazo es usar métodos de Monte Carlo con base en Cadenas de Markov (MCMC). Con este algoritmo se produce una Cadena de Markov de la distribución estacionaria $\pi(\theta|\rho((x', x_0))) \leq \epsilon$ que converge a la distribución a posteriori aproximada. Según [45], este método tiene la desventaja de que las correlaciones naturales entre las cadenas de Markov y las bajas tasas de rechazo pueden hacer que se produzcan cadenas largas que queden atascadas en regiones de baja probabilidad para varias iteraciones. El siguiente enfoque busca solucionar los problemas mencionados anteriormente.

Computación Bayesiana Aproximada usando Monte Carlo Secuencial

Esta técnica usa los valores de los parámetros $\{\theta^{(1)}, \dots, \theta^{(N)}\}$ muestreados de la distribución a priori. Este algoritmo opera de la siguiente forma:

1. Inicializar los valores de tolerancia $\epsilon_1, \dots, \epsilon_T$
Establecer el indicador poblacional $t = 0$
2. (a) Establecer el indicador $i = 1$
(b) Si $t = 0$, muestrear θ'' independientemente de la distribución a priori $\pi(\theta)$
De otra forma, muestrear θ' de la población previa $\{\theta_{t-1}^{(i)}\}$ con pesos w_{t-1} y se perturba para obtener $\theta'' \sim K_t(\theta|\theta')$, donde K_t es el kernel de la perturbación.
Si $\pi(\theta'') = 0$, vuelva al paso 2(b) Simule un conjunto de datos candidatos $x' \sim f(x|\theta'')$
Si $\rho(x', x_0) \geq \epsilon_t$, volver al paso 2(b)
(c) Definir $\theta_t^{(i)} = \theta''$ y calcular los pesos para $\theta_t^{(i)}$ dados por:

$$w_t^{(i)} = \begin{cases} 1, & \text{si } t = 0 \\ \frac{\pi(\theta_t^{(i)})}{\sum_{j=1}^N w_{t-1}^j K_t(\theta_t^{(j)}, \theta_t^{(i)})} & \text{si } t > 0 \end{cases}$$

Si $i < N$, sea $i = i + 1$ y vuelva al paso 2(b)

3. Normalice los pesos
Si $t < T$, sea $t = t + 1$ y retome el paso 2(a)

El kernel de perturbación suele escogerse como una caminata aleatoria. Nótese que cuando $T = 1$ el anterior algoritmo corresponde al de muestreo por rechazo.

Los desarrollos actuales no permiten incluir la interacción entre individuos dentro de los procesos de inferencia estadística basados en MBA. Sin embargo, estas técnicas han sido usadas en aplicaciones en donde la interacción entre los agentes hace parte de la naturaleza del fenómeno de estudio (ver [75] y [45]). Esta metodología busca estudiar la distribución del parámetro o conjunto de parámetros de interés a partir de los datos simulados. Si los procesos de validación de entrada, análisis de sensibilidad y validación de salida fueron realizados correctamente, puede realizarse esta etapa sin tener en cuenta la interacción entre los individuos durante el proceso de validación.

3.3.3. Implementación de la metodología

Para implementar la metodología propuesta se utilizan el ambiente de modelamiento programable para simular fenómenos sociales y naturales *NetLogo*, especializado en MBA ([84]), y el software estadístico R ([71]), en particular el paquete RNetLogo ([50]), el cual permite realizar análisis de sensibilidad y calibración.

APLICACIÓN

4.1. Planteamiento y entendimiento del problema

Con el fin de llevar la metodología a la práctica, este capítulo presenta el planteamiento de una aplicación relacionada con el uso de métodos anticonceptivos modernos en Colombia y la reducción del embarazo inoportuno y no deseado.

Los métodos anticonceptivos reducen o impiden la posibilidad de embarazos; son parte fundamental de la planificación familiar. Según lo presenta la Organización Mundial de la Salud¹, la planificación familiar permite a las personas controlar el número de hijos y el intervalo entre embarazos, además, tiene importantes consecuencias en distintos aspectos, como lo son: lograr bienestar y autonomía de las mujeres; prevención de los riesgos para la salud relacionados con el embarazo en las mujeres; reducción de la mortalidad de los menores de un año; prevención de la infección por el VIH y el SIDA; poder de decisión y una mejor educación; y disminución del embarazo de adolescentes.

En un reporte de las Naciones Unidas de 2012 denominado “Sí a la opción, no al azar: planificación de la familia, derechos humanos y desarrollo”², señala que la planificación familiar es una importante inversión económica, que reduce la pobreza y la exclusión, mejora la salud y la vida de las mujeres. En el mismo reporte, se estima que en el mundo es necesario agregar 4,100 millones de dólares a la financiación para satisfacer las necesidades actuales en términos de métodos anticonceptivos, mientras que calcula que con esto se reducirían unos 11,300 millones de dólares en atención de salud materna y neonatal.

En esta misma dirección Profamilia, “entidad privada sin ánimo de lucro especializada en salud sexual y salud reproductiva que ofrece servicios médicos, educación y venta de productos a la población colombiana”³, señala que este tipo de entidades y los gobiernos, desde los años sesenta, han influenciado en América Latina los temas de planificación familiar, con base en cuatro razones

¹<http://www.who.int/mediacentre/factsheets/fs351/es/> Consultado por última vez el 17 de octubre de 2014

²<http://nacionesunidas.org.co/blog/2012/11/14/222-millones-de-mujeres-no-tienen-acceso-a-la-planificacion-familiar-en-el-mundo/> Consultado por última vez el 4 de octubre de 2015

³http://profamilia.com/index.php?option=com_content&view=article&id=124&Itemid=28 Consultado por última vez el 26 de octubre de 2014.

básicas ([70]): (a) Razones de orden demográfico y socioeconómico, (b) Razones relativas a los Derechos Humanos, (c) Razones de Salud y (d) Razones de Justicia y Equidad.

Los métodos anticonceptivos suelen clasificarse en dos grupos: por un lado, los métodos anticonceptivos tradicionales como el *coitus interruptus* y el método del ritmo; por otro lado, los métodos modernos, que corresponden a los demás métodos presentados en el gráfico 4.1.

La importancia de estudiar el uso de distintos métodos anticonceptivos radica en cambios drásticos en la efectividad que tienen para prevenir embarazos y enfermedades de transmisión sexual y los efectos que esto tiene sobre el empoderamiento de las mujeres, su acceso al mercado laboral y las decisiones al interior de las parejas. Según el Centro para el Control y Prevención de Enfermedades del Departamento de Salud y Servicios Humanos de los Estados Unidos, la efectividad entre los métodos más conocidos varía entre 72 % y 99.85 %⁴. Esta efectividad fue estudiada en términos de embarazos no deseados entre mujeres dentro del primer año de uso típico de métodos anticonceptivos. El gráfico 4.1 presenta los porcentajes reportados de falla en dichos métodos.

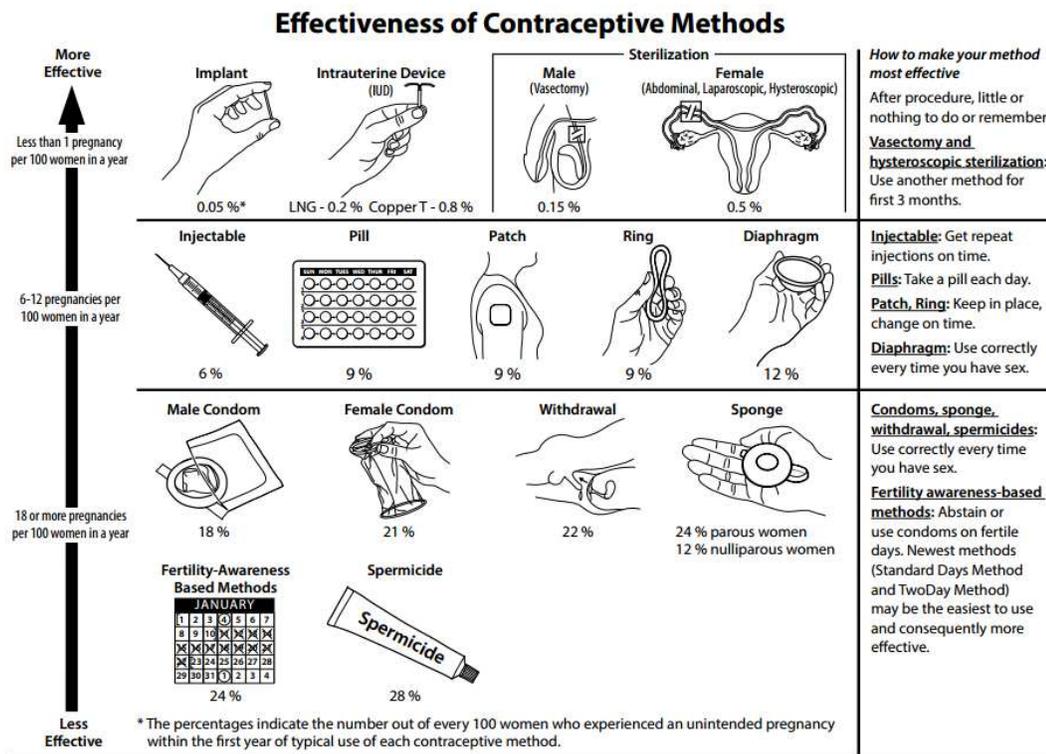


FIGURA 4.1. Efectividad de los métodos anticonceptivos. Tomada de CDC de los Estados Unidos.

Dentro de la demografía una herramienta bien conocida son las encuestas de demografía y salud (DHS por sus siglas en inglés *Demographic and Health Surveys*). Este tipo de encuestas recolectan información más específica que en los censos, principalmente relacionada con características demográficas de la población (etnia, género, edad, composición familiar, nupcialidad, entre otras) y factores relacionados con la salud (medidas antropométricas, salud sexual y reproductiva, fecundidad, mortalidad, estado nutricional, entre otros).

⁴http://www.cdc.gov/reproductivehealth/UnintendedPregnancy/PDF/Contraceptive_methods_508.pdf Consultado por última vez el 20 de octubre de 2014

4.2. Datos preliminares

La Encuesta Nacional de Demografía y Salud (ENDS) en Colombia es un proyecto que se ha replicado cada 5 años desde 1990 con la activa participación de Profamilia. Según se presenta en [70], se espera que la información de la ENDS “sirva de base para la evaluación de diferentes políticas, planes, proyectos y programas que se desarrollan en el país, especialmente los orientados a los grupos de población más necesitados”. La ENDS 2005 encuestó efectivamente a 37,211 hogares, mientras que la ENDS 2010 encuestó 51,447 hogares. Estas encuestas tienen cobertura nacional con representatividad urbana - rural, seis regiones (Caribe, Oriental, Central, Bogotá, Pacífica y Amazonía - Orinoquía), 16 subregiones y cada uno de los departamentos de manera independiente ([69, 70]). Ambas encuestas incluyen una sección con temas relacionados con la planificación familiar y el conocimiento y uso de métodos anticonceptivos (Ver [69, 70]). Adicionalmente recolectaron información sobre las preferencias relacionadas con fecundidad y demanda de planificación familiar, el conocimiento sobre enfermedades de transmisión sexual y la percepción sobre la educación sexual. Algunos de los resultados más relevantes, y que coinciden en la ENDS 2005 y 2010, son los siguientes:

- El conocimiento de métodos anticonceptivos modernos y tradicionales es universal en Colombia. El conocimiento se refiere a haber escuchado de los métodos pero esto no significa que los hayan utilizado ni que sepan de su uso correcto. Este comportamiento viene observándose desde la ENDS 1995.
- Entre las mujeres los métodos anticonceptivos más conocidos en su orden son: el condón, la píldora, la inyección y la esterilización femenina.
- Se observa que en todos los grupos poblacionales de mujeres el conocimiento de los métodos anticonceptivos modernos es mayor que el de los tradicionales.
- No se encontraron evidencias de diferencias en el conocimiento de métodos anticonceptivos ni por zona (urbana o rural), ni por niveles de educación, ni por estrato socioeconómico o índice de riqueza⁵.

En cuanto al uso de métodos anticonceptivos se destaca que:

- El uso de métodos anticonceptivos se está volviendo universal en Colombia, especialmente el uso de métodos modernos. En la ENDS 2005, 81 % de las mujeres manifestaron que han utilizado alguna vez en su vida un método anticonceptivo. Este porcentaje es cercano a 85 % de las mujeres en la ENDS 2010. Entre 2005 y 2010 se observan cambios importantes en el uso de métodos anticonceptivos modernos. Los tres métodos que presentan mayor crecimiento entre 2005 y 2010 son el condón (pasando de 45.6 % a 59.6 %), la inyección (pasando de 27.1 % a 37.6 %) y la anticoncepción de emergencia (pasando de 3.2 % a 10.8 %).
- Entre las mujeres en edades de 15 a 49 años de edad, el uso actual de métodos de planificación familiar se encontraba en 2005 en 56.4 % y para 2010 aumentó a 61 %. Para 2010, entre las mujeres de estas edades que se encontraban unidas⁶ este porcentaje corresponde a 79 %,

⁵El índice de riqueza medido en la ENDS corresponde a un enfoque que sintetiza el nivel socio-económico en términos de activos o riqueza de los hogares encuestados. Para ello se usa una técnica de Análisis de Componentes Principales, los niveles del índice corresponden a los quintiles.

⁶En este trabajo, de acuerdo a los términos utilizado en la ENDS 2010, se consideran mujeres unidas a las unidas o casadas sin ningún tipo de discriminación entre ellas.

mientras que este porcentaje alcanza 82 % para las mujeres no unidas sexualmente activas⁷. En este caso también se observan cambios importantes en el uso de métodos anticonceptivos al comparar los resultados de 2005 y 2010. Los dos métodos modernos que presentan mayor aumento son la esterilización femenina (pasando de 20.6 % a 24.2 %) y la inyección (pasando de 4.9 % a 7.9 %). En ambas encuestas se observaron diferencias considerables por grupo de edad y por condición entre mujeres unidas y mujeres no unidas sexualmente activas.

- Se observan grandes diferencias en los métodos anticonceptivos usados por región, subregión, nivel educativo e índice de riqueza. Por ejemplo, en 2010 la esterilización masculina alcanza un máximo de uso de 12.4 % en Medellín y su área metropolitana, siendo mucho más alto que en cualquier otra subregión. Por su parte, este método resultó ser poco usado en la región caribe (0.3 %). Adicionalmente, el uso de la esterilización masculina aumenta con el nivel educativo y el índice de riqueza.

Por último, es importante destacar que en 2010, 41.4 % de las mujeres que usaban algún tipo de método anticonceptivo dejaron de usarlo en el primer año, lo cual representa una leve reducción en comparación con el mismo dato para 2005 (42 %). Entre este porcentaje de mujeres, las fallas del método produjeron una tasa de discontinuación de 7.2 % en 2005 y 4.6 % en 2010, por su parte, el deseo de quedar embarazada fue la razón para 3.5 % de las mujeres en 2005 y 3.3 % en 2010, el cambio a un método más efectivo fue la razón para discontinuar para 18.3 % en 2005 y 16.1 % en 2010. Otras razones (como efectos secundarios, razones de salud y sexo poco frecuente) fueron señaladas por 13 % de las mujeres en 2005 y 17.3 % en 2010.

Otro dato interesante es el relacionado con el uso futuro de métodos anticonceptivos. Entre las mujeres que no usan métodos anticonceptivos, en 2005, 68.9 % manifiestan la intención de usar métodos en el futuro. Para 2010 este porcentaje se reduce a 66.3 %. Para 2010 se encuentra que, entre quienes no tienen intención de uso, menos de 0.1 % tienen razones de acceso o costo, 2.4 % tienen miedo a algún efecto secundario, mientras que 81.2 % tienen razones relacionadas con la fecundidad como menopausia o histerectomía (44 %), infertilidad de la mujer (15.2 %) o quiere más hijos (11.6 %).

Finalmente, la ENDS también provee información sobre planificación de la fecundidad. Se destaca que en 2005, 27 % de las mujeres reportan que los nacimientos experimentados se pueden clasificar como inoportunos, mientras que 27 % fueron reportados abiertamente como no deseados. En 2010 estos valores corresponden respectivamente a 29 % y 23 %.

Por su parte, según las Naciones Unidas⁸, en Colombia 7 % de las mujeres tienen demanda insatisfecha de anticoncepción, el cual sube a 20 % para las adolescentes. Este informe destaca que Colombia es un ejemplo en América Latina en cuanto al conocimiento y uso de métodos anticonceptivos. Se señala que tanto en el régimen subsidiado como en el contributivo, están incluidos la mayoría de los métodos anticonceptivos, sin embargo, 48 % de las mujeres no sabe que las instituciones de salud están obligadas a proporcionar los métodos anticonceptivos sin costo adicional a los costos de consulta.

⁷Para la ENDS 2010, se considera mujer sexualmente activa aquella que haya tenido relaciones sexuales durante los últimos 30 días antes de la encuesta.

⁸<http://nacionesunidas.org.co/blog/2012/11/14/222-millones-de-mujeres-no-tienen-acceso-a-la-planificacion-familiar-en-el-mundo/> Consultado por última vez el 4 de octubre de 2015.

Para la presentación de la aplicación se presentarán las generalidades relacionadas con la forma en cómo se aborda el fenómeno de estudio desde tres perspectivas diferentes. En primer lugar, se presenta el ejemplo abordado a través del Análisis de Redes; en segundo lugar, se aborda el estudio a través de un MBA; y por último, se resuelve el problema haciendo uso de la metodología propuesta en este documento.

4.3. Una mirada desde distintos enfoques

4.3.1. Análisis de redes

Estudiar el uso y la selección de métodos anticonceptivos y sus consecuencias sobre la ocurrencia de embarazos inoportunos y no deseados desde el análisis de redes, implica que se está entendiendo que este es un problema esencialmente social y que las decisiones tomadas por las personas guardan un grado de dependencia con las decisiones tomadas por las personas conectadas con ellas. Para llevar a cabo este estudio desde el análisis de redes, sería necesario recolectar información de la población de interés, de las conexiones que existen entre los individuos y de los comportamientos y acciones relacionados con la planificación familiar y el uso de anticonceptivos. Sin embargo esto no sería suficiente, pues para adelantar estudios de formación, crecimiento, difusión y aprendizaje, es necesario contar con datos longitudinales que permitan evaluar la incidencia de los individuos en el cambio de acciones y comportamientos.

Existen varios estudios que se han enfocado en el rol de las redes sociales en las decisiones sobre uso de métodos anticonceptivos: en [94] se realiza un estudio con un número reducido de mujeres de EEUU perteneciente a minorías. Los autores concluyen que la red social de las mujeres incluidas en el estudio tenía una influencia en las decisiones relacionados con métodos anticonceptivos. La familia, amigos y compañeros de la red social fueron claves en relación con los mitos, las ideas erróneas y las experiencias indirectas. En un estudio longitudinal realizado en Kenya, cuyos resultados se presentan en [11], se concluye que las redes sociales de las mujeres y los hombres tienen efectos significativos y substanciales sobre el uso de métodos anticonceptivos, incluso después de controlar por otros factores observados y no observados. Otro estudio realizado en Polonia ([18]) concluye que los comportamientos anticonceptivos de amigos y familiares son más influyentes que las propias características socio-demográficas de las mujeres. Además se encuentra que hay una influencia adicional de las características de la comunidad. Se encuentra que el nivel de educación de los vecinos aceleran el uso de métodos anticonceptivos mientras que la alta religiosidad de estos retrasa el uso de métodos anticonceptivos modernos.

4.3.2. MBA

Para estudiar el uso y las decisiones en torno a los métodos anticonceptivos y la ocurrencia de embarazos no deseados o inoportunos desde los MBA es necesario simular un conjunto de individuos y sus comportamientos a través de un sistema que busque reproducir características y patrones similares a los observados en la realidad. Los agentes en este MBA, corresponden a las mujeres en edad fértil entre 15 y 34 años de edad. Esta decisión tiene una razón básica y es que, tal como se ha presentado anteriormente, es sobre este grupo poblacional que se suelen realizar los estudios sobre fecundidad, fertilidad y uso de métodos anticonceptivos, lo cual aporta información previa para inicializar el modelo computacional. La decisión de reducir el grupo de edad, se debe a que las mujeres de 35 y más años suelen asociarse con problemas de infertilidad, menopausia

e histerectomía, razón por la cual no entran a jugar un papel determinante en el uso de métodos anticonceptivos y reducción de embarazos inoportunos y no deseados.

VARIABLES DE ENTRADA: (a) Grupo de edad. (b) Estado conyugal. (c) Uso de métodos anticonceptivos y tipo de método usado. (e) Experiencia de fecundidad. (f) Actividad sexual. (g) Fertilidad.

VARIABLES DE SALIDA: (a) Ocurrencia de embarazos. (b) Ocurrencia de embarazos inoportunos o no deseados.

La información para inicializar el MBA se basará principalmente en los resultados de la ENDS 2005 y la ENDS 2010.

4.3.3. MBA y Análisis de redes

El estudio sobre uso y selección de métodos anticonceptivos y su relación con la ocurrencia de embarazos no deseados e inoportunos, reúne una serie de características de mucho interés para poder implementar la metodología propuesta. En particular, las relaciones que se establecen entre las variables de entrada son complejas y esto también ocurre con las variables de salida. Además existen factores que no afectan a todas las mujeres y algunos de estos factores tienen un efecto diferencial por características observables. Por otro lado, la experiencia y la ocurrencia de eventos son determinantes para el cambio de comportamientos de los agentes, lo que hace indispensable que se considere una estructura dinámica para el estudio de este fenómeno.

- Entendimiento del problema: Durante este capítulo se ha adelantado el entendimiento del problema, orientado principalmente a identificar las variables del sistema. Sin embargo, en este punto se desea profundizar un aspecto importante: el tratamiento simultáneo de los factores interpersonales y los impersonales. Como se presentó anteriormente, en [63] se señala que las decisiones relacionadas con planificación personal dependen de factores interpersonales e impersonales. El análisis de redes es una aproximación para el tratamiento de los primeros, mientras que los MBA lo son para los segundos. La combinación de estas dos metodologías permiten tratar de manera simultánea los dos tipos de factores. Para ello, adicional a lo presentado en el segmento anterior, la arquitectura de los agentes incluye una estructura social entre ellos, lo que permite una mejor aproximación al fenómeno real.
- Búsqueda de datos empíricos y planteamiento de las relaciones existentes entre las variables. La información para inicializar el MBA se basa principalmente en los resultados de la ENDS 2005 y la ENDS 2010. Las variables que se presentaron en la sección anterior se mantienen para este análisis. Las relaciones entre las variables se derivan de análisis estadísticos que se presentan en el Capítulo 5.
- Planteamiento del modelo: Para dar respuesta a este aspecto se parte del planteamiento del MBA del segmento anterior. Sin embargo, a este modelo es necesario incluirle aspectos relacionados con las relaciones sociales entre los agentes. Para ello, sería útil enfocarse en las siguientes variables adicionales como:
 - El método más utilizado por la vecindad del agente.
 - La efectividad experimentada con los métodos anticonceptivos utilizados por la vecindad.
 - Factores que ayuden a determinar una influencia diferencial de las opiniones de los otros agentes sobre la decisión del agente.

-
- Simulación de los agentes y su ambiente: Se simulan un total de 100 agentes que en este caso corresponden a mujeres en edades de 15 a 34 años. Esta simulación inicialmente conserva las distribuciones en las características demográficas encontradas en la ENDS 2005 en cuanto a grupo de edad, experiencia de fecundidad, actividad sexual y estado conyugal.
 - Validación y ajustes: Para llevar a cabo el proceso de construcción del MBA, el proceso de calibración, análisis de sensibilidad, validación y ajustes, se utilizan los software NetLogo y R project. Para estos análisis se hace uso del materia suplementario presentado en [78].

RESULTADOS Y ANÁLISIS

5.1. Datos empíricos y planteamiento de relaciones existentes entre las variables

La aplicación propuesta estudia el uso de métodos anticonceptivos y la ocurrencia de embarazos inoportunos o no deseados en una población en la cual se conservan características socio-demográficas y de fecundidad de las mujeres colombianas entre 15 y 34 años. El esquema que se sigue para la aplicación es el siguiente: en primer lugar, se definen algunas características sociodemográficas con las cuales se van a simular los agentes; en segundo lugar, se derivan algunas reglas con ayuda de análisis de regresión; en tercer lugar se llevan a cabo algunos procesos de validación y de sensibilidad; finalmente, se plantean algunas alternativas de políticas públicas enfocadas a la reducción de embarazos no deseados.

Los datos para llevar a cabo esta aplicación se derivan de los resultados de la ENDS 2005 y la ENDS 2010, disponibles en el portal web *The DHS Program*¹. Para algunas variables es necesario hacer la recodificación con base en los manuales que se encuentran en la misma página². En esta aplicación se hace uso principalmente de los datos de la ENDS 2005 para las características de la simulación y para la derivación de reglas. Los datos de la ENDS 2010 son utilizados para verificar si el MBA logra reproducir las características del sistema.

Como se presentó en el capítulo 4, el uso de métodos anticonceptivos está asociado con factores personales, interpersonales e impersonales. Los factores personales están relacionados con características socio-demográficas, creencias, percepciones y experiencias. Los factores interpersonales son los relacionados con las interacciones con otros individuos, mientras que los factores impersonales están relacionados con temas de acceso, costos, políticas públicas, campañas publicitarias, entre otros.

Información para inicializar el modelo: Para inicializar el modelo se usan, además de los resultados de la ENDS 2005 y 2010, la siguientes fuente de información: estudio sobre efecti-

¹<http://www.dhsprogram.com/> Consultado por última vez el 23 de noviembre de 2015.

²<https://dhsprogram.com/pubs/pdf/DHSG4/Recode4DHS.pdf> Consultado por última vez el 23 de noviembre de 2015

vidad de métodos anticonceptivos del Centro para el Control y Prevención de Enfermedades del Departamento de Salud y Servicios Humanos de los Estados Unidos.

En cuanto a la especificación del modelo, las variables a considerar son los siguientes:

VARIABLES DE ENTRADA: (a) Grupo de edad. (b) Estado conyugal. (c) Uso de métodos anticonceptivos y tipo de método usado. (e) Experiencia de fecundidad. (f) Actividad sexual. (g) Fertilidad.

VARIABLES DE SALIDA: (a) Ocurrencia de embarazos. (b) Ocurrencia de embarazos inoportunos o no deseados.

Las variables de entrada serán las características que serán asignadas a los agentes simulados, que en este caso corresponden a mujeres colombianas en edad fértil. Incluir otras variables socio-demográficas como la región, el nivel de educación o el índice de riqueza, en el grupo de variables de entrada, puede contribuir a la explicación del tipo de método anticonceptivo usado, pero no a la ocurrencia de embarazos no deseados o inoportunos. Las relaciones encontradas podrían ser espurias y depender de variables no observables sobre las cuales no es adecuado realizar supuestos. Además, los escenarios de política pública evaluados a través de esta metodología no pueden enfocarse en variables difícilmente modificables y que no estén influenciando directamente la variable de interés.

A partir de las variables de entrada se generan algunos indicadores de especial importancia dentro de la simulación y la formulación de escenarios de una política pública orientada a disminuir los embarazos inoportunos y no deseados. En la ENDS 2010 se encuentra que 7 % de las mujeres en unión que no están usando métodos anticonceptivos tienen demanda insatisfecha. En 5.1 se presentan cómo las variables de entrada ayudan a definir la demanda insatisfecha y la necesidad de uso de métodos anticonceptivos, ya sea para espaciar o limitar los embarazos, es decir disminuir la ocurrencia de embarazos inoportunos o no deseados.

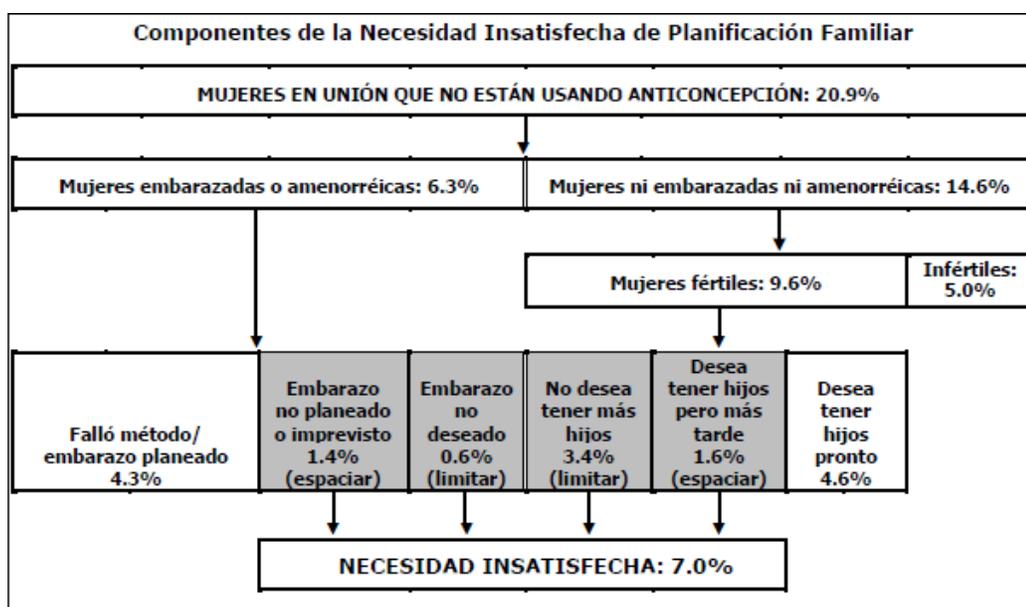


FIGURA 5.1. .

Componentes de la Necesidad Insatisfecha de Planificación Familiar. Tomada de la ENDS 2010.

Para el estudio se define la siguiente tipología jerárquica de métodos anticonceptivos, yendo de los menos efectivos a los más efectivos: (a) Métodos tradicionales o folclóricos (abstinencia periódica, retiro, métodos folclóricos). (b) Métodos de barrera (Espumas/jaleas, condón femenino o masculino). (c) Métodos hormonales (MELA, píldora, inyección, implantes, anticoncepción de emergencia). (d) Métodos intrauterinos (dispositivos intrauterinos) (e) Métodos definitivos (esterilización masculina y femenina).

Teniendo en cuenta la información presentada en 5.1 se pueden plantear tres grupos de mujeres con necesidades diferentes en términos de anticoncepción. En este trabajo se amplían los grupos de demanda insatisfecha considerando a las mujeres no unidas sexualmente activas. En primer lugar, las mujeres no unidas que no han tenido hijos ni están embarazadas y que no usan métodos anticonceptivos, para las cuales los métodos anticonceptivos a usar deberían ser los métodos de barrera (como mínimo), con el fin de disminuir el riesgo de embarazos inoportunos y no deseados y a la vez disminuir el riesgo de contagio de enfermedades de transmisión sexual. En segundo lugar, las mujeres unidas y no unidas que no están usando anticoncepción, que se encuentran embarazadas y no lo tenían planeado o es imprevisto o aquellas que no están embarazadas pero desean tener hijos pero más tarde. Este grupo de mujeres corresponde a aquellas que necesitan espaciar los embarazos, es decir la necesidad de disminuir los embarazos inoportunos y para las cuales se recomienda el uso de métodos anticonceptivos hormonales o intrauterinos. Por último, un tercer grupo de mujeres conformado por mujeres unidas o no unidas que no están utilizando anticoncepción, que se encuentran embarazadas y no lo deseaban o que no están embarazadas pero no desean tener más hijos. Este grupo corresponde a mujeres que necesitan limitar los embarazos, es decir la necesidad de disminuir los embarazos inoportunos y para las cuales se recomienda el uso de métodos definitivos. De manera análoga, aunque para las mujeres que si usan métodos anticonceptivos no se puede hablar de necesidad insatisfecha, se introduce el concepto de uso de método idóneo, el cual corresponde al método recomendado para cada uno de los anteriores grupos.

Las siguientes tablas presentan información útil para inicializar los parámetros de la simulación. Las distribuciones son presentadas con respecto al total fila:

TABLA 5.1. Distribución por grupo de edad y estado conyugal.

ENDS 2005			ENDS 2010		
	No unidas	Unidas		No unidas	Unidas
Entre 15 y 19 años	86.5 %	13.5 %	Entre 15 y 19 años	86.3 %	13.7 %
Entre 20 y 24 años	56.9 %	43.1 %	Entre 20 y 24 años	58.7 %	41.3 %
Entre 25 y 29 años	40.8 %	59.2 %	Entre 25 y 29 años	39.8 %	60.2 %
Entre 30 y 34 años	33.5 %	66.5 %	Entre 30 y 34 años	32.4 %	67.6 %

Fuente: ENDS 2005, ENDS 2010. Cálculos propios.

Para complementar la información sobre uso pasado y presente de métodos anticonceptivos, en el siguiente resultado se adopta una división entre los métodos anticonceptivos dependiendo de sus características y haciendo una distinción entre métodos tradicionales y modernos, de acuerdo a la tipología presentada al principio de este capítulo. Debido a que en el uso pasado las mujeres pudieron haber señalado varias opciones, se define como método pasado aquel que se encuentre más arriba en la escala jerárquica definida, siendo los métodos tradicionales y folclóricos los que

TABLA 5.2. Distribución por estado conyugal y actividad sexual.

ENDS 2005				ENDS 2010			
	Nunca ha tenido relaciones sexuales	No es activa sexualmente	Si es activa sexualmente		Nunca ha tenido relaciones sexuales	No es activa sexualmente	Si es activa sexualmente
Entre 15 y 19 años	56.2 %	20.1 %	23.7 %	Entre 15 y 19 años	50 %	23.9 %	26.1 %
Entre 20 y 24 años	12.8 %	29.7 %	57.5 %	Entre 20 y 24 años	9.3 %	32.2 %	58.5 %
Entre 25 y 29 años	5.4 %	27.3 %	67.3 %	Entre 25 y 29 años	2.8 %	26.9 %	70.3 %
Entre 30 y 34 años	2.7 %	27.5 %	69.9 %	Entre 30 y 34 años	1.5 %	27.4 %	71.1 %

Fuente: ENDS 2005, ENDS 2010. Cálculos propios.

TABLA 5.3. Distribución por uso actual y uso pasado de métodos anticonceptivos

ENDS 2005			ENDS 2010		
	Usa en el presente	No usa en el presente		Usa en el presente	No usa en el presente
Ha usado	67.2 %	32.8 %	Ha usado	69.1 %	30.9 %
No ha usado	0 %	100 %	No ha usado	0 %	100 %

Fuente: ENDS 2005, ENDS 2010. Cálculos propios.

se encuentran en el primer nivel y los métodos definitivos en el último. Esto significa que si una mujer señala, por ejemplo, haber usado un método de barrera y uno hormonal, se usará el método hormonal como referencia para los análisis relacionados con uso pasado de métodos anticonceptivos. Para facilitar la lectura de los resultados de la ENDS 2005, a continuación se presentan las frecuencias relativas con respecto al total fila:

TABLA 5.4. Distribución relativa por tipo de método anticonceptivo usado actualmente y uso pasado.

¿Ha usado alguna vez?	Uso presente de métodos anticonceptivos					
	No usa en el presente	Método tradicional	Método de barrera	Método hormonal	Método intrauterino	Método definitivo
No ha usado	100 %	0 %	0 %	0 %	0 %	0 %
Método tradicional	56.5 %	42.4 %	0 %	1.1 %	0 %	0 %
Método de barrera	58.5 %	11.8 %	29.5 %	0.3 %	0 %	0 %
Método hormonal	38.8 %	8.8 %	10.7 %	41.7 %	0 %	0 %
Método intrauterino	16.9 %	5.5 %	6.4 %	15.5 %	55.8 %	0 %
Método definitivo	0.8 %	0 %	0.3 %	0.5 %	0.2 %	98.2 %

Fuente: ENDS 2005. Cálculos propios.

5.2. Planteamiento del modelo

Intuitivamente, la ocurrencia de embarazos no deseados o inoportunos debería estar relacionado únicamente con la experiencia de fecundidad, expresada como una combinación de tres variables: número de hijos vivos, número de hijos deseados y tiempo que les gustaría esperar desde ahora hasta el nacimiento de un hijo. A partir de estas variables se define la experiencia de fecundidad a partir del método de Lightbourne³, con 4 niveles: (a) Número de hijos deseados menor a número de hijos vivos, (b) Número de hijos deseados igual a número de hijos vivos, (c) Número de hijos deseados mayor a número de hijos vivos y quiere espaciar el próximo nacimiento, (d) Número de hijos deseados mayor a número de hijos vivos y desea tener pronto un hijo. Sin embargo, existen otros factores diferentes a la experiencia de fecundidad, que determinan la ocurrencia de un embarazo inoportuno o no deseado.

5.2.1. Definición de reglas para la simulación

Un primer análisis a realizar para incluir en la especificación de la simulación, es la relación de algunas variables observables con la ocurrencia de embarazos. Para ello, se plantea otra regresión logística utilizando como variable respuesta la ocurrencia de embarazos y como variables explicativas el tipo de uso pasado de métodos anticonceptivos y el grupo de edad.

En un modelo de regresión logístico se tiene que:

$$\pi_i = P(Y_i = 1 | \mathbf{X}_i = \mathbf{x}_i) = \frac{\exp \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}{1 + \exp \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}} \quad (5.1)$$

lo que es equivalente a:

$$\text{logit}(\pi_i) = \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad (5.2)$$

de la anterior ecuación se derivan las razones de chances (OR por sus siglas en inglés, *odds ratio*) al calcular $\exp(\beta_j)$.

Los resultados del análisis de regresión permiten identificar qué tanto el tipo de método anti-conceptivo usado en el pasado y el grupo de edad, tienen un efecto significativo en la explicación del comportamiento de la ocurrencia de embarazos. A continuación se presenta el resumen del análisis exploratorio realizado:

Como se presenta en la siguiente tabla, al calcular las razones de chances se encuentra que para los grupos de edad de 20 a 24 años y de 25 a 29 años, estas no son significativamente diferentes de 1. Por el contrario, se observa una disminución en la razón de chances relacionada con el grupo de edad de 30 a 34 años.

El comportamiento observado en el anterior análisis de regresión, en cuanto al grupo de edad, coincide con diferencias importantes observadas en la fertilidad entre las mujeres de 20 a 29 años y las de 30 a 39 años, lo que hace pensar que la fertilidad es el mecanismo que genera estas diferencias. Este supuesto no es posible probarlo debido a que la pregunta sobre problemas de

³Lightbourne, R. (1985). Desired Number of Births and Prospects for Fertility Decline in 40 Countries. *International Family Planning Perspectives* 11 (2) pp. 34-39.

TABLA 5.5. Análisis exploratorio de potenciales determinantes de la ocurrencia de embarazos.

Variable	Estimación	Desv. Est.	Valor z	Pr(> z)
Intercepto	-1.3483	0.0848	-15.91	0.0000
Uso pasado trad.	-0.3446	0.1269	-2.72	0.0066
Uso pasado barrera	-0.7499	0.0997	-7.52	0.0000
Uso pasado horm.	-0.6510	0.0867	-7.51	0.0000
Uso pasado intraut.	-0.9918	0.1069	-9.28	0.0000
Uso pasado def.	-16.7152	139.8514	-0.12	0.9049
Edad 20-24	-0.2443	0.0723	-3.38	0.0007
Edad 25-29	-0.4651	0.0806	-5.77	0.0000
Edad 30-34	-0.5814	0.0887	-6.56	0.0000

Fuente: ENDS 2005. Cálculos propios.

Variable	OR	2.5 %	97.5 %
(Intercept)	0.26	0.22	0.31
Uso pasado trad.	0.71	0.55	0.91
Uso pasado barrera	0.47	0.39	0.57
Uso pasado horm.	0.52	0.44	0.62
Uso pasado intraut.	0.37	0.30	0.46
Uso pasado def.	0.00	0.00	0.00
Edad 20-24	0.78	0.68	0.90
Edad 25-29	0.63	0.54	0.74
Edad 30-34	0.56	0.47	0.66

Fuente: ENDS 2005. Cálculos propios.

fertilidad solo se hace a mujeres que no están embarazadas y que han buscado tener hijos. Por su parte, en la variable de uso pasado, además de tener un problema en su definición, no se puede determinar si el tipo de método era el usado en el momento de quedar embarazada. Por estas razones, en la simulación no puede incluirse una especificación determinística de la ocurrencia de embarazo. La alternativa que se adopta es simular la ocurrencia de embarazo a través de dos factores: la efectividad del método y la fertilidad de las mujeres.

Por su parte, en cuanto a la ocurrencia de embarazos no deseados o inoportunos, en una primera simulación se especifica dependiendo únicamente de la experiencia de fecundidad, sin embargo después de correr 100 simulaciones con 1000 mujeres, se observan tasas de embarazos no deseados e inoportunos bastante altas en comparación con lo observado en la ENDS 2010. Lo anterior hace sospechar de posibles sesgos de esta variable y que en la ENDS 2010 ([70]) se asocia a los problemas en la medición de la fecundidad deseada por ser preguntas que “se refieren al pasado y por lo tanto conllevan riesgos de memoria y de veracidad o de racionalización de la respuesta, frente a hechos ya consumados y posiblemente de trascendencia afectiva”.

Para contrastar el anterior hallazgo, a continuación se presentan los resultados de una regresión logística que usa como variable respuesta la ocurrencia de embarazos inoportunos o no deseados y como variables explicativas la experiencia en fecundidad, el estado marital y el grupo de edad. Las anteriores son variables con información disponible en la ENDS 2010 y que entrarían a jugar el papel de variables de entrada en el sistema simulado. Estas variables además están relacionadas

con esa racionalización de la respuesta por ser determinantes de la situación afectiva y de percepción frente al embarazo por parte de la mujer. Esta regresión se hace con la información de las mujeres que reportaron estar actualmente embarazadas.

TABLA 5.6. Análisis exploratorio de potenciales determinantes de la ocurrencia de embarazos no deseados o inoportunos.

Variable	Estimación	Desv. Est.	Valor z	Pr(> z)
Intercepto	0.2406	0.4229	0.57	0.5693
Fecund. deseada mayor observada espac.	0.6301	0.4225	1.49	0.1359
Fecund. deseada igual observada	1.4996	0.4211	3.56	0.0004
Fecund. deseada menor observada	3.1471	0.4822	6.53	0.0000
Unida	-1.0008	0.1355	-7.39	0.0000
Edad 20-24	-0.3443	0.1467	-2.35	0.0189
Edad 25-29	-0.5704	0.1669	-3.42	0.0006
Edad 30-34	-1.0328	0.1905	-5.42	0.0000

Fuente: ENDS 2005. Cálculos propios.

Los resultados de la regresión logística permiten comprobar que las tres variables definidas como explicativas tienen un efecto significativo en la explicación de la ocurrencia de embarazos no deseados e inoportunos. Al calcular las razones de chances, se encuentra que la experiencia de fecundidad relacionada con número de hijos igual o menor al número de hijos vivos, aumenta la probabilidad de que el embarazo haya sido no deseado o inoportuno frente a la probabilidad de que no lo haya sido. No se encuentran diferencias relacionadas entre los dos niveles que discriminan por el tiempo en que le gustaría tener el próximo hijo, sin embargo este nivel se mantiene debido a su importancia en el estudio de la ocurrencia de embarazos inoportunos. También se encuentra que entre mayor es el grupo de edad al cual pertenecen las mujeres embarazadas, disminuye la razón entre la probabilidad de que el embarazo haya sido no deseado o inoportuno frente a la probabilidad de que no lo haya sido. Por su parte, el hecho de estar unidas disminuye la razón entre la probabilidad de embarazo no deseado o inoportuno frente a la probabilidad de embarazo deseado.

TABLA 5.7. Razón de chances para los diferentes niveles de las variables explicativas incluidas en el análisis.

Variable	OR	2.5 %	97.5 %
Intercepto	1.27	0.54	2.87
Fecund. deseada mayor observada espac.	1.88	0.84	4.45
Fecund. deseada igual observada	4.48	2.01	10.61
Fecund. deseada menor observada	23.27	9.27	62.02
Unida	0.37	0.28	0.48
Edad 20-24	0.71	0.53	0.94
Edad 25-29	0.57	0.41	0.78
Edad 30-34	0.36	0.24	0.52

Fuente: ENDS 2005. Cálculos propios.

A partir de los resultados de la regresión logística, se implementa un análisis discriminante para definir dentro de la simulación, dependiendo los valores en las 3 variables explicativas, si el embarazo de la mujer es no deseado o inoportuno. Adicionalmente, se lleva a cabo un análisis para definir el umbral (o *cutoff*) a ser usado, buscando maximizar la sensibilidad (tasa de verdade-

ros positivos) y minimizar la especificación (tasa de verdaderos negativos). En 5.2 se presenta el análisis y se define como umbral una probabilidad de 0.625.

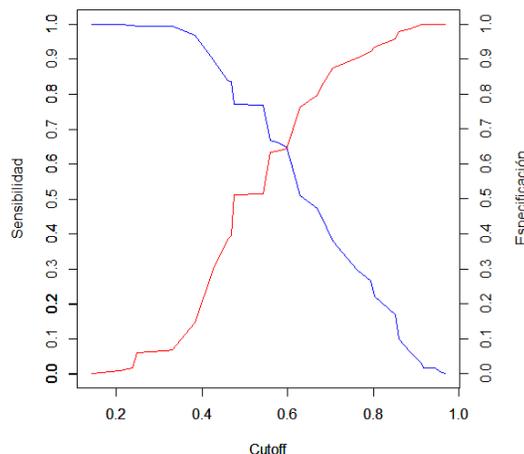


FIGURA 5.2. Umbral para regresión logística.

Un segundo ejercicio preliminar de otras 100 simulaciones con 1000 mujeres, permite identificar porcentajes muy altos de embarazo en comparación con los hallazgos de la ENDS 2010. Una segunda revisión de las variables incluidas permite identificar que no se ha considerado la terminación del embarazo por causas naturales o artificiales. En el gráfico 5.3 se presenta el esquema general del sistema que se está estudiando en esta aplicación, después de haber incluido los hallazgos preliminares, y las relaciones generales entre las variables.

Para continuar con el ejercicio de la aplicación, se toman las siguientes decisiones relacionadas con la simulación: (a) se simulan 1000 mujeres en edades entre 15 y 34 años conservando las características de las mujeres colombianas en esas edades, relacionadas con estado conyugal, actividad sexual y experiencia de fecundidad. (b) Se asigna de manera aleatoria el tipo de método anticonceptivo usado en el presente para liberar esta variable de otros factores y estudiar sus efectos sobre la ocurrencia de embarazos. (c) Para cada tipo de método anticonceptivo, la ocurrencia de embarazo está determinada por la efectividad del método y la fertilidad de las mujeres. La efectividad del método se asigna como un valor entre 0 y la mayor tasa de efectividad encontrada para cada tipo de métodos anticonceptivos. Una aproximación de dicha efectividad es calculada con base en la información del gráfico 4.1. (d) En el momento 0 la totalidad de las mujeres no se encuentran embarazadas. Cada instante de tiempo que pasa en la simulación corresponde a un mes. En cada mes un porcentaje de mujeres cambian algunas de las variables de la simulación, comenzando por el tipo de método anticonceptivo y las variables relacionadas. (e) La experiencia de fecundidad se deriva de las preguntas de la ENDS relacionadas con el número de hijos deseados, número de hijos vivos y el tiempo que le gustaría esperar antes del próximo nacimiento y la diferencia entre número de hijos deseados y número de hijos vivos. (f) Se incluye la ocurrencia de problemas de fertilidad y terminación de embarazos para mejorar la explicación del fenómeno.

En cuanto a la fertilidad, hay que aclarar que esta variable tiene censura, esto es, solo es observable cuando la mujer ha intentado tener hijos y no lo ha logrado. Sin embargo esta no es una dificultad para la simulación porque esta variable solo tendrá efecto cuando la mujer vaya a quedar embarazada. A continuación se presenta la distribución de los problemas de fertilidad por grupo de edad.

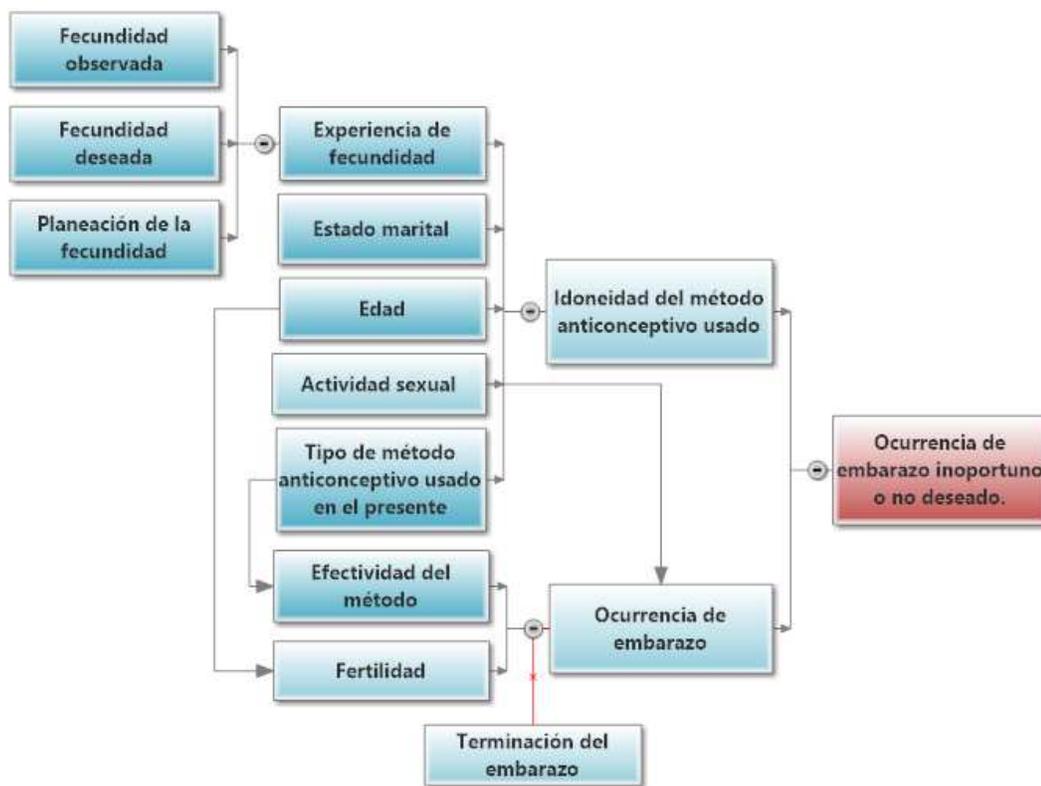


FIGURA 5.3. .

Esquema general de la aplicación.

TABLA 5.8. Distribución de problemas de fertilidad por grupo de edad.

	Sin problemas de fertilidad	Con problemas de fertilidad
Entre 15 y 19 años	98.6 %	1.4 %
Entre 20 y 24 años	93.3 %	6.7 %
Entre 25 y 29 años	86.1 %	13.9 %
Entre 30 y 34 años	82 %	18 %

Fuente: ENDS 2005. Cálculos propios.

En cuanto a la terminación del embarazo, la información contenida en la ENDS no permite relacionar su ocurrencia con la edad en la que el evento sucedió. Sin embargo, se estima que 14 % de los embarazos de las mujeres entre 15 y 34 años, que alguna vez han estado embarazadas, presentaron terminación. Cerca de 50 % de estas terminaciones de embarazo fueron pérdidas, aproximadamente 30 % abortos o interrupciones y el resto corresponden a muerte fetal intrauterina y embarazo extrauterino. Es importante aclarar que esta variable suele tener sesgos asociados a omisión de respuesta o veracidad por tratarse de información sensible. Otro aspecto que debe mencionarse es la entrada en vigencia de la sentencia C-355 del 10 de mayo de 2006 de la Corte Constitucional de Colombia, en donde se explican los motivos por los cuales se despenaliza en el país el aborto voluntario en tres casos. En la ENDS 2010, primera medición después de la sentencia, no se encontró una prevalencia alta de abortos cobijados con esta sentencia, razón por la cual este factor no se incluye en la simulación. Para la simulación se va suponer que las terminaciones de embarazo ocurren con una tasa de 14 % y se presentan a los 3 meses de embarazo.

Un nuevo ejercicio preliminar con 100 simulaciones, permite encontrar un número muy alto de embarazos por mujer, en el período de simulación. Esto conduce a pensar que en la simulación no se están reproduciendo de manera adecuada las características de fecundidad de las mujeres colombianas en edad fértil. El intervalo intergenésico corresponde al tiempo que transcurre entre los nacimientos, es decir, el factor que podría explicar el comportamiento atípico observado en la simulación. Este intervalo se simula, para cada grupo de edad, con una distribución *poisson* con λ igual a la mediana observada en la ENDS 2005, esto es, 24.4 meses para mujeres entre 15 y 19 años; 33.8 meses para mujeres entre 20 y 29 años y 57.1 meses para mujeres entre 30 y 39 años. Dentro de la simulación se asume que la terminación del embarazo ocurre en el tercer mes de embarazo. En cuanto al tiempo que transcurre entre una terminación de un embarazo y el siguiente embarazo, no se tiene ninguna información. Sin embargo, en la ENDS 2005 y ENDS 2010 se encuentra que el tiempo entre un embarazo y otro es menor en cerca de 15 meses cuando no hay sobrevivencia del hijo anterior. Esto hace pensar que el tiempo hasta el siguiente embarazo cuando se presenta terminación podría estar alrededor de 15 meses.

Para el desarrollo de la aplicación se presentarán algunos escenarios y se analizarán los cambios en los resultados. Progresivamente en cada escenario se incluirán aspectos metodológicos y se irá volviendo más complejo el esquema de la simulación y de los análisis estadísticos relacionados con análisis de sensibilidad, calibración, validación e inferencia estadística.

5.3. Implementación de la metodología

Escenario 1: En este escenario se incluyen dentro de la simulación las probabilidades de transición asociadas a cambios en el estado marital y el tipo de métodos anticonceptivos usados. Para inicializar las matrices de transición de estas dos variables, se usa información de la ENDS 2005 del calendario que es incluido al final del cuestionario individual, en donde las mujeres debían registrar eventos relacionados con 4 aspectos: (a) Nacimientos, embarazos, uso de métodos anticonceptivos, (b) Discontinuidad de uso anticonceptivo, (c) Matrimonio/Unión, (d) Movimientos y tipos de localidades.

Usando información del primer aspecto del calendario y tomando en cuenta a las mujeres que hubieran iniciado su vida sexual, que no tuvieron embarazos, nacimientos o terminaciones de embarazo, se estima la matriz de transición entre métodos anticonceptivos utilizando la tipología usada a lo largo de este capítulo. Adicionalmente se excluyen las mujeres mayores de 34 años por no hacer parte de la población de estudio y porque su comportamiento puede ser diferente al de las mujeres menores.

En cuanto al estado conyugal, se calcula la siguiente matriz de transición, la cual será un insumo para definir la parte dinámica del escenario 1 en la simulación. En este caso también se excluyen las mujeres mayores de 34 años.

El uso de métodos anticonceptivos entre mujeres de 15 a 34 años que alguna vez han tenido relaciones sexuales y no se encuentran embarazadas era, con base en cálculos de la ENDS 2005, de 68.5 %, mientras que para 2010 este valor se estima en 71.7 %. Al realizar un primer ejercicio de simulación, se encuentra que este porcentaje está siendo en promedio 20 % menor. Sin embargo, al observar la matriz de transición asociada a esta variable, se puede inferir que esta es la causa del comportamiento, debido a que las probabilidades de transición entre los diferentes métodos y ningún método, son relativamente altas. Por otro lado, hay un rezago en la información debido a

TABLA 5.9. Matriz de transición entre métodos anticonceptivos.

	Ninguno	Tradicional	Barrera	Hormonal	Intrauterina	Definitiva
Ninguno	0.995	0.001	0.002	0.002	0.000	0.000
Tradicional	0.030	0.959	0.005	0.006	0.001	0
Barrera	0.043	0.004	0.945	0.007	0.001	0
Hormonal	0.0020	0.003	0.007	0.969	0.001	0.000
Intrauterino	0.007	0.001	0.001	0.003	0.988	0.000
Definitiva	0.004	0.000	0.000	0.002	0.001	0.993

Fuente: ENDS 2005. Cálculos propios.

TABLA 5.10. Matriz de transición entre estados conyugales.

	No unida	Unida
No unida	0.997	0.003
Unida	0.014	0.986

Fuente: ENDS 2005. Cálculos propios.

que no se tienen datos sobre la transición entre los estados de la actividad sexual. Para corregir esta inconsistencia, se propone volver a calcular la matriz de transición pero sin considerar las mujeres que descontinuaron el uso de métodos anticonceptivos por factores ajenos a los relacionados con el método o con fecundidad (por ejemplo porque la pareja desaprobó, acceso o disponibilidad, costo, entre otros). Esto se hace debido a que esos factores podrían estar sobreestimando las probabilidades de transición de pasar a usar ningún método. Al excluir estos individuos, la matriz de transición queda definida como se presenta en la siguiente tabla.

TABLA 5.11. Matriz de transición entre métodos anticonceptivos excluyendo mujeres que hayan descontinuado el uso por factores ajenos a los métodos.

	Ninguno	Tradicional	Barrera	Hormonal	Intrauterina	Definitiva
Ninguno	0.970	0.003	0.009	0.016	0.002	0
Tradicional	0.023	0.970	0.001	0.005	0.001	0
Barrera	0.033	0.004	0.956	0.006	0.001	0
Hormonal	0.0014	0.002	0.006	0.976	0.001	0.001
Intrauterino	0.001	0.001	0.001	0.002	0.995	0.000
Definitiva	0.001	0.000	0.000	0.002	0.001	0.995

Además de implementar dentro de la simulación la matriz de transición entre métodos anticonceptivos, excluyendo las mujeres que hayan descontinuado su uso por factores ajenos a los métodos, en el escenario 2 se incluye un esquema dinámico que simula el envejecimiento de las mujeres. Para ello, en cada instante de tiempo la mujer suma un mes más en su edad. Cuando la mujer cumple 20, 25 ó 30 años se define de nuevo su grupo de edad y la actividad sexual. Con el fin de seguir estudiando el comportamiento de las mujeres menores a 35 años, el grupo inicial de mujeres se simula de acuerdo a las características de mujeres de 15 a 30 años. De esta forma, los resultados después de una simulación de 60 meses corresponderá a mujeres en edades entre 20 y 35 años.

En este segundo escenario se prueba el efecto que tiene sobre los resultados de la simulación el cambio en los siguientes parámetros: el porcentaje de terminación de embarazos, el tiempo que pasa entre la terminación de un embarazo y volver a intentar quedar embarazada, el momento en el que las mujeres cambian su actividad sexual. El análisis sobre este último parámetro se incluye debido a que, en la simulación inicial del escenario 1, la actividad sexual cambia justo en el momento en que cambia de grupo de edad, pero se espera que esta presente cambios meses antes de ese momento. Para el análisis de cada parámetro, se fijan los otros dos en los valores iniciales disponibles, esto es, 14 % de terminación de embarazos, 15 meses entre la terminación de un embarazo y volver a intentar quedar embarazadas y 0 meses antes del momento en que cambian el grupo de edad.

Los siguientes gráficos resumen los resultados de la simulación bajo diferentes escenarios en cuanto a porcentaje de terminación de embarazos. Para cada parámetro se presentan los resultados que reflejan los cambios tanto para el porcentaje de mujeres embarazadas como para el porcentaje de embarazos no deseados o inoportunos:

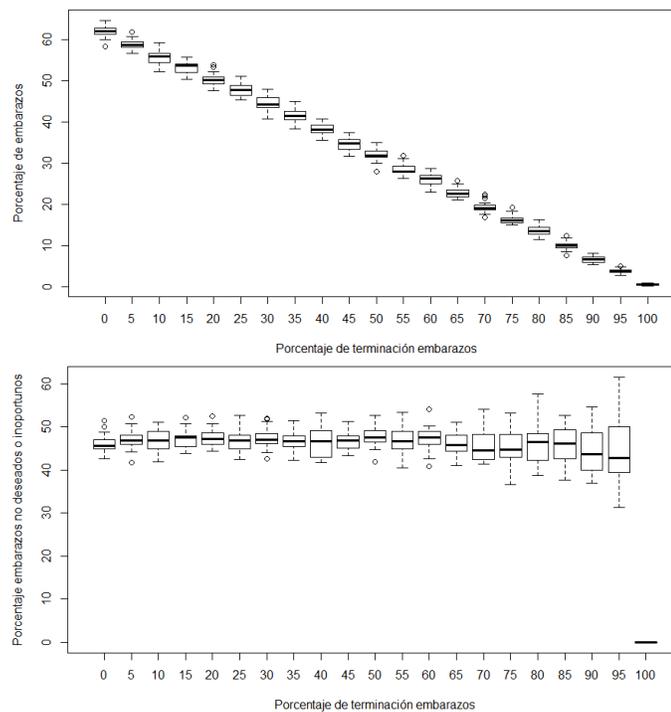


FIGURA 5.4. .

Análisis de sensibilidad. Escenario 1.

Este análisis se repite con valores cercanos a 14 % que es el porcentaje de mujeres que manifiestan haber tenido problemas de fertilidad en la ENDS 2005. Los gráficos con los resultados se presentan a continuación.

Estos análisis permiten identificar el papel que están desempeñando las variables en el modelo computacional para así determinar si corresponde con el modelo teórico planteado para el MBA. Se encuentra que el porcentaje de embarazos disminuye a medida que el porcentaje de terminaciones de embarazos aumenta en la población de estudio. Sin embargo, de acuerdo a lo esperado, esta variable no tiene un efecto directo sobre el porcentaje de embarazos no deseados o inoportunos, excepto para valores cercanos a 100 %. Al realizar los análisis en valores cercanos al porcentaje

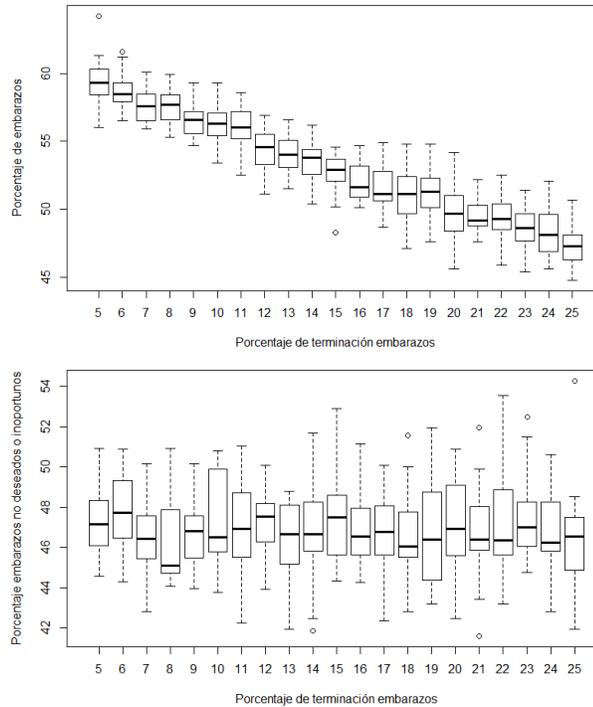


FIGURA 5.5. .

Análisis de sensibilidad. Escenario 1.

estimado, es decir 14 %, los efectos en los cambios de este parámetro se hacen más pequeños, a pesar de que la tendencia decreciente continúa.

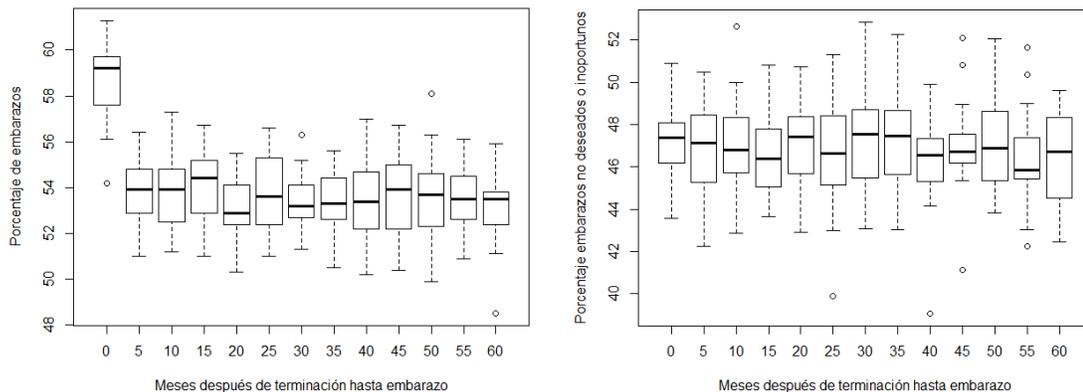


FIGURA 5.6. .

Análisis de sensibilidad. Escenario 1.

En 5.6 se presenta el análisis de sensibilidad del modelo computacional bajo cambios en el parámetro que define cuántos meses pasan en promedio después de la terminación de un embarazo hasta el próximo. Dado que primero se fija el porcentaje de terminación de embarazos, este parámetro resulta poco relevante en la simulación, excepto cuando toma un valor de 0 meses. En este caso también se evidencia que esta variable no tiene un efecto directo sobre la ocurrencia de embarazos no deseados o inoportunos.

Finalmente en 5.7 se muestran los resultados asociados a los cambios en el número de meses, antes de cambiar de grupo de edad, en que cambia la actividad sexual. No se observa una tenden-

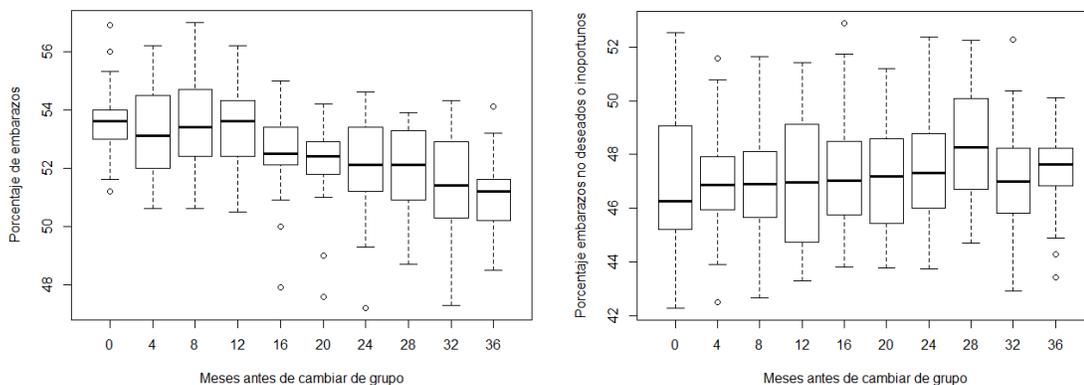


FIGURA 5.7. .

Análisis de sensibilidad. Escenario 1.

cia cuando aumenta el valor del parámetro ni en el porcentaje de mujeres embarazadas ni en la ocurrencia de embarazos no deseados o inoportunos.

Escenario 2: En este escenario se repite el escenario 1 pero en este caso, en cada instante de tiempo se va simulando la entrada a la población de estudio de mujeres de 15 años. Esto permite que los análisis después de los 60 meses de la simulación se orienten a mujeres de 15 a 34 años. Para mantener la distribución por grupos de edad, se simulan mujeres con exactamente 15 años, las cuales también hacen parte del proceso de envejecimiento pero que después de la observación de 60 meses siguen perteneciendo al primer grupo de edad. Como se viene trabajando la simulación con 1000 mujeres, en cada instante de tiempo se crean 6 mujeres, lo que garantiza que la distribución por grupo de edad se mantenga.

Para continuar con el último análisis del escenario 1, en este caso se presentan los cambios en las variables de interés cuando el número de meses en que ocurre el cambio va entre 0 y 12. Los resultados en 5.8 no son concluyentes sobre el posible efecto de este parámetro.

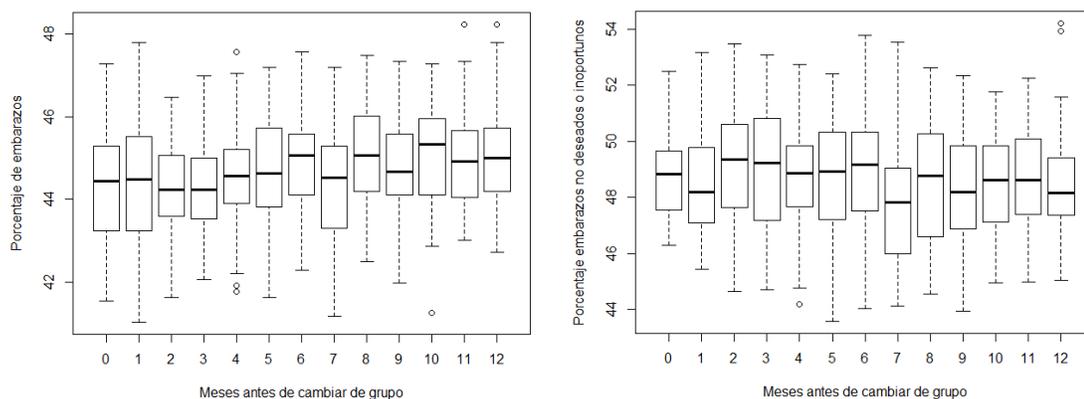


FIGURA 5.8. .

Análisis de sensibilidad. Escenario 2.

Como análisis complementario, se evalúa el efecto que tiene sobre el resultado, cambios simultáneos en los parámetros asociados con el mes en que usualmente terminan los embarazos y los meses, antes de cambiar de grupo de edad, en que cambia la actividad sexual.

En la ENDS se les pregunta a las mujeres embarazadas si el embarazo actual era deseado para este momento o no y se indaga a las mujeres, que tuvieron hijos en los últimos 5 años, si querían quedar embarazadas en el momento o no querían tener más hijos. Con estas dos variables puede calcularse la tasa de embarazos no deseados e inoportunos, que es la cifra que se ha presentado en los diferentes análisis a lo largo de este capítulo. Con el fin de presentar los análisis de sensibilidad y calibración, la tasa de embarazos no deseados e inoportunos es calculada con información de la ENDS 2010, dando un valor de 51.8 % para las mujeres entre 15 y 34 años. En cuanto al porcentaje de mujeres embarazadas o que han estado embarazadas en los últimos 5 años, el porcentaje calculado a partir de los datos del calendario de la ENDS 2010 es 42.8 %. Se definen dos criterios de calibración para cada uno de estos parámetros, estar alrededor de 3 % y de 5 % de los valores estimados en los datos reales. Los resultados se presentan en el gráfico 5.9.

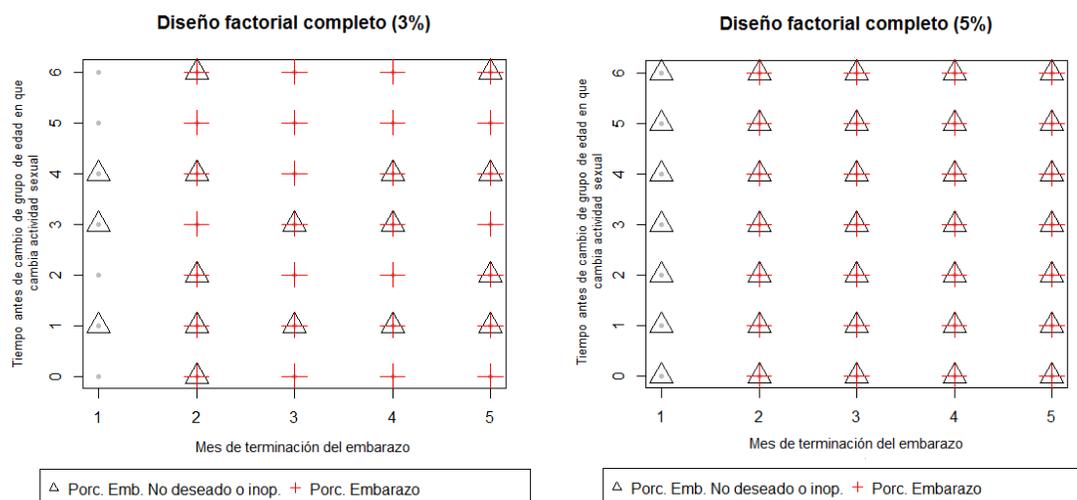


FIGURA 5.9. .

Análisis de sensibilidad. Escenario 2.

Los símbolos que se incluyen en el arreglo sirven para identificar cuáles alternativas dan como resultado valores, para las variables respuesta, dentro de los criterios de calibración. En este caso, se busca identificar qué niveles de las variables mes de terminación del embarazo y mes en el que cambia la actividad sexual antes del cambio de grupo de edad, producen valores en las variables porcentaje de embarazos y porcentaje de embarazos no deseados o inoportunos dentro de un 3 % y 5 % alrededor de los valores observados en la ENDS 2010.

En cuanto a la calibración de los parámetros sobre los que se tiene cierto nivel de incertidumbre, se usa un método de calibración categórica combinado con un algoritmo genético. Este análisis se hace sobre los 4 parámetros que se han analizado durante este escenario y se definen los siguientes dominios para buscar la mejor solución:

- Tiempo en meses en que sucede la terminación de los embarazos. Dominio: entre 1 y 5 meses.
- Tiempo en meses en que cambia la actividad sexual antes de que cambie el grupo de edad. Dominio: entre 0 y 12 meses.
- Tasa de terminación de embarazos. Dominio: entre 10 % y 20 %.

- Tiempo en meses entre la terminación de un embarazo y un próximo embarazo. Dominio: entre 5 y 25 meses.

Los resultados del análisis arrojan que la mejor solución está dada por los valores: 2.8 meses, 6 meses, 15.4 % y 12 meses, respectivamente.

En 5.10 se presentan los resultados del valor de evaluación entre las diferentes generaciones del algoritmo genético y el comportamiento de la función de costo en los diferentes individuos de la población sobre los cuales se desarrollo el algoritmo genético. Al realizar estos análisis se encuentra que los resultados del algoritmo genético son altamente sensibles a los intervalos dados para los parámetros y al tamaño de la población inicial. En la literatura no se encuentran criterios claros respecto a estas dos definiciones dentro del algoritmo. Trabajos futuros pueden explorar este aspecto.

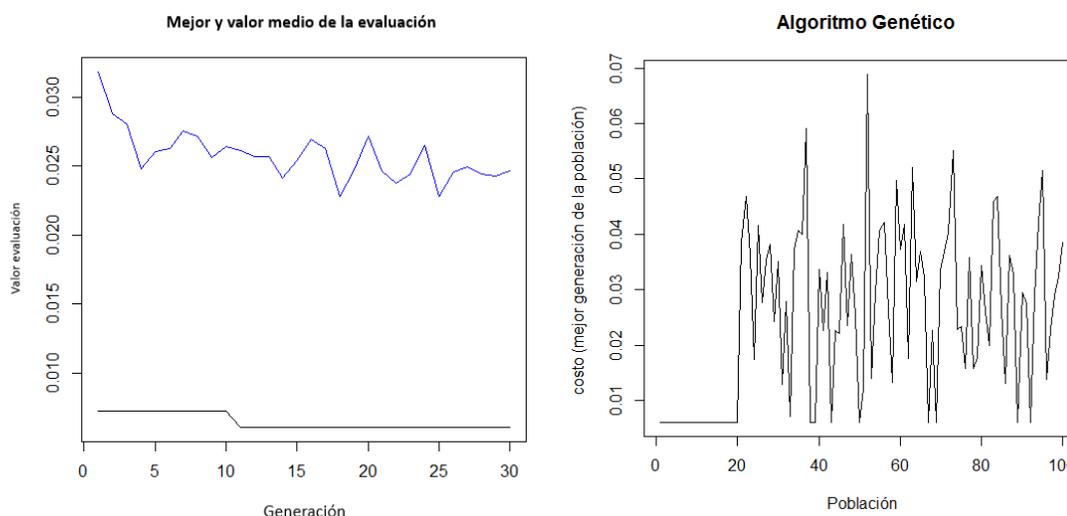


FIGURA 5.10. .

Análisis de calibración - Algoritmo genético. Escenario 2.

Escenario 3: En este escenario se incluyen los fenómenos de formación, crecimiento y aprendizaje en redes. Dentro de la simulación, para la formación de la red se parte del modelo de Watts-Strogatz para generar una red que incorpora el *small world phenomenon*, es decir una red con diámetro pequeño y coeficiente de agrupamiento grande. En cuanto al crecimiento de la red, se simula el comportamiento de un modelo de conexión preferencial sencillo, determinado por la mayor probabilidad de formación de enlaces con los nodos que tienen más enlaces. En cuanto al aprendizaje, lo que se hace es seguir con un modelo de aprendizaje bayesiano que está determinado por la efectividad experimentada por los demás nodos de la red. En este caso, esa efectividad está determinada por la experiencia, con determinado método, de las mujeres de la red en el tiempo $t - 1$. El aprendizaje dentro de la red resulta ser el fenómeno que hace explícito el efecto de la interacción social en el sistema e incorpora simultáneamente los fenómenos de formación y crecimiento, los cuales son los determinantes de la estructura de la red y la conformación de la vecindad de cada individuo. Para implementar el aprendizaje dentro del sistema, los cambios en el tipo de método anticonceptivo usado no dependerán directamente de la matriz de transición entre métodos sino de la efectividad experimentada por las mujeres que conforman la vecindad de cada mujer. Esto significa que se parte de la idea de que la selección de método anticonceptivo está determinada por una función de recompensa $r(x, y)$, siendo x el método seleccionado y y la efectividad

del método. Por lo tanto, el aprendizaje ocurre cuando se busca optimizar la efectividad, dada la experiencia de las mujeres de la vecindad.

En la práctica, se encuentra que el fenómeno de aprendizaje dentro de la red logra reproducir las transiciones entre tipos de métodos anticonceptivos desde el enfoque de la efectividad observada por la vecindad del agente. Esto permite que haya una definición local de los cambios en los métodos anticonceptivos. Además, el hecho de poder explicar los cambios en los métodos anticonceptivos desde el análisis de redes, provee una herramienta para plantear intervenciones basados en la red, por ejemplo, identificar mujeres influyentes (con mayor número de enlaces) e intensificar en ellas campañas para que usen métodos intrauterinos o definitivos. Otra alternativa bajo esta aproximación podría ser utilizar el conocimiento de las redes para difundir información. En este resultado es importante aclarar que las herramientas del análisis de redes lograrían complementar la simulación cuando no hay suficiente información de la toma de decisiones en cada instante de tiempo. En este caso particular, como se cuenta con información de las matrices de transición, el nivel de especificidad del escenario 2 es lo suficientemente alto como para no necesitar el uso de fenómenos de redes sociales. Sin embargo, en fenómenos donde no se cuente con suficiente información o donde se quieran explotar más las relaciones entre individuos, se podría profundizar en este aspecto.

5.4. Alternativas de políticas públicas

Para finalizar la aplicación, se formulan tres alternativas de políticas públicas buscando la reducción del porcentaje de embarazos no deseados o inoportunos. Estas alternativas de política pública serán probadas en los escenarios 2 y 3, una vez se han ajustado los valores de los parámetros de acuerdo a los resultados del análisis de sensibilidad y el análisis de calibración. Sin embargo, los resultados son presentados para el escenario 2 debido a que resultan con un alto grado de similitud con el escenario 3. Además los resultados son presentados en este escenario porque el nivel de especificación es más alto que el del escenario 3. Es importante recordar que para el escenario en el que se incluyen fenómenos de redes fue necesario incluir algunas consideraciones basados en los modelos usuales de crecimiento, formación y aprendizaje.

Una primera alternativa consiste en un aumento en el período intergenésico. Este aumento se simula en número de meses y se aplica sobre el período intergenésico definido para cada grupo de edad. Una segunda alternativa consiste en cambiar en determinado porcentaje la distribución entre tipos de métodos, disminuyendo ese porcentaje para los grupos de mujeres que no usan, usan métodos tradicionales y usan métodos de barrera y aumentando ese mismo porcentaje en los métodos hormonales, intrauterinos y definitivos. Por ejemplo, si la intervención es de 1 %, este porcentaje se reduce en la distribución inicial de mujeres que no usan ningún método, usan un método tradicional o uno de barrera y se aumenta en 1 % en la distribución inicial de los métodos hormonales, intrauterinos y definitivos. Esta intervención se aplicará en la distribución por tipos de método anticonceptivo de cada tipo de actividad sexual. Una tercera alternativa consiste en retrasar los cambios en la actividad sexual asociados al cambio en el grupo de edad. Esta alternativa tiene la dificultad del rezago de información que se ha mencionado en este capítulo.

Alternativa 1:

Al implementar esta alternativa en el escenario 2 se observa que no hay un efecto claro del aumento en el período intergenésico sobre las variables de interés. Una razón podría ser que el

tiempo de observación es muy corto, razón por la cual se simula de nuevo el modelo computacional para un total de 120 meses. Los resultados se presentan en el gráfico 5.11, sin embargo no son concluyentes.

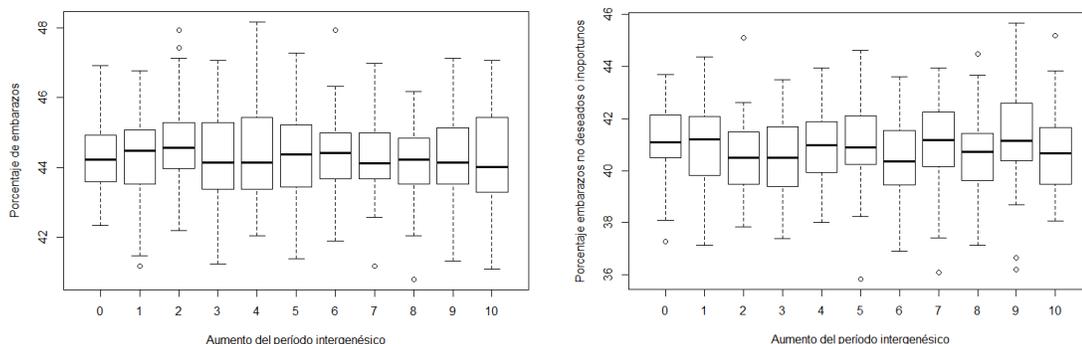


FIGURA 5.11. .

Resultados descriptivos. Escenario 2 - Alternativa 1

Alternativa 2:

En el escenario 2, los resultados descriptivos (5.12) evidencian que esta alternativa tiene un efecto sobre la disminución del porcentaje de embarazos pero no tiene un efecto claro sobre la ocurrencia de embarazos no deseados o inoportunos. Este resultado permite pensar en una alternativa 4 que se concentre en la idoneidad del método, la cual es un factor que no ha sido considerado en los análisis. Una cuarta alternativa de política pública consiste en aumentar el porcentaje de mujeres que usan un método idóneo de acuerdo a su experiencia de fecundidad y su actividad sexual.

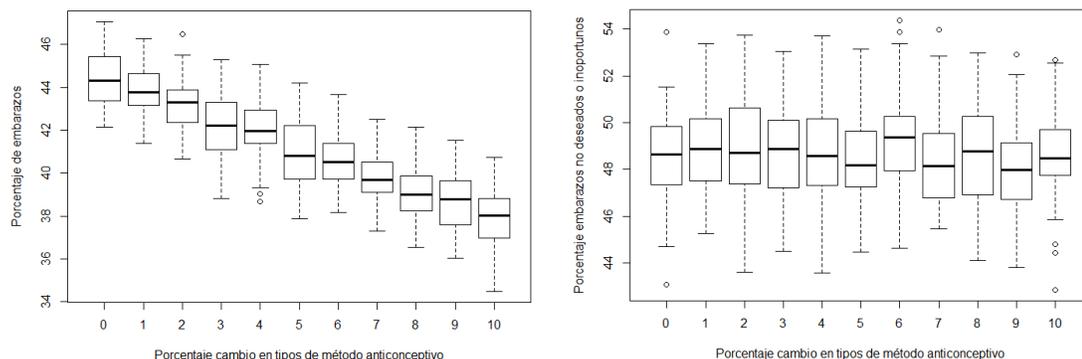


FIGURA 5.12. .

Resultados descriptivos. Escenario 2 - Alternativa 2

Alternativa 3:

Al implementar la intervención en el escenario 2, se observa que existe un efecto del número de meses que se pospone el cambio en la actividad sexual de las mujeres, con respecto a su grupo de edad. La principal razón de este efecto es que las mujeres del grupo de edad de 15 a 19 años pasan de ser en su mayoría mujeres que no han iniciado su vida sexual, a ser mujeres mayoritariamente con una vida sexual activa en el grupo de edad de 20 a 24 años. Una intervención orientada a posponer estos cambios en la actividad sexual parece disminuir tanto los porcentajes de embarazos como la ocurrencia de embarazos no deseados e inoportunos.

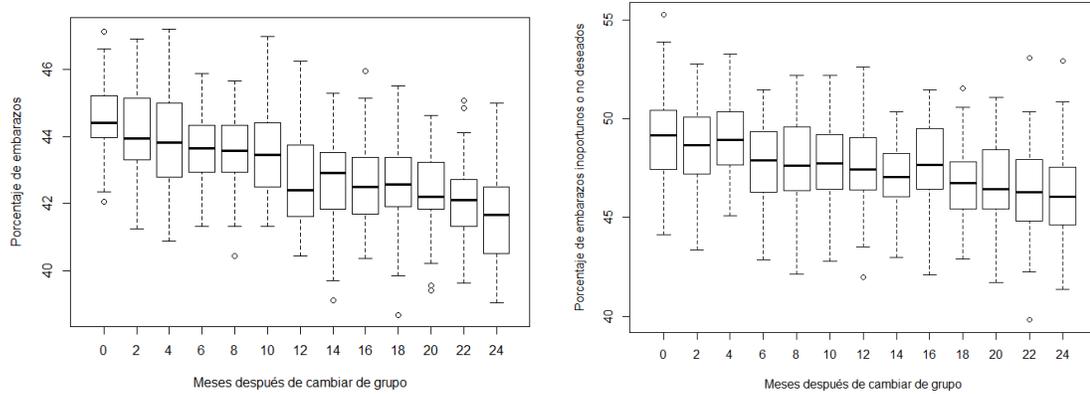


FIGURA 5.13. .

Resultados descriptivos. Escenario 2 - Alternativa 3

Alternativa 4:

Esta alternativa surge como respuesta a lo observado en la alternativa 2 y consiste en aumentar el porcentaje de mujeres que usan un método idóneo de acuerdo a su experiencia de fecundidad, actividad sexual y su estado marital. En 5.14 se presenta un primer resultado interesante de esta alternativa y es el aumento progresivo del porcentaje de mujeres que usan un método anticonceptivo a medida que aumenta el porcentaje de mujeres a las que se les asigna directamente el método anticonceptivo idóneo.

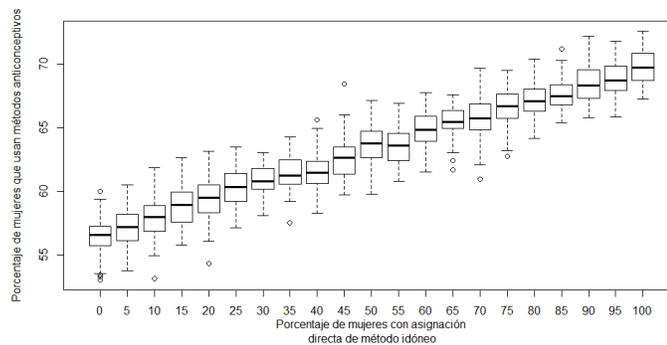


FIGURA 5.14. .

Resultados descriptivos. Escenario 2 - Alternativa 4

En 5.15 se presentan los cambios observados en las simulaciones en las variables respuesta ante cambios en el porcentaje de mujeres a las que se les asigna directamente el método anticonceptivo idóneo, de acuerdo a sus características. Es evidente que esta alternativa de política pública tendría consecuencias importantes en la reducción del porcentaje de embarazos y en el porcentaje de embarazos no deseados o inoportunos.

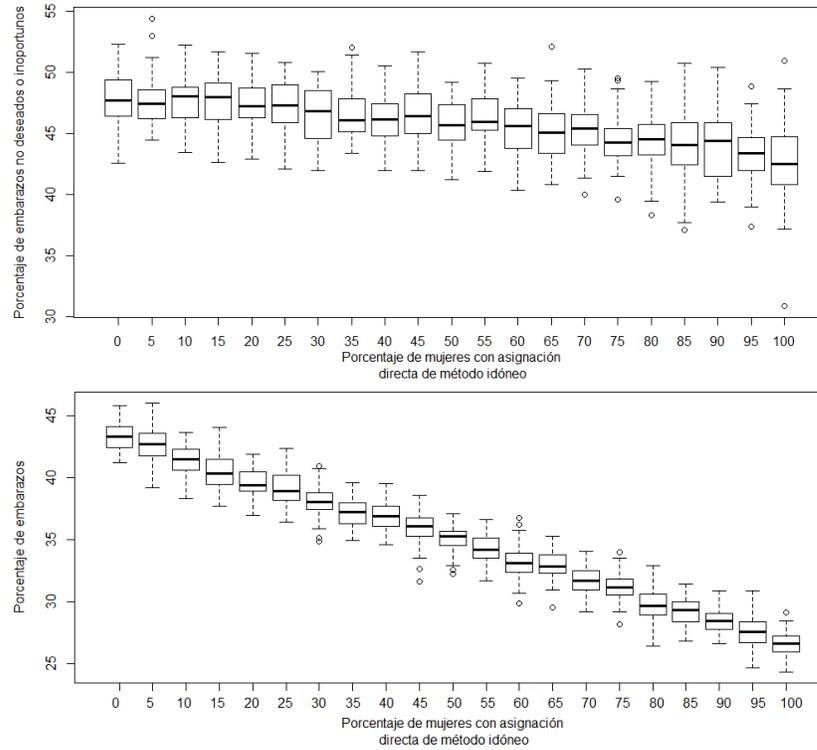


FIGURA 5.15. .

Resultados descriptivos. Escenario 2 - Alternativa 4

Se encontró que el modelo computacional propuesto replica los resultados del sistema. Los resultados de la modelación con base en la información de la ENDS 2005 se ajustan a los resultados encontrados en la ENDS 2010 luego de simular 60 meses. Dentro de las 4 alternativas, la última es la que tiene efectos sobre el porcentaje de uso de métodos anticonceptivos, el porcentaje de embarazos y el porcentaje de embarazos no deseados o inoportunos.

CONCLUSIONES

- La metodología propuesta en este documento combina el análisis de redes con los MBA. Los fenómenos de formación, crecimiento, difusión y aprendizaje en redes aportan en la definición del modelo computacional a través de la definición de la estructura de interacción entre los agentes y la derivación de reglas y comportamientos a partir de las relaciones que se establecen entre ellos. El documento presenta una propuesta metodológica que define las diferentes etapas para abordar el estudio de sistemas complejos en donde la interacción entre los individuos juega un papel relevante dentro del sistema.
- La principal ventaja del análisis de redes es que provee las bases teóricas y prácticas para estudiar fenómenos en donde la interacción entre individuos juega un papel relevante sobre el comportamiento de las variables de interés.
- La principal ventaja de los MBA es que es una metodología que permite estudiar sistemas complejos en los cuales el comportamiento de los individuos, las reglas que se establecen dentro del sistema, la relación entre los individuos y la relación con el ambiente pueden ser incorporadas para mejorar el modelamiento del sistema. Por tratarse de una metodología de micro-simulación, es apropiada para explicar el comportamiento de sistemas en donde la heterogeneidad entre individuos, las interacciones a nivel micro y el comportamiento adaptativo son importantes para estudiar el comportamiento a nivel macro.
- Existe una desventaja común a las dos metodologías y es la demanda de información con alto nivel de especificidad. Sin embargo esta desventaja es inherente a su naturaleza. Por un lado, el análisis de redes requiere como mínimo que se tenga información de los individuos y de las relaciones que se establecen entre ellos. Por su parte, los MBA por tratarse de una metodología cuyo soporte se encuentra en el modelamiento basado en patrones y en la micro-simulación, demanda información a nivel de individuos y a nivel del sistema, para poder abordar de manera adecuada el estudio del fenómeno con base en patrones.
- La implementación computacional de la propuesta metodológica se facilita debido a la existencia de paquetes dentro del ambiente de *R project* que se aproximan bastante a los diferentes procesos y etapas tanto de los MBA como del análisis de redes. Adicionalmente, el programa *NetLogo* provee las herramientas básicas para trabajar con los MBA. Estas metodologías, por tratarse de alternativas para estudiar fenómenos complejos, demandan mayor complejidad en la especificación del modelo y una importante carga computacional.

- Para el análisis de redes se presentan un conjunto de herramientas estadísticas que contribuyen al estudio de al menos 4 aspectos: muestreo y estimación, modelamiento, inferencia estadística en redes y modelamiento y predicción de redes conocidas. En el caso de los MBA, en esta propuesta metodológica se plantea una batería de herramientas estadísticas que contribuyen a la mejora de la calidad y la confiabilidad en las 5 etapas de los MBA: (a) inicialización, formulación de patrones y parametrización, (b) construcción de reglas, (c) modelamiento inverso, (d) análisis de sensibilidad e incertidumbre y (e) análisis de resultados a partir de herramientas descriptivas e inferenciales.
- La propuesta metodológica desarrollada permite abordar el estudio de escenarios de políticas públicas desde un enfoque técnico basado en evidencias. La sinergia entre los MBA y el análisis de redes permite incorporar dentro del análisis diversas herramientas para el estudio de un sistema en donde la interacción entre individuos tiene relevancia. Además, posibilita el planteamiento de escenarios de política pública que explotan la estructura de interacción entre individuos en redes.
- La metodología fue implementada en una aplicación sobre uso de métodos anticonceptivos y ocurrencia de embarazo no deseado o inoportuno en Colombia. Se utilizaron datos de las ENDS 2005 y 2010. La metodología propuesta logra replicar las características y el comportamiento del sistema. El uso de análisis de redes logra replicar el comportamiento sobre cambio de métodos anticonceptivos a través del aprendizaje basado en la efectividad observada por las mujeres de la red. Se formularon 4 alternativas de política pública, encontrando que el aumento en el porcentaje de uso del método anticonceptivo idóneo reduce el porcentaje de embarazos, reduce el porcentaje de embarazos no deseados o inoportunos y aumenta el porcentaje de mujeres que usan métodos anticonceptivos.
- Los análisis realizados en el desarrollo de la aplicación evidencian la potencia teórica y práctica de la metodología y su uso dentro del contexto de la evaluación de escenarios de políticas públicas con un enfoque técnico basado en evidencias.

GLOSARIO

- **Caminata:** Secuencia de nodos que comienza en un nodo i y termina en j , tales que, cada pareja consecutiva está conectada.
- **Ciclo:** Es una caminata que comienza y termina en el mismo nodo.
- **Diámetro:** Distancia más grande que hay entre cualquier par de nodos.
- **Distancia:** Longitud de la caminata más corta que hay entre ellos.
- **Grado:** Número de enlaces que tiene un nodo con otros nodos de la red, lo cual equivale a la cardinalidad de su vecindad.
- **Grafos:** Son representaciones de relaciones binarias que se establecen entre vértices, denominados nodos, y que se representan a través de líneas, también llamadas aristas o enlaces. Los grafos son el objeto de estudio de un campo de las matemáticas conocido como teoría de grafos.
- **Modelo:** Es una representación o abstracción de un problema o sistema, a través de una formulación que involucra variables de entrada y de salida.
- **Simulación:** Experimento computacional para resolver un problema, basado en un modelo de un sistema.
- **Red dirigida:** Es una red en la cual las relaciones pueden establecerse de manera unilateral, es decir que un nodo puede tener una relación con otro sin que éste tenga conexión con el otro.
- **Red no dirigida:** Es una red en la cual las relaciones se establecen de manera recíproca, de forma que una relación de un nodo con otro implica una relación entre ambos.
- **Vecindad:** La vecindad de un nodo es el número de nodos con los cuales está conectado.

Bibliografía

- [1] Andrey Popov, *Genetic algorithms for optimization*, MATLAB, Hamburg, Alemania, 2005.
- [2] Peter Aronow and Cyrus Samii, *Estimating average causal effects under general interference*, 2013, Working paper.
- [3] Robert Axelrod, *Handbook of Computational Economics. Chapter 33: Agent-based Modeling as a Bridge Between Disciplines*, Gerald R. Ford School of Public Policy, University of Michigan, USA, 2006.
- [4] Venkatesh Bala and Sanjeev Goyal, *Learning from Neighbors*, Review of Economics Studies (1998), no. 65, 569–621.
- [5] Osman Balci, *Validation, verification, and testing techniques throughout the life cycle of a simulation study*, Annals of Operations Research **53** (1994), 121–173.
- [6] Sven Banish, Ricardo Lima, and Tanya Araújo, *Agent based models and opinion dynamics as markov chains*, Working Papers. School of Economics and Management. Technical University of Lisbon.
- [7] Albert-László Barabási, *Network Science*, northeastern University, Boston, USA, 2012.
- [8] Frank Bass, “A dynamic model of market share and sales behaviour”, Proceedings, Winter Conference American Marketing Association (1963), 269.
- [9] ———, “A new product growth model for consumer durables ”, Management Science **15** (1969), 215–227.
- [10] Leonardo Bastos and Anthony O’Hagan, *Diagnostics for Gaussian Process Emulators*, Technometrics. American Statistical Association and American Society for Quality **51** (2009), no. 1.
- [11] Jere Behrman, Hans-Peter Kohler, and Susan Cotts, *Social Networks and changes in contraceptive use over time: evidence from a longitudinal study in rural Kenya*, Demography **39** (2002), no. 4, 713–738.
- [12] Jakub Bijak, Jason Hilton, Eric Silverman, and Viet Cao, *Reforging the Wedding Ring: Exploring a Semi-Artificial Model of Population for the United Kingdom with Gaussian process emulators*, Demographic Research **29** (2013), 729–766.
- [13] Francesco Billari, Alexia Prskawetz, Belinda Aparicio, and Thomas Fent, *The “Wedding-Ring”: An agent-based marriage model based on interaction*, Demographic Research **17** (2007), 59–82.

-
- [14] Liliana Blanco, *Probabilidad*, second ed., Universidad Nacional de Colombia, Bogotá, Colombia, 2010.
- [15] Richard Bookstaber, *Using Agent-Based Models for Analyzing Threats Stability*, Office of Financial Research, U.S. Department of the Treasury (2012).
- [16] Christian Breunig, Xun Cao, and Adam Luedtke, *International Migration and its Political Sources: A Network Analysis*, Migration and its Sources. Princeton publisher (2008).
- [17] Nicholas Christakis and James Fowler, *The Spread of Obesity in a Large Social Network over 32 years*, *The New England Journal of Medicine* **4** (2007), no. 357, 370–380.
- [18] Heidi Colleran and Ruth Mace, *Social network- and community-level influences on contraceptive use: evidence from rural Poland*, Royal Society Publishing **282** (2015).
- [19] Morris DeGroot, “*Reaching a consensus*”, *Journal of the American Statistical Association* (1974), no. 69, 118–121.
- [20] Peter DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel, *Persuasion Bias, Social Influence, and Unidimensional Opinions*, *Quarterly Journal of Economics* **118** (2003), no. 2, 909–968.
- [21] S. N. Dorogotsev and J.F. Mendes, *Evolution of networks*, *Advances in Physics* **51** (2002), no. 4, 1079–1187.
- [22] David Easley and Jon Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, first ed., Cambridge University Press, New York, USA, 2010.
- [23] Dean Eckles, Brian Karrer, and Johan Ugander, *Design and analysis of experiments in Networks: reducing bias from interference*, 2014.
- [24] Abdulrahman El-Sayed, Peter Scarborough, Lars Seemann, and Sandro Galea, *Social network analysis and agent-based modeling in social epidemiology*, *Epidemiologic perspectives & innovations* **9** (2012), no. 1.
- [25] Corinna Elsenbroich and Nigel Gilbert, *Modelling Norms*, Springer, 2014.
- [26] Paul Erdos and Alfred Renyi, *On Random Graphs*, *Publicationes Mathematicae* **6** (1959), 290–297.
- [27] Ulf Ewert, Mathias Rohel, and Adelinde Uhrmacher, *Physica Verl*, ch. Consequences of mortality crises in pre-modern european towns: a multiagent-based simulation approach, pp. 175–196, 2003.
- [28] Annalisa Fabretti, *On the problem of calibrating an agent based model of financial markets*, *Journal of Economic Interaction and Coordination* **8** (2013), 277–293.
- [29] Kai-Tai Fang, Runze Li, and Agus Sudjianto, *Design and Modeling for Computer Experiments*, Chapman and Hall/CRC, New York, USA, 2006.
- [30] Karen Foreit, Marco de Castro, and Eliane Duarte, *The impact of Mass Media Advertising on a Voluntary Sterilization Program in Brasil*, *Studies in Family Planning* **20** (1989), no. 2, 107–116.
- [31] Stephanie Forrest, *Genetic Algorithms: Principles of Natural Selection Applied to Computation*, *Science, New Series* **261** (1993), no. 872-878.

-
- [32] James Fowler and Nicholas Christakis, *Dynamic spread of happiness in a large social network: longitudinal analysis of the Framingham Heart Study social network*, British Medical Journal **338** (2009), no. 7685, 23–27.
- [33] Anastasia Gage, *Sexual Activity and Contraceptive Use: The Components of the Decision-making Process*, Studies in Family Planning **29** (1998), no. 2, 154–166.
- [34] Douglas Gale and Shachar Kariv, *Bayesian Learning in Social Networks*, Games and Economic Behaviour **2** (2003), no. 45, 329–346.
- [35] Yongqin Gao, Vince Freech, and Greg Madey, *Conceptual framework for agent-based modeling and simulation*, NAACSOS (2003).
- [36] Sandra Girón, Julio Mateus, and Carlos Castellar, *Análisis costo-efectividad de dos intervenciones para el control de la malaria en el área urbana de Buenaventura, Colombia*, Biomédica - Revista del Instituto Nacional de Salud **26** (2006), no. 3.
- [37] Iván Gómez-Sánchez and Yaira Pardo, *Percepción del uso de anticonceptivos en Bogotá (Colombia) 2009. Estudio Cualitativo.*, Revista Colombiana de Obstetricia y Ginecología **61** (2010), no. 1.
- [38] Lina González, Sofia Claros, Germán Cano, and Héctor Flechas, *Comparación evaluación costo-beneficio Programas nutricionales en Colombia Familias en Acción y Hogares comunitarios*, Archivos de Economía - Dirección Nacional de Planeación (DNP) (2009), no. 353.
- [39] Sanjeev Goyal, “*Learning in Networks*”, Handbook of Social Economics (2012).
- [40] Jakob Grazzini and Matteo Richiardi, “*Estimation of ergodic agent-based models by simulated minimum distance*”, Journal of Economic Dynamics and Control **51** (2015), 148–165.
- [41] Volker Grimm, *Individual-Based Models*, Ecological Models (2008), 1959–1968.
- [42] Anthony O'Hagan, *Curve Fitting and Optimal Design for Prediction*, Journal of the Royal Statistical Society **40** (1978), no. 1, 1–42.
- [43] Kathrin Happe, *Agent-based modelling and sensitivity analysis by experimental design and metamodelling: an application to modelling regional structural change*, Paper prepared for the XIth International Congress of the European Association of Agricultural Economists, The future of Rural Europe in the Global Agri-Food System (Copenhagen, Dinamarca), Agosto 2005.
- [44] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, Springer books, New York, USA, 2001.
- [45] Daniel Heard, *Statistical Inference Utilizing Agent Based Models*, Ph.D. thesis, Duke University, 2014.
- [46] Scott Heckbert, Tim Baynes, and Andrew Reeson, *Agent-based modelling in ecological economics*, Annals of the New York Academy of Sciences **1185** (2010), no. 489.
- [47] Mevin Hooten and Christopher Wikle, *Statistical Agent-Based Models for Discrete Spatio-Temporal Systems*, American Statistical Association **105** (2010), 39–53.
- [48] Jeffrey Horn, Nicholas Nafpliotis, and David Goldberg, *A Niche Pareto Algorithm for Multiobjective Optimization*, Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence., 1994, pp. 82–87.

-
- [49] Matthew Jackson, *Social and Economic Networks*, first ed., Princeton University Press, Princeton, USA, 2008.
- [50] Jan Thiele, *Package 'rnetlogo'*, Georg-August-Universität Göttingen, Göttingen, Alemania, 2014.
- [51] Marco Janssen and Elinor Ostrom, *Empirical based agent-based modelling*, *Ecology and society* **11** (2006), no. 2.
- [52] Harsimran Kaur, *Algorithm used in Intrusion Detection Systems: a review*, *International Journal of Innovative Research in Computer and Communication Engineering* **2** (2014), no. 5.
- [53] Marc Kennedy and Anthony O'Hagan, *Bayesian calibration of computer models*, *Journal of the Royal Statistical Society* **63** (2001), no. 3, 425–464.
- [54] Aln Kirman, *Learning in Agent Based Models*, Groupement de Recherche en Economie (2010), no. 57.
- [55] Jack Kleijnen, Susan Sanchez, Thomas Lucas, and Thomas Cioppa, *A user's guide to the brave new world of designing simulation experiments*, *Journal of Computing* **17** (2005), no. 3, 263–289.
- [56] Eric Kolaczyk, *Statistical Analysis of Network Data. Methods and Models*, first ed., Springer series in Statistics, New York, USA, 2009.
- [57] Eugenio Lahera, *Política y Políticas Públicas*, CEPAL - Serie políticas sociales (2004), no. 95.
- [58] Roberto Leombruni and Matteo Richiardi, *Why are economists sceptical about agent-based simulations*, *Physica A* **355** (2005), 103–109.
- [59] Xitong Li and Lynn Wu, “*Measuring Effects of Observational Learning and Social-Network Word-of-Mouth (WOM) on the Sales of Daily-Deal Vouchers*”, *System Sciences (HICSS)* (2013), 2908–2917.
- [60] Ribin Lye, James Lung, and Siew Cheong, *Understanding agent-based models of financial markets: a bottom-up approach based on order parameters and phase diagrams*, School of Physical and Mathematical Sciences. Nanyang Technological University (2012).
- [61] Monica Magadi and Sian Curtis, *Trends and Determinants of Contraceptive Method Choice in Kenya*, *Studies in Family Planning* **34** (2003), no. 3, 149–159.
- [62] Charles Manski, *Identification of treatment response with social interactions*, *The Econometrics Journal* **16** (2013), no. S1-S23.
- [63] Mark Montgomery and John Casterline, *Social learning, social influence and New Models of Fertility*, *Population and Development Review* **22** (1996), 151–175.
- [64] The An Ngo and Linda See, *Agent-Based Models of Geographical Systems. Chapter 10: Calibration and Validation of Agent-Based Models of Land Cover Change.*, Springer, 2012.
- [65] David Olaleye and Akinrinola Bankole, *The impact of mass media family planning promotion on contraceptive behavior of women in Ghana*, *Population Research and Policy Review* **13** (1994), 161–177.

-
- [66] Matthew Oremland and Reinhard Laubenbacher, *Optimization of Agent-Based Models: Scaling Methods and Heuristic Algorithms*, Journal of Artificial Societies and Social Simulation **2** (2014), no. 17.
- [67] Cesar Paez, *Cuatro décadas de Guerra contra las drogas ilícitas: un balance costo-beneficio*, Centro de Pensamiento Estratégico - Ministerio de Relaciones Exteriores (2012).
- [68] Cyril Piou, Uta Berger, and Volker Grimm, *Proposing an information criterion for individual-based models developed in a pattern-oriented modelling framework*, Ecological Modelling **220** (2009), no. 17, 1957–1967.
- [69] PROFAMILIA, *Encuesta Nacional de Demografía y Salud (ENDS)*, 2005.
- [70] ———, *Encuesta Nacional de Demografía y Salud (ENDS)*, 2010.
- [71] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [72] William Rand, *Machine Learning meets Agent-Based Modeling: When not to go to a bar*, Northwestern Institute on Complex Systems (2006).
- [73] Matteo Richiardi, “*Estimation of Agent-Based Models*”, Material preparado para la conferencia titulada “Calibration, estimation and validation of agent-based models” dentro del curso “An stochastic process approach to agent-based modeling and simulation” ofrecido por The Max Planck Institute for Demographic Research, October 2015.
- [74] Alex Rogers and Peter von Tessin, *Multi-objective calibration for agent based models*, Multi-Objective Calibration for Agent-Based Models, SCS Europe BVBA, 2004, Proceedings 5th Workshop on Agent-Based Simulation.
- [75] Andrea Sottoriva and Simon Tavarã, *Integrating Approximate Bayesian Computation with Complex Agent-Based Models for Cancer Research*, Proceedings in Computational Statistics (2010), 57–66.
- [76] Forrest Stonedahl, *Genetic algorithms for the Exploration of Parameter Spaces in Agent-Based Models*, Ph.D. thesis, Northwestern University, 2011.
- [77] Mujahid Tabassum and Kuruvilla Mathew, *A Genetic Algorithm analysis toward optimization solutions*, International Journal of Digital Information and Wireless Communications **4** (2014), no. 1, 124–142.
- [78] Jan Thiele, *R marries NetLogo: Introduction to the RNetLogo package*, Journal of Statistical Software **53** (2014), no. 2.
- [79] Jan Thiele, Winfried Kurth, and Volker Grimm, *Facilitating parameter estimation and sensitivity analysis of Agent-Based Models: A cookbook using NetLogo and R*, Journal of Artificial Societies and Social Simulation **17** (2014), no. 3.
- [80] Elizabeth Thomson, Maria Winkler-Dowrak, Martin Spielauer, and Alexia Prskawetz, *Union Instability as an Engine of Fertility? A Microsimulation Model for France*, Demography **19** (2012), no. 1, 175–195.
- [81] Stefan Thurner, *Systemic financial risk: agent based models to understand the leverage on national scales and its consequences*, OECD. Multi-Disciplinary Issues. Future Global Shocks (2011).

-
- [82] Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg, *Graph cluster randomization: network exposure to multiple universes*, Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. (New York, USA), 2013, pp. 329–337.
- [83] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow, *The anatomy of the Facebook social graph*, Technical report.
- [84] Uri Wilensky, *Netlogo itself*, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, USA, 1999.
- [85] Thomas Valente and Walter Saba, *Mass Media and Interpersonal Influence in a Reproductive Health Communication Campaign in Bolivia*, Communication Research **25** (1998), no. 1, 96–124.
- [86] Koen van Dam, Igor Nikolic, and Zofia Lukszo, *Agent-Based Modeling of Socio-Technical Systems*, Springer, 2013.
- [87] Evert van Imhoff and Wendy Post, *Microsimulation methods for population projection*, Population. English selection **10** (1998), no. 1, 97–138.
- [88] Stephen Walsh, Ronald Rindfuss, David Guilkey, Philip Page, Brian Frizzelle, Carlos Mena, Diego Quiroga, and Francesco Pizzitutti, *Modeling Population-Environment Interactions in a World Heritage Site: Comparison of Statistical and Agent Based Modeling Approaches to Study Complex Systems*, Tech. report, James McDonell Foundation, 2011.
- [89] Duncan Watts and Steven Strogatz, *Collective dynamics of 'small-world' networks*, Nature **393** (1998), 440–442.
- [90] Michael Widener, Sara Metcalf, and Yaneer Bar-Yam, *Agent-based modeling of policies to improve urban food access for low-income populations*, Applied Geography **40** (2013), 1–10.
- [91] Janusz Wojtusiak, Tobias Warden, and Otthein Herzog, *Machine learning in agent-based stochastic simulation: inferential theory and evaluation in transportation logistics*, Computers and Mathematics with Applications **64** (2012), 3658–3665.
- [92] Ling Xu and Jian-Bo Yang, *Introduction to Multi-Criteria Decision Making and Evidential Reasoning Approach*, Working Paper - Manchester School of Management. University of Manchester Institute of Science and Technology (2001), no. 106.
- [93] Yong Yang, Diez Roux, Amy Auchincloss, Daniel Rodriguez, and Daniel Brown, *A spatial Agent-Based Model for the simulation of adults' daily walking within a city*, American Journal of Preventive Medicine **3** (2011), no. 40, 353–361.
- [94] Lynn Yee and Melissa Simon, *The role of the social network in contraceptive decision-making among young, African American and Latina women*, Adolesc Health. National Institutes of Health **47** (2010), no. 4, 374–380.
- [95] Gordan Zitković, *Introduction to Stochastic Processes*, Department of Mathematics. The University of Texas, Texas, USA, 2010.