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

MECANISMOS DE ADMISIÓN ESCOLAR EN CHILE

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN
GESTIÓN DE OPERACIONES
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

NATALIE EPSTEIN ROSENBERG

PROFESOR GUÍA:
JOSÉ RAFAEL CORREA HAEUSSLER

MIEMBROS DE LA COMISIÓN:
JUAN ESCOBAR CASTRO
IGNACIO RÍOS URIBE

SANTIAGO DE CHILE
2017

RESUMEN DE LA MEMORIA PARA OPTAR
AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES
POR: NATALIE EPSTEIN ROSENBERG
FECHA: 2017
PROF. GUÍA: JOSÉ RAFAEL CORREA HAEUSSLER

MECANISMOS DE ADMISIÓN ESCOLAR EN CHILE

El 29 de mayo del 2015 se promulgó la Ley de Inclusión Escolar que, entre otras cosas, regula la admisión de los estudiantes y elimina la selección arbitraria en establecimientos educacionales que reciben aportes del Estado, con el fin de dar a las familias igualdad de oportunidades para ingresar a los colegios de su preferencia.

Una forma de resolver el fin a la selección escolar es a través de un mecanismo centralizado que asigne a los alumnos a colegios en base a sus preferencias, bajo ciertos criterios de prioridad y desempates aleatorios entre los postulantes. Esto puede resolverse mediante diferentes algoritmos que entregan asignaciones distintas en cuanto al resultado y sus propiedades.

En este trabajo explicamos el desarrollo del algoritmo de asignación escolar en Chile, las razones que nos llevaron a elegir el algoritmo de Aceptación Diferida (DA) con sus propiedades y los resultados obtenidos en el proceso de postulación para el año escolar 2017 en la Región de Magallanes. Dentro de los resultados destaca que un 86.4 % de los alumnos quedaron asignados en alguna de sus preferencias y un 58 % en la primera preferencia declarada. Los alumnos que no quedaron asignados en sus preferencias fueron asignados por distancia a su colegio más cercano con cupos disponibles, o bien al colegio donde estaban matriculados antes de postular.

Usando los datos obtenidos, realizamos simulaciones para comparar alternativas de reglas de rompimiento de empates, que pueden ser una lotería única para cada alumno, o una diferente por cada colegio. Al comparar estas dos alternativas, los resultados corroboran lo que expone la literatura, en el sentido de que no hay dominancia estocástica entre los resultados con ambos tipos de loterías, ya que con una lotería única más alumnos quedan en sus primeras preferencias, pero también un mayor porcentaje no es asignado a ninguna de sus preferencias. En el caso de Chile, la Ley de Inclusión Escolar establece que cada colegio puede hacer su propia lotería, por lo que se implementó de esta manera.

Finalmente, bajo la hipótesis de que las preferencias de los alumnos en la elección de colegios están correlacionadas entre sí, desarrollamos un modelo probabilístico que permite simular dichas preferencias. Este modelo se basa en el problema de la Urnas de Polya y modelos de Preferential Attachment, en los cuales los elementos del proceso, en este caso las preferencias, siguen una distribución de ley de potencia. Esto, con el objetivo de poder simular instancias de mayor tamaño y anticiparse a lo que ocurrirá en los próximos años cuando el mecanismo se extienda a todas las regiones del país.

Agradecimientos

A mi familia, en especial a mis papás y mis hermanos, por su apoyo incondicional, su preocupación y todo el cariño.

A mi profesor guía José Correa, por invitarme a participar en el proyecto y acompañarme durante esta importante etapa de mi vida. A los profesores de mi comisión Juan Escobar e Ignacio Ríos por su dedicación y apoyo durante el proyecto y la realización de esta tesis.

A mis amigas y amigos del colegio, de la universidad y de la oficina en especial, por el constante interés y apoyo en mi tesis, y por todos los buenos momentos que compartimos.

A los funcionarios del DII y del MGO por su dedicación y preocupación cada día.

Al Núcleo Milenio Información y Coordinación en Redes por el financiamiento otorgado para llevar a cabo este trabajo.

Tabla de Contenido

1. Introducción **1**
1.1. Objetivos 2
1.2. Resultados 2
1.3. Estructura del trabajo 3

2. Problema de Asignación Escolar **4**
2.1. Descripción del problema 4
2.1.1. Propiedades 5
2.1.2. Algoritmos 6
2.2. Aceptación Diferida 7
2.3. Experiencia internacional 9

3. Asignación escolar en Chile **11**
3.1. Diseño del algoritmo 11
3.1.1. Características específicas 11
3.1.2. Desarrollo del algoritmo 13
3.2. Análisis de propiedades 16
3.2.1. A prueba de estrategias 17
3.2.2. Postulación en bloque 21
3.2.3. Estabilidad 23
3.2.4. Matrícula asegurada 25

4. Implementación Región de Magallanes **26**
4.1. Descripción de la instancia 26
4.2. Resultados computacionales 27
4.3. Análisis de resultados 28
4.3.1. Prioritarios 28
4.3.2. Reglas de rompimiento de empates 30

5. Modelo de preferencias **33**
5.1. Preferencias en Magallanes 33
5.2. Preferential Attachment y Polya's Urn Problem 35
5.3. Ajuste del modelo para Magallanes 37
5.3.1. Modelo con niveles agregados 38
5.3.2. Modelo Pre Kínder 44
5.4. Modelo con colegios fijos 45

6. Conclusiones **48**
6.1. Conclusiones generales 48
6.2. Trabajo futuro 48

Bibliografía **50**

Índice de Tablas

2.1. Preferencias y prioridades 5

3.1. Preferencias y prioridades nivel 9 14
3.2. Preferencias y prioridades nivel 1 14
3.3. Preferencias y prioridades actualizadas nivel 9 15
3.4. Preferencias y prioridades reales nivel 9 18
3.5. Preferencias y prioridades reales nivel 1 18
3.6. Preferencias y prioridades alteradas nivel 9 18
3.7. Preferencias y prioridades actualizadas nivel 1 19
3.8. Preferencias y prioridades reales nivel 9 19
3.9. Preferencias y prioridades reales nivel 1 20
3.10. Preferencias y prioridades actualizadas nivel 1 20
3.11. Preferencias y prioridades alteradas nivel 9 20
3.12. Preferencias y prioridades actualizadas nivel 1 21
3.13. Preferencias y prioridades reales nivel 9 22
3.14. Preferencias y prioridades reales nivel 1 22
3.15. Preferencias y prioridades actualizadas nivel 1 22
3.16. Preferencias y prioridades reales nivel 9 24
3.17. Preferencias y prioridades reales nivel 1 24
3.18. Preferencias y prioridades actualizadas nivel 1 25

4.1. Alumnos y colegios por nivel 26
4.2. Cumplimiento de la cuota de prioritarios por nivel 30
4.3. Comparación de reglas de desempate 32

5.1. Preferencias por nivel 34

Índice de Ilustraciones

3.1. Extracto de la Ley de Inclusión Escolar 13

4.1. Porcentaje acumulado de alumnos asignados por preferencia 27
4.2. Resumen de la asignación por nivel, Punta Arenas 28
4.3. Porcentaje de alumnos asignados por preferencia, nivel pre-kínder, Punta Arenas 29
4.4. Porcentaje acumulado de alumnos asignados por preferencia 32

5.1. Preferencias por RBD 34
5.2. Preferencias por RBD, nivel pre-kínder 35
5.3. Postulantes por RBD simulación con 3436 alumnos y $p_0 = 1{,}83\,\%$ 38
5.4. Postulantes por RBD simulación 1674 alumnos, $c = 43$ y $p_0 = 3{,}7\,\%$ 42
5.5. Postulantes por RBD simulación con 3436 alumnos, $c = 43$ y $p_0 = 1{,}83\,\%$ . . 43
5.6. Postulantes por RBD simulación con 3436 alumnos, $c = 95$ y $p_0 = 1{,}83\,\%$ . . 43
5.7. Postulantes por RBD simulación con 1195 alumnos y $p_0 = 3{,}51\,\%$ 44
5.8. Postulantes por RBD simulación con 586 alumnos, $c = 31$ y $p_0 = 7{,}2\,\%$ . . . 45
5.9. Postulantes por RBD simulación con 1195 alumnos, $c = 31$ y $p_0 = 3{,}51\,\%$ . . 45
5.10. Postulantes por RBD simulación con 1195 alumnos, $c = 64$ y $p_0 = 3{,}51\,\%$ . . 46

Capítulo 1

Introducción

El 29 de mayo del 2015 se promulgó la Ley de Inclusión Escolar que regula la admisión de los estudiantes, elimina el financiamiento compartido y prohíbe el lucro en establecimientos educacionales que reciben aportes del Estado. Estas medidas buscan entregar a las familias una educación de calidad y la posibilidad de elegir libremente el proyecto educativo de su preferencia.

Para llevar a cabo la implementación de esta ley y en específico solucionar el problema del fin de la selección, el Ministerio de Educación (MINEDUC) decidió desarrollar un mecanismo centralizado, a través de una plataforma web en la cual las familias postulan a sus hijos a los colegios de su preferencia y un algoritmo que asigna a los postulantes a dichos colegios. Esta asignación se realiza en base a la disponibilidad de cupos y en caso de que no los hubiera, con un desempate aleatorio entre los postulantes, logrando así una mayor equidad y transparencia en el proceso.

En esta tesis abordamos, en primer lugar, el desarrollo de un algoritmo de asignación escolar, diseñado en base a los requerimientos reales para el caso chileno, la aplicación del algoritmo y los resultados obtenidos en el proceso de admisión escolar 2017 en la Región de Magallanes. Y, en segundo lugar, el modelamiento de las preferencias de las familias para realizar simulaciones a mayor escala que representen lo que ocurrirá en los procesos de admisión durante los próximos años en todo Chile.

El algoritmo de asignación escolar se desarrolló en base al algoritmo de Aceptación Diferida (DA) propuesto por Gale and Shapley [1962], que actualmente es utilizado y aplicado en diferentes ámbitos para la resolución de problemas de políticas públicas. El concepto de un sistema centralizado de asignación escolar no es nuevo, ya que ha sido implementado con algunas adaptaciones en distintas ciudades alrededor del mundo. Entre los principales ejemplos destacan:

- Nueva York, donde Abdulkadiroğlu et al. [2009] rediseñan y analizan el cambio desde un proceso descentralizado y estrategizable por un sistema centralizado que asigna más de 90.000 alumnos a enseñanza media cada año.
- Boston, ciudad en la que comenzaron con un algoritmo de asignación conocido como

el "Mecanismo de Boston", pero luego lo cambiaron a DA dado que, como muestran Abdulkadiroglu et al. [2006], bajo el primer algoritmo las familias tenían incentivos a no declarar sus verdaderas preferencias.

- Ámsterdam, donde al comienzo utilizaron DA con una regla de rompimiento de empates múltiple y al año siguiente la cambiaron a una única lotería para mayor eficiencia del algoritmo, como muestra Gautier et al. [2016].

1.1. Objetivos

El objetivo general del trabajo es desarrollar el algoritmo de asignación escolar que utilizamos para el proceso de admisión del año 2017 en la Región de Magallanes. Entre los objetivos específicos destacan desarrollar un algoritmo de asignación escolar que cumpla con los requisitos de la Ley de Inclusión Escolar y del MINEDUC, demostrar en qué casos el algoritmo implementado cumple con las propiedades deseables de una asignación y en cuáles no, elaborar estadísticas descriptivas de los resultados de la implementación en la Región de Magallanes y simular las preferencias de los apoderados con modelos probabilísticos.

El algoritmo centralizado de asignación escolar se utilizará para los procesos de admisión de todo Chile, de manera gradual, según un cronograma elaborado por el MINEDUC, que incorpora año tras año más regiones y más niveles por región.

Para efectos de este trabajo utilizamos los datos de la Región de Magallanes para el año escolar 2017 en Pre-Kínder, Kínder, Primero Básico, Séptimo Básico y Primero Medio, que son los niveles de inicio de ciclo y en que los colegios generalmente empiezan a funcionar.

1.2. Resultados

En primer lugar, estudiamos las propiedades del algoritmo diseñado, demostrando las condiciones bajo las cuales el algoritmo es a prueba de estrategias y estable. Además probamos cómo influye el hecho de hacer postulación en bloque para las familias.

En general, probamos que el algoritmo es a prueba de estrategias, excepto en el caso en que alguna de las familias que realizan postulación en bloque quede en colegios diferentes. En este caso, las familias podrían lograr que ambos hermanos queden asignados en el mismo colegio alterando sus preferencias. En el caso de la estabilidad ocurre algo similar, ya que el algoritmo es estable, excepto en parejas de hermanos que realizan postulación en bloque y quedan en colegios diferentes, en cuyo caso el algoritmo no es necesariamente estable. En cuanto a la postulación en bloque, demostramos que es la estrategia óptima para las familias que tienen como primera preferencia que ambos hermanos queden en el mismo colegio.

Luego analizamos los resultados obtenidos al utilizar el algoritmo en la Región de Magallanes. Dentro de los principales resultados destacamos que, de los 3357 alumnos con postulaciones válidas, un 58 % de los alumnos queda en su primera preferencia y un 86.4 % queda en alguna de sus preferencias. Visto por nivel, en pre-kínder más de un 60 % de los alumnos queda en su primera preferencia y cerca de un 10 % de los 1195 alumnos queda asignado por distancia.

Además, a partir de los resultados obtenidos con múltiples loterías (MTB) y realizando simulaciones utilizando una lotería única (STB) observamos que ninguna de las dos reglas de rompimiento de empates domina estocásticamente a la otra, tal como muestra Ashlagi and Nikzad [2015], siendo STB la que asigna más alumnos a las primeras preferencias, pero deja más alumnos sin asignación.

Finalmente, y bajo la hipótesis que las preferencias de los alumnos están correlacionadas, encontramos que adaptando el problema de las Urnas de Polya al caso de alumnos y colegios, podemos simular las postulaciones de los alumnos. Más aun, demostramos que la cantidad de alumnos que postulan a cada colegio sigue una distribución de ley de potencia de parámetro $\left(1 + \frac{1+c \cdot p_0}{1-p_0}\right)$, donde c y p_0 son constantes positivas.

1.3. Estructura del trabajo

Esta tesis está estructurada de la siguiente forma: en el capítulo 2 presentamos la descripción del problema de asignación escolar, la notación, propiedades y algoritmos utilizados. Luego explicamos el algoritmo de Aceptación Diferida (DA) y finalmente mostramos resultados teóricos y prácticos conocidos.

En el capítulo 3 detallamos los elementos que hacen único el problema en Chile y describimos la implementación del algoritmo que finalmente resuelve dicho problema. Luego presentamos resultados para demostrar cuándo las propiedades del algoritmo se pierden y cuándo no, ante las modificaciones realizadas a DA.

En el capítulo 4 presentamos los datos y características de la Región de Magallanes, los resultados obtenidos tras la ejecución del algoritmo y análisis de cambios en el desarrollo de algunos elementos del algoritmo para evaluar implementaciones alternativas para los próximos años.

En el capítulo 5 analizamos las preferencias de las familias, estudiamos modelos conocidos como Preferencial Attachment y Polya's Urn Problem y proponemos un modelo probabilístico que permite replicar las postulaciones de las familias y simular nuevas instancias.

Finalmente, en el capítulo 6 presentamos las conclusiones relacionadas al trabajo realizado y algunas propuestas para continuar el desarrollo del algoritmo y extender los resultados.

Capítulo 2

Problema de Asignación Escolar

2.1. Descripción del problema

Un problema de asignación escolar considera un conjunto S de alumnos y un conjunto C de colegios donde cada colegio $c \in C$ tiene una determinada capacidad $q_c \in \mathbb{N}$ para recibir alumnos. Una asignación μ consiste en una relación entre los agentes de ambos conjuntos de modo que cada alumno $s \in S$ es asignado a lo más a un colegio o se mantiene no asignado, y a cada colegio $c \in C$ se le asigna una cantidad de alumnos menor o igual a su capacidad.

Se denomina $\mu(s)$ el colegio al cual fue asignado el alumno s bajo la asignación μ, que puede en efecto ser vacío, y de manera equivalente $\mu(c)$ es el subconjunto de alumnos asignados al colegio c, el cual también puede ser vacío.

Para encontrar la asignación, cada alumno s declara preferencias estrictas sobre los colegios. El ranking del colegio c para el alumno s es el número de colegios que el alumno débilmente prefiere ante c. El alumno incluye dentro de sus preferencias el no quedar asignado y no tiene necesidad de agregar a su lista de preferencias colegios donde no quiere ser asignado.

Los colegios por su parte son indiferentes entre los alumnos, es decir, no tienen preferencias sobre los mismos, pero deben realizar desempates aleatorios entre los postulantes para poder encontrar una asignación, por lo cual se hablará de las prioridades de los alumnos en los colegios.

A partir de estas definiciones, hay variados algoritmos que permiten encontrar asignaciones y con ello solucionar el problema. Sin embargo, dependiendo del algoritmo que se utilice se obtienen asignaciones que satisfacen distintas propiedades.

2.1.1. Propiedades

Estabilidad

Para que una asignación μ sea estable no pueden presentarse simultáneamente las siguientes dos condiciones:

- El alumno s prefiere el colegio c a su asignación actual $\mu(s)$.
- El colegio c o bien tiene algún asiento disponible, es decir, $|\mu(c)| < q_c$, o bien tiene en su asignación $\mu(c)$ algún alumno s' con peor prioridad que s.

Si estas condiciones se dieran simultáneamente para algún par (s, c), entonces la asignación no sería estable pues ambas partes acordarían realizar el cambio y mejorar su estado actual.

Eficiencia

La eficiencia se refiere a que, dada una asignación, no hay grupos de alumnos que puedan intercambiar su asignación haciendo que la nueva asignación mejore al menos a alguno de ellos y no empeore a ningún alumno de S. Entendemos por mejorar el cambiarse a un colegio con mejor ranking para el alumno en su lista de preferencias.

La eficiencia se entiende desde el punto de vista de los estudiantes, para quienes se busca una asignación óptima, ya que los colegios son indiferentes ante los postulantes y realizan desempates aleatorios para generar sus listas de preferencias.

En los casos en que la asignación obtenida no es eficiente, hay mecanismos para recuperar la eficiencia. Por ejemplo, existen algoritmos que permiten encontrar ciclos (grupos de alumnos) que al intercambiarse de colegios mejoran la asignación de algunos de ellos y no perjudican a otros. Sin embargo, esto tiene costos como la pérdida de estabilidad o la estrategización del algoritmo.

En el siguiente ejemplo, se ilustra el trade-off que existe entre la eficiencia y la estabilidad de una asignación, quedando en evidencia por qué, en ciertos casos, no se puede tener ambas propiedades.

Supongamos que $S = \{s_1, s_2, s_3\}$ y $C = \{c_1, c_2, c_3\}$. Las preferencias y prioridades se presentan en la tabla 2.1 donde en cada casilla el número de la izquierda representa la preferencia del alumno por el colegio y el número de la derecha representa la prioridad del alumno en el colegio.

	c_1	c_2	c_3
s_1	2,1	1,2	3,2
s_2	1,3	2,1	3,1
s_3	1,2	2,3	3,3

Tabla 2.1: Preferencias y prioridades

En este caso, una asignación estable es:

$$\mu_{\text{estable}} = \begin{pmatrix} c_1 & c_2 & c_3 \\ s_1 & s_2 & s_3 \end{pmatrix}.$$

Pero podemos ver que hay al menos una asignación en la que los alumnos quedarían mejor, es decir, un asignación más eficiente, la que se obtiene cuando s_1 y s_2 intercambian sus colegios. Esta sería:

$$\mu_{eficiente} = \begin{pmatrix} c_1 & c_2 & c_3 \\ s_2 & s_1 & s_3 \end{pmatrix}.$$

Sin embargo, si analizamos la asignación eficiente, podemos ver que s_3 prefiere el colegio c_1 que su actual asignación c_3, y además c_1 prefiere a s_3 sobre s_2, que es el alumno que actualmente le fue asignado. Por ende, c_1 y s_3 podrían negociar para quedar juntos. Esto ilustra la pérdida de estabilidad al mejorar la eficiencia.

Además, en el caso de asignar a s_3 en c_1 se estaría perjudicando a s_2 quien quedaría en c_2 que es peor para él que el caso estable inicial.

A prueba de estrategias

Este concepto se refiere a que la mejor estrategia que tienen los alumnos para quedar en colegios de su preferencia es declarar sus verdaderas preferencias al postular.

Esta es una propiedad deseable del algoritmo ya que incentiva a los alumnos a postular declarando sus verdaderas preferencias y hace que su postulación sea independiente de las creencias que se tengan sobre los demás postulantes. Esto permite elaborar estadísticas coherentes en el sentido de que si calculamos la cantidad de alumnos asignados a cada preferencia, estas preferencias son reales, lo que permite a largo plazo realizar políticas públicas y estudios de factores que influyen en las postulaciones.

2.1.2. Algoritmos

Hay diversos algoritmos para encontrar asignaciones y cada uno logra mantener algunas de las propiedades descritas anteriormente. Sin embargo, ninguno logra necesariamente cumplir con las tres propiedades simultáneamente como muestra Roth [1982] y como se evidencia en el ejemplo mostrado anteriormente. Es por esto que para decidir el algoritmo a utilizar, es necesario definir las propiedades que queremos mantener.

Los algoritmos utilizados para resolver problemas de asignación que son a prueba de estrategias son: Aceptación Diferida (DA) y Top Trading Cycles (TTC). El primero es estable y el segundo eficiente. Abdulkadiroglu and Sönmez [2003] proponen estos dos algoritmos y muestran que DA adaptado al caso de asignación escolar entrega una asignación que si bien no es eficiente, cumple con que no hay ningún alumno que pueda mejorar su asignación sin perjudicar a otro estudiante o perdiendo estabilidad.

Específicamente, el algoritmo DA entrega una asignación Pareto-óptima dentro de las asignaciones estables. En este sentido, Abdulkadiroğlu et al. [2009] muestran que ante cualquier regla de rompimiento de empates, no hay ningún algoritmo que sea a prueba de estrategias y que entregue una asignación que domine la que se obtiene utilizando el algoritmo DA bajo dicha regla de rompimiento de empates.

A modo general, hay dos formas de realizar el rompimiento aleatorio de empates:

- Una única lotería (single tie-breaking o STB) en la que los alumnos obtienen el mismo orden aleatorio en todos los colegios.
- Una lotería por cada colegio (multiple tie-breaking o MTB) en que cada colegio ordena aleatoriamente a sus postulantes y los órdenes de cada colegio no están relacionados.

Dependiendo del algoritmo y del tipo de lotería que se utilice, se pueden obtener distintas asignaciones. Más detalles con respecto a las loterías se presentan en la sección 4.3.2.

Dado lo anterior, decidimos utilizar el algoritmo de Aceptación Diferida ya que es a prueba de estrategias y estable.

2.2. Aceptación Diferida

El algoritmo de Aceptación Diferida fue propuesto inicialmente por Gale and Shapley [1962], quienes resolvieron con este método el problema conocido como emparejamiento de hombres con mujeres.

En el problema original de emparejamiento, los hombres invitan a salir a la mujer que prefieren en primer lugar, y las mujeres que reciben propuestas aceptan tentativamente al hombre que más les gusta y rechazan al resto.

A modo general, en un paso k:

- Todos los hombres que fueron rechazados en el paso anterior se proponen a la siguiente mujer en su lista de preferencias.
- Las mujeres que tengan una o más propuestas eligen al hombre de su preferencia de entre las propuestas que recibieron y el hombre que tenían aceptado tentativamente del paso anterior y lo aceptan tentativamente, rechazando a todos los demás.

Esto continúa hasta que todos los hombres son asignados o los hombres que quedan sin asignar no tienen más mujeres en su lista de preferencias.

Este modelo fue extendido a lo que se conoce como "many-to-one matching", en el cual una de las partes tiene capacidad de recibir a más de uno de los que se proponen. Este es exactamente el caso del problema de asignación escolar, en el cual los colegios tienen una capacidad determinada de alumnos a los que pueden recibir, mientras que los alumnos son asignados a lo más a un establecimiento.

En este caso, el algoritmo de DA sigue funcionando, entrega una solución estable y óptima para los postulantes a los colegios. Aun más, el algoritmo sigue siendo a prueba de estrategias para los alumnos.

En el caso de la asignación escolar, en el Algoritmo 1 mostramos el pseudo-código que se utiliza para resolver el problema.

Algoritmo 1 Many-to-one matching

```
while hay algún estudiante s no asignado y su lista de preferencias es no vacía do
    c := primer colegio en la lista de s al cual s no ha postulado;
    if c tiene asientos disponibles then
        se asigna s a c {tentativamente}
    else
        if c prefiere a s por sobre el peor alumno tentativamente asignado s' then
            se asigna s a c tentativamente y s' queda libre
        else
            c rechaza a s {y s sigue libre}
        end if
    end if
end while
return  la asignación estable de los colegios con sus estudiantes
```

Los principales resultados de la literatura en relación al problema de selección escolar se enfocan en demostrar que el algoritmo presentado entrega una asignación óptima, estable y a prueba de estrategias, todo desde el punto de vista de las familias que postulan y dadas las reglas de desempate utilizadas.

Gale and Shapley [1962] demuestran que en el problema de emparejamiento siempre existe una asignación estable. Para los casos de "many-to-one matching", demuestran que bajo DA los postulantes reciben una asignación igual o mejor que cualquier otra asignación estable.

Luego, Roth [1982] muestra que la estrategia óptima de los agentes que proponen en el algoritmo, que en este caso serían los estudiantes (S), es declarar sus verdaderas preferencias.

Por otro lado, Roth [1984] prueba que no existe un algoritmo de asignación estable que haga que la estrategia dominante de los agentes que reciben las postulaciones, en este caso los colegios (C), sea declarar sus verdaderas preferencias. Sin embargo, el proceso chileno no se ve afectado porque los colegios son indiferentes a los alumnos y realizan desempates aleatorios para crear las listas de prioridades, por lo que el algoritmo no es estrategizable por parte de los colegios.

En los casos en que los colegios no tienen preferencias estrictas sobre los alumnos, para poder implementar los algoritmos se realizan desempates aleatorios entre los postulantes. Erdil and Ergin [2008] muestran que utilizar desempates aleatorios genera una pérdida de bienestar en los alumnos, y más aun, muestran que no hay algoritmos a prueba de estrategias que cumplan con que no hay alumnos que puedan mejorar su asignación sin perjudicar a otro alumno o perdiendo estabilidad.

2.3. Experiencia internacional

El concepto de asignación escolar centralizada no es nuevo. Diversas ciudades alrededor del mundo la utilizan y cada caso tiene características e incluso implementaciones diferentes. Aun así, en la mayor parte de los casos se usa como base el algoritmo de Aceptación Diferida, el cual se adapta para cumplir requerimientos específicos de cada ciudad, como criterios de prioridad, división por zonas e incluso el tipo de reglas de rompimiento de empates a utilizar.

A continuación presentamos algunos de los casos más relevantes de la literatura en los que se utiliza el algoritmo de DA para realizar la asignación escolar, y presentamos las características que distinguen cada caso.

Boston, EE.UU.

Es uno de los casos más conocidos ya que en un comienzo, cuando decidieron realizar una asignación centralizada, implementaron un algoritmo de asignación, conocido como el "Mecanismo de Boston", que no era a prueba de estrategias ya que los postulantes tenían incentivos a declarar colegios que les gustaran pero no fueran excesivamente demandados para maximizar la probabilidad de quedar asignados. Esto se daba porque los alumnos eran aceptados de manera definitiva en un colegio en cada ronda del algoritmo.

Dado lo anterior, en el año 2005 se decidió cambiar el algoritmo a uno que fuera a prueba de estrategias y se optó por utilizar DA. Este trabajo fue desarrollado por Abdulkadiroglu et al. [2006] y en él explican las razones para cambiar el algoritmo a uno que fuera a prueba de estrategias.

Además, el caso de Boston se caracteriza por dar prioridad a alumnos que tienen hermanos en un determinado colegio y por introducir el concepto de zonas, dando cierta prioridad a los alumnos que postulan a colegios que están en sus barrios o a una distancia apta para caminar.

Cabe señalar que en el caso de Boston, la asignación se realiza en 3 niveles de ingreso a los colegios. Cada uno de estos niveles consta de 3.000 a 4.000 alumnos.

Nueva York, EE.UU.

Es un caso conocido por su gran tamaño, dado que cuenta con más de 90.000 alumnos que deben ser asignados a colegios de enseñanza media. Abdulkadiroğlu et al. [2009] estudian el problema desde el año 2003 cuando comienzan con el rediseño del proceso de asignación basándose en DA.

El proceso de asignación originalmente consistía en 3 rondas descentralizadas, con listas de espera, e incluso permitía a los colegios darle mayor prioridad a alumnos que los declaraban entre sus primeras preferencias, lo que hacía que el algoritmo no fuera a prueba de estrategias para los postulantes.

Ámsterdam, Holanda

Se caracteriza por haber cambiado la regla de rompimiento de empates tras un año de implementación. El trabajo se llevó a cabo por Gautier et al. [2016] y muestra que el primer año se utilizaron múltiples loterías para romper los empates, lo que implicó ineficiencias en la asignación desde el punto de vista de los estudiantes.

Esto llevó a que al siguiente año implementaran el algoritmo con una única lotería, disminuyendo los ciclos de mejora y con ello los reclamos de las familias.

Capítulo 3

Asignación escolar en Chile

3.1. Diseño del algoritmo

El objetivo del algoritmo desarrollado es ser implementado en Chile, por lo que debe cumplir con las características específicas que se indican en la Ley de Inclusión Escolar y los Reglamentos relacionados a este proceso.

En esta sección, detallamos los elementos específicos que debe tener el algoritmo de asignación escolar. Asimismo explicamos la manera de implementar cada uno de estos y demostramos en qué casos el algoritmo sigue cumpliendo las propiedades explicadas en la sección anterior.

3.1.1. Características específicas

A continuación presentamos y explicamos los requerimientos establecidos por el MINEDUC para el proceso de admisión escolar 2017:

- Múltiples loterías (MTB): cada colegio puede realizar una lotería independiente entre sus postulantes, por lo que la regla de rompimiento de empates con que funcionará el algoritmo será de múltiples loterías.

  Este elemento no tiene implicancias directas en la elaboración del algoritmo. Sin embargo, influye de manera significativa en los resultados de la asignación que se obtiene.

  Abordamos esto con más detalle en la sección 4.3, donde analizamos los resultados obtenidos por Ashlagi et al. [2015] quienes demuestran que al utilizar múltiples loterías, los alumnos quedan, en promedio, asignados a preferencias más bajas, pero un porcentaje menor de ellos quedan sin asignar, en comparación con una lotería única (STB). Además vemos que Abdulkadiroğlu et al. [2009] comparan los resultados obtenidos en Nueva York utilizando ambos tipos de loterías.

- Grupos o criterios de prioridad: al postular a un colegio, cada alumno es identificado como perteneciente a un grupo o criterio de prioridad, dentro del cual los alumnos son rankeados aleatoriamente.

Los criterios son los siguientes:

1. Alumnos que tienen un hermano matriculado en el colegio al que postulan y los que tienen un hermano mayor participando del proceso el cual ya fue asignado a dicho colegio durante el proceso de asignación.

2. Una cuota de alumnos prioritarios, que se calcula como un 15 % de los cupos del colegio menos los prioritarios actualmente matriculados. El resto de los alumnos prioritarios que no participan de la cuota participan como alumnos regulares.

3. Hijos o familiares directos de un funcionario del establecimiento.

4. Haber sido alumno del colegio anteriormente y no haber sido expulsado del mismo.

5. Todos los demás alumnos.

Esto lo podemos ver directamente en la Ley de Inclusión Escolar en la figura 3.1.

Cabe señalar que estos elementos no influyen directamente en el algoritmo, ya que este ordenamiento es el input del algoritmo.

- Prioridad a hermanos del mismo nivel: hermanos que postulan al mismo nivel tienen prioridad tal como si fueran hermanos de niveles diferentes. Para esto se les asigna la misma prioridad en cada colegio al que ambos postulan. En caso de que admitir a ambos alumnos haga que se exceda la capacidad del colegio, entonces uno de los hermanos no podrá ser asignado en dicho colegio y los hermanos quedarán en colegios diferentes.

- Postulación en bloque: dos o más hermanos pueden seleccionar la opción de postular en bloque, lo cual es equivalente a declarar que su primera preferencia es quedar ambos en el mismo colegio, aunque no sea el primero de su lista de preferencias.

- Matrícula asegurada (M.A.): a los alumnos matriculados en un establecimiento, que participan del proceso para cambiarse de colegio, se les asegura la continuidad en su establecimiento de origen en caso de no quedar asignado a algún colegio de los que postuló. Cabe señalar que los alumnos que aplican al colegio por primera vez no pueden tener matrícula asegurada ya que no provienen de ningún colegio.
  Si un colegio declara menos cupos que alumnos (por ejemplo porque quiere reducir el tamaño de sus cursos) los alumnos que actualmente están matriculados en dicho establecimiento y que querían cambiarse de colegio pero no pudieron tienen asegurado su cupo aunque las vacantes declaradas por el colegio no sean suficientes.

- Asignación por distancia (Dist.): los alumnos que no queden asignados a ninguno de los colegios de su lista de preferencias y que no tengan matrícula asegurada en algún colegio, serán asignados al colegio más cercano a su hogar con cupos disponibles y sin copago.

Artículo 7º ter.- La etapa de admisión propiamente tal será realizada por los establecimientos educacionales.

Todos los estudiantes que postulen a un establecimiento educacional deberán ser admitidos, en caso de que los cupos disponibles sean suficientes en relación al número de postulaciones.

Sólo en los casos de que los cupos disponibles sean menores al número de postulantes, los establecimientos educacionales deberán aplicar un procedimiento de admisión aleatorio definido por éstos, de entre los mecanismos que ponga a su disposición el Ministerio de Educación, que deberán ser objetivos y transparentes. Dicho procedimiento de admisión deberá considerar los siguientes criterios de prioridad en orden sucesivo, para su incorporación directa a la lista de admisión del establecimiento:

a) Existencia de hermanas o hermanos que postulen o se encuentren matriculados en el mismo establecimiento.

b) Incorporación del 15% de estudiantes prioritarios, de conformidad al artículo 6º, letra a) ter.

c) La condición de hijo o hija de un profesor o profesora, asistente de la educación, manipulador o manipuladora de alimentos o cualquier otro trabajador o trabajadora que preste servicios permanentes en el establecimiento educacional.

d) La circunstancia de haber estado matriculado anteriormente en el establecimiento educacional al que se postula, salvo que el postulante hubiere sido expulsado con anterioridad del mismo.

Si aplicando el procedimiento señalado en el inciso anterior, se presentara el caso que el número de postulantes que cumple con un mismo criterio es superior al número de vacantes que informa el establecimiento, se aplicará respecto de dichos postulantes el sistema de admisión aleatorio definido por el establecimiento.

Figura 3.1: Extracto de la Ley de Inclusión Escolar

3.1.2. Desarrollo del algoritmo

En lo que sigue, explicamos cómo se implementaron los diferentes requerimientos y las razones por las que se tomaron ciertas decisiones para llevar a cabo el modelo.

En primer lugar, cada nivel se resuelve de manera independiente, pero en orden desde el nivel más alto al nivel más bajo. Esta implementación hace que haya interacción entre los niveles, a través de la prioridad de hermanos y la postulación en bloque. Las consecuencias de estas interacciones se detallan a continuación y en la sección 3.2 con mayor detalle.

En cuanto a la cuota de alumnos prioritarios, se define la cantidad de cupos que equivalen al 15 % de los cupos del colegio y a partir del orden aleatorio original se sube a los primeros alumnos prioritarios en orden relativo en la lista. Estos alumnos quedan con prioridad menor que los alumnos con prioridad de hermanos.

Por otra parte, la prioridad de hermanos viene dada como un input del sistema en los casos en que el hermano mayor ya está matriculado en un colegio. Sin embargo, cuando un alumno es asignado a un establecimiento y tiene un hermano menor postulando a dicho colegio, entonces el hermano menor adquiere prioridad de hermano en el colegio donde fue asignado su hermano mayor, pero se mantiene con menor prioridad que los alumnos que tenían prioridad de hermano por un alumno matriculado antes de comenzar el proceso.

En cuanto a la postulación en bloque, si un alumno queda asignado a un colegio y estaba realizando postulación en bloque, su hermano menor no sólo adquiere mejor prioridad en el colegio por el punto anterior, sino que también se le actualizan sus preferencias, pasando a ser su primera preferencia el colegio donde quedó asignado su hermano mayor.

Debido a los dos puntos anteriores, los niveles de los colegios no son completamente independientes y esto hace que el algoritmo en su conjunto no sea necesariamente a prueba de estrategias como se verá en la sección 3.2.1. Más aún, las asignaciones que se obtienen a partir de iterar desde el nivel más alto o desde el nivel más bajo no son necesariamente las mismas, lo que se ilustra en el siguiente ejemplo.

Supongamos que $C = \{c_1, c_2\}$ tienen cursos en los niveles 1 y 9 y cada colegio tiene 1 cupo en cada nivel, $S_9 = \{s_1, s_2\}$ son alumnos que postulan al nivel 9 y $S_1 = \{s_3, s_4\}$ postulan al nivel 1. Los alumnos s_1 y s_3 son hermanos haciendo postulación en bloque. s_4 tiene prioridad de hermano en el colegio c_1 por un hermano que no participa del sistema. Las preferencias y prioridades originales se muestran en las tablas 3.1 y 3.2.

Nivel 9	c_1	c_2
s_1	1,1	2,1
s_2	1,2	2,2

Tabla 3.1: Preferencias y prioridades nivel 9

Nivel 1	c_1	c_2
s_3	1,2	2,1
s_4	1,1	2,2

Tabla 3.2: Preferencias y prioridades nivel 1

Según esto, la asignación del nivel 9 sería:

$$\mu_9 = \begin{pmatrix} c_1 & c_2 \\ s_1 & s_2 \end{pmatrix}.$$

Como s_1 queda en c_1, su hermano s_3 adquiere prioridad de hermano, pero s_4 también tiene prioridad de hermano por lo que las prioridades no se ven alteradas y la asignación del nivel 1 sería:

$$\mu_1 = \begin{pmatrix} c_1 & c_2 \\ s_4 & s_3 \end{pmatrix}.$$

Es decir, s_1 y s_3 quedan en colegios diferentes. Ahora bien, si nuevamente utilizamos las preferencias y prioridades de las tablas 3.1 y 3.2 pero se asigna primero el nivel 1 y luego el nivel 9, la asignación del nivel 1 sería:

$$\mu_1 = \begin{pmatrix} c_1 & c_2 \\ s_4 & s_3 \end{pmatrix}.$$

Modificamos las prioridades y preferencias del nivel 9, como se muestra en la tabla 3.3.

Nivel 9	c_1	c_2
s_1	2,1	1,1
s_2	1,2	2,2

Tabla 3.3: Preferencias y prioridades actualizadas nivel 9

Con lo que la asignación del nivel 9 sería:

$$\mu_1 = \begin{pmatrix} c_1 & c_2 \\ s_2 & s_1 \end{pmatrix}.$$

Es decir, en este caso los hermanos sí quedan en el mismo colegio.

Es importante destacar que tal como ocurre en este caso, hay otros casos en que familias que quedan juntas iterando desde el nivel más alto quedarían separadas si empezamos iterando desde el nivel más bajo. El ejemplo es análogo.

Tomamos la opción de iterar desde el nivel más alto ya que estos tienen, en general, menos cupos (sólo de alumnos que se cambian de colegios o algunos cupos remanentes) y al iterar desde niveles altos se da mayor igualdad de oportunidades entre todos los postulantes. Esto, porque si iteramos desde niveles inferiores, los alumnos que llenarían los cupos de cursos superiores serían en su mayoría alumnos con hermanos en niveles inferiores.

La asignación por distancia se realiza al final de cada nivel para los alumnos que no fueron asignados a ninguna de sus preferencias. Para esto, se selecciona el colegio más cercano con cupos disponibles y sin copago de cada alumno y se asigna al alumno que tenga la mayor distancia a su colegio más cercano. Se itera este proceso hasta asignar a todos los alumnos.

Implementamos el algoritmo de asignación de la siguiente manera:

Algoritmo 2 Algoritmo de asignación escolar en Chile

```
for nivel de entrada desde el más alto do
  while algún estudiante s no ha sido asignado y tiene colegios en su lista de preferencias
  a los que no ha postulado do
    c := primer colegio en la lista de s al cual s no ha postulado;
    if c tiene cupos disponibles then
      se asigna s a c {tentativamente}
    else
      se toma el alumno s′ con peor prioridad asignado tentativamente a c
      if s′ tiene matrícula asegurada then
        if s tiene matrícula asegurada then
          se asigna s a c {tentativamente}
        else
          c rechaza a s {y s permanece libre}
        end if
      else
        if c prefiere s a s′ then
          se asigna s a c tentativamente y s′ queda libre
        else
          c rechaza a s {y s permanece libre}
        end if
      end if
    end if
  end while
  while algun estudiante s no ha sido asignado do
    se asigna s a su colegio más cercano con cupos disponibles.
  end while
  for c en los colegios de niveles más bajos que el nivel actual do
    Se actualizan las prioridades de los hermanos.
  end for
  for s en los estudiantes que postulan a niveles más bajos que el nivel actual do
    Se actualizan las preferencias de los alumnos que teníán postulación en bloque.
  end for
end for
return  la asignación estable que contiene a los colegios con sus estudiantes asignados.
```

3.2. Análisis de propiedades

Como explicamos en la sección 2.1.1 el algoritmo de Aceptación Diferida cumple con las propiedades de ser a prueba de estrategias y estable, sin embargo como estamos utilizando una regla de rompimiento de empates de múltiples loterías, la asignación no es eficiente.

Dadas las especificaciones de la Ley de Inclusión Escolar y las particularidades mencionadas en esta sección, buscamos demostrar de qué manera el algoritmo diseñado para resolver el problema mantiene las propiedades deseables y en qué casos estas no se cumplen.

3.2.1. A prueba de estrategias

El algoritmo DA utilizado para asignaciones many-to-one es a prueba de estrategias para el lado que propone, en este caso los alumnos, ya que no tienen una mejor estrategia al postular que declarar sus verdaderas preferencias. Cabe señalar que el algoritmo es a prueba de estrategias para los alumnos y no para los colegios, como muestra Roth [1982] al demostrar que no hay algoritmos que entreguen asignaciones estables en los cuales la estrategia óptima de todos los agentes (alumnos y colegios) sea declarar las verdaderas preferencias. Más aun, Roth [1984] demuestra que no hay algoritmos que entreguen asignaciones estables en los cuales la estrategia óptima de todos los agentes sea declarar las verdaderas preferencias, aunque DA, cuando los alumnos son quienes postulan a los colegios, hace que la estrategia dominante de los alumnos sea declarar las verdaderas preferencias.

A pesar de lo anterior, en este caso los colegios no declaran sus preferencias sino que éstas son aleatorias por lo que no hay estrategias que los colegios puedan utilizar para recibir alumnos que les son preferidos y por ende no se pierde la no selección.

En el caso del algoritmo DA adaptado al caso chileno, hay que separar la demostración en varios casos particulares. En primer lugar, si se analiza cada nivel por separado, cada nivel sigue siendo a prueba de estrategias ya que no posee diferencias con el caso original, pero si se consideran las interacciones entre niveles, hay casos en que el algoritmo puede ser estrategizable.

Teorema 3.1 *El Algoritmo 2, de asignación escolar en Chile, es a prueba de estrategias para los alumnos dentro de cada nivel.*

DEMOSTRACIÓN. Si se realiza la asignación de un nivel aislado, el algoritmo es exactamente DA con múltiples loterías, y como se ha mencionado en secciones anteriores y como demuestra Roth [1984], DA es un algoritmo a prueba de estrategias para los alumnos. □

Teorema 3.2 *Al utilizar el Algoritmo 2, de asignación escolar en Chile, puede haber comportamiento estratégico por parte de los alumnos considerando las interacciones entre niveles.*

DEMOSTRACIÓN. Hay que notar que las interacciones entre los niveles vienen dadas por la actualización de prioridades a hermanos menores en los colegios y por la actualización de preferencias de los hermanos menores con respecto a los hermanos mayores.

Un alumno que tiene un hermano en un nivel inferior podría mentir en sus verdaderas preferencias, declarando colegios que sean de preferencia de su hermano menor para quedar asignado en dicho colegio y que su hermano menor obtenga prioridad de hermano. Sin embargo el hermano mayor estaría mintiendo a costa de su propia pérdida de bienestar. Esto lo podemos ver en el siguiente ejemplo.

Supongamos que $C = \{c_1, c_2\}$ son colegios con cursos en los niveles 1 y 9 y cada colegio tiene 1 cupo en cada nivel, $S_9 = \{s_1, s_2\}$ y $S_1 = \{s_3, s_4\}$. Los alumnos s_1 y s_3 son hermanos. Las tablas 3.4 y 3.5 muestran las preferencias y prioridades.

Nivel 9	c_1	c_2
s_1	2,1	1,1
s_2	1,2	2,2

Tabla 3.4: Preferencias y prioridades reales nivel 9

Nivel 1	c_1	c_2
s_3	1,2	2,1
s_4	1,1	2,2

Tabla 3.5: Preferencias y prioridades reales nivel 1

Si todos declaran sus verdaderas preferencias la asignación resultante del nivel 9 sería:

$$\mu_9 = \begin{pmatrix} c_1 & c_2 \\ s_2 & s_1 \end{pmatrix}.$$

Luego, la asignación del nivel 1 sería:

$$\mu_1 = \begin{pmatrix} c_1 & c_2 \\ s_4 & s_3 \end{pmatrix}.$$

Sin embargo, el alumno s_1 puede mentir en sus preferencias para hacer que su hermano quede en un colegio que prefiere más. Las preferencias que declara para lograr esto se muestran en la tabla 3.6.

Nivel 9	c_1	c_2
s_1	1,1	2,1
s_2	1,2	2,2

Tabla 3.6: Preferencias y prioridades alteradas nivel 9

Con lo que la asignación del nivel 9 sería:

$$\mu_9 = \begin{pmatrix} c_1 & c_2 \\ s_1 & s_2 \end{pmatrix}.$$

Y se deben actualizar las prioridades del nivel 1 como mostramos en la tabla 3.7.

Y finalmente la asignación del nivel 1 sería:

$$\mu_1 = \begin{pmatrix} c_1 & c_2 \\ s_3 & s_4 \end{pmatrix}.$$

Nivel 1	c_1	c_2
s_3	1,1	2,1
s_4	1,2	2,2

Tabla 3.7: Preferencias y prioridades actualizadas nivel 1

Con lo que el alumno s_3 mejora su asignación a costa de que su hermano mayor mienta y quede en una preferencia más baja. □

Teorema 3.3 *Si un par de hermanos (postulando a niveles diferentes) realizan postulación en bloque, declaran sus verdaderas preferencias y quedan asignados al mismo colegio utilizando el Algoritmo 2, entonces no existe para ellos una estrategia que domine declarar las verdaderas preferencias.*

DEMOSTRACIÓN. En el nivel del hermano mayor la asignación se realiza utilizando DA, con esto, el alumno queda en la mejor preferencia posible con cupos disponibles por lo que no tiene incentivos a mentir en sus preferencias.

Luego, el hermano menor actualiza sus preferencias, declarando que su primera preferencia es el colegio donde fue asignado su hermano mayor, por lo que si queda asignado en dicho colegio, no hay forma de asignarlo mejor y tampoco tiene incentivos a mentir. □

Teorema 3.4 *Si un par de hermanos (postulando a niveles diferentes) realizan postulación en bloque, declaran sus verdaderas preferencias y quedan asignados a colegios distintos utilizando el Algoritmo 2, hay casos en que el hermano mayor puede no declarar las verdaderas preferencias y quedar en el mismo colegio que su hermano menor.*

DEMOSTRACIÓN. Primero hay que notar que al realizar postulación en bloque, los hermanos declaran que su primera preferencia es quedar juntos, aunque sea en preferencias bajas de su lista.

En este sentido, y como la asignación se realiza desde el nivel más alto al nivel más bajo, el hermano mayor puede mentir. En el siguiente ejemplo se muestra un caso donde el hermano mayor miente en sus preferencias para quedar junto a su hermano menor.

Supongamos que $C = \{c_1, c_2\}$ son colegios con cursos en los niveles 1 y 9 y cada colegio tiene 1 cupo en cada nivel, $S_9 = \{s_1, s_2\}$ y $S_1 = \{s_3, s_4\}$. El alumno s_4 tiene un hermano en el colegio c_1 por lo que tiene prioridad de hermano. Los alumnos s_1 y s_3 son hermanos haciendo postulación en bloque. Las tablas 3.8 y 3.9 muestran las preferencias y prioridades.

Nivel 9	c_1	c_2
s_1	1,1	2,2
s_2	1,2	2,1

Tabla 3.8: Preferencias y prioridades reales nivel 9

Si todos declaran sus verdaderas preferencias la asignación resultante del nivel 9 sería:

Nivel 1	c_1	c_2
s_3	2,2	1,2
s_4	1,1	2,1

Tabla 3.9: Preferencias y prioridades reales nivel 1

$$\mu_9 = \begin{pmatrix} c_1 & c_2 \\ s_1 & s_2 \end{pmatrix}.$$

Luego en el nivel 1 el alumno s_3 actualiza sus preferencias pero el colegio c_1 no cambia sus prioridades porque el alumno s_4 también tiene prioridad de hermano. Quedando como sigue:

Nivel 1	c_1	c_2
s_3	1,2	2,2
s_4	1,1	2,1

Tabla 3.10: Preferencias y prioridades actualizadas nivel 1

Luego la asignación resultante del nivel 1 será:

$$\mu_1 = \begin{pmatrix} c_1 & c_2 \\ s_4 & s_3 \end{pmatrix}.$$

Por lo que s_1 y s_3 no quedan en el mismo colegio.

Ahora analizamos el caso en que el alumno s_1 miente en sus preferencias:

Nivel 9	c_1	c_2
s_1	2,1	1,2
s_2	1,2	2,1

Tabla 3.11: Preferencias y prioridades alteradas nivel 9

Las preferencias y prioridades del nivel 1 siguen siendo las de la tabla 3.9.

Luego la asignación del nivel 9 resulta:

$$\mu_9 = \begin{pmatrix} c_1 & c_2 \\ s_2 & s_1 \end{pmatrix}.$$

Con esto, para el nivel 1, el alumno s_3 adquiere mayor prioridad en el colegio c_2 por la postulación en bloque. Las prioridades actualizadas se muestran en la tabla 3.12.

Y finalmente la asignación del nivel 1 queda como sigue:

$$\mu_1 = \begin{pmatrix} c_1 & c_2 \\ s_4 & s_3 \end{pmatrix}.$$

Nivel 1	c_1	c_2
s_3	2,2	1,1
s_4	1,1	2,2

Tabla 3.12: Preferencias y prioridades actualizadas nivel 1

Dado esto, concluimos que hay posibilidad de que el hermano mayor mienta y así una pareja de hermanos que hizo postulación familiar obtenga una mejor asignación. Cabe señalar que el hermano mayor queda en un colegio de menor preferencia que en el caso que declara las verdaderas preferencias, pero queda mejor porque está en el mismo colegio que su hermano menor. □

3.2.2. Postulación en bloque

La postulación en bloque tiene como objetivo darle a las familias mayor probabilidad de que sus hijos queden en el mismo colegio. En lo que viene, buscamos demostrar que con la implementación actual del algoritmo, esta probabilidad efectivamente aumenta.

Teorema 3.5 *Si, utilizando el Algoritmo 2, dos hermanos quedan asignados en el mismo colegio sin realizar postulación en bloque, entonces necesariamente quedarían juntos si la realizan.*

DEMOSTRACIÓN. Supongamos que los alumnos s_1 y s_2 son hermanos, donde s_1 postula al nivel 9 y s_2 al nivel 1. Los hermanos no realizan postulación en bloque y quedan asignados al colegio c_4 que estaba en la lista de preferencias de ambos.

Dado esto, se busca mostrar que si s_1 y s_2 realizaran postulación en bloque no quedarían juntos.

En primer lugar, es directo que el alumno s_1 queda asignado en c_4 ya que para él no hay ninguna diferencia en el algoritmo al realizar o no postulación en bloque.

Luego para el alumno s_2 el colegio c_4 sube a estar en su primera preferencia, y como en el caso sin postulación en bloque quedaba asignado, el colegio tiene cupos para él, por lo que necesariamente será asignado. Lo cual contradice el hecho de que quedan separados con postulación en bloque.

Por ende, si los hermanos quedan asignados sin postulación en bloque, necesariamente quedarán asignados juntos con ella. □

Teorema 3.6 *Si una familia prefiere quedar junta en un colegio de su preferencia a quedar separados en colegios que hayan declarado con mejor ranking en su lista de preferencias, entonces la decisión óptima es realizar postulación en bloque si la asignación se realiza según el Algoritmo 2.*

Demostración. Primero, utilizamos el Teorema 3.5 para mostrar que realizando postulación en bloque, las familias obtienen al menos el mismo resultado que sin realizarla.

En segundo lugar, mostramos que hay al menos un caso en que dos hermanos quedan en colegios diferentes cuando no hacen postulación en bloque, pero que al hacerla quedan juntos.

Esto ocurre en casos como el siguiente, supongamos $C = \{c_1, c_2, c_3\}$ tienen cursos en los niveles 1 y 9 y cada colegio tiene 1 cupo en cada nivel. $S_9 = \{s_1, s_2\}$ son alumnos que postulan al nivel 9 y $S_1 = \{s_3, s_4\}$ postulan al nivel 1. Los alumnos s_1 y s_3 son hermanos. Las preferencias y prioridades se muestran en las tablas 3.13 y 3.14.

Nivel 9	c_1	c_2	c_3
s_1	1,2	3,1	2,1
s_2	1,1	3,2	2,2

Tabla 3.13: Preferencias y prioridades reales nivel 9

Nivel 1	c_1	c_2	c_3
s_3	1,1	2,2	3,2
s_4	3,2	2,1	1,1

Tabla 3.14: Preferencias y prioridades reales nivel 1

Si los hermanos no realizan postulación en bloque, la asignación sería:

$$\mu_9 = \begin{pmatrix} c_1 & c_2 & c_3 \\ s_2 & \emptyset & s_1 \end{pmatrix}$$

$$\mu_1 = \begin{pmatrix} c_1 & c_2 & c_3 \\ s_3 & \emptyset & s_1 \end{pmatrix}.$$

En cambio, si s_1 y s_3 realizaran postulación en bloque, entonces la asignación del nivel 9 sería:

$$\mu_9 = \begin{pmatrix} c_1 & c_2 & c_3 \\ s_2 & \emptyset & s_1 \end{pmatrix}.$$

Pero luego las preferencias y prioridades del nivel 1 cambiarían como se muestra en la tabla 3.15.

Nivel 1	c_1	c_2	c_3
s_3	2,1	3,2	1,1
s_4	3,2	2,1	1,2

Tabla 3.15: Preferencias y prioridades actualizadas nivel 1

Quedando la asignación del nivel 1 como sigue:

$$\mu_1 = \begin{pmatrix} c_1 & c_2 & c_3 \\ \emptyset & s_2 & s_3 \end{pmatrix}.$$

□

Por ende, queda demostrado que siempre que dos hermanos queden en el mismo colegio, también quedarían juntos realizando postulación en bloque, y que en casos de hermanos que quedan separados, quedarían juntos realizando postulación en bloque. Es decir, si la verdadera primera preferencia de las familias es que los hermanos queden en el mismo colegio, deben hacer postulación en bloque.

3.2.3. Estabilidad

La idea de que el algoritmo desarrollado sea estable es que no se genere un conjunto de alumnos que prefieran colegios en los que tienen mejor prioridad que el último alumno matriculado. Para evaluar si el algoritmo sigue entregando una asignación estable, es necesario evaluar dos casos, con y sin postulación en bloque.

Teorema 3.7 *Si ninguna familia realiza postulación en bloque, la asignación entregada por el Algoritmo 2 es estable.*

Demostración. En un nivel donde no hay hermanos, es directo que no hay cambios. En niveles con hermanos, se ve que el nivel del hermano mayor tampoco se ve afectado de ninguna manera en sus propiedades y asignación y que únicamente le otorgará prioridad de hermano a su hermano menor, quedando así demostrado tal como en el caso genérico de DA.

En el nivel del hermano menor, se conserva estabilidad si se considera la asignación como producto de las prioridades actualizadas del colegio, es decir, se considera que el alumno que tiene un hermano asignado en un nivel superior pasa a tener prioridad de hermano no hay cambios con respecto a DA. □

Teorema 3.8 *Si hay familias que realizan postulación en bloque, la asignación entregada por el Algoritmo 2 es estable si los hermanos de cada familia con postulación en bloque quedan asignados en el mismo colegio.*

Demostración. Si el bloque queda junto, el hermano mayor no puede mejorar su asignación por efecto del algoritmo y, más aun, no quiere cambiarse ya que su hermano menor quedó en dicho colegio.

El hermano menor quedó en su nueva primera preferencia (el colegio de su hermano mayor) y por ende no puede mejorar su asignación.

También hay estabilidad desde el punto de vista de los demás alumnos (que no eran parte de la postulación familiar). En particular, en el nivel superior no hay cambios con respecto al algoritmo original, y en el nivel inferior, se debe considerar la estabilidad bajo las nuevas prioridades del colegio sobre el hermano menor. □

Teorema 3.9 *Si hay familias que realizan postulación en bloque, la asignación entregada por* *el Algoritmo 2 no es necesariamente estable si los hermanos de las familias con postulación en bloque quedan asignados a colegios diferentes.*

Demostración. En el caso de que el bloque quede separado, es decir, que los hermanos sean asignados a colegios diferentes, se puede ver que en el nivel inferior la asignación es estable para todos los alumnos. En particular para el alumno que era hermano menor y quedó en un colegio diferente que su hermano, esto se debe a que aún con el aumento de prioridad en dicho colegio, entraron alumnos con mejor prioridad que él y por ende no habían cupos disponibles, por lo que no puede mejorar su asignación. Para los demás alumnos no hay diferencias en la asignación a lo que pasaría sin postulación en bloque.

Sin embargo, la asignación no es necesariamente estable en el caso del nivel superior. Para los alumnos que no están involucrados en la postulación en bloque no se presentan problemas. Pero para el hermano mayor que hizo postulación en bloque y no quedó con su hermano, podría darse el caso de que el hermano mayor pudiera entrar en el colegio en el que fue asignado su hermano menor, y esto sería mejor que su asignación actual ya que la familia declaró que su primera preferencia era que los hermanos quedaran juntos.

La razón de que ocurra esto es la ejecución del algoritmo en orden descendente por nivel. El ejemplo que sigue ilustra un caso en que esto puede ocurrir.

Supongamos que los alumnos $S_9 = \{s_1, s_2\}$ postulan al nivel 9 y los alumnos $S_1\{s_3, s_4\}$ postulan al nivel 1. Los colegios $C = \{c_1, c_2, c_3\}$ tienen un cupo en cada nivel. Los alumnos s_1 y s_3 son hermanos con postulación familiar. Las preferencias y prioridades se muestran en las tablas 3.16 y 3.17. s_4 tiene prioridad de hermano en el colegio c_1.

Nivel 9	c_1	c_2	c_3
s_1	1,1	2,1	3,2
s_2	1,2	2,2	3,1

Tabla 3.16: Preferencias y prioridades reales nivel 9

Nivel 1	c_1	c_2	c_3
s_3	2,2	3,1	1,2
s_4	1,1	2,2	3,1

Tabla 3.17: Preferencias y prioridades reales nivel 1

Bajo estas condiciones, la asignación del nivel 9 sería:

$$\mu_9 = \begin{pmatrix} c_1 & c_2 & c_3 \\ s_1 & s_2 & \emptyset \end{pmatrix}.$$

Luego actualizamos la preferencia de s_3 por el colegio c_1 pero no se actualizan las prioridades ya que s_4 tenía prioridad de hermano. Esto se muestra en la tabla 3.18.

Y la asignación queda como sigue:

Nivel 1	c_1	c_2	c_3
s_3	1,2	3,1	2,2
s_4	1,1	2,2	3,1

Tabla 3.18: Preferencias y prioridades actualizadas nivel 1

$$\mu_1 = \begin{pmatrix} c_1 & c_2 & c_3 \\ s_4 & \emptyset & s_3 \end{pmatrix}.$$

Es decir, s_1 y s_3 no quedan juntos a pesar de haber realizado postulación en bloque. Sin embargo, considerando que a partir de esta asignación s_1 prefiere el colegio c_3 (porque su hermano fue asignado) y a su vez este colegio tiene disponibilidad, s_3 podría cambiarse al colegio c_3. Es decir, la asignación anterior, representada por μ_9 y μ_1, no es estable.

Es importante notar que esto no ocurre sólo en el caso en que haya un colegio con cupos disponibles, sino que también podría darse en el caso que haya un alumnos asignado a dicho colegio (c_3 en este caso) con peor prioridad que el alumno que desea ingresar (en este caso s_1). $\square$

3.2.4. Matrícula asegurada

Buscamos explicar por qué las propiedades de DA no se ven alteradas al asignar alumnos con matrícula asegurada de la forma en que explicamos en la sección 3.1.2.

Teorema 3.10 *La implementación de la asignación por matrícula asegurada no afecta las propiedades del Algoritmo 2.*

Demostración. Cuando un colegio declara cupos suficientes al menos para recibir a todos los alumnos que tienen matrícula asegurada en dicho establecimiento, entonces el problema no cambia con respecto a la situación sin matrícula asegurada ya que simplemente se los considera como los alumnos con mayor prioridad del establecimiento.

Si el colegio no declarase suficientes cupos para estos alumnos, y los alumnos no fuesen aceptados en alguno de los colegios a los que postularon, entonces el colegio de origen tiene la obligación de aceptarlos aunque se sobrepasen los cupos declarados.

Esto no afecta el algoritmo ya que ningún alumno nuevo sería aceptado y el colegio en cuestión estaría en la misma situación que antes de haber participado del proceso, mientras que los alumnos tienen asegurada la continuidad de sus estudios. $\square$

Cabe señalar que quedar asignado por matrícula asegurada no implica necesariamente un resultado negativo. Un alumno puede participar del proceso de admisión para intentar cambiarse a un colegio que le gusta más y en caso de no ser aceptado quedarse en su colegio y no declarar más preferencias. En estos casos, la matrícula asegurada sería equivalente a quedar asignado a la segunda preferencia y por ende sería un resultado positivo.

Capítulo 4

Implementación Región de Magallanes

4.1. Descripción de la instancia

La implementación del algoritmo en la Región de Magallanes se realizó entre septiembre y diciembre del 2016 para el proceso de admisión escolar 2017. Durante este período se realizaron diversas etapas del proceso, pero para efectos de este trabajo nos centramos en la etapa principal de asignación que se describe a continuación.

Dado el calendario de puesta en marcha del sistema centralizado de asignación escolar, la instancia corresponde a la Región de Magallanes para los 5 niveles de ingreso. Éstos son pre-kínder, kínder, primero básico, séptimo básico y primero medio. Esto considera a 3436 alumnos en total y 63 colegios, de los cuales no todos incluyen todos los niveles mencionados. En la tabla 4.1 se muestran los datos específicos por nivel.

Nivel	# alumnos	# colegios
Pre-Kínder	1195	42
Kínder	572	43
Primero básico	455	53
Séptimo básico	174	52
Primero medio	1040	24

Tabla 4.1: Alumnos y colegios por nivel

Un total de 140 alumnos que postularon en este proceso de admisión tenían un hermano menor postulando en el sistema y 8 alumnos tenían 2 hermanos menores postulando, a quienes le otorgarían prioridad de hermano en el colegio donde fueron asignados.

Además, 40 pares de hermanos hicieron postulación en bloque. La postulación en bloque afecta únicamente a los colegios de la intersección de las listas de preferencia de ambos alumnos, es decir, sólo se aumenta la preferencia del hermano menor en el colegio donde el hermano mayor fue asignado, en caso de que dicho colegio esté en su lista de preferencias.

En cuanto a las postulaciones, la Ley de Inclusión Escolar establece que cada alumno debe postular al menos a dos colegios, excepto en zonas rurales, donde pueden postular solamente a un colegio. Las familias postularon a 3.46 colegios en promedio.

4.2. Resultados computacionales

En primer lugar, en la figura 4.1 podemos ver que un 58 % de los postulantes de la región queda en su primera preferencia y que un 86.4 % queda en alguna de sus preferencias. Un 9 % de los alumnos es asignado por distancia y un 4.6 % se mantiene en su colegio de origen, lo cual representamos en el gráfico por Dist. y M.A. respectivamente.

Cabe señalar que, en lo que sigue, los resultados se calculan a partir de un total de 3357 alumnos ya que se excluyen aquellos que postulan a su colegio de origen en primera preferencia. Dichos alumnos quedarán asignados sin duda a su colegio y no influyen en el proceso de asignación de ninguna manera.

Al analizar el caso de la comuna de Punta Arenas, se ve que proporcionalmente menos alumnos quedan en su primera preferencia y más alumnos son asignados por distancia o a su colegio de origen.

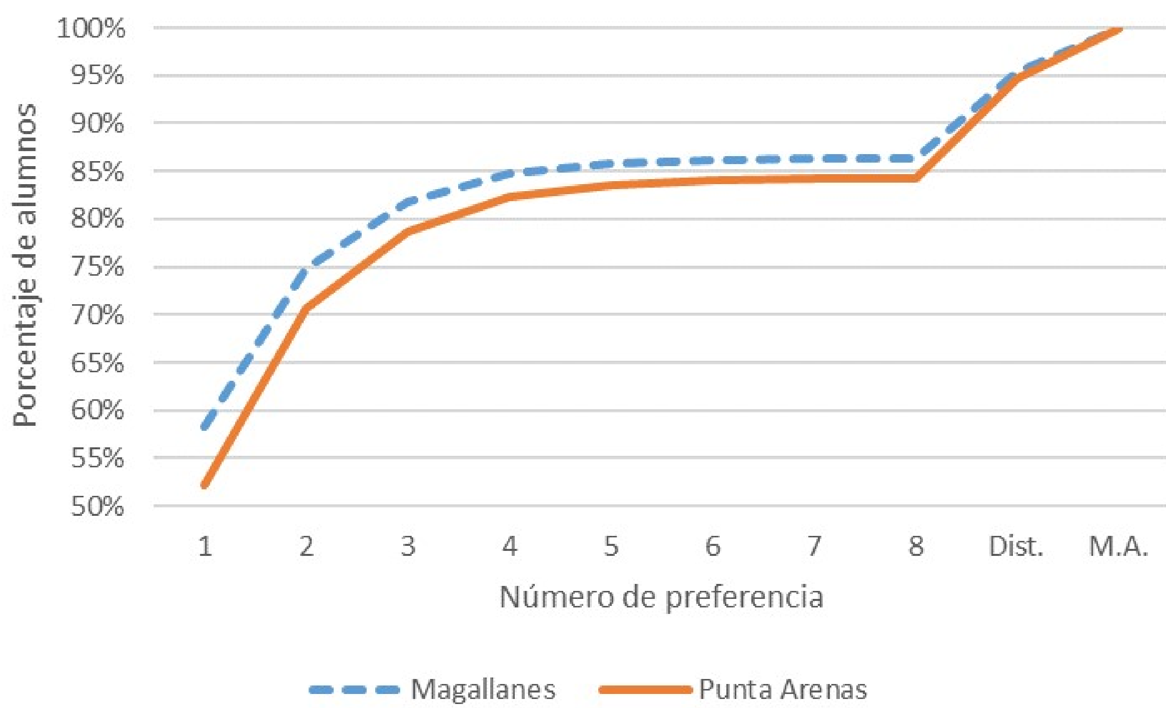

Figura 4.1: Porcentaje acumulado de alumnos asignados por preferencia

Esto puede darse por varias razones, pero destacamos la hipótesis de la composición geográfica de la región. La comuna de Punta Arenas está considerablemente más concentrada que el resto de la región, lo que puede provocar mayores correlaciones en las postulaciones y un empeoramiento de los resultados. En cambio, en comunas menos habitadas y con menos colegios, como son las zonas rurales, las personas postulan a colegios donde es casi seguro que hay cupos para ellos.

Si revisamos los resultados para cada nivel en Punta Arenas, se puede ver que estos varían de manera significativa. En la figura 4.2 vemos, por ejemplo, que en el nivel de pre-kínder queda más de un 60 % de los alumnos en su primera preferencia, mientras que en primero básico lo hace menos de un 40 %. En cuanto a la matrícula asegurada destaca séptimo básico, donde más de un 20 % de los alumnos queda en su colegio de origen.

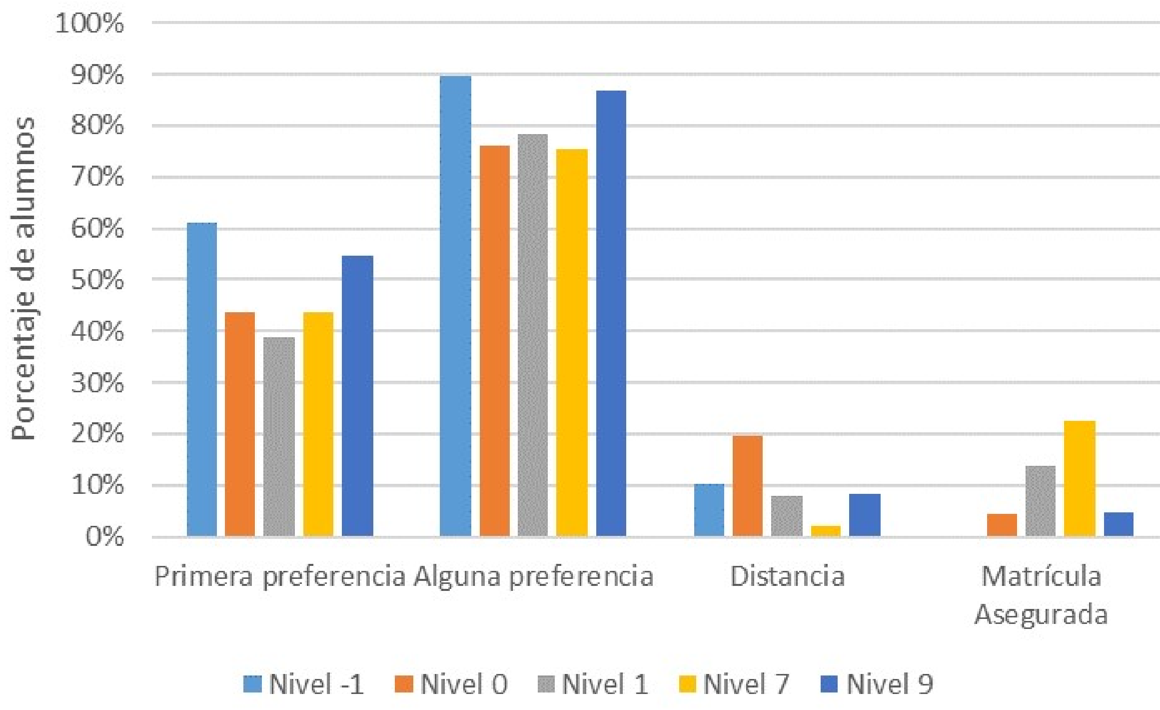

Figura 4.2: Resumen de la asignación por nivel, Punta Arenas

4.3. Análisis de resultados

4.3.1. Prioritarios

Los criterios de prioridad son un tema complejo que no tiene una única forma de resolverse. En particular, la implementación descrita en este trabajo no asegura que se cumpla con el 15 % de alumnos prioritarios por curso, les da ventaja sobre el resto de los alumnos.

En este sentido, analizamos la diferencia entre la asignación de los alumnos prioritarios y los que no lo son. Esto se realiza para el nivel de pre-kínder en la comuna de Punta Arenas, donde hay 956 alumnos de los cuales 197 son prioritarios y 56 son de la cuota de prioritarios, es decir, quedaron asignados en el colegio en el que clasificaron dentro del criterio de prioritarios.

Dado esto, en la figura 4.3 podemos ver que una menor proporción de alumnos no prioritarios queda en su primera preferencia en relación a los prioritarios en general y a aquellos que están en la cuota de prioritarios. En particular, más de un 11 % de alumnos no prioritarios quedan asignados por distancia, mientras que para los prioritarios esta cifra es sólo un 4.6 %. Los alumnos denominados en la cuota de prioritarios quedan asignados en su totalidad a sus tres primeras preferencias.

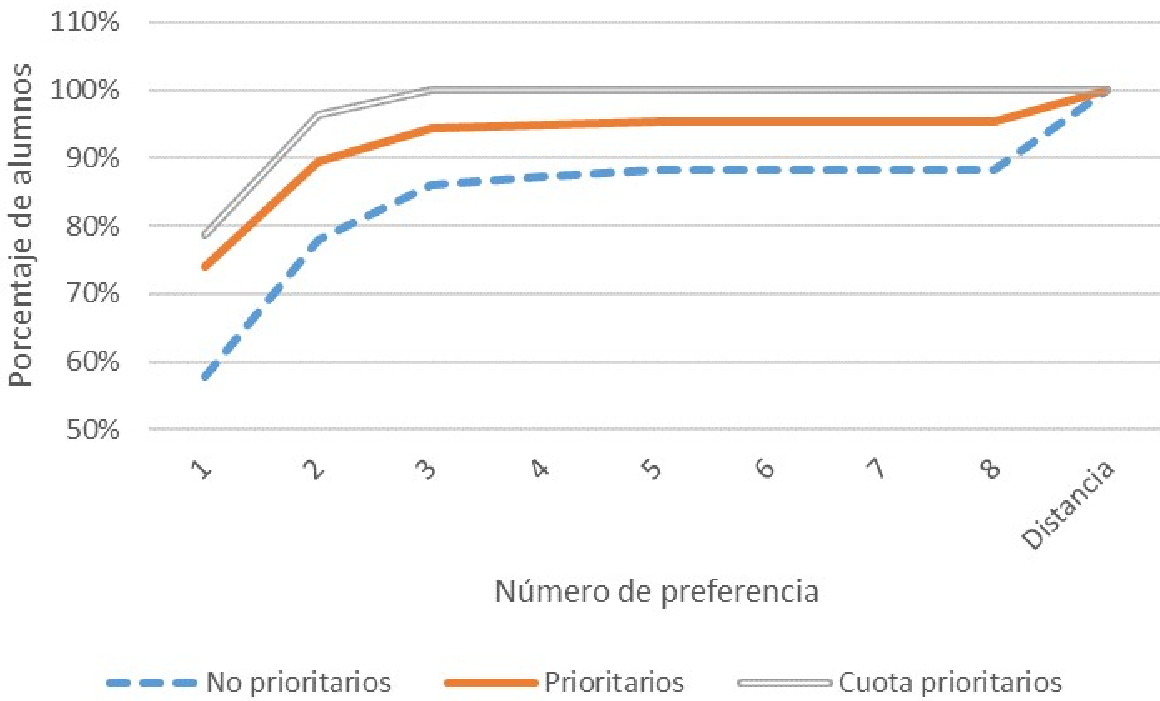

Figura 4.3: Porcentaje de alumnos asignados por preferencia, nivel pre-kínder, Punta Arenas

Analizando desde el punto de vista de los colegios, queremos identificar los cursos que no cumplen con la cuota establecida de alumnos prioritarios pero que tienen alumnos prioritarios en su lista de espera. La razón para que la cuota no se cumpla, aún cuando hay prioritarios en la lista de espera, es que los alumnos a los que se les dio mayor prioridad en dicho colegio están en un colegio mejor para ellos o que el colegio completó sus cupos con alumnos que tenían prioridad de hermanos.

Estos colegios se muestran en la tabla 4.2, donde la columna "# colegios que no cumplen cuota" se refiere al total de colegios que cumplen la condición anterior en dicho nivel. La columna "# colegios que pueden cumplirla" se refiere a los colegios que, dada la cantidad de alumnos prioritarios que tienen en su lista de espera, bajo un algoritmo de asignación diferente podrían cumplir la cuota de alumnos prioritarios aunque no se asegura que la cumplirán.

Cabe señalar que hay colegios en los que no se cumple la cuota pero que no tienen alumnos prioritarios en su lista de espera, estos colegios no se consideran en el análisis ya que no hay un algoritmo de asignación que les permita cumplir la cuota de prioritarios sin perjudicar la asignación de los alumnos.

Dada la implementación actual de la cuota de prioritarios, realizamos 10.000 simulaciones del algoritmo bajo el supuesto de que se elimina la cuota de prioritarios en los colegios. Los

Nivel	# colegios que no cumplen cuota	# colegios que pueden cumplirla
Pre-Kínder	6	4
Kínder	7	5
Primero básico	3	3
Séptimo básico	2	0
Primero medio	2	2

Tabla 4.2: Cumplimiento de la cuota de prioritarios por nivel

resultados muestran que al eliminar la cuota de prioritarios, en promedio, sólo un colegio más cumple la cuota de prioritarios.

Si bien esta diferencia no es significativa, puede indicar que el algoritmo actual no es efectivo para cumplir la cuota de prioritarios y se cumple porque cerca de un 28 % de los alumnos totales son alumnos prioritarios.

El hecho de que al subir una cantidad fija de alumnos prioritarios, menos colegios cumplan la cuota, se debe a que darles mayor prioridad no implica aumentar las probabilidades de cumplir la cuota, sino aumentar las probabilidades de que queden en mejores prioridades.

4.3.2. Reglas de rompimiento de empates

La elección de la regla de rompimiento de empates no tiene una única respuesta. En primer lugar, los resultados dependen no sólo de la regla de desempate en sí, sino también del largo de las listas de preferencias de los estudiantes y cupos de los colegios en relación a la cantidad de alumnos postulando.

La pregunta sobre qué regla de desempate utilizar para DA fue formulada por primera vez por Abdulkadiroglu and Sönmez [2003], sugiriendo que MTB (múltiples loterías) podría generar ineficiencias innecesarias. Sin embargo, en la práctica, MTB se ve como una opción más justa y equitativa para las familias, pues con STB (única lotería) si se obtiene una mala lotería, es poco probable quedar asignado en cualquier colegio de la lista de preferencias.

Abdulkadiroğlu et al. [2009] muestran, con datos de Nueva York, que utilizar STB permite que más alumnos queden en sus primeras preferencias pero a la vez, que más alumnos queden sin asignación. Es decir, no hay dominancia estocástica entre STB y MTB.

Esto se explica porque si un alumno fue asignado a sus primeras preferencias bajo STB, entonces debe tener una buena lotería y es poco probable que otro alumno se quede con su lugar, mientras que si un alumno es rechazado de varios colegios bajo STB entonces debe tener una mala lotería y pocas probabilidades de ser asignado a los siguientes colegios, por lo que recibirá un colegio que está bajo en su lista de preferencias o no será asignado. MTB evita esto y es más justo en cuanto a las oportunidades que se les da a los alumnos.

En particular, Ashlagi et al. [2015] dan evidencia teórica que muestra que no hay dominancia estocástica entre ambas reglas de rompimiento de empates y además demuestra que las listas de preferencias cortas mejoran la asignación promedio de los alumnos bajo MTB. Intuitivamente esto se da porque si los alumnos postulan a muchos colegios, pueden aparecer con buena prioridad en las listas de prioridades de colegios que están bajos en sus preferencias, bloqueando a alumnos que si quieren estar en esos colegios pero obtuvieron baja prioridad, y así generar largas cadenas de bloqueo que empeoran los resultados.

El problema de obligar a los alumnos a declarar pocos colegios, es que algunos alumnos podrían no declarar colegios de alta demanda bajo el supuesto de que es poco probable que queden asignados y prefieren ocupar el lugar de la lista en un colegio donde es más probable ser asignado, lo que hace que el algoritmo no sea a prueba de estrategias.

El caso chileno no restringe el número de preferencias de las familias, por ende no se pierde la propiedad de ser a prueba de estrategias. Aun así, las listas de preferencias efectivas de los alumnos son relatitvamente cortas, por lo que un 13.6 % de los alumnos queda sin asignar.

Por último, Ashlagi and Nikzad [2015] separan el problema de elección de reglas de rompimiento de empates en dos escenarios, cuando el mercado está sobredemandado y cuando no lo está. Para entender las principales conclusiones que entrega, es necesario definir 3 elementos:

- Dominancia estocástica: la distribución de preferencias de los alumnos es una función $R : [1; m] \to [0; n]$ donde $R(\mathrm{i})$ se refiere al número de estudiantes que son asignados a alguna de sus primeras i preferencias. Fijamos una constante $\varepsilon \geq 0$. Decimos que la distribución de preferencias R domina estocásticamente la distribución R' si para cualquier entero $\mathrm{i} \in [1, m]$, $\sum_{j=1}^{\mathrm{i}} R(j) \geq \sum_{j=1}^{\mathrm{i}} R'(j)$. Y R casi domina estocásticamente a R' si para cualquier $\mathrm{i} \in [1, m]$, o bien $\sum_{j=1}^{\mathrm{i}} R(j) \geq \sum_{j=1}^{\mathrm{i}} R'(j)$ o $\sum_{j=1}^{m} R(j) \leq (\log n)^{1+\varepsilon}$.

- Pares de mejora: asumimos una asignación μ. Sea $\mu(x)$ la asignación de x y supongamos que $\mu(x) = \emptyset$ si x no es asignado en μ. Un par de estudiantes $s, s' \in S$ son considerados un par de mejora en μ si s prefiere $\mu(s')$ a $\mu(s)$ y s' prefiere $\mu(s)$ a $\mu(s')$. Definimos $\ddot{\mu}(s)$ como el número de pares de mejora en que participa s en μ.

- Varianza del ranking: para explicar este concepto primero se necesita definir ciertos elementos. Consideremos una asignación μ, luego para cualquier alumno s, $\mu^{\#}(s)$ se refiere al lugar que ocupa el colegio $\mu(s)$ en la lista de preferencias de s. Definimos el ranking promedio de los alumnos asignados en μ como $Ar(\mu) = \frac{1}{|\mu(C)|} \sum_{s \in \mu(C)} \mu^{\#}(s)$, donde $\mu(C)$ son todos los alumnos asignados.

  Además, definimos como μ_π la asignación que se obtiene bajo un orden aleatorio π. Con esto, la varianza del ranking se define de la siguiente manera:

$$Var[\mu^{\#}(s)] = E_{(\pi(c):c \in C)}[(Ar(\mu_\pi) - \mu_\pi^{\#}(s))^2]\mu_\pi(s) \neq \emptyset.$$

Con esto, los autores demuestran que en mercados con sobredemanda, STB casi domina estocásticamente a MTB, MTB genera muchos pares de mejora y la varianza del ranking es mayor con MTB, mientras que en el caso subdemandado, no hay dominancia estocástica, MTB genera pocos pares de mejora y la varianza del ranking es mayor con STB.

Dado lo anterior, proponen utilizar STB en un subconjunto de colegios que denominan populares y MTB en el resto de los colegios. La popularidad se da en colegios que tienen más postulantes que cupos, pero el umbral desde el que se considera que un colegio es popular dependerá de cada caso.

Los resultados obtenidos en la Región de Magallanes van en línea con la literatura. Para mostrar esto, realizamos 1.000 simulaciones de la instancia con STB para comparar los resultados con los obtenidos en el caso real con MTB. En STB mantenemos los criterios de prioridad por lo que, en cada colegio, mejoramos la prioridad a los alumnos que cumplen los criterios de prioridad de la Ley de Inclusión Escolar detallados en la sección 3.1.1.

En la figura 4.4 vemos que con STB más alumnos quedan en sus 3 primeras preferencias mientras que con MTB menos alumnos quedan sin asignar, lo que implica que más alumnos quedan en preferencias más bajas. El punto de intersección de ambas loterías es entre las preferencias 3 y 4.

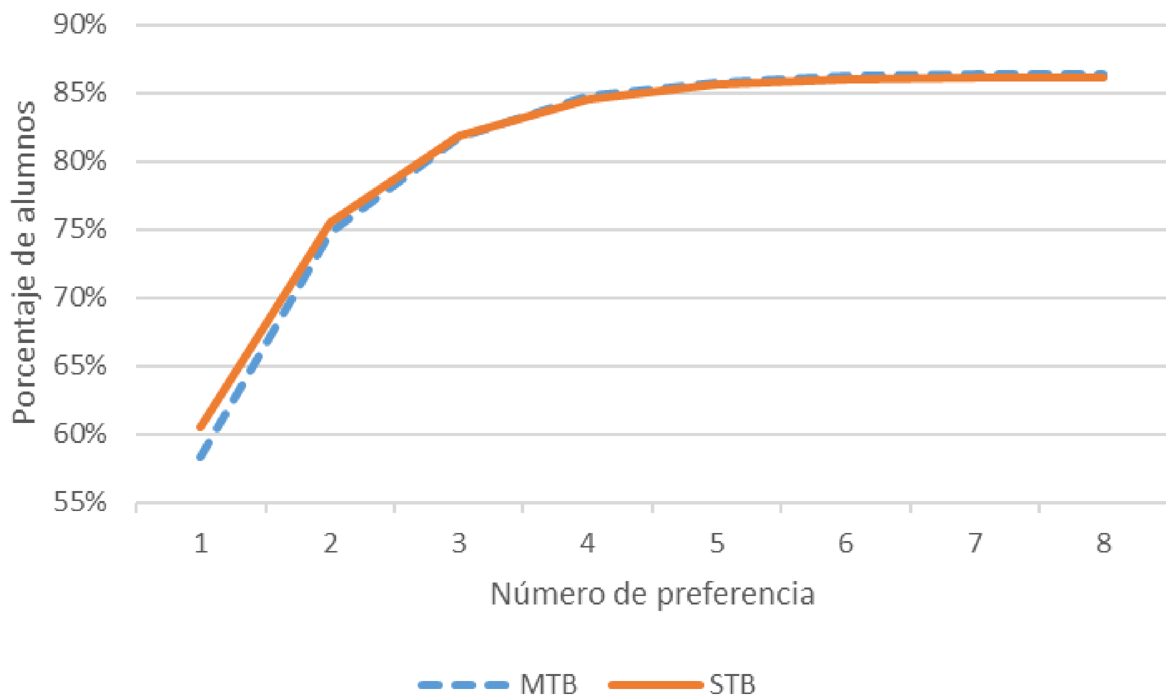

Figura 4.4: Porcentaje acumulado de alumnos asignados por preferencia

Sin embargo, los resultados muestran que bajo el criterio de única lotería, más alumnos quedan sin asignar, lo que en este caso equivale a ser asignados a su colegio de origen o por distancia. Los valores se muestran en la tabla 4.3.

Regla de desempate	Asignados a preferencia	Distancia o matrícula asegurada
STB	86.08 %	13.92 %
MTB	86.36 %	13.64 %

Tabla 4.3: Comparación de reglas de desempate

Capítulo 5

Modelo de preferencias

En esta sección analizamos numéricamente las preferencias de los apoderados a partir de los datos obtenidos en la Región de Magallanes y buscamos un modelo probabilístico que se ajuste a los datos obtenidos.

El objetivo de modelar las preferencias es poder realizar simulaciones de instancias más grandes que reflejen lo que ocurrirá en los próximos años con el Mecanismo de Admisión Escolar en todo Chile.

Utilizar un modelo probabilístico consiste en modelar no a partir de las decisiones particulares de cada persona, sino que a partir de las consecuencias de las decisiones, es decir, observando las elecciones y no las razones de la elección.

La hipótesis principal es que las preferencias de los apoderados están correlacionadas, ya sea porque las familias tienden a copiar las elecciones de los demás o porque se valoran elementos similares a la hora de elegir colegios.

Otra opción sería utilizar un modelo de elección, como un modelo logit o logit multinomial, los cuales permiten entender las motivaciones de las familias detrás de la elección de los establecimientos. No desarrollamos este tipo de modelos ya que dado el tamaño de la instancia los resultados no son significativos e incluso el modelo podría quedar mal identificado.

5.1. Preferencias en Magallanes

Para comenzar, es interesante analizar las preferencias de las familias. La Ley de Inclusión Escolar indica que las familias deben postular como mínimo a dos colegios, excepto en zonas rurales donde pueden postular solo a uno.

Los alumnos que postulan sólo a un colegio representan menos del 1 % del total. Considerando al total de las familias, observamos que postulan a 3.46 colegios en promedio. En la tabla 5.1 presentamos las postulaciones promedio por nivel.

Nivel	# postulantes	# preferencias
Pre-Kínder	1195	3.14
Kínder	572	3.36
Primero básico	455	3.74
Séptimo básico	174	3.8
Primero medio	1040	3.7

Tabla 5.1: Preferencias por nivel

Aun así, destaca que si bien un 99 % de los alumnos declara 2 preferencias o más, sólo un 70 % declara 3 preferencias o más, y este número se reduce a menos de un 40 % de alumnos que declaran 4 o más preferencias. Es por esto que para efectos del análisis de preferencias se analizarán a lo más las 3 primeras preferencias declaradas por las familias, haciendo énfasis en la primera.

En la figura 5.1 podemos ver la cantidad de postulantes en primera preferencia que recibe cada colegio, luego los postulantes en sus dos primeras preferencias y en sus 3 primeras preferencias. Vemos que si se ordenan los colegios según los más demandados en primera preferencia, este orden no se mantiene al revisar las preferencias acumuladas. Cabe destacar que hay 2 colegios que no reciben ninguna postulación en ningún nivel ni lugar en la lista de prioridades de los alumnos.

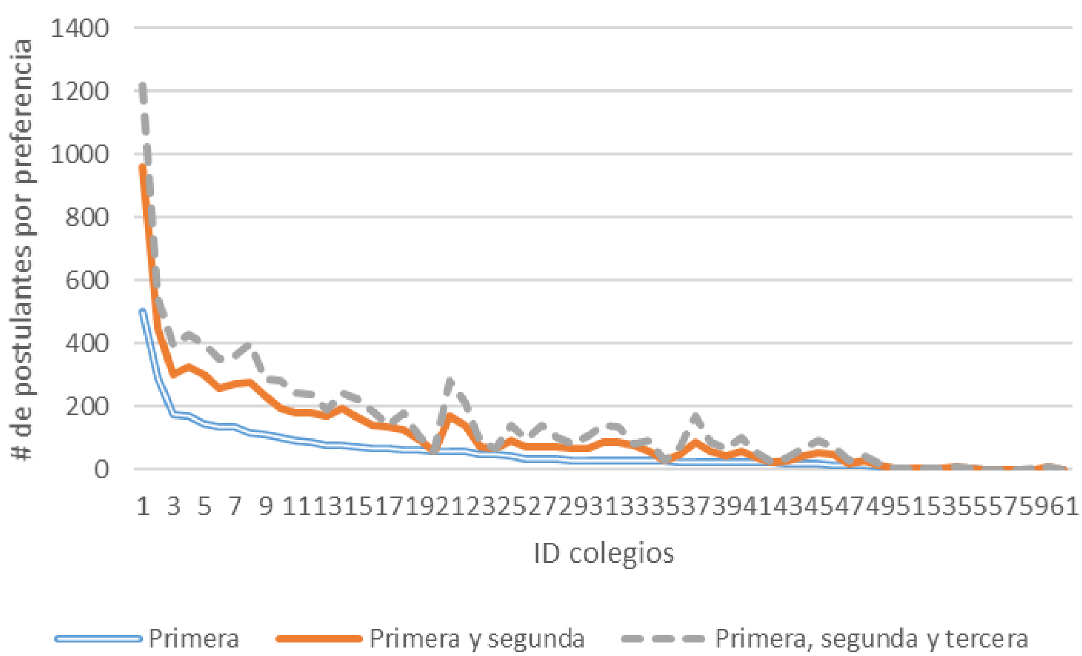

Figura 5.1: Preferencias por RBD

Luego analizamos el resultado anterior para el nivel de pre-kínder, y se ve que la tendencia es similar. Los resultados se muestran en la figura 5.2. Analizamos dicho nivel porque representa más de un tercio de los alumnos totales del sistema (exactamente un 34.8 %) y porque en este nivel los alumnos están en igualdad de condiciones ya que ninguno tiene colegio con matrícula asegurada pues están postulando al colegio por primera vez.

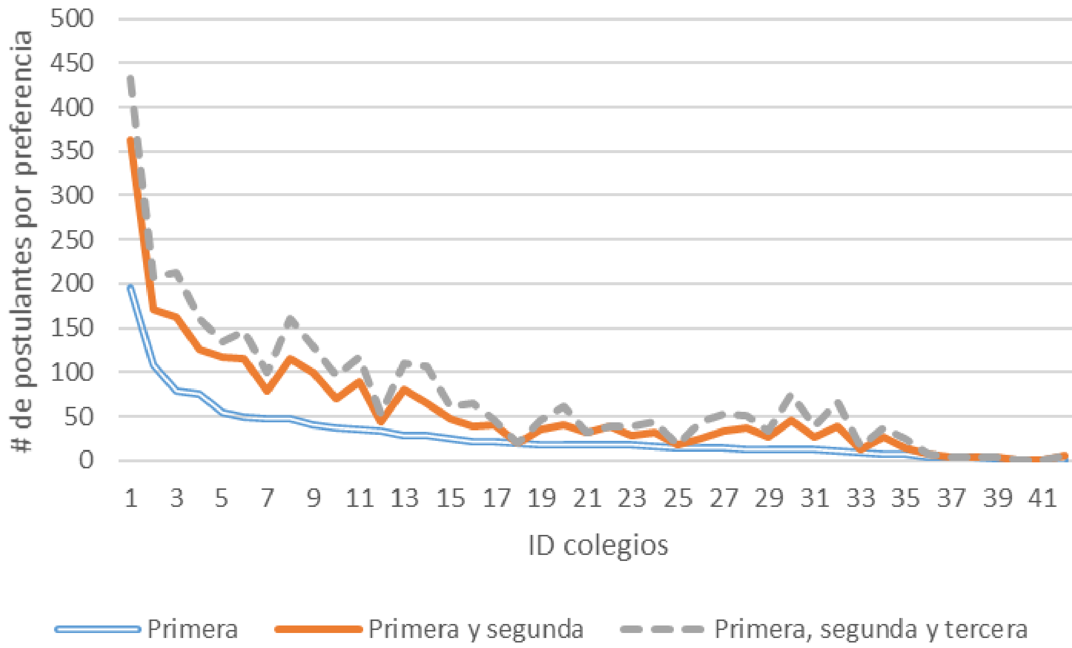

Figura 5.2: Preferencias por RBD, nivel pre-kínder

Los dos gráficos anteriores van en línea con la hipótesis de que las preferencias están correlacionadas, ya que se puede ver que hay un grupo de colegios que son más populares y otros que reciben pocas postulaciones incluso mirando todas las preferencias de las familias.

5.2. Preferential Attachment y Polya's Urn Problem

En esta sección explicamos en qué consisten los modelos de Preferential Attachment, que generan distribuciones de ley de potencia, las cuales podrían representar los datos de este problema. Luego estudiamos cómo las Urnas de Polya generan distribuciones equivalentes y representan el problema de mejor manera.

Una distribución de ley de potencia es una relación entre dos cantidades, donde el cambio en una implica un cambio proporcional en la otra. Independiente de los valores iniciales, una cantidad cambia como potencia de la otra.

Tal como muestran Barabási and Albert [1999], variados sistemas pueden ser modelados como redes con topologías complejas, las cuales tienen una propiedad común: los vértices siguen una distribución de ley de potencia. La principal conclusión del autor es que lo que produce dicho fenómeno es que al modelar la red como nodos y aristas, el sistema se comporta añadiendo constantemente nuevos vértices y a su vez, uniendo dichos vértices con preferencia hacia aquellos que ya están más conectados, es decir, que tienen mayor cantidad de aristas previas.

Dado esto, buscamos modelar el problema de la elección de colegios como un grafo y evaluar si el proceso sigue una distribución de ley de potencia.

Dentro de los principales ejemplos de sistemas que se comportan de esta forma destacan la World Wide Web, las citas de artículos científicos e incluso complejas redes genéticas.

Por su parte, Easley and Kleinberg [2010] desarrollan una idea similar pero basada en el concepto de popularidad. Proponen que ciertos procesos, como la compra de libros, y los link a páginas web, pueden tener correlación y producir la llamada popularidad de ciertos elementos de la red.

El estudio que realizan muestra que, por ejemplo, la cantidad de páginas web con k links apuntando hacia ella es proporcional a $\frac{1}{k^2}$. En general también muestran que en otros casos las cantidades son proporcionales a $\frac{a}{k^c}$ donde a y c son constantes.

Esto es importante ya que estas cantidades decresen lento cuando k aumenta, haciendo factible la existencia de algunos elementos muy populares. Estas funciones, que decrecen según k^c generan una distribución de ley de potencia.

A partir de lo anterior, proponen un algoritmo que permite obtener dicha distribución en los datos. En particular, esto se logra al utilizar modelos conocidos como preferential attachment o rich-get-richer, que se refieren al hecho de que la probabilidad de elegir una opción es directamente proporcional a la cantidad de veces que esa opción ha sido elegida.

Para explicar el modelo básico se utiliza como ejemplo la creación de páginas web, donde cada vez que una página es creada, debe crear un link a una página existente. El modelo funciona como sigue:

Paso j: se crea la página j, la cual crea un link a una página anterior según la siguiente regla:

a) Con probabilidad p, la página j elige una página i a partir de una distribución uniforme.

b) Con probabilidad $1-p$, la página j elige una página i con probabilidad proporcional a la popularidad de i, es decir, a la cantidad de links que la página i ya tiene.

Con el algoritmo anterior pueden recuperar una aproximación determinística que permite identificar el parámetro c con el que se caracteriza la distribución de ley de potencia propuesta.

Para esto, primero definen la variable aleatoria $X_j(t)$ como el número de links a la página j en el tiempo t, y asumen como condición inicial $X_j(j) = 0$. Además definen la probabilidad que el nodo $t+1$ se una a j como:

$$\frac{p}{t} + \frac{(1-p)X_j(t)}{t}.$$

Luego a través de transformar el modelo a uno determinístico, se llega a la conclusión de que la fracción de nodos con k conexiones es aproximadamente:

$$f(k) = \frac{1}{p}\left(\frac{(1-p)}{p}k + 1\right)^{-(1+\frac{1}{1-p})}.$$

En otras palabras, el modelo determinístico predice que la fracción de nodos con k conexiones es proporcional a $k^{-(1+\frac{1}{1-p})}$, es decir, una distribución de ley de potencia con exponente $1+\frac{1}{1-p}$.

En el caso de la elección de colegios por parte de los alumnos, vemos que el problema no es equivalente ya que los alumnos no se unen a otros alumnos, sino que van generando lazos hacia colegios. Es decir, en este caso el problema no puede ser representado simplemente a través de nodos y arcos como en los casos anteriormente mencionados.

Es por esto que buscamos una aproximación diferente a este problema, y mostramos los resultados de Chung et al. [2003], quien generaliza el problema de las Urnas de Polya y encuentra que este proceso también se distribuye como una ley de potencia.

En general, estos procesos consisten en que con probabilidad p se crea una urna con una bola y con probabilidad $1-p$ se pone una nueva bola en alguna de las urnas ya existentes tal que la probabilidad de ponerla en una urna sea proporcional a la cantidad de bolas que la urna tiene.

Esta aproximación sirve para representar mejor el problema de elección de colegios, ya que el problema sería equivalente a crear colegios con un alumno postulando con probabilidad p y con probabilidad $1-p$ el alumno postula a un colegio existente. Cuando postula a un colegio existente lo hace con probabilidad proporcional a la cantidad de postulantes que tiene el colegio.

El problema que se presenta con este modelo es que la creación de colegios puede no ser un elemento realista a la hora de simular el modelo, pues en la práctica el número de colegios está fijo. Abordamos este tema con mayor profundidad en la sección 5.4.

5.3. Ajuste del modelo para Magallanes

La idea de extender los modelos anteriores al caso de elección de colegios se basa en la hipótesis de que hay correlación entre las postulaciones de las familias, y que dado esto, se observa un grupo de colegios populares que definitivamente están sobre demandados mientras que hay otro grupo de colegios que reciben pocas postulaciones.

Creemos que esto puede ocurrir por dos factores, el primero es por características de los establecimientos educacionales, tales como infraestructura, resultados académicos, proyecto educativo u otros. Y el segundo son las postulaciones de otras familias, conocidas o desconocidas para cada postulante, que le agregan o quitan valor a cada colegio.

5.3.1. Modelo con niveles agregados

Para realizar las simulaciones, primero determinamos los parámetros a utilizar en el modelo. En particular, definimos el valor de p_0, que es la probabilidad de crear un colegio nuevo con un alumno.

Comenzamos simulando la primera preferencia de los alumnos, por lo que el valor de p_0 debe ser tal que al realizar la simulación para los 3436 alumnos, el total de colegios creados sea cercano a 63 y de esta forma representar el caso real obtenido de los datos en la Región de Magallanes. Esto se logra tomando $p_0 = \frac{63}{3436} = 1{,}83\,\%$.

Sin embargo, al simular los datos con el parámetro encontrado y realizando 10.000 simulaciones, obtenemos un conjunto muy pequeño de colegios populares. En efecto, 18 colegios se llevan más del 98 % de los postulantes, mientras que en los datos observados son 46 los colegios que cumplen dicha condición. Más aún, en las simulaciones hay un colegio que tiene más de 3000 postulantes, mientras que en los datos el colegio más popular tiene 501 postulantes solamente. En la figura 5.3 comparamos los datos reales con la simulación realizada.

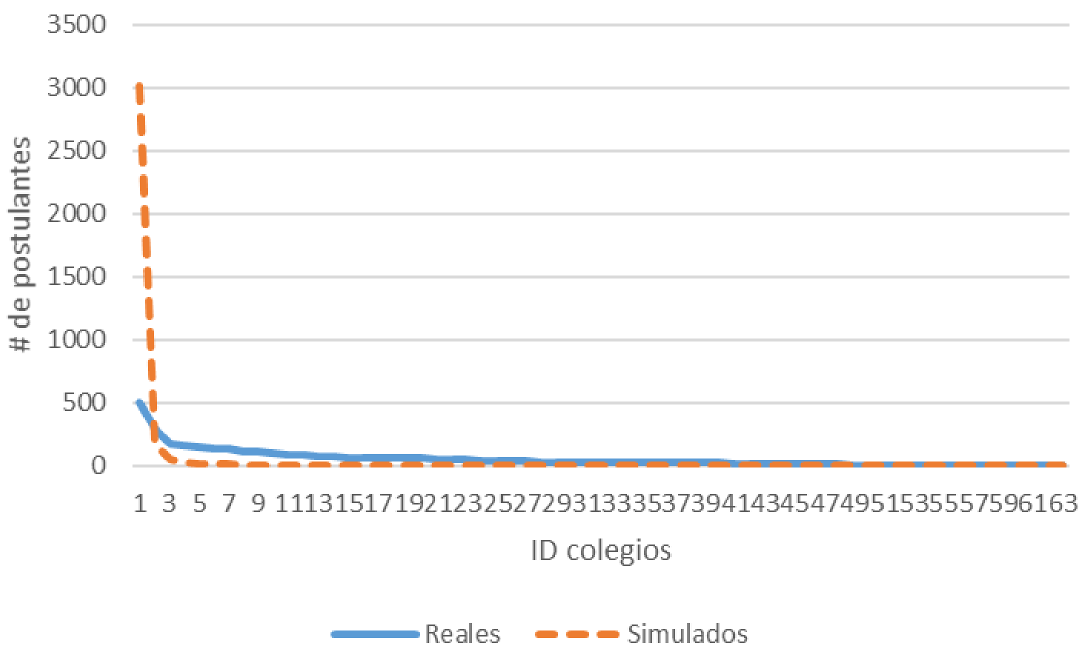

Figura 5.3: Postulantes por RBD simulación con 3436 alumnos y $p_0 = 1{,}83\,\%$

Dado esto, la nueva hipótesis es que si bien existe una correlación, esta no es tan fuerte como para ser modelada como un proceso puramente proporcional a los postulantes anteriores. Por ende buscamos una forma de suavizar la correlación a simular. Una posible solución es cambiar el modelo tal que cumpla lo siguiente:

Paso j: el alumno j postula a un colegio según la siguiente regla:

a) Con probabilidad p_0, se crea un nuevo colegio con un postulante (j).

b) Con probabilidad $1-p_0$, el alumno j postula al colegio i con probabilidad proporcional a los postulantes de i más una constante c. Es decir, con probabilidad $\frac{postulantes_{\mathrm{i}}+c}{j-1+c\cdot colegios}$ donde colegios es la cantidad de colegios existentes hasta el paso $j-1$ y $postulantes_{\mathrm{i}}$ son la cantidad de alumnos postulando a i en el paso $j-1$.

Buscamos demostrar que este modelo entrega una distribución de ley de potencia y estimar el exponente que caracteriza dicha distribución. Para esto utilizamos como base el trabajo realizado por Chung et al. [2003], en particular en los Teoremas 4.1 y 4.2.

Teorema 5.1 *El modelo propuesto, en el que la probabilidad de elegir un colegio es proporcional a los postulantes del colegio más una constante, sigue una distribución de ley de potencia.*

DEMOSTRACIÓN. Consideremos el instante en que ya han postulado $n-1$ alumnos al sistema y estamos considerando la postulacion del n-ésimo estudiante. Definamos los siguientes términos:

- $p_{\mathrm{i}}(n)$: probabilidad de que el n-ésimo alumno postule a un colegio con i postulantes ($\mathrm{i} \geq 1$).

- p_0: probabilidad que el n-ésimo alumno cree un nuevo colegio.

- $F_{\mathrm{i}}(n)$: número de colegios con i postulantes antes de asignar al n-ésimo estudiante.

- $N(n)$: número total de colegios antes de asignar al n-ésimo estudiante.

- $f_{\mathrm{i}}(n)$: fracción de colegios con i postulantes antes de asignar al n-ésimo estudiante.

- $E_{\mathrm{i}}(n)$: cambio en el número esperado de colegios con i postulantes al asignar al n-ésimo estudiante.

- $\hat{F}_{\mathrm{i}}(n)$: variable aleatoria del número de colegios con i postulantes después de asignar al n-ésimo alumno.

- $\hat{N}(n)$: variable aleatoria del número total de colegios después de asignar al n-ésimo alumno.

- $\hat{f}_{\mathrm{i}}(n)$: variable aleatoria de la fracción de colegios con i postulantes después de asignar al n-ésimo alumno.

- $A_n = \begin{cases} \mathrm{i} & \text{si n postula a colegio con i alumnos} \\ 0 & \text{si crea colegio nuevo} \end{cases}$

- K, c: constantes positivas.

A partir de esto tenemos que: $f_{\mathrm{i}}(n) = \frac{F_{\mathrm{i}}(n)}{N(n)}$, $\mathbb{E}(\hat{F}_{\mathrm{i}}(n)) = F_{\mathrm{i}}(n) + E_{\mathrm{i}}(n)$, $E_{\mathrm{i}}(n) = p_{\mathrm{i}-1}(n) - p_{\mathrm{i}}(n)$ y $p_{\mathrm{i}}(n) = \frac{(1-p_0)\cdot f_{\mathrm{i}}(n)\cdot(\mathrm{i}+c)}{\sum_{j=1}^{\infty} f_j(n)\cdot(j+c)}$.

En forma intuitiva postulamos que en el estado estacionario se debería verificar que $f_\mathrm{i} = \mathbb{E}(\hat{f}_\mathrm{i})$. Formalmente esto lo podemos demostrar usando la siguiente igualdad:

$$\hat{f}_\mathrm{i}(n) = \frac{\hat{F}_\mathrm{i}(n)}{\hat{N}(n)} = \frac{F_\mathrm{i}(n)}{\hat{N}(n)} + \frac{\mathbb{1}_{\{A_n=\mathrm{i}-1\}}}{\hat{N}(n)} - \frac{\mathbb{1}_{\{A_n=\mathrm{i}\}}}{\hat{N}(n)}$$

Luego tomando esperanza sobre las variables aleatorias tenemos que:

$$\begin{aligned}\mathbb{E}(\hat{f}_\mathrm{i}(n)) &= \frac{F_\mathrm{i}(n)}{N(n)} \cdot (1-p_0) + \frac{F_\mathrm{i}(n)}{N(n)+1} \cdot p_0 + \frac{p_{\mathrm{i}-1}(n)}{N(n)} - \frac{p_\mathrm{i}(n)}{N(n)} \\ &= f_\mathrm{i}(n) \cdot (1-p_0) + f_\mathrm{i}(n) \cdot \frac{N(n)}{N(n)+1} \cdot p_0 + \frac{p_{\mathrm{i}-1}(n) - p_\mathrm{i}(n)}{N(n)}\end{aligned}$$

Tomamos el límite cuando $n \to \infty$ y llegamos a que $\mathbb{E}(\hat{f}_\mathrm{i}(n)) = f_\mathrm{i}(n)$. En la demostración hemos utilizado las siguientes identidades:

$$\lim_{n\to\infty} \frac{N(n)}{N(n)+1} = 1,$$

$$\lim_{n\to\infty} \frac{p_{\mathrm{i}-1}(n) - p_\mathrm{i}(n)}{N(n)} = 0.$$

Con esto, pasamos a desarrollar $\mathbb{E}(\hat{f}_\mathrm{i}(n))$:

Para el caso en que $\mathrm{i} \geq 2$:

$$\begin{aligned}\mathbb{E}(\hat{f}_\mathrm{i}(n)) = \mathbb{E}\left(\frac{\hat{F}_\mathrm{i}(n)}{\hat{N}(n)}\right) &= \mathbb{E}\left(\frac{\hat{F}_\mathrm{i}(n)}{\hat{N}(n)} | A_n \neq 0\right)\mathbb{P}(A_n \neq 0) + \mathbb{E}\left(\frac{\hat{F}_\mathrm{i}(n)}{\hat{N}(n)} | A_n = 0\right)\mathbb{P}(A_n = 0) \\ &= \frac{\mathbb{E}(\hat{F}_\mathrm{i}(n) | A_n \neq 0)}{N(n)} \cdot (1-p_0) + \frac{F_\mathrm{i}(n)}{N(n)+1} \cdot p_0 \\ &= \frac{F_\mathrm{i}(n) + \frac{E_\mathrm{i}(n)}{(1-p_0)}}{N(n)} \cdot (1-p_0) + \frac{F_\mathrm{i}(n)}{N(n)} \frac{N(n)}{(N(n)+1)} \cdot p_0 \\ &= \frac{F_\mathrm{i}(n)}{N(n)} \cdot (1-p_0) + \frac{E_\mathrm{i}(n) \cdot (1-p_0)}{N(n) \cdot (1-p_0)} + \frac{F_\mathrm{i}(n)}{N(n)} \frac{N(n) \cdot p_0}{(N(n)+1)} \\ &= f_\mathrm{i}(n) \cdot (1-p_0) + \frac{E_\mathrm{i}(n)}{N(n)} + f_\mathrm{i} \cdot p_0 \frac{N(n)}{N(n)+1} \\ &= f_\mathrm{i}(n) - \frac{f_\mathrm{i}(n) \cdot p_0}{N(n)+1} + \frac{E_\mathrm{i}(n)}{N(n)}\end{aligned}$$

Al igualar $\mathbb{E}(\hat{f}_\mathrm{i}(n)) = f_\mathrm{i}(n)$ y tomar límite cuando $n \to \infty$ obtenemos que:

$$\frac{E_\mathrm{i}(n)}{f_\mathrm{i}(n)} = \frac{N(n)}{N(n)+1} \cdot p_0 = p_0$$

Para el caso $\mathrm{i} = 1$:

$$\begin{aligned}
\mathbb{E}(\hat{f}_{\mathrm{i}}(n)) = \mathbb{E}\left(\frac{\hat{F}_{\mathrm{i}}(n)}{\hat{N}(n)}\right) &= \mathbb{E}\left(\frac{\hat{F}_{\mathrm{i}}(n)}{\hat{N}(n)}|A_n \neq 0\right)\mathbb{P}(A_n \neq 0) + \mathbb{E}\left(\frac{\hat{F}_{\mathrm{i}}(n)}{\hat{N}(n)}|A_n = 0\right)\mathbb{P}(A_n = 0) \\
&= \frac{\mathbb{E}(\hat{F}_{\mathrm{i}}(n)|A_n \neq 0)}{N(n)} \cdot (1-p_0) + \frac{\mathbb{E}(\hat{F}_{\mathrm{i}}(n)|A_n = 0)}{N(n)+1} \cdot p_0 \\
&= \frac{F_{\mathrm{i}}(n) + \frac{E_{\mathrm{i}}(n)-p_0}{(1-p_0)}}{N(n)} \cdot (1-p_0) + \frac{F_{\mathrm{i}}(n)+1}{N(n)+1} \cdot p_0 \\
&= \frac{F_{\mathrm{i}}(n)}{N(n)} \cdot (1-p_0) + \frac{E_{\mathrm{i}}(n)-p_0}{N(n)} + \frac{F_{\mathrm{i}}(n)\cdot p_0}{N(n)+1} + \frac{p_0}{N(n)+1} \\
&= f_{\mathrm{i}}(n) \cdot (1-p_0) + \frac{E_{\mathrm{i}}(n)-p_0}{N(n)} + \frac{F_{\mathrm{i}}(n)\cdot p_0}{N(n)+1} + \frac{p_0}{N(n)+1}
\end{aligned}$$

Al igualar $\mathbb{E}(\hat{f}_{\mathrm{i}}(n)) = f_{\mathrm{i}}(n)$ y tomar límite cuando $n \to \infty$ obtenemos que:

$$E_{\mathrm{i}}(n) = \frac{N(n)}{N(n)+1} \cdot f_{\mathrm{i}}(n) \cdot p_0 + \frac{p_0}{N(n)+1} = f_{\mathrm{i}}(n) \cdot p_0$$

Porque $\text{lím}_{n\to\infty} \frac{p_0}{N(n)+1} = 0$.

Lo cual es equivalente al caso anterior y concluimos que en el largo plazo $\forall \mathrm{i} : \frac{E_{\mathrm{i}}(n)}{f_{\mathrm{i}}(n)} = p_0$.

A partir de esto, calculamos E_{i} en función de p_{i} lo que nos permite encontrar una recurrencia para f_{i}, omitimos el subíndice n y asumimos que estamos en el largo plazo:

$$E_{\mathrm{i}} = p_{\mathrm{i}-1} - p_{\mathrm{i}} = \frac{(1-p_0)(f_{\mathrm{i}-1}\cdot(\mathrm{i}-1+c) - f_{\mathrm{i}}\cdot(\mathrm{i}+c))}{\sum_{j=1}^{\infty} f_j \cdot (j+c)}$$

Igualando $E_{\mathrm{i}} = f_{\mathrm{i}} \cdot p_0$ y despejando llegamos a que:

$$f_{\mathrm{i}} = \left(\frac{\mathrm{i}-1+c}{\mathrm{i}+c+K}\right) \cdot f_{\mathrm{i}-1}$$

Donde $K = \frac{p_0\cdot\sum_{j=1}^{\infty} f_j\cdot(j+c)}{(1-p_0)} = \frac{1+c\cdot p_0}{1-p_0}$. Resolviendo la recurrencia obtenemos que:

$$\begin{aligned}
f_{\mathrm{i}} \propto \prod_{j=2}^{\mathrm{i}} \frac{j-1+c}{j+c+K} &\propto \frac{\Gamma(c+\mathrm{i})}{\Gamma(K+c+\mathrm{i}+1)} \\
&\propto \mathrm{i}^{(c+\mathrm{i}-K-c-\mathrm{i}-1)} \propto \mathrm{i}^{-(K+1)} \\
&\propto \mathrm{i}^{-\left(1+\frac{1+c\cdot p_0}{1-p_0}\right)}
\end{aligned}$$

Lo que equivale a una distribución de ley de potencia con exponente $\left(1+\frac{1+c\cdot p_0}{1-p_0}\right)$. $\square$

Intuitivamente, la constante c representa la popularidad del colegio posterior a las postulaciones, pues si sabemos de antemano que en promedio habrán 63 colegios es porque estos colegios tienen algún grado de popularidad. Visto de otra forma, el colegio gana cierta popularidad por el hecho de que cuando los alumnos sean asignados, el curso tendrá una cantidad de alumnos y esto se asegura porque el colegio tiene historia y tiene alumnos en otros niveles.

Para el caso de la primera preferencia, se busca ajustar el parámentro c de manera de minimizar los errores al cuadrado. Esto se realiza para la mitad de los alumnos, seleccionados de manera aleatoria, con el fin de encontrar la constante y luego evaluar si sirve para simular el problema completo.

Realizando 10.000 simulaciones con $p_0 = \frac{63}{1674} = 0{,}037$, obtenemos que con $c = 43$ los datos simulados son los más similares al caso real, lo que se muestra en la figura 5.4.

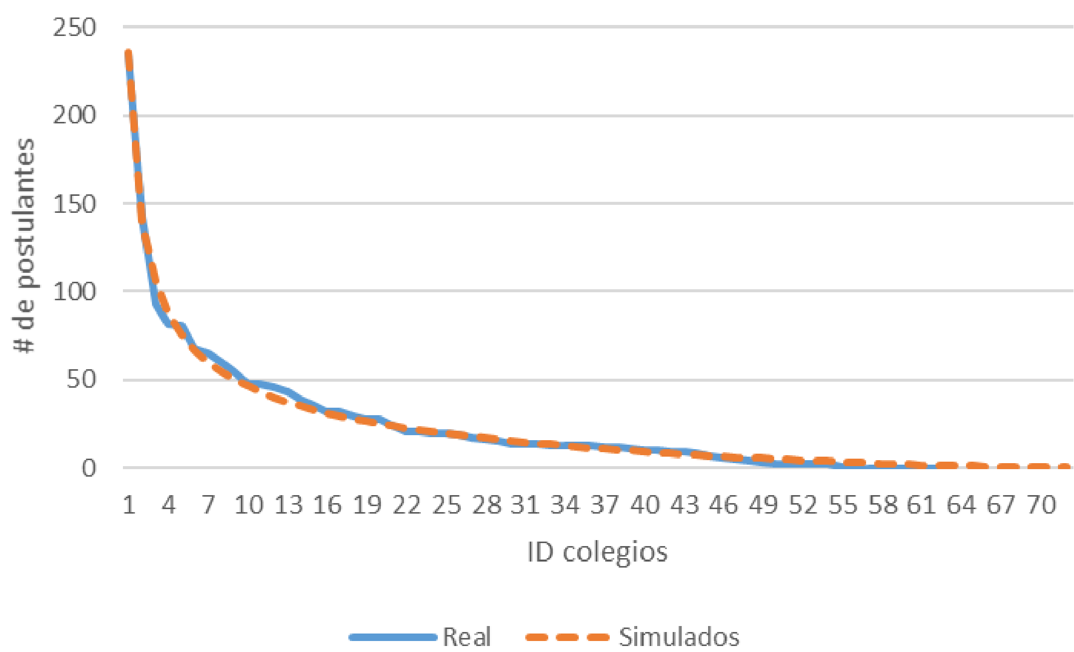

Figura 5.4: Postulantes por RBD simulación 1674 alumnos, $c = 43$ y $p_0 = 3{,}7\,\%$

Luego se simula para el total de alumnos con $p_0 = 0{,}018$ para crear en promedio 63 colegios. Los resultados de las 10.000 simulaciones se muestran en la figura 5.5 donde vemos que no se ajusta de manera óptima a los datos reales.

Dado lo anterior, se busca la constante que ajusta los datos totales, obteniendo que es $c = 95$, muy cercano al doble de la constante encontrada para la mitad de la muestra. Los resultados se muestran en la figura 5.6.

Esto se explica porque el término $c \cdot p_0$ de la expresión de la fracción de colegios $f_{\mathrm{i}} = \mathrm{i}^{-\left(1+\frac{1+c \cdot p_0}{1-p_0}\right)}$ se mantiene constante en ambos casos cuando variamos p_0 reduciéndolo a la mitad y c aumentándolo al doble. Esta constante $c \cdot p_0$ es el término que finalmente se suma al exponente en comparación con una distribución de ley de potencia clásica.

Desde un punto de vista intuitivo, podemos ver que a medida que aumenta c la distribución es menos pronunciada, es decir, las probabilidades de postular a cada colegio se hacen

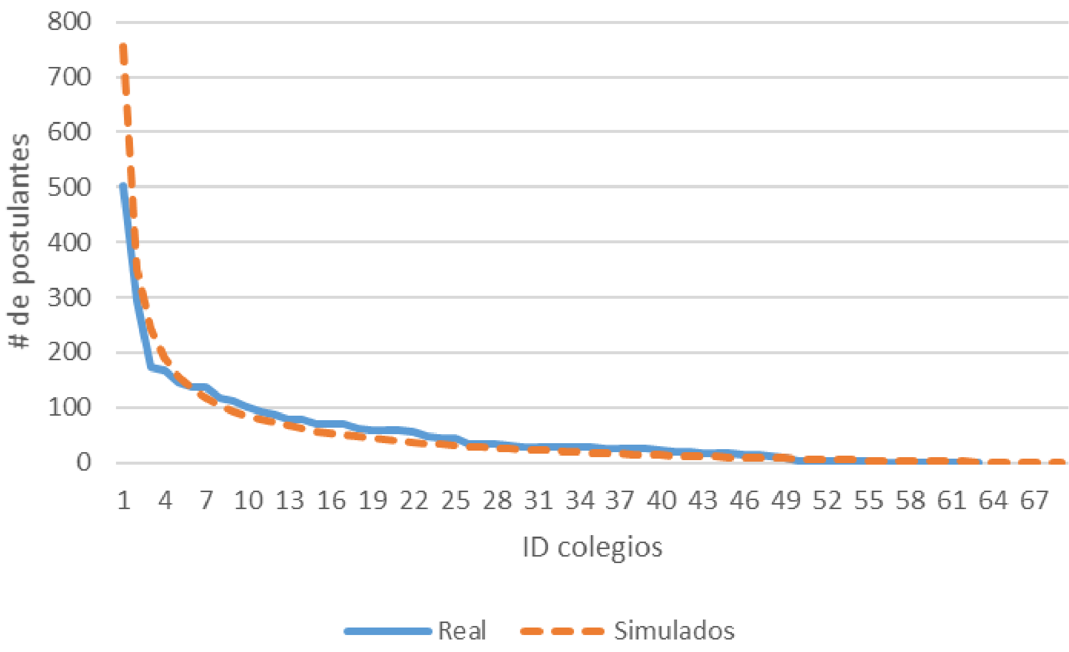

Figura 5.5: Postulantes por RBD simulación con 3436 alumnos, $c = 43$ y $p_0 = 1{,}83\,\%$

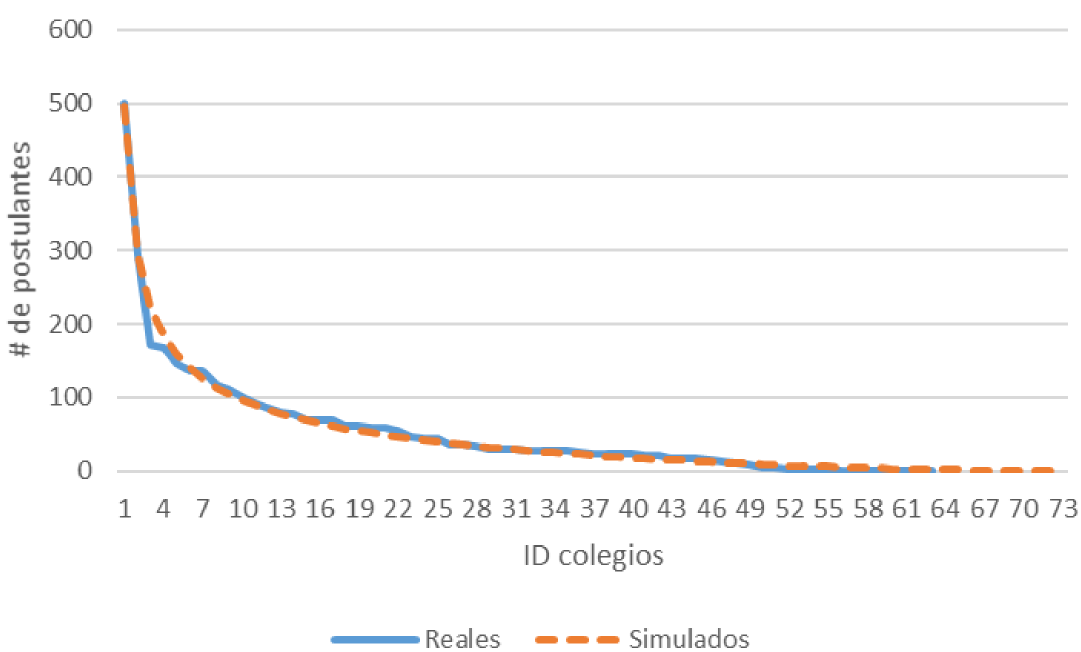

Figura 5.6: Postulantes por RBD simulación con 3436 alumnos, $c = 95$ y $p_0 = 1{,}83\,\%$

más similares. Esto indica que al aumentar el número de alumnos, todos los colegios ganan popularidad por el hecho de que al haber más alumnos, una mayor cantidad de colegios tendrá muchos alumnos asignados.

5.3.2. Modelo Pre Kínder

En el caso del primer nivel de ingreso, calculamos el valor de p_0 que permite crear 42 colegios en promedio como $p_0 = \frac{42}{1195} = 3{,}51\,\%$. Al realizar las simulaciones vemos que ocurre lo mismo que en el caso con los niveles agregados, lo que mostramos en la figura 5.7. Por ende, buscamos una constante que permita modelar las preferencias de las familias minimizando los errores al cuadrado.

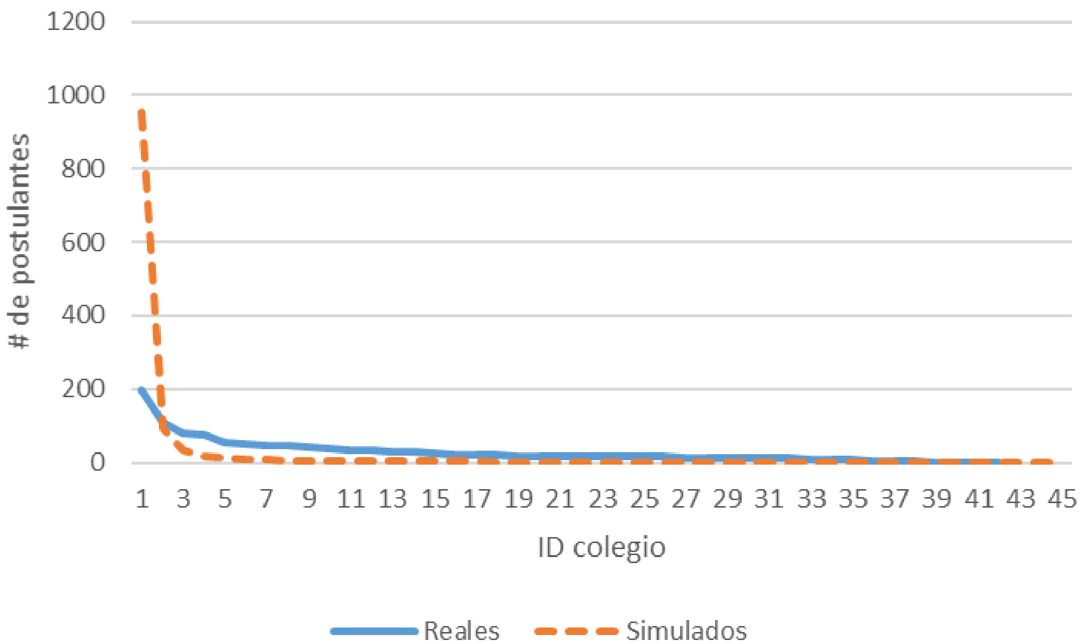

Figura 5.7: Postulantes por RBD simulación con 1195 alumnos y $p_0 = 3{,}51\,\%$

Primero, tomamos la mitad de la muestra, es decir, 586 alumnos y calculamos $p_0 = \frac{42}{586} = 0{,}072$. Realizando simulaciones y minimizando los errores al cuadrado, llegamos a que la constante $c = 31$. Los resultados obtenidos se pueden ver en la figura 5.8

Luego simulando con la constante $c = 31$ para el total de los alumnos vemos que las simulaciones no reflejan el caso real de manera óptima. Esto se ve en la figura 5.9.

Entonces, calculamos el valor de la probabilidad p_0 de crear un colegio nuevo con todos los alumnos, obteniendo que $p_0 = \frac{42}{1195} = 3{,}51\,\%$.

Luego realizamos las simulaciones de postulaciones en primera preferencia, obteniendo que con $c = 64$ se minimiza el error cuadrático entre los datos reales y los simulados. En la figura 5.10 podemos ver los resultados de la simulación en comparación con los datos reales.

Vemos que nuevamente la constante en el caso con todos los alumnos es cercana al doble del valor de la constante encontrada para la mitad de la muestra.

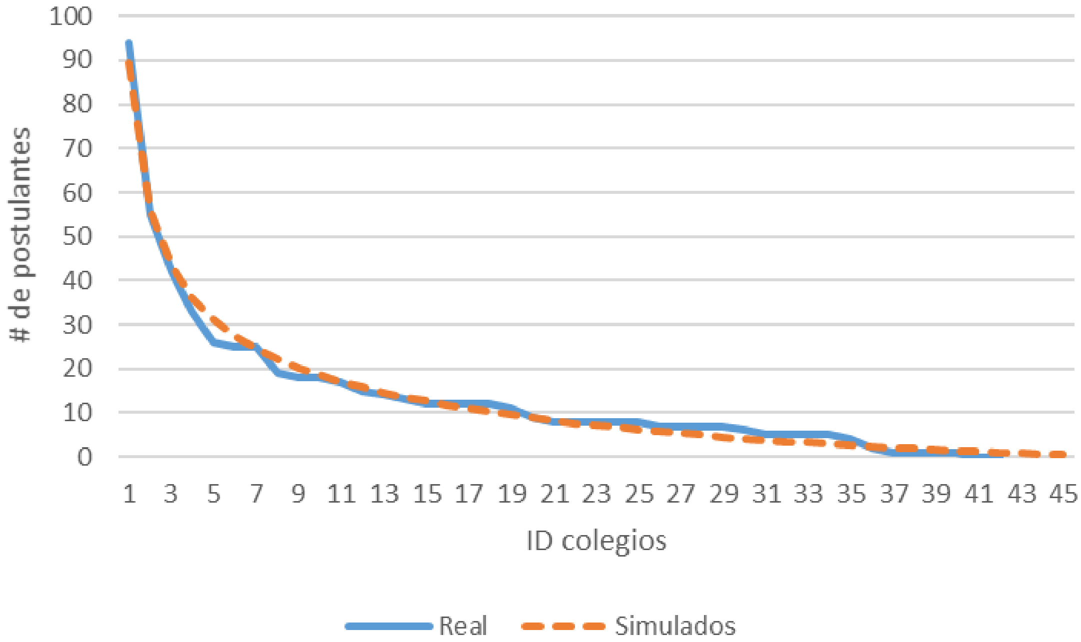

Figura 5.8: Postulantes por RBD simulación con 586 alumnos, $c = 31$ y $p_0 = 7{,}2\,\%$

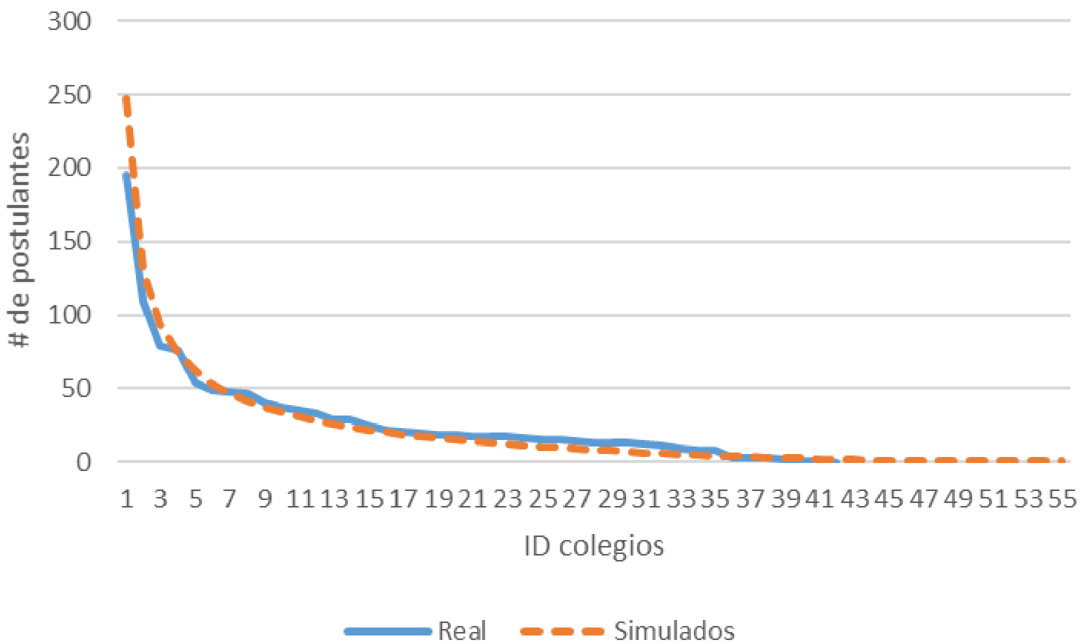

Figura 5.9: Postulantes por RBD simulación con 1195 alumnos, $c = 31$ y $p_0 = 3{,}51\,\%$

5.4. Modelo con colegios fijos

El hecho de crear colegios con probabilidad positiva podría ser un elemento poco realista del modelo, por ende, proponemos realizar un modelo en que en vez de crear un nuevo colegio, el alumno postula con probabilidad uniforme a uno de los colegios que ya existen.

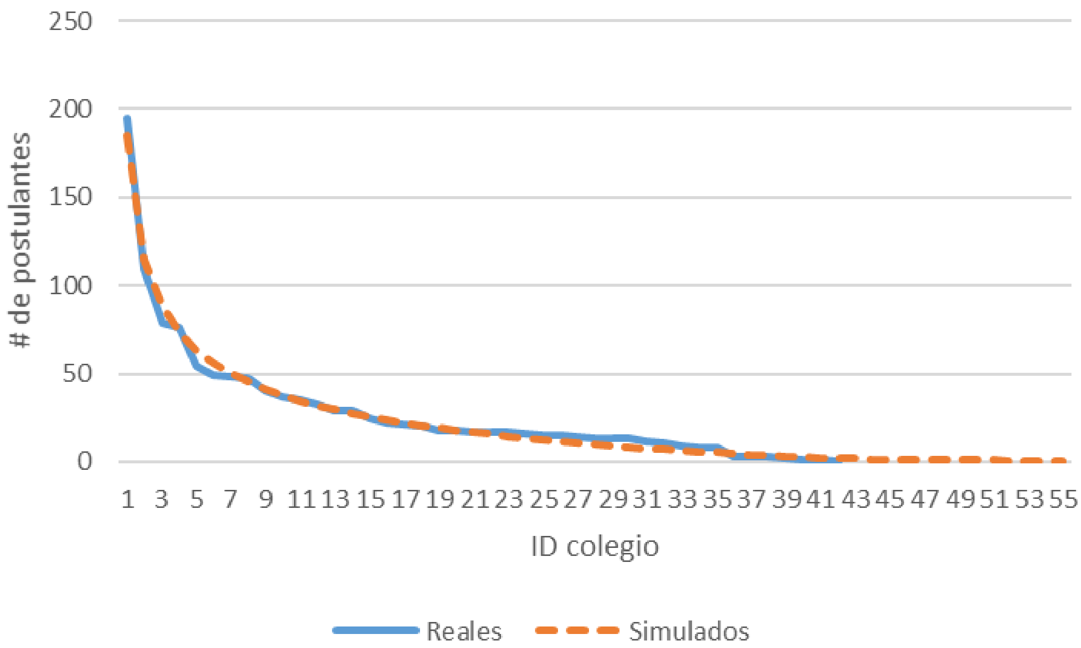

Figura 5.10: Postulantes por RBD simulación con 1195 alumnos, $c = 64$ y $p_0 = 3{,}51\,\%$

El proceso que proponemos queda como sigue:

Paso j: el alumno j postula a uno de los M colegios según la siguiente regla:

a) Con probabilidad p, el alumno j postula al colegio i con probabilidad $\frac{1}{M}$.

b) Con probabilidad $1 - p$, el alumno j postula al colegio i con probabilidad proporcional a los postulantes de i. Es decir, con probabilidad $\frac{postulantes_{\mathrm{i}}}{j-1}$ donde $postulantes_{\mathrm{i}}$ son la cantidad de alumnos postulando a i actualmente.

Teorema 5.2 *El modelo propuesto, en el que la probabilidad de elegir un colegio tiene una componente proporcional y una componente uniforme, entrega una distrubución uniforme de los datos en el largo plazo.*

Demostración. Definiciones:

- N_n^{i}: número de estudiantes que tiene el colegio i en el tiempo n.

- M: número total de colegios (fijo).

- X_n^{i}: fracción de estudiantes que tiene el colegio i en n.

- S_n: colegio al cual postula el alumno n.

A partir de esto obtenemos que: $X_n^{\mathrm{i}} = \frac{N_n^{\mathrm{i}}}{n}$.

Con lo anterior, tenemos que:

$$\mathbb{1}_{\{S_{n+1}=\mathrm{i}\}} = \begin{cases} 1 & \text{con probabilidad } \frac{p}{M} + (1-p)X_n^{\mathrm{i}} \\ 0 & \text{si no} \end{cases}$$

lo que nos entrega que:

$$X_{n+1}^{\mathrm{i}} = \begin{cases} \frac{N_n^{\mathrm{i}}}{n+1} = \frac{X_n^{\mathrm{i}} \cdot n}{n+1} & \text{si } \mathbb{1}_{\{S_{n+1}=\mathrm{i}\}} = 0 \\ \frac{N_n^{\mathrm{i}}+1}{n+1} = \frac{X_n^{\mathrm{i}} \cdot n+1}{n+1} & \text{si } \mathbb{1}_{\{S_{n+1}=\mathrm{i}\}} = 1 \end{cases}$$

Reescribiendo nos queda que:

$$\begin{aligned} X_{n+1}^{\mathrm{i}} &= X_n^{\mathrm{i}} + \frac{1}{n+1}\left[\mathbb{1}_{\{S_{n+1}=\mathrm{i}\}} - X_n^{\mathrm{i}}\right] \\ &= X_n^{\mathrm{i}} + \frac{1}{n+1}\left[\frac{p}{M} + (1-p)X_n^{\mathrm{i}} - X_n^{\mathrm{i}} + U_{n+1}^{\mathrm{i}}\right] \end{aligned}$$

A partir de esto tenemos que:

$$\frac{X_{n+1}^{\mathrm{i}} - X_n^{\mathrm{i}}}{\frac{1}{n+1}} = p\left[\frac{1}{M} - X_n^{\mathrm{i}} + \hat{U}_{n+1}^{\mathrm{i}}\right]$$

Donde vemos que el ruido es martingala y acotado.

Analizando la ecuación diferencial $\dot{X}^{\mathrm{i}} = p\left[\frac{1}{M} - X^{\mathrm{i}}\right]$ vemos que tiene un único atractor global que es $A = \frac{1}{M}$. El conjunto límite de X_n, $L(X_n)$ es un ICT de la ecuación diferencial.

Según los resultados de Benaïm [1999], dicho atractor global debe ser el único ICT. Y por ende, casi seguro, $L(X_n) \subseteq A$.

Con lo que concluimos que $X^{\mathrm{i}} = \frac{1}{M}$.

Es decir, que la fracción de estudiantes que tiene el colegio i en el largo plazo es igual para todos los colegios. □

Es por esto que se opta por utilizar el modelo en el que el número de colegios es variable ya que, dados los datos, la hipótesis es que hay correlación en las postulaciones, lo que contradice que en el largo plazo todos los colegios reciban la misma cantidad de postulantes.

Capítulo 6

Conclusiones

6.1. Conclusiones generales

El problema de asignación escolar es un tema que actualmente está siendo ampliamente estudiado en diversas ciudades alrededor del mundo, cada una de las cuales introduce elementos específicos necesarios para implementar el algoritmo de acuerdo a sus requerimientos.

Este trabajo de tesis aborda el problema de Chile, especificando la forma en que se desarrolla e implementa el algoritmo para realizar la asignación escolar y de qué manera se mantienen las propiedades deseables del algoritmo.

En particular, vemos que hay casos en que la estrategia óptima de los alumnos con hermanos menores puede ser no declarar sus verdaderas preferencias, lo cual se da cuando los hermanos hacen postulación en bloque y no quedan en el mismo colegio. Si esto ocurre, también se pierde estabilidad bajo la definición de entregada en la sección 2.

Sobre el modelamiento de las preferencias, extendimos el modelo de la Urnas de Polya y los procesos de tipo Preferential Attachment para el caso de colegios y postulantes, lo cual se traduce en la posibilidad de simular instancias de mayor tamaño.

La extensión que realizamos al modelo de las Urnas de Polya fue agregar una constante que le diera mayor probabilidad de ser elegido a cualquier colegio que se creara en un iteración. Esto con el fin de no otorgar una diferencia de popularidad tan grande entre colegios.

6.2. Trabajo futuro

En relación a los análisis de resultados, vemos que en el caso de la cuota de alumnos prioritarios las respuestas no son concluyentes y por ende proponemos implementar el mecanismo con reservas minoritarias como proponen Hafalir et al. [2013] y evaluar si esta implementación tiene un impacto positivo en los resultados. Cabe señalar que el impacto positivo se considera como el cumplimiento de las cuotas de alumnos prioritarios en los colegios y no como una mejora en la asignación.

En cuanto a las reglas de rompimiento de empates a utilizar, en la sección 4.3.2 explicamos las ventajas y desventajas de utilizar STB y MTB, y explicamos la regla de rompimiento de empates híbrida propuesta por Ashlagi and Nikzad [2015]. En este sentido, se propone implementar el algoritmo utilizando esta regla de desempate y evaluar cuál es el umbral que separa colegios populares y no populares en el caso de Chile.

Por otro lado, proponemos extender el modelo de preferencias al caso donde hay más de un tipo de alumno. Esto implica tener, por ejemplo, alumnos de tipo A que solo afectan las postulaciones de otros alumnos tipo A y alumnos tipo B que influyen en otros alumnos tipo B.

La idea detrás de esto es que no todos los alumnos de una región se comportan igual. En particular, estos comportamientos pueden darse por varios factores como distintas creencias religiosas, habiendo alumnos que postulan a colegios católicos y otros que no, o por la distribución de la población en la región, pudiendo ver que los alumnos postulan a las ciudades en las que viven y no hay una postulación generalizada en la región completa.

Otra línea de desarrollo del modelo de preferencias, es utilizar modelos estructurales, lo cual permitiría entender los factores que llevan a una determinada familia a tomar decisiones. Estos factores pueden incluir la distancia al colegio, la infraestructura del mismo, los resultados del Simce y la PSU, o características propias de la familia, como el nivel de ingreso o la religión entre otros.

Bibliografía

Atila Abdulkadiroglu and Tayfun Sönmez. School choice: A mechanism design approach. *The American Economic Review*, 93(3):729–747, 2003.

Atila Abdulkadiroglu, Parag Pathak, Alvin E Roth, and Tayfun Sonmez. Changing the boston school choice mechanism. Technical report, National Bureau of Economic Research, 2006.

Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the nyc high school match. *The American Economic Review*, 99(5):1954–1978, 2009.

Itai Ashlagi and Afshin Nikzad. What matters in tie-breaking rules? how competition guides design. *Unpublished working paper*, 2015.

Itai Ashlagi, Afshin Nikzad, and Assaf I Romm. Assigning more students to their top choices: A tiebreaking rule comparison. 2015.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. ISSN 0036-8075. doi: 10.1126/science.286.5439.509. URL `http://science.sciencemag.org/content/286/5439/509`.

Michel Benaïm. Dynamics of stochastic approximation algorithms. In *Seminaire de probabilites XXXIII*, pages 1–68. Springer, 1999.

Fan Chung, Shirin Handjani, and Doug Jungreis. Generalizations of polya's urn problem. *Annals of combinatorics*, 7(2):141–153, 2003.

David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

Aytek Erdil and Haluk Ergin. What's the matter with tie-breaking? improving efficiency in school choice. *The American Economic Review*, 98(3):669–689, 2008. ISSN 00028282. URL `http://www.jstor.org/stable/29730091`.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Pieter A Gautier, Monique De Haan, Bas Van der Klaauw, and Hessel Oosterbeek. Eerste analyse matching en loting voortgezet onderwijs amsterdam 2016*. *Evaluation Paper, URL: http://www. verenigingosvo. nl/wp-content/uploads/2013/09/Rapport-loting-*

en-matching-2016-beschrijvend-rapport. pdf, R etrieved June, 14:2016, 2016.

Isa E Hafalir, M Bumin Yenmez, and Muhammed A Yildirim. Effective affirmative action in school choice. *Theoretical Economics*, 8(2):325–363, 2013.

Alvin E Roth. The economics of matching: Stability and incentives. *Mathematics of operations research*, 7(4):617–628, 1982.

Alvin E Roth. Misrepresentation and stability in the marriage problem. *Journal of Economic theory*, 34(2):383–387, 1984.