MODULARIDAD EN REDES DE APAREAMIENTO EN PLANTAS POLINIZADAS POR INSECTOS BAJO FACTORES ABIOTICOS VARIABLES

Por

Leoncio Rodriguez Quiñones

Tesis sometida en cumplimiento parcial de los requerimientos para el grado de

MAESTRÍA EN CIENCIAS

en

MATEMÁTICA APLICADA

UNIVERSIDAD DE PUERTO RICO RECINTO UNIVERSITARIO DE MAYAGÜEZ

November, 2010

Aprobada por:

Juan Romero, Ph.D Miembro, Comité Graduado

Edgardo Lorenzo, Ph.D Miembro, Comité Graduado

Luis Gordillo, Ph.D Presidente, Comité Graduado

Mario Córdova, Ph.D Representante de Estudios Graduados

Silvestre Colón, M.Sc Director del Departamento Fecha

Fecha

Fecha

Fecha

Fecha

Resumen de Disertación Presentado a Escuela Graduada de la Universidad de Puerto Rico como requisito parcial de los Requerimientos para el grado de Maestría en Ciencias

MODULARIDAD EN REDES DE APAREAMIENTO EN PLANTAS POLINIZADAS POR INSECTOS BAJO FACTORES ABIOTICOS VARIABLES

Por

Leoncio Rodriguez Quiñones

November 2010

Consejero: Luis Gordillo Departamento: Ciencias Matemáticas

Este trabajo está enfocado en el estudio de la propiedad estructural de la modularidad en redes complejas que surgen cuando los eventos de polinización son modelados usando un kernel de dispersión de cola gruesa. Los valores de la modularidad son obtenidos desde diferentes redes cuyos nodos tienen dos configuraciones espaciales diferentes. Estos valores de la modularidad son usados para entender algunos aspectos acerca del flujo genético en una población de árboles y cómo son las interacciones en términos de apareamiento entre los individuos de la población. Además, esta información acerca de la modularidad de las redes de apareamiento es importante porque permite conocer algunos aspectos de carácter ecológico y biológico que ocurren dentro de la población de árboles. Este conocimiento puede ser usado como base para el manejo de la conservación de especies en peligro de extinción. Abstract of Dissertation Presented to the Graduate School of the University of Puerto Rico in Partial Fulfillment of the Requirements for the Degree of Master of Sciences

MODULARITY IN MATING NETWORKS OF INSECT-POLLINATED PLANTS UNDER VARIABLE ABIOTIC FACTORS

By

Leoncio Rodriguez Quiñones

November 2010

Chair: Luis Gordillo Major Department: Mathematical Sciences

This work is focused in the study of the structural property of modularity in complex networks that arise when the pollination events are modeled using a thick dispersal kernel. The values of modularity are obtained from differents networks whose nodes have two differents spatial configurations. These values of the modularity are used to understand some aspects about the genetic flow in a population of trees and how are the interactions in mating terms between individuals of the population. In addition, this information about the modularity of the mating networks is important because it allows to know about some ecological and biological aspects that occur inside the population of trees. This knowledge can be used as foundation for the management of the conservation of the species at risk of extinction. Copyright © 2010

por

Leoncio Rodriguez Quiñones

A mis viejos, hermanos y sobrinos.

AGRADECIMIENTOS

A Dios, mi familia, mis amigos y especialmente a los Doctores Luis Gordillo, Luis Cáceres, Juan Ortiz, Juan Romero y Edgardo Lorenzo.

TABLA DE CONTENIDO

			pá	ágina
RES	UMEN	EN ESPAÑOL		. ii
ABS	TRAC	T ENGLISH		. iii
AGF	RADEC	CIMIENTOS		. vi
LIST	TA DE	TABLAS		. ix
LIST	TA DE	FIGURAS		. x
LIST	TA DE	ABREVIATURAS		. xii
LIST	TA DE	SÍMBOLOS		. xiii
1	INTR	ODUCCIÓN		. 1
	$1.1 \\ 1.2 \\ 1.3 \\ 1.4$	Justificación	•	$ \begin{array}{c} 3 \\ 4 \\ 5 \\ 6 \end{array} $
2	REDE	CS COMPLEJAS		. 8
	2.1 2.2 2.3 2.4	Historia de las Redes Complejas	•	. 8 . 11 . 13 . 13 . 13 . 16 . 20
3	ASPE	CTOS BIOLOGICOS		. 24
	3.1 3.2 3.3	Descripción del Problema	•	. 24 . 26 . 27
4	REDE	CS COMPLEJAS PARA MODELAR EL PROCESO BIOLOGIC	O	29
	4.1 4.2 4.3 4.4	Construcción de las Redes Complejas	•	. 30 . 32 . 38 . 39

		4.4.1Del Criterio de Metropolis a la distribución Boltzmann-Gibbs44.4.2Algoritmo de Simulated Annealing4	2 4
5	RESU	JLTADOS	6
	$5.1 \\ 5.2$	Las redes y su uso para modelar el problema4Resultados4	16 18
6	CONC	CLUSIONES	57
APÉ	NDIC	\mathbf{ES}	68
7	TABL PAI RE	LAS DE LOS VALORES DE LA MODULARIDAD OBTENIDOSRA LAS REDES CONSTRUIDAS EN LOS EXPERIMENTOSALIZADOS6	59

LISTA DE TABLAS

Tabla	pág	gina
7–1	Modularidad de redes con nodos configurados mediante una distribución uniforme bivariada	69
7–2	Modularidad de redes cuyos nodos se distribuyen uniforme en X y en Y siguen la distribución dada por la ecuación 4.1 con $\lambda = 1$	69
7–3	Modularidad de redes cuyos nodos se distribuyen uniforme en X y en Y siguen la distribución dada por la ecuación 4.1 con $\lambda = 0.75$	70
7–4	Modularidad de redes cuyos nodos se distribuyen uniforme en X y en Y siguen la distribución dada por la ecuación 4.1 con $\lambda = 0.65$	70
7–5	Modularidad de redes cuyos nodos se distribuyen uniforme en X y en Y siguen la distribución dada por la ecuación 4.1 con $\lambda = 0.5$	70

LISTA DE FIGURAS

Figura

4–1	Red cuyos nodos siguen una distribución dada por la ecuación 4.1 con parámetro $\lambda = 1$ en Y y una distribución uniforme en X	32
4–2	Red cuyos nodos siguen una distribución dada por la ecuación 4.1 con parámetro $\lambda = 0.75$ en Y y una distribución uniforme en X	33
4–3	Red cuyos nodos siguen una distribución dada por la ecuación 4.1 con parámetro $\lambda = 0.65$ en Y y una distribución uniforme en X	34
4–4	Red cuyos nodos siguen una distribución dada por la ecuación 4.1 con parámetro $\lambda = 0.5$ en Y y una distribución uniforme en X	35
4–5	Red donde los nodos están ubicados espacialmente siguiendo una dis- tribución uniforme bivariada	36
5–1	Gráfica de los valores de la modularidad mostrados en la tabla 7–1	51
5–2	Gráfica de los valores promedio de la modularidad mostrados la tabla 7–1	52
5–3	Gráfica de los valores de la varianza mostrados en la tabla 7–1	53
5–4	Red donde los nodos están ubicados espacialmente siguiendo una dis- tribución uniforme bivariada y parámetro $\alpha = 75.7$ en el kernel de dispersión	54
5–5	Gráfica de los valores de la modularidad mostrados en la tabla 7–2	55
5–6	Gráfica de los valores promedio de la modularidad mostrados en la tabla 7–2	56
5–7	Gráfica de los valores de las varianzas mostrados en la tabla 7–2 $\ .\ .$	57
5–8	Red donde se muestran los módulos detectados por NETCARTO y cuyos nodos en el eje horizontal siguen una distribución uniforme y en el eje vertical siguen una distribución determinada por la ecuación 4.1 cuyo parámetro es $\lambda = 1$ y parámetro $\alpha = 75.7$ en el kernel de dispersión	58
5–9	Gráfica de los valores de la modularidad mostrados en la tabla 7–3	59

5–10 Gráfica de los valores promedio de la modularidad mostrados en la tabla 7–3	60
5–11 Gráfica de los valores de las varianzas mostrados en la tabla 7–3 $\ .$.	60
5–12 Red donde se muestran los módulos detectados por NETCARTO y cuyos nodos en el eje horizontal siguen una distribución uniforme y en el eje vertical siguen una distribución determinada por la ecuación 4.1 cuyo parámetro es $\lambda = 0.75$ y parámetro $\alpha = 225.7$ en el kernel de dispersión	61
5–13 Gráfica de los valores de la modularidad mostrados en la tabla 74	62
5–14 Gráfica de los valores promedio de la modularidad mostrados en la tabla 7–4	62
5–15 Gráfica de los valores de las varianzas mostrados en la tabla 74	63
5–16 Red donde se muestran los módulos detectados por NETCARTO y cuyos nodos en el eje horizontal siguen una distribución uniforme y en el eje vertical siguen una distribución determinada por la ecuación 4.1 cuyo parámetro es $\lambda = 0.65$ y parámetro $\alpha = 175.7$ en el kernel de dispersión	63
5–17 Gráfica de los valores de la modularidad mostrados en la tabla $7–5$	64
5–18 Gráfica de los valores promedio de la modularidad mostrados en la tabla 7–5	64
5–19 Gráfica de los valores de las varianzas mostrados en la tabla 7–5 $\ .$.	65
5–20 Red donde se muestran los módulos detectados por NETCARTO y cuyos nodos en el eje horizontal siguen una distribución uniforme y en el eje vertical siguen una distribución determinada por la ecuación 4.1 cuyo parámetro es $\lambda = 0.5$ y parámetro $\alpha = 325.7$ en el kernel de dispersión	65
5–21 Curvas de la modularidad promedio obtenidas para las diferentes con- figuraciones espaciales de los árboles	66
5–22 Superfice en tres dimensiones obtenida a partir de los parámetros λ en la distribución vertical de los nodos y α en el kernel de dispersión	66

LISTA DE ABREVIATURAS

IDE Integrated Development Environment.

LISTA DE SÍMBOLOS

- \mathbb{R} Conjunto de los números reales
- \mathbb{R}^p Espacio real euclídeo *p*-dimensional
- L(p) Longitud de un camino.
- C(p) Coeficiente de agrupamiento

 L_{random} Longitud de un camino en un grafo aleatorio

 C_{random} Coeficiente de agrupamiento en un grafo aleatorio

>> Mucho mayor que.

CAPÍTULO 1 INTRODUCCIÓN

La reproducción en los seres vivos es un fenómeno que constantemente llama la atención de los científicos. En algunas especies este fenómeno ha llegado a ser ampliamente estudiado y bien documentado, mientras que en otras es casi desconocido como se lleva a cabo el proceso de apareamiento entre dos individuos y los hechos que rigen el acto de reproducción. En ciertos casos, la dinámica de algunas especies y su evolución han permitido que el estudio de este fenómeno sea relativamente sencillo y su entendimiento ha permitido la creación de estrategias de conservación de especies que pueden correr el riesgo de ir a la extinción.

En otros casos, estudiar y documentar este fenómeno puede llegar a ser una tarea difícil debido a la complejidad de los procesos que intervienen en la reproducción. Dicha complejidad está dada por el hecho de que para algunas especies, el contacto directo entre los individuos ha sido descartado. Esto trae como consecuencia que el apareamiento entre dos individuos dependa de otras especies y en ocasiones llegue a depender de factores ambientales y climáticos.

En el caso particular de las plantas que requieren de un insecto polinizador, el proceso de apareamiento entre dos individuos está determinado por el comportamiento de los insectos que realizan la polinización. Factores espaciales, ambientales y climáticos pueden afectar de manera considerable el número de apareamientos de uno o varios individuos, debido a que cambios en estos factores pueden alterar el comportamiento de los insectos polinizadores. El propósito de este trabajo es estudiar y documentar como ocurre la interacción entre los individuos de una población de árboles en términos del proceso de apareamiento; dicho proceso está determinado por la actividad polinizadora de los insectos que transportan el polen de un árbol a otro. La forma en la que se va a modelar este proceso de apareamiento, es mediante el uso de redes complejas. Usando está herramienta, se representan los árboles de la población como los nodos de una red y la ocurrencia de un enlace entre estos nodos, representa un vínculo de apareamiento entre los árboles; un enlace entre pares de nodos se ubica con una probabilidad determinada por una función de densidad de probabilidad obtenida en [1] y que en adelante se conocerá como el *kernel de dispersión*.

Con la ayuda de la teoría de las redes complejas se podrá medir la tendencia que tienen los individuos de la población a formar grupos de nodos disjuntos bien definidos y conocidos como *modulos*. Esta tendencia a la formación de modulos disjuntos en las redes construidas es conocida como la modularidad de las redes[1] y su formación está principalmente motivada por factores espaciales. En particular, la distancia entre pares de árboles juega un papel de gran importancia por la limitación que impone a los insectos en su radio de acción, de tal manera que la aparición de un enlace entre individuos de la población decrecerá a medida que las distancias entre ellos aumente.

Debido a que la modularidad de la red mide la tendencia de los nodos a formar módulos en donde la densidad de enlaces internos es alta comparada con la densidad de enlaces externos hacia otros grupos, se estudia como esta medida puede variar al considerar diferentes configuraciones espaciales de los individuos y al cambiar la configuración de los enlaces debido a un cambio en el parámetro de escala del kernel de dispersión. En este trabajo se está suponiendo que este cambio en el parámetro de escala está reflejando cambios debido a factores abióticos.

1.1 Justificación

Debido al aumento en la desaparición de bosques y selvas en el mundo, muchas especies de plantas y animales diariamente alargan la lista de especies en peligro de extinción. Es por eso que resulta importante desarrollar técnicas que permitan estudiar el impacto causado por la intervención de diversos factores, en especial aquellos factores que repercuten sobre los procesos que contribuyen a la sobrevivencia de una especie. En particular, el proceso de reproducción que se lleva cabo entre individuos de una población.

La importancia de este trabajo, radica en la utilización de las redes complejas para simular eventos de polinización de árboles teniendo en cuenta factores abióticos. El uso de las redes complejas permitirá cuantificar características generales de las redes de apareamiento (*conectividad y modularidad*) de los árboles, de tal manera que se pueda obtener un entendimiento de como es la interacción entre individuos en términos de reproducción y las consecuencias de esta interacción sobre el flujo genético en la población.

Para el caso de este trabajo, se tiene principal interés en la tendencia que tienen los árboles de la población a formar grupos en donde los eventos de apareamiento ocurren en mayor número, comparado con los que ocurren entre estos mismos grupos.

Estudiar la aparición de estos grupos dentro de la población, permitirá entender cuáles son los posibles efectos sobre el flujo genético dentro de la población, ya que los genes que se acentúen dentro de un grupo pueden ser distintos a los que se acentúen en otro grupo, contribuyendo esto a que eventualmente dichos grupos se distancien genéticamente unos de otros y provocando esto una extinción de la especie original.

De esta manera al desarrollar una técnica que permita estudiar como los árboles dentro de una población, están estableciendo vínculos de apareamiento para responder a las diferentes condiciones impuestas por el entorno, permitirá desarrollar estrategias enfocadas a la conservación y preservación de algunas especies de plantas, principalmente aquellas que se están viendo afectadas por los diferentes cambios que se están generando en sus hábitats. En particular las especies de plantas donde el proceso de reproducción está determinado por la actividad polinizadora de insectos.

1.2 Evolución en Plantas

Aunque existen diferentes definiciones para el término evolución biológica, la definición dada por **Futuyma** en [2] es considerada por **Stebbins** en [3] como una definición "intermedia" entre aquellas que hacen especial énfasis entre el componente genético de una población y las que enfatizan en las interacciones poblaciónambiente. Esta definición considera que la evolución biológica consiste en los cambios que trascienden el tiempo de vida de un individuo en una población, y donde el término población debe entenderse como un grupo de organismos o individuos de una misma especie que ocupan un mismo espacio o área geográfica.

Para hablar a grandes rasgos de cómo evolucionaron las plantas, primero hay que decir que sus orígenes están relacionados con la aparición de las primeras células vegetales como antecesores de las algas y de las plantas verdes. En el caso de las algas verdes, fueron éstas las que dieron origen a todas las plantas terrestres.

Las plantas terrestres aparecen aproximadamente hace 450 a 470 millones de años como lo indica el fósil de una planta vascular llamada *Cooksonia*. Este se convirtió en una evidencia importante para tener una idea de cuando las plantas ya poblaban la superficie terrestre. Con el transcurrir del tiempo las plantas que estaban en la superficie terrestre empezaron a adaptarse de forma tal que poseían raíces que les permitían obtener nutrientes de tal manera que necesitaban menos de la presencia del agua para su crecimiento y sobrevivencia.

De esta manera en el periodo jurásico aparecen las coníferas y los helechos como las plantas con mayor presencia en la superficie terrestre. Sin embargo, las angiospermas, aunque ya existían, desarrollaron flores como una adaptación que les permitió llevar a cabo una reproducción de tipo sexual. Esta flores que servían y sirven aún como atractores, son polinizadas por insectos y en algunos casos por ciertas especies de aves. Luego de que la fecundación ocurre dentro de ellas, la flor llega a convertirse en un fruto que tiene una semilla, que después al caer al suelo germinará dando origen a una nueva planta. Esta adaptación le permitió a las plantas con flores llevar a cabo una colonización del planeta después de la época del eoceno ¹.

1.3 Polinización

En las angiospermas, la polinización es el proceso por el cual se transfiere el polen desde los órganos masculinos de la flor o *estambres* hacia el *estigma*, que es la parte de la flor en donde se lleva a cabo la fecundación de los óvulos permitiendo así la producción de semillas y frutos. El proceso de polinización puede llevarse a cabo generalmente por medio de las siguientes formas: polinización zoófila, polinización anemófila, polinización hidrófila o autopolinización. En el caso de la polinización zoófila el polen es transportado por un animal, que en ocasiones puede ser un insecto (*polinización entomófila*) o un ave (*polinización ornitofila*). En el caso de la polinización anemófila, el encargado del transporte del polen es el viento, mientras que en la polinización hidrófila el transporte del polen es realizado por el agua.

En el caso de las plantas que utilizan animales como polinizadores, cabe resaltar que la polinización es un resultado de una relación simbiótica entre la planta y el animal. Por ejemplo, mientras un insecto se alimenta o se refugia en una flor, los granos de polen se adhieren a su cuerpo para ser transportados hacia otras flores en diferentes plantas, contribuyendo así al proceso de reproducción de la población de plantas.

¹ Eoceno es una época geológica de la Tierra, la segunda del período Paleógeno en la Era Cenozoica. Comprende el tiempo entre el final del Paleoceno (hace $55, 8 \pm 0, 2$ millones de años) y el principio del Oligoceno (hace $33, 9 \pm 0, 1$ millones de años).

1.4 Flujo Genético

El flujo genético es el intercambio o traspaso de material genético de una población a otra generalmente dentro de la misma especie. Este traspaso puede ser causado por la inmigración de uno o varios individuos de una población a otra o en el caso de las plantas debido al intercambio de polen de una población a otra.

Es importante resaltar que este intercambio de material genético entre poblaciones puede llegar a causar cambios considerables en las frecuencias de los genes de una población debido a la introducción de nuevo material genético, proveniente de individuos inmigrantes o del polen recibido desde otras poblaciones según sea el caso.

Existen diferentes factores que pueden afectar el ritmo de intercambio de material genético entre poblaciones. Uno de los factores que más afecta este ritmo es el de la movilidad: entre más posibilidades de movimiento tenga los individuos de una población, mayor será la capacidad migratoria, y por lo tanto, el traspaso de información genética será mayor.

En el caso de las plantas, éstas presentan una desventaja en términos de movilidad con respecto a los animales. En éstas el transporte del material genético entre poblaciones de plantas es encargado a terceros.

Por otro lado cabe resaltar que debido a diferentes factores, ya sean de tipo geográfico, climático, etc, el flujo genético puede verse afectado entre las poblaciones que constituyen una misma especie. Como consecuencia, hay una exposición de la especie a estar en peligro de extinción.

De esta manera el flujo genético es un factor al cual se le debe prestar especial importancia, ya que puede brindar un entendimiento de cómo la variabilidad en el patrimonio genético de una especie puede incrementar o disminuir las posibilidades de sobrevivencia de la misma. En este sentido, es de gran utilidad descubrir herramientas que permitan adquirir conocimientos acerca del flujo de la información genética entre las poblaciones.

CAPÍTULO 2 REDES COMPLEJAS

Una de las herramientas más utilizadas en la actualidad para el estudio de muchos sistemas en la naturaleza y en problemas del mundo real son las redes complejas, las cuales se pueden pensar intuitivamente como grafos con un gran número de vértices (*nodos*) y de aristas (*enlaces*) obtenidos al modelar algunos fenómenos del mundo real y cuyas carácteristicas topológicas no son triviales. Alrededor de las redes complejas se vienen desarrollando avances importantes de carácter teórico y constantemente se están produciendo artículos con aportes significativos para el entendimiento de las mismas.

Muchos científicos están concentrando sus esfuerzos en el correcto entendimiento de las redes complejas, la utilidad que estas han mostrado para modelar algunos fenómenos del mundo real ha motivado el desarrollo de diversos estudios que están contribuyendo al fortalecimiento de lo que actualmente se conoce como la teoría de redes complejas.

En este capítulo se presenta un descripción general acerca del desarrollo de las redes complejas. También se presenta un marco general acerca de lo que es una red compleja (*complex network*), las principales características que ellas poseen, en particular la estructura de módulos o comunidades dentro de éstas.

2.1 Historia de las Redes Complejas

Los inicios de las redes complejas se remontan principalmente a los inicios de la teoría de grafos en 1736 con el trabajo del gran matemático Leonhard Euler sobre el problema de los puentes de Königsberg, el problema era el siguiente: El rio Pregel atraviesa la ciudad de Königsberg y a lo largo de su cauce hay dos islas que se comunican entre ellas y con la ciudad mediante siete puentes, entonces se quería saber si existía la posibilidad de comenzar una caminata por la ciudad y las islas, de tal manera que cada puente solo se pudiera cruzar una y solamente una vez.

Euler demostró que tal caminata no era posible y para llegar a esta conclusión, se dio cuenta que la elección de la ruta dentro de cualquier porción de tierra era irrelevante. Sin embargo, lo que debía tenerse en cuenta en una ruta, era la sucesión de puentes que eran atravesados.

Euler al notar estos hechos, reformuló el problema en un contexto abstracto y para eso introdujo un objeto matemático llamado *grafo*, el cual está constituido por un conjunto de nodos o vértices y por un conjunto de aristas o enlaces que conectan esos nodos.

De esta manera al crear un grafo en donde las porciones de tierra eran consideradas como los nodos y los puentes eran considerados como las aristas, Euler logró una representación del problema en la que se tenían en cuenta los principales detalles del problema original, para así demostrar la no existencia de un camino que atravesara todos los siete puentes exactamente una vez.

La estrategia de Euler al hacer uso de un grafo para solucionar el problema, dio paso a definiciones y teoremas acerca de los caminos y recorridos dentro de un grafo.¹. Esto dio origen a la *teoría de grafos*; una teoría profunda e interesante, que ha servido como base para el estudio de las propiedades de una red.

 $^{^{1}}$ Un camino en un grafo es una sucesión finita en la que aparecen alternadamente vértices y aristas de dicho grafo, por ejemplo caminos *Eulerianos y Hamiltonianos* en donde cada arista es atravesada exactamente una vez y donde cada vértice es visitado exactamente una vez respectivamente

Como los vértices pueden representar casi cualquier entidad y las aristas las conexiones entre estos (*personas y amistades, ciudades y carreteras, aeropuertos y rutas aéreas entre ellos, etc.*), la teoría de grafos ha permitido generar estudios acerca de las características topológicas de los grafos. De esta manera, dichos estudios permiten buscar soluciones a los diferentes problemas que aparecen tanto en las matemáticas puras como en problemas del mundo real que se abordan con el uso de los grafos.

Es precisamente su uso fuera de las matemáticas puras, lo que ha hecho que la teoría de grafos haya sido ampliamente difundida y utilizada en las décadas pasadas, especialmente en la ingeniería, ciencias de la computación, investigación de operaciones y más recientemente en biología y sociología.

De hecho, a mediados del siglo XX, cuando se presentó un gran interés en la forma de cuantificar métodos en sociología, fue la teoría de grafos la herramienta que adoptaron los estudiosos de esta área para analizar datos empíricos. Inclusive, se adoptaron ciertos términos que en un principio eran exclusivos de la teoría de grafos.

Fue también con ayuda de los grafos que algunos matemáticos comenzaron a estudiar cómo se podían propagar enfermedades o rumores dentro de un grupo de personas. Esto trajo como consecuencia que en algunas ocasiones resultara más útil observar los grafos desde un punto de vista estocástico y dinámico, más que desde el punto de vista determinístico y estático; de tal manera que algunas propiedades de los mismos ahora se pudieran ver como distribuciones de probabilidad.

Este paso de lo estático a lo dinámico, trajo como consecuencia un cambio en la teoría tradicional de los grafos y exigió por parte los científicos nuevos descubrimientos que dieron paso al concepto de red compleja y a la teoría de redes complejas simultáneamente.

2.2 Aspectos Generales sobre las Redes Complejas

Los resultados y aportes que se han obtenido en la teoría de grafos han surgido del estudio de un grafo visto como un objeto matemático, estático e inmerso en la misma armonía de las matemáticas.

Al modelar situaciones del mundo real con el uso de grafos, hay que tener en cuenta los diferentes aspectos dinámicos que este trae consigo. Por ejemplo, al modelar fenómenos de reproducción, dispersión de enfermedades, relaciones sociales, etc., es importante tener en cuenta los diferentes comportamientos que presentan estos individuos como factores generadores de cambios en el tiempo.

Esta exigencia y la importancia de poder obtener resultados lo más fiel posible a la realidad, origina la teoría de las redes complejas. De hecho, en la actualidad se está mostrando como un nuevo tipo de ciencia conocida como la "ciencia de las redes complejas" y está desempeñando un papel importante en el entendimiento de situaciones del mundo real que hasta hace unos años eran difíciles de estudiar y describir.

La teoría de las redes complejas se diferencia de la teoría tradicional de grafos y de los trabajos anteriores en redes, principalmente en tres aspectos según se indica en [4]:

- Se enfoca en las propiedades de las redes del mundo real, preocupándose tanto de las cuestiones empíricas como de las teóricas.
- Frecuentemente tiene un punto de vista de que las redes no son estáticas y que por el contrario evolucionan de acuerdo a diferentes reglas dinámicas.
- 3. Entender las redes como las estructuras sobre las cuales se construyen los sistemas dinámicos distribuidos, y no solamente como un objeto topológico.

Es importante resaltar que el desarrollo de las redes complejas está enfocado hacia el estudio de las características estructurales de la red. Así, primero se debe estar en la capacidad de identificar cuales de esas características son las de principal importancia. Además por ser obtenidas desde el mundo real, se debe tener en cuenta cuales datos empíricos son esenciales para el desarrollo formal de la teoría desde una base matemática sólida.

Por ejemplo, en el estudio de una red social debe ser importante diferenciar los datos empíricos relevantes para así establecer una conexión entre un par de individuos.

Una de las metas que se han trazado los investigadores de la nueva ciencia de las redes complejas está relacionada con el entendimiento de como la estructura global de la red puede verse afectada por los procesos dinámicos que ocurren a una escala local. Por ejemplo, la modularidad de la red o el número de enlaces en la misma puede verse afectada por la aparición o remoción de vértices.

Por otro lado según [4], considerar las redes como la estructura sobre la cual están construidos los sistemas dinámicos distribuidos implica una visión de los sistemas dinámicos en donde los vértices de un grafo representan entidades dinámicas discretas.

Dicho lo anterior, se tienen las siguientes definiciones de lo que es una red y una red compleja respectivamente.

Definición 1 (Red). Una red es un conjunto de elementos discretos (vértices) y un conjunto de conexiones (aristas) que enlazan dichos elementos por pares.

Definición 2 (Red Compleja). Una red compleja es una red con características topológicas no triviales y donde dichas características no ocurren en redes como los grafos aleatorios.

Algunas de estas características topológicas son la cola gruesa o pesada en la distribución de grado de los nodos, un alto coeficiente de agrupamiento, la aparición de estructura de módulos, entre otras.

2.3 Tipos de Redes Complejas

Los dos tipos de redes que han sido ampliamente estudiadas son las redes *mundo* pequeño, y las redes *libres de escala*. A continuación se muestra una descripción general de lo que es cada una de estas redes y sus características más importantes. Esta descripción está basada en los artículos originales donde fueron incialmente propuestas.

2.3.1 Redes Mundo-Pequeño

Para hablar de redes mundo pequeño es importante primero tener en cuenta las siguiente definiciones:

Definición 3 (Red Regular). Una red regular, es una red en la que todos los nodos tienen el mismo grado y donde cada nodo está conectado a sus nodos adyacentes por enlaces no dirigidos

Definición 4 (Red Aleatoria). *Es una colección de puntos, o nodos, con líneas, o enlaces que conectan pares de nodos al azar.*

Para empezar a hablar de las redes mundo pequeño, primero se puede decir que el término "*mundo pequeño*" está motivado por la idea de seis grados de separación, propuesta por el escritor húngaro **Frigyes Karinthy** en 1929 y que posteriormente fue probada mediante un experimento por el psicólogo **Stanley Milgram** en los años 60.

En el esquema propuesto por Milgram, cada individuo estaba representado por un nodo conectado a otro en caso de que tales individuos se conocieran mutuamente. La pregunta que surgía era ¿Qué número de individuos o nodos intermedios separan en promedio a dos personas escogidas al azar dentro de la población norteamericana? , la respuesta a esta pregunta fue ¡solamente seis!. Desde entonces este resultado se conoce como "seis grados de separación", en otras palabras es el soporte teórico del tradicional dicho "el mundo es un pañuelo". Luego en 1998, **Duncan J. Watts** y **Steven H. Strogratz** en su artículo llamado *Collective dynamics of 'small-world' networks*, propusieron un modelo sencillo de red denominado redes mundo pequeño. En términos generales Watts y Strogatz introducen el modelo de red mundo pequeño, como una red que no es completamente regular pero que tampoco es completamente aleatoria. Ejemplos de este tipo de redes citados en [5] son la red neuronal del gusano *Caenorhabditis elegans*, la red eléctrica del oeste de Estados Unidos y la red de colaboración de los actores de cine.

Por otro lado, Watts y Strogatz mostraron que los modelos de sistemas dinámicos con *small-world coupling* muestran una velocidad elevada de propagación de señales, poder computacional y sincronizabilidad, como una consecuencia de esto las enfermedades infecciosas se propagan de manera más fácil en redes del tipo mundo pequeño que en redes o retículos regulares.

Para lograr una configuración de una red entre regular y aleatoria, el procedimiento que llevaron a cabo Watts y Strogatz consistía en reconfigurar un grafo regular con n nodos dispuestos en forma de anillo y con k enlaces por vértices. La manera en que esta reconfiguración se realizaba era renovando cada enlace al azar con probabilidad p hasta "sintonizar" una red entre la regularidad (p = 0) y el desorden (p = 1), de tal manera que con estas nuevas redes obtenidas se pudiera estudiar lo que pasaba cuando 0 .

De igual forma, en [5] Watts y Strogatz cuantifican las propiedades estructurales de las redes construidas por el proceso antes mencionado. Dicha cuantificación se realiza a partir de algunas características de las redes, tales como la longitud de un camino L(p) y coeficiente de agrupamiento C(p).

En [5], L(p) y C(p) se definen de la siguiente manera: L(p) se define como el número de lados en el camino más corto entre dos nodos, promediado sobre todos los pares de nodos y para definir C(p), suponga que un nodo v tiene k_v nodos adyacentes; así a lo más $k_v(k_v - 1)/2$ enlaces pueden existir entre ellos. Sea C_v la fracción de estos enlaces admisibles que actualmente existen. Se define C(p) como el promedio de C_v sobre todos los nodos.

Por otro lado, dentro del estudio que llevaron a cabo Watts y Strogratz en [5], las redes en las cuales ellos estaban interesados eran redes con muchos vértices y con pocas conexiones, en las cuales se requería que $n \gg k \gg \ln(n) \gg 1$, donde $k \gg \ln(n)$ garantiza la conexidad de un grafo aleatorio, según se cita en la referencia número 16 de [5] y donde n es el número de vértices en el grafo y k es el número de lados por vértice. El símbolo \gg denota *mucho mayor que* (ver lista de simbolos).

Además, Watts y Strogatz descubrieron que $L \sim \frac{n}{2k} >> 1$ y $C \sim 3/4$ cuando $p \rightarrow 0$, es decir, cuando la red se aproxima a la característica de regularidad, mientras que $L \approx L_{random} \sim \frac{\ln(n)}{\ln(k)}$ y $C \approx C_{random} \sim k/n \ll 1$ cuando $p \rightarrow 1$, es decir, cuando la red se aproxima al desorden.

En este caso L_{random} y C_{random} representan la longitud de camino y el coeficiente de agrupamiento, respectivamente, de un grafo aleatorio que tiene el mismo número de vértices n y el mismo número promedio k de lados por vértices que la redes reales utilizadas en [5].

Estos resultados indican que cuando la red es totalmente regular se presenta un alto grado de agrupamiento, en donde L crece linealmente con el número de nodos; mientras que en una red aleatoria se tiene un bajo grado de agrupamiento, donde L crece logarítmicamente con el número de nodos.

Dados los anteriores resultados, se podría esperar que grandes valores de C están siempre vinculados con grandes valores de L, y pequeños valores de C están vinculados con pequeños valores de L.

Sin embargo, uno de los resultados más importantes obtenidos por Watts y Strogatz en [5], reveló que existe un amplio rango de valores de p en donde las redes

mundo pequeño presentan un valor de L(p) que es casi tan pequeño como L_{random} y en donde se mantiene que $C(p) >> C_{random}$.

Esta implicación sobre el coeficiente de agrupamiento está relacionada con el hecho de que al hacer la reconfiguración de una red regular hacia una red aleatoria pueden aparecer nuevos enlaces que conectan dos nodos directamente y que en principio estaban alejados por determinada cantidad de enlaces intermedios entre ellos (es decir, un enlace nuevo puede convertir dos nodos que no eran adyacentes en nodos adyacentes, haciendo que L(p) disminuya).

En general, las propiedades más relevantes que presenta una red mundo pequeño son:

- Un alto grado de agrupamiento como en una red regular.
- Mantienen una longitud de camino tan pequeña como una red aleatoria.

2.3.2 Redes Libres de Escala

El término redes libres de escala (en inglés *scale-free networks*) es debido a Albert-László Barabási y Réka Albert, por su trabajo *Emergence of Scaling in Random Networks* publicado en 1999. Este término hace referencia a las redes que tienen la característica de que la distribución de los grados de los nodos sigue una *Ley de Potencias.*²

Ejemplos de estas redes son la World Wide Web [6], la red de actores de películas [5] y las redes de citación [7]. Estos ejemplos fueron usados por Barabási y Albert para afirmar que la Ley de Potencias es una característica presente en muchas redes que modelan sistemas del mundo real.

² Una relación en forma de ley de potencia entre dos escalares x y y es aquella que puede expresarse en la forma $y = ax^k$ donde a y k son constantes y son la constante de proporcionalidad y el exponente de la potencia respectivamente.

Si se tiene en cuenta que muchas redes en el mundo real son creadas a partir de procesos dinámicos, de adición de vértices y reconfiguración de enlaces, resulta importante estudiar las propiedades estructurales de este tipo de redes. Es así como Barabási y Albert en [6], luego de examinar los ejemplos antes mencionados, proponen que la Ley de Potencias en la distribución de los grados de los nodos de una red podría llegar ser una propiedad genérica de muchas redes. De hecho, es quizás debido a [6] que actualmente hay un gran interés en las redes libres de escala.

Dentro de otras propuestas hechas por Barabási y Albert en su trabajo, se tiene que las propiedades de las redes libres de escala pueden ser explicadas considerando la red como una entidad dinámica que crece con el paso del tiempo, en contraste con estudios anteriores donde la red era considerada como un grafo estático.

La otra propuesta estaba basada en un modelo específico para el crecimiento de la red, de tal manera que al construir la red y al darse su crecimiento en el tiempo, la distribución de los grados de los nodos que se obtiene, es una Ley de Potencias similares a las que se observan en la World Wide Web y otras redes.

Uno de los aspectos más interesantes del trabajo de Barabási y Albert, es mostrar que redes aparentemente diferentes como la World Wide Web y la redes de citaciones, en realidad tienen la característica de que la distribución de los grados de los nodos dentro de éstas sigue una Ley de Potencias.

Esto llevó a Barabási y Albert a sugerir que todas las redes de este tipo tienen un mecanismo común que rige la estructura de estas redes; el mecanismo consiste de dos componentes:

- 1. La red crece de tal forma que nuevos nodos son agregados continuamente.
- Los nodos adquieren nuevos lados en proporción al número de lados que ya tienen (preferential-attachment).

Para demostrar esto, Barabási y Albert en [6] proponen un modelo, en el cual el crecimiento de la red, está dado por *pasos de tiempo*, donde en cada paso de tiempo

se agrega a la red un nuevo nodo con m enlaces saliendo de él y conectándolo a nodos ya existentes que son elegidos al azar con una probabilidad proporcional al grado.

El análisis hecho por Barabási y Albert en términos generales y mostrado en [4] consiste en lo siguiente:

Sea k_i el grado del nodo i. Entonces la propabilidad de que un lado perteneciendo a un nuevo nodo que aparece en la red se conecte a i es:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{2.1}$$

Donde $\sum_{j} k_{j}$ es la suma total de los grados de los nodos incluyendo los grados de los nuevos nodos que aparecieron.

En promedio, el grado k_i del nodo *i* crece a una tasa proporcional a k_i . De esta forma si se ignoran las fluctuaciones y se asume que k_i siempre toma un valor medio, la evolución en el tiempo de k_i está dada por:

$$\frac{dk_i}{dt} = m\Pi(k_i) = \frac{mk_i}{\sum_j k_j} = \frac{k_i}{2t}$$
(2.2)

Donde se considera que $\sum_{j} k_{j} = 2mt$ debido a que si un nodo aparece en cada paso de tiempo, después de t pasos habrán t nuevos nodos, cada uno conectado con m nodos ya existentes escogidos por el *preferential attachment*. De esta forma, el aporte de enlaces por parte de estos nuevos nodos es mt y como cada uno de los mnodos escogidos tiene t nuevos enlaces, se tiene un total de 2mt en la suma de los grados de todos los nodos actuales en la red.

Al integrar la ecuación 2.2 y considerando la condición inicial de que en $t = t_i$ se tiene $k_i = m$ se obtiene:

$$k_i(t) = m \left(\frac{t}{t_i}\right)^{1/2} \tag{2.3}$$

Donde t_i es el tiempo en el cual el vértice i es agregado a la red.

Entonces el modelo sugerido por Barabási y Albert de la distribución de los grados al tiempo t está dada por:

$$P(k) = -\left(\frac{dk_i}{dt_i}\right)^{-1} = \frac{2m^2t}{k^3}$$

$$(2.4)$$

Donde P(k) es la probabilidad de que un nodo de la red elegido al azar, tenga grado ky donde la igualdad de la derecha en la ecuación 2.4 se obtiene al derivar la ecuación 2.3 respecto a t_i y reemplazando en dicha derivada las ecuaciones $(t/t_i)^{1/2} = k/m$ y $t_i^2 = (t^2m^4)/k^4$ (obtenidas de la ecuación 2.3).

De este modelo se observa que la distribución de los grados, sigue una Ley de Potencias, en donde el grado del exponente es $\gamma = 3$ como se muestra en [6] y donde dicho sea de paso, es este exponente un problema que presenta el modelo; debido a que $\gamma = 3$ es independiente de los parámetros del modelo y por lo tanto no se puede variar.

Al surgir la inquietud de si el modelo puede ser modificado para dar origen a otros exponentes, Barabási y Albert, resaltaron que existen diversas formas para darle forma a una red. Dentro de estas formas se puede tener la adición y reconfiguración o la adición y remoción de nodos o lados.

La idea que ellos presentaron en un trabajo posterior en el año 2000, incorpora la adición de nuevos lados entre los nodos que ya existen y el movimiento de algunos lados ya existentes.

Según se muestra en [4] en este nuevo modelo propuesto por Barabási y Albert uno de tres eventos podían pasar en cada paso del tiempo:

- Con probabilidad p, m enlaces nuevos son agregados a la red. Un extremo de cada uno de los nuevos lados es conectado a un nodo seleccionado uniformemente al azar de la red y el otro extremo elegido usando el proceso preferential attachment
- 2. Con probabilidad q, m lados son "reconfigurados", de tal manera que un vértice i es seleccionado al azar y uno de sus lados elegido al azar es removido y reemplazado

con un nuevo lado cuyo otro extremo está conectado a un vértice elegido de acuerdo al proceso de *preferential attachment*

3. Con probabilidad 1 - p - q, un nuevo nodo es agregado a la red. El nuevo nodo tiene m nuevos enlaces que están conectados a nodos ya existentes en la red por medio del preferential attachment.

Al tener en cuenta la construcción de una red bajo la propuesta de este nuevo modelo, la red obtenida tendrá una distribución de grados de los nodos que sigue una Ley de Potencia, con un exponente γ que depende de los parámetros p,q y m.

2.4 Algunas Características Topológicas de una Red Compleja

Quizás la consecuencia más importante que se presenta al intentar modelar sistemas del mundo real con ayuda de las redes, es que en muchos casos las redes obtenidas, tienen características topológicas no triviales donde la forma en la que se presentan las conexiones entre los nodos de la red, no es puramente regular ni puramente aleatorio.

Son precisamente estas características no triviales las que diferencian a las redes complejas de los retículos y los grafos aleatorios; dentro de estas características tenemos: La distribución de los grados de los nodos en la red es de cola pesada (*distribución de grado*), un alto coeficiente de agrupamiento, preferencia de los nodos por conectarse con nodos que son similares o diferentes en algún sentido (*assortativity*), estructura de comunidad entre otras.

Mucho del trabajo llevado a cabo por los teóricos de las redes complejas se ha concentrado en poder cuantificar estas características, de tal forma que se puedan medir los efectos de la reconfiguración de una red a lo largo de un proceso de cambio en los sistemas que estas representan. A continuación se definen algunas de estas características: **Definición 5** (Distribución de Grado). Sea p_k la probabilidad de que un nodo elegido uniformemente al azar en una red tenga grado k, es decir, $p_k = n_k/n$ donde n_k es el número de nodos con grado k y n el número total de nodos en la red. Se define la **distribución de grado** para la red como el histograma de los grados de los nodos [8].

Al decir que la distribución de grado en las redes complejas es de cola pesada, se está haciendo referencia a las distribuciones de probabilidad cuya cola es más pesada que la cola de una distribución exponencial.

En el caso de la característica de *assortativity* más que una definición puede decirse que es la correlación entre propiedades de nodos adyacentes [9]. Una forma de medir la *assortativity* es el *coeficiente de assortativity* definido por M. E. J. Newman en [9].

Una de las propiedades que ha sido más estudiada en la estructura de las redes, es la estructura de comunidad. Detectar esta característica y poder cuantificarla, ha sido el objeto de estudio de muchos científicos que en los últimos años, han desarrollado algoritmos para encontrar la mejor manera de poder medir esta propiedad.

Poder detectar esta propiedad en las redes que surgen al modelar situaciones del mundo real permite entender la interacción que ocurre entre las entidades que conforman la red. **M. E. J. Newman** y **M. Girvan** definen la estructura de comunidad en [10] de la siguiente manera:

Definición 6 (Estructura de Comunidad). La estructura de comunidad se define como la división de la red en grupos de nodos dentro de los cuales las conexiones de red son densas pero entre estos grupos dichas conexiones son escasas.

Una de las medidas que ha sido ampliamente citada por diferentes autores ([1],[11], [12],[13] entre otros) en trabajos sobre teoría de redes y de sus aplicaciones a problemas del mundo real, es la medida cuantitativa de modularidad definida por Newman y Girvan en [10].

La idea de esta aproximación propuesta por Newman y Girvan está basada en el siguiente hecho: Dado que una red se puede particionar de diferentes maneras, se está interesado en saber, ¿Cuál de esas particiones es la mejor?.

Para dar respuesta a esa pregunta Newman y Girvan definen la medida de *Modularidad*, como aquella que permite cuantificar la calidad de una partición de una red dada.

En términos generales esta medida está basada en la medida de *assortativity* propuesta también por Newman en [9]. La construcción de esta medida se muestra en [10] y es hecha de la siguiente manera: Se define una matriz simétrica e de tamaño $k \times k$ donde las entradas e_{ij} representan la fracción de enlaces en la red que conectan nodos en un módulo i con nodos en un módulo j.

De esta forma la $Tr(\mathbf{e}) = \sum_{i} e_{ii}$ representa la fracción de enlaces en la red que conectan nodos dentro de un mismo módulo.

Sin embargo, Newman y Girvan son conscientes de que la traza en si misma no es un buen indicador de la calidad de la partición dada, entonces definen también $a_i = \sum_j e_{ij}$ como la fracción de enlaces que se conectan a nodos en el módulo *i*.

Destacan además, que si en una red los enlaces se conectan a nodos sin tener en cuenta el módulo al cual ellos pertenecen, entonces $e_{ij} = a_i a_j$.

De esta manera la medida de modularidad está definida como [10]:

$$Q = \sum_{i} (e_{ii} - a_i^2) = Tr(\mathbf{e}) - \|\mathbf{e}^2\|$$
(2.5)

Donde $||\mathbf{x}||$ indica la suma de los elementos en la matriz \mathbf{x} . En algunos artículos ([1], [13]) esta medida se presenta de la siguiente forma:

$$M = \sum_{s=1}^{N_M} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right]$$
(2.6)

Donde N_M es el número de módulos, L es el número de enlaces en la red, l_s es el número de enlaces entre nodos en el módulo s, y d_s es la suma de los grados de los nodos en el módulo s.
CAPÍTULO 3 ASPECTOS BIOLOGICOS

El propósito del presente capítulo es describir en términos generales cómo es que ocurre la interacción entre los individuos de una población de árboles desde el punto de vista del proceso de la polinización. Dicha interacción permite llevar a cabo la reproducción, para que cada uno de los individuos pueda difundir su información genética a través de la población.

3.1 Descripción del Problema

El problema propuesto en este trabajo surge del estudio hecho por Miguel Fortuna, Cristina Garcia, Paulo Guimarães y Jordi Bascompte en [1]. En ese trabajo, Fortuna y sus colaboradores modelan mediante las redes complejas el proceso de apareamiento entre los individuos de una población de *Prunus Mahaleb*, cuya polinización es llevada a cabo por insectos.

En este proceso de modelamiento de las redes de apareamiento, se tienen en cuenta dos *kernels de dispersión* (uno de cola gruesa y otro de cola delgada) para simular los eventos de apareamiento entre pares de árboles de la población. Además se tienen en cuenta diferentes configuraciones en la ubicación de los árboles que constituyen la población.

Teniendo en cuenta estos factores, se lleva a cabo un estudio de la modularidad y conectividad de las redes construidas, con el propósito de obtener información que pueda ser interpretada desde un punto de vista ecológico y biológico. Al igual que en [1], en este trabajo también se construyen diferentes redes de apareamiento entre una población de árboles, con el propósito de estudiar propiedades estructurales de las mismas.

A diferencia del trabajo presentado en [1], en este trabajo se tiene en cuenta solo un *kernel de dispersión de cola gruesa*, el cual está determinado por una función de densidad de probabilidad Weibull y cuya construcción y ajuste se muestra de manera detallada en [1].

El propósito de usar el kernel de dispersión de cola gruesa en la construcción de las redes permite a los individuos separados por largas distancias (400 mts o más) tener mayor posibilidad de generar un vínculo de apareamiento. Esto trae como consecuencia que las redes construidas presenten ciertas diferencias estructurales con las redes construidas utilizando un *kernel de dispersión de cola delgada*.

Por ejemplo en [1], las redes construidas con el kernel de dispersión de cola gruesa presentan una mayor conectividad y una menor modularidad que las construidas con un kernel de dispersión de cola delgada. En el caso de este trabajo el objetivo es estudiar cómo cambian algunas propiedades estructurales de las redes construidas, haciendo énfasis en un análisis de la modularidad cuando el parámetro de escala en el kernel de dispersión (*función de densidad de probabilidad Weibull*) cambia en un cierto valor constante, después de que un número fijo de redes han sido construidas.

Los datos e información acerca de la población de árboles que son utilizadas para crear las redes y hacer las simulaciones hechas en este trabajo son obtenidos de [1]. La población de árboles estudiados es una población constituida por 195 árboles, 131 de los cuales son árboles que tienen flores con funciones masculinas y femeninas (hermafroditas) y 64 son árboles que poseen solamente funciones femeninas (árboles femeninos).

3.2 Kernel de Dispersión

Con el propósito de establecer un evento de apareamiento entre dos individuos de la población de árboles (*hermafrodita-hermafrodita ó hermafrodita-hembra*) y así generar un enlace entre los nodos de la red que los representan se utiliza el concepto de *kernel de dispersión*.

Este concepto es de suma importancia en el presente trabajo, ya que los enlaces entre los nodos de la red se generan con la ayuda del kernel de dispersión. Para generar estos enlaces se calcula la distancia entre dos nodos de una red y se asigna una probabilidad de apareamiento como una función del kernel dispersión. Si esa probabilidad de apareamiento es mayor que un número aleatorio uniformemente distribuido entre 0 y 1, el enlace es creado entre los dos nodos [1].

La definición de kernel de dispersión según M. A. Lewis y J. Bullock en [14] se muestra a continuación.

Definición 7 (Kernel de Dispersión). Es la distribución de frecuencia de las distancias de dispersión de polen, basada en la identidad y ubicación de los árboles de apareamiento.

Como se hace referencia en [14], el kernel de dispersión ha sido objeto de estudio por parte de algunos matemáticos y biólogos, en donde la teoría matemática clásica de las ondas viajeras, ecuaciones diferenciales parciales no lineales, modelos integrales relacionados y estudios biológicos detallados, han mostrado que es el kernel de dispersión, el que determina la tasa a la cual las plantas pueden esparcirse espacialmente cuando son introducidas dentro de un nuevo ambiente [15], o cuando responden a cambios en las condiciones ambientales [16].

Ya que el *kernel de dispersión* no es otra cosa que una función de densidad de probabilidad, se presenta a continuación la definición de una *distribución de cola gruesa* (o pesada) como se presenta en [17].

Definición 8 (Distribución de Cola Pesada). La distribución de una variable aleatoria X con función de distribución B sobre $(0, \infty)$ se dice tener cola pesada, si

$$\int_0^\infty e^{\epsilon x} B dx = \infty \quad para \ todo \ \epsilon > 0 \tag{3.1}$$

De manera intuitiva, se puede decir que una distribución de cola gruesa (o cola pesada) hace referencia a una distribución cuya cola no es acotada exponencialmente. Es decir, es una distribución que tiene una cola más gruesa que la cola de una distribución exponencial.

Aunque en muchas aplicaciones son las distribuciones de cola derecha pesada las que son utilizadas, también pueden haber distribuciones de cola izquierda pesada o distribuciones con ambas colas pesadas.

Dicho lo anterior, se puede decir entonces que al hacer referencia a un kernel de dispersión de cola gruesa se está hablando de una función de densidad de probabilidad, cuya cola no es acotada exponencialmente. En el caso de este trabajo, el kernel de dispersión que se utilizó está determinado por una función de densidad de probabilidad Weibull.

3.3 Variación del Parámetro de Escala en el Kernel de Dispersión

La idea de variar el parámetro de escala en el kernel de dispersión tiene como propósito observar cómo cambian las propiedades estructurales de las redes construidas, enfocándose principalmente en cómo cambia la modularidad de dichas redes complejas.

Al aumentar el valor del parámetro de escala en el kernel de dispersión, la cola de la distribución se hace más gruesa. Esto trae como consecuencia que la masa de la distribución no esté concentrada en valores cercanos a cero.

En términos del proceso de apareamiento que se estudia en este trabajo, si se aumenta el valor del parámetro de escala en el kernel de dispersión, las probabilidades de ocurrencia de un evento de apareamiento entre árboles distantes también aumentarán; esto trae como consecuencia la aparición de un mayor número de enlaces entre los nodos de la red y por lo tanto redes más conectadas.

La pregunta que surge es: ¿Qué representa en términos biológicos un aumento en el valor del parámetro de escala en el kernel de dispersión? Una posible respuesta está basada en la idea que se propone en este trabajo. Esta consiste en considerar el parámetro de escala en el kernel de dispersión como un factor que permite tener en cuenta las influencias del medio ambiente en el cual se desarrollan y prosperan los polinizadores de las flores.

Dentro de estas influencias se tienen las de tipo abiótico; por ejemplo: la temperatura, el viento, la humedad, fragmentación de la población por diversas causas etc. En las de tipo biótico, una de las más relevantes, es la exposición de los polinizadores ante sus depredadores naturales cuando dejan sus lugares de protección para la recolección del polen.

De lo anterior, se tiene que una posible aproximación, es considerar el hecho de que un aumento en los valores del parámetro de escala estaría representando un aumento en las condiciones que favorecen el desempeño de los polinizadores en el transporte del polen. No sólo a los árboles en las vecindades de sus lugares de partida, sino también a aquellos árboles que se encuentran ubicados a distancias mayores, a las que normalmente no llegarían en situaciones desfavorables para la actividad de recolección de polen.

CAPÍTULO 4 REDES COMPLEJAS PARA MODELAR EL PROCESO BIOLOGICO

El propósito de este capítulo es describir de manera detallada todo el trabajo que se realizó para la construcción de las redes complejas, que modelan el proceso biológico de apareamiento entre los árboles de la población. Posteriormente son estas redes las que sirven como datos de entrada a un software llamado **NETCARTO**, desarrollado por **Roger Guimerà** y **Luis A.N. Amaral** ([18],[13]) y el cual implementa un algoritmo creado por ellos mismos basado en la idea de *Simulated Annealing*.

Este programa permite obtener datos como los módulos encontrados por el algoritmo e información sobre estos módulos tales como: el número de nodos en el módulo, los nodos que constituyen ese módulo y el número total de enlaces en el módulo (esto incluye los enlaces dentro del mismo módulo y enlaces desde un módulo hacia otros módulos).

También brinda información sobre el valor de la modularidad para la partición de la red encontrada por el algoritmo y los roles de cada uno de los nodos en la red.

El software brinda la oportunidad de encontrar un valor promedio de la modularidad, para diferentes redes creadas aleatoriamente con la misma distribución de grado que la red original (*"randomizations"*). En este trabajo, sólo se tiene en cuenta el valor de la modularidad encontrado para la red original que se da como dato de entrada al software.

4.1 Construcción de las Redes Complejas

Con el fin de estudiar la relación que hay entre la modularidad de una red y la ecología de una población de árboles que son polinizados por insectos, se construyeron varias redes teniendo en cuenta dos factores que pueden influenciar la modularidad.

El primero de estos factores es la distribución espacial de los nodos que representan los árboles de la población, y para esto se realizaron dos experimentos. En el primer experimento se ubicaron los árboles a lo largo de una superficie totalmente plana y siguiendo una distribución uniforme bivariada.

En el segundo experimento, se realizó una nueva ubicación de los nodos en el plano, pero esta vez la distribución de los nodos a lo largo del eje horizontal sigue una distribución uniforme, y a lo largo del eje vertical, los nodos siguen una distribución de probabilidad creada con el objetivo de simular algunas configuraciones espaciales, en donde los nodos tienden a agruparse hacia un lugar particular a medida que el parámetro de esta distribución cambia.

En este caso la distribución que se consideró tiene función de densidad

$$f(x;\lambda) = \begin{cases} -\frac{2x}{\lambda^2} + \frac{2}{\lambda} & si \qquad 0 \le x \le 1\\ 0 & si \quad x \in (-\infty, 0) \cup (1, \infty) \end{cases}$$
(4.1)

Para los experimentos realizados, los valores que se dieron al parámetro λ fueron de 1, 0.75, 0.65, 0.5, de tal manera que los nodos van a estar más agrupados hacia el cero a medida que el parámetro varia de 1 a 0.5.

Posteriormente, para estas configuraciones espaciales de los nodos se generan enlaces entre estos, de acuerdo a una función de densidad de probabilidad Weibull, en donde el parámetro de escala α se incrementa por un valor de 50 cada vez que 5 redes son creadas para un mismo valor de α (desde 75.7 hasta 325.7). Para cada configuración espacial se construyen 30 redes, teniendo así 150 redes creadas en total. La función de densidad de probabilidad de una variable aleatoria que se distribuye Weibull está determinada por:

$$f(x;\alpha;\beta) = \begin{cases} \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-(x/\alpha)^{\beta}} & si \quad x \ge 0\\ 0 & si \quad x < 0 \end{cases}$$
(4.2)

Donde $\beta > 0$ es el parámetro de forma y $\alpha > 0$ es el parámetro de escala de la distribución; para poder visualizar las redes construidas se utilizó el programa *PAJEK* versión 1.25, este fue desarrollado por *Vladimir Batagelj* del departamento de matemáticas, FMF y *Andrej Mrvar* de la facultad de ciencias Sociales, ambos de la Universidad de Ljubljana. Este programa permite la visualización y el análisis de redes que poseen miles de vértices (aunque puede llegar a soportar millones de vértices).

Las figuras 4–1, 4–2, 4–3,4–4, muestran distintas redes construidas para la configuración espacial, en la que horizontalmente los nodos siguen una distribución uniforme y verticalmente siguen la distribución descrita en la ecuación 4.1, con parámetros $\lambda = 1, 0.75, 0.65, 0.5$ respectivamente.

En estas figuras, el valor del parámetro de escala que se utiliza en el kernel de dispersión para generar los enlaces es de $\alpha = 75.7, 225.7, 175.7, 325.7$ respectivamente.

Por otro lado la figura 4–5 muestra una red construida en la que la configuración espacial de los nodos, está determinada por una distribución uniforme bivariada y en la que de nuevo el valor del parámetro de escala utilizado en el kernel de dispersión es $\alpha = 75.7$.

En todas las redes mostradas en las figuras 4–1, 4–2, 4–3, 4–4, 4–5; los nodos de color verde representan los árboles que tiene flores solo con funciones femeninas



Figura 4–1: Red donde los nodos están ubicados espacialmente siguiendo una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1 con parámetro $\lambda = 1$ en el eje vertical, en este caso el valor del parámetro α en el kernel de dispersión es 75.7

(hembras), y los nodos de color negro representan los árboles que poseen flores con funciones masculinas y femeninas (hermafroditas).

4.2 Implementación

El software en el que se desarrolló el programa para la construcción de las redes complejas, que modelan el fenómeno de apareamiento entre los árboles, fue MATLAB 7.0. a través de su entorno de desarrollo integrado (IDE).

El lenguaje de programación empleado es el lenguaje M, propio de MATLAB el cual permite un manejo sencillo para el trabajo con matrices.

El programa en realidad es una función que recibe como parámetros de entrada, el número de árboles de la población y el número de redes que se quieren crear para cada uno de los valores del parámetro de escala en el kernel de dispersión.



Figura 4–2: Red donde los nodos están ubicados espacialmente siguiendo una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1 con parámetro $\lambda = 0.75$ en el eje vertical, en este caso el valor del parámetro alfa en el kernel de dispersión es 225.7

Los datos de salida de esta función son cinco matrices, tres de la cuales son de tamaño $n \times n$ y dos son de tamaño $1 \times n$ donde n es el número de árboles que constituyen la población. La función al momento de ejecutarse en la ventana de comandos de MATLAB tiene el siguiente aspecto:

[Q,D,C,logico,t,t1]=posicionesgenerouniforme(n,repe)

La función anterior se utilizó para generar las redes cuyos nodos han sido configurados espacialmente mediante una distribución uniforme bivariada. Un ejemplo de una red creada con esta función es la que se observa en la Figura 4–5.

En el caso de las redes en donde la configuración espacial de los nodos sigue horizontalmente una distribución uniforme y verticalmente una distribución dada por la ecuación 4.1, con parámetro $\lambda = 1, 0.75, 0.65, 0.5$, la función en la ventana de comandos de MATLAB se ve de la siguiente manera:



Figura 4–3: Red donde los nodos están ubicados espacialmente siguiendo una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1 con parámetro $\lambda = 0.65$ en el eje vertical, en este caso el valor del parámetro alfa en el kernel de dispersión es 175.7

function [Q,D,C,logico,t,t1]=posicionesgeneroexponencial(n,repe)

Para ambas funciones, las variables **n** y **repe** almacenan los datos de entrada. La variable **n**, almacena el número de árboles de la población y la variable **repe** almacena el número de redes que se construyen para cada valor del parámetro de escala del kernel de dispersión (en el caso de los experimentos realizados, se tiene n=195 y **repe**=5).

La variable \mathbf{Q} almacena una matriz de tamaño $1 \times n$, sus entradas representan las identidades de cada uno de los árboles de la población; esta matriz está constituida por ceros y unos, en donde los unos representan árboles hermafroditas y los ceros representan árboles con funciones netamente femeninas. Esta asignación de identidad a los árboles está basada en la proporción total de árboles hermafroditas dentro de la población.



Figura 4–4: Red donde los nodos están ubicados espacialmente siguiendo una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1, con parámetro $\lambda = 0.5$ en el eje vertical, en este caso el valor del parámetro alfa en el kernel de dispersión es 325.7

La forma en la que se realiza la asignación de identidad a cada árbol es la siguiente: para cada árbol se genera un número aleatorio uniformemente distribuido entre 0 y 1, este número se compara con el valor de la proporción p de árboles hermafroditas. Si el valor de p es mayor que el número aleatorio, entonces al árbol se le asigna la identidad hermafrodita, si no la identidad del árbol será solamente femenina.

Las variables \mathbf{t} y $\mathbf{t1}$ almacenan matrices de tamaño $1 \times n$ cuyas entradas determinan las abscisas y las ordenadas respectivamente, de las parejas ordenadas que determinan la posición de un árbol en el plano. Por ejemplo la pareja ordenada (t(i), t1(i)) determina la posición del árbol *i* de la población.

La variable **D** almacena una matriz de tamaño $n \times n$ cuyas entradas d_{ij} (en el lenguaje de MATLAB estas entradas se representan como D(i, j)) son las distancias

98



Figura 4–5: Red donde los nodos están ubicados espacialmente siguiendo una distribución uniforme bivariada

entre el árbol i y el árbol j. En este caso, la escala que se usa para estas distancias es la mostrada en [1] en donde 0.85 cm equivalen a 100 m. Por otro lado, $d_{ij} = -1$ en el caso de que el árbol i y el árbol j tengan solamente funciones femeninas y $d_{ij} = 0$ cuando i = j. La forma en la que esta matriz se crea dentro del programa se muestra a continuación:

```
for i=1:n
    if(Q(i)==1)
    for j=1:n
        D(i,j)=norm([t(i) t1(i)]-[t(j) t1(j)],2)*(100/.85);
    end
    else
    for j=1:n
        if(Q(j)==0)
        D(i,j)=-1;
        else
        D(i,j)=norm([t(i) t1(i)]-[t(j) t1(j)],2)*(100/.85);
        end
    end
    end
end
end
```

La variable C por su parte, almacena una matriz de tamaño $n \times n$, esta matriz tiene como función mantener las mismas entradas de la matriz almacenada en la variable **D**, con la diferencia de que sus entradas cumplen $c_{ij} \ge 0$. Además es la matriz que se utiliza para la creación de la matriz que se almacenará posteriormente en la variable **logico**.

Por último se tiene la variable **logico**. Esta variable almacena una matriz de tamaño $n \times n$ que consta solamente de ceros y unos; $logico_{ij} = 1$ representa la existencia de un enlace entre el árbol i y el árbol j y $logico_{ij} = 0$ significa que no hay conexión mediante un enlace entre los nodos i y j. La forma en la que se crea esta matriz dentro del programa es la siguiente:

Esta parte del código es quizás una de las más importantes debido a que es donde se utiliza el kernel de dispersión para la creación de enlaces entre pares de árboles *hermafrodita-hermafrodita* y *hermafrodita-hembra*. La forma en la que se crea un enlace entre el árbol i y el árbol j es la misma empleada en [1].

La descripción de como en [1] se crea un enlace entre un par cualquiera de nodos es como sigue: Una probabilidad de apareamiento es asignada como una función del kernel de dispersión; si dicha probabilidad es mayor que un número aleatorio uniformemente distribuido entre 0 y 1 el enlace es creado. En el caso de este trabajo, este proceso se realizó hasta mil veces con el propósito de asegurar que el grado de cada uno de los nodos en la red sea mayor que cero.

La variable \mathbf{r} que aparece en la implementación del kernel de dispersión, es la encargada de la variación del parámetro de escala (en este caso un valor constante de 50 cada vez que el número de redes determinadas por **repe** hayan sido creadas).

Como las matrices de adyacencia que representan las redes construidas son simétricas. Se hace la asignación logico(j,i)=0 para evitar de nuevo la creación de otro enlace entre los árboles *i* y *j* en el caso de que ya se tenga que logico(i,j)=1.

4.3 El Software NETCARTO

El programa **NETCARTO** fue desarrollado por **Roger Guimerà** y **Luis A.N. Amaral**, este programa implementa un algoritmo basado en la idea de optimización por *Simulated Annealing* la cual a su vez fue desarrollada por **S. Kirkpatrick**, **C.D Gellat Jr** y **M.P Vecchi** en el año de 1983 y cuyo trabajo se muestra en [19].

Los conceptos sobre los cuales está apoyado el desarrollo del programa **NET-CARTO**, se exponen detalladamente en [13] y [18]; dentro de un marco general, ellos demuestran que pueden encontrar módulos en una red compleja y a su vez clasificar los nodos de una red dentro de ciertos roles definidos de manera heurística.

Según lo expuesto en [13], la forma en la que **Roger Guimerà** y **Luis A.N. Amaral** identifican los módulos de la red, es por medio de la optimización de la modularidad (ecuación 2.6) usando simulated annealing. El hecho de usar el concepto de *simulated annealing*, les permite no tener la necesidad de conocer una partición a priori de la red, o lo que es lo mismo, no necesitan saber con anterioridad el número de módulos. De hecho, este número de módulos resulta ser una de las salidas del algoritmo.

Como dato importante se debe decir que el algoritmo es capaz de identificar módulos en una red cuyos nodos, tienen hasta el 50% de sus enlaces fuera de su propio módulo [13]. Además cabe resaltar que el método propuesto en [13] utiliza el *simulated annealing*, para encontrar la partición de la red con la modularidad más alta.

4.4 Simulated Annealing

Según se expone en [13], el *simulated annealing* es una técnica de optimización estocástica desarrollada por S.Kirkpatrick, C.D Gellat Jr y M.P Vecchi en [19], esta permite obtener una buena aproximación al valor óptimo global de una función, cuando el número de estados es muy difícil de tratar de forma combinatoria.

En este caso, la palabra estados hace referencia a los elementos del conjunto de partida de la función para el problema de encontrar la modularidad de una red. Dichos estados son las diferentes particiones de la red y la función es precisamente la función de modularidad descrita por la ecuación 2.6. Cuando se trata de encontrar la modularidad de una red lo que se busca mediante el *simulated annealing* es encontrar para que partición de la red, el valor de la función de modularidad es el más alto.

El término *annealing* proviene de la analogía del proceso físico de enfríamiento de una sustancia; el análisis del comportamiento de una sustancia cuando esta se enfría, es un tema de especial atención para la mecánica estadística.

Dicho esto, la idea del simulated annealing expuesta en [20] es la siguiente: Cuando una sustancia es sometida a una temperatura alta, las moléculas que la conforman presentan un alto grado de movilidad, cuando esta temperatura empieza a decrecer, la movilidad de estas moléculas va disminuyendo, llevando a éstas a una posible disposición, en la que se encuentren alineadas permitiendo una estructura cristalina en la sustancia. Esta disposición de alineamiento de las moléculas es el estado de menor energía para el sistema.

Es importante decir que al usar la palabra "posible" se está indicando que la temperatura no es la única que rige el sistema cuando la sustancia ha alcanzado un estado de mínima energía. Como la idea es alcanzar dicho estado, el enfríamiento debe realizarse a una tasa lo suficientemente lenta, de tal forma que las moléculas puedan tener una mayor oportunidad de alinearse correctamente antes de una disminución considerable en su movilidad.

Si por el contrario, la tasa de enfríamiento es demasiado rápida, se corre el riesgo de que las moléculas no logren una alineación correcta y como resultado se obtenga un estado policristalino. Es decir, un estado en el que las moléculas carecen de una estructura ordenada y donde el estado de la sustancia, no presenta un valor mínimo de energía. De esta forma, se tiene que el principio de annealing es aquel que busca enfríar lentamente una sustancia para poder alcanzar un estado de mínima energía.

Es importante decir que el algoritmo basado en el simulated annealing está diseñado para ir a través de los mínimos locales con el propósito de poder ir hacia el mínimo globlal de la función $L = L(\theta)$, algunas veces conocida como función objetivo o función de pérdida.

Como la idea era buscar un análogo en optimización de lo que es un estado de mínima energía para un sistema, se introdujo el concepto de función de pérdida de tal forma que la idea consiste en reducir al valor mínimo dicha función. De esto se tiene que la técnica del *simulated annealing* tiene como propósito intentar matemáticamente capturar el proceso de enfríamiento controlado asociado al proceso físico y donde el objetivo principal es poder alcanzar el valor más bajo de la función de pérdida, de cara a un posible mínimo local [20]. Al igual que en el proceso de enfríamiento de una sustancia, en el que el sistema puede llegar a alcanzar estados temporales de alta energía cuando las moléculas están intentando alinearse; el simulated anneling permite que la función de pérdida, presente algunos incrementos temporales cuando se está llevando a cabo el proceso de capturar información para alcanzar el mínimo global.

Dicho lo anterior, se puede decir que la parte complicada del algoritmo de *simulated annealing*, es el análogo matemático de la tasa de enfríamiento en el proceso físico.

Una de las principales propiedades que tiene el simulated annealing, es la disposición para sacrificar la ganancia de un decrecimiento rápido en la función de pérdida, de tal forma que se permitan incrementos temporales de los valores de la función. Dicha propiedad es derivada de la distribución de probabilidad Boltzmann-Gibbs, donde la probabilidad de que un sistema tenga un estado discreto de energía x está dada por:

$$P(\text{estado de energía} = x) = C_T \exp\left(-\frac{x}{C_b T}\right)$$
 (4.3)

Donde $C_T > 0$ es una constante normalizadora, $C_b > 0$ es la constante de Boltzmann y T es la temperatura del sistema. De la ecuación 4.3 se tiene que cuando las temperaturas son altas, el sistema tendra más probabilidad de encontrarse en un estado de alta energía que cuando las temperaturas son bajas. Según [20] la analogía de optimización se deriva de el hecho de que aún cuando las temperaturas son bajas, existe una probabilidad distinta de cero de que el sistema alcance un estado de alta energía.

En [21] N. Metropolis introdujo la distribución Boltzmann-Gibbs (ecuación 4.3) con la idea de poder construir un medio para simular un sistema que se encuentra a alguna temperatura fija. De esta forma, si un sistema se encuentra en un estado de energía ϵ_{curr} y algunos aspectos se cambian, de tal manera que el sistema pueda

alcanzar un nuevo estado de energía ϵ_{new} , entonces la simulación de Metropolis, siempre hace que el sistema vaya hacia el nuevo estado de energía, si $\epsilon_{new} < \epsilon_{curr}$ en el caso de que $\epsilon_{curr} \ge \epsilon_{new}$ entonces la propabilidad de que el sistema vaya hacia el nuevo estado está dada por:

$$\exp\left(-\frac{\epsilon_{new} - \epsilon_{curr}}{C_b T}\right) \tag{4.4}$$

La expresión en 4.4 es conocida como el *criterio de Metropolis*, la importancia de este criterio rádica en el hecho de que después de que se han realizado un número suficiente de dichas desiciones y respuestas, el sistema eventualmente estará en un equilibrio, donde el sistema además estará gobernado por la distribución de Boltzmann-Gibbs cuando está a una temperatura fija T.

4.4.1 Del Criterio de Metropolis a la distribución Boltzmann-Gibbs

A continuación se muestra un bosquejo de como se obtiene la distribución de Boltzmann-Gibbs 4.3 a partir de repeticiones del criterio de metropolis. Este bosquejo es tomado de [20], y es mostrado a continuación por la pertinencia que este hecho tiene en la construcción del algoritmo de simulated annealing.

Suponga que ϵ_{new} y ϵ_{curr} puede tomar uno de los N valores x_1, x_2, \ldots, x_N , el propósito es mostrar que después de llevar a cabo varias repeticiones de la ecuación 4.4 se obtiene $P(\text{estado de energia} = x_i) = C_T \exp[-x_i/(C_bT)]$ para cada $i = 1, 2, \ldots, N$.

Se debe tener en cuenta que la probabilidad de ir de un estado i a cualquier estado j resulta ser el producto de dos probalidades: (i) la probabilidad de elegir a jcomo el próximo estado, dado que el sistema se encuentra en el estado i y (ii) dado que el estado j ha sido elegido, la probabilidad de que realmente vaya al estado jdesde el estado i. Con el objetivo de llevar a cabo una simplificación, se supone que la probabilidad en (i) de elegir un estado cualquiera desde cualquier estado es igual y está dada por 1/N. Para el caso (ii) la probabilidad de aceptación está dada por la ecuación 4.4.

Dicho lo anterior, el objetivo es encontrar las probabilidades de transición de estado, p_{ij} de ir desde un estado *i* hacia un estado *j*, estas probabilidades son los componentes de una matriz de transición. Sea a_{ij} la probabilidad de aceptar la transición desde el estado x_i hacia el estado x_j . En particular, la probabilidad de aceptación para $i \neq j$ desde la ecuación 4.4 está dada por:

$$a_{ij} = \left\{ \begin{array}{ccc} 1, & si & x_j < x_i \\ \exp\left(-\frac{x_j - x_i}{C_b T}\right), & si & x_j \ge x_i \end{array} \right\}$$
(4.5)

Se tiene entonces, que la matriz $\mathbf{P} = [p_{ij}]$ es la matriz de transición. Para $i \neq j$, se tiene que $p_{ij} = a_{ij}/N$. Ya que existe una probabilidad positiva de que el proceso permanezca en el estado *i* aun cuando la energía es más baja en el estado *j*, la ecuación 4.5 no aplica a los elementos p_{ii} . Sin embargo, como \mathbf{P} es una matriz de transición se debe cumplir que la suma de los elementos de un fila cualquiera de *P* debe ser igual a uno, por lo tanto se tiene:

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij}$$

$$= 1 - \frac{1}{N} \sum_{j \neq i} a_{ij}$$
(4.6)

De esta manera, la matriz de transición que da las probabilidades de ir de un estado i hacia un estado j es:

$$P = \begin{bmatrix} 1 - \frac{1}{N} \sum_{k \neq 1} a_{1k} & \frac{1}{N} a_{12} & \dots & \frac{1}{N} a_{1N} \\ \frac{1}{N} a_{21} & 1 - \frac{1}{N} \sum_{k \neq 2} a_{2k} & \ddots & \frac{1}{N} a_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N} a_{N1} & \frac{1}{N} a_{N2} & \dots & 1 - \frac{1}{N} \sum_{k \neq N} a_{Nk} \end{bmatrix}$$
(4.7)

La importancia de la matriz que se muestra en 4.7, está dada por el hecho de que dicha matriz, se puede utilizar para encontrar la distribución estacionaria que representa las probabilidades de los estados de energia del sistema, estando en cualquiera de los N estados.

Si esta distribución estacionaria N-dimensional se denota mediante el vector $\overline{\mathbf{p}}$, donde la *i*-ésima entrada corresponde a la probabilidad de que el estado de energía sea igual a x_i . Este vector cumple que $\overline{\mathbf{p}}^T = \overline{\mathbf{p}}^T \mathbf{P}$, y por lo tanto esta expresión da origen a N ecuaciones y N incognitas las cuales son los elementos \overline{p}_i .

Si se asume que $\sum_{i=1}^{N} x_i = constante$, entonces la solución a cada uno de estas ecuaciones corresponde a la distribución Boltzmann-Gibbs. Es decir, $\overline{p}_i = P(\text{estado de energia} = x_i) = C_T \exp\left[-x_i/(C_bT)\right]$ para todo *i*.

La idea de utilizar el simulated annealing para identificar los módulos de una red está explicada de manera detallada en la sección de $M\acute{e}todos$ en [13].

4.4.2 Algoritmo de Simulated Annealing

A continuación se muestra la forma general del algoritmo de simulated annealing, la idea principal de este se debe al trabajo realizado por Kirkpatrick S, Gelatt CD. Jr, Vecchi MP en [19], ellos introdujeron la forma moderna del algoritmo de simulated annealing usando el criterio de Metropolis con el propósito de optimización.

Se tiene que $L(\theta), \theta \in \Theta \subseteq \mathbb{R}^p$, de está forma los pasos generales del algoritmo de simulated annealing cuando hay medidas perfectas de L son los siguientes [20]:

- Paso 1 (**Inicialización**) Establezca una temperatura inicial T y un vector parámetro inicial $\hat{\theta}_0 = \theta_{curr} \in \Theta$; determine $L(\theta_{curr})$.
- Paso 2 Relativo al valor actual de θ_{curr} , aleatoriamente determine un nuevo valor de θ , $\theta_{new} \in \Theta$ y determine $L(\theta_{new})$
- Paso 3 Compare los dos valores de L obtenidos usando el criterio de Metropolis. Sea $\delta = L(\theta_{new}) - L(\theta_{curr})$. Si $\delta < 0$ entonces acepte θ_{new} . Si $\delta \ge 0$ entonces, acepte θ_{new} solamente si una variable aleatoria uniformemente distribuida en (0, 1)

(generada por el método de Montecarlo) satisface $U \leq \exp[-\delta/(C_bT)]$ (sin pérdida de generalidad se puede suponer $C_b = 1$ ya que el usuario tiene control total sobre T). Si θ_{new} es aceptada, entonces θ_{curr} es reemplazada por θ_{new} ; si no entonces θ_{curr} permanece como está.

- Paso 4 Repita los pasos 1 y 2 por algún periodo hasta que el presupuesto de las evaluaciónes de la función asignadas para ese T hayan sido usadas o el sistema alcance algún estado de equilibrio.
- Paso 5 Disminuya el valor de T según el annealing acordado y retorne al paso 1. Continue el proceso hasta que el presupuesto total para las evaluaciones de la función haya sido usado o alguna indicación de convergencia sea satisfecha (Análogo al sistema siendo congelado en su estado de energía mínimo). La estimación final es $\hat{\theta}_n$ (que se toma como el más reciente θ_{curr}) representando el valor θ después de n iteraciones.

Es importante resaltar que haciendo los respectivos cambios, el algoritmo de simulated annealing provee una herramienta para localizar los valores máximos de una función objetivo.

CAPÍTULO 5 RESULTADOS

En este capítulo se presentan los resultados obtenidos en los diferentes experimentos realizados. Estos tienen como propósito detectar los cambios estructurales en las redes que surgen a partir del intento por modelar el proceso de apareamiento en una población de árboles, cuya polinización es llevada a cabo por insectos y en donde se tienen en cuenta factores abióticos que influyen en la actividad polinizadora de dichos insectos; el énfasis principal se hace en la propiedad estructural de la modularidad. Con ayuda de un software especializado se logran obtener valores numéricos de la modularidad para cada una de las redes y estos valores son los que permitirán describir los posibles efectos biológicos que se pueden presentar en las poblaciones de árboles cuyos individuos tienden a formar grupos por causa de diversos factores, tales grupos en ocasiones tienden a distanciarse genéticamente unos de otros lo que contribuye a la exposición de la especie al peligro de extinción.

5.1 Las redes y su uso para modelar el problema

Las redes creadas surgen a partir de la idea de construir un mecanismo que permitiera cuantificar diferentes procesos que se llevan dentro de una población de árboles. En el caso de este trabajo, el único proceso que se tuvo en cuenta fue el de apareamiento entre individuos dentro de la población, dicho apareamiento se lleva a cabo a través de la polinización por medio de insectos; de esta manera si se tienen en cuenta los factores que pueden influenciar la actividad polinizadora de los insectos la idea es poder detectar que implicaciones trae esto sobre el flujo de polen dentro de la población, permitiendo esto a su vez estudiar cómo los individuos comparten su material genético con los demás.

Ya que los árboles no presentan movimiento, la idea es considerarlos como los nodos de una red que en principio está totalmente desconectada, las conexiones entre pares de nodos se pondrán de manera estocástica usando un kernel de dispersión y cada conexión o enlace, representará el proceso de apareamiento entre pares de individuos dentro de la población. Es decir, para un par de nodos cualesquiera dentro de la red se tiene que, dado que estos representan árboles, la aparición de un enlace entre ellos determina un evento exitoso de apareamiento entre dicho par de árboles; en este caso las identidades de los individuos que se tienen son árboles hermafroditas y árboles hembras, cada uno de estos identificados dentro de las redes con nodos de diferentes colores (negro para los hermafroditas y verdes para las hembras).

Se estudian diferentes configuraciones espaciales para la misma población de árboles, es decir un mismo árbol se ubica en distintas posiciones dentro de un área determinada fija, dichas posiciones están determinadas por coordenadas obtenidas a partir de la simulación de dos variables aleatorias una de las cuales determina la coordenada horizontal y la otra la coordenada vertical. Las variable aleatoria que determina la coordenada horizontal en todos los experimentos sigue una distribución uniforme. En el caso de la variable aleatoria que determina la coordenada vertical ésta sigue una distribución uniforme en uno de los experimentos y una distribución determinada por la ecuación 4.1 en los demás experimentos.

Este tipo de modelaje espacial en el que son simuladas diferentes posiciones de un mismo árbol dentro de la población o en el que las posiciones de los nodos están determinadas por las coordenadas geográficas de los arboles de una población en estudio, es utilizada también en [1], [11].

El objetivo de utilizar esta ultima distribución consiste en poder simular diferentes condiciones de agrupamiento de los árboles en la población, este agrupamiento está determinado por los valores del parámetro λ en la ecuación 4.1; fueron considerados los valores $\lambda = 1, 0.75, 0.65, 0.5$. A medida que estos valores decrecen hacia cero, los nodos tienden a concentrarse hacia cero; intentando simular con esto algunos posibles escenarios espaciales que pueden presentarse en la naturaleza debido a la competencia por la búsqueda de condiciones favorables para el desarrollo y crecimiento de cada individuo dentro de la población.

Para el caso de la generación de un enlace entre un par de nodos en la red que representa un evento exitoso de apareamiento entre los árboles representados por dichos nodos, la idea de tener en cuenta los factores abióticos que influyen en la actividad polinizadora de los insectos se intenta simular mediante la variación del parámetro de escala α en el kernel de dispersión.

Para este trabajo, el kernel de dispersión está determinado por la función de densidad de una distribución de probabilidad Weibull. Este kernel de dispersión está catalogado como un kernel de cola gruesa que para este caso trae como consecuencia la generación de enlaces entre pares de nodos que representan árboles separados por largas distancias [1].

A medida que el valor de dicho parámetro es aumentado se obtiene una cola más gruesa en el kernel de dispersión y por lo tanto un aumento en la ocurrencia de eventos de apareamiento entre pares de árboles, esto se ve reflejado en la aparición de nuevos enlaces dentro de la red y por lo tanto la obtención de redes más conectadas que presentan menores valores de la modularidad, significando esto una menor fragmentación de dichas redes en grupos dentro de los cuales los eventos de apareamiento tienen mayor probabilidad de ocurrir entre los individuos que los conforman.

5.2 Resultados

Con ayuda del programa **NETCARTO** se obtienen los niveles de modularidad para cada una de las redes, además se obtienen los promedios, las varianzas y los coeficientes de variación de los niveles de modularidad de las redes creadas, estos datos son obtenidos para cada uno de los valores del parámetro de escala de la función de densidad de probabilidad, que determina los enlaces entre pares de nodos de una red. La tabla 7–1 muestra esta información cuando los nodos son configurados espacialmente siguiendo una distribución uniforme bivariada, las gráficas de estos datos se muestran en las figuras 5–1, 5–2 y 5–3.

En las figuras 5–1, 5–5, 5–9, 5–13 y 5–17 se puede observar cómo la modularidad de las redes construidas, disminuye a medida que los valores del parámetro α aumentan, debido a que mientras mayor sea el valor en el parámetro de escala, la cola del kernel de dispersión se hace más gruesa causando un incremento en los eventos de polinización entre árboles ubicados a larga distancia y esto a su vez trae como consecuencia una reducción de la estructura modular de las redes construidas.

En las figuras 5–2, 5–6, 5–10, 5–14 y 5–18 la curva azul une los valores promedios de la modularidad obtenidos para cada uno de los valores del parámetro α considerados. Como se puede observar en las tablas 7–1, 7–2, 7–3, 7–4 y 7–5 estos valores promedios de la modularidad disminuyen conforme el valor del parámetro aumenta.

Es decir, que cuando las condiciones del entorno aumentan en favor de los polinizadores, su actividad de polinización se hará con mayor efectividad llegando a árboles que bajo condiciones adversas no llegarían. Esto permite la aparición de nuevos enlaces y una disminución en los valores promedios de la modularidad. Dicha disminución en los valores de la modularidad significa que se está favoreciendo el apareamiento entre los individuos de la población y como consecuencia se permitirá un incremento en el flujo genético entre individuos a través de la población.

La importancia de que exista un aumento en el flujo genético dentro de la población de árboles, radica en el hecho de que se podrán mitigar los efectos de la *deriva genética* es decir, que se pueda aumentar la variación genética dentro de la

población de tal forma que se tenga mayor oportunidad de aumentar las frecuencias de aquellos alelos que están cercanos a desaparecer dentro del patrimonio genético de la población y por lo tanto permitir una mayor diversidad genética.

Para el caso de la población estudiada en este trabajo, resulta de gran importancia la reaparición de nuevos alelos que diversifiquen el mapa genético de la población, esta nueva información genética puede provenir de poblaciones cercanas que transmiten su información genética a través de polinizadores o a partir de la vinculación de nuevos individuos provenientes de semillas de otras poblaciones y que fueron introducidos a través de algun medio dentro de la población en estudio, permitiendo esto una posibilidad para incrementar la diversificación del mapa genético de la población estudiada.

De esta manera, la inclusión de nuevo material genético dentro de la población en estudio tendrá mayor oportunidad de prevalecer y expandirse en la medida que las redes construidas presenten una menor estructura modular, ya que si la red está constituida por varios módulos puede que esa nueva información genética arrive dentro de un módulo que presente muy poca conexión a los otros módulos de la red y de esta manera se disminuya la posibilidad de compartir esa nueva información genética con el resto de la población.

De igual forma se puede decir que una reducción en la estructura modular de las redes, permite que no se presente una fragmentación de la población en poblaciones de menor tamaño que con el transcurrir del tiempo, se distanciarán unas de otras genéticamente debido a que los alelos que se pierdan en una población podrán ser aquellos que se acentúen en otra, generando con el transcurrir del tiempo un aislamiento entre dichas poblaciones.

De la ultima columna en las tablas 7–1, 7–2, 7–3, 7–4 y 7–1 se puede concluir que los valores de la modularidad no presentan una variabilidad alta cuando una red nueva es construida utilizando el mismo valor del parámetro α y la misma disposición espacial de los nodos en la población ya que el coeficiente de variación es menor al 30%, esto quiere decir que mientras las condiciones del entorno para los polinizadores permanezcan constantes (un mismo valor del parámetro) su actividad de polinización será similar y los eventos de apareamiento que se generen contribuirán a la generación de redes estructuralmente semejantes en términos de la modularidad.

En la figura 5–4 se muestra la misma red que la mostrada en la figura 4–5, pero en este caso se muestran los módulos detectados por el programa **NETCARTO**, en este caso los tres módulos detectados están representados mediante los colores amarillo, azul y rojo. Esto indica que los árboles representados con un mismo color presentan un mayor número de interacciones entre ellos que con árboles representados con un color diferente. Por ejemplo un árbol marcado con color azul, tiene más posiblidad de compartir su polen con árboles de color azul que con árboles marcados con color amarillo o rojo.



Figura 5–1: Los puntos rojos son los datos de la Modularidad mostrados en la tabla 7–1, en esta gráfica se puede observar como los valores de la modularidad decrecen a medida que el valor del parámetro α aumenta de valor, esto indica que a medida que los individuos logran establecer vinculo de apareamiento con individuos a distancias lejanas, la red estará más conectada y su estructura modular se reducirá.

Para las redes cuyos nodos siguen una distribución de probabilidad uniforme a lo largo del eje horizontal y una distribución dada por la expresión 4.1 con parámetros



Figura 5–2: La curva azul une los puntos del valor promedio de la modularidad para cada valor del parámetro α , la disminución en los valores de la modularidad están dados por el aumento en el grosor de la cola del kernel dispersión, en tal caso las redes están más conectadas y el flujo genético entre los árboles de la población aumentará en favor de controlar la aparición de genes que pueden poner en peligro la integridad de la población.

 $\lambda=1, 0.75, 0.65, 0.5$ en el eje vertical, esta información se muestra en las tablas 7–2, 7–3, 7–4, 7–5 respectivamente.

Como se puede observar en las tablas 7–2,7–3,7–4,7–5 al igual que en los experimentos en donde la disposición de los árboles estaba determinada por una configuración espacial uniforme bivariada, los valores de la modularidad también decrecen a medida que el valor del parámetro de escala α aumenta su valor. En estas tablas se puede observar que los valores promedios de la modularidad también decrecen a medida que el valor del parámetro λ en la distribución determinada por la ecuación 4.1 disminuye, es decir que a medida que los árboles tienden a estar más agrupados hacia un lugar particular del área en donde ellos se encuentran, la estructura modular de la red disminuye, es decir la tendencia de la red a fragmentarse en grupos es menor debido a que la cercania entre los individuos de la población permite una mayor interacción entre todos los individuos.



Figura 5–3: Los puntos verdes son los valores de la varianza de la modularidad para cada valor del parámetro α mostrados en la tabla 7–1.

También se puede concluir de los resultados mostrados en las tablas anteriores, que existe consistencia con el hecho de que si los polinizadores tienen condiciones favorables para llevar a cabo la actividad de polinización, los eventos de apareamiento entre pares de árboles ocurrirán en un mayor número, trayendo esto como consecuencia un mayor flujo de polen entre los individuos de la población y por lo tanto la obtención de redes más conectadas.

Las gráficas de los datos mostrados en las tablas 7–2 se muestran en las figuras 5–5,5–6 y 5–7

En la figura 5–8 se puede observar la misma red creada en la figura 4–1, en este caso se muestran los nodos con tres diferentes colores, estos colores corresponden a los tres módulos detectados por el programa **NETCARTO**, se debe decir que los nodos de color verde en esta figura son nodos que pertenecen a un mismo módulo y no son necesariamente árboles con flores cuyas funciones son netamente femeninas como en la figura 4–1.

Para el caso de los datos mostrados en la tabla 7–3 las gráficas se muestran en las figuras 5–9, 5–10, y 5–11



Figura 5–4: Los diferentes módulos detectados por el programa **NETCARTO** se representan por medio de los colores amarillo, azul y rojo. Es decir, los nodos de un mismo color, presentan mayor número de enlaces entre ellos comparado con el número de enlaces hacia nodos con otro color.

Para el caso en el que los nodos siguen una distribución en el eje vertical determinada por la ecuación 4.1 con parámetro $\lambda = 0.75$ la figura 5–12 muestra los módulos detectados por el programa **NETCARTO**, en este caso el parámetro de escala en el kernel de dispersión tiene un valor de $\alpha = 225.7$. A diferencia de la figura 4–2 en este caso lo nodos de color verde corresponden a uno de los módulos detectados por el programa, estos no necesariamente coinciden con los árboles cuyas flores presentan funciones netamente femeninas.

Respectivamente en las figuras 5–13, 5–14 y 5–15 se pueden observar las gráficas de los datos que aparecenen la tabla 7–4

La red que se muestra en la 5–16, es la misma red de la figura 4–3 pero en este caso, se muestran los nodos detectados por el programa **NETCARTO**, en este caso el programa detecto dos módulos en la red, uno de ellos conformado por los nodos de color rojo y el otro conformado por los nodos de color amarillo, está red



Figura 5–5: Los puntos rojos son los datos de la Modularidad mostrados en la tabla 7–2, se puede observar como los valores de la modularidad decrecen cuando ocurre un aumento en el valor del parámetro α , este decrecimiento de los valores de la modularidad reflejan una disminución en la estructura modular de la red, lo que significa mayores oportunidades para que los individuos puedan compartir su material genético a través de toda la población

fue construida con un parámetro de escala del kernel de dispersión $\alpha = 175.7$ y la distribución en el eje vertical está determinada por la ecuación 4.1 con parámetro $\lambda = 0.65$. El hecho de que en está red solo aparezcan dos módulos se debe a que el agrupamiento entre los árboles es mayor que en los anteriores valores dados para el parámetro λ , está disminución en el número de módulos indica que existe una mayor interacción entre todos los árboles de la población, esto trae como consecuencia un mayor intercambio de información genética entre todos los individuos que constituyen la población, mitigando esto los efectos de la deriva genética y permitiendo una mayor diversidad en el patrimonio genético, de tal forma que esto resulte en beneficio de la preservación de la especie.

Por último, para los datos mostrados en la tabla 7–5 las gráficas se pueden observar en las figuras 5-17, 5-18 y 5-19.

De manera similar a los casos anteriores, en la figura 5-20 se muestra la misma red que se visualiza en la 4-4 pero con los módulos detectados por el programa



Figura 5–6: La curva azul une los puntos del valor promedio de la modularidad para cada uno de los valores del parámetro α mostrados en la tabla 7–2, la disminución en los valores de la modularidad están dados por el aumento en el grosor de la cola del kernel de dispersión, en tal caso las redes están más conectadas y el flujo genético entre los árboles de la población aumentarán en favor de controlar la aparición de genes que puedan poner en peligro la integridad de la población.

NETCARTO, en este caso la distribución en el eje vertical de los árboles está determinada por la ecuación 4.1 con parámetro $\lambda = 0.5$ y el parámetro en el kernel de dispersión es se tomó como $\alpha = 325.7$

Igual que en el caso de la figura 5–16, la aparición de sólo dos módulos en la red es consecuencia del agrupamiento que presentan los nodos, permitiendo la posibilidad de que un individuo dentro de la población pueda establecer contactos con un número mayor de individuos debido a la cercania a la que se encuentran unos de otros.

La figura 5–21 muestra las curvas de los valores promedio de la modularidad para las diferentes configuraciones espaciales de los individuos de la población, en esta se puede observar como los valores promedios de la modularidad para la configuración espacial determinada por la distribución uniforme bivariada, son mayores que los valores promedios obtenidos para las configuraciónes espaciales cuya distribución



Figura 5–7: Los puntos verdes son los valores de la varianza de la modularidad para cada valor del parámetro α mostrados en la tabla 7–2. en el eje vertical estaba determinada por la ecuación 4.1 y seguian una distribución uniforme en el eje horizontal.

Una posible explicación a esta situación es que en el caso en el que los árboles tengan una mayor dispersión en el espacio (la cual está simulada por la distribución uniforme bivariada) será más difícil para ellos establecer vínculos de apareamiento con individuos que se encuentra ubicados a largas distancias desde sus posiciones, debido a que los polinizadores preferirán recorrer distancias cortas en la búsqueda de alimento y refugio. Esto contribuye a que se conformen grupos de árboles en los que los eventos de apareamiento tengan una mayor posibilidad de ocurrir debido a la cercanía de árboles vecinos; son estos grupos los que contribuyen al aumento en la estructura modular de las redes. Cabe resaltar que estos módulos están principalmente determinados por el número de enlaces que ocurren entre los nodos que conforman cada uno de estos.

Por otro lado en el caso de la población de árboles en donde el mayor número de los individuos se encuentran cercanos unos de otros, debido a las condiciones que el entorno les ofrece (simulada por la distribución dada por la ecuación 4.1 con



Figura 5–8: Red cuyos nodos siguen una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1 en el eje vertical con parámetro $\lambda = 1$ y con parámetro $\alpha = 75.7$ en el kernel de dispersión. Los módulos detectados por el programa NETCARTO corresponden a los colores verde, rojo y amarillo.

parámetro $\lambda = 0.5$ en el eje vertical y la distribución uniforme en el eje horizontal). Se obtienen datos de la modularidad más bajos debido a que los eventos de apareamiento entre los individuos de la población, podrán ocurrir con mayor probabilidad debido a la cercanía de estos. Esta cercanía favorece el transporte del polen por parte de los polinizadores y por consiguiente las redes resultantes reflejarán una disminución en la tendencia a la formación de grupos ya que se facilitan los vínculos de apareamiento entre todos los individuos.

Por ejemplo, puede ocurrir que cerca de un suelo rico en minerales el número de árboles de una población sea mayor que el número de árboles presentes en lugares en donde esa riqueza de minerales va disminuyendo, en tal caso se tendrá que la formación de un grupo numeroso de árboles cerca de la fuente de minerales, resultará desfavorable para los individuos que se encuentren a largas distancias del grupo más



Figura 5–9: Los puntos rojos son los datos de la Modularidad mostrados en la tabla 7–3, se puede observar como los valores de la modularidad decrecen cuando ocurre un aumento en el valor del parámetro α , este decrecimiento de los valores de la modularidad reflejan una disminución en la estructura modular de la red, lo que significa mayores oportunidades para que los individuos puedan compartir su material genético a través de toda la población

grande, poniendo en riesgo su intercambio de polen con el grupo más numeroso y contribuyendo a la fragmentación de la población en grupos que eventualmente estarán aislados unos de otros.

En la gráfica que se muestra en la figura 5–22 se puede observar como es el cambio de los valores de la modularidad, en términos de los parámetros λ y α en el caso de los nodos que presentan una configuración espacial determinada por una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1 en el eje vertical.


Figura 5–10: La curva azul une los puntos del valor promedio de la modularidad para cada uno de los valores del parámetro α mostrados en la tabla 7–3, la disminución en los valores de la modularidad están dados por el aumento en el grosor de la cola del kernel de dispersión, en tal caso las redes están más conectadas y el flujo genético entre los árboles de la población aumentarán en favor de controlar la aparición de genes que puedan poner en peligro la integridad de la población.



Figura 5–11: Los puntos verdes son los valores de la varianza de la modularidad para cada valor del parámetro α mostrados en la tabla 7–3.



Paiek

Figura 5–12: Red cuyos nodos siguen una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1 en el eje vertical con parámetro $\lambda = 0.75$ y con parámetro $\alpha = 225.7$ en el kernel de dispersión. Los módulos detectados por el programa NETCARTO corresponden a los colores verde, rojo y amarillo.



Figura 5–13: Los puntos rojos son los datos de la Modularidad mostrados en la tabla 7–4, se puede observar como los valores de la modularidad decrecen cuando ocurre un aumento en el valor del parámetro α , este decrecimiento de los valores de la modularidad reflejan una disminución en la estructura modular de la red, lo que significa mayores oportunidades para que los individuos puedan compartir su material genético a través de toda la población



Figura 5–14: La curva azul une los puntos del valor promedio de la modularidad para cada uno de los valores del parámetro α mostrados en la tabla 7–4, la disminución en los valores de la modularidad están dados por el aumento en el grosor de la cola del kernel de dispersión, en tal caso las redes están más conectadas y el flujo genético entre los árboles de la población aumentarán en favor de controlar la aparición de genes que puedan poner en peligro la integridad de la población.



Figura 5–15: Los puntos verdes son los valores de la varianza de la modularidad para cada valor del parámetro α mostrados en la tabla 7–4.



Paiek

Figura 5–16: Red cuyos nodos siguen una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1 en el eje vertical con parámetro $\lambda = 0.65$ y con parámetro $\alpha = 175.7$ en el kernel de dispersión. Los módulos detectados por el programa NETCARTO corresponden a los colores rojo y amarillo.



Figura 5–17: Los puntos rojos son los datos de la Modularidad mostrados en la tabla 7–5, se puede observar como los valores de la modularidad decrecen cuando ocurre un aumento en el valor del parámetro α , este decrecimiento de los valores de la modularidad reflejan una disminución en la estructura modular de la red, lo que significa mayores oportunidades para que los individuos puedan compartir su material genético a través de toda la población



Figura 5–18: La curva azul une los puntos del valor promedio de la modularidad para cada uno de los valores del parámetro α mostrados en la tabla 7–5, la disminución en los valores de la modularidad están dados por el aumento en el grosor de la cola del kernel de dispersión, en tal caso las redes están más conectadas y el flujo genético entre los árboles de la población aumentarán en favor de controlar la aparición de genes que puedan poner en peligro la integridad de la población.



Figura 5–19: Los puntos verdes son los valores de la varianza de la modularidad para cada valor del parámetro α mostrados en la tabla 7–5.



Pajek

Figura 5–20: Red cuyos nodos siguen una distribución uniforme en el eje horizontal y una distribución determinada por la ecuación 4.1 en el eje vertical con parámetro $\lambda = 0.5$ y con parámetro $\alpha = 325.7$ en el kernel de dispersión. Los módulos detectados por el programa NETCARTO corresponden a los colores verde y amarillo.



Figura 5–21: Curva de la modularidad promedio obtenidas para las diferentes configuraciones espaciales de los árboles. En esta gráfica se puede observar como los valores de la modularidad disminuyen a medida que la dispersión entre los árboles disminuye, la curva de color fuxia representa los valores promedio de la modularidad detectados por el programa **NETCARTO** para la red cuya configuración de nodos está determinada por una distribución uniforme bivariada, mientras que la curva de color azul representa los valores promedio de la modularidad para la red cuya configuracion espacial de los nodos está determinada por la ecuación 4.1 en el eje vertical con parámetro $\lambda = 0.5$ y una distribución uniforme en el eje horizontal.



Figura 5–22: Superficie obtenida a partir de los parámetros λ y α , se puede observar como los valores promedios de la modularidad decrecen a medida que la cola del kernel de dispersión se hace más gruesa y cuando los nodos presentan un mayor agrupamiento. Es decir, cuando los valores del parámetro de escala en el kernel de dispersión aumentan y el parámetro λ disminuye su valor de uno a cero.

CAPÍTULO 6 CONCLUSIONES

- Se confirma el hecho de que cuando los polinizadores tienen condiciones favorables para llevar a cabo la actividad de polinización, los eventos de apareamiento ocurrirán en mayor número causando la aparición de nuevos enlaces y una disminución en los valores de la modularidad. Esto implica un incremento en el flujo genético dentro de la población.
- Poblaciones que presentan un grado de dispersión mayor, presentarán una estructura modular mayor, debido a la dificultad impuesta sobre los polinizadores en términos de distancia.
- Los árboles dentro de una población no establecerán vínculos de apareamiento con individuos al azar, sino que existirá una tendencia a la formación de grupos dentro de la población. Dentro de estos grupos, la ocurrencia de eventos de apareamiento entre los árboles que los conforman será mayor comparada con los eventos de apareamiento que pueden ocurrir entre dichos grupos.
- El decrecimiento en los valores promedio de la modularidad indica la posibilidad de que los individuos dentro de la población, pueden compartir su material genético con un mayor número de individuos. Esto puede influir en la diversificación del mapa genético de la población.
- Resulta importante desarrollar estudios acerca de los procesos que afectan la sobrevivencia de algunas especies de plantas, en particular el proceso de reproducción.
 Estos estudios deben contribuir en el desarrollo de estrategias de conservación enfocadas hacia la no fragmentación de las poblaciones.

APÉNDICES

CAPÍTULO 7 TABLAS DE LOS VALORES DE LA MODULARIDAD OBTENIDOS PARA LAS REDES CONSTRUIDAS EN LOS EXPERIMENTOS REALIZADOS

Las tablas que se muestran a continuación muestran los valores de la modularidad para cada una las redes construidas con los diferentes valores del parámetro α . En estas tablas se muestra también los valores promedios, la varianza y el coeficiente de variación de dichos valores de la modularidad.

α	Mod. Red 1	Mod. Red 2	Mod. Red 3	Mod. Red 4	Mod. Red 5	Mod. Promedio	Varianza	Coef. Var
75.7	0.224598	0.226968	0.228932	0.227118	0.222035	0.225930	$7.11151 \cdot 10^{-6}$	1.180339%
125.7	0.146954	0.143445	0.145568	0.145597	0.148166	0.145946	$3.11604 \cdot 10^{-6}$	1.209509%
175.7	0.109223	0.108123	0.107374	0.106896	0.104871	0.107297	$2.61099 \cdot 10^{-6}$	1.505965%
225.7	0.086283	0.085200	0.085542	0.084330	0.085798	0.085431	$5.34619 \cdot 10^{-7}$	0.855863%
275.7	0.069809	0.074194	0.072506	0.073690	0.074518	0.072943	$3.65414 \cdot 10^{-6}$	2.620649%
325.7	0.064582	0.064551	0.063855	0.062089	0.064736	0.063963	$1.21249 \cdot 10^{-6}$	1.721512%

Tabla 7–1: Modularidad de redes con nodos configurados mediante una distribución uniforme bivariada

	1	1						1
α	Mod. Red 1	Mod. Red 2	Mod. Red 3	Mod. Red 4	Mod. Red 5	Mod. Promedio	Varianza	Coef. Var
75.7	0.178514	0.176417	0.170730	0.179395	0.173312	0.175674	$1.31214 \cdot 10^{-5}$	2.061971%
125.7	0.101968	0.104727	0.103494	0.099147	0.102845	0.102436	$4.39297 \cdot 10^{-6}$	2.046098%
175.7	0.071049	0.071999	0.072372	0.071261	0.070991	0.071534	$3.80768 \cdot 10^{-7}$	0.862616%
225.7	0.060033	0.055911	0.054591	0.056761	0.057115	0.056882	$4.04731 \cdot 10^{-6}$	3.536782%
275.7	0.046789	0.049142	0.051026	0.050131	0.048992	0.049216	$2.51483 \cdot 10^{-6}$	3.222166%
325.7	0.044387	0.043612	0.044596	0.042966	0.042499	0.043612	$8.06241 \cdot 10^{-7}$	2.058858%

Tabla 7–2: Modularidad de redes con nodos configurados mediante una distribución uniforme en el eje horizontal y en el eje vertical siguen la distribución mostrada en la ecuación 4.1 con parámetro $\lambda = 1$

α	Mod. Red 1	Mod. Red 2	Mod. Red 3	Mod. Red 4	Mod. Red 5	Mod. Promedio	Varianza	Coef. Var
75.7	0.163730	0.158124	0.161772	0.159046	0.157736	0.160082	0.000007	1.652747%
125.	7 0.086684	0.083475	0.081887	0.084221	0.080898	0.083433	0.000005	2.680076%
175.	7 0.052685	0.057339	0.058193	0.055076	0.057899	0.056238	0.000005	3.976080%
225.	7 0.045287	0.044200	0.044567	0.046070	0.043814	0.044788	0.000001	2.232740%
275.	7 0.038519	0.038044	0.040350	0.038570	0.038186	0.038734	0.000001	2.581711%
325.	7 0.035204	0.034307	0.033704	0.037253	0.037153	0.035524	0.000003	4.875720%

Tabla 7–3: Modularidad de redes con nodos configurados mediante una distribución uniforme en el eje horizontal y en el eje vertical siguen la distribución mostrada en la ecuación 4.1 con parámetro $\lambda = 0.75$

α	Mod. Red 1	Mod. Red 2	Mod. Red 3	Mod. Red 4	Mod. Red 5	Mod. Promedio	Varianza	Coef. Var
75.7	0.131880	0.129494	0.128448	0.130872	0.129943	0.130127	$1.7204 \cdot 10^{-6}$	1.007969%
125.7	0.072994	0.071367	0.067304	0.069498	0.072561	0.070745	$5.53456 \cdot 10^{-6}$	3.325414%
175.7	0.048759	0.049723	0.047622	0.046924	0.049211	0.048448	$1.32734 \cdot 10^{-6}$	2.378018%
225.7	0.035539	0.039655	0.039322	0.040782	0.039316	0.038923	$3.9392 \cdot 10^{-6}$	5.099149%
275.7	0.034230	0.033819	0.034020	0.035897	0.036304	0.034854	$1.33663 \cdot 10^{-6}$	3.317057%
325.7	0.031265	0.032533	0.030900	0.032125	0.030472	0.031459	$7.3033 \cdot 10^{-7}$	2.716530%

Tabla 7–4: Modularidad de redes con nodos configurados mediante una distribución uniforme en el eje horizontal y en el eje vertical siguen la distribución mostrada en la ecuación 4.1 con parámetro $\lambda = 0.65$

α	Mod. Red 1	Mod. Red 2	Mod. Red 3	Mod. Red 4	Mod. Red 5	Mod. Promedio	Varianza	Coef. Var
75.7	0.126026	0.124866	0.124506	0.122322	0.120951	0.1237342	0.0000042	1.656284%
125.7	0.066352	0.069998	0.066503	0.068984	0.067969	0.0679612	0.0000025	2.326531%
175.7	0.043774	0.045042	0.047724	0.047400	0.044824	0.0457528	0.0000030	3.785671%
225.7	0.038245	0.037442	0.037029	0.037103	0.038164	0.0375966	0.0000003	1.456840%
275.7	0.032088	0.033798	0.032898	0.032752	0.033410	0.0329892	0.0000004	1.917159%
325.7	0.027297	0.030586	0.030475	0.028708	0.029614	0.0293360	0.0000019	4.698680%

Tabla 7–5: Modularidad de redes con nodos configurados mediante una distribución uniforme en el eje horizontal y en el eje vertical siguen la distribución mostrada en la ecuación 4.1 con parámetro $\lambda = 0.5$

REFERENCIAS BIBLIOGRÁFICAS

- Fortuna MA, García C, Guimar aes PR, and Bascompte J. Spatial mating networks in insect-pollinated plants. *Ecol Lett*, 11:490–498, 2008.
- [2] D.J. Futuyma. *Evolutionary Biology*. Sinauer Assoc, 2nd edition, 1986.
- [3] S.K Gottlieb, L.D.and Jain. *Plant Evolutionary Biology*. Chapman and Hall Ltd, 1st edition, 1988.
- [4] M. E. J. Newman, A.-L. Barabási, and D. J. Watts. The Structure and Dynamics of NETWORKS. Princeton University Press, 1st edition, 2006.
- [5] Watts DJ and Strogratz SH. Collective dynamics of 'small-world networks'. Nature, 393:440–442, 1998.
- [6] A.-L. Barabási and D. J. Watts. Emergence of Scaling in Random Networks. Science, 286:509–512, 1999.
- [7] S Redner. How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4:131–134, 1998.
- [8] M. E. J Newman. The structure and function of complex networks. SIAM Review, 45:167–256, 2003.
- [9] M. E. J Newman. Mixing patterns in networks. *Phys. Rev*, E 67:026126, 2003.
- [10] Newman M.E.J and Girvan M. Finding and evaluating community structure in networks. *Phys. Rev*, E 69:026113, 2004.
- [11] Fortuna MA, Albaladejo RG, Ferández Laura, Aparicio Aberlado, and Bascompte J. Networks of spatial genetic variation across species. Proc Natl Acad Sci USA, 106:19044–19049, 2009.
- [12] Kashtan N and Alon Uri. Spontaneous evolution of modularity and networks motifs. Proc Natl Acad Sci USA, 102:13773–13778, 2005.

- [13] Guimerà R and Nunes Amaral LA. Functional cartography of complex metabolic networks. *Nature*, 433:895–900, 2005.
- [14] M. A. Lewis and J. Bullock. Workshop: Mathematical models for plant dispersal. *Final Report*, pages 1–5, Sept 21, 2003-October 2, 2003.
- [15] H.F. Weinberger. Long-time behavior of a class of biological models. SIAM J. Appl. Math., 13:353–396, 1982.
- [16] J.S. Clark, M.A. Lewis, and L. Horvath. Invasion by extremes: Population spread with variation in dispersal and reproduction. Am. Nat, 157:537–554, 2001.
- [17] Søren Asmussen. Applied probability and queues. Springer-Verlag New York, Inc, 2nd edition, 2003.
- [18] R. Guimerà and L.A.N Amaral. Functional cartography of complex metabolic networks. *Nature*, 433:895–900, 2005.
- [19] Kirkpatrick S, Gelatt CD. Jr, and Vecchi MP. Optimization by simulated annealing. *Science*, 220:671–680, 1983.
- [20] James C. Spall. Introduction to stochastic search and optimization: estimation, simulation, and control. Wiley-Interscience series in discrete mathematics, 1st edition, 2003.
- [21] M. Rosenbluth A. Teller E. Teller N. Metropolis, A. Rosenbluth. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys, 21,6:1087– 1092, 1953.

MODULARIDAD EN REDES DE APAREAMIENTO EN PLANTAS POLINIZADAS POR INSECTOS BAJO FACTORES ABIOTICOS VARIABLES

Leoncio Rodriguez Quiñones leoncio.rodriguez@upr.edu Departamento de Ciencias Matemáticas Consejero: Luis Gordillo Grado: Maestría en Ciencias Fecha de Graduación: November 2010

Este trabajo está enfocado en el estudio de la propiedad estructural de la modularidad en redes complejas que surgen cuando los eventos de polinización son modelados usando un kernel de dispersión de cola gruesa. Los valores de la modularidad son obtenidos desde diferentes redes cuyos nodos tienen dos configuraciones espaciales diferente, estos valores de la modularidad son usados para entender algunos aspectos acerca del flujo genético en una población de árboles y como son las interacciones en términos de apareamiento entre los individuos de la población. Además esta información acerca de la modularidad de las redes de apareamiento es importante porque permite conocer acerca de algunos aspectos de caracter ecológico y biológico que ocurren dentro de la población de árboles, especialmente porque este conocimiento puede ser usado como base para el manejo de la conservación de especies que corren el riesgo de estár en peligro de extinción.