Tesis Doctoral

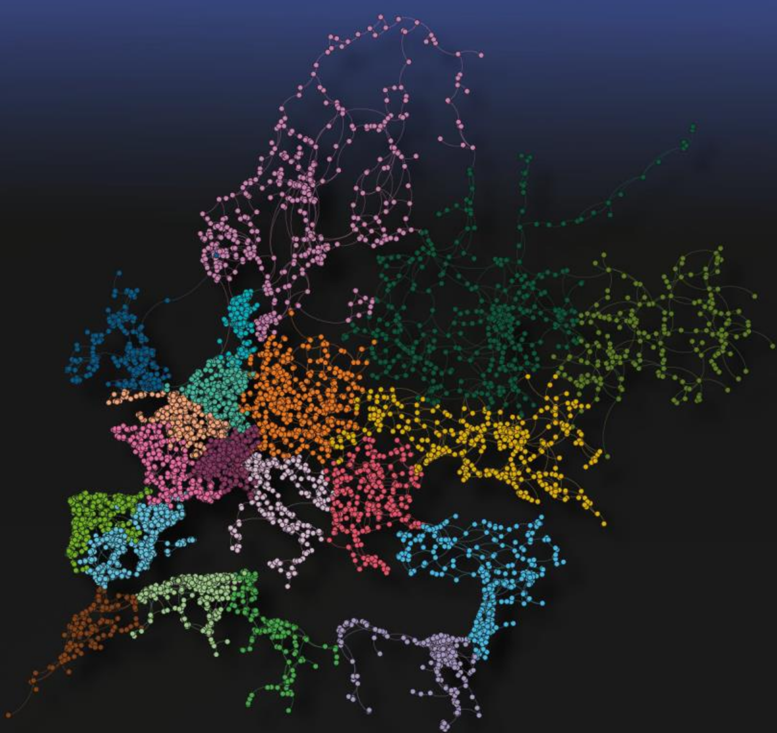
Optimización Multiobjetivo para la Detección de Comunidades en Redes Complejas

Multi-Objective Optimization for Community Detection in Complex Networks

Manuel Alejandro Guerrero López

Directores

Dra. Consolación Gil Montoya
Dr. Francisco Gil Montoya



Departamento de Informática
Universidad de Almería

Diciembre de 2020

Departamento de Informática

Escuela Superior de Ingeniería

TESIS DOCTORAL

Optimización Multiobjetivo para la Detección de Comunidades en Redes Complejas

Multi-Objective Optimization for Community Detection in Complex Networks

AUTOR

Manuel Alejandro Guerrero López

Tesis Doctoral presentada para optar al Grado de Doctor por la Universidad de Almería. Programa de Doctorado en Tecnología de Invernaderos e Ingeniería Industrial y Ambiental (RD99/11).

DIRECTORES

Dra. Consolación Gil Montoya

Dr. Francisco Gil Montoya

Universidad de Almería, diciembre 2020

A Cristina y a mis padres,

Manuel y Carmen, a quienes debo

el universo.

Agradecimientos

Durante el largo camino recorrido en esta tesis, son muchas las personas que de una manera u otra han contribuido y me han apoyado. Me gustaría trasladarles la más sincera de las gratitudes en éstas humildes palabras.

En primer lugar, quiero agradecer a mis directores Consolación G. y Francisco G. la oportunidad que me han dado de seguir formándome junto a enormes profesionales, de los que he aprendido mucho. Quiero agradecerles la predisposición y ayuda que siempre he encontrado en ellos cuando la he necesitado. Son muchos los momentos que hemos vivido a lo largo de este tiempo y por los que siempre les guardaré un cariño especial.

Tampoco olvido de los coautores de los artículos que han hecho posible esta tesis, en especial a Raúl Baños, una gran persona, cuya experiencia, orientación y colaboración han sido vitales para lograr lleva a cabo con éxito los métodos y experimentos desarrollados en las publicaciones.

Agradecer, a la Universidad de Almería, por ser la institución que ha contribuido tanto en mi formación y sigue haciéndolo.

A los amigos, a todos ellos, sin excepción. Todos han puesto su granito de arena con sentido del humor, ánimo, apoyo, y sobe todo y muy importante, compañía.

Y en especial, éste último párrafo quiero dedicarlo a ellos, los que siempre están ahí, los incondicionales, los que personifican y transmiten todos los valores que siempre he admirado, la familia, padres, hermanos y pareja. Infinitas gracias.

Gracias, a todos y cada uno de vosotros.

Índice general

Figuras

Tablas

Preámbulo

La investigación realizada en la presente tesis se ha materializado en 4 publicaciones científicas internacionales. Tres de ellas corresponden a artículos publicados en revistas incluidas en el Journal Citation Reports (JCR) de la Web of Science (WOS), las cuales tratan el enfoque evolutivo mono-objetivo. La cuarta publicación, en revisión en una revista internacional del JCR en el momento de escritura de esta tesis, presenta un nuevo algoritmo evolutivo multiobjetivo, así como nuevos objetivos a optimizar.

A continuación, se detallan cada una de las publicaciones y sus referencias bibliográficas:

PUBLICACIÓN CIENTÍFICA 1

En el Capítulo 2 se describe el artículo titulado "*Adaptive community detection in complex networks using genetic algorithms*", publicado en la revista Neurocomputing. Describe la implementación de un nuevo algoritmo genético evolutivo mono-objetivo (GGA+) que incluye eficientes métodos de inicialización y operadores avanzados de búsqueda bajo la guía del índice modularidad. Además, el algoritmo incorpora una estrategia que permite un análisis flexible y adaptativo de las características de una red desde diferentes niveles de detalle de acuerdo con las necesidades de un analista. Los resultados obtenidos en diferentes redes de prueba ampliamente utilizadas en la literatura han sido comparados con otros cinco algoritmos evolutivos.

PUBLICACIÓN CIENTÍFICA 2

En el Capítulo 3 se presenta el artículo "Community detection in national-scale high voltage transmission networks using genetic algorithms", publicado en la revista Advanced Enginering Informatics. En él, se ha realizado un estudio basado en el análisis de detección de comunidades sobre redes eléctricas de transmisión de alta tensión a escala nacional (Italia, Alemania, Francia, Península Ibérica y Texas), y del caso de prueba IEEE-118. Los resultados obtenidos de la aplicación de distintos algoritmos genéticos evolutivos demuestran el buen rendimiento ofrecido por estás técnicas en la detección de comunidades de sistemas complejos. Además, gracias a un análisis con distintos niveles de detalle, se ha conseguido revelar interesante información topológica. Finalmente, se ha discutido, desde un punto de vista técnico, la utilidad de la detección de comunidades en análisis de contingencias basados en flujos de potencia.

PUBLICACIÓN CIENTÍFICA 3

En el Capítulo 4 se presenta el artículo "*Evolutionary Algorithms for Community Detection in Continental-Scale High-Voltage Transmission Grids*", publicado en la revista *Symmetry*. Se ha realizado un estudio pionero de detección de comunidades sobre redes eléctricas de alta tensión a escala continental (red europea y red norte americana), con el fin de determinar la aportación que los métodos evolutivos pueden ofrecer en diseño y análisis de redes complejas de grandes dimensiones. Para ello, se han comparado varios algoritmos genéticos evolutivos como GGA+, y el bien conocido método de Louvain. Los resultados confirman la escalabilidad y buen rendimiento de los algoritmos evolutivos en sistemas complejos.

PUBLICACIÓN CIENTÍFICA 4

En el Capítulo 5 se describe el artículo "*Multi-objective evolutionary algorithms to find community structures in large networks*", en revisión en una revista internacional del JCR. Este trabajo describe los inconvenientes del uso de Modularidad como único objetivo a optimizar y el beneficio de considerar otros objetivos alternativos como "*conductancia*" y "*desbalanceo*" (nuevo objetivo presentado en esta publicación). Además, presenta un nuevo algoritmo multiobjetivo (MOGGA+) basado en la dominancia de Pareto, el cual incluye novedosos y eficientes métodos de inicialización y operadores genéticos. Los resultados obtenidos en redes eléctricas de escala nacional han permitido obtener Frentes de Pareto de alta calidad en comparación con otras aproximaciones multiobjetivo, y, además, abrir una nueva vía de investigación en detección de comunidades mediante la optimización simultanea de diferentes objetivos en redes de gran tamaño.

REFERENCIAS DE LAS PUBLICACIONES

PUBLICACIÓN CIENTÍFICA 1:

Título: "Adaptive community detection in complex networks using genetic algorithms."
Autores: Manuel A. Guerrero, Francisco G. Montoya, Raúl Baños, Alfredo Alcayde, Consolación Gil
Revista científica: Neurocomputing.
Volumen: 266.
Páginas: 101-113.
Año: 2017.
ISSN: 0925-2312.
Doi: 10.1016/j.neucom.2017.05.029

Datos JCR (Journal Citation Reports):

Factor de impacto (2017): 3,241.
Categoría: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE.
Ranking categoría: 27/132.
Cuartil: Q1.
Editor: ELSEVIER SCIENCE BV.
País: Países Bajos.

PUBLICACIÓN CIENTÍFICA 2:

Título: "Community detection in national-scale high voltage transmission networks using genetic algorithms."
Autores: Manuel A. Guerrero, Francisco G. Montoya, Raúl Baños, Alfredo Alcayde, Consolación Gil
Revista científica: Advanced Engineering Informatics.
Volumen: 38.
Páginas: 232-241.
Año: 2018.
ISSN: 1474-0346.
Doi: 10.1016/j.aei.2018.07.001

Datos JCR (Journal Citation Reports):

Factor de impacto (2017): 3,772.
Categoría: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE.
Ranking categoría: 32/134.
Cuartil: Q1.
Editor: ELSEVIER SCI LTD.
País: Inglaterra.

PUBLICACIÓN CIENTÍFICA 3:

Título: "Evolutionary Algorithms for Community Detection in Continental-Scale High-Voltage Transmission Grids."
Autores: Manuel A. Guerrero, Raúl Baños, Consolación Gil, Francisco G. Montoya, Alfredo Alcayde.
Revista científica: Symmetry-Basel.
Volumen: 11
Páginas: 1472.
Año: 2019.
ISSN: 2073-8994.
Doi: 10.3390/sym11121472

Datos JCR (Journal Citation Reports):

Factor de impacto (2018): 2,143.
Categoría: MULTIDISCIPLINARY SCIENCES.
Ranking categoría: 30/69.
Cuartil: Q2.
Editor: MDPI AG.
País: Suiza.

PUBLICACIÓN CIENTÍFICA 4:

Título: "Multi-objective evolutionary algorithms to find community structures in large network."
Autores: Manuel A. Guerrero, Raúl Baños, Consolación Gil, Francisco G. Montoya, Alfredo Alcayde.
Revista: Impacto Internacional

Resumen

A raíz de las bases de la teoría de grafos que surgieron en el siglo XVIII y sus desarrollos posteriores, es posible estudiar las relaciones o vínculos existentes en un conjunto de elementos y determinar las características y propiedades de dichas relaciones. En las últimas décadas, los importantes y continuos avances en el ámbito de las ciencias de la computación han permitido modelar y analizar grafos de cualquier escala y tipología correspondientes a sistemas reales de alta complejidad que proceden de multitud de áreas de conocimiento como ingeniería, física, ciencias sociales, etc. En la actualidad, gracias al modelado y análisis de grafos por computador, es posible obtener información valiosa sobre las relaciones funcionales o las características de un determinado sistema, por complejo que éste sea. Ello ha dado lugar a que investigadores de todo el mundo muestren un gran interés por resolver una amplia variedad de problemas relacionados con grafos, como pueden ser: partición de grafos, coloreado de grafos, problemas de enrutado, problemas de flujos en redes, etc. Todos estos problemas son altamente complejos, hasta el punto de estar incluidos en la categoría de problemas NP-completos, lo que implica que instancias lo suficientemente grandes de dichos problemas no pueden ser resueltos en un tiempo limitado.

En los últimos años, un problema relativamente nuevo que puede ser abordado mediante las técnicas de análisis basadas en grafos, ha despertado un gran interés dentro de la comunidad científica debido al potencial de analizar sistemas complejos que éste ofrece. Este problema, llamado detección de comunidades, nace a raíz de una característica común inherente a todos los sistemas complejos, la presencia de patrones de nodos más densamente conectados entre sí que con el resto de nodos de la red. De los nodos que muestran estos patrones de conexión, llamados comunidades, se espera que compartan ciertas propiedades que permitirán detectar nuevas características o relaciones funcionales de la red. La búsqueda de estos patrones o estructuras de comunidades es conocida como problema de la detección de comunidades, el cual ha sido clasificado en la literatura científica como problema NP-completo. Cómo encontrar la estructura de comunidades óptima que mejor representa las características de una red, se ha convertido en todo un hándicap, pues se han propuesto multitud de algoritmos y funciones objetivo para resolver el problema. Entre ellos, los algoritmos evolutivos y el índice de Modularidad han destacado como las principales soluciones aceptadas por la comunidad científica.

Además de la importancia de la búsqueda de un algoritmo capaz de resolver el problema de la detección de comunidades eficientemente, el análisis de redes complejas a menudo se encuentra limitado a estudiar las características de una red desde una única perspectiva debido al enfoque propuesto por la mayoría de los algoritmos que encontramos en la literatura científica. Con el fin de lograr un análisis más completo y detallado de las redes objeto de estudio, esta tesis doctoral propone un enfoque novedoso, flexible y adaptativo para el análisis de comunidades de cualquier complejidad y tipología de red. En particular, se analizan redes de diferente topología y extensión comúnmente utilizadas en la literatura científica, y, además, se lleva a cabo un estudio pionero en el ámbito de ingeniería eléctrica sobre redes de transmisión de alta tensión. La importancia de este estudio radica en que estamos tratando con una de las mayores infraestructuras creadas por el hombre, con cientos de miles de kilómetros de líneas de alta tensión que interconectan no sólo regiones y países, sino también áreas continentales. Además, debido a la creciente demanda de electricidad y a la aparición de nuevos formatos de generación eléctrica, resulta imprescindible abordar el estudio sobre la ampliación y modificación de redes eléctricas mediante nuevos procedimientos complementarios a las técnicas tradicionales utilizadas en el estudio de sistemas eléctricos de potencia. La incorporación de estos nuevos formatos de generación de energía eléctrica renovables, está provocando que el diseño clásico de las redes eléctricas pase de un modelo centralizado, basado en disponer de un número relativamente acotado de centrales térmicas, nucleares e hidroeléctricas, a un modelo distribuido con un peso creciente en energías renovables, originando así un aumento de la conectividad en las redes y, por consiguiente, un incremento de su complejidad. En este sentido, se hace indispensable la necesidad de aplicar estrategias de control más robustas, así como nuevas técnicas de optimización para gestionar sistemas a gran escala.

Actualmente, aunque existen procedimientos de análisis y diseño de redes eléctricas, como es el caso de la resolución de flujos de potencia que permite analizar múltiples configuraciones del sistema eléctrico, el inmenso abanico de posibles configuraciones existentes provoca que sea necesario aplicar nuevos procedimientos complementarios de cara a hacer factible la búsqueda de soluciones. En este sentido, en la presente tesis doctoral se aplica la detección de comunidades en el estudio y análisis de redes eléctricas de gran tamaño a fin de ampliar las conclusiones obtenidas por los procedimientos tradicionales, además de analizar los resultados como posibles mejoras de control de la red, que supondrían de una mayor fiabilidad y capacidad de restauración ante graves perturbaciones como desastres naturales (tormentas, terremotos, etc.) mediante el desarrollo de nuevos planes de conmutación para proteger islas o desconexiones parciales de la red, evitando así una mayor degradación durante los incidentes.

En resumen, el objetivo de esta tesis se centra en afrontar el problema de la detección de comunidades en ámbitos que precisan de la aparición de nuevas técnicas de diseño y control, a través del desarrollo de nuevos algoritmos evolutivos (mono-objetivo y multiobjetivo) que permitan analizar las características topológicas de una red desde distintas perspectivas, con mayor o menor nivel de detalle en función de las necesidades del analista. Para lograr este objetivo, en primer lugar, se han diseñado métodos de inicialización poblacional eficientes y operadores evolutivos avanzados basados en intercambios entre comunidades, que han sido incorporados al algoritmo genético mono-objetivo propuesto en esta tesis, el cual optimiza el popular Índice de Modularidad como función objetivo. Dicho algoritmo, denominado "Generational Genetic Algorithm" (GGA+), ha sido evaluado mediante un análisis de rendimiento sobre multitud de benchmarks populares basados en redes sociales, además de sobre grafos de gran escala con miles de nodos y aristas que representan redes reales de transmisión de alta tensión de escala nacional y continental. Tras el éxito de los resultados mostrados por la implementación del algoritmo mono-objetivo, con la finalidad de evaluar el rendimiento que los métodos de detección de comunidades evolutivos pueden ofrecer en el ámbito de ingeniería, distintos algoritmos como el mencionado GGA+ y otros algoritmos evolutivos referenciados en la literatura, como

"Modularity and Improved Genetic Algorithm" (MIGA) o el bien conocido método de "Louvain", se han aplicado sobre múltiples redes eléctricas, demostrando que los resultados obtenidos avalan su utilidad y buen rendimiento.

Actualmente, la mayoría de métodos destinados a resolver el problema de la detección de comunidades, utilizan un enfoque mono-objetivo, siendo el índice Modularidad la función objetivo más extendida. Sin embargo, debido a la propia definición de comunidad en la que ésta puede ser vista desde un enfoque multiobjetivo, donde un objetivo puede ser maximizar el número de conexiones dentro de una comunidad y otro objetivo minimizar el número de conexiones con nodos externos a la comunidad, y, debido a que algunos estudios recientes han demostrado que al considerar Modularidad como único objetivo pueden surgir inconvenientes relacionados con el límite de resolución y desbalanceo de las soluciones, parece adecuado plantear el problema de la detección de comunidad desde un punto de vista multiobjetivo.

Por estos motivos, en segundo lugar, basándonos en el buen rendimiento demostrado por GGA+, se ha diseñado una nueva versión multiobjetivo llamada MOGGA+, la cual incluye nuevos objetivos a optimizar en la detección de comunidades, abriendo de esta manera una nueva vía de investigación mediante la aplicación de algoritmos evolutivos multiobjetivo que optimizan simultáneamente diferentes objetivos. Más concretamente, MOGGA+ ha sido diseñado con nuevos métodos de inicialización avanzados y operadores evolutivos. Además, utiliza un conjunto de soluciones no dominadas basado en la dominancia de Pareto, y una estructura adicional para modificar dinámicamente la probabilidad de aplicar distintos operadores evolutivos en tiempo de ejecución. En cuanto a los diferentes objetivos optimizados, se ha analizado la combinación del índice de *Modularidad* con un novedoso objetivo propuesto en esta tesis, “*Desbalanceo”* de comunidades, propiedad a tener en cuenta a la hora de diseñar redes eléctricas resistentes a contingencias, donde es importante que la red pueda ser separada en islas (subredes) de escala aproximadamente similar, para evitar una mayor degradación de la red. También se ha analizado el objetivo *Conductancia* (como alternativa a M*odularidad)* junto con *Desbalanceo*. Conductancia es una medida de la fracción del volumen total de aristas en un subgrafo que están conectados a vértices o nodos de otros subgrafos de una red. Por último, cabe señalar que MOGGA+ incorpora técnicas de paralelismo para mejorar su rendimiento.

Abstract

Following the foundations of graph theory that emerged in the eighteenth century, as well as its subsequent developments, it is possible to study the relationships or links existing in a set of elements and identify the characteristics and properties of those relationships. In the recent decades, the important and continuous advances in the field of computer science have made it possible to model and analyze graphs of any scale and typology corresponding to highly complex real systems that come from many areas of knowledge, such as engineering, physics and social sciences. At present, thanks to computer graphics modeling and analysis, it is possible to obtain valuable information on the functional relationships or characteristics of a specific system, however complex it may be. This has led researchers around the world to show great interest in solving a wide variety of graph-related problems, among which can be found: graph partitioning, graph coloring, routing problems and network flow problems. All these problems are highly complex, to the extent that they are included in the category of NP-complete problems, which implies that sufficiently large instances of such problems cannot be solved in a limited time.

In recent years, a relatively new problem that can be addressed through graphical analysis techniques has sparked great interest within the scientific community due to the potential to analyze the complex systems it offers. This problem, called community detection, arises from a common characteristic inherent in all complex systems, the presence of node patterns more densely connected to each other than to the rest of the nodes in the network. The nodes that show these connection patterns, called communities, are expected to share certain properties that will allow them to detect new features or functional relationships in the network. The search for these community structures is known as the community detection problem, which has also been classified in the scientific literature as an NP-complete problem. How to find the optimal community structure that best represents the characteristics of a network has become a real handicap, since many objective algorithms and functions have been proposed to solve the problem. Among them, evolutionary algorithms and the Modularity Index have stood out as the main solutions accepted by the scientific community.

In addition to the importance of searching for an algorithm capable of solving the community detection problem efficiently, complex network analysis is often limited to studying the characteristics of a network from a single perspective due to the approach proposed by most of the algorithms found in the scientific literature. In order to achieve a more complete and detailed analysis of the networks under study, this doctoral thesis proposes a novel, flexible and adaptive approach for the analysis of communities of any complexity and network typology. More specifically, networks of different topology and extension commonly used in the scientific literature are analyzed, and, in addition, a pioneering study in the field of electrical engineering on high-voltage transmission networks is carried out. The importance of this study lies in the fact that we are dealing with one of the largest infrastructures created by humanity, with hundreds of thousands of kilometers of high-voltage lines that interconnect regions, countries and continental areas. What is more, due to the growing demand for electricity and the appearance of new formats of electricity generation, it is essential to undertake the study on the expansion and modification of electrical networks through new complementary procedures to the traditional techniques used in the study of electrical power systems. The incorporation of these new renewable electricity generation formats is causing the classic design of electrical networks to change from a centralized model, based on having a relatively limited number of thermal, nuclear and hydroelectric plants, to a distributed model with an increasing weight in renewable energy, thus causing an increase in network connectivity and, consequently, an increase in its complexity. In this sense, the need to apply more robust control strategies, as well as new optimization techniques to manage systems on a large scale, is essential.

Currently, although there are analysis and design procedures for electrical networks, such as the resolution of power flows that allow multiple configurations of the electrical system to be analyzed, the immense range of possible existing configurations means that it is necessary to apply new complementary procedures in order to make the search for solutions feasible. In this sense, in this doctoral thesis, community detection is applied in the study and analysis of large electrical networks in order to expand the conclusions obtained by traditional procedures and analyze the results as possible improvements in network control, thus enhancing its reliability and allowing a faster and more efficient restoration in the face of serious disturbances such as natural disasters (storms, earthquakes, etc.), all through the development of new switching plans to protect islands or partial disconnections from the network, thus avoiding further degradation during incidents.

In summary, the objective of this thesis is based on facing the problem of detecting communities in areas that require the appearance of new design and control techniques, through the development of new evolutionary algorithms that allow analyzing the topological characteristics of a network from different perspectives, with a greater or lesser level of detail depending on the analyst's needs. To achieve this objective, firstly, efficient population initialization methods and advanced genetic operators based on exchanges between communities have been designed, which have been incorporated into the single-objective genetic algorithm proposed in this thesis, which optimizes the popular Modularity Index as an objective function. This algorithm, called "*Generational Genetic Algorithm*" (GGA+), has been evaluated through a performance analysis on a multitude of popular benchmarks based on social networks, as well as on large-scale graphs with thousands of nodes and edges that represent real high-voltage transmission networks on a national and continental scale. After the success of the results shown by the implementation of the single-objective algorithm, in order to evaluate the performance that evolutionary community detection methods can offer in the engineering field, different algorithms such as the aforementioned GGA+, along with other evolutionary algorithms referenced in the literature, such as "*Modularity and Improved Genetic Algorithm*" (MIGA) or the well-known "*Louvain*" method, have been applied to multiple electrical networks, showing that the results obtained support their usefulness and good performance.

Currently, most of the methods aimed at solving the problem of community detection use a mono-objective approach, with the Modularity

index being the most widely chosen objective function to optimize. However, due to the definition of community in which it can be viewed from a multi-objective approach, where one objective may be to maximize the number of connections within a community and another objective to minimize the number of connections with nodes external to the community, and since some recent studies have shown that when considering Modularity as the only objective, problems related to the resolution limit and imbalance of the solutions may arise, it seems appropriate to approach the problem of community detection from a multi-objective point of view.

For these reasons, secondly, based on the good performance demonstrated by GGA+, the decision was made to design the multi-objective version called MOGGA+, which includes new objectives to be optimized in the detection of communities, thus opening a new avenue of research by applying multi-objective evolutionary algorithms that simultaneously optimize different objectives. More specifically, MOGGA+ has been designed with new advanced initialization methods and evolutionary operators. In addition, it uses a set of non-dominated solutions based on Pareto dominance, and an additional structure to dynamically modify the probability of applying different evolutionary operators at run time. Regarding the different optimized objectives, the combination of the *Modularity* index has been analyzed with a novel objective proposed in this thesis, *Imbalance*, a property to take into account when designing electrical networks resistant to contingencies, in which it is important that the network can be separated into islands (subnets) of approximately similar scale, to avoid further degradation of the network. The objective *Conductance* (as an alternative to Modularity), along with *Imbalance,* has also been analyzed. Conductance is a measure of the fraction of the total volume of edges in a subgraph that are connected to vertices or

nodes of other subgraphs in a network. Finally, it should be noted that MOGGA+ incorporates parallelism techniques to improve its performance.

Capítulo 1
Introducción

1.1. Antecedentes

La teoría de grafos, propuesta por Leonardo Euler en el año 1736, no solo permitió afrontar el muy conocido problema de los puentes de Königsberg, sino que desencadenó un estudio persistente de las propiedades de los grafos [1] gracias a la capacidad que ofrecían para modelar sistemas reales en multitud de áreas como la sociología [2], biología [3], física [4], informática [5] [6], etc. A raíz de la aparición de los ordenadores fue cuando surgió la posibilidad de representar, manejar y procesar grandes cantidades de datos de manera eficiente, generando un gran interés en la comunidad científica por analizar diversidad de sistemas complejos como las redes de colaboración, Internet, Word Wide Web, redes biológicas, redes eléctricas [7], redes de comunicación y transporte [5], redes sociales (Facebook, Twitter, Skype, Instagram, WhatsApp, Flickr) [8] o redes e-commerce [9] con el objetivo de descubrir nuevas relaciones funcionales o características de los sistemas. Una de las características más importantes compartida por todos los sistemas complejos es la estructura de comunidades [10], la cual se basa en el principio de que dos nodos tienen una mayor probabilidad de estar interconectados si ambos son miembros de la misma comunidad, y menor probabilidad si no comparten comunidades. Estas estructuras (comunidades) están formadas por grupos de nodos que se encuentran más densamente conectados entre sí que con el resto de nodos de la red. Los nodos pertenecientes a la misma comunidad tienen más probabilidad de compartir propiedades en común, por lo que pueden ayudar a revelar nuevas características o relaciones funcionales de la red. Cómo encontrar la estructura de comunidades óptima que mejor representa las características de una red, define al problema de detección de comunidades, uno de los problemas más importantes y complejos en el campo del análisis de redes complejas.

En consecuencia, la búsqueda de la estructura de comunidad óptima, ha generado la propuesta de multitud de métodos de optimización basados fundamentalmente en dos tendencias:

1. Detección de comunidades no solapadas. Los nodos de una red pertenecen a una única comunidad. Dentro de este enfoque encontramos la mayoría de los algoritmos propuestos, como son:

    - Algoritmos basados en métodos tradicionales [11]: particionamiento en clúster, clustering jerárquico, clustering espectral o particionamiento de grafos.

    - Algoritmos basados en optimización de funciones objetivo: optimización extrema [12], enfriamiento simulado [13, 14], optimización espectral [15], métodos divisivos [16], particionamiento heurístico [17], optimización voraz [18] o algoritmos basados en técnicas inspiradas en la biología como Algoritmos Evolutivos (EA) [19], Algoritmo de murciélago (BA) [20]) o Algoritmo de luciérnaga (FA) [21].

2. Detección de comunidades solapadas [22]. Los nodos de una red pueden pertenecer a varias comunidades por la similitud entre ellos (por ejemplo, redes de autores escribiendo sobre distintos topics). En este tipo de algoritmos encontramos los basados en: optimización y expansión local, algoritmos de factorización de matriz no negativa, método de Clique Percolation (CPM) [23], etc.

A pesar de la cantidad de métodos y algoritmos propuestos para resolver el problema de la detección de comunidades (como podemos encontrar en [24] o [25]), debido a la naturaleza compleja del problema, los métodos que más popularidad han alcanzado han sido aquellos basados en heurísticas o metaheurísticas, capaces de hallar la solución óptima o cercana a ella, en un período de tiempo limitado. Las heurísticas son métodos o algoritmos exploratorios que no aseguran encontrar las soluciones óptimas, sino soluciones aproximadas. La mayoría de algoritmos heurísticos son diseñados específicamente para resolver problemas particulares, aunque en algunos casos dichas técnicas pueden ser generalizadas para resolver cualquier problema de optimización, en cuyo caso reciben el nombre de metaheurísticas [26].

Existe un gran número de técnicas metaheurísticas, desde técnicas de búsqueda local no poblacionales como el Ascenso de Colinas [27], la Búsqueda Tabú [28], o Enfriamiento Simulado [13, 14], a técnicas poblacionales como la optimización mediante colonias de hormigas [29], optimización de enjambre de partículas (PSO) [30], búsqueda dispersa [31], algoritmos evolutivos [32], algoritmos meméticos [33], etc. Siendo, sin lugar a dudas, los Algoritmos

Evolutivos (AE) [19] la metaheurística poblacional más utilizada y referenciada en la literatura. En particular, un algoritmo evolutivo consiste en una población de individuos que evolucionan continua y selectivamente hasta que se cumpla un criterio de terminación (por ejemplo, el número de generaciones).

Entre las técnicas evolutivas más populares, los algoritmos genéticos (AG) [16] son probablemente los más utilizados. Un algoritmo genético imita la selección natural al desarrollar una población de soluciones individuales a lo largo del tiempo hasta que se cumpla una condición de terminación y se tome al mejor individuo como una solución aceptable. Las características más importantes de los AG son la representación de los individuos utilizada (por ejemplo, binaria o real), los operadores genéticos empleados (por ejemplo, mutación y cruce) para evolucionar a los individuos y la función objetivo que cuantifica la optimalidad del individuo.

Cuando se hace uso de técnicas de optimización evolutiva [34, 35], como los AG, toma gran importancia la función objetivo, cuya labor es determinar la optimalidad de una solución y guiar al algoritmo durante el proceso de búsqueda hacia la solución óptima. En este sentido, es importante definir los objetivos que se deben optimizar, ya que reflejarán los aspectos fundamentales de una buena estructura de comunidades. Estos objetivos han sido extensamente estudiados por la comunidad científica, dando lugar a un grupo diverso de funciones objetivo, entre las que encontramos: Conductancia (Conductance) [36], Expansión [37], Densidad Interna (Internal Density) [38], Radio de Corte (Cut Ratio) o Máximo-ODF (Maximum-ODF) [39], Modularidad (Modularity) [40], Puntuación de Comunidad (Community Score) [41], Aptitud de Comunidad (Community Fitness) [42] o Sorpresa (Surprise) [43].

Dentro del amplio espectro de funciones objetivo propuestas, una de las que sin duda más relevancia ha alcanzado por su simplicidad computacional y buenos resultados, ha sido Modularidad (Modularity). Esta popular función de calidad propuesta por Newman [40], considera que cuanto mayor sea el valor de Modularidad, más precisa será la estructura de comunidades detectada. La aparición de esta función objetivo ha dado lugar a mejoras constantes [44] de uno de los algoritmos más conocidos hasta la fecha, Girvan-Newman (GN) [16], algoritmo basado en un método divisivo. Como consecuencia, la búsqueda del valor óptimo de Modularidad, convierte a la detección de comunidades en un problema de optimización NP-completo ideal para el empleo de métodos heurísticos y metaheurísticos.

La popularización de las técnicas metaheurísticas evolutivas para la optimización de Modularidad, ha permitido a los AG convertirse en la técnica más utilizada en el ámbito de la detección de comunidades. Sin embargo, estas metaheurísticas de propósito general necesitan ser adaptadas para lograr un algoritmo más preciso y convergente. Diferentes autores han optado por adaptar estos algoritmos bajo un enfoque mono-objetivo, simplificando así el diseño del algoritmo y mejorando su rendimiento como es el caso de los algoritmos de Girvan-Newman (GN) [16] y GA-Net [41]. Otros métodos, como MIGA [45], optan por complementar el AG mono-objetivo con técnicas de búsqueda local como el enfriamiento simulado [13, 14] o Meme-Net con ascenso de colinas [46], buscando una mejor y rápida convergencia del algoritmo, a la vez que evitan posibles óptimos locales. También, basados en el algoritmo de Girvan-Newman, se han diseñado algoritmos mediante técnicas de optimización voraz, como por ejemplo el algoritmo de Louvain [47], capaz de analizar redes de gran tamaño. Otros autores, debido a la propia definición de comunidad en la que esta puede ser vista como un conjunto de nodos fuertemente conectados entre sí y débilmente conectados con el resto de nodos de la red, optan por un enfoque multiobjetivo en el que un objetivo puede ser maximizar el número de conexiones dentro de una comunidad y otro objetivo minimizar el número de conexiones con nodos externos a la comunidad. Es el caso de algoritmos como MOGA-Net [48] basado en NSGA-II [49] o MOCD [50] basado en PESA-II [51]. Este enfoque multiobjetivo, provoca la necesidad de optimizar varios objetivos de forma simultánea, generalmente contrapuestos, por lo que además de ser necesarias más funciones objetivo para evaluar dichos objetivos, también es necesario un proceso de selección más complejo que establezca un criterio de preferencia de una solución frente a las demás. Es en este punto, donde aparece el concepto de dominancia de Pareto y su utilidad como proceso de selección multiobjetivo [52, 53]. Más específicamente, la dominancia de Pareto se define como:

Dados dos individuos i_1 e i_2, el criterio de Pareto-dominancia determina que i_1 *domina* a i_2 si i_1 es mejor que i_2 en al menos un objetivo y no es peor que i_2 en el resto, que i_1 *es dominado* por i_2 si i_1 es peor que i_2 en al menos un objetivo y no es mejor que i_2 en ningún otro, y que i_1 e i_2 son *indiferentes* si i_1 es mejor que i_2 en algún objetivo pero es peor que i_2 en otro u otros.

Esta definición, da lugar a un conjunto de soluciones llamado frente de Pareto, formados por aquellas que no son dominadas por ninguna otra, como podemos ver en la figura 1.1.

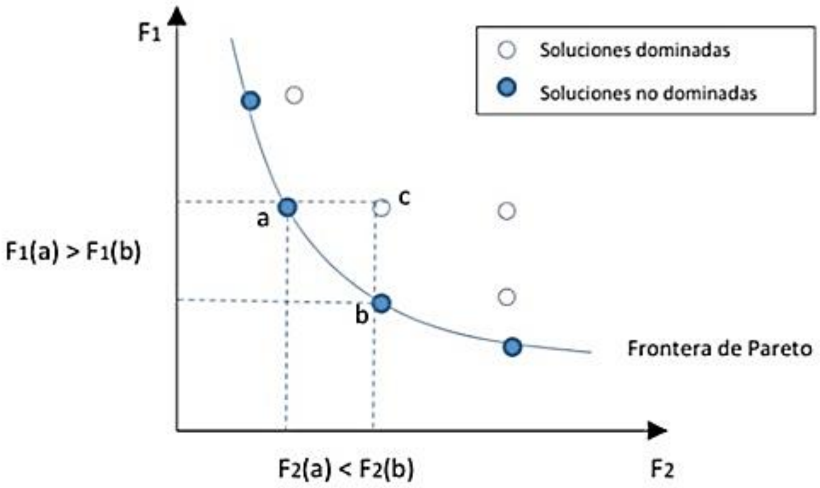


Figure 1.1 Frente de Pareto.

1.2. Motivación y Justificación

La principal justificación que ha motivado el desarrollo de esta tesis ha sido la necesidad de comprender o analizar sistemas reales cada vez más complejos que resulta imposible por el ser humano sin ayuda externa. Un ejemplo de ello, es el aumento de la complejidad de las redes eléctricas de alta tensión, pues la aparición de nuevas fuentes de energías renovables y aumento de la demanda, han dado lugar a una mayor conectividad de la red, provocando la necesidad de nuevos métodos de diseño y control más eficientes. El análisis de sistemas complejos puede lograrse mediante la búsqueda de la estructura de comunidad que mejor representa las características de una red, también conocido como detección de comunidades.

Sin embargo, a la hora de aplicar la detección de comunidades sobre sistemas complejos, se han detectado ciertas limitaciones que han motivado a la investigación realizada en la presente tesis. Estas limitaciones son:

- Necesidad de optimización algorítmica que produzca soluciones estables y precisas en entornos más complejos. Se han propuesto multitud de algoritmos heurísticos basados en el índice de Modularidad, pero debido a la complejidad del problema (NP-completo), los resultados obtenidos varían considerablemente de una ejecución a otra.

- Necesidad de metaheurísticas comunes que puedan ser adaptadas a diversidad de áreas. Existen multitud de algoritmos diseñados específicamente ad-hoc para adaptarse al problema en cuestión.

- Limitación en el análisis de detección de comunidades sobre redes complejas. A menudo, el análisis de redes se encuentra limitado tanto por el tamaño y complejidad de las redes, como por el estudio de las características de una red desde una única perspectiva.

Resolver estas limitaciones permitirá optimizar el análisis de sistemas complejos mediante la detección de comunidad, pudiendo aplicar esta técnica en distintos ámbitos, como por ejemplo al apoyo de procedimientos de estudio de sistemas eléctricos con distintos fines como el diseño de redes eléctricas (cada vez más interconectadas por el fenómeno de la globalización), o como a la prevención de contingencias provocadas por catástrofes naturales, cortes de línea, ataques terroristas, etc.

1.3. Hipótesis

La hipótesis que se plantea en esta tesis, teniendo en cuenta las limitaciones indicadas en el punto anterior, es diseñar nuevos métodos de optimización evolutiva que sean estables y eficientes, y que aporten una nueva perspectiva flexible desde la que analizar las características de una red, sean cual sean las propiedades de la misma. Una vez logrado este objetivo, los métodos desarrolladas serán puestos en práctica sobre escenarios reales de grandes dimensiones donde la detección de comunidades aún no ha sido aplicada y podría revelar valiosa información.

Para lograr este cometido, primero será necesario realizar una investigación del estado del arte de los algoritmos y técnicas existentes en el ámbito de la detección de comunidades. Una vez realizada, se diseñarán e implementarán diversos métodos que incluirán el enfoque de análisis en diferentes perspectivas, es decir, analizar la red desde distintos niveles de detalle. Finalmente, se comprobará y evaluará el rendimiento de los métodos propuestos para verificar su validez antes de ser aplicados a problemas reales relacionados con el diseño y control de sistemas de redes eléctricas de alta tensión de escala nacional y continental.

1.4. Objetivos

El objetivo general que se persigue en esta tesis doctoral es analizar, diseñar, implementar y validar nuevos algoritmos de optimización computacional que permitan detectar comunidades en redes complejas de manera flexible y adaptable a cualquier área de conocimiento. Las comunidades obtenidas por dichos métodos facilitarán que los analistas de este tipo de redes lleven a cabo estudios de las mismas desde diferentes niveles de detalle, o como se ha denominado, desde diferentes "*grados de abstracción*". Es decir, mientras que la mayoría de las técnicas y procedimientos actuales obtienen una única solución procedente de la optimización de un único objetivo, se pretende que los procedimientos diseñados obtengan un conjunto de soluciones optimizando simultáneamente distintos objetivos que permitan al analista estudiar de forma más completa las características de la red.

Como punto de partida, esta tesis tomo como referencia el modelo de los algoritmo genético mono-objetivo, a partir del cual es necesario diseñar una versión estable, eficiente y convergente, capaz de generar una solución que permita estudiar las características de una red desde distintos grados de abstracción. Para ello, se ha tomado la decisión de diseñar tanto un proceso de inicialización eficiente, como nuevos operadores genéticos avanzados basados en el intercambio de nodos entre comunidades. Asimismo, el algoritmo utiliza información a priori sobre el número de comunidades a detectar, permitiendo así acotar el espacio de búsqueda del algoritmo y, por tanto, mejorar su convergencia obteniendo resultados más precisos y estables. En un gran número de situaciones reales, el número de estructuras de comunidades que forman una red es conocido de manera aproximada, lo que ha permitido acotar el espacio de búsqueda del problema. En los casos en que inicialmente el número de estructuras de comunidades es desconocido, es habitual utilizar un algoritmo genético básico que proporcione una estimación inicial del número de estructuras de comunidades.

Una vez diseñado el algoritmo mono-objetivo (GGA+), se ha llevado a cabo su implementación mediante técnicas de programación adecuadas que, además, incorporan la capacidad de exportar la solución obtenida (integrada por los distintos grados de abstracción) para poder ser visualizada por softwares externos especializados en grafos. Posteriormente, mediante benchmarks de diferentes características y tamaños, se han validado los resultados obtenidos, permitiendo así, iniciar la siguiente etapa, el diseño del algoritmo multiobjetivo (MOGGA+).

Como objetivo final, una vez comprobada la validez de los algoritmos implementados, se han modelado redes reales eléctricas complejas de alta tensión, sobre las cuales, tras la aplicación de los métodos propuestos, se han expuesto las conclusiones obtenidas. Además, se ha discutido si realmente el enfoque flexible de detección de comunidades propuesto aporta beneficios, especialmente en el ámbito de las redes eléctricas.

A continuación, se desglosan los objetivos alcanzados más específicamente:

- Diseñar e implementar un algoritmo genético mono-objetivo (GGA+). En este objetivo no sólo se incluye el diseño del algoritmo, sino también el diseño de métodos de inicialización poblacional y operadores genéticos eficientes, así como estructuras adicionales para obtener el máximo rendimiento de los operadores propuestos.

- Depurar y evaluar GGA+ mediante estudios empíricos basados en benchmarks utilizados popularmente en la detección de comunidades. Se incluye, la corrección de los posibles errores de implementación, así como diferentes optimizaciones que han sido introducidas en el mismo. Esta labor se ha realizado mediante una comparativa entre los procedimientos propuestos y otras técnicas evolutivas bien conocidas.

- Estudiar, identificar e implementar los componentes esenciales y estrategias seguidas por los algoritmos genéticos multiobjetivo, para diseñar la versión multiobjetivo (MOGGA+) utilizando como base el algoritmo GGA+. El enfoque multiobjetivo permite evaluar otras características interesantes de las estructuras de comunidades, como por ejemplo el balanceo de individuos entre comunidades, el cual, desde un punto de vista técnico, aporta información valiosa sobre el diseño de redes eléctricas, ya que una condición necesaria para la estabilidad de la red es que los flujos de potencia estén equilibrados y no sobrecarguen zonas específicas. MOGGA+, incluye un proceso de inicialización eficiente, operadores genéticos avanzados, la técnica de selección basada en la dominancia de Pareto, nuevas funciones objetivo y técnicas de paralelismo para mejorar su rendimiento.

- Depurar y evaluar el rendimiento de MOGGA+ frente a otras metaheurísticas multiobjetivo conocidas, con la ayuda de distintos benchmarks y redes reales.

- Aplicar los distintos métodos desarrollados sobre redes eléctricas complejas, cuyo análisis de resultados, determina la aportación global de la tesis doctoral.

1.5. Metodología

El proceso de desarrollo de la tesis se divide principalmente en cuatro etapas claramente diferenciadas entre sí. Se ha comenzado con un estudio inicial para adquirir los conocimientos previos necesarios al desarrollo de algoritmos evolutivos, concretamente algoritmos genéticos mono-objetivo y multiobjetivo. La siguiente etapa se ha centrado en el diseño e implementación de varios algoritmos mono-objetivo y multiobjetivo que incorporan el concepto “grado de abstracción”. En la tercera etapa, los algoritmos propuestos se han testeado frente a benchmarks de gran tamaño y, comparado con otros métodos conocidos en la literatura para evaluar su rendimiento. Finalmente, en la cuarta etapa, se ha modelado un conjunto de redes eléctricas reales, con el objetivo de demostrar el rendimiento, versatilidad y adaptabilidad de los métodos de detección de comunidades propuestos.

Estas cuatro etapas se pueden desglosar en múltiples tareas cuya función es cumplir la hipótesis y los objetivos establecidos:

- La primera etapa ha incluido tareas como la formación e investigación previa para conocer tanto el estado del arte, como las técnicas y tecnologías que se pueden utilizar en el ámbito de la detección de comunidades. Se ha profundizado principalmente, en el estudio de los métodos heurísticos y metaheurísticos, para conocer y comprender los conceptos fundamentales de los algoritmos evolutivos. Además, se ha realizado un análisis del problema de detección de comunidades para comprender su complejidad.

- La segunda etapa se ha centrado en el diseño e implementación de los algoritmos evolutivos propuestos (mono-objetivo y multiobjetivo) que integran el análisis flexible mediante grado de abstracción. Para realizar la implementación, se ha utilizado el lenguaje de programación C# .Net framework 4. En primer lugar, se han implementado las estructuras de datos que serán compartidas por todos los algoritmos. A continuación, se han implementado los algoritmos mono-objetivo, formados por procesos de inicialización eficientes y operadores genéticos específicos que han utilizado las estructuras de datos implementadas anteriormente. Se han implementado varios diseños mono-objetivos, combinando distintos procesos de inicialización y operadores genéticos, con el fin de compararlos entre ellos y detectar los más eficientes. Posteriormente, tomando como referencia el diseño mono-objetivo más eficiente, se ha implementado la versión multiobjetivo basada en dominancia de Pareto y métodos de inicialización y operadores evolutivos más avanzados. Todas las implementaciones realizadas han incluido un estándar para exportación de grafos (GraphML) que permite a otros softwares externos de visualización de grafos importar las soluciones obtenidas por los métodos propuestos.

- La tercera etapa se ha basado en la validación y optimización de los algoritmos implementados a través del uso de benchmarks populares en el campo de la detección de comunidades. Estos benchmarks se han dividido en reales (Books about US politics, American College footbal, IEEE 118, redes eléctricas, etc.), y sintéticos, generados por el método propuesto por Santo Fortunato [54] que permite simular redes reales de gran tamaño con distintas características. Para realizar las validaciones de los resultados obtenidos, se ha recurrido a procedimientos estadísticos estándar en este tipo de estudios como ANOVA y métricas de convergencia como NMI [55] o Hyper-Volumen [56]. Una vez se ha comprobada la validez de los resultados obtenidos, y de cara a evaluar el rendimiento de los métodos propuestos, se han comparado los resultados con los de otros métodos conocidos en detección de comunidades.

- La cuarta etapa se ha centrado en la aplicación de los métodos implementados sobre redes eléctricas de alta tensión de escala nacional y continental. Esta etapa ha requerido la obtención de los datos necesarios para elaborar las redes eléctricas, construir los grafos que las modelan y, finalmente, aplicar los métodos evolutivos para obtener las conclusiones pertinentes que reflejan la aportación realizada en esta tesis.

Capítulo 2

Adaptive community detection in complex networks using genetic algorithms

Community detection is a challenging optimization problem that consists in searching for communities that belong to a network or graph under the assumption that the nodes of the same community share properties that enable the detection of new characteristics or functional relationships in the network. A large number of methods have been proposed to address this problem in many research fields, such as power systems, biology, sociology or physics. Many of those optimization methods use modularity to identify the optimal network subdivision. This chapter presents a new generational genetic algorithm (GGA+) that includes efficient initialisation methods and search operators under the guidance of modularity. Further, this approach enables a flexible and adaptive analysis of the characteristics of a network from different levels of detail according to an analyst's needs. Results obtained in networks of different sizes and characteristics show the good performance of GGA+ in comparison with other five genetic algorithms, including efficient algorithms published in recent years.

2.1. Introduction

Network analysis is a powerful tool for describing many real systems, such as sociology [2], biology [57] or power systems [58], among many others. In particular, it has been proven that many real networks have a structure of modules which are characterized by groups of densely interconnected nodes. These modules or communities are subgraphs of the network such that nodes within a community are densely linked, while connections between subgraphs are sparser. These communities represent functional units in their networks, for example, a community in a social network corresponds to people that often communicate with other people of the same community, i.e., they share similar interests or backgrounds [57].

Having in mind the importance of Genetic Algorithms (GAs) [19] in computational optimization, this chapter presents a new generational genetic algorithm (GGA+) for solving the community detection problem, which is guided by the modularity index [40] and considers different degrees of abstraction. This concept provides the analyst the capacity to perform a flexible and adaptive analysis of the network using graphical information with different levels of detail.

The remainder of the chapter is organized as follows: Section 2.2 briefly describes the problem of community detection in graphs, and some previous approaches applied to this problem. Section 2.3 describes in detail the generational genetic algorithm here proposed. Section 2.4 presents the empirical study, where GGA+ and other five genetic algorithms are evaluated using four networks of different sizes and characteristics. The conclusions of the work are provided in Section 2.5.

2.2. Related work

The study developed by Leonhard Euler in 1736 about the Seven Bridges of Königsberg problem laid the foundations of graph theory and initiated an ongoing study of the properties of graphs [1]. Graphs are used to model real systems in many areas, such as power systems [58], sociology [2], biology [57], physics [4], and informatics [6, 53] thanks to the computers, which allow to efficiently representing, managing and processing large amounts of data, including graph-based structures. Among other emerging scientific areas, the analysis of communities in complex systems has gained importance due to their application to multiple contexts, including collaboration networks, biological systems, social networks in the Internet, transportation networks, or electrical networks [7]. All complex systems share a common characteristic: community structures [10] that consist of groups of nodes inside a network that are more densely connected than with the remaining nodes of the network. As the nodes that belong to the same community have a higher probability of shared properties, community detection can reveal new characteristics or functional relationships of a network. The community detection problem has been analyzed by the scientific community [57, 59] and consists in searching for the community structure that better represents the characteristics of a network. The complexity of this problem comes from the difficulty of determining the optimum community structure that best represents the characteristics of a network.

Modularity [40] has become one of the most extensively applied objective functions in community detection due to is simplicity and ease of calculation. Modularity provides a numerical value that represents the quality of the solution, such that the greater the value is, the more accurate the community structure. Modularity (Q) is defined as:

$$Q = \frac{1}{2M} \sum \left(a_{ij} - \frac{K_i K_j}{2M} \right) \delta(i, j)$$

where *M* represents the total number of edges in the network; the sub-indices *i* and *j* indicate two nodes of the network; Ki and Kj are the degree of the i th and *j* th nodes, respectively; the parameter a_{ij} is the element of the *i* th row and the *j* th column of the adjacency matrix; and $\delta(i, j)$ represents the relationship between the *i* th node and the *j* th node, such that if node *i* and node *j* are in the same community, $\delta(i, j) = 1$; otherwise, $\delta(i, j) = 0$.

2.3. GGA+: A new generational genetic algorithm for community detection

Many heuristic and metaheuristic methods have been proposed to solve community detection problems, including simulated annealing [14], swarm intelligence [60], and genetic and evolutionary algorithms [19]. This section presents a new generational genetic algorithm (GGA+) that is guided by the modularity index (MI) [40] and includes efficient strategies and search operators to detect communities in networks.

2.3.1. Initialisation

In many real-world situations, the number of community structures that form a network is known beforehand and, therefore, the search space to be explored by the algorithms can be reduced. In other cases, the number of community structures is initially unknown, but the algorithms can estimate a different number of community structures. Therefore, by determining beforehand the number of initial community structures to be detected, the search space can be reduced and the performance of the algorithm can be improved.

Despite random initialisation is often used, it can generate infeasible solutions, that is, nodes that are not interconnected due to a lack of relationship between them. To overcome this inconvenience, a safe initialisation is considered, such that each node *i* is connected with a neighbouring node *j* of the original graph [41]. Undesired clusters with disconnected nodes are prevented and, therefore, the search space of possible solutions is restricted to feasible individuals, which helps to improve the algorithm convergence. To prevent the generation of individuals having unbalanced communities, the concept of *community density* is used, which is incorporated into the structure of individual data to balance the loads between communities (*balanced initialisation*). In this phase, the community density vector of the individuals of the population is configured. The number of nodes that each community contains (community size) is a function of the number of communities to be detected. Subsequently, these sizes are employed to initialise each component of the *community density* vector with the number of nodes in each community. The genetic initialisation in GGA+ is based on the *safe and balanced initialisation* concept, in which a maximum node size is assigned to each community, as it is described below:

- Populate communities with neighbouring nodes: a community to be populated and the node n_i are selected. Based on n_i, the community is populated with this node and its neighbours nei_{ij} until the maximum community size is attained. If the community is not completed with the values nei_{ij} of the node n_i, another node n_{i+n} (not included in the nodes that already exist in the community) is selected to repeat the process. This procedure is repeated until the community is completed.

- Populate communities with adjacent neighbouring nodes: a community to be populated and a node n_i are selected, and a neighbouring node nei_{ij} of the node n_i is subsequently selected. Based on the node nei_{ij}, the community is completed with this node and its neighbours nei_{ijk} until the maximum community size is attained. If the community is not completed with the neighbouring nodes nei_{ijk} of the node nei_{ij}, another neighbouring node $nei_{i(j+1)}$ of the node n_i (not included in the nodes that already exist in the community) is selected to repeat the process. This procedure is repeated until the community is completed.

Figure 2.1 displays an example that illustrates both concepts: *safe initialisation* and *balanced initialisation*. The number of communities to be detected in Figure 2.1 (two communities in this case) is used to calculate the node size for each community in the balanced initialisation stage. Once the size of each community is defined, the safe initialisation stage begins; in this case, the *populate communities with neighbouring nodes* method is used to initiate the individual. Therefore, the community structure is generated by the combination of the safe and balanced initialisations, which coincides with the optimum problem solution in the case study presented in Figure 2.1.

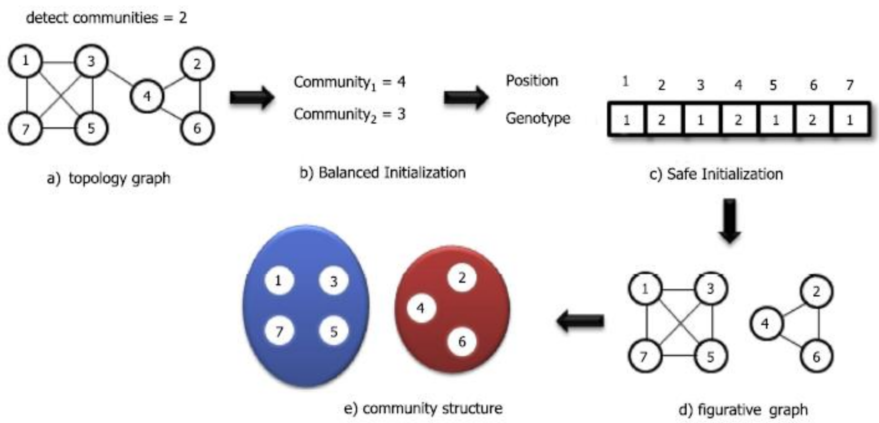


Figure 2.1. Safe and balanced initialisation scheme.

On the other hand, initialisation of the migration vector between boundaries is based on the idea of migration towards the most attractive destination, which coincides with the community that has the greatest number of nodes directly connected to a certain node. Given a node, the community that contains the highest number of nodes connected to this one will be selected as the destination boundary community to which the selected node will migrate.

The scheme described for the initialisation of the migration vector between boundaries is shown in Figure 2.2. Figure 2.2(d) shows the resulting vector, which elements represents nodes of the graph, its values, and the destination community to which each node could migrate. Table 2.1 shows the boundary community generation method, in which the columns *Communities 1* and *2* indicate, for the corresponding community, the number of nodes directly connected to the node of the column *Nodes*. The column *Migration Vector* shows the destination community selected for each node based on the highest recorded value among all community counters (in this example, the counters are communities 1 and 2).

Table 2.1. Generation method of border community.

Nodes	**Communiry 1**	**Community 2**	**Migration Vector**
1	2	1	1
2	0	2	2
3	3	1	1
4	0	3	2
5	2	1	1
6	0	2	2
7	2	1	1

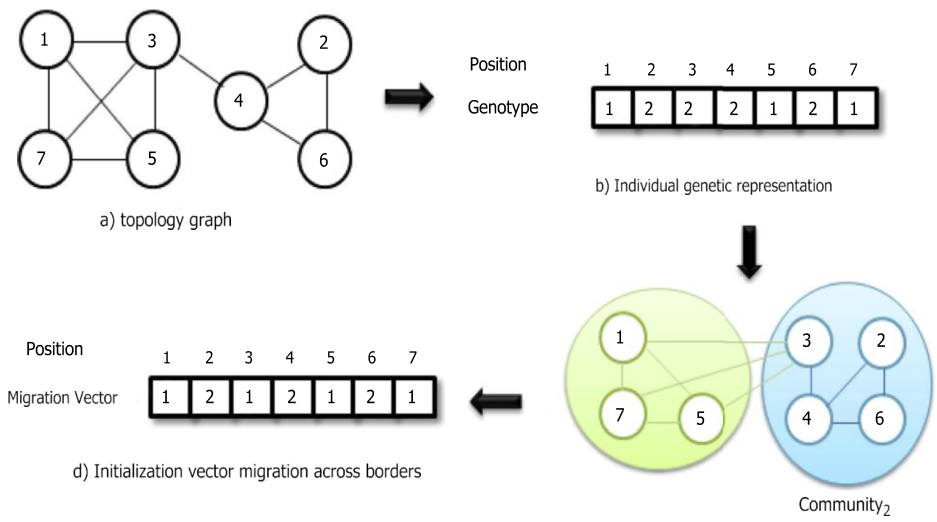


Figure 2.2. Initialisation migration vector scheme.

2.3.2. Genetic operators

The algorithm here proposed (GGA+) uses genetic search operators that have been especially designed to obtain the maximum performance of the proposed data structure.

The crossover operator is based on the exchange of nodes between boundaries, such that it is selected a first node with a destination boundary community that differs from the current boundary community to perform an exchange with a second node from the destination boundary community of the first selected node. The second node satisfies the criterion of being to the best related node to the actual community of the first node selected. For the calculation of the second node of the best related destination community, the first selected node is disregarded. An advantage of this operator is that it maintains the balance in the number of nodes between communities. Figure

2.3 shows the phases of the mutation operator. The set of nodes suitable for mutation are nodes with a destination community that differs from the actual community. Subsequently, a random node of this set is selected, which will be exchanged with a node from the destination community. In this example, node 3 is selected for the exchange; its destination community is node 1. Figure 2.3(c) shows how the node of the destination community (community 1) that is best related to the origin community (community 2) is selected to be exchanged. The column “Neighbours: origin community” lists the number of neighbours (directly connected nodes). The node selected for the exchange is the node with the highest number of neighbours; node 5, which has one neighbour (node 4), is selected in this case. This calculation disregards the node selected in the origin community (node 3), as this node migrates to the destination community and no longer belongs to the origin community. Phase d) shows the state of the individual after mutation.

The mutation operator is based on the migration of nodes between boundaries of two different communities. As it is observed in Figure 2.4, this operator is applied to a part of an individual’s genes, depending on a previously defined reproduction ratio, which will reflect the amount of genes that could migrate in this operation. A value that corresponds to the *migration between boundaries* vector is assigned to each selected gene, such that each node migrates to the destination community with the best relationship.

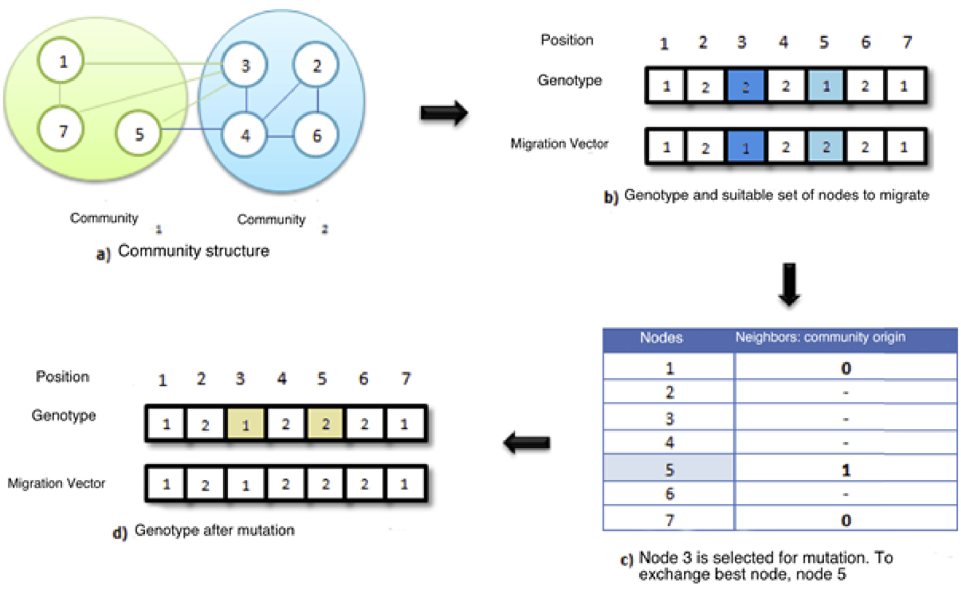


Figure 2.3. Crossover operator based on the exchange of nodes between boundaries.

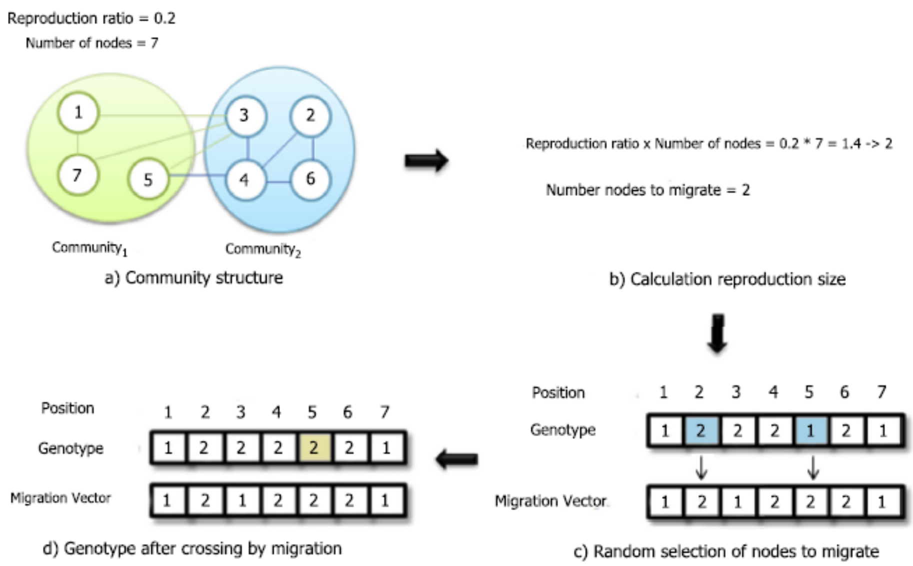


Figure 2.4. Mutation operator based on migration of nodes between boundaries.

2.3.3. Degree of abstraction

The objective of this work is not only to design an efficient robust algorithm to generate a solution that enables an analysis of the characteristics of a network from a general point of view. Moreover, the genetic representation included in this study enables the number of communities of the solution to be controlled in order to obtain different of details that could facilitate the analysis of the network from different perspectives. To do this, GGA+ is designed to allow that the individuals of the population search for different number of communities (*network degree of abstraction*), such that a set of solutions is obtained that enables the network to be analysed from the most general perspective (lowest number of detected communities) to the most detailed perspective (highest number of detected communities). Figure 2.5 illustrates the concept of *degree of abstraction* considering the interval [2,6]: given a set of solutions with a different number of detected communities, the solution with the lowest number of communities (two communities) is equivalent to the maximum degree of abstraction, whereas the solution with the highest number of communities (six communities) is equivalent to the minimum degree of abstraction. The maximum and minimum values of the degree of abstraction define the *range of available solutions* as all solutions from the interval [maximum degree of abstraction, minimum degree of abstraction], with an integer number of detected communities.

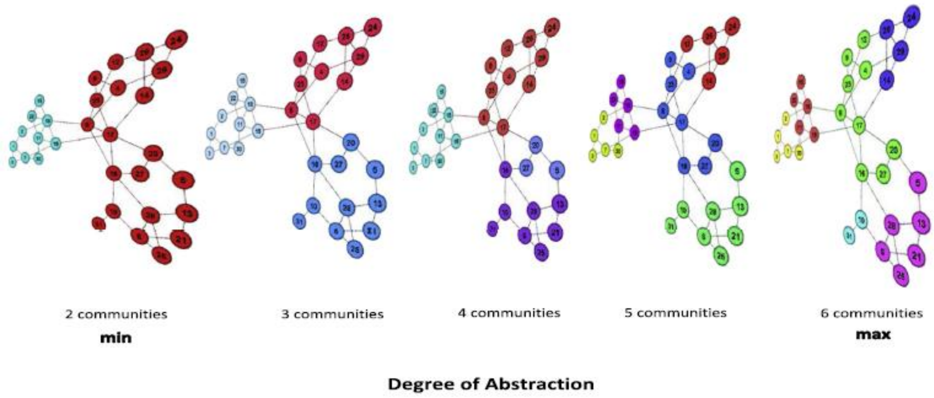


Figure 2.5. Degree of abstraction of a network.

2.4. Empirical study

In this section, they are analysed the results obtained by GGA+ and other genetic algorithms when detecting communities in four networks of different sizes and characteristics. The experiments were performed on a personal computer with an Intel Core i7 3630Q processor (2.4 GHz, 8 GB DDR3 RAM), which executes the application developed in C# .Net Framework 4.

2.4.1. Benchmarks

To conduct the performance analysis of the different genetic algorithms, four real networks of different sizes and characteristics are considered.

- Books about US politics: this network is composed of 105 books (nodes) and 441 edges, and the optimal solution proposed by its author is composed by three communities [61].
- American college football: this network is composed of 115 nodes and 613 edges, and the optimal solution is composed by 12 communities [62].

- 250N heterogeneous network: this network is composed of 250 nodes and 2436 edges, and the optimal solution is formed by 13 communities [54]. This graph introduces features of real networks, i.e., the heterogeneity in the distributions of node degree and community size.
- 500N heterogeneous network: based on the same idea than the previous 250N benchmark, this network is composed of 500 nodes and 5020 edges, and the optimal solution is composed by 24 communities [54].

2.4.2. Algorithms

The GA functional model defines the strategies for selecting the reproduction and replacement strategy to create new individuals in the population. In this study different stationary and generational models have been designed and implemented. Moreover, some of the GAs included in the comparison (SA-SGA, SA-GGA, and MIGA) use simulated annealing [14] is applied to escape from local optima [63]. The methods apply the local search method after performing a genetic operation.

The characteristics of GGA+ and the different versions of the GAs are listed in Table 2.2. Moreover, this table also includes the main characteristics of the Modularity and Improved Genetic Algorithm (MIGA) [45]. MIGA, which was proposed a few years ago, outperformed to another genetic algorithm and also to a memetic algorithm.

Table 2.2. Characteristics of proposed genetic algorithms.

Algorithm	Type	Crossover operator	Mutation operator	Simulated annealing
SGA	Stationary	2-point	Random	No
GGA	Generational	2-point	Random	No
SA-SGA	Stationary	2-point	Random	Yes
SA-GGA	Generational	2-point	Random	Yes
MIGA	Generational	2-point	Random	Yes
GGA +	Generational	Border exchange	Border migration	No

2.4.3. Performance measures

Normalised Mutual Information (NMI) is a performance measure for community detection. NMI is used to evaluate the performance of the proposed GAs. It is defined in the equation below according to Danon et al. [55]

$$NMI(A,B) = \frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B} C_{ij}\log\left(\frac{C_{ij}N}{C_iC_j}\right)}{\sum_{i=1}^{C_A} C_i\log\left(\frac{C_i}{N}\right) + \sum_{j=1}^{C_B} C_j\log\left(\frac{C_j}{N}\right)}$$

where two parts - A and B - of the network C form the confusion matrix. C_{ij} is the number of nodes that belong to community *i* of part A and community *j* of part B; C_A is the number of communities of part A(B); C_i is the number of nodes in community *i* of part A; and N is the total number of nodes. If $A = B$, then $NMI(A,B) = 1$; if A is completely different from B, $NMI(A,B) = 0$. The closer the NMI value is to 1, the more similar is the detected community structure to the best structure of the network.

2.4.4. Parameter configuration

To perform a fair comparison of all design algorithms, the same parametric configuration is considered, after a preliminary sensitivity analysis. Table 2.3

shows the values of these general (common) parameters, and also the specific parameters used by SA-SGA, SA- GGA, and MIGA to apply Simulated Annealing [14], which have been established according to [45].

Table 2.3. Parameter settings.

General parameters	
Population size	200
Iterations	200
Maximum iterations without improvement	50
Crossover probability	0.8
Mutation probability	0.2
Specific parameters (algorithm)	
SA-SGA/SA-GGA/MIGA	Initial temperature = 800,000 Cooling rate = 0.99
GGA +	Loop count l = 10 Reproduction ratio = 0.2

2.4.5. Results and discussion

To conduct the performance analysis, a total of 50 independent runs have been performed for each algorithm on each benchmark. The algorithms have been configured to find the same optimal number of communities according to the author of each benchmark. This strategy allows us to compare the solutions obtained by the algorithms with the optimal solution using the performance measure NMI.

The accuracy of the algorithms has been evaluated according to the mean NMI values (Table 2.4) and the mean modularity values (Table 2.5). From these results, it is observed how GGA+ achieves the best mean and standard deviation in all test instances performance in all test cases, both using the NMI and the modularity index. It is observed that the larger the problem instance, the greater the advantage of GGA+ with respect to the other methods. Moreover, the standard deviation (shown between parentheses) is lower in GGA+ than in other methods, which denotes the robustness of the results obtained by GGA+.

Table 2.4. NMI values obtained by each algorithm in all the benchmarks (mean and standard deviation of 50 independent runs).

Benchmark	SGA	GGA	SA-SGA	SA-GGA	MIGA	GGA +
Books about	0.5398	0.1902	0.5188	0.5282	0.5500	0.5538
US politics	(0.0417)	(0.0552)	(0.0521)	(0.0541)	(0.0037)	(0.0000)
American college	0.5775	0.3452	0.8262	0.8188	0.8486	0.9259
football	(0.0548)	(0.0270)	(0.0376)	(0.0356)	(0.0295)	(0.0041)
250N heterogeneous	0.3061	0.1801	0.8862	0.8626	0.7373	0.9996
network	(0.0272)	(0.0151)	(0.0396)	(0.0401)	(0.0197)	(0.0028)
500N heterogeneous	0.2441	0.1905	0.6591	0.5566	0.6533	0.9943
network	(0.0127)	(0.0201)	(0.0222)	(0.0194)	(0.0134)	(0.0072)

Table 2.5. Modularity values obtained by each algorithm in all the benchmarks (mean and standard deviation of 50 independent runs).

Benchmark	SGA	GGA	SA-SGA	SA-GGA	MIGA	GGA +
Books about	0.4655	0.2533	0.4982	0.4963	0.5168	0.5221
US politics	(0.0212)	(0.0372)	(0.0205)	(0.0185)	(0.0080)	(0.0000)
American college	0.3617	0.1170	0.5468	0.5402	0.5497	0.6008
football	(0.0331)	(0.0183)	(0.0242)	(0.0251)	(0.0170)	(0.0014)
250N heterogeneous	0.1524	0.0488	0.4618	0.4482	0.3601	0.5126
network	(0.0138)	(0.0074)	(0.0198)	(0.0233)	(0.0138)	(0.0 0 07)
500N heterogeneous	0.0665	0.0293	0.3214	0.2484	0.2990	0.5525
network	(0.0056)	(0.0051)	(0.0142)	(0.0129)	(0.0074)	(0.0023)

In the smaller network - Books about US politics - GGA+, with a value of 0.5538, is much better than GGA and slightly better than the other methods. In addition to the numerical results, another main point to consider in the performance analysis is the algorithm stability, which indicates the similarity among the solutions obtained in each independent run. According to the Figure 2.6, GGA+ not only obtains the best mean results in this network, but it also obtains solutions with high level of stability (the 50 independent runs performed with GGA+ obtain the same result).

We must notice that the despite of real community structure is not obtained by GGA+ (as the mean NMI values is never 1), this does not mean the algorithm performance is poor. The reason of this difference comes from the fact that the optimal communities proposed by the author of this network seem not to consider the modularity, while the algorithms here analysed are guided by modularity. This situation is described in Figure 2.7, where the optimal solution defined for the authors of this network is compared with MIGA and GGA+. From this figure it is possible to conclude that according to modularity the solution obtained by GGA+ has a very high quality, but it is somewhat different to the optimal configuration of communities defined by the authors of this network, which is the reference used to calculate NMI.

The results obtained in the network Books about US politics (Figure 2.8) reveals that the greater the degree of abstraction, the more abstract the solution is compared with the optimum solution with two communities. Therefore, Figure 2.8(c) and (d) can detect new more specific communities, up to a total of 5 communities. Similar to the remaining graphs, with the minimum degree of abstraction of six communities, the community detection limit is attained.

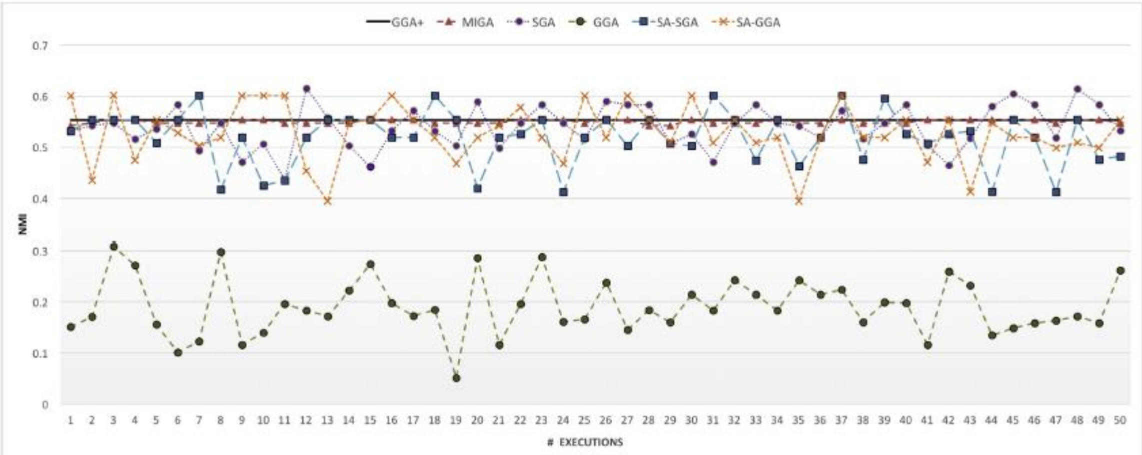


Figure 2.6. Comparison of stability based on NMI for 50 executions on Books about US politics network.

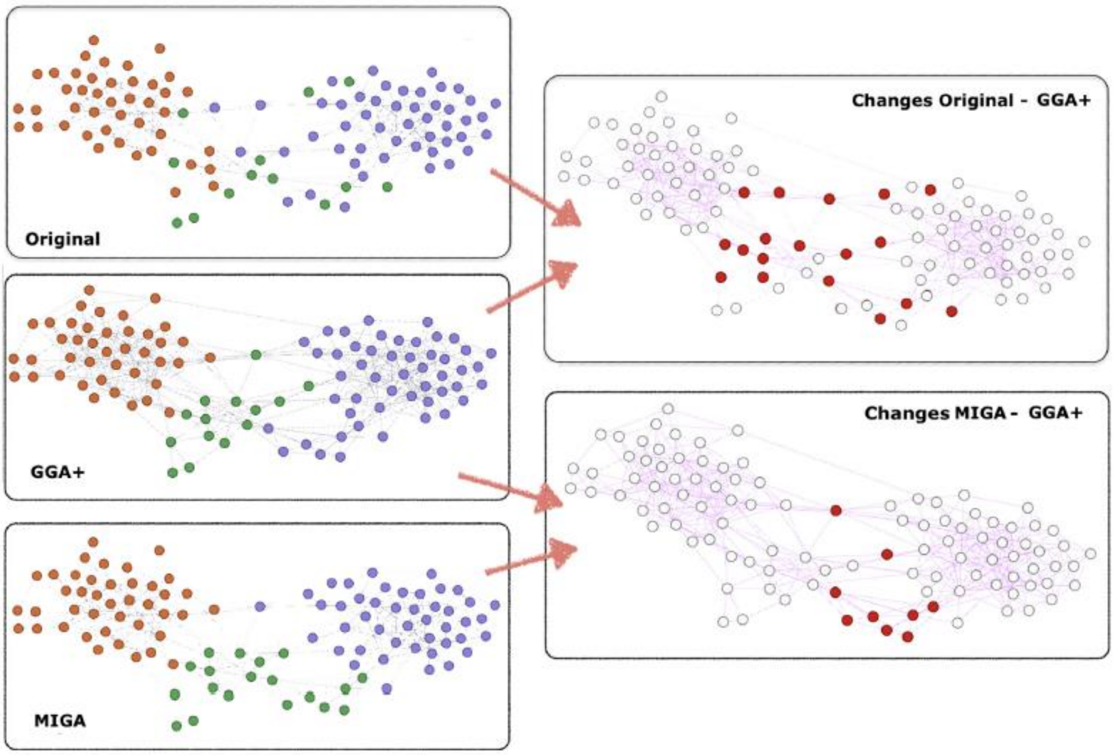
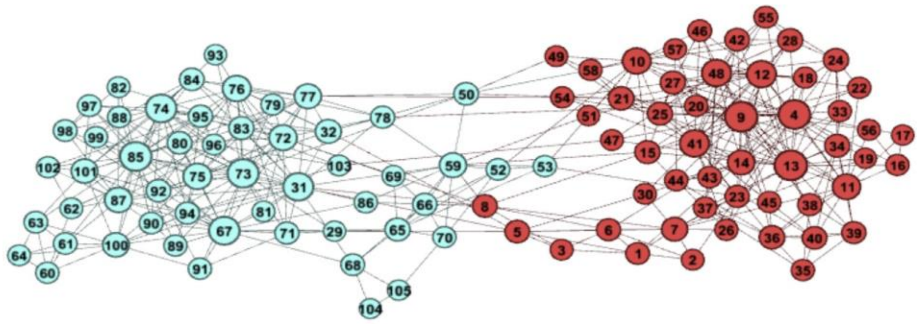
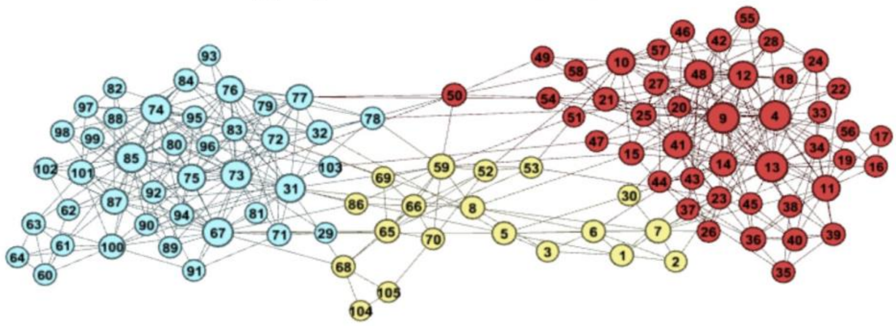


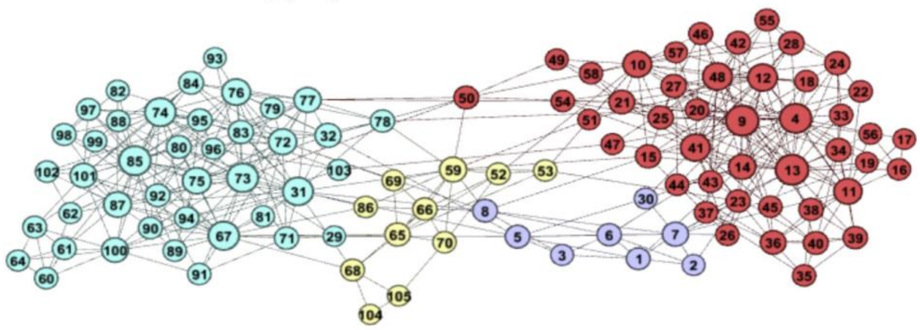
Figure 2.7. On comparing MIGA and GGA+ with the optimal solution taken as reference to calculate NMI (Books about US politics).



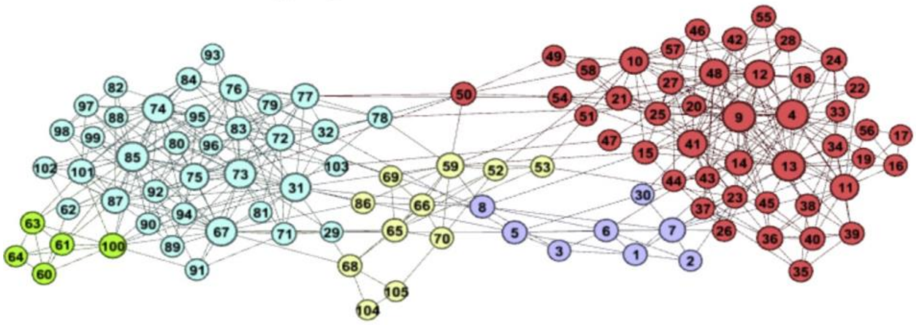
(a) Degree of abstraction 2 communities



(b) Degree of abstraction 3 communities



(c) Degree of abstraction 4 communities



(d) Degree of abstraction 5 communities

Figure 2.8. Set of obtained solutions for Books about US politics network in the range [2,5].

The results obtained with the second network in size - American College football network - show that GGA+ and MIGA obtain a solution near the optimum solution, while the rest of methods, especially GGA, obtain a considerably different solution, which highlights the features of GGA+. Figure 2.9(c) shows how GGA+ is clearly the method with higher level of stability (only 47 of the 50 independent runs obtain the same best result). However, the rest of methods are not so stable, as it can be clearly seen from the oscillations in the solutions obtained, especially in the case of SGA, which values of NMI oscillate between 0.45 and 0.70.

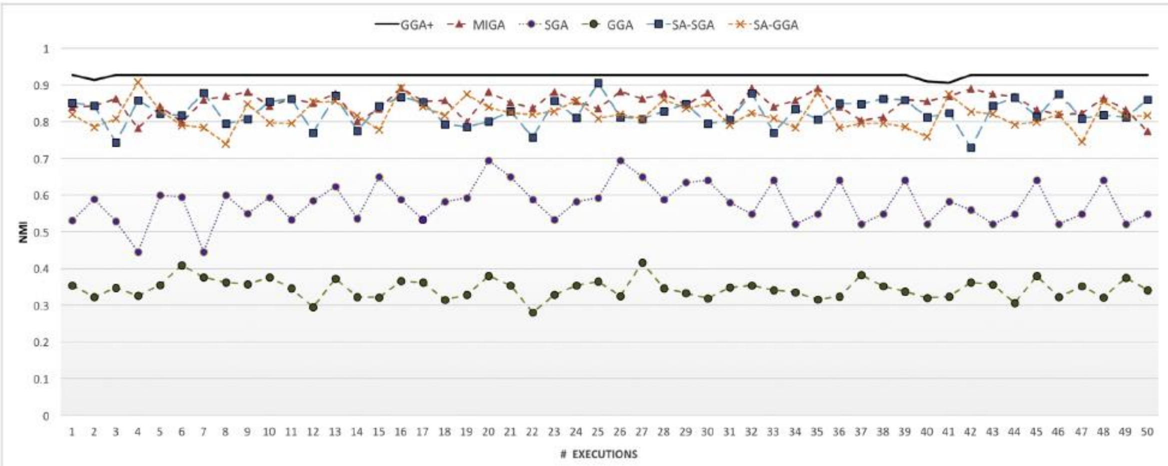
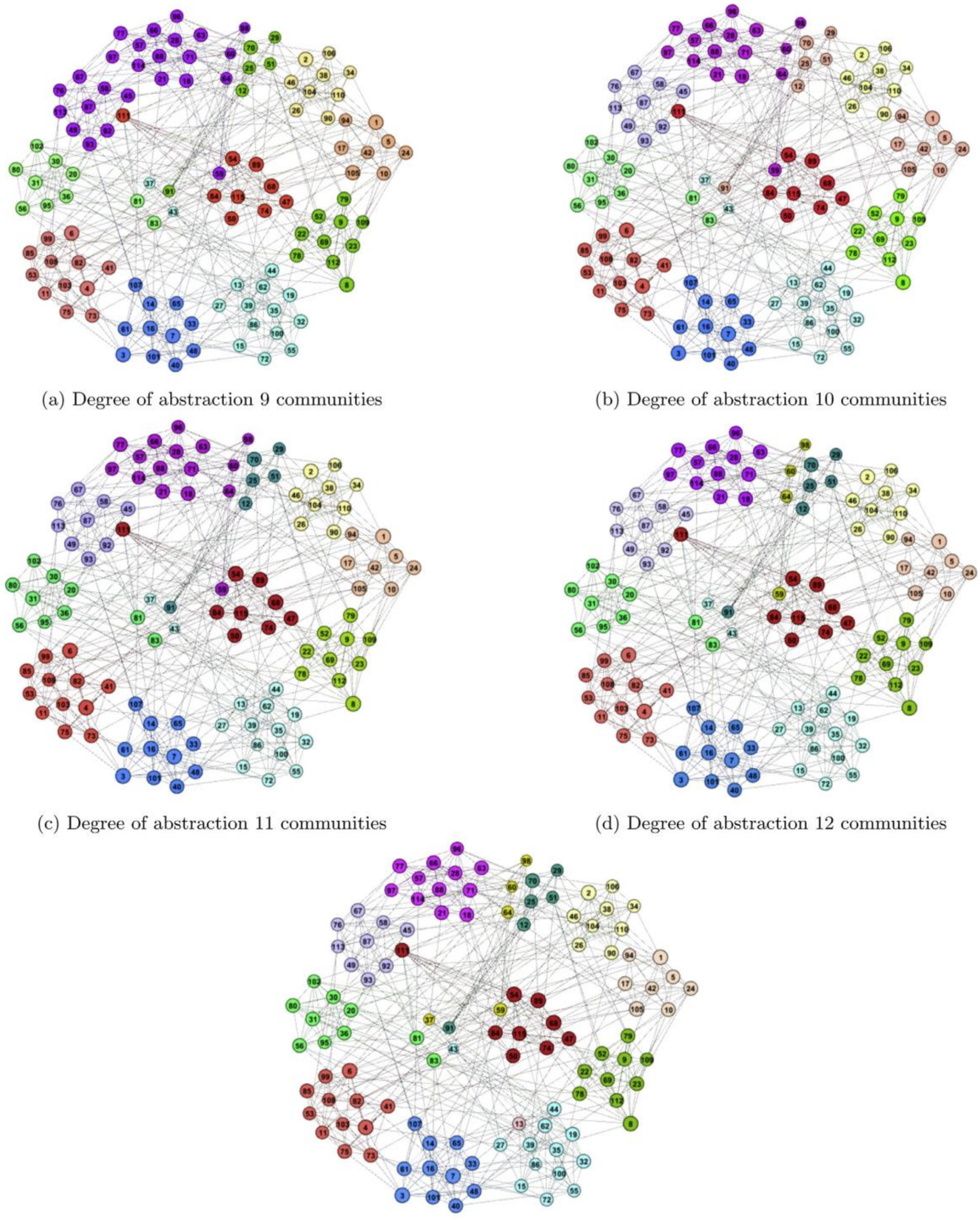


Figure 2.9. Comparison of stability based on NMI for 50 executions on American college football network.

Figure 2.10 also shows the results obtained by GGA+ using different degrees of abstraction (from 9 to 13 communities). It is observed that the greater the degree of abstraction is, the more abstract the solution is compared with the solution obtained for nine communities.



(a) Degree of abstraction 9 communities

(b) Degree of abstraction 10 communities

(c) Degree of abstraction 11 communities

(d) Degree of abstraction 12 communities

(e) Degree of abstraction 13 communities

Figure 2.10. Set of obtained solutions for American college football network in the range [9,13].

Table 2.4 shows that GGA+ also obtains the best results in the largest networks (250N and 500N). Figs. 2.11 and 2.12 compare the stability of the six algorithms in these largest networks. In both cases, GGA+ is the most stable algorithm, and also the method that obtains the best results in the different independent runs. The communities detected by GGA+ in the heterogeneous networks of 250 and 500 nodes with the optimal number of communities are displayed in Figs. 2.13 and 2.14, respectively. As it can be seen in both figures, GGA+ is able to obtain the optimal configuration according to NMI. Having in mind these figures, and the results provided in Table 2.4 and Table 2.5, it can be concluded that GGA+ not only outperforms to the other five genetic algorithms, but it also has a good performance in large networks.

As shown in Table 2.5, GGA+ is the algorithm that obtains better mean results for modularity. These results show that the genetic operators implemented in GGA+ - *Mutation operator based on migration of nodes between boundaries* and *crossover operator based on the exchange of nodes between boundaries* - achieve a better maximisation of the modularity objective function compared with the basic genetic operators, which indirectly allows to GGA+ to obtain the best results in terms of NMI.

Whenever several algorithms are compared it is necessary to determining if there a significant difference between the results obtained by these methods. With this aim, it is applied the one-way ANOVA, which results should be interpreted as follows: given the typical confidence level of 95%, the null hypothesis is rejected if the probability value is smaller than or equal to 0.05 (p - value $\leq$ 0.05). Table 2.6 shows the p-values obtained by applying the one-way ANOVA to the 50 independent runs from which Table 2.4 has been built, where it is possible to conclude that the null hypothesis is rejected in all cases, i.e., there is a significant variation between at least some of the means of the different groups, which denotes that the use of different algorithms has a real effect in the

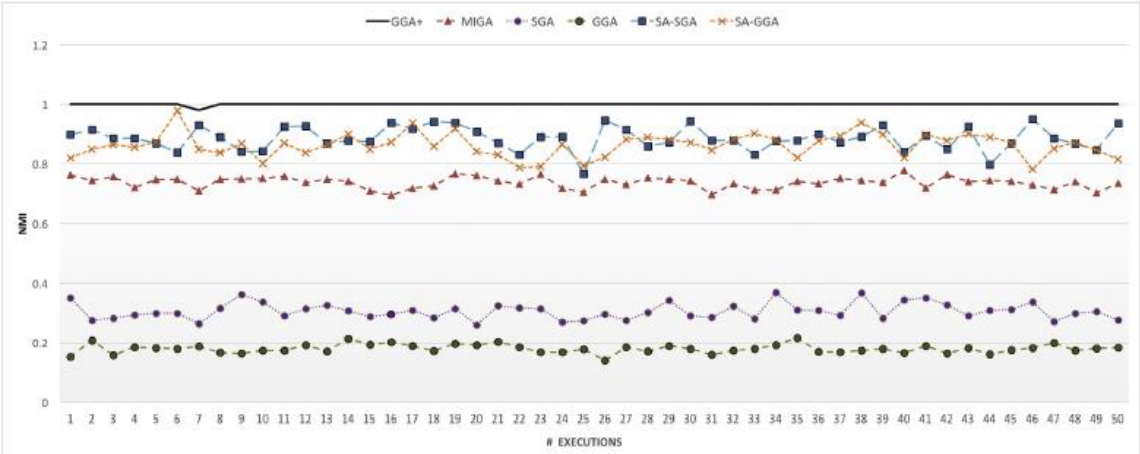


Figure 2.11.Comparison of stability based on NMI for 50 executions on the 250N heterogeneous network.

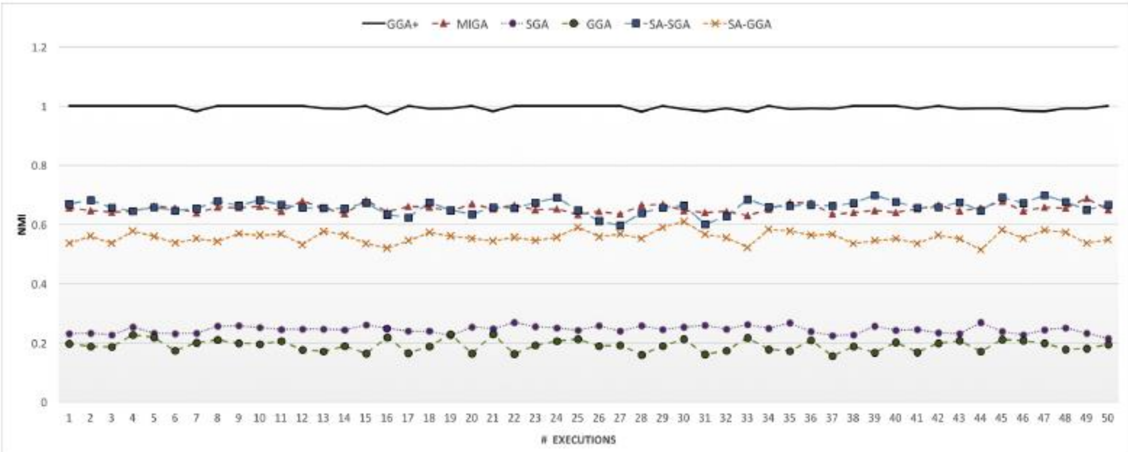


Figure 2.12. Comparison of stability based on NMI for 50 executions on the 500N heterogeneous network.

Table 2.6. Analysis of variance (ANOVA) using the primary results of the 50 independent runs.

Benchmark	*p-value*	*Null hypothesis*
Books about US politics	*1,9152E −150*	*Reject*
American college football	*7,3829E −224*	Reject
250N heterogeneous network	*7,3205E −308*	Reject
500N heterogeneous network	≈ 0	Reject

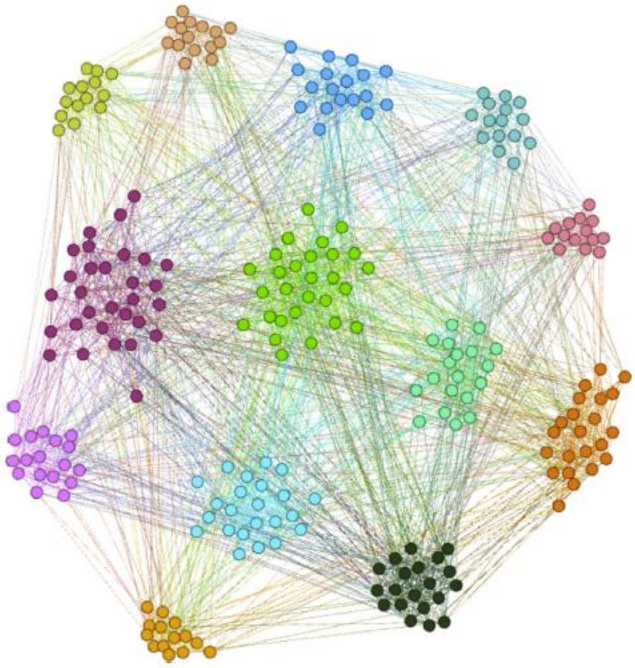


Figure 2.13. Communities detected by GGA+ in the heterogeneus network of 250 nodes (13 communities).

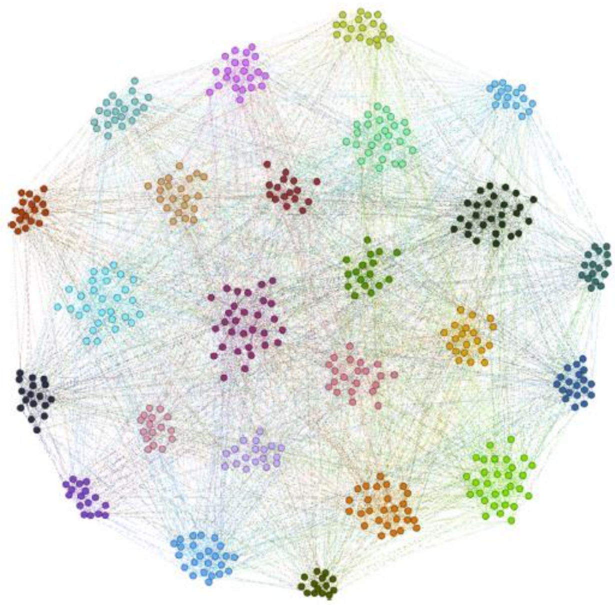


Figure 2.14. Communities detected by GGA+ in the heterogeneous network of 500 nodes (24 communities).

results. Since the null hypothesis of ANOVA is rejected in all cases, Tukey's honestly significant difference (HSD) test [64] is then conducted. This is a single-step multiple comparison procedure that is often used in conjunction with an ANOVA to find means that are significantly different from each other [65]. Table 2.7 provides the test results for four test problems. Based on these results, it is realized that using different algorithms often produce significantly differences from each other independently of the characteristics of the test problems. In particular, results obtained by GGA+ are significantly different to the other cases in the three larger networks, which allow to validate the mean values displayed in Table 2.4.

Table 2.7. Tukey's multiple-comparison test (× / ✓): the means of the 50 independent runs of the pair of algorithms compared are different / similar).

Algorithm	*Differences*				
(Books about US politics)	*GGA*	*SA-SGA*	*SA-GGA*	*MIGA*	*GGA +*
SGA	✓	×	×	×	×
GGA		✓	✓	✓	✓
SA-SGA			×	✓	✓
SA-GGA				×	✓
MIGA					×
(American college football)	*GGA*	*SA-SGA*	*SA-GGA*	*MIGA*	*GGA +*
SGA	✓	✓	✓	✓	✓
GGA		✓	✓	✓	✓
SA-SGA			×	✓	✓
SA-GGA				✓	✓
MIGA					✓
(250N heterogeneous network)	*GGA*	*SA-SGA*	*SA-GGA*	*MIGA*	*GGA +*
SGA	✓	✓	✓	✓	✓
GGA		✓	✓	✓	✓
SA-SGA			✓	✓	✓
SA-GGA				✓	✓
MIGA					✓
(500N heterogeneous network)	GGA	SA-SGA	SA-GGA	MIGA	GGA +
SGA	✓	✓	✓	✓	✓
GGA		✓	✓	✓	✓
SA-SGA			✓	×	✓
SA-GGA				✓	✓
MIGA					✓

It is important to indicate that the computational times required to perform each of the 50 executions oscillate between a few seconds in those methods that do not apply simulated annealing (SGA, GGA, GGA+) to a few minutes in the case of the algorithms that apply simulated annealing (SA-SGA, SA-GGA and MIGA).

Once the performance analysis of the different presented algorithms is performed based on the accuracy, stability and computational time, it is possible to conclude that GGA+ achieves the best performance. At this point, it is interesting to determine which algorithmic components included in GGA+ are more important to obtain high quality results. With this aim, Table 2.8 shows the results obtained by GGA+ and three different variants when solving the largest network (500N heterogeneous network), where:

- GGA+ (11): the same GGA+ algorithm described above.
- GGA+ (10): GGA+ with random initialisation and advanced search operators.
- GGA+ (01): GGA+ with balanced initialisation and random search operators.
- GGA+ (00): GGA+ with random initialisation and random search operators.

The results obtained show that both, the initialisation strategy and, especially, the search operators are key factors to outperform the other five genetic algorithms used in the comparison. Therefore, it is possible to conclude that these strategies included in GGA+ could be used by many other heuristic and metaheuristics that could be proposed to solve community detection problems.

Table 2.8. NMI values obtained by GGA+ and algorithm in 500N heterogeneous network benchmarks (mean and standard deviation of 50 independent runs).

GGA + (00)	GGA + (01)	GGA + (10)	GGA + (11)
0,2489 (0,0091)	0,5953 (0,0097)	0,8880 (0,0214)	0,9943 (0,0072)

2.5. Conclusions

This chapter analyses the performance of genetic algorithms for solving community detection problems, and presents a new generational genetic algorithm (GGA+) that includes efficient initialisation methods and search operators under the guidance of modularity. The empirical study includes the comparison of GGA+ with other five genetic algorithms, some of which use generational and stationary strategies and/or simulated annealing during the optimization process. We adopt a flexible and adaptive analysis of the characteristics of a network from different perspectives at a greater or lower level of detail (degree of abstraction), which enables an analyst to evaluate a network using this efficient and scalable approach, which encompasses the data structure, genetic operators, data exportation format and graphical analysis by the graph visualisation software. Four different benchmarks, ranging from 105 to 500 nodes, and from 441 to 5020 edges are considered in the study. The numerical and graphical results, supported by ANOVA and Tukey's statistical tests, indicate that the results obtained by GGA+ outperform the other five methods in all test problems, especially in larger problem instances. Therefore, the initialisation strategies and search operators included in GGA+ could be incorporated in other heuristic and metaheuristic methods to efficiently detect communities in complex networks.

Capítulo 3

Community detection in national-scale high voltage transmission networks using genetic algorithms

The large-scale interconnection of electricity networks has been one of the most important investments made by electric companies, and this trend is expected to continue in the future. One of the research topics in this field is the application of graph-based analysis to identify the characteristics of power grids. In particular, the application of community detection techniques allows for the identification of network elements that share valuable properties by partitioning a network into some loosely coupled sub-networks (communities) of similar scale, such that nodes within a community are densely linked, while connections between different communities are sparser. This chapter proposes the use of competitive genetic algorithms to rapidly detect any number of community structures in complex grid networks. Results obtained in several national-scale high voltage transmission networks, including Italy, Germany, France, the Iberian peninsula (Spain and Portugal), Texas (US), and the IEEE 118 bus test case that represents a portion of the American Electric Power System (in the Midwestern US), show the good performance of genetic algorithms to detect communities in power grids. In addition to the topological analysis of power grids, the implications of these results from an engineering point of view are discussed, as well as how they could be used to analyze the vulnerability risk of power grids to avoid large-scale cascade failures.

3.1. Introduction

The growing demand for electricity has involved that high voltage transmission networks have become one of the most important infrastructures in our society societies [66]. Furthermore, the complexity of power grids has increased with economic development, necessitating the application of robust control and optimisation strategies to manage large-scale systems [67].

Different studies have shown that graph-based network analysis is a powerful tool for describing many real systems in a variety of fields [68], including engineering tasks [69, 70, 71]. Community structure is an important feature of graphs representing real systems, since many real networks have clusters, with many edges connecting nodes within the same cluster, and comparatively few edges connecting to nodes in different clusters. Finding the optimal partition of the vertices of a graph into clusters such that the corresponding modularity [40] is maximized is an NP-hard problem [11, 72]. As community detection is a difficult problem, complex computational and mathematical techniques are needed.

Some authors have applied community detection techniques to manage small and medium-sized power grids [58, 73, 74, 75], but a little attention has been paid to solve this problem in national-scale electrical networks. To cover this gap in the literature, this chapter analyses the performance of evolutionary approaches for solving the community detection problem with applications to several national-scale high voltage power grids. These algorithms, which are guided by the modularity index [40] and consider different degrees of abstraction (i.e. the methods are able to detect any number -k- of communities), allow for a flexible and adaptive analysis of the grid by considering different levels of detail.

The remainder of the chapter is organised as follows: Section 3.2 introduces the problem of community detection in graphs, and revises some previous studies that have been applied to electrical grids. Section 3.3 presents the main characteristics of two genetic algorithms used to solve the community detection problem in graphs. Section 3.4 presents an empirical study that compares the performance of these methods in detecting communities in five national-scale grids, and also in a well-known benchmark. Section 3.5 discusses the results and implications of this work into the engineering field, emphasizing the possibility of using the proposed methods for line outage contingency analysis, including the system islanding scenario. Finally, the conclusions of this work are presented in Section 3.6.

3.2. Related work

Electricity networks have been built since the end of the XIX century [76]. Electrification still continues today, leading to a high degree of interconnection spanning states and now reaching a continental scale. In the past, the distribution system was unidirectional, distributing electricity from a small number of large power plants down to end users, whose demand was generally regarded as rigid and exogenous. However, grid operation has changed significantly in recent decades for several reasons, including the integration of variable output renewable energy sources [77]. The popularity of renewable energy has led electricity generated in power plants to be complemented by renewable power sources, some of which are located in industrial installations and residential buildings, with the result that the distribution system has become bidirectional. Moreover, the use of renewable sources has presented an alternative to grid extension for remote village electrification [78]. Unfortunately, the uncertainty and variability of wind and solar generation affects the grid operations, although some recent studies have shown that these inconveniences can be mitigated by balancing the variability of renewable sources using the transmission grid and balancing with storage [79].

The growing worldwide demand for electricity, together with the inclusion of new power plants, requires increasing grid connectivity and applying complex control methods. Some recent investigations have proposed the analysis of power grid infrastructures using graph-based complex network analysis techniques [74], such that the nodes of the network often represent the power plants, distribution substations and transmission substations, while the edges correspond to transmission lines. Advances in computer science have allowed for the efficient representation, management and processing of large amounts of data, including graph-based networks representing real systems. The application of graph-based analysis techniques has allowed for the analysis of the topological structure of networks representing power grids [80]. For example, some studies have analyzed the vulnerability of power grids to blackouts using graph topological indexes [81].

All complex systems share a common characteristic: community structures [10]. Communities consist of groups of nodes inside a network that are more densely connected with each other than with the remaining nodes of the network. As the nodes belonging to the same community have a higher likelihood of interaction, detecting those communities can reveal characteristics or functional relationships in a given network. Therefore, the community detection problem consists of partitioning the nodes in a network into groups such that there are many edges connecting nodes within the same group, and comparatively few edges connecting nodes in different groups. In the case of power grids, communities represent high-voltage lines that are densely connected.

Some studies have applied community detection to power grids. In [73], it was introduced a node similarity index to assign each node to the community sharing maximum similarity, which exhibited a good performance in a set of experiments performed on several IEEE standard power grids. Other authors have analyzed the optimal phasor measurement unit placement problem using algorithms for community detection to identify coherent groups based on an equivalent graph of generator nodes [82]. Another study presented a hierarchical spectral clustering method that reveals the internal connectivity structure of the power transmission capability of islanding systems using a network with nodes and links representing buses and electrical transmission lines, respectively [75]. Some investigations have analyzed the possibility of using community detection for islanding power systems as an emergency response to isolate failures that might propagate and lead to major disturbances [58]. Community detection has also been applied to analyze the vulnerability of the power systems under terrorist attacks [83], among other applications.

3.3. Evolutionary algorithms

Evolutionary computation [32] is a research field closely related to computational intelligence that is focused on designing algorithms to solve complex global optimisation problems. Evolutionary algorithms are problem-solving procedures that include evolutionary processes as the key design elements. In particular, an evolutionary algorithm consists of a population of individuals that continually and selectively evolve until a termination criterion is fulfilled.

Among evolutionary techniques, Genetic Algorithms (GAs) [19] are likely the most widely used. A genetic algorithm mimics natural selection by evolving a population of individual solutions to the problem at hand over time until a termination condition is fulfilled and the best individual is taken as an acceptable solution. Two of the most important characteristics of GAs are the representation used (e.g., binary or real) and the genetic operators employed (e.g., mutation and crossover).

As the community detection problem is highly complex, researchers have applied heuristics and metaheuristics to obtain high quality solutions with a reduced runtime. In particular, GAs are selected because they have been used to solve many electrical problems [84, 85, 86, 87, 88]. In this study, two genetic algorithms have been adapted to solve community detection problems in power grids. These recently proposed algorithms (MIGA and GGA+) have been shown to be more effective than other approaches to community detection based on benchmarks typically used to compare algorithms for this problem.

- The Modularity and Improved Genetic Algorithm (MIGA) [45] takes the modularity (Q) as the objective function, and uses the number of community structures as prior information to improve the stability and accuracy of community detection. MIGA also uses Simulated Annealing [14] as a local search strategy.

- The Generational Genetic Algorithm (GGA+) [89] includes efficient and safe initialisation methods in which a maximum node size is assigned to each community. Several operators are applied to migrate or exchange nodes between communities while using the modularity function as the objective function. An important feature of GGA+ is that it is able to rapidly obtain community partitions with different degrees of abstraction.

3.4. Empirical study

This section analyses the performance of the MIGA and GGA+ algorithms in detecting communities in several national-scale high voltage transmission networks with different characteristics. Neglecting complex electrical properties, the nodes of the network represent the power plants, distribution and transmission substations, while the edges correspond to transmission lines. In this way, the power grid is simplified as a unidirectional and unweighted network.

3.4.1. **Modularity**

Most optimisation methods apply modularity to detect communities in networks. Modularity [40] may be the most extensively applied objective function in community detection problems due to its simplicity and ease of calculation. Modularity provides a numerical value that represents the quality of the solution, such that the higher the value, the more accurate the community structure. Therefore, the aim of the algorithms is to maximize the value of modularity (Q), defined as:

$$Q = \frac{1}{2M}\sum\left(a_{ij} - \frac{K_i K_j}{2M}\right)\delta(i,j)$$

where *M* represents the total number of edges in the network; the sub-indices *i* and *j* indicate two nodes of the network; Ki and Kj are the degree of the i th and *j* th nodes, respectively; the parameter a_{ij} is the element of the *i* th row and the *j* th column of the adjacency matrix; and $\delta(i,j)$ represents the relationship between the *i* th node and the *j* th node, such that if node *i* and node *j* are in the same community, $\delta(i,j) = 1$; otherwise, $\delta(i,j) = 0$.

Therefore, taking into account the previous definition, the community detection problem consists of finding a network partition that maximizes modularity. This problem has been proven to be NP-hard [11, 72]. This is indeed the reason why heuristics and meta-heuristics are used [90].

3.4.2. **Test case**

To analyse the performance of the genetic algorithms, several case studies have been considered. On the one hand, it is used the IEEE 118 bus test case, which represents a portion of the American Electric Power System (in the

Midwestern US) as of December 1962. This network contains 118 buses, 186 branches, 54 thermal generators and 91 load sides. It has been selected to compare the performance of GGA+ with the node similarity index proposed in [73]. On the other hand, the graph models of a five national-scale power grids are considered.

Four are European power grids: Italy, including Sardinia and Sicily, Germany, the continental territory of France, and the Iberian Peninsula, including the Balearic Islands. The graph models were obtained from the Transmission System Map, which includes information about the transmission system network operated by members of the European Network of Transmission System Operators for Electricity (ENTSO-E) [91]. The ENTSO-E data is managed and displayed using Gephi software [92] that includes filters to extract only the nodes that belong to each country. The output is exported from Gephi to spreadsheet files (.csv) that include the nodes and the relationships between nodes (edges or links) of each country separately. These graphs include transmission lines designed for 220 kV voltage and higher and generation stations with a net generation capacity of more than 100 MW. Furthermore, it is also analyzed the graph model of the Texas power grid, the second largest state in the United States (US) by both area and population. It is well-known that America’s 48 contiguous states (and most of Canada’s population) receive the bulk of their electricity from three separate electric grids: the huge Eastern Interconnection, the Western Interconnection, and the relatively small Texas grid [93], which is almost entirely managed by the Electric Reliability Council of Texas (ERCOT). This network consists of a 2000 bus power system with electrical transmission lines of 345 kV, 115 kV and 13.8 kV. The synthetic network model was built based on the statistical analysis of real power systems and public information obtained from the Illinois Center for a Smarter Electric Grid (ICSEG) [94].

Table 3.1 describes some graph characteristics of these five high voltage power grids. The surface covered by these networks range between about 300,000 and 700,000 km^2, and the number of nodes and edges are almost proportional to the surface area. The dimensions of these networks are significantly larger than those of other power grids considered in recent studies (see e.g. [73]).

Table 3.1. Graph description of the power grids used.

Feature – Power Grid	IEEE 118	Italy	Germany	France	Iberian peninsula	Texas
Approx. Surface (km^2)	N/A	301,338	357,376	551,695	582,918	695,662
Nodes	118	352	438	904	1104	2007
Edges	177	462	662	1163	1416	2607
Average degree	3.03	2.63	3.03	2.57	2.57	2.60
Network diameter	14	39	21	28	40	39
Average path length	6.31	12.67	9.22	12.09	13.17	15.50
Average clustering coefficient	0.18	0.04	0.23	0.05	0.09	0.02

3.4.3. Parameter configuration

There is no a consensus for establishing an optimal population size, although a larger population size is often desirable to obtain high quality solutions. The selection of probabilities of applying crossover and mutation operators also depend of the problem at hand. However, a high crossover operator is generally preferable. To perform a fair comparison between the two genetic algorithms, the parameters were adjusted by means of a sensitivity analysis that consists in carrying out multiple runs of the algorithms with different combinations of population size and probabilities of applying the evolutionary operators: population size (50, 100, and 200 individuals), probability of crossover (50%, 80%, and 95%), and probability of mutation (10%, 20%, and 50%). To accomplish the sensitivity analysis, several independent runs of each of these 27 scenarios are executed. Table 3.2 shows the parameters in common and the model-specific parameters used by MIGA and GGA+, taking into account the results of the sensitivity analysis. The experiments were performed on a personal computer with an Intel Core i7 3630Q processor (2.4 GHz, 8 GB DDR3 RAM), which executes the application developed in C# .Net Framework 4. The computational time required to perform each independent run ranged from a few seconds for GGA+ to a few minutes in the case of the MIGA algorithm, since the latter uses Simulated Annealing in the search process.

Table 3.2. Parameter settings.

General parameters	
Population size	200
Iterations	200
Maximum iterations without improvement	50
Crossover probability	0.8
Mutation probability	0.2
Specific parameters (algorithm)	
MIGA	Initial temperature = 800000 Cooling rate =0.99 loop count l = 10
GGA+	Reproduction ratio = 0.2

3.4.4. Results and discussion

The accuracy of MIGA and GGA+ is evaluated according to the modularity (Q) values. Table 3.3 shows the mean and standard deviation values obtained by both algorithms after 50 independent runs when detecting from $k = 2$ to $k = 10$ communities. Based on these results, GGA+ achieves the best mean and standard deviation in all test instances independently of the number of communities to be detected. These results also indicate that the larger the problem instance, the greater the advantage of GGA+ with respect to MIGA. Moreover, the standard deviation (shown between parentheses) obtained from the results of these 50 independent runs is lower in GGA+ than in MIGA, which denotes the robustness of the results obtained by GGA+.

To compare the two algorithms, it is necessary to determine if there are significant differences between the results obtained by the different methods. With this aim, a one-way ANOVA has been applied, and the results indicate that the $p - value < 0.05$ in all cases, i.e., the null hypothesis is always rejected, indicating that there is a significant variation between at least some of the means of the different groups. Therefore, the results obtained by GGA+ are significantly different from those obtained by MIGA, supporting the conclusion based on the mean values displayed in Table 3.3.

Figure 3.1 shows the communities detected by MIGA and GGA+ in the Iberian Peninsula power grid. These results reveal the significant differences between the two algorithms, especially when the number of communities increases. Even when detecting two communities, MIGA has some difficulty in assigning communities in the central part of the graph, while GGA+ obtains clearly differentiated communities. Based on the data included in these figures and the results in Table 3.3, GGA+ not only outperforms MIGA, but it also exhibits good performance in these large networks.

Table 3.3. Mean (and standard deviation) of the modularity obtained by MIGA and GGA+ after 50 independent runs.

		Number of communities								
Test case	Method	2	3	4	5	6	7	8	9	10
Italy	MIGA	0,482 (0,002)	0,626 (0,005)	0,701 (0,005)	0,736 (0,005)	0,757 (0,007)	0,774 (0,005)	0,784 (0,004)	0,788 (0,004)	0,788 (0,003)
	GGA+	0,491 (0,000)	0,638 (0,000)	0,715 (0,001)	0,743 (0,002)	0,770 (0,005)	0,787 (0,001)	0,798 (0,001)	0,806 (0,002)	0,810 (0,002)
Germany	MIGA	0,478 (0,003)	0,637 (0,005)	0,707 (0,004)	0,737 (0,005)	0,759 (0,007)	0,769 (0,006)	0,781 (0,006)	0,789 (0,005)	0,793 (0,004)
	GGA+	0,485 (0,001)	0,643 (0,001)	0,714 (0,001)	0,754 (0,003)	0,784 (0,002)	0,800 (0,001)	0,810 (0,001)	0,814 (0,003)	0,817 (0,002)
France	MIGA	0,477 (0,002)	0,636 (0,003)	0,709 (0,004)	0,750 (0,003)	0,776 (0,005)	0,795 (0,005)	0,804 (0,004)	0,811 (0,003)	0,818 (0,003)
	GGA+	0,486 (0,001)	0,645 (0,001)	0,720 (0,001)	0,765 (0,002)	0,794 (0,003)	0,818 (0,001)	0,830 (0,001)	0,838 (0,002)	0,842 (0,002)
Iberian Peninsula	MIGA	0,480 (0,002)	0,637 (0,006)	0,712 (0,003)	0,758 (0,005)	0,787 (0,005)	0,812 (0,003)	0,823 (0,003)	0,828 (0,002)	0,832 (0,003)
	GGA+	0,489 (0,001)	0,650 (0,001)	0,727 (0,002)	0,773 (0,001)	0,804 (0,001)	0,829 (0,001)	0,836 (0,002)	0,843 (0,002)	0,851 (0,002)
Texas	MIGA	0,485 (0,002)	0,635 (0,005)	0,704 (0,004)	0,748 (0,004)	0,772 (0,004)	0,789 (0,003)	0,804 (0,004)	0,814 (0,003)	0,826 (0,003)
	GGA+	0,493 (0,001)	0,654 (0,001)	0,731 (0,003)	0,776 (0,002)	0,804 (0,003)	0,823 (0,002)	0,838 (0,004)	0,851 (0,002)	0,858 (0,003)

The results obtained by GGA+ are analyzed in more detail. Figure 3.2(a) displays the communities detected by GGA+ in the Iberian Peninsula power grid with different levels of detail. These data reveal that this algorithm is able to obtain quality solutions even when the number of communities increases. Moreover, Figure 3.2(b) provides a different layout based on the Force Atlas2 [95] plugin in Gephi. This visualisation method builds a force directed layout by simulating a physical system in order to accommodate nodes and links in a spatial network. Nodes repel each other like charged particles, while edges attract their nodes like springs. The aim of this method is to help construct a balanced state network that facilitates data interpretation. The analysis of Figure 3.2 demonstrates the good performance of GGA+ independently of the degree of abstraction.

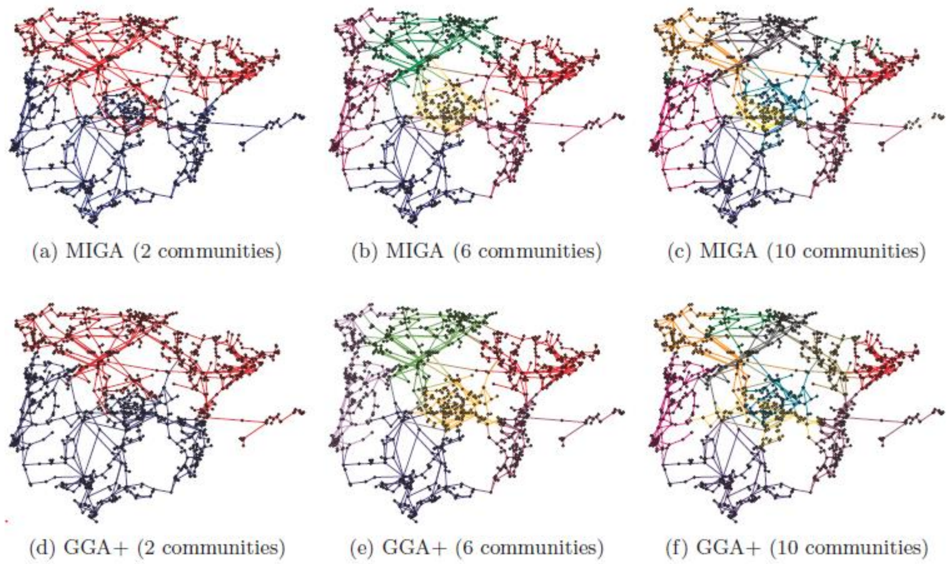
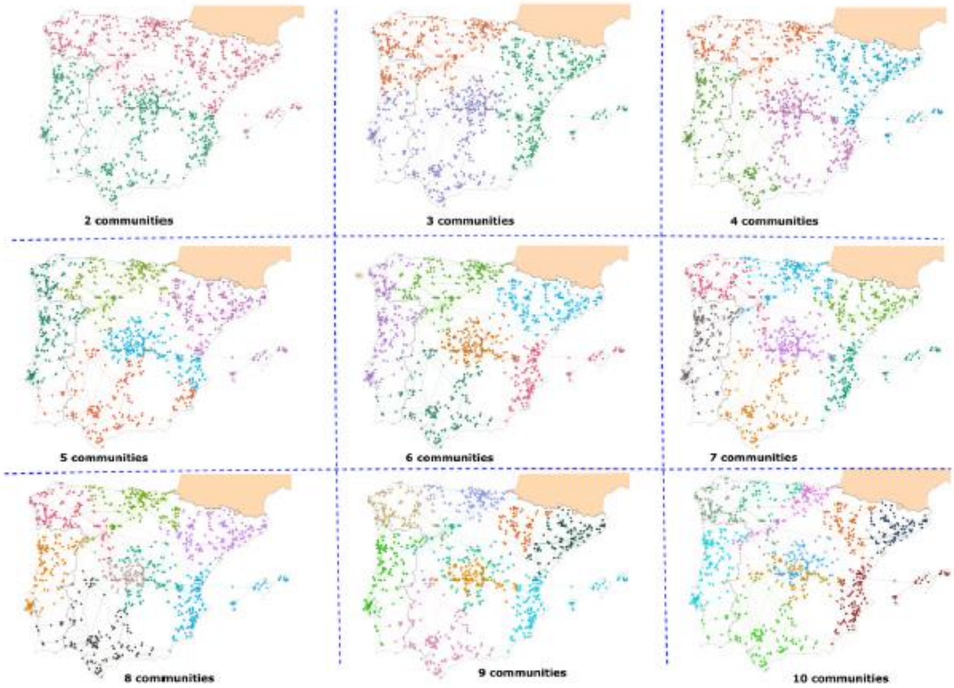
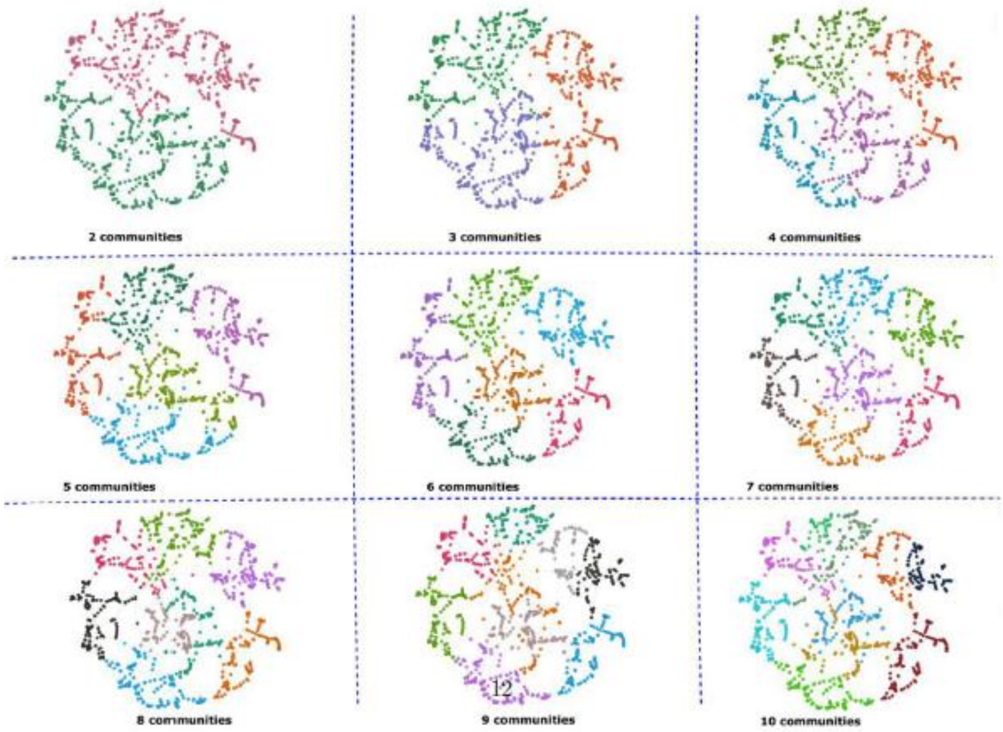


Figure 3.1. Results obtained by MIGA and GGA+ in the Iberian Peninsula network considering k=2, k=6 and k=10 communities.

Figure 3.3 displays the communities detected by GGA+ in the networks representing the power grids of Italy, Germany and France with 3 and 8 communities, respectively. These results clearly show that independently of the number of communities, GGA+ detects clearly differentiated communities. In the case of the Italian network (Figure 3.3(a)), the solution provided by GGA+ clearly divides the Italian power grid into three communities: north, central and south. The analysis of the communities detected in the German power grid (Figure 3.3(b)) also provides interesting results. In particular, the structure of three communities obtained by GGA+ is coincident with the historical division of Germany during the Cold War, with one community covering a large area in



(a) Physical layout.



(b) Layout using Force Atlas 2.

Figure 3.2. Results obtained by GGA+ in the Iberian Peninsula and Balearic Islands considering from k = 2 to k = 10 communities.

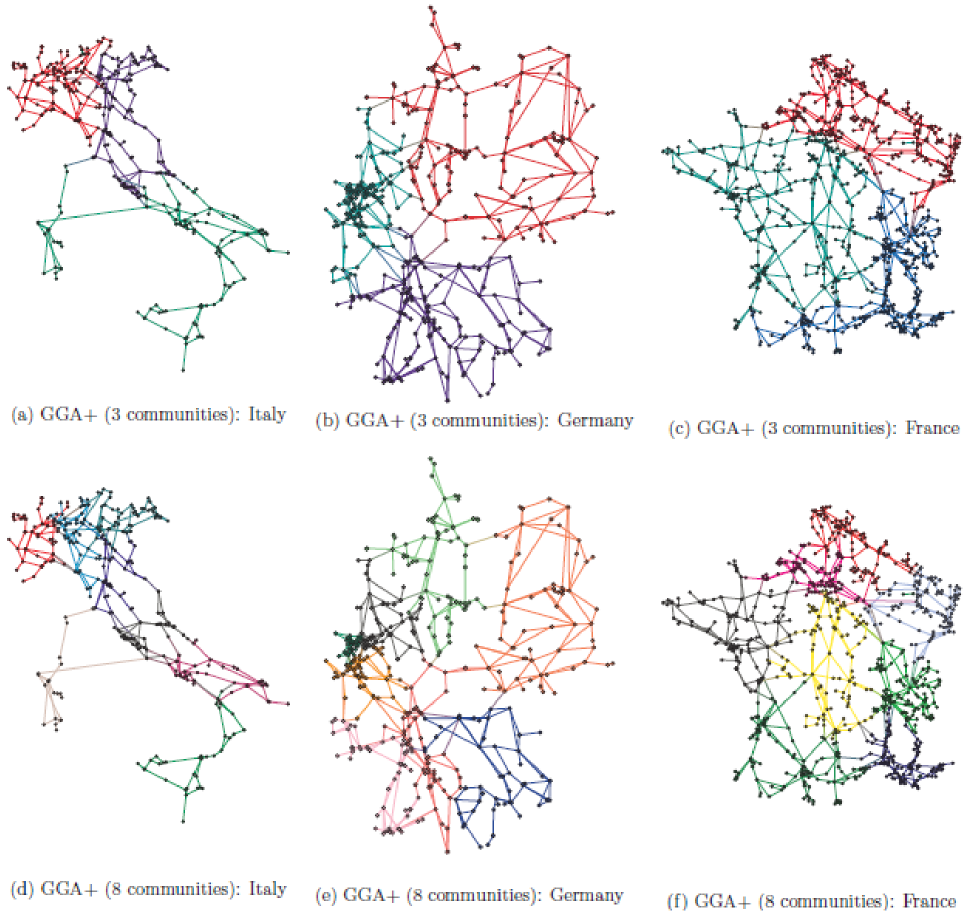


Figure 3.3. Results obtained by GAA +in the Italian, German and French power grids considering k = 3 and k = 8 communities.

East Germany and two communities in West Germany covering the north-central area (British and French sectors) and the south area (American sector).

These results coincide with previous studies that have highlighted that the evolution of electric systems in Europe was largely affected by political conflicts such as the two world wars, the Cold War and the political separation of Eastern and Western Europe [96]. The network representing the French power grid exhibits a homogeneous geographical distribution of high voltage transmission lines, except for the central zone, where the density is slightly lower than in the south-east and north. When the network is divided into three communities (Figure 3.3(c)), they are of similar size, although the north-east and southeast communities cover a smaller area than the central community. When these three networks are divided into eight communities (Figure 3.3(d), (e), and (f)) the connected regions representing each community are of similar size.

Figure 3.4 presents some solutions obtained by GGA+ in the Texas network with different levels of detail (3, 4, 5, and 6 communities). A significant concentration of substations and electrical lines is observed around the major cities in Texas, i.e., Houston, San Antonio, Dallas, and Austin. In fact, Figure 3.4 shows that when GGA+ obtains a set of five communities, four communities include these important metropolitan cities, while the fifth region covers the sparsely populated west region of Texas.

In addition to the comparison using the five previous networks, Figure 3.5 compares the results obtained by GGA+ with those obtained by the method recently proposed by [73] in the IEEE 118 bus network. Figure 3.5(a), shows the one-line diagram of IEEE 118-bus test system including the communities detected by [73], while Figure 3.5(b) shows the communities detected by GGA+. Figure 3.5(c) displays the graph with the communities detected by [73], which obtains a modularity value of 0.689, while the modularity obtained by GGA+ (Figure 3.5(d)) is 0.726. Figure 3.5(e) shows the difference between both solutions, such that those nodes (buses) vary from community between the solutions are highlighted with red color with a bigger size. Therefore, it has also been proved that the genetic algorithm outperforms the method proposed by [73].

3.5. Implications for power grid design and operation

The results obtained in the previous section show how genetic algorithms are capable of determining communities in graphs that represent the power grids on a national scale. In addition to information about the topological characteristics of these systems, understanding their community structure can be valuable information to consider in the planning, design and operation of energy systems.

Contingency analysis is an important tool in power systems planning and design stages to evaluate their security. A contingency can be defined as an unplanned outage due to the loss or failure of a small part of the power system (e.g. a transmission line). Since many years ago, power systems engineers guarantee the security of the system, planning studies are able to withstand sudden disturbances such as electric short circuits or unanticipated loss of system elements [97]. The (n-1)-security criterion is verified by analyzing power flows considering one component failure (e.g.: a transmission line is out of service) and determining how this contingency would affect the network operation.

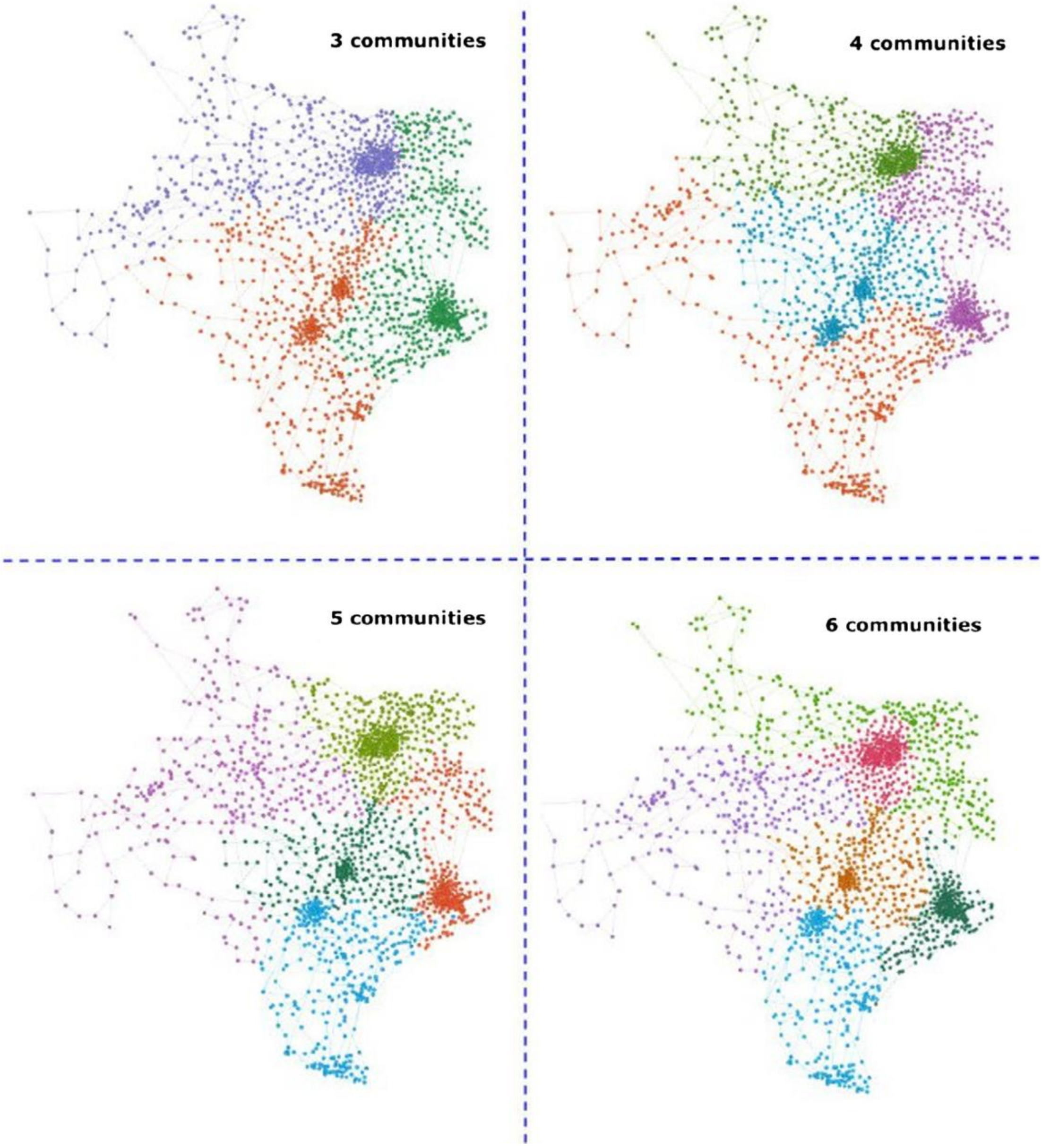


Figure 3.4. Results obtained by GGA+ in the Texas network considering 3, 4, 5 and 6 communities.

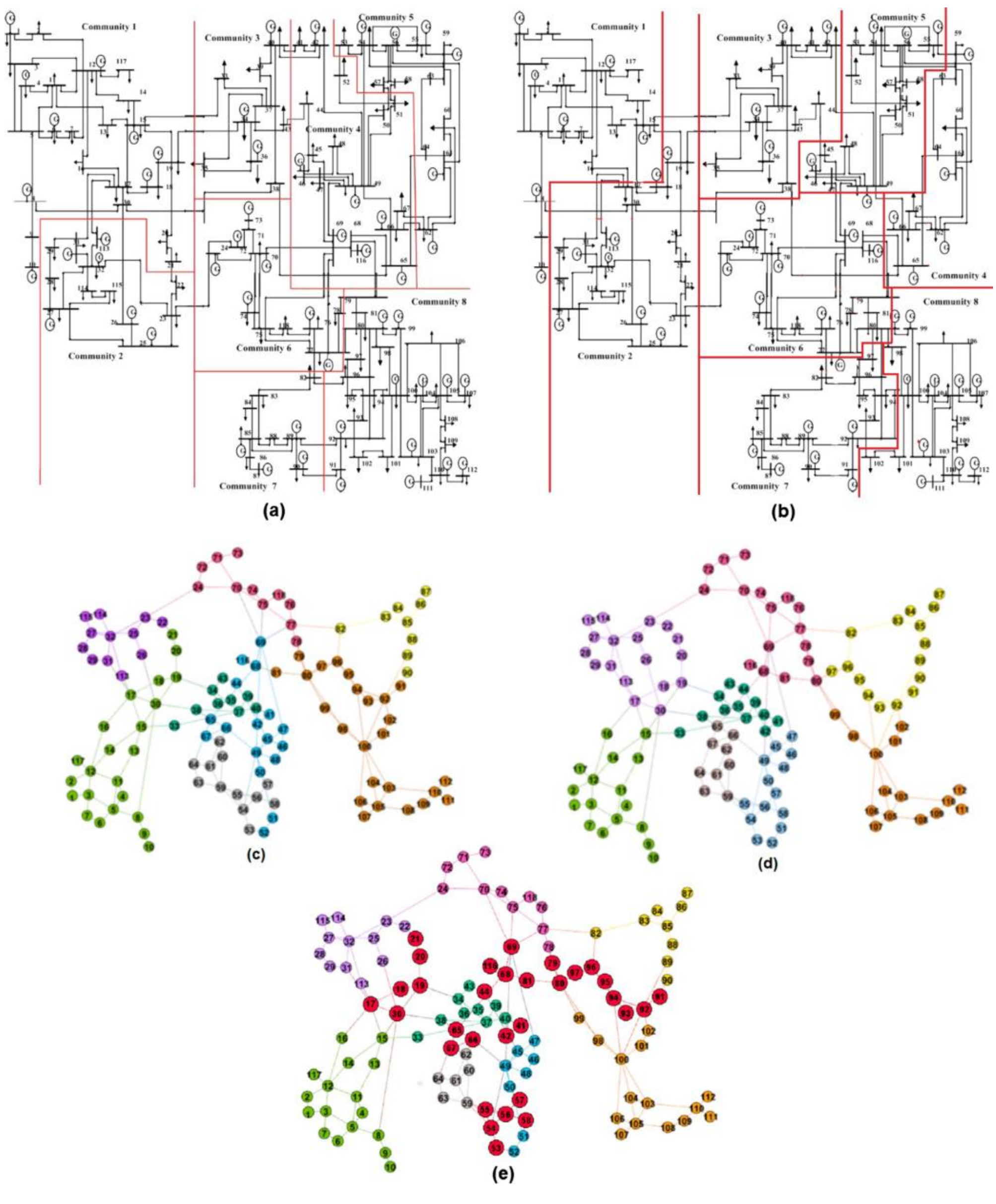


Figure 3.5. One-line diagram of the IEEE 118-bus test system with eight communities detected by [73] (a) and GGA+ (b). Graph with the nodes of each community detected by [73] (c) and GGA+ (d), and its differences (e).

Table 3.4 shows some results obtained from a (n-1)-security contingency analysis in the network IEEE 118. Since this system has 186 branches, the (n-1)-security contingency analysis requires to calculate 186 power flows (in each of these executions one branch is disabled). With this aim, a Matlab power system simulation package (Matpower) is used. Table 3.4 shows the five most critical transmission lines (branches) in terms of active power losses (Ploss). When these results are compared with the communities detected by GGA+, it is observed an interesting pattern: these branches of the power grid which outage provoke higher total active power losses are often edges that connect nodes of different communities obtained by the genetic algorithm. Therefore, as future work, it is planned to perform an exhaustive study to confirm that the detection of communities in power grids would provide useful information to those power systems engineers that perform contingency analyses.

On the other hand, it is important to remark that the operation of power systems has changed significantly in the last years. Traditional electric power distribution systems have been designed on the assumption that the sources of power are the primary substations close to conventional power plants [98]. This assumption is invalidated by the entrance of distributed generation, which allows to generate electricity in plants that are connected to a distribution network rather than the transmission network. The application of distributed generation has considerably increased in last years and it is expected to be higher in the near future due to the emerging utilization of renewable energy resources, including solar panels, micro wind turbines, and combined heat and power units. This tendency has led to a demand for a new electricity distribution paradigm, such that there is an increasing interest in the concepts of smart grids and microgrids. A microgrid is a cluster of both distributed generators and loads which act to cooperate with the main grid or autonomously from it [99]. In case of an upstream network disturbance, the microgrid is able to disconnect from the main grid then operating as a self-controlled entity. With this aim, several authors have proposed methods for island partitioning as a solution to rapidly restore the energy supply of important loads, and to reduce the outage time [98]. But it is not an easy task, since large power imbalance in the island can lead to frequency instability, then resulting in a desenergization of part or all of the system [100]. Therefore, as future work, it will be considered the use of community detection as a partitioning strategy to implement islanding strategies in power systems.

Table 3.4. Critical transmission lines -sorted from highest to lowest active power losses (Ploss) due to failure in one transmission line (branch)- that are boundary edges between different communities -graph obtained by GGA+ considering from k=2 to k=20 communities-.

POWER FLOW ANALYSIS (n-1)-security criterion					COMMUNITY DETECTION ANALYSIS
Branch (failure)	From bus	To bus	Ploss (MW)	Qloss (MVAr)	Is boundary line in the community detected?
8	8	5	197,03	1136,16	YES (k = {10, 11, 12, 13, 14, 15, 16, 17,18, 19, 20})
51	38	37	169,57	919,93	YES (k = {18, 19, 20})
96	38	65	167,84	906,67	YES (k = {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20})
38	26	30	165,90	936,08	YES (k = {2, 4})
36	30	17	150,14	859,82	YES (k = {18, 19, 20})

3.6. Conclusions

Power systems design is a critical engineering task that aims to deliver electric energy from power plants to network loads with high reliability, while the system is operated economically at maximum efficiency. However, these systems are commonly affected by contingencies due to uncontrollable technical issues.

This chapter analyses the performance of two efficient genetic algorithms (MIGA and GGA+) for solving community detection problems in national-scale high voltage transmission networks. These approaches use powerful initialisation methods and evolutionary search operators under the guidance of modularity, and are able to obtain good quality solutions in different networks. These algorithms enable a flexible and adaptive analysis of the characteristics of power grids with different levels of detail (number of communities). The empirical study considers six test cases representing the national-scale high voltage power grids of Italy, Germany, the continental territory of France, the Iberian peninsula (Portugal and Spain, including the Balearic islands), Texas (US), and the IEEE 118 bus test case that represents a portion of the American Electric Power System (in the Midwestern). Despite several studies have previously applied community detection to networks representing power grids, to our knowledge, it is the first study that has empirically analyzed the performance of evolutionary methods in network representing national-scale power grids. Moreover, this study is particularly interesting if we take into account that the development of the power grid in the U.S. was organized according to federal and state regulation, while the power grid in Europe has traditionally organized country by country.

These results are useful in showing that genetic algorithms are fast and powerful tools to detect communities in national-scale high voltage transmission networks, and they also provide interesting topological information about the physical distribution and concentration of these grids. In addition, this new knowledge is useful for engineering decision makers since it provides information about the critical elements (lines in our case) which damage may trigger cascades of failures across the grid and lead to a large blackout. Future work will be focused to study with more detail how partitions obtained with community detection algorithms could provide useful information for contingency analysis and islanding strategies. Moreover, it is expected to apply parallel and multi-objective optimisation methods to solve this problem in national and continental-scale power grids.

Capítulo 4

Evolutionary Algorithms for Community Detection in Continental-Scale High-Voltage Transmission Grids

In order to investigate the topological characteristics of power grids, this chapter use of evolutionary algorithms for community detection using modularity density measures on networks representing supergrids in order to discover densely connected structures. Two evolutionary approaches (generational genetic algorithm, GGA+, and modularity and improved genetic algorithm, MIGA) were applied. The results obtained in two large networks representing supergrids (European grid and North American grid) provide insights on both the structure of the supergrid and the topological differences between different regions. Numerical and graphical results show how these evolutionary approaches clearly outperform to the well-known Louvain modularity method. In particular, the average value of modularity obtained by GGA+ in the European grid was 0.815, while an average of 0.827 was reached in the North American grid. These results outperform those obtained by MIGA and Louvain methods (0.801 and 0.766 in the European grid and 0.813 and 0.798 in the North American grid, respectively).

4.1. Introduction

The optimal design and management of these supergrids is a difficult task, since it is necessary to manage large systems that include heterogeneous power grids from different countries. Most investigations in power systems often analyse optimisation problems such as optimal power flow, unit commitment, and economic dispatch, among others [101, 102]. The solutions to these problems are often determined by the symmetry of the admittance and Jacobian matrices [103, 104], and the topology of high-voltage transmission lines that connect the power produced at generating stations to substations, at which point the power flow is derived to other transmission lines or stepped down in voltage, and then submitted across power distribution lines into the end users. Many publications have addressed the factors that constrain the development of electricity infrastructure [105, 106]. In particular, experts have highlighted that existing electric grids are inadequate to cope with increasing volumes of renewable electricity [107]. For example, the transmission systems in European countries are old, and a many miles of lines need to be replaced, upgraded, and even expanded to secure market integration, ensure supply security, and cope with the expansion in renewable energy planned for the next few years [108]. A similar challenge is faced in the United States, where renewable energy generation also accounts for an increasingly high percentage of annual demand [109].

Taking into account the fact that worldwide demand for electricity has been increasing and will continue to, it is necessary to ensure the reliable and secure operation of electricity transmission networks to efficiently transport energy from generation sources to electricity consumers. To achieve this goal, decisions need to be supported by expert systems able to process a large number of variables. Graph-based network analysis is a powerful tool for describing many real systems in a variety of fields [68]. Topological analysis provides the infrastructural information of power systems that is essential to assess network robustness or to generate synthetic power grids [110]. For example, some studies have detected complex symmetric subgraphs in large-scale power grids [111], and have provided a list of symmetric subgraphs with respect to reference nodes observed in the US grid.

Most real networks (graphs) representing real systems have clusters, such that many edges connect nodes within the same cluster, and comparatively few edges connect nodes in different clusters. This is why community detection [10, 11, 69] has gained popularity in recent years, especially among researchers working with complex networks [112, 113, 114]. In particular, community detection has been applied in field of electrical engineering, including the management of power grids [15, 58, 74, 75]. However, keeping the complexity of the problem in mind, more work is needed to develop efficient algorithms to enable rapid community detection. With that aim, this chapter evaluates the performance of evolutionary approaches for community detection in supergrids. These algorithms, which are guided by the modularity index [40] and consider different *degrees of abstraction* (i.e., detect any number of communities), enable a flexible and adaptive analysis of the power grid.

The remainder of the chapter is organized as follows: Section 4.2 briefly describes the problem of community detection using graphs, and revises some previous studies that have been applied to electrical grids. Section 4.3 presents the main characteristics of two evolutionary algorithms used to solve the community detection problem using graphs [66]. Section 4.4 presents an empirical study that compares these methods for community detection in two supergrids. The conclusions of this work are provided in Section 4.5.

4.2. Community Detection

This section introduces the use of community detection in different research areas and discusses how community detection methods contribute to the analysis of power grids.

4.2.1. Community Detection: General Overview

Communities, also named clusters, are dense subgraphs which are well separated from each other. The community structure of complex networks reveals both their organisation and hidden relationships, among other elements [115]. In practice, a simple idea that has attained great popularity is that a community is a subgraph such that many edges connect nodes within the same group, and comparatively few edges connect nodes in different groups [69].

Many studies in different disciplines have shown that the community structure of complex networks reveals both their organisation and hidden relationships among their elements [115]. In particular, identifying communities can be useful for classifying the nodes in different groups [11]. So, nodes located at a central position in their community may have an important function of control and stability within the cluster, while those nodes located at the proximity of other communities can play a role of mediation or information exchange with these neighbouring communities.

An important consideration to be determined here is the number of communities to be detected. Some algorithms allow one to include a pre-established number of communities to be detected, while other approaches aim to infer the adequate number of communities depending of the characteristics of the networks [69]. A recently published survey paper [24] has reviewed a large number of community detection algorithms in multidisciplinary applications considering both disjoint and overlapping community detection problems. These applications include the study of social networks [115, 116], communication networks [117, 118], engineering systems and networks [114, 119] biology and ecology [120, 121], health sciences [122], scientometrics [123, 124], economics [125], etc.

4.2.2. **Community Detection in Power Grids**

In recent years, the interest in the development of supergrids has grown remarkably. The supergrid concept was born as a solution to allow large-scale electrical power exchanges over continent-wide areas. This concept has been considered both a potential solution to transmission bottlenecks and an opportunity to trade higher volumes of electricity across longer distances [126]. In particular, we show the complexity of several high-voltage transmission topologies intended to connect two or more subsystems here, and note that supergrids have a meshed form to provide redundancy. In addition to the use of complex control methods [127], the variability of renewable sources [128] at continental scales can be mitigated by using the transmission grid and balancing locally with storage [79]. Some of the future major transmission projects around the world are described in [126]. For example, different projects aim to promote an efficient and reliable transmission grid in North America, including the *Tres Amigas superstation*.

This superstation is the first version of this supergrid vision, since it is projected as a high-voltage direct current (HVDC) super-node asynchronously connecting the existing alternating current (AC) networks intended to link the three North American grids: the Eastern Interconnection, the Western Interconnection, and Texas Interconnection. This project involves a three-way alternating current/direct current (AC/DC) transmission superstation with several miles of underground superconducting DC cable, which will eliminate the market separation between the three asynchronous interconnections in the continental U.S. [129]. In the case of Europe, these authors indicate that an important number of major HVDC interconnections are being promoted to establish intercontinental interconnections with neighbouring regions with the aim of integrating regional energy markets into a single European market to achieve the European Union's (EU) renewable energy goals. Some authors have introduced the concept of global grid as the future stage of the electricity network, in which most of the large power plants in the world will be connected [130].

Some recent studies have proposed the analysis of the power grid infrastructure using graph-based network analysis techniques [74]. Usually, the nodes of the network represent the power plants and distribution and transmission substations, while the edges correspond to transmission lines. The application of graph-based analysis techniques has allowed for an analysis of the topological structure of networks representing power grids [80]. As commented above, a typical characteristic of all complex networks is the existence of community structures [10, 11], such that detecting those communities can reveal the characteristics or functional relationships in a given network. In the case of power grids, communities represent substations densely connected by high-voltage transmission lines.

The importance of community detection in power grids comes from the fact that it is necessary to maintain grid reliability and enable more efficient restoration from severe disturbances. In particular, it is necessary to prepare a distribution grid for natural disasters (e.g., a storm), by developing switching plans to safely islands or disconnecting portions of the grid, preventing further degradation during incidents and enabling faster restoration after the disturbance. For example, reference [58] applied community detection to island power systems as an emergency response method to isolate failures that could propagate and lead to major disturbances. These authors developed two approaches based on modularity, with the DC power flow model incorporated into them, for islanding in medium and large networks and tested them in networks having 14, 30, 57, 118, and 247 nodes [58]. Other approaches use node similarity indexes to assign each node to the community sharing maximum similarity [73], and have demonstrated the good performance of this method in two IEEE standard power grids (39-bus standard power grid and 118-bus standard power grid). The IEEE 118-bus was also studied in [131]. Other researchers have presented a hierarchical spectral clustering method to reveal the internal connectivity of power transmission, establishing the possibility of islanding systems using a network with nodes and links representing buses and electrical transmission lines, respectively [75]. That approach was evaluated in several test systems of small, medium, and large sizes, including a model of Great Britain's transmission network [75]. Community detection has also been applied to analyse the vulnerability of the power systems under terrorist attacks [83], among other applications. However, none of these previous approaches have analysed supergrids.

4.3. Methodology

Two genetic algorithms designed to detect communities in graphs were applied in large networks modelling supergrids. These evolutionary approaches are guided by *modularity* [40], which is an applied objective function extensively used in community detection due to its simplicity and ease of calculation. Modularity provides a numerical value that represents the quality of the solution, with greater values corresponding to a more accurate community structure. Therefore, the aim was to find communities that maximise the value of *modularity* (*Q*), defined in section 3.4.1.

Finding these communities by maximising the modularity [40] or another objective function is an NP-hard problem [72]. Brandes et al. [132] proved that modularity maximisation is an NP-hard problem, even for the restricted version with a bond of two clusters, and suggested further investigation of approximation algorithms and heuristics for solving this problem. More recently, other authors have demonstrated the high complexity of calculating modularity on sparse graphs and dense graphs separately [133]. Due to the high complexity of the community detection problem, researchers have applied heuristic and metaheuristic methods to obtain high quality solutions, in a reasonable computational time.

The field of evolutionary computation [32] is closely related to computational intelligence, with a focus on designing algorithms to solve complex global optimisation problems. Evolutionary algorithms are problem-solving procedures that include evolutionary processes as the key design elements, such that a population of individuals is continually and selectively evolved until a termination criteria is fulfilled. Genetic algorithms (GAs) [19] are probably the most widely used evolutionary techniques. A genetic algorithm mimics natural selection by evolving, over time, a population of individual solutions to the problem at hand until a termination condition is fulfilled and the best individual is taken as an acceptable solution. Two important characteristics of GAs are the representation used (e.g., binary or real) and the genetic operators employed (e.g., mutation and crossover). GAs have been successfully applied to solve electrical problems [86, 87, 88].

In this study, two genetic algorithms were adapted to solve community detection problems in power grids. These algorithms (MIGA and GGA+) were recently proposed and shown to be more effective than other approaches to community detection, as assessed by benchmarks typically used to compare algorithms. Figure 4.1 shows the flowchart of MIGA and GGA+. The main characteristics of both methods are briefly described below.

- The modularity and improved genetic algorithm (MIGA) [45] takes the modularity (Q) as the objective function, and uses the number of community structures as prior information to improve stability and accuracy of community detection. MIGA also uses simulated annealing [14] as local search strategy. Note that many authors have previously considered the use of local search strategies [134] for solving hard optimisation problems.

- The generational genetic algorithm (GGA+): GGA+ [89] includes efficient and safe initialisation methods in which a maximum node size is assigned to each community. Several operators are applied to migrate or exchange nodes between communities while using the modularity function as the objective function. An important feature of GGA+ is that it is able to rapidly obtain community partitions with different degrees of abstraction.

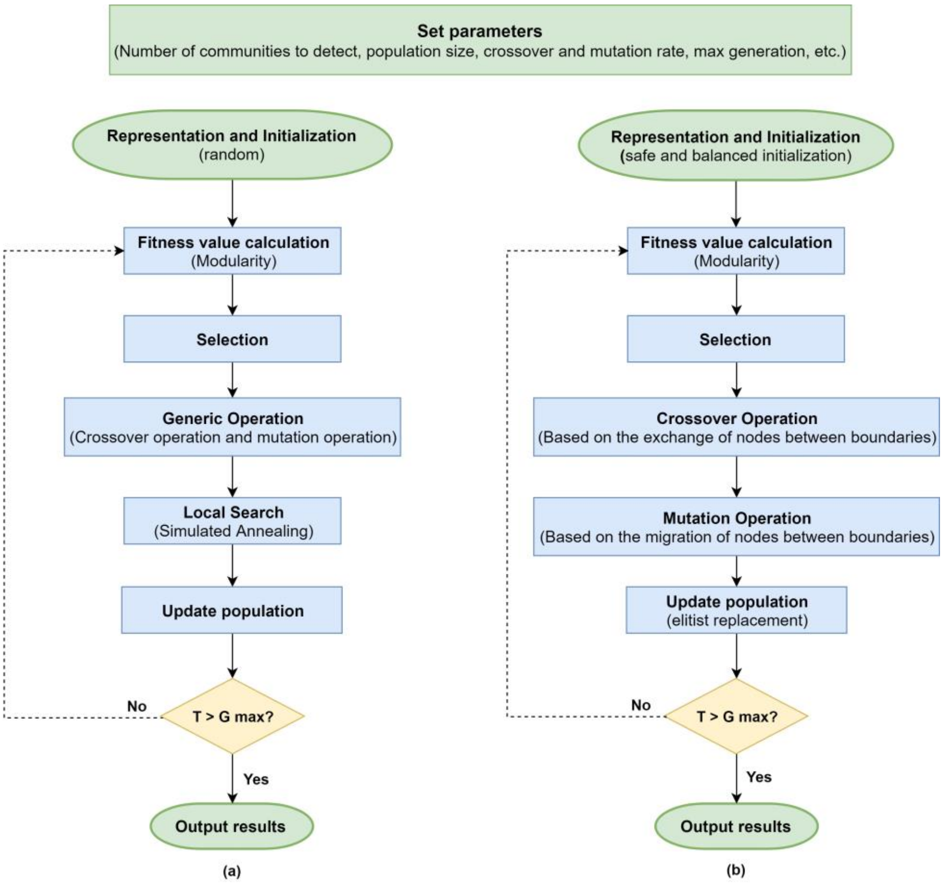


Figure 4.1. Flowchart of (a) MIGA and (b) GGA+.

4.4. Empirical Study

This section analyses the performance of the MIGA and GGA+ algorithms in detecting communities in two networks representing supergrids. Neglecting complex electrical properties, the nodes of the network represent the power plants and distribution and transmission substations, while the edges correspond to transmission lines. In this way, the power grid is simplified as an undirected and unweighted network.

4.4.1. Test Cases

To analyse the performance of the genetic algorithms, graphical models of two supergrids were considered: the European grid, including part of Russia, North Africa, and part of the Near East; and the North American grid.

- The European electric network was obtained from the European Network of Transmission System Operators (ENTSO-E) [91]. The ENTSO-E group consists of 43 electricity transmission system operators from 36 countries across Europe who are responsible for the bulk transmission of electric power on the main high-voltage electric networks. This power network, which also includes data from North African and Near Eastern countries, is formed by transmission lines designedfor 220 kV voltage and higher and generation stations with a net generation capacity of more than 100 MW.
- The North American electric network was obtained using the GridKit 1.0 toolkit, which was developed in the context of the SciGRID project at the NEXT ENERGY-EWE Research Centre for Energy Technology [135]. GridKit is a power grid extraction tool that converts geographical objects representing elements of power systems in OpenStreetMap to model the

electric network. This network covers the United States, Canada, and Mexico and includes transmission lines that operate at relatively high-voltages varying from 69 kV up to 765 kV. The power grid of the United States is probably the best system studied in the literature, due to the particular characteristics of the network. The regions covered receive their bulk electricity from three separate electric grids: the huge Eastern Interconnection, the Western Interconnection, and the relatively small Texas grid [93], which is almost entirely managed by the Electric Reliability Council of Texas (ERCOT).

Table 4.1 describes some graphical characteristics of these networks. The number of nodes and edges is very large, which denotes the high complexity of community detection in these networks. Note that the dimensions of these networks are significantly larger than other power grids considered in recent studies (see, e.g., [73]). In fact, to our knowledge, no previous paper has applied graph-based analytical methods to power networks of these dimensions.

Table 4.1. Graph description of the European (EU) and North American (NA) power grids.

Feature	EU Grid	NA Grid
Nodes	7893	16,063
Edges	10,346	20,169
Average degree	2.62	2.51
Network diameter	108	158
Avg. path length	41.27	48.99
Avg. clustering coefficient	0.07	0.01
Eigenvector centrality	1.70	1.87

4.4.2. **Parameter Configuration**

To perform a fair comparison between the two evolutionary algorithms, the parameters were set to the values established in the original publications. The influence of GGA+ parameters was adjusted by means of a sensitivity analysis method based on executing the algorithms with different number of individuals and probabilities of using evolutionary search operators [131]. Statistical results obtained from these independent runs were considered to select the following parameters: the population size was set to 200 individuals; the number of iterations (generations) of the algorithm was set to 200; and the probability of applying the search operators was set between 20% and 80%. As commented above, MIGA also uses simulated annealing [14] as a local search strategy, with the following parameters: initial temperature 800,000, cooling rate 0.99, and minimum temperature 0.01. The experiments were performed on a personal computer with an Intel Core i7 3630Q processor (2.4 GHz, 8 GB DDR3 RAM), which executed the application we developed in the C# .Net Framework 4.

4.4.3. **Results and Discussion**

The accuracies of MIGA and GGA+ were evaluated according to the Q values. Table 4.2 shows the maxima, means, minima, and standard deviations (SD) of the modularities obtained by MIGA and GGA+ in the European (EA) and the North American (NA) grids considering c = {2, 3, 4, 5, 10, 20, 30, 40, 50} communities. A number of communities within the range of 2 to 50 were used to show how evolutionary algorithms are able to work under different levels of abstraction. However, these algorithms could be applied to obtain a greater number of communities, although their size would decrease considerably. The median runtimes (in minutes) of these 30 independent runs are also provided.

Furthermore, the communities detected by Louvain modularity method implemented in Gephi are also shown in this table. To conduct the performance analysis and to statistically compare the quality of the solutions obtained by the two algorithms, a total of 30 independent runs were performed with each algorithm on each network.

Table 4.2. Results obtained by MIGA and GGA+ after 30 independent runs and comparison with Louvain method implemented in Gephi (runtime in minutes).

Grid	Method	Metric	c = 2	c = 3	c = 4	c = 5	c = 10	c = 20	c = 30	c = 40	c = 50	AVG
EU	MIGA	Max(Q)	0.496	0.660	0.740	0.785	0.877	0.911	0.921	0.924	0.922	0.804
		Mean(Q)	0.495	0.654	0.736	0.781	0.872	0.909	0.918	0.922	0.921	0.801
		Min(Q)	0.491	0.612	0.730	0.777	0.866	0.907	0.916	0.920	0.919	0.793
		SD(Q)	0.001	0.009	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.002
		Mean time	515	632	404	409	289	296	205	237	246	359
	GGA+	Max(Q)	0.498	0.665	0.746	0.793	0.889	0.929	0.941	0.943	0.948	0.817
		Mean(Q)	0.498	0.663	0.744	0.793	0.887	0.927	0.938	0.942	0.947	0.815
		Min(Q)	0.496	0.661	0.742	0.791	0.885	0.925	0.937	0.940	0.946	0.814
		SD(Q)	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.001
		Mean time	661	614	556	417	443	462	314	228	244	438
	Louvain	(Q)	0.291	0.599	0.677	0.699	0.874	0.929	0.939	0.944	0.945	0.766
NA	MIGA	Max(Q)	0.498	0.663	0.743	0.788	0.882	0.924	0.939	0.945	0.948	0.814
		Mean(Q)	0.498	0.659	0.740	0.789	0.878	0.922	0.937	0.944	0.947	0.813
		Min(Q)	0.497	0.656	0.736	0.785	0.872	0.921	0.936	0.943	0.946	0.810
		SD(Q)	0.000	0.002	0.002	0.002	0.003	0.001	0.001	0.000	0.000	0.001
		Mean time	1228	770	708	747	425	317	423	559	244	602
	GGA+	Max(Q)	0.499	0.670	0.753	0.804	0.900	0.943	0.956	0.963	0.967	0.828
		Mean(Q)	0.499	0.669	0.753	0.802	0.898	0.940	0.955	0.962	0.966	0.827
		Min(Q)	0.498	0.667	0.751	0.801	0.893	0.936	0.954	0.961	0.965	0.825
		SD(Q)	0.000	0.001	0.001	0.001	0.001	0.002	0.001	0.000	0.000	0.001
		Mean time	1909	1335	1221	568	521	764	593	552	298	862
	Louvain	(Q)	0.478	0.608	0.652	0.754	0.881	0.935	0.952	0.959	0.963	0.798

These results show that GGA+ achieved the best mean and maximum values in both grids, regardless of the number of communities to be detected. These results also indicate that the greater the number of communities, the greater the advantage of GGA+ over MIGA. In addition, the standard deviation obtained from the results of these 30 independent runs was often smaller for GGA+ than for MIGA. The modularity values obtained by both algorithms increased with the number of communities without degradation of the standard deviation, indicating the robustness of these evolutionary approaches.

Table 4.2 shows that the runtime required by both evolutionary algorithms is of the same order of magnitude in the North American and European networks, while the differences come from the fact that the former has approximately double the number of nodes and edges as the latter (see Table 4.1). On the other hand, in both cases the runtime tends to decrease when the number of communities is greater. This is due to the crossover and mutation operators moving a given percentage of the nodes between a community and a neighbouring one. Therefore, the bigger the communities are, the higher that number of nodes that are moved between neighbouring communities, and therefore, the runtime increases. It can be concluded that GGA+ is scalable both in terms of network size and in terms of number of communities. Both algorithms require a few hours to complete the search process with these parameter settings, which is not a critical issue since the goal is to find solutions with greater modularity regardless of the execution time. Of course, the execution time could be reduced considerably by modifying the parameter settings or applying parallel processing techniques.

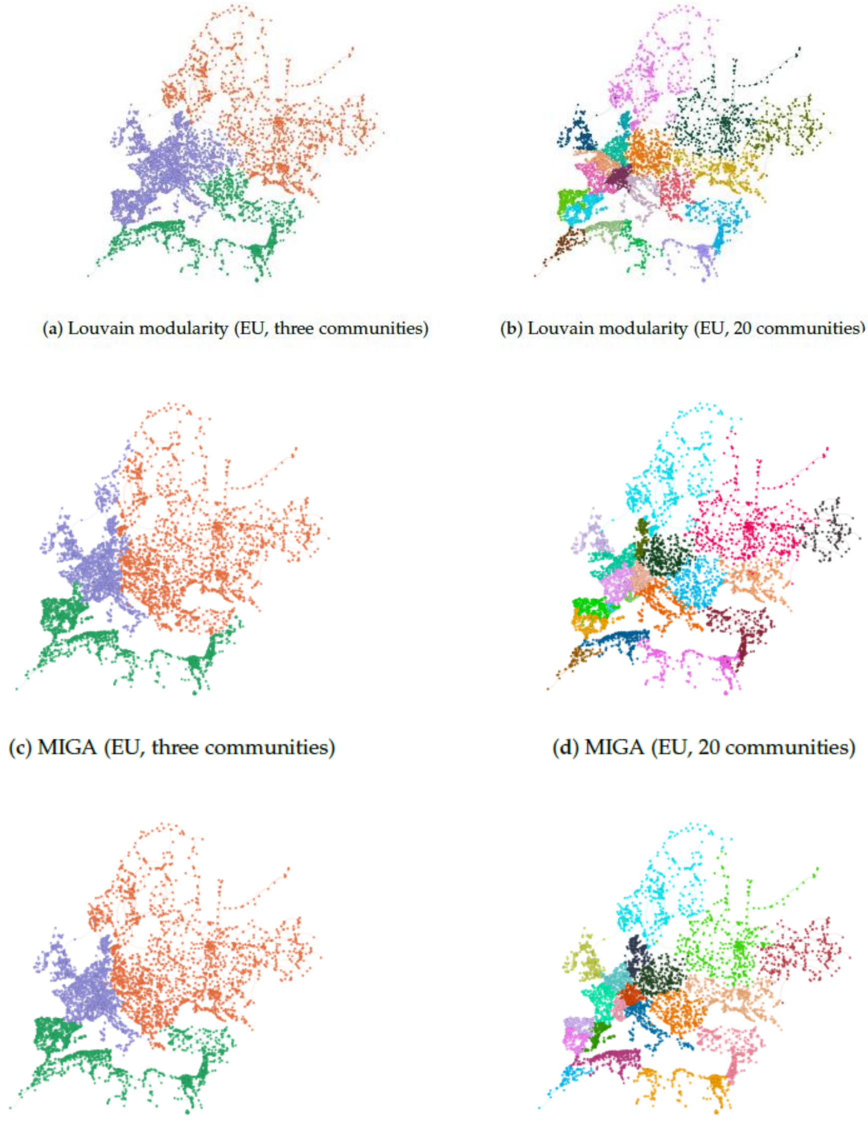
When two algorithms are compared, it is common to determine whether there are significant differences between the solutions they obtain. With this aim, a one-way ANOVA was applied, with the results indicating that the *p*-value was <0.05 in all cases; i.e., the null hypothesis was always rejected,

which means that there was a significant difference between at least some of the means of the different groups. Thus, the results obtained by GGA+ were significantly different from those obtained by MIGA, validating the mean values in Table 4.2.

The analysis of Figure 4.2 reveals that there are some differences between the results obtained by the two methods when detecting three and twenty communities, especially when the number of communities increases. Each community in these networks is represented with a random colour. Even in the case of detecting only three communities, MIGA has some difficulties in assigning communities in some parts of the graph, while GGA+ obtains clearly differentiated communities. Considering these graphical results and the results provided in Table 4.2, it can be concluded that GGA+ not only outperforms to MIGA, but also exhibits good performance in these large networks.

The results obtained by GGA+ are analysed in more detail here. Figure 4.3(a-c) display the communities detected by GGA+ in the European power grid with 5, 10, and 30 communities. These results reveal that this algorithm is able to obtain good quality solutions even when the number of communities increases. Moreover, Figure 4.3(d-f) provide a different layout based on the ForceAtlas2 [95] plugin in Gephi for these three networks. While the results presented in Figure 4.3(a-c) correspond to the coordinates of each node, the results displayed in Figure 4.3(d-f) cannot be read as a Cartesian projection. Instead, ForceAtlas2 was in a drawing mode that has the specificity of placing each node depending on the other nodes. This visualisation method builds a force directed layout by simulating a physical system in order to accommodate nodes and links in a spatial network. Nodes repel each other like charged particles, while edges attract their nodes like springs. Note that the same colour is used to represent the physical layout and the distribution obtained by ForceAtlas2. Moreover, the number of nodes in each community is often balanced (e.g., the five communities obtained in the European grid have a percentage of nodes between 19.16% and 20.47% of the total of nodes), although there are some significant imbalances between clusters when the number of communities increases (e.g., 30 communities). The analysis of Figure 4.3 demonstrates the good behavior of GGA+ independent of the degree of abstraction. Figure 4.4 shows how geographical structures change with the number of communities.

The analysis of the North American network supports similar conclusions. Thus, Figure 4.5(a-c) display the results obtained by GGA+ in that network when 5, 10, and 30 communities are detected. These data reveal that this algorithm is able to obtain good quality solutions not only with a few communities, but when the number of communities increases. The results obtained by the layout provided by ForceAtlas2 for these configurations (Figure 4.5(d-f)) demonstrate the good behavior of GGA+ independent of the degree of abstraction. Finally, Figure 4.5 shows that the algorithm is able to obtain differentiated clusters, even when the number of communities increases significantly. The results obtained here are of particular interest, bearing in mind that the North American electrical grid is made up of three interconnections: the Western Interconnection, the Eastern Interconnection, and the ERCOT (Texas) Interconnection, which are not synchronised, and alternating current (AC) power must be converted to direct current (DC) power for transfer across any of the interconnections. To overcome these limitations, the *Tres Amigas* superstation has been planned in New Mexico (U.S.), a 1.6 billion dollar project that aims to connect these three primary interconnections to facilitate the smooth, reliable, and efficient transfer of green power from region to region while integrating substantial renewable energy sources [136]. Figure 4.6 shows how geographical structures in the North American grid change with the number of communities.



(a) Louvain modularity (EU, three communities)

(b) Louvain modularity (EU, 20 communities)

(c) MIGA (EU, three communities)

(d) MIGA (EU, 20 communities)

(e) GGA+ (EU, three communities)

(f) GGA+ (EU, 20 communities)

Figure 4.2. Results obtained by Louvain modularity method, MIGA, and GGA+ for the European grid (three and 20 communities).

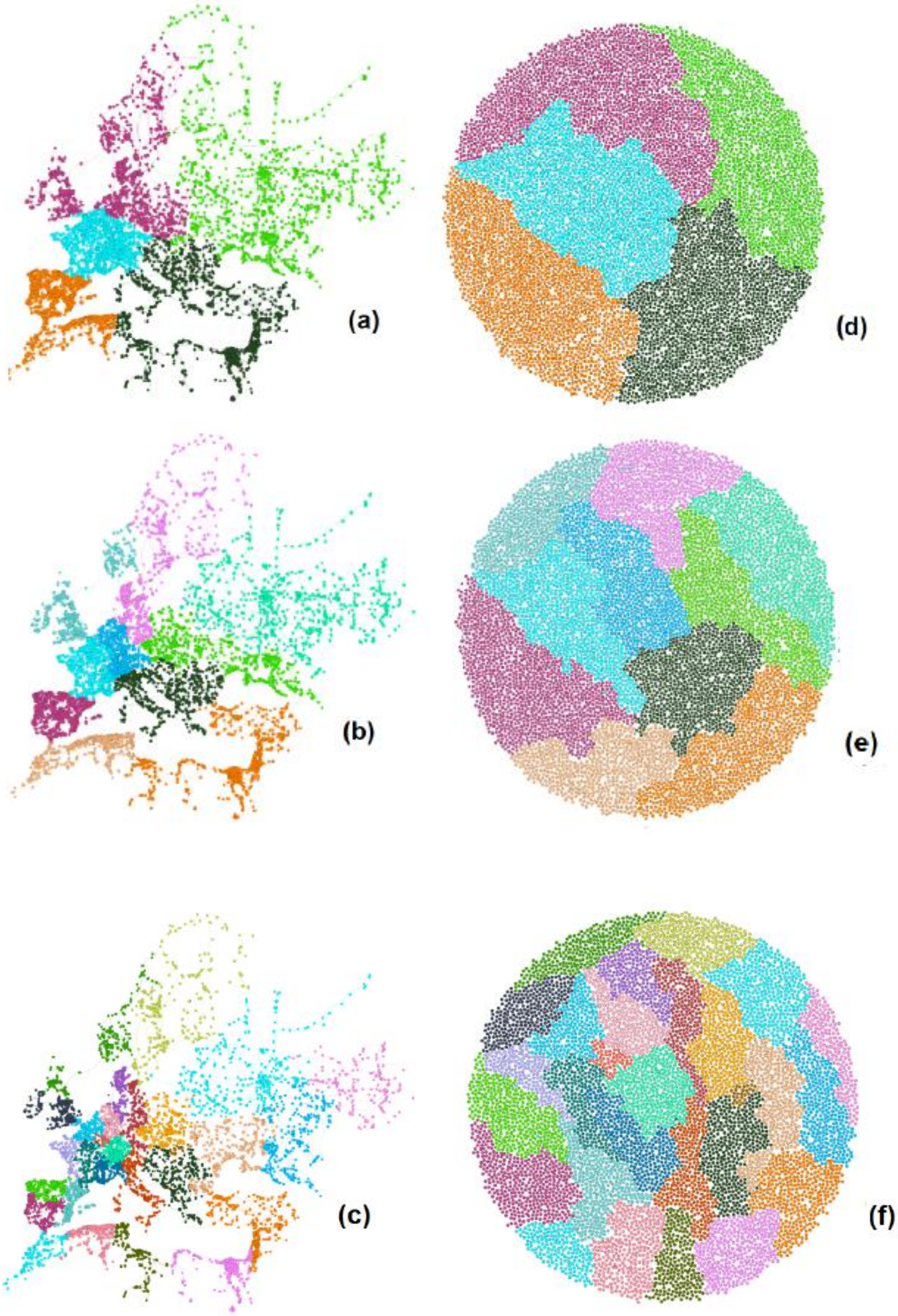


Figure 4.3. Results obtained by GGA+ for the European power grid: physical layout with (a) five communities, (b) 10 communities, and (c) 30 communities. Distribution obtained by ForceAtlas2 with (d) five communities, (e) 10 communities, and (f) 30 communities.

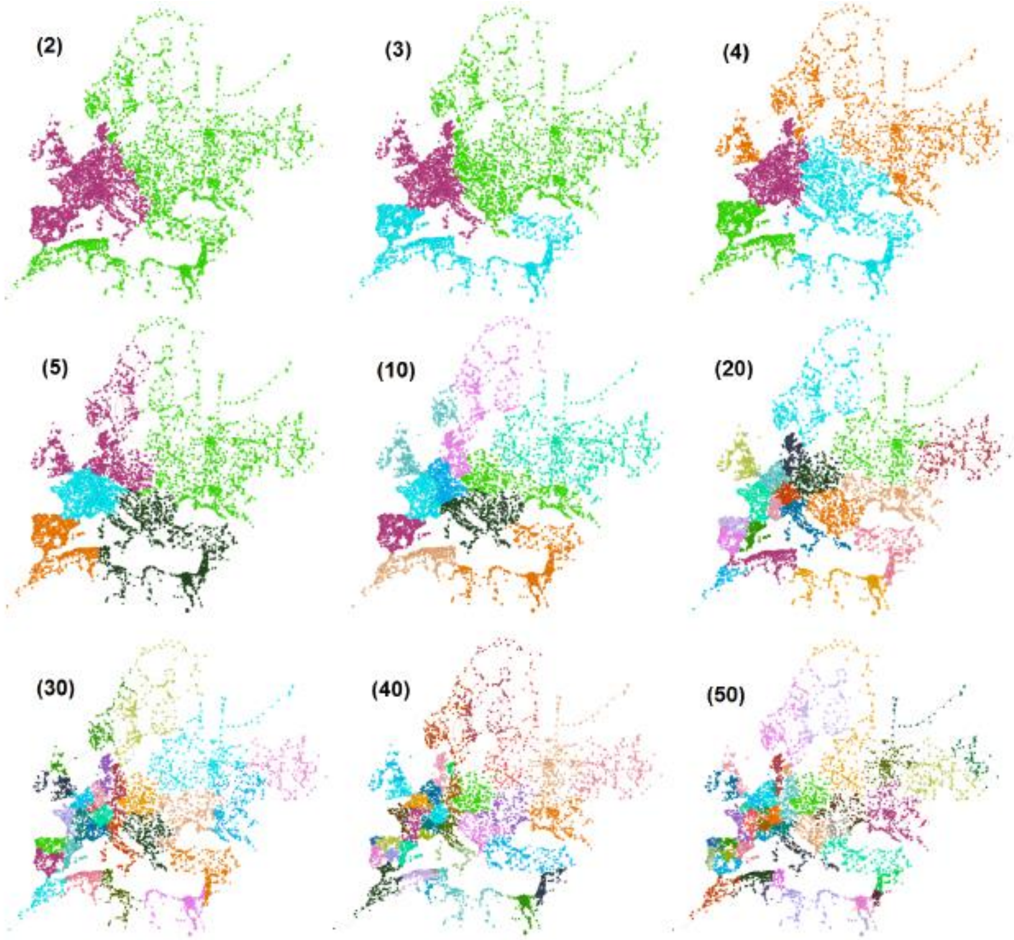


Figure 4.4. Physical layout of the communities detected by GGA+ in the European network using different degrees of abstraction (the number of communities is indicated in parentheses).

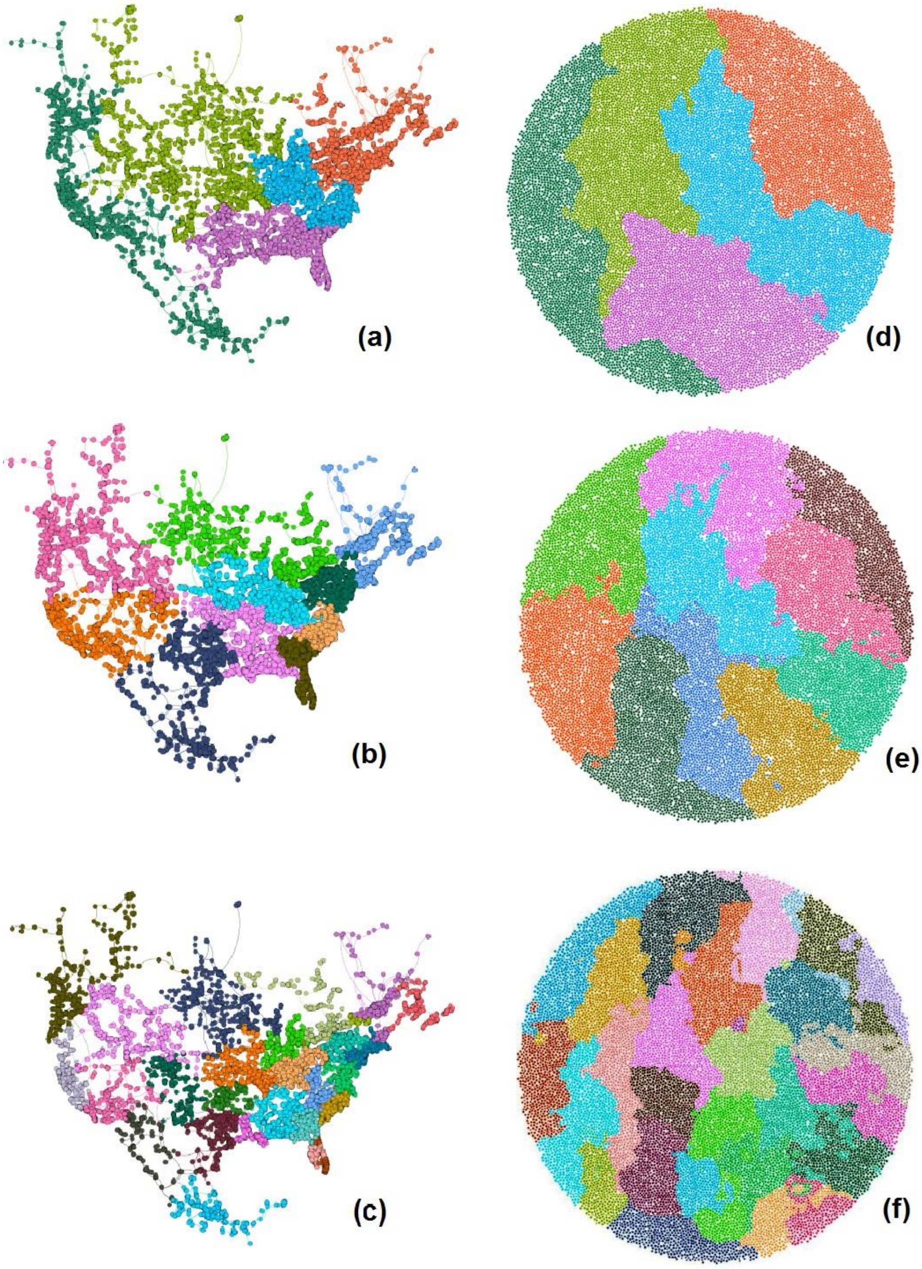


Figure 4.5. Results obtained by GGA+ for the North American power grid: physical layout with (a) five communities, (b) 10 communities, and (c) 30 communities. Distribution obtained by ForceAtlas2 with (d) five communities, (e) 10 communities, and (f) 30 communities.

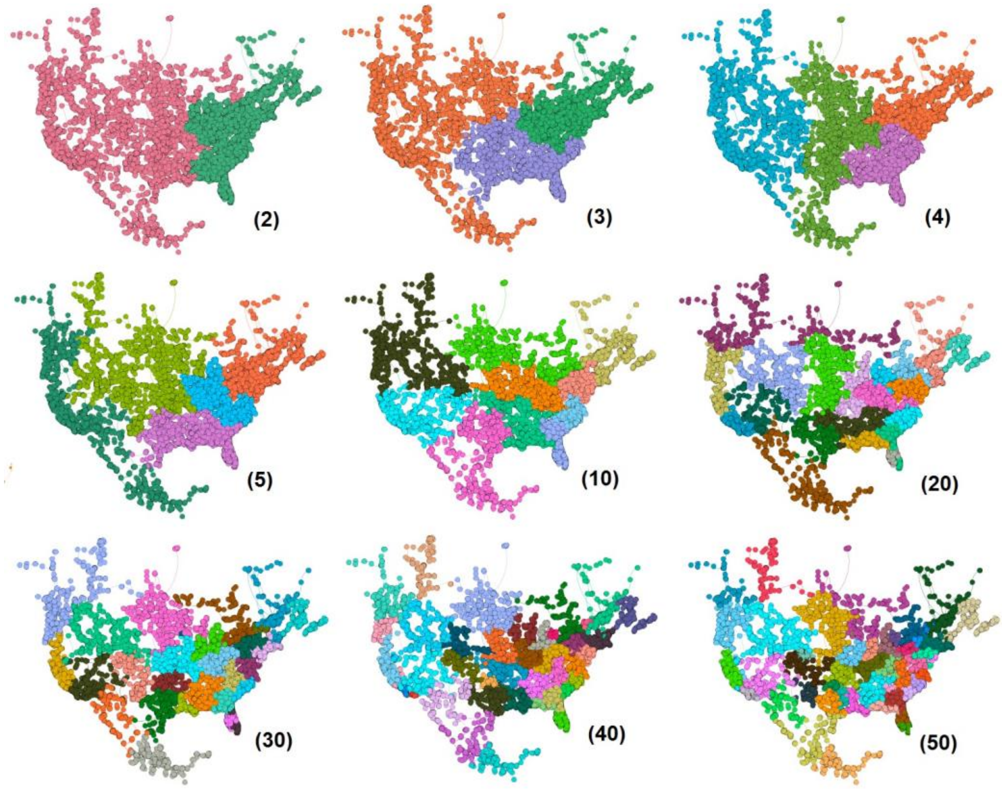


Figure 4.6. Physical layout of the communities detected by GGA+ in the North American network using different degrees of abstraction (the number of communities is indicated in parentheses).

4.5. Conclusions

The optimal design of high-voltage transmission networks is a critical issue to supply electrical energy to residential areas and industries. In fact, the growing integration of power grids across regions requires investment in more transmission power supply systems to ensure system stability and guarantee power supplies. To reach that aim, it is important to investigate the topological characteristics of these supergrids. This chapter opens a new avenue of research by analysing the community structures in supergrids in a fast and effective way. In particular, it is shown that solving the community detection problem with evolutionary algorithms allows one to obtain some key ideas about the structure of these networks. In particular, two evolutionary methods that include powerful initialisation methods and evolutionary search operators under the guidance of modularity were used to detect communities in large-scale networks. The evolutionary algorithms adopted a flexible and adaptive analysis of the characteristics of the power grids with different levels of detail (number of communities). The empirical study considered two large networks representing supergrids: (i) Europe, including Russia, North Africa, and part of the Near East (7893 nodes and 10346 branches); and (ii) North America (16063 nodes and 20169 edges). In particular, these methods were able to partition the networks into some loosely coupled sub-networks (communities) of similar scale, such that nodes within a community were densely linked, while connections between different communities were sparser. Numerical and graphical analysis using graph visualisation tools showed that GGA+ slightly outperformed MIGA, especially when the number of communities increased. Both evolutionary approaches outperformed the modularity values of the communities detected by the Louvain method implemented in Gephi. The results obtained show that evolutionary approaches are efficient methods for detecting communities in supergrids having thousands of nodes, and provide interesting topological information about the physical distribution and concentration of these elements of the grids. Future work will apply parallel and multi-objective optimisation methods and include the electrical properties of the power networks.

Capítulo 5

Multi-objective evolutionary algorithms to find community structures in large networks

Real-world complex systems are often modeled by networks such that the elements are represented by vertices and their interactions by edges. An important characteristic of these networks is they contain clusters of vertices densely linked amongst themselves and more sparsely connected to nodes outside the cluster. Community detection in networks has become an emerging area of investigation in recent years, but most papers aim to solve single-objective formulations, often focused to optimize the modularity. However, several studies have highlighted that considering modularity as a unique objective often involves resolution limit and imbalance inconveniences. This chapter opens a new avenue of research in the study of multi-objective variants of the classical community detection problem by applying multi-objective evolutionary algorithms that simultaneously optimize different objectives. In particular, they are analyzed two multi-objective variants involving not only modularity but also the conductance metric and the imbalance in the number of nodes of the communities. With this aim, it is presented a new Pareto-based multi-objective evolutionary algorithm that includes advanced initialization strategies and search operators. The results obtained when solving large-scale networks representing real-life power systems show the good performance of these methods and demonstrate that it is possible to obtain balanced number of nodes in the clusters formed, while also having high modularity and conductance values.

5.1. Introduction

Graph theory is one of the most important branches of mathematics. Graphs are often used to model networks such that nodes (vertices) are the elements and links (edges) denote interactions between these elements. In practice, graph theory is used to model real-life complex systems using graphs and to understand the role of the nodes within a given network. Some applications of graph theory are found in the study of transportation networks, computer and interconnection networks, telecommunication networks, electrical networks, biological systems, social networks, etc. [137].

Community detection is an emerging area of research that is attracting interest among scientists studying complex networks. The aim here is to detect community structures, that is, groups of densely interconnected nodes such that connections between the nodes are denser than connections with the rest of the network. The interest in detecting these groups or communities comes from the fact that the elements of each community potentially share similar features. Most research papers dealing with community detection consider single-objective formulations in which only one objective, usually *modularity* [40], is optimized. However, recent investigations have shown some drawbacks derived from the only use of *modularity*. For example, in [138] it is demonstrated the existence of resolution limit problems, while other authors have detected the existence of imbalance problems [139] which implies that classical measures tend to overemphasize either the interior or the exterior of a community. Some approaches have been proposed to mitigate the latter inconvenience, including symmetric frameworks to maintain a balance between the interior and the exterior of a community [140].

Given the above, it seems suitable to design algorithms to detect communities that do not only consider the *modularity*, but also other objective functions. A large number of methods have been proposed for solving multi-objective optimization problems (MOPs). Among these approaches, Multi-objective Evolutionary Algorithms (MOEAs) are probably the most widely applied strategies. MOEAs are stochastic optimization methods that simulate the process of natural evolution for solving MOPs. By using different strategies, MOEAs are able to obtain a set of solutions of the Pareto-optimal set, or at least good approximations to them [141]. This chapter proposes a new MOEA, called Multi-objective Generational Genetic Algorithm+ (MOGGA+), which extends the features of the Generational Genetic Algorithm+ (GGA+) [89] that has successfully been applied to the classical single-objective formulation of the community detection problem. The performance of the proposed method has been compared with a high-performance MOEA often used in the literature for solving large-scale benchmarks and network data taken from large-scale power grids.

The remainder of the chapter is organized as follows: Section 2 describes the problem of community detection in graphs, including an overview of some multi-objective formulations applied to this problem. Section 3 includes the formal description of the bi-objective formulations proposed in the chapter. Section 4 presents in detail the algorithm proposed to solve these bi-objective problems. Section 5 presents the empirical study, which compares the proposed method with other approaches in several case studies of different size and topology. The conclusions of the work are provided in Section 6.

5.2. Multi-objective community detection: An overview

Community detection is a problem closely related to that of classical graph partitioning [17]. In fact, graph partitioning is often used for community detection in different areas of application [142]. But while the goal of graph partitioning is to minimize the number of edges connecting nodes from different graphs, community detection consists of finding community structures [10], that is, groups of densely interconnected nodes such that connections between the nodes are denser than connections with the rest of the network. The interest in detecting these groups or communities comes from the fact that the elements of each community potentially share similar features [57]. It is important to remark that, in some real-world situations, the number of community structures that form a network is known beforehand and, therefore, the search space to be explored by the algorithms can be reduced. However, in many other cases, however, the number of community structures is initially unknown and the algorithms must obtain several solutions featuring different number of community structures.

Most research papers dealing with community detection consider single-objective formulations in which only one objective (usually *modularity* [40]) is optimized. But determining the optimum community structure that best represents the characteristics of a network is a complex task. Indeed, *modularity* maximization (see Equation 1) is an NP-hard problem [132], which means it is not possible to guarantee that the optimum solution will be found within a limited execution time [143]. Despite the generalized use of *modularity* to evaluate the quality of community detection algorithms, some authors have detected resolution limit [138] and imbalance problems [139].

Many problems in science and engineering are multi-objective since they involve the simultaneous optimization of two or more conflicting objectives, that is, the improvement of an objective often involves the deterioration of another or others. Typically, these MOPs have been addressed using scalarization techniques that combine the different objective functions into a single one that can then be solved by single-objective algorithms. Two typical scalarization techniques are: linear weighting and ε-constrained methods [144]. On the one hand, the idea behind weighting methods is to assign non-negative weights (at least one of them positive) to each objective function and then to minimize (or maximize) the weighted sum of the objectives. On the other hand, ε-constrained methods are based on ranking the objective functions in order of importance, such that each objective function is optimized individually, subject to the restriction that the higher ranked functions cannot exceed a certain percentage of the optimal values reached in previous generations (iterations). Despite their popularity, scalarization methods have certain drawbacks, for example, the assigning of weights or rankings to the objectives is often arbitrary. Furthermore, these methods only obtain a single (global) trade-off solution. An interesting way to overcome these drawbacks is to use Pareto-based optimization techniques. Pareto-based multi-objective algorithms aim to obtain not one but a set of solutions that are evaluated in terms of Pareto-dominance relations [145]. A solution A is said to be non-dominated or Pareto-optimal if no other feasible solution B dominates it, i.e., B is not better than A in at least one objective. The set of all non-dominated solutions found in the solution space forms the so-called Pareto-optimal front that represents the optimal trade-off between all objectives considered. This approach is very useful in decision-making processes since it provides a set of solutions to experts, who will choose the one that best suits their preferences. It is important to remark that it is very difficult to obtain the Pareto-optimal front in complex optimization problems, which is why the aim is to obtain a set of non-dominated solutions as an approximation to that set [53]. Many authors have also considered decomposition algorithms that decompose the task of approximating the Pareto-optimal front into a set of subtasks such that each task is a subproblem which can be single-objective or multi-objective. This strategy becomes useful when dealing with many objectives, but it seems to be more accurate for obtaining the entire Pareto-optimal front in bi-objective formulations since it provides a set of possible solutions to the decision-maker.

In recent years, some researchers have proposed solving the community detection problem by considering several objectives simultaneously. For example, [48] proposed the Multi-objective Genetic Algorithm for Networks (MOGA-Net), which maximizes the intra-connections inside each community and minimizes inter-connections between different communities. These objectives were also considered in [146], which proposed a MOEA with decomposition (MOEA/D-Net). Other authors have proposed decomposing *modularity* into two terms that represent the intra-link strength and the inter-link strength of a partition [50]. Similarly, [147, 148] proposed multi-objective evolutionary frameworks for solving multi-objective community detection approaches that consider the *intra-neighbour score* and the *inter-neighbour score* as objectives to optimize. In [149], a multi-objective algorithm was presented to optimize the *community score* and the *community fitness*. A MOEA based on Affinity Propagation (APMOEA) was presented in [138] to optimize the *ratio association* and the *ratio cut*, obtaining good results in comparison with MOEA/D-Net. In [150], the label-based dynamic multi-objective genetic algorithm (L-DMGA) was proposed for maximizing the *snapshot quality* and minimizing the *temporal cost*. In [151], a Multi-objective Genetic Algorithm (MOGA-OCD) was proposed for detecting overlapping communities such that the *internal connectivity* of the communities is maximized, whereas the number of *external connections* to the rest of the graph is minimized. In [152], the *negative ratio association* and *ratio cut* were optimized using the Discrete Inverse Modelling-based MOEA with Decomposition algorithm (DIM-MOEA/D), which obtains a similar or better performance than other approaches, including MOCD [50], MOGA-Net [153], MOEA/D-net [146] and MODPSO/D [154].

In addition to MOEAs, other meta-heuristic approaches for solving multi-objective community detection problems have been proposed. This is the case of the Particle Swarm Optimization (PSO) algorithm with decomposition (MOPSO/D) proposed in [154] for minimizing the *kernel k-means* and *ratio cut*. These objectives were also considered years later in [155], which proposed the so-called MOPSO-Net, also based on PSO. In [156], the researchers presented the so-called Multi-objective Immune Algorithm for Multi-Resolution Community Detection (MICD), which aims to optimize the *modified ratio association* and the *ratio cut*. In [115], a Multi-objective Biogeography based Optimization Algorithm with Decomposition (MBBOD) was presented to simultaneously optimize *modularity* and a metric that measures the similarity of attributes of the nodes of a community. Other MBBOD implementation was introduced in [157], where the two objectives to optimize were *modularity* and the *normalized mutual information*. This pair of objectives were also considered in [158] and optimized by the Multi-objective Discrete Teaching-Learning-based Optimization with Decomposition (MODTLBO/D), which obtained good results in comparison with MOCD [50], MOGA-Net [153], MOEA/D-Net [146] and MODPSO/D [154] in different problem instances. Other researchers proposed a multi-objective optimization community detection algorithm with attribute information (MOCDA) to simultaneously optimize *modularity* and *homogeneity* [159]. This pair of objectives has also been considered in [160], which proposed the so-called Multi-objective Attributed Community Detection Algorithm with Node Importance Analysis (MANIA). Another investigation [161] proposed a local information based MOEA (L-MOEA) that adopts a decomposition strategy to optimize the *negative ratio association* and *ratio cut*.

Therefore, taking into account these and other previous studies, it is obvious that it is possible to create a large number of multi-objective community detection formulations by combining different objectives [25]. The overview of the state of the art in this field shows that most papers dealing with community detection from a multi-objective perspective still currently use scalarization or decomposition approaches. Only a few approaches have solved multi-objective community detection approaches considering Pareto-based methods, for example, as in [162] which proposed the Multi-objective Adaptive Fast Evolutionary Algorithm for community detection (F-SGCD) that optimizes *community score* and *community fitness* using Pareto-dominance comparisons. This chapter aims to advance in the design of efficient Pareto-based multi-objective algorithms for solving multi-objective formulations of the community detection problem that consider additional objectives to the *modularity*.

5.3. Problem formulation

This chapter proposes the analysis of two bi-objective formulations of the community detection problem: (a) Maximize the *modularity* (Equation 1) and minimize the *imbalance* (Equation 3); (b) Maximize the *conductance* (Equation 2) and minimize the *imbalance* (Equation 3).

- Modularity (Q) [40]: Modularity considers that a solution is good if there are many edges within communities and only a few between them. A solution with a Q value close to 1 indicates strong community structure from a topological perspective [152].

$$Q = \frac{1}{2M} \sum \left(a_{ij} - \frac{K_i K_j}{2M} \right) \delta(i, j)$$

  where *M* is the total number of edges in the network; the sub-indices *i* and *j* denote two nodes of the network; K_i and K_j are the degree of the *i-th* and *j-th* nodes, respectively; the parameter a_{ij} is the element of the *i-th* row and the *j-th* column of the adjacency matrix; and $\delta(i, j)$ represents the relationship between the *i-th* node and the *j-th* node, such that if node *i* and node *j* are in the same community, $\delta(i, j) = 1$; otherwise, $\delta(i, j) = 0$. The aim is, therefore, to maximize the modularity (Q) value.

- Conductance (CON) [163, 164]: The conductance is a measure of the fraction of total edge volume that points outside the community. The aim here is to minimize the conductance (CON) value.

$$CON = \sum_{i=1}^{N} \frac{L_i}{L_i + 2Z_i}$$

where L_i is the number of edges of the *i-th* community that are linked to nodes from other communities, Z_i is the number of edges of the *i-th* community that are not linked to nodes from other communities, and *N* is the number of communities considered.

- Imbalance (IMB): The imbalance represents the difference in the number of nodes included in the communities detected. The aim is, therefore, to minimize the imbalance (IMB) value.

$$IMB = \sqrt{\frac{\sum_{i=I}^{N}(X_i - \overline{X})^2}{N}}$$

where X_i is the number of nodes of the *i-th* community, $\overline{X}$ is the number of nodes of the network divided by the number of communities considered, and *N* is the number of communities considered.

5.4. Multi-objective Generational Genetic Algorithm: MOGGA+

This section presents MOGGA+, a Pareto-based MOEA that solves the multi-objective formulations described above. MOGGA+ extends the single-objective algorithm GGA+ [89] by including some strategies such as *radial initialization*, the use of n set of non-dominated solutions and the dynamic modification of the probability of applying evolutionary operators in runtime.

In particular, it incorporates a new data structure that assigns some probabilities of executing a genetic operator that can be dynamically adjusted (within a given range) according to the evolution of the operator during the successive generations of the algorithm. Furthermore, MOGGA+ considers different degrees of abstraction, which allows the user to perform a flexible and adaptive analysis of the network using graphical information with different levels of detail.

Figure 5.1 shows the operation of MOGGA+. The input data required by MOGGA+ is a graph *G* modelling a given network, the number of communities to be detected, the population size, crossover rate, mutation rate and the termination criterion (maximum number of generations). With this information, MOGGA+ creates a new population initialized based on the radial initialization process and initializes the migration vector between boundaries. Then, the algorithm evaluates the individuals through the objective functions and creates a set of non-dominated solutions that will store all the individuals that satisfy the constraints. After that, the individuals of the population are sorted by assigning a rank according to Pareto dominance comparisons. Then a new population is generated by applying the dynamically selected genetic operator (crossover and/or mutation) and the individuals are checked in order to update the non-dominated set. Finally, the replacement operator is applied to the new population. At the end of the execution, the algorithm returns the set of non-dominated solutions as an approximation to the true Pareto-optimal front.

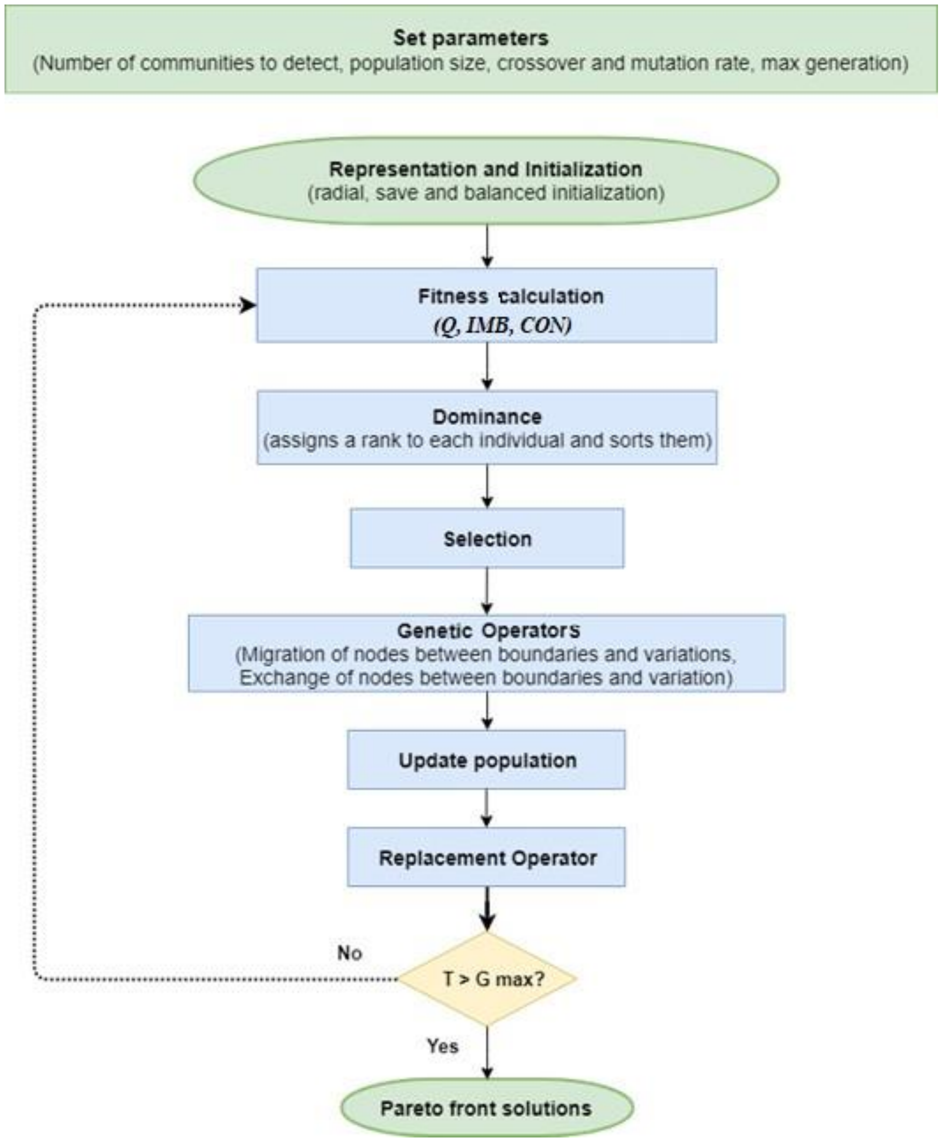


Figure 5.1. MOGGA+ operation.

5.4.1. Genetic representacion

The genetic representation used by MOGGA+ is based in the following notation:

$$r_m = \left[r_m^1\ r_m^2 \dots r_m^i \dots r_m^N\right]$$

where vector r_m represents the *m-th* individual of the population and r_m^i indicates the community to which the *i-th* node belongs. All the nodes in this individual contain a positive integer value, where *N* is the total number of nodes of the network.

5.4.2. Initialisation of the population

Although random initialisation is often used, it can generate unfeasible solutions, that is, isolated nodes or groups of nodes that are not interconnected with nodes of the same community. To guarantee that the individuals generated during the initialization process are feasible, the initialization process used by MOGGA+ is based on the concept of *safe initialization* [41]. In addition, to avoid the generation of unbalanced size communities, the concept of *balanced initialization* used in [89] is incorporated to establish the community size. It must also be noted that MOGGA+ allows the P_{size} individuals of the population search for a different number of communities (*degree of abstraction*), such that the network can be analyzed from a global perspective (a few communities) to a higher detail (many communities).

Furthermore, to improve the quality of the initial communities, a *radial initialization* is used which functions as follows: Given a set of communities $S_1, S_1, \dots S_i, \dots S_c$, where *C* is the number of communities to detect, each community S_i is created by including that node n_j not previously assigned to another community having the higher degree of connectivity with other nodes. Then, those neighbouring nodes of the node n_j not previously assigned to community S_i are incorporated until it reaches the community size previously

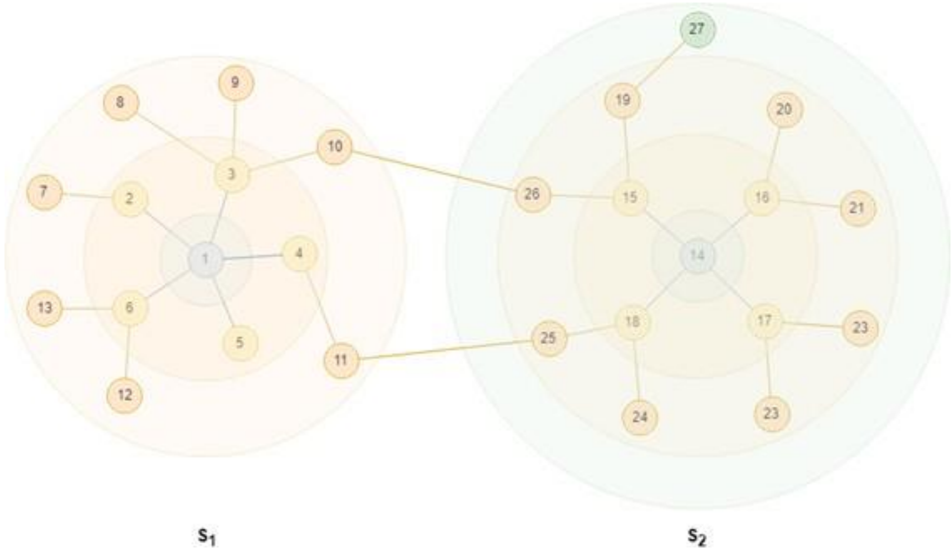


Figure 5.2. Radial Initialisation.

established. If the community size is not reached after including node n_j and its neighbours, n_{jk}, the process is repeated with the first neighbour node, n_{j1}. If this criteria is not reached, the process is then repeated with the second neighbour node, n_{j2}, and so on. In the hypothetical case that the community does not reach the pre-established size, the process would be repeated considering the neighbour nodes located at the next level of distance, that is, using concentric ratios where the neighbour nodes are added within that distance. This process can be observed in the example shown in Figure 5.2. Here, the maximum number of communities to be detected is 2 and the number of nodes is 27. The initialization vector will establish the size of the communities at $|S_1| = 13$ and $|S_2| = 14$. Next, a non-assigned node is selected considering the concept of *node with higher degree*. In this case, node 1 (degree 5) is selected and then included in community S_1. As S_1 has not been completed (current size 1 with respect to $|S_1| = 13$), the neighbouring nodes to node 1 are included, that is, community S_1 now contains the nodes 1 to 6, but the maximum size $|S_1| = 13$ has still not been reached, which is the reason why the next level is analyzed, until the community is completed (S_1 contains nodes 1 to 13). Once community S_1 is completed, this process is repeated to complete community S_2.

5.4.3. Migration vector and genetic operators

MOGGA+ uses a migration vector between boundaries [89]. This vector determines the most attractive destination of each boundary node between different communities, such that given a node, the community that contains the highest number of nodes connected to the former will be selected as the destination boundary community to which the selected node will migrate. Based on this migration vector, the algorithm herein proposed (MOGGA+) applies several genetic search operators that have been especially designed to obtain the maximum performance of the proposed data structure.

5.4.3.1 Mutation operator

MOGGA+ uses three mutation operators that consist of the migration of boundary nodes to a different community. These operators are randomly applied.

- Migration of a boundary node to the best destination community (M1): moves boundary node *j* located at community S_i to the best neighbouring community S_{best} [89].

- Migration of N nodes to the best destination community (M2): moves boundary node *j* located at community S_i to the best destination community S_{best}. Furthermore, a random number of neighbouring nodes of node *j* are also moved to community S_{best}.

- Migration of N nodes to a random destination community (M3): moves boundary node *j* located at community Si to a random destination community S_{random}. Furthermore, a random number of neighbouring nodes of node *j* are also moved to community S_{random} .

5.4.3.2 Crossover operator

Some studies have shown that typical crossover operators are not suitable for community detection problems since they lead to the disruption of good communities or may even cause the generated communities to be disconnected, thus significantly degrading the search capability of the algorithms [165]. Our implementation consists of the exchange of communities between boundary nodes of different communities. Two crossover/exchange operators are considered by MOGGA+:

- Best exchange of boundary nodes (EX1): moves boundary node *j* located at community S_i to the best neighbouring community S_{best}, and then moves from S_{best} to S_i the node k which obtains the best result from moving to the S_i community.

- Random exchange of boundary nodes (EX2): moves the boundary node *j* located at the community S_i to a random community S_{random}, and then moves from S_{random} to S_i the node *k* that gets the best result from moving to the S_i community.

5.4.3.2 Selector operator

MOGGA+ uses an elitist replacement procedure for substituting a percentage of individuals of the main population for some individuals of the non-dominated set. The number of individuals to replace in each generation is a random number calculated in the range $\left[\left((minRatio + incrementRatio) * P_{size}\right),\ \left(maxRatio * P_{size}\right)\right]$ where:

- *minRatio*: minimum number of individuals to replace.
- *maxRatio:* maximum number of individuals to replace.
- *incrementRatio:* parameter that dynamically increases the number of individuals to replace, that is, the algorithm becomes more elitist when the number of generations performed increases. Let G_{max} be the number of generations that the population will evolve, the value of *incrementRatio* is calculated as follows:

$$incrementRatio = (maxRatio - minRatio)/G_{max}$$

5.4.4. Termination criteria

The termination criteria used here is to perform a maximum number of generations (G_{max}).

5.5. Empirical study

This section analyzes the performance of MOGGA+ in networks of different sizes and characteristics. The experiments were performed on a personal computer with an Intel Core i7 3630Q processor (2.4 GHz, 8 GB DDR3 RAM), which executes the application developed in C# .Net Framework 4.

5.5.1. Algorithms

The performance of MOGGA+ is compared with MOGA-Net [153]. This algorithm has been used for two reasons: MOGA-Net is a multi-objective algorithm developed by Clara Pizzuti which is often used in the context of multi-objective community detection; MOGA-Net adapts the well-known Non-dominated Sorting Genetic Algorithm (NSGA-II) proposed by Deb et al. [49] to the problem of community detection. The genetic representation of MOGA-Net is based on the locus-based adjacency representation [166]. The initialization process is based on random generation of individuals, but taking into account the effective connections of the nodes in the network. MOGA-Net uses a standard uniform crossover operator, while the mutation operator is implemented in order to guarantee that each node is linked only to one of its neighbours in the mutated child. The operation of MOGA-Net consists of creating a new population initialized at random and repaired to produce safe individuals. After that, the algorithm evaluates the individuals through the objective functions, then applying NSGA-II to each one. NSGA-II assigns a rank according to Pareto dominance and sorts the individuals. Then, a new population is generated by applying the genetic operators (uniform crossover and mutation). Finally, once the termination criterion is achieved, the set of non-dominated solutions obtained is returned as output of the algorithm. It is worth noting that it was necessary to implement a new version of MOGA-Net that considers the objectives of modularity, imbalance and conductance since the original MOGA-Net implementation maximizes the intra-connections inside each community and minimizes inter-connections between different communities.

5.5.2. **Test problems**

To conduct the performance analysis of the different algorithms herein presented, networks representing five national-scale power grids proposed in [131] are utilized. Four graphs were obtained from the transmission system map, available on the webiste of the European Network of Transmission System Operators for Electricity (ENTSO-E) [91]: Italy, including Sardinia and Sicily, Germany, the continental territory of France, and the Iberian peninsula, including the Balearic islands. Further, the comparison also covers a graph model of the Texas power grid [93], which is almost entirely managed by the Electric Reliability Council of Texas (ERCOT). Table 5.1 describes some graph characteristics of these five networks representing high voltage power grids. Some previous studies have shown that community detection in national-scale high voltage transmission networks provides interesting topological information about the physical distribution and concentration of these grids [131].

Table 5.1. Description of the graphs used to model the five power grids.

Feature Power grid	Italy	Germany	France	Iberian peninsula	Texas
Nodes	352	438	904	1104	2007
Edges	462	662	1163	1416	2607
Average degree	2.63	3.03	2.57	2.57	2.60
Network diameter	39	21	28	40	39

5.5.3. **Parameter settings**

Table 5.2 shows the parameters settings used by MOGGA+, including the population size, the number of generations (termination criterion), and the probabilities of applying the different variants of the mutation, crossover and selection operators. The same population size and number of generations are considered for MOGA-Net.

Table 5.2. MOGGA+ parameters.

Population size (Psize)		200
Generations (Gmax)		200
Mutation probability (initial / min / max)		
	M1	0.35 / 0.20 / 0.60
	M2	0.35 / 0.20 / 0.60
	M3	0.30 / 0.20 / 0.60
Crossover probability (initial / min / max)		
	EX1	0.50 / 0.20 / 0.80
	EX2	0.50 / 0.20 / 0.80
Selection probability (minRatio / maxRatio)		0.15 / 0.35

5.5.4. Performance metrics

As commented above, the aim of multi-objective optimization algorithms is to obtain the true Pareto-optimal front or, alternatively, an approximation to it. However, an important issue here is the intrinsic trade-off between the goals of proximity and diversity preservation [167], that is, the selection mechanisms should select a diverse set of solutions close to the set of non-dominated solutions. A large number of performance metrics have been proposed in the past [168]. Two widely used metrics have been used in our study: the Hyper-volume and the Schott's spacing metric. The Hyper-volume (HV) metric is the only unary indicator that is Pareto compliant [56] and often used as a measure of convergence towards the Pareto front as well as the maximum spread of the solutions obtained. The Schott's Spacing (SS) metric [169] measures the spread of solutions in a non-dominated set according to the relative distance between the nearest solutions in the non-dominated set.

5.5.5. Results and discussion

To conduct the performance analysis, a total of 30 independent runs have been performed for each algorithm on the five networks representing the power grids of Italy, France, Germany, the Iberian Peninsula and Texas. The accuracy of the algorithms has been evaluated according to the HV and SS metrics described above. Table 5.3 shows the results obtained by both algorithms in these five networks when optimizing Modularity and Imbalance, while Table 5.4 shows the same comparison when optimizing Conductance and Imbalance. As can be observed, in both multi-objective formulations MOGGA+ clearly outperforms the results obtained by MOGA-Net in all these networks in terms of HV metric, that is, the former provides a better approximation to the (unknown) true Pareto-optimal front than the latter.

Table 5.3. Results obtained by MOGA-Net and MOGGA+ considering Modularity and Imbalance objectives.

	Method	Hyper-volume		Schott's spacing	
		Best	Mean	Best	Mean
Italy	MOGA-Net	1.683	1.644	0.024	0.046
	MOGGA+	2.186	2.168	0.076	0.096
Germany	MOGA-Net	1.667	1.594	0.022	0.061
	MOGGA+	2.166	2.156	0.060	0.093
France	MOGA-Net	1.894	1.779	0.016	0.030
	MOGGA+	2.929	2.919	0.077	0.165
Iberian Peninsula	MOGA-Net	1.533	1.426	0.124	0.010
	MOGGA+	2.578	2.546	0.168	0.018
Texas	MOGA-Net	1.763	1.605	0.012	0.021
	MOGGA+	3.939	3.926	0.159	0.194

Table 5.4. Results obtained by MOGA-Net and MOGGA+ considering Conductance and Imbalance objectives.

	Method	Hyper-volume		Schott's spacing	
		Best	Mean	Best	Mean
Italy	MOGA-Net	5936.804	5723.398	0.120	0.254
	MOGGA+	6160.020	6160.020	66.202	66.202
Germany	MOGA-Net	2680.895	2616.242	0.096	0.254
	MOGGA+	2831.827	2831.735	3.891	3.891
France	MOGA-Net	29984.032	28639.357	0.155	0.328
	MOGGA+	35615.731	35614.983	0.00	193.953
Iberian Peninsula	MOGA-Net	43316.548	41260.761	0.154	0.421
	MOGGA+	53646.681	53646.653	6.584	7.992
Texas	MOGA-Net	12556.285	11639.416	0.262	0.653
	MOGGA+	19685.936	19685.928	0.216	34.303

The analysis of the results in terms of SS metric indicates, however, that MOGA-Net outperforms MOGGA+. These results are due to the characteristics of spacing metrics. In particular, some previous studies have highlighted that if the solutions of the non-dominated set are clustered in small groups, the distance between the groups is not considered since only the shortest distances are computed [170]. Our results denote that MOGGA+ obtains a better approximation to the true Pareto-optimal front but these solutions are relatively dispersed in the solution space, while MOGA-Net often obtains a set of non-dominated solutions distant from the Pareto-optimal front, but more concentrated, so that the spacing between solutions is smaller, thus obtaining a better result in the SS metric. Figure 5.3. Comparison between MOGA-Net and MOGGA+ in the Italian network considering (a) modularity and imbalance. shows the non-dominated fronts obtained by MOGA-Net and MOGGA+ in the Italian network when optimizing both formulations. It is important to remark that the non-dominated fronts shown in both figures are concave because the minimization of the conductance and imbalance objectives has been implemented as the maximization of the inverse values of CON (Equation 2) and IMB (Equation 3). In these figures, it can be observed that the non-dominated sets obtained by MOGGA+ show a better approximation to the (unknown) Pareto-optimal front, but there are several solutions which are far from the others, which is why the SS metric obtained by MOGA-Net is better.

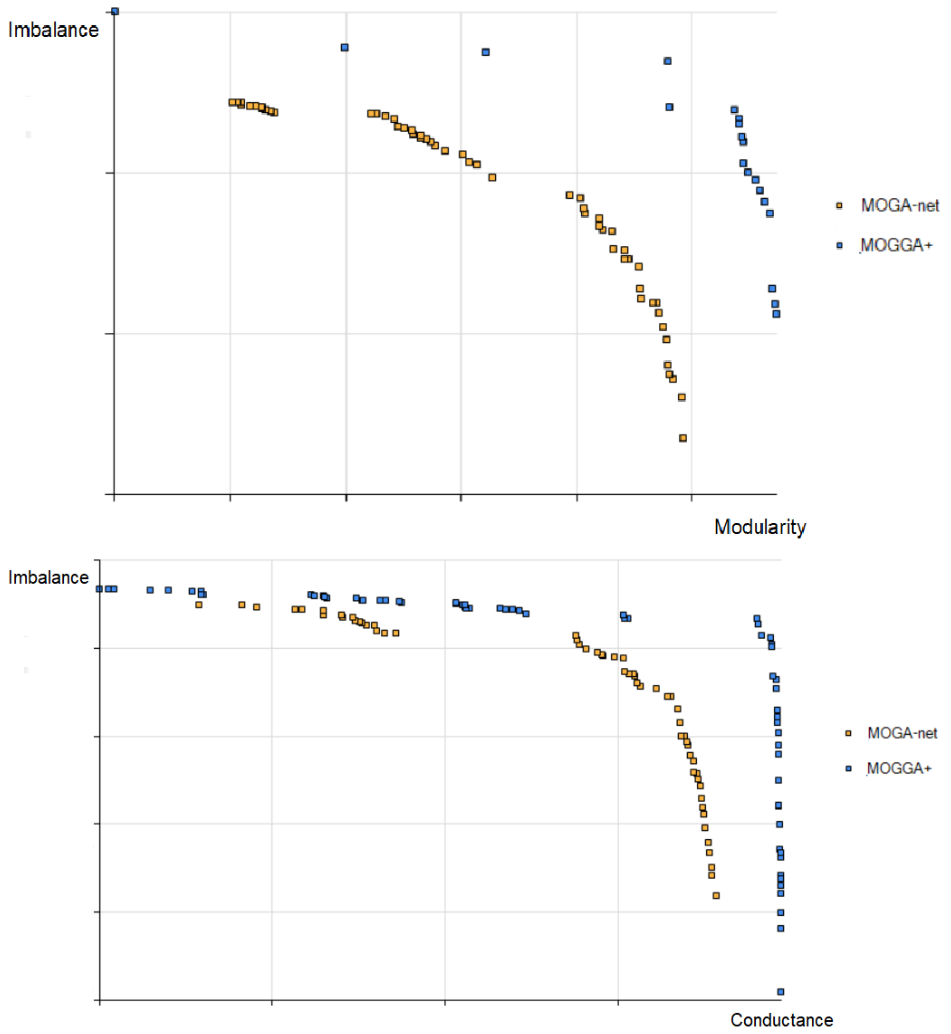
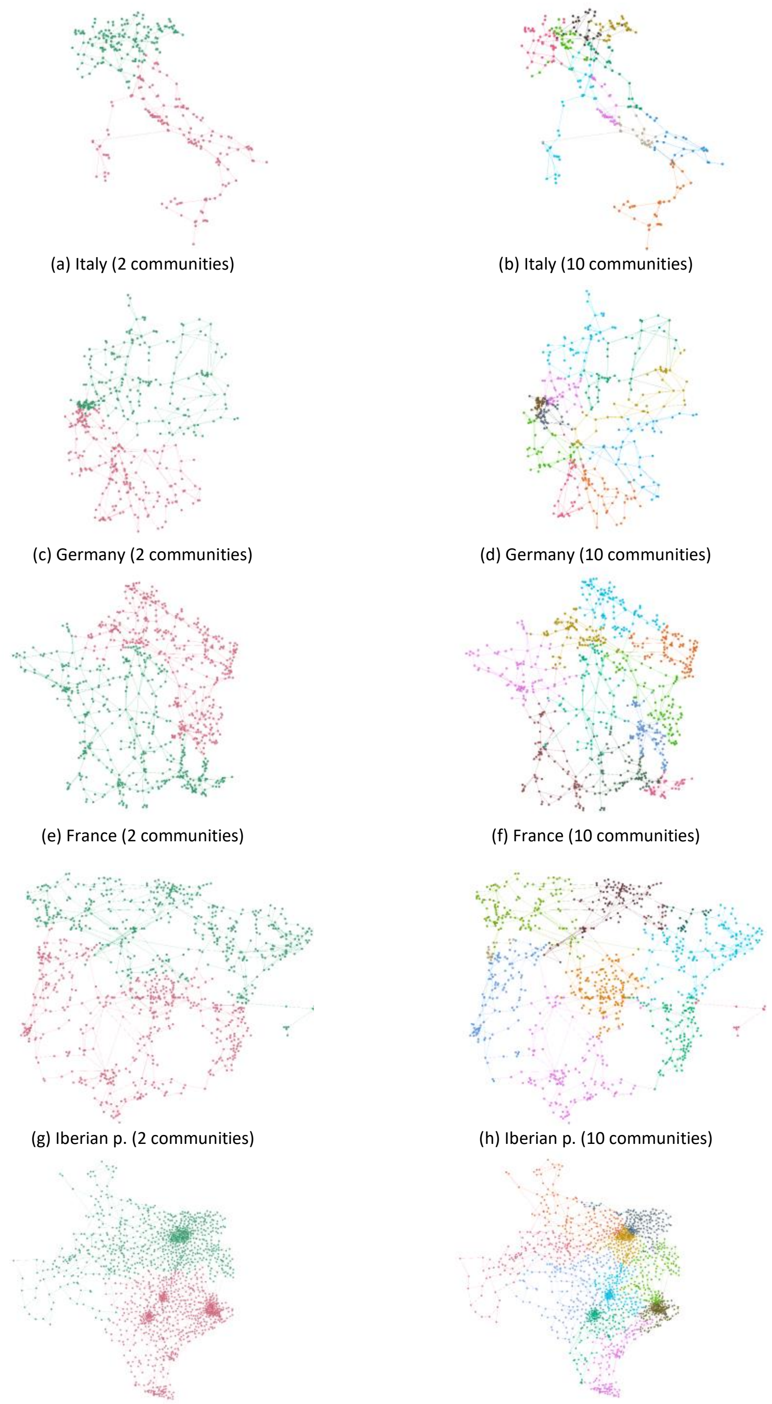


Figure 5.3. Comparison between MOGA-Net and MOGGA+ in the Italian network considering (a) modularity and imbalance.



(a) Italy (2 communities) (b) Italy (10 communities)

(c) Germany (2 communities) (d) Germany (10 communities)

(e) France (2 communities) (f) France (10 communities)

(g) Iberian p. (2 communities) (h) Iberian p. (10 communities)

(i) Texas (2 communities) (j) Texas (10 communities)

Figure 5.4. Physical layout of the communities detected in the networks representing the power grids of Italy, Germany, France, Iberian P. and Texas considering 2 and 10 communities.

Finally, Figure 5.4 displays some examples of communities detected by MOGGA+ in the five networks representing power grids considering different levels of resolution (2 and 10 communities). As can be expected, solutions with only two communities return balanced solutions, while increasing the number of communities often increases the imbalance. Figure 5.4 is useful to understand how the implemented algorithms could provide high quality solutions considering different objectives, such that the decision-maker could later decide which one is the best option according to the particular characteristics of the study at hand.

5.6. Conclusions

Community detection is an important area of investigation in the field of complex networks. An overview of the state of the art in this field shows that most published papers aim to maximize the *modularity* value. However, considering *modularity* as a lone objective can involve resolution limit and imbalance inconveniences. This chapter is the first to propose the use of Pareto-based MOEAs for solving two different multi-objective formulations: (a) the maximization of the modularity and the minimization of the imbalance; and (b) the maximization of the conductance and the minimization of the imbalance. More specifically, a new multi-objective evolutionary algorithm is described that includes efficient initialization methods and search operators to obtain a set of non-dominated solutions as an approximation to the unknown Pareto-optimal front. The empirical study includes the comparison with a well-known MOEA used for community detection. Different networks representing real power systems having hundreds of vertices and edges are used to evaluate the performance of the algorithms. The numerical and graphical results show the high performance of these Pareto-based MOEAs for solving both formulations. This chapter opens a new avenue of research by detecting community structures considering different objectives in a large variety of large-scale networks. Moreover, the results obtained in these graphs representing national-scale high voltage transmission networks provide interesting topological information about the physical distribution and concentration of these grids according to different criteria.

Capítulo 6
Conclusiones

Como resultado de la investigación realizada se han obtenido diferentes publicaciones en revistas especializadas de impacto internacional, destacando principalmente las tres publicaciones (capítulos 2, 3 y 4) aceptadas en revistas incluidas en el listado Journal Citation Reports (JCR) de la Web of Science (WOS), y una cuarta publicación en proceso de revisión (capítulo 5), que supone una interesante aportación en el campo de la optimización computacional evolutiva multiobjetivo al demostrar la eficiencia y utilidad de los métodos desarrollados mediante su aplicación en problemas reales de ingeniería eléctrica.

A continuación, se presentan más detalladamente las principales aportaciones del trabajo de investigación.

6.1. Resultados y discusión

Aplicar la detección de comunidades sobre sistemas reales complejos es una poderosa herramienta analítica para estudiar, comprender y obtener información valiosa sobre las relaciones y características que las estructuras de comunidades ocultan en los sistemas modelados, pero a la vez, se trata de un problema de alta complejidad algorítmica clasificado como problema NP-Completo. En esta tesis se han propuesto varios algoritmos evolutivos mono-objetivo y multiobjetivo para abordar la detección de comunidades de manera general y, se han puestos en práctica en infraestructuras tan críticas como las redes eléctricas de transmisión de alta tensión, pues suministran la energía necesaria para la sociedad actual en pueblos, ciudades, regiones, países y continentes. Los principales resultados y conclusiones obtenidos son agrupados acorde a las publicaciones realizadas.

6.1.1. Publicación científica 1: “Adaptive community detection in complex networks using genetic algorithms” (Capítulo 2)

Como punto de partida, en esta publicación se ha diseñado e implementado el algoritmo genético generacional mono-objetivo (*GGA+*). GGA+ toma como referencia el índice de modularidad e incorpora un novedoso método de inicialización poblacional eficiente, basado en los conceptos de “*inicialización segura*” y “*balanceada*”, además, utiliza nuevos operadores genéticos avanzados basados en el intercambio entre nodos frontera. El desarrollo de estos importantes componentes ha contribuido individualmente al excepcional rendimiento que ofrece el algoritmo en su conjunto. Además, durante el diseño de GGA+ se ha puesto especial interés en la incorporación de un enfoque denominado "*Grado de abstracción*" el cual permite un análisis flexible y adaptativo de las características de una red desde diferentes niveles de detalle de acuerdo con las necesidades de un analista.

En cuanto al rendimiento de GGA+, se han utilizado distintos benchmarks reales y sintéticos (con tamaños de hasta varios cientos de nodos y miles de aristas) para analizar y comparar los resultados obtenidos frente a otros cinco algoritmos evolutivos diseñados con diferentes estrategias, entre los que cabe destacar "*Modularity and Improved Genetic Algorithm*" (MIGA) el cual incluye a Simulated Annealing como algoritmo de búsqueda local. Los resultados obtenidos muestran que GGA+ supera al resto de algoritmos, con lo que se demuestra el alto rendimiento alcanzado por el algoritmo.

Las decisiones de diseño tomadas sobre varios componentes de GGA+, han sido clave para lograr el alto rendimiento que este ha mostrado. Entre ellas, destacan la estrategia estructural basada en un diseño generacional, el método de inicialización poblacional basado en "inicialización segura y balanceada", y especialmente, los operadores evolutivos avanzados. Gracias a estos componentes, GGA+ ha podido superar el rendimiento del resto de algoritmos genéticos incluidos en la comparación. Por tanto, es posible concluir que las estrategias de diseño seguidas en estos componentes, pueden ser utilizadas por otras heurísticas y metaheurísticas de detección de comunidades para beneficiarse de su eficiencia en cualquier ámbito.

6.1.2. Publicación científica 2: “Community detection in national-scale high voltage transmission networks using genetic algorithms” (Capítulo 3)

Al hilo de esta tesis, una vez se ha desarrollado y presentado una técnica evolutiva eficiente y contrastada, capaz de realizar un análisis más detallado de las redes, en este artículo se demuestra la capacidad y versatilidad de los métodos evolutivos de detección de comunidades al aplicarlos sobre sistemas eléctricos reales complejos.

Es precisamente en el ámbito de ingeniería eléctrica donde se enmarca el estudio realizado en esta tesis sobre una de las infraestructuras más importantes para la sociedad actual, las redes eléctricas de transmisión de alta tensión a escala nacional y continental. Hasta la fecha se trata del primer estudio que empíricamente ha analizado el rendimiento de los métodos evolutivos sobre sistemas eléctricos de tales dimensiones. En la actualidad, las redes eléctricas siguen siendo una infraestructura de enorme inversión para compañías eléctricas y gobiernos en un esfuerzo conjunto por abastecer la creciente demanda energética requerida por una sociedad inmersa en la globalización, en la que, además, la irrupción de nuevas fuentes de energía renovables distribuidas provocan un cambio de paradigma en el diseño de las redes tradicionales y, por tanto, la necesidad de adaptación a ellas. Todo ello, ha derivado en un aumento y actualización de las líneas de transmisión encargadas de interconectar regiones cada vez más extensas, provocando que el diseño de redes eléctricas se vuelva más y más complejo. Este aumento de la complejidad requiere de la aparición y uso de nuevas técnicas de diseño y control para prevenir el aumento de posibles contingencias imprevistas como cortes de líneas, desastres naturales, ataques terroristas, etc., permitiendo el restablecimiento de la red con el menor impacto posible en la cadena de suministro energético.

En esta publicación se estudia la aplicación de métodos evolutivos sobre redes eléctricas de alta tensión a escala nacional. Concretamente, se ha evaluado el rendimiento de dos algoritmos evolutivos mono-objetivo (MIGA y GGA+) sobres redes nacionales de distintas características como la Península Ibérica, Alemania, Francia e Italia y, la red americana de Texas formada por 2.007 nodos y 2.607 aristas. Ambos algoritmos incorporan el concepto "*Grado de abstracción*", el cual, gracias a la capacidad que ofrece de analizar redes desde distintos niveles de detalle (distinto número de comunidades), ha permitido revelar interesantes patrones de conexión. Un ejemplo de ello, es el observado en la red alemana, donde es posible apreciar la subdivisión que la red sufrió debido a conflictos políticos durante la Guerra Fría. Otro ejemplo es observado en la red de Texas, donde se revela una clara y diferenciada acumulación de subestaciones y líneas eléctricas en las zonas localizadas por las mayores ciudades (Houston, San Antonio, Dallas y Austin) del estado. En particular, los resultados obtenidos por GGA+ permiten observar comunidades más claramente diferenciadas que las obtenidas por MIGA, incluso cuando el número de comunidades aumenta puesto que también aumenta la dificultad para detectar comunidades. El estudio empírico llevado a cabo, permite demostrar que GGA+ supera a MIGA en todos los escenarios analizados, mostrando a su vez una gran precisión y estabilidad.

A raíz de los resultados obtenidos por los algoritmos analizados, se confirma el buen rendimiento que los métodos genéticos evolutivos ofrecen para resolver el problema de la detección de comunidades en redes eléctricas de escala nacional y, por consiguiente, en problemas reales de ámbito industrial. Incluso, gracias al análisis flexible y adaptativo, se pueden obtener datos valiosos sobre la distribución física de la red que pueden ser útiles tanto en el diseño topológico de la misma, como en el diseño de controles de contingencias. El análisis de contingencias es una importante cuestión a tener en cuenta a la hora de diseñar una red segura y fiable. En este sentido, en esta publicación se introduce una nueva vía de investigación al aplicar la detección de comunidades como apoyo a los procedimientos basados en el estudio de sistemas eléctricos de flujo de potencia tales como el criterio de seguridad (n-1) o el particionamiento de islas para la desconexión parcial de partes de la red.

6.1.3. Publicación científica 3: “Evolutionary Algorithms for Community Detection in Continental-Scale High-Voltage Transmission Grids” (Capítulo 3)

Hasta este punto, se ha confirmado el buen rendimiento que ofrecen los métodos evolutivos como herramientas útiles y complementarias a los procedimientos actuales de análisis y detección de comunidades sobre redes eléctricas de transmisión de alta tensión de escala nacional, pero es en este artículo donde se analiza el rendimiento de estos métodos en escenarios más globales de escala continental, también llamados superredes. El concepto de superred nació como una solución para permitir intercambios de energía eléctrica a gran escala en áreas de todo el continente. Este estudio es particularmente atractivo si observamos el enorme interés generado en los gobiernos, traducido en enormes cantidades de recursos financieros destinados a proyectos millonarios como la unificación de las redes nacionales en un único mercado europeo con el fin de cumplir los objetivos de energía renovable marcados por la Unión Europea debidos al cambio climático.

El diseño y manejo de estas superredes es una tarea compleja que requiere garantizar la fiabilidad y seguridad de las líneas de transmisión eléctrica para poder trasportar eficientemente la energía desde fuentes generadoras hasta el cliente final. Para alcanzar este objetivo es necesario el uso de herramientas que permitan modelar los sistemas complejos y analizar las infraestructuras. Esta tarea puede llevarse a cabo mediante la técnica de análisis de grafos basada en detección de comunidades, pues proporciona información valiosa sobre aspectos topológicos. La importancia de la detección de comunidades en redes eléctricas procede del hecho de que es necesario mantener la fiabilidad de la red y prepararla para posibles contingencias como desastres naturales (ej.: tormentas) mediante el diseño de planes de conmutación para proteger islas o desconectar partes de la red, evitando una mayor degradación durante los incidentes y permitiendo una restauración más rápida después de la perturbación.

Teniendo en cuenta las consideraciones anteriores, en esta publicación se ha realizado un estudio pionero en redes eléctricas de escala continental, con el objetivo de evaluar el rendimiento que los métodos de detección de comunidades evolutivos pueden ofrecer en redes de gran tamaño y gran complejidad. Para ello, se ha realizado un estudio empírico utilizando distintos métodos de alto rendimiento como MIGA, GGA+ o el bien conocido método de Louvain, sobre dos superredes, la red eléctrica europea que incluye Rusia, Norte de África y parte de Oriente Próximo formada por 7.893 nodos y 10.346 aristas, y la red norteamericana con 10.063 nodos y 20.169 aristas. Ambas redes cuentan con un tamaño considerablemente mayor que otras típicas redes eléctricas de prueba como la "IEEE 118-bus test system" formada por 118 nodos y 186 aristas.

Tras los resultados obtenidos en ambas superredes, se constata la capacidad ofrecida por los algoritmos evolutivos para detectar estructuras de comunidades bien definidas, tanto en soluciones genéricas que van desde 2 comunidades hasta en soluciones más detalladas con 50 comunidades, proporcionando así la posibilidad de realizar análisis con distintos niveles de detalle (grado de abstracción), y demostrando la flexibilidad y adaptabilidad de los métodos evolutivos. En cuanto al rendimiento obtenido por estos métodos, es destacable la estabilidad de los resultados obtenidos entre distintas ejecuciones de GGA+ y MIGA, y el alto rendimiento mostrado por GGA+ frente a MIGA y el método de Louvain en términos cualitativos de la solución encontrada. Empíricamente, se ha podido concluir que a medida que aumenta el número de comunidades a detectar, GGA+ es capaz de detectar estructuras mejor diferenciadas, de lo que MIGA es capaz. A raíz de estos resultados, se puede concluir que GGA+ supera a MIGA y que, además, demuestra un buen rendimiento en superredes independientemente del nivel de detalle, consiguiendo resultados precisos y estables, gracias a sus métodos de inicialización eficientes y operadores genéticos avanzados.

Por último, como se ha mencionado anteriormente, un diseño óptimo de redes de transmisión eléctrica, necesita de la inversión en sistemas de suministro para garantizar la cantidad y estabilidad del suministro energético. Esta inversión conlleva un aumento de la complejidad de la red, dando lugar a la necesidad de análisis y estudios topológicos más completos que pueden ser llevados a cabo mediante la detección de comunidades. En el estudio realizado en este artículo se abre una nueva vía de investigación al demostrar el buen rendimiento que los métodos genéticos evolutivos ofrecen para resolver el problema de la detección de comunidades de manera rápida y efectiva sobre superredes de miles de nodos.

6.1.4. Publicación científica 4: "Multi-objective evolutionary algorithms to find community structures in large networks" (cap. 5)

Una visión general del estado del arte en este campo muestra que la mayoría de métodos destinados a resolver el problema de la detección de comunidades utilizan un enfoque mono-objetivo, siendo el índice "*Modularidad*" la función objetivo más ampliamente utilizada. Sin embargo, considerar "modularidad" como único objetivo puede implicar inconvenientes en el límite de resolución y problemas de desbalanceo. Por estos motivos, y como objetivo final de esta tesis, y concretamente en esta publicación, tras los resultados obtenidos previamente que avalan el buen rendimiento ofrecido por los métodos evolutivos en problemas reales de ingeniería eléctrica, se ha tomado como referencia GGA+ para diseñar la versión evolutiva multiobjetivo (MOGGA+) basada en la dominancia de Pareto. Esta versión ofrece la posibilidad de estudiar redes de gran tamaño en función de varios objetivos, abriendo de esta manera una nueva vía de investigación mediante el diseño y aplicación de nuevos algoritmos evolutivos multiobjetivo que optimizan simultáneamente diferentes objetivos en redes de gran tamaño.

MOGGA+, no es solamente la versión multiobjetivo de GGA+ con la capacidad añadida de evaluar varios objetivos apoyándose en el concepto de dominancia de Pareto, sino que además, introduce conceptos novedosos con respecto a su predecesor como: un novedoso método de inicialización poblacional más eficiente denominado "*inicialización radial*", la "*ejecución de distintos operadores genéticos basada en probabilidades dinámicas*", un "*método de reemplazo elitista*", y el ya comentado concepto "*grado de abstracción*", el cual permite un análisis flexible y adaptativo.

A la hora de seleccionar los pares de objetivos a optimizar por el enfoque multiobjetivo presentado, no solo se ha tomado como referencia el índice de "*Modularidad*", sino que alternativamente a él, se han considerado otros objetivos como la "*conductancia*" o el propuesto en esta tesis, "*desbalanceo de comunidades*". Este último objetivo, proporciona un factor clave a tener en cuenta a la hora de diseñar redes eléctricas resistentes a contingencias, pues es importante que la distribución de la red permita ser separada en islas (subredes o comunidades), de tamaño aproximadamente similar (balanceadas), para evitar una mayor degradación o desenergización del sistema, ya sea parcial o total, y así permitir una rápida restauración frente a la perturbación. En este sentido, este artículo es el primero en considerar la optimización de las duplas de objetivos formadas por ["Modularidad", "Desbalanceo"] y ["Conductancia", "Desbalanceo"].

Los resultados obtenidos procedentes del análisis de rendimiento realizado sobre redes eléctricas nacionales como la Península Ibérica, Alemania, Francia, Italia o Texas, permiten concluir la idoneidad del uso de métodos evolutivos multiobjetivo para el estudio topológico de sistemas eléctricos y la efectividad del objetivo "*desbalanceo*" al generar comunidades más balanceadas. Además, la comparación de los frentes de Pareto obtenidos por MOGA-Net y MOGGA+, ha permitido concluir que MOGGA+ supera a su competidor en base a la métrica Hyper-Volumen, la cual es utilizada como medida de convergencia al desconocido frente óptimo de Pareto. El diseño de todos los componentes introducidos en MOGGA+ ha contribuido notablemente a mejorar el rendimiento del algoritmo y a superar a MOGA-Net, basado en el popular NSGA-II.

El buen rendimiento demostrado numérica y gráficamente por los métodos evolutivos multiobjetivo, abre una nueva vía de investigación en gran variedad de sistemas complejos al considerar diferentes objetivos a optimizar en detección de comunidades. Finalmente, los frentes de Pareto resultantes revelan información interesante sobre la distribución física de las redes eléctricas de transmisión de alta tensión, que puede ser utilizada como apoyo en la toma de decisiones relacionadas con la planificación de métodos de prevención de contingencias o el diseño topológico de redes eléctricas de gran tamaño.

6.1.5. **Conclusiones generales**

La revisión de las conclusiones obtenidas a lo largo de todas las publicaciones científicas derivadas de esta tesis, permiten afirmar que los objetivos marcados en la presente tesis, han sido alcanzados, pues se han logrado desarrollar métodos evolutivos mono-objetivo y multiobjetivo precisos, eficientes y estables, que pueden ser aplicados a diversidad de áreas como deportes, política e ingeniería eléctrica entre otras, con la garantía de obtener siempre buenos resultados, independientemente del nivel de detalle requerido (número de comunidades detectadas). Además, gracias al enfoque basado en el "*grado de abstracción*" se han podido analizar sistemas complejos en mayor profundidad. Prueba de ello, es la aportación realizada en el campo del análisis de sistemas eléctricos reales de transmisión de alta tensión, al proporcionar nuevos métodos de apoyo a la toma de decisiones, donde, además, queda demostrada la escalabilidad que proporcionan los métodos evolutivos al ser aplicados eficientemente a sistemas eléctricos complejos de gran tamaño.

6.2. Trabajo futuro

El trabajo futuro se centrará en estudiar con más detalle cómo las estructuras de comunidades obtenidas con algoritmos de detección de comunidades podrían proporcionar información útil para el análisis de contingencia y las estrategias de isla, incorporando las propiedades eléctricas de las redes de energía.

Además, se espera aplicar métodos de optimización paralelos con múltiples objetivos como "puntuación de comunidad", "aptitud de comunidad", "puntuación intravecinos", "similaridad", "asociación de relación negativa", "ratio de corte", etc., para la detección de comunidades en redes eléctricas de escala nacional y continental. La incorporación de técnicas de paralelismo en los métodos evolutivos reducirá los tiempos de ejecución, especialmente en las redes continentales que, por sus dimensiones, precisan de altos requerimientos computacionales que pueden ser afrontados aprovechando el uso de las arquitecturas paralelas de los procesadores actuales formados por diferentes núcleos de procesamiento.

Otras publicaciones

- Poster para el Congreso Gecco 2017, titulado "Community detection in power grids by an evolutionary method" sobre la aplicación de métodos evolutivos para la detección de comunidades en redes eléctricas [171].

Referencias Bibliográficas

[1] Bollobas, B., Modern Graph Theory, New York, USA: Springer Verlag, 1998.

[2] Freeman, L. C., The development of social network analysis, Empirical, 2004.

[3] Barberán, A., Bates, S.T., Casamayor, E.O., Fierer, N., «Using network analysis to explore co-occurrence patterns in soil,» *The ISME journal,* vol. 6, nº 2, pp. 343-351, 2012.

[4] Aggarwal, C., Subbian, K., «Evolutionary network analysis: A survey,» *ACM Computing Surveys (CSUR),* vol. 47, nº 1, 2014.

[5] Baños, R., Ortega, J., Gil, C., Márquez, A. L., De Toro, F., «A hybrid meta-heuristic for multi-objective vehicle routing,» *Computers & Industrial Engineering,* vol. 65, nº 2, pp. 286-296, 2013.

[6] Hu, D.L., Wang, N., Jia, Y., Ge, S. , A literature review on cloud computing by using social network analysis, Engineering, London, 2015.

[7] Montoya, F.G., Baños, R., Gil, C., Espín, A., Alcayde, A., Gómez, J., «Minimization of voltage deviation and power losses,» *Engineering Applications of Artificial Intelligence,* vol. 23, nº 5, pp. 695-703, 2010.

[8] Ellison, N.B, «Social Network Sites: Definition, History, and Scholarship,» *Journal of Computer-Mediated Communication,* vol. 13, nº 1, pp. 210-230, 2007.

[9] Ríos, S.A., Videla-Cavieres, I.F., «Generating Groups of Products Using Graph Mining Techniques,» *Procedia Computer Science,* vol. 35, pp. 730-

738, 2014.

[10] Newman, M., «Communities, modules and large-scale structure in networks,» *Nature Physics,* vol. 8, nº 1, pp. 25-31, 2012.

[11] Fortunato, S., «Community detection in graphs,» *Physics Reports,* vol. 486, nº 3-5, pp. 75-174, 2010.

[12] Boettcher, S., Percus, A.G, «Extremal optimization for graph partitioning,» *Physical Review E,* vol. 64, nº 2, 2001.

[13] Gil, C., Ortega, J., Díaz, A.F., Montoya, M.D.G., «Annealing-based heuristics and genetic algorithms for circuit partitioning in parallel test generation,» *Future Generation Computer Systems,* vol. 14, nº 5, pp. 439-451, 1998.

[14] Kirkpatrick, S., Gelatt, C., Vecchi, M., «Optimization by simulated annealing,» *Science,* vol. 220, nº 4598, pp. 671-680, 1983.

[15] Chen, M., Kuzmin, K., Szymanski, B.K., «Community Detection via Maximization of Modularity and Its Variants,» *IEEE Transactions on Computational Social Systems,* vol. 1, nº 1, pp. 46-65, 2014.

[16] Newman, M.E.J, Girvan, M., «Community structure in social and biological networks,» *PNAS,* pp. 7821-7826, 2002.

[17] Gil, C., Ortega, J., Montoya, M., Baños, R. , «A mixed heuristic for circuit partitioning,» *Computational Optimization and Applications,* vol. 23, nº 3, pp. 321-340, 2002.

[18] Newman, M.E.J., «Fast algorithm for detecting community structure in networks,» *Physical review E,* vol. 69, nº 6, 2004.

[19] Holland, J., «Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence,» de *Ann Arbor*, University of Michigan, 1975.

[20] Yang, X.S., «A new metaheuristic bat-inspired algorithm,» *Nature inspired cooperative strategies for optimization,* vol. 284, pp. 65-74, 2010.

[21] Yang, X.-S., «Firefly algorithm, stochastic test functions and design optimisation,» *International Journal of Bio-Inspired Computation,* vol. 2, nº 2, pp. 78-84, 2010.

[22] Gregory, S., «An algorithm to find overlapping community structure in networks,» de *European Conference on Principles of Data Mining and Knowledge Discovery*, vol. 4702, Springer, 2007, pp. 91-102.

[23] Palla, G., Derényi, I., Farkas, I., Vicsek, T., «Uncovering the overlapping community structure of complex networks in nature and society,» *Nature,* vol. 435, nº 7043, pp. 814-818, 2005.

[24] Javed, M.A., Younis, M.S., Latif, S., Qadir, J., Baig, A., «Community detection in networks: A multidisciplinary review,» *Journal of Network and Computer Applications,* vol. 108, pp. 87-111, 2018.

[25] Osaba, E., Del Ser, J., Camacho, D., Bilbao, M.N., Yang, X.-S., «Community detection in networks using bio-inspired optimization: Latest developments, new results and perspectives with a selection of recent meta-heuristics,» *Applied Soft Computing,* vol. 87, p. 106010, 2020.

[26] Talbi, E.-G., Metaheuristics: From Design to Implementation, Wiley, 2009.

[27] Greiner, R., «PALO: a probabilistic hill-climbing algorithm,» *Artificial Intelligence,* vol. 84, nº 1-2, pp. 177-208, 1996.

[28] Glover, F., Laguna, M., Tabu Search, Boston: Kluwer Academic Publishers, 1997.

[29] Dorigo, M., Stutzle, T., Ant colony optimization, Bradford Book, 2004.

[30] Kennedy, J., «Particle swarm optimization,» *Encyclopedia of machine learning,* pp. 760-766, 2011.

[31] Martí, R., Laguna, M., Glover, F., «Principles of Scatter Search,» *European Journal of Operations Research,* vol. 169, nº 2, pp. 359-372, 2006.

[32] Fogel, D.B., Evolutionary Computation: Toward a New Philosophy of Machine Intelligence, 3rd ed., Hoboken, NJ, USA: Wiley-IEEE Press, 2006.

[33] Moscato, P., Memetic Algorithms, In Handbook of Applied Optimization, In P.M. Pardalos y M.G.C. Resende (eds.), Oxford University Press, 2000.

[34] Manzano-Agugliaro, F., San Antonio Gómez, C., López, S., Montoya, F., Gil, C., «Pareto-based evolutionary algorithms for the calculation of transformation parameters and accuracy assessment of historical maps,» *Computers & Geosciences,* vol. 57, pp. 124-132, 2013.

[35] Montoya, F.G., Manzano-Agugliaro, F., López-Márquez, S., Hernández-Escobedo, Q., Gil, C., «Wind turbine selection for wind farm layout using multi-objective evolutionary algorithms,» *Expert Systems with Applications,* vol. 41, nº 15, pp. 6585-6595, 2014.

[36] Ke, Q., Dong, Y., Wu, B., «Effcient search in networks using conductance,» de *International Conference on Advances in Social Networks Analysis and Mining*, Kaohsiung, Taiwan, 2011.

[37] Chen, R., Mo, Y., Pan, Z., «Performance improvement of edge expansion technique for bdd-based network reliability analysis,» *Journal of Computers,* vol. 8, nº 9, pp. 2190-2196, 2013.

[38] Mishra, N., Schreiber, R., Stanton, I., Tarjan, R.E., «Clustering social networks,» de *Algorithms and Models for the Web-Graph*, Springer, 2007.

[39] Leskovec, J., Lang, K. J., Mahoney, M., «Empirical comparison of algorithms for network community detection,» de *Proceedings of the 19th international conference on World wide web*, ACM, 2010.

[40] Newman, M.E.J, Girvan, M., «Finding and evaluating community structure in networks,» *Physics Review E,* vol. 69, nº 2, 2004.

[41] Pizzuti, C., «Ga-net: a genetic algorithm for community detection in social networks,» de *In Proc. of the 10th International Conference on Parallel Problem Solving from Nature (PPSN 2008)*, Dortmund, Germany, 2008.

[42] Lancichinetti, A., Fortunato, S., Kertész, J., «Detecting the overlapping and hierarchical community structure in complex networks,» *New Journal of Physics,* vol. 11, nº 3, p. 33015, 2009.

[43] Aldecoa, R., Marín, I., «Deciphering network community structure by surprise,» *PLoS ONE,* nº 6(9), 2011.

[44] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D., «Defining and identifying communities in networks,» *Proceedings of the National Academy of Sciences of the United States of America,* vol. 101, nº 9, pp. 2658-2663, 2004.

[45] Shang, R., Bai, J., Jiao, L., Jin, C., «Community detection based on modularity and an improved genetic algorithm,» *Physica A: Statistical Mechanics and its Applications,* vol. 392, nº 5, pp. 1215-1231, 2013.

[46] Gong, M., Fu, B., Jiao, L., Du, H., «Memetic algorithm for community detection in networks,» *Physical Review E,* vol. 84, nº 5, 2011.

[47] Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., «Fast unfolding of communities in large networks,» *Journal of Statistical Mechanics: Theory and Experiment,* vol. 2008, 2008.

[48] Pizzuti, C., «A Multi-Objective Genetic Algorithm for Community Detection in Networks,» de *Proc. Of the 21th International Conference on Tools with Artificial Intelligence*, Newark, New Jersey, USA., 2009.

[49] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., «A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II,» *IEEE Transactions on Evolutionary Computation,* vol. 6, nº 2, pp. 182 - 197, 2002.

[50] Shi, C., Yan, Z., Cai, Y., Wu, B., «Multi-objective community detection in complex networks,» *Applied Soft Computing,* vol. 12, nº 2, pp. 850-859, 2012.

[51] Corne, D., Jerram, N., Knowles, J., Oates, M., «Pesa-II: region-based selection in evolutionary multiobjective optimization,» *Genetic and Evolutionary Computation Conference,* pp. 283-290, 2001.

[52] Deb, K., Multi-objective Optimization Using Evolutionary Algorithms, New

York: John Wiley & Sons, 2002.

[53] Baños, R., Gil, C., Paechter, B., Ortega, J., «A hybrid meta-heuristic for multi-objective optimization: Mosats,» *Journal of Mathematical Modelling and Algorithms,* vol. 6, nº 2, pp. 213-230, 2007.

[54] Lancichinetti, A., Fortunato, S., Radicchi, F., «Benchmark graphs for testing community detection algorithms,» *Physical,* vol. 78, 2008.

[55] Danon, L., Diaz Guilera, A., Duch, J., Arenas, A., «Comparing community structure identification,» *Journal of Statistical Mechanics: Theory and Experiment,* vol. 09, 2005.

[56] Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Da Fonseca, V.G., «Performance assessment of multiobjective optimizers: An analysis and review,» *IEEE Transactions on evolutionary computation,* vol. 7, pp. 117-132, 2003.

[57] Wang, R.-S., Zhang, S., Wang, Y., Zhang, X.-S., Chen, L., «Clustering complex networks and biological networks by nonnegative matrix factorization with various similarity measures,» *Neurocomputing,* vol. 72, nº 1-3, pp. 134-141, 2008.

[58] Pahwa, S., Youssef, M., Schumm, P., Scoglio, C., Schulz, N., «Optimal intentional islanding to enhance the robustness of power grid networks,» *Physica A: Statistical Mechanics and its Applications,* vol. 392, nº 17, p. 3741–3754, 2013.

[59] Rossi, F., Villa-Vialaneix, N., «Optimizing an organized modularity measure for topographic graph clustering: a deterministic annealing approach,» *Neurocomputing,* vol. 73, nº 7-9, pp. 1142-1163, 2010.

[60] Rosset, V., Paulo, M.A., Cespedes, J.G., Nascimento, M.C., «Enhancing the reliability on data delivery and energy efficiency by combining swarm intelligence and community detection in large-scale,» *Expert Systems with Applications,* vol. 78, pp. 89-102, 2017.

[61] Newman, M.E.J., «Modularity and community structure in networks,» *PNAS,* vol. 103, nº 23, pp. 8577-8582, 2006.

[62] Jiang, J.Q., McQuay, L.J., «Modularity functions maximization with nonnegative relaxation facilitates community detection in networks,» *Physica A: Statistical Mechanics and its Applications,* vol. 391, nº 3, pp. 854-865, 2012.

[63] Boseniuk, T., Ebeling, W., «Darwin-and Haeckel-strategies in optimization problems,» de *Proceedings of the International Conference on Parallel Problem Solving from Nature*, Springer, 1991.

[64] Tukey, J.W., «Comparing individual means in the analysis of variance,»

Biometrics, vol. 5, nº 2, pp. 99-114, 1949.

[65] Pereira, M.d.A., Junior, .C.A.D , Gontijo Carrano, E., De Vasconcelos, J.A., «A niching genetic programming-based multi-objective algorithm for hybrid data classification,» *Neurocomputing,* vol. 133, pp. 342-357, 2014.

[66] Agency, I.E., «Large-scale Electricity Interconnection. Technology and Prospects for Cross-regional Networks,» 2016. [En línea]. Available: https://euagenda.eu/upload/publications/untitled-69070-ea.pdf. [Último acceso: 05 05 2018].

[67] Mahmoud, M., Hussain, S., Abido, M., «Modeling and control of microgrid: an overview,» *Journal of the Franklin Institute,* vol. 351, nº 5, pp. 2822-2859, 2014.

[68] Bornholdt, S., Schuster, H.G., Handbook of Graphs and Networks: From the Genome to the internet, New York, NY, USA: John Wiley & Sons, 2003.

[69] Fortunato, S., Hric, D., «Community detection in networks: a user guide,» *Physics Reports,* vol. 659, pp. 1-44, 2016.

[70] Tang, W., Allman, A., Pourkargar, D.B., Daoutidis, P., «Optimal decomposition for distributed optimization in nonlinear model predictive control through community detection,» *Computers & Chemical Engineering,* vol. 111, pp. 43-54, 2018.

[71] Zhou, C., Ding, L., Zhou, Y., Luo, H., «Topological mapping and assessment of multiple settlement time series in deep excavation: a complex network perspective,» *Advanced Engineering Informatics,* vol. 36, pp. 1-19, 2018.

[72] Nascimento, M.C., Pitsoulis, L., «Community detection by modularity maximization using GRASP with path relinking,» *Computers & Operations Research,* vol. 40, nº 12, pp. 3121-3131, 2013.

[73] Chen, Z., Xie, Z., Zhang, Q., «Community detection based on local topological information and its application in power grid,» *Neurocomputing,* vol. 170, pp. 384-392, 2015.

[74] Pagani, G.A., Aiello, M., «The Power Grid as a complex network: A survey,» *Physica A: Statistical Mechanics and its Applications,* vol. 392, nº 11, pp. 2688-2700, 2013.

[75] Sánchez-García, R.J., Fennelly, M., Norris, S., Wright, N., Niblo, G., Brodzki, J., Bialek, J.W., «Hierarchical Spectral Clustering of Power Grids,» *IEEE Transactions on Power Systems,* vol. 29, nº 5, pp. 2229-2237, 2014.

[76] Morton, D.L., «Reviewing the history of electric power and electrification,» *Endeavour,* vol. 26, pp. 60-63, 2002.

[77] Etxegarai, A., Eguia, P., Torres, E., Iturregi, A., Valverde, V., «Review of

grid connection requirements for generation assets in weak power grids,» *Renewable and Sustainable Energy Reviews,* vol. 31, pp. 1501-1514, 2015.

[78] Mahapatra, S., Dasappa, S., «Rural electrification: optimising the choice between decentralised renewable energy sources and grid extension,» *Energy for Sustainable Development,* vol. 16, nº 2, pp. 146-154, 2012.

[79] Schlachtberger, D., Brown, T., Schramm, S., Greiner, M., «The benefits of cooperation in a highly renewable european electricity network,» *Energy,* vol. 134, pp. 469-481, 2017.

[80] Barthelemy, M., «Spatial Networks,» *Physics Reports,* vol. 499, nº 1-3, pp. 1-101, 2011.

[81] Kim, C.J., Obah, O.B., «Vulnerability metrics and analysis for communities in complex networks,» *International Journal of Emerging Electric Power Systems,* vol. 8, nº 6, pp. 1-15, 2007.

[82] Gómez, O., Rios, M.A., «ILP-based multistage placement of PMUs with dynamic monitoring constraints,» *International Journal of Electrical Power & Energy Systems,* vol. 53, pp. 95-105, 2013.

[83] Wang, S., Zhang, J., Zhao, M., Min, X., «Vulnerability analysis and critical areas identification of the power systems under terrorist attacks,» *Physica A: Statistical Mechanics and its Applications,* vol. 473, pp. 156-165, 2017.

[84] Clarke, J., McLay, L., McLeskey, J.T., «Comparison of genetic algorithm to particle swarm for constrained simulation-based optimization of a geothermal power plant,» *Advanced Engineering Informatics,* vol. 28, nº 1, pp. 81-90, 2014.

[85] Saleh, A.I., Rabie, A.H., Abo-Al-Ez, K.M., «A data mining based load forecasting strategy for smart electrical grids,» *Advanced Engineering Informatics,* vol. 30, pp. 422-448, 2014.

[86] Sánchez, P., Montoya, F.G., Manzano-Agugliaro, F., Gil, C., «Genetic algorithm for stransform optimisation in the analysis and classification of electrical signal perturbations,» *Expert Systems with Applications,* vol. 40, nº 17, pp. 6766-6777, 2013.

[87] Thakur, M., Kumar, A., «Optimal coordination of directional over current relays using a modified real coded genetic algorithm: a comparative study,» *International Journal of Electrical Power & Energy Systems,* vol. 82, pp. 484-495, 2016.

[88] Wu, Y., Tang, Y., Han, B., Ni, M., «A topology analysis and genetic algorithm combined approach for power network intentional islanding,» *International Journal of Electrical Power & Energy Systems,* vol. 71, pp. 174-183, 2015.

[89] Guerrero, M.A., Montoya, F.G., Baños, R., Alcayde, A., Gil C., «Adaptive community detection in complex networks using genetic algorithms,» *Neurocomputing,* vol. 266, pp. 101-113, 2017.

[90] Rocco, C.M., Ramirez-Marquez, J.E., «Vulnerability metrics and analysis for communities in complex networks,» *Reliability Engineering & System Safety,* vol. 96, nº 10, pp. 1360-1366, 2011.

[91] ENTSO-E, «Entso-e Transmission System Map,» 2017. [En línea]. Available: https://www.entsoe.eu/ map/Pages/default.aspx. [Último acceso: 05 05 2018].

[92] Bastian, M., Heymann, S., Jacomy,M., «Gephi: an open source software for exploring and manipulating networks,» de *International AAAI Conference on Weblogs and Social Media*, Leuven, Belgium, 2009.

[93] Mazur, A., Metcalfe, T., «America's three electric grids: are efficiency and reliability functions of grid size?,» *Electric Power Systems Research,* vol. 89, pp. 191-195, 2012.

[94] ICSEG, «Texas Synthetic Power Grid Test Case,» 2016. [En línea]. Available: https://icseg.iti.illinois.edu/synthetic-power-cases/texas2000-june2016/. [Último acceso: 05 05 2018].

[95] Jacomy, M., Venturini, T., Heymann, S., Bastian, M., «Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software,» *PloS One,* vol. 9, 2014.

[96] Ponce-Jara, M., Ruiz, E., Gil, R., Sancristóbal, E., Pérez Molina, C., Castro, M., «Smart grid: assessment of the past and present in developed and developing countries,» *Energy Strategy Reviews,* vol. 18, pp. 38-52, 2017.

[97] Wiest, P., Rudion, K., Probst, A., «Efficient integration of (n-1)-security into probabilistic network expansion planning,» *International Journal of Electrical Power & Energy Systems,* vol. 94, pp. 151-159, 2018.

[98] Oboudi, M.H., Hooshmand, R., Karamad, A., «Feasible method for making controlled intentional islanding of microgrids based on the modified shuffled frog leap algorithm,» *International Journal of Electrical Power & Energy Systems,* vol. 78, pp. 745-754, 2016.

[99] Hosseinnezhad, V., Rafiee, M., Ahmadian, M., Siano, P., «Optimal island partitioning of smart distribution systems to improve system restoration under emergency conditions,» *International Journal of Electrical Power & Energy Systems,* vol. 97, pp. 155-164, 2018.

[100] Li, B., Sansavini, G., «Effective multi-objective selection of inter-subnetwork power shifts to mitigate cascading failures,» *Electric Power Systems Research,* vol. 134, pp. 114-125, 2016.

[101] Alqurashi, A., Etemadi, A.H., Khodaei, A., «Treatment of uncertainty for next generation power systems: State-of-the-art in stochastic optimization,» *Electric Power Systems Research,* vol. 141, pp. 233-245, 2016.

[102] Wang, Y., Wang, S., Wu, L., «Distributed optimization approaches for emerging power systems operation: A review,» *Electric Power Systems Research,* vol. 144, pp. 127-135, 2017.

[103] Lin, W.M., Teng, J.H., «Three-phase distribution network fast-decoupled power flow solutions,» *International Journal of Electrical Power & Energy Systems,* vol. 22, nº 5, pp. 375-380, 2000.

[104] Kamel, S., Jurado, F., «Power flow analysis with easy modelling of interline power flow controller,» *Electric Power Systems Research,* vol. 108, pp. 234-244, 2014.

[105] Battaglini, A., Komendantova, N., Brtnik, P., Patt, «Perception of barriers for expansion of electricity grids in the European Union,» *Energy Policy,* vol. 47, pp. 254-259, 2012.

[106] Codetta-Raiteri, D., Bobbio, A., Montani, S., Portinale, L., «A dynamic Bayesian network based framework to evaluate cascading effects in a power grid,» *Engineering Applications of Artificial Intelligence,* vol. 25, pp. 683-697, 2012.

[107] IEA, OECD., «Security of Supply in Electricity Markets,» Evidence and Policy Issues, 2002. [En línea]. Available: https://www.oecd-ilibrary.org/energy/security-of-supply-in-electricity-markets_9789264174504-en. [Último acceso: 2019 11 11].

[108] ENTSO-E, «Ten-Year Network Development Plan 2010-2020,» [En línea]. Available: https://www.entsoe.eu/fileadmin/user_upload/_library/SDC/TYNDP/TYNDP-final_document.pdf. [Último acceso: 2019 11 11].

[109] Cochran, J., Denholm, P., Speer, B., Miller, «Grid Integration and the Carrying Capacity of the U.S. Grid to Incorporate Variable Renewable Energy,» *National Renewable Energy Laboratory: Golden,* 2015.

[110] Espejo, R., Lumbreras, S., Ramos, A., «Analysis of transmission-power-grid topology and scalability, the European case study,» *Physica A: Statistical Mechanics and its Applications,* vol. 509, pp. 383-395, 2018.

[111] MacArthur, B.D., Sánchez García, R.J., Anderson, J.W., «Symmetry in complex networks,» *Discrete Applied Mathematics,* vol. 156, nº 18, pp. 3525-531, 2008.

[112] Xiao, J., Li, X., Chen, S., Wang, Y., Han, J., Zhou, Z., «Complex network measurement and optimization of Chinese domestic movies with internet of things technology,» *Computers & Electrical Engineering,* vol. 66, pp. 246-

259, 2017.

[113] Xiao, Z., Cao, B., Zhou, G., Sun, J., «The monitoring and research of unstable locations in eco-industrial networks,» *Computers & Industrial Engineering,* vol. 105, pp. 234-246, 2017.

[114] He, X., Dong, Y., Wu, Y., Wei, G., Xing, L., Yan, J., «Structure analysis and core community detection of embodied resources networks among regional industries.,» *Physica A: Statistical Mechanics and its Applications,* vol. 479, pp. 137-150, 2017.

[115] Reihanian, A., Feizi-Derakhshi, M.R., Aghdasi, H.S., «Community detection in social networks with node attributes based on multi-objective biogeography based optimization,» *Engineering Applications of Artificial Intelligence,* vol. 62, pp. 51-67, 2017.

[116] Kanavos, A., Perikos, I., Hatzilygeroudis, I., Tsakalidis, A., «Emotional community detection in social networks,» *Computers & Electrical Engineering,* vol. 65, pp. 449-460, 2018.

[117] Nguyen, N.P., Dinh, T.N., Shen, Y., Thai, M.T., «Dynamic social community detection and its applications,» *PLoS ONE,* vol. 9, 2014.

[118] Soliman, S.S., El-Sayed, M.F., Hassan, Y.F., «Semantic clustering of search engine results,» *Sci. World J.,* 2015.

[119] Kicsi, A., Csuvik, V., Vidács, L., Horváth, F., Beszédes, Á., Gyimóthy, T., Kocsis, F., «Feature analysis using information retrieval, community detection and structural analysis methods in product line adoption.,» *Journal of Systems and Software,* vol. 155, pp. 70-90, 2019.

[120] Atay, Y., Koc, I., Babaoglu, I., Kodaz, H., «Community detection from biological and social networks: A comparative analysis of metaheuristic algorithms.,» *Applied Soft Computing,* vol. 50, pp. 194-211, 2017.

[121] Rozylowicz, L., Nita, A., Manolache, S., Popescu, V.D., Hartel, T., «Navigating protected areas networks for improving diffusion of conservation practices,» *Journal of Environmental Management,* vol. 230, pp. 413-421, 2019.

[122] Soofi, S., Ahmed, S., Fox, M.P., MacLeod, W.B., Thea, D.M., Qazi, S.A., Bhutta, Z.A., «Effectiveness of community case management of severe pneumonia with oral amoxicillin in children aged 2–59 months in Matiari district, rural Pakistan: A cluster-randomised controlled trial.,» *Lancet,* vol. 379, pp. 729-737, 2012.

[123] Gómez Núñez, A.J., Batagelj, V., Vargas Quesada, B., Moya Anegón, F., Chinchilla Rodríguez, Z., «Optimizing SCImago Journal & Country Rank classification by community detection,» *Journal of Informetrics,* vol. 8, nº 2, pp. 369-383, 2014.

[124] Carusi, C., Bianchi, G., «Scientific community detection via bipartite scholar/journal graph co-clustering,» *Journal of Informetrics,* vol. 13, nº 1, pp. 354-386, 2019.

[125] Jackson, M.O., «An overview of social networks and economic applications,» de *In Handbook of Social Economics*, Amsterdam, The Netherlands, North-Holland (Elsevier), 2011, pp. 511-585.

[126] Arcia Garibaldi, G., Cruz Romero, P., Gómez Expósito, A., «Future power transmission: Visions, technologies and challenges,» *Renewable and Sustainable Energy Reviews,* vol. 94, pp. 285-301, 2018.

[127] DeMarco, C.L., Baone, C.A., «Control of Power Systems with High Penetration Variable Generation,» de *IRenewable Energy Integration: Practical Management of Variability, Uncertainty, and Flexibility in Power Grids, 2nd ed.*, San Diego, CA, USA, Academic Press (Elsevier), 2017, pp. 385-394.

[128] Wu, H., Krad, I.m Florita, A., Hodge, B.M., Ibanez, E., Zhang, J., Ela, E., «Stochastic multi-timescale power system operations with variable wind generation,» *IEEE Transactions on Power Systems,* vol. 32, pp. 3325-3337, 2016.

[129] Teichler, S.L., Levitine, I., «HVDC transmission: A path to the future?,» *The Electricity Journal,* vol. 23, nº 4, pp. 27-41, 2010.

[130] Chatzivasileiadis, S., Ernst, D., Andersson, G., «The global grid,» *Renewable Energy,* vol. 57, pp. 372-383, 2013.

[131] Guerrero, M.A., Montoya, F.G., Baños, R., Alcayde, A., Gil, C., «Community detection in national-scale high voltage transmission networks using genetic algorithms,» *Advanced Engineering Informatics,* vol. 38, pp. 232-241, 2018.

[132] Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., Wagner, D., «On modularity clustering,» *IEEE Transactions on Knowledge and Data Engineering,* vol. 20, p. 172–188, 2008.

[133] DasGupta, B., Desai, D., «On the complexity of Newman's community finding approach for biological and social networks,» *Journal of Computer & System Sciences,* vol. 79, pp. 50-67, 2013.

[134] Kizys, R., Juan, A.A., Sawik, B., Calvet, L., «A biased-randomized iterated local search algorithm for rich portfolio optimization.,» *Appl. Sci.,* vol. 9, p. 3509, 2019.

[135] Wiegmans, B., «GridKit 1.0 'for Scientists' (Version v1.0),» Zenodo, 2016. [En línea]. Available: https://zenodo.org/record/47263#.XeT6e2ZumiM. [Último acceso: 11 11 2019].

[136] TresAmigas, «Tres Amigas Superstation Project,» 2019. [En línea]. Available: http://www.tresamigasllc.com/. [Último acceso: 11 11 2019].

[137] Arumugam, S., Brandstädt, A., Nishizeki, T., andbook of graph theory, combinatorial optimization, and algorithm, 2016.

[138] Shang, R., Luo, S., Zhang, W., Stolkin, R., Jiao, L., «A multiobjective evolutionary algorithm to find community structures based on affnity propagation,» *Physica A: Statistical Mechanics and its Applications,* vol. 453, pp. 203-227, 2016.

[139] Sun, P. G., «Imbalance problem in community detection,» *Physica A: Statistical Mechanics and its Applications,* vol. 457, pp. 364-376, 2016.

[140] Sun, P. G., Sun, X., «Complete graph model for community detection,» *Physica A: Statistical Mechanics and its Applications,* vol. 471, pp. 88-97, 2017.

[141] Wang, L., Ng, A.H., Deb, K., Multi-objective evolutionary optimisation for product design and manufacturing, Springer, 2011.

[142] Tripathy, B., De-Anonymization Techniques for Social Networks, London: Academic Press an imprint of Elsevier, 2019.

[143] Garey, M.R., Johnson, D.S., Computers and Intractability: A Guide to the Theory of NP-Completeness, New York: W. H. Freeman & Co, 1979.

[144] Emmerich, M.T., Deutz, A.H., «A tutorial on multiobjective optimization: fundamentals and evolutionary methods,» *Natural computing,* vol. 17, pp. 585-609, 2018.

[145] Coello, C.A.C., Lamont, G.B., Van Veldhuizen, D.A., Evolutionary algorithms for solving multi-objective problems, Springer, 2007.

[146] Gong, M., Ma, L., Zhang, Q., Jiao, L., «Community detection in networks by using multiobjective evolutionary algorithm with decomposition,» *Physica A: Statistical Mechanics and its Applications,* vol. 391, pp. 4050-4060, 2012.

[147] Bara'a, A.A., Khoder, H.S., «A new multi-objective evolutionary framework for community mining in dynamic social networks,» *Swarm and Evolutionary Computation,* vol. 31, pp. 90-109, 2016.

[148] Hariz, W.A., Abdulhalim, M.F., «Improving the performance of evolutionary multi-objective co-clustering models for community detection in complex social networks,» *Swarm and Evolutionary Computation,* vol. 26, pp. 137-156, 2016.

[149] Amiri, B., Hossain, L., Crawford, J.W., Wigand, R.T., «Community detection in complex networks: Multi-objective enhanced fireply algorithm,» *Knowledge-Based Systems,* vol. 46, pp. 1-11, 2013.

[150] Niu, X., Si, W., Wu, C. Q., «A label-based evolutionary computing approach to dynamic community detection,» *Computer Communications,* vol. 108, pp. 110-122, 2017.

[151] Bello-Orgaz, G., Salcedo-Sanz, S., Camacho, D., «A multi-objective genetic algorithm for overlapping community detection based on edge encoding,» *Information Sciences,* vol. 462, pp. 290-314, 2018.

[152] Zou, F., Chen, D., Huang, D.-S., Lu, R., Wang, X., «Inverse modelling-based multi-objective evolutionary algorithm with decomposition for community detection in complex networks,» *Physica A: Statistical Mechanics and its Applications,* vol. 513, pp. 662-674, 2019.

[153] Pizzuti, C., «A multiobjective genetic algorithm to find communities in complex networks,» *IEEE Transactions on Evolutionary Computation,* vol. 16, pp. 418-430, 2011.

[154] Gong, M., Cai, Q., Chen, X., Ma, L., «Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition,» *IEEE Transactions on Evolutionary Computation,* vol. 18, pp. 82-97, 2013.

[155] Rahimi, S., Abdollahpouri, A., Moradi, P., «A multi-objective particle swarm optimization algorithm for comcommunity detection in complex networks,» *Swarm and Evolutionary Computation,* vol. 39, pp. 297-309, 2018.

[156] Gong, M., Chen, X., Ma, L., Zhang, Q., Jiao, L., «Identification of multi-resolution network structures with multi-objective immune algorithm,» *Applied Soft Computing,* vol. 13, pp. 1705-1717, 2013.

[157] Zhou, X., Liu, Y., Li, B., Sun, G., «Multiobjective biogeography based optimization algorithm with decomposition for community detection in dynamic networks,» *Physica A: Statistical Mechanics and its Applications,* vol. 436, pp. 430-442, 2015.

[158] Chen, D., Zou, F., Lu, R., Yu, L., Li, Z., Wang, J., «Multi-objective optimization of community detection using discrete teaching-learning-based optimization with decomposition,» *Information Sciences,* vol. 369, pp. 402-418, 2016.

[159] Wu, P., Pan, L., «Multi-objective community detection method by integrating users' behavior attributes,» *Neurocomputing,* vol. 210, pp. 13-25, 2016.

[160] Moayedikia, A., «Multi-objective community detection algorithm with node importance analysis in attributed networks,» *Applied Soft Computing,* vol. 67, pp. 434-451, 2018.

[161] Cheng, F., Cui, T., Su, Y., Niu, Y., Zhang, X., «local information based multi-objective evolutionary algorithm for community detection in complex

networks,» *Applied Soft Computing,* vol. 69, pp. 357-367, 2018.

[162] Li, Q., Cao, Z., Ding, W., Li, Q., «A multi-objective adaptive evolutionary algorithm to extract communities in networks,» *Swarm and Evolutionary Computation,* vol. 52, p. 100629, 2020.

[163] Gao, Y., Zhang, H., Zhang, Y., «Overlapping community detection based on conductance optimization in large-scale networks,» *Physica A: Statistical Mechanics and its Applications,* vol. 522, pp. 69-79, 2019.

[164] Pattanayak, H.S., Sangal, A. L., Verma, H.K., «Community detection in social networks based on fire propagation,» *Swarm and evolutionary computation,* vol. 44, pp. 31-48, 2019.

[165] He, D., Wang, Z., Yang, B., Zhou, C., «Genetic algorithm with ensemble learning for detecting community structure in complex networks,» *IEEE,* pp. 702-707, 2009.

[166] Park, Y., Song, M., «A genetic algorithm for clustering problems,» *Proceedings of the third annual conference on genetic programming,* vol. 1998, pp. 568-575, 1998.

[167] Bosman, P. A., Thierens, D., «The balance between proximity and diversity in multiobjective evolutionary algorithms,» *IEEE transactions on evolutionary computation,* vol. 7, pp. 174-188, 2003.

[168] Menchaca-Mendez, A., Coello, C.A.C., «Selection mechanisms based on the maximin fitness function to solve multi-objective optimization problems,» *Information Sciences,* vol. 332, pp. 131-152, 2016.

[169] Schott, J.R., Fault tolerant design using single and multicriteria genetic algorithm optimization, AFB OH: Technical Report Air Force Inst of Tech Wright-Patterson, 1995.

[170] Taghavi, T., Pimentel, A.D., «Design metrics and visualization techniques for analyzing the performance of moeas in dse,» *International Conference on Embedded Computer Systems: Architectures, Modeling and Simulation,* pp. 67-76, 2011.

[171] Guerrero, M.A., Gil, C., Montoya, F. G., Alcayde, A., Baños. R., «Community detection in power grids by an evolutionary method,» de *GECCO*, Berlin, 2017.