

Práctica entregable

Análisis de redes sociales y Computación Evolutiva y Bioinspirada hito 2



Introducción

En esta práctica detectaremos comunidades en grafos usando algoritmos evolutivos multiobjetivo. Se trabajará con una red con datos reales de Amazon. En dicha red cada nodo es un artículo de Amazon y dos nodos están conectados entre sí cuando suficientes usuarios han comprado los dos artículos juntos. La red cuenta con una estructura de comunidad conocida, en concreto, se conoce la categoría de cada producto que será usada como comunidad. La comunidad a la que pertenece cada nodo está almacenada en el atributo “community” dentro de cada nodo. Respecto al algoritmo, cada grupo trabajará en base al algoritmo genético desarrollado en el hito 1.

A menudo al trabajar con datos reales, las comunidades de la red no son equilibradas. Aquí el número de nodos en cada comunidad:

Comunidad	Número de nodos
4	327
1	42
0	35
2	24
6	16
5	14
7	9
3	8

Objetivo

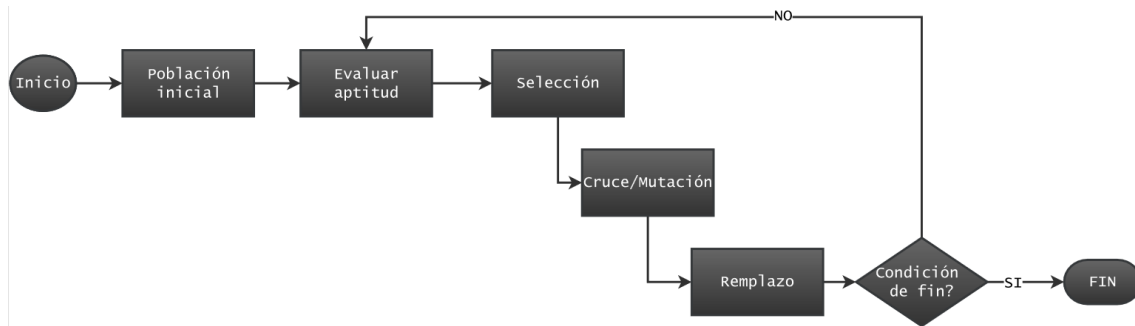
El objetivo de este ejercicio es doble, por un lado, estudiaremos las limitaciones de la Modularidad a la hora de detectar comunidades con tamaños muy desbalanceados; por otro lado, estudiaremos la capacidad de los algoritmos multiobjetivo para analizar una red en distintos niveles jerárquicos.

Ejercicio A – Estructura de comunidad con Leiden

Utiliza el algoritmo de Leiden para optimizar la modularidad y buscar la estructura de comunidad de la red. Compara cuanto se parece la estructura de comunidad encontrada con

Leiden con la estructura real de los grupos de la red ¿La estructura de comunidad encontrada por el algoritmo de Leiden se acerca a la correcta? ¿En qué se diferencia?

Ejercicio B – Algoritmo Evolutivo multiobjetivo



Diseño un algoritmo evolutivo multiobjetivo capaz de detectar comunidades. Debes tener en cuenta las siguientes cuestiones:

- **Codificación:** Locus Adjacency
- **Algoritmo MultiObjetivo:** cada grupo partirá del genético implementado en el hito 1.
- **Objetivos:** Seguir las indicaciones propuestas en “*Comparison and selection of objective functions in multiobjective community detection*” para elegir dos funciones de fitness que den buen resultado al optimizarse conjuntamente.
 - **Objetivo I:** A la elección del alumno
 - **Objetivo II:** A la elección del alumno
- **Método de Selección:** A la elección del alumno, modifique el método de selección o mantenga el ya implementado.
- **Método de Cruce:** A la elección del alumno, modifique el método de cruce o mantenga el que tiene sabiendo que la codificación es entera y se permiten valores repetidos.
- **Método de Mutación:** Sugerido el cambio de alelo por vecino al azar. Para realizar la mutación se selecciona un porcentaje de los alelos de un individuo y se cambia el valor de ese alelo por uno de los vecinos del nodo que representa elegido al azar. También puede probar los que ya tiene implementado.

Nota: tenga en cuenta que el algoritmo debe explorar la mayor parte del frente en lugar de explotar una zona.

- **Reemplazo:** A la elección del alumno, pero se aconseja alguno de los métodos vistos en teoría en los algoritmos MOGA, SPEA y NSGA-II.
- **Condición de parada:** el algoritmo tendrá un presupuesto de 50k llamadas a la función de fitness $f(f1, f2)$.
- **Hiperparámetros:**
 - Al igual que en el hito anterior, tenéis libertad para encontrar el conjunto de hiperparámetros que más le favorezca a vuestro algoritmo.

En este ejercicio se debe incluir la metodología seguida y los experimentos realizados para tunear cada una de las piezas que forman el genético, así como detallar la configuración final elegida.

Ejercicio C – Análisis de las soluciones del Pareto

La principal ventaja del método multiobjetivo es que devuelve un Pareto de soluciones en vez de una única solución, lo que permite analizar la red desde distintos niveles de jerarquía. Esta última sección consiste en analizar el Pareto de soluciones devuelto por el algoritmo. ¿Qué forma tienen las distintas comunidades del Pareto? ¿Hay alguna que se parezcan más a la estructura real de la red? ¿En qué zona del Pareto se localizan?

Normativa de entrega

- Para realizar la entrega se debe realizar una memoria donde se respondan las cuestiones propuestas en los ejercicios A, B y C.
- Además, se debe realizar un repositorio de github publico dónde se incluya todo el código necesario para reproducir los resultados plasmados en la memoria. El repositorio debe contener instrucciones de como ejecutar el código y las dependencias que necesita. No es necesario entregar el código en el Moodle, con incluir un enlace al repositorio en la memoria es suficiente.
- La entrega se realizará en los Moodle de las asignaturas de “Análisis de Redes Sociales” y “Computación Evolutiva y Bioinspirada” en los espacios habilitados para ello. Se entregará la misma práctica en ambas asignaturas, aunque la nota puede variar ya que en cada asignatura se valorará más las partes más relacionadas con su asignatura.
- Cada grupo tendrá que realizar una presentación en la semana 17 (Enero), donde se justifique las decisiones tomadas en el apartado B, así como presentar los resultados y las conclusiones de los Hitos A y C.