



El procesado de grafos es un reto en la gran mayoría de áreas de ciencias, biología, ingeniería e incluso humanidades ya que los grafos permiten representar y, posteriormente, procesar todo tipo de estructuras de datos y relaciones. Sin embargo el análisis de la información relacionada con estos problemas, y representada mediante grafos, se hace menos útil cuando, o bien se tiene un elevado volumen de información o bien se desea extraer gran variedad de posibles relaciones, similitudes, etc., presentes entre los grafos que componen estas grandes bases de datos. Así, los problemas relacionados con el mantenimiento, explotación y minería de grandes bases de datos y problemas relacionados con el isomorfismo de grafos y subgrafos, son actualmente un desafío de los investigadores dado que se requieren nuevos modelos y algoritmos capaces de dar soluciones adecuadas a la "extracción" de información de grandes bases de datos con un coste computacional adecuado.

El objetivo de nuestro proyecto es la propuesta y desarrollo de nuevas soluciones computacionales haciendo uso del isomorfismo de grafos y subgrafos para la construcción, clasificación y minería de grandes bases de datos en las que, entre otros tipos de objetos, la información se representa mediante el uso de grafos.

Dado que la estructura o topología de un grafo determina las características y propiedades de la información que representan y, basándonos en trabajos previos en los que se han desarrollado modelos y algoritmos adecuados para la extracción de características estructurales de los grafos y obtener medidas de similitud basadas en el isomorfismo de subgrafos, en este proyecto, se pretende ampliar y mejorar estas propuestas existentes dado que: (a) el desarrollo de nuevos algoritmos basados en medidas "difusas" de isomorfismo de subgrafos permitirá obtener medidas más fiables de similitud, (b) el uso de medidas de similitud aproximada permitirá la construcción de métodos de clasificación y screening más eficaces para medias y grandes bases de datos (c) estas medidas de isomorfismo aproximado permitirá obtener modelos mejorados de predicción de propiedades d) el uso de estas medidas y de descriptores topológicos permitirá la construcción de bases de datos combinatorias (virtuales) de un alto interés en investigación farmacológica y e) los nuevos algoritmos permitirán disminuir el coste computacional en estos problemas combinatorios y NP-completos.

Nuestros desarrollos darán lugar a un conjunto de sistemas software que serán publicados en un portal Web para su uso, bajo registro, por la comunidad científica. Estos sistemas (cálculo de isomorfismo, clustering, screening y construcción de bases de datos combinatorias) construidos con lenguaje Java, SGBD Oracle y procedimientos en lenguajes nativos (C, PL/SQL, etc.) podrán ser susceptibles de ser registrados o patentados.

Los objetivos del proyecto son:

- Cálculo de similitudes estructurales mediante algoritmos softcomputing de isomorfismo aproximado de grafos / subgrafos.
- 
- Desarrollo de algoritmos basados en satisfacción de restricciones para la generación de bases de datos combinatorias de grafos.
- Desarrollo de un modelo de clasificación multidimensional y evolutivo basado en medidas difusas de similitud—diversidad a partir del isomorfismo aproximado de subgrafos.
- Construcción de una herramienta para HTS de bases de datos.
- Documentación. Difusión de los resultados. Coordinación .