

# Trabajo Práctico 2

Marcelo A. Soria

## Descripción de los datos

Los datos que vamos a utilizar para este trabajo práctico provienen de un estudio de las relaciones interpersonales en un estudio de abogacía de Estados Unidos.

Los datasets y una breve descripción del trabajo y las variables están aquí:

<http://moreno.ss.uci.edu/data.html#lazega>

El archivo **lazega.dat** contiene tres matrices de adyacencia, una a continuación de la otra, y el archivo **lazatt.dat** tiene información extra para cada uno de los individuos que participaron del estudio.

Las matrices de adyacencia describen tres tipos de relaciones interpersonales diferentes. La primera describe las relaciones de colaboración en proyectos laborales. La segunda informa a quien o quienes recurren los profesionales dentro de la firma por consejo profesional. Finalmente, la tercera red describe relaciones de amistad que exceden al ámbito de trabajo.

El objetivo de este trabajo es estudiar las relaciones entre los profesionales de esta firma utilizando grafos. Estos grafos pueden surgir del análisis de las matrices de adyacencia, o de matrices que combinen información de más de una. También se debe utilizar la información extra de los abogados (sexo, edad, universidad de origen, etc.) para caracterizar las comunidades del grafo o para establecer patrones que caractericen a los vértices con mayor o menor centralidad.

## Indicaciones generales para la realización del TP:

- El TP se realizará en grupos de **tres o cuatro** personas, preferentemente manteniendo los grupos del TP1.
- La herramienta recomendada para los análisis es *igraph* (<http://igraph.org/>), en cualquiera de sus implementaciones, R, Python o C (para temerarios).
- La fecha de entrega es el lunes 21 de diciembre a las 24:00.

## Tareas específicas

1. Construir al menos dos grafos a partir de cualquiera de las tres matrices de adyacencia, o creando nuevas matrices que combinen información de al menos dos de las originales. En el caso de matrices combinadas describir qué están tratando de analizar con esa red.
2. Caracterización topológica de las redes (tipo de red, números de nodos, vértices, densidad, análisis de los grados y de medidas de centralidad, etc.).
3. Utilizar la información disponible de los profesionales para determinar si las medidas de centralidad se relacionan con alguna característica personal (edad, años en la empresa, etc)
4. Buscar comunidades en los grafos con al menos dos métodos de búsqueda de comunidades. Determinar si existe alguna asociación entre comunidades y características personales. Esto puede mirarse a nivel de variables individuales o integrando las información de todas las variables mediante un análisis de clusters. Las preguntas a responder son del tipo ¿Existe asociación entre variables personales (edad, sexo, etc.) y las comunidades, o entre pertenencia a clusters y a comunidades?
5. Proponer un escenario donde todo el conocimiento generado en los puntos anteriores pudiera tener una aplicación para este estudio. Por ejemplo, para mejorar el flujo de información, para acercar abogados de más experiencia a los más nuevos o determinar si hay segregación por sexo, edad, universidad de origen. Les puede servir de ayuda ver el libro que escribió uno de los autores sobre este estudio (<http://elazega.fr/wp-content/uploads/2012/10/Lazega-Collegial-Phenomenon-2001-Introduction.pdf>). El autor no utiliza grafos, así que una posibilidad es ver si utilizando grafos pueden confirmar o rechazar algunas de las hipótesis propuestas.

## Importante

**Enviar el informe de este trabajo a [soria@agro.uba.ar](mailto:soria@agro.uba.ar) antes de las 24 hs. del lunes 21 de diciembre. Los trabajos que ingresen después no serán considerados.**