



## Enunciado General del Proyecto Analytics en Redes Sociales

## Tabla de contenido

|  |          |
|--|----------|
| <b>OBJETIVOS .....</b>                 | <b>2</b> |
| <b>CONTEXTO .....</b>                  | <b>3</b> |
| <b>DATOS .....</b>                     | <b>4</b> |
| <b>METODOLOGÍA Y ENTREGABLES .....</b> | <b>5</b> |
| <b>INSUMOS .....</b>                   | <b>8</b> |
| <b>EVALUACIÓN .....</b>                | <b>8</b> |
| <b>REFERENCIAS .....</b>               | <b>8</b> |

## Objetivos

El objetivo general del proyecto es analizar las redes sociales de diferentes inversionistas para caracterizar su comportamiento e identificar los atributos significativos a la hora de financiar compañías. Se espera que los estudiantes sean capaces de aplicar los conocimientos vistos en el curso en un contexto real y llevar a cabo análisis e interpretaciones en el contexto del problema dado, utilizando las herramientas adecuadas.

La siguiente lectura tiene como propósito presentar el enunciado general del proyecto que se desarrollará a lo largo del curso. En primer lugar, se presenta el contexto en el que se desarrolla el proyecto, luego se describe la base de datos con la que se trabajará, después se presentan las fases y entregables y, finalmente, le evaluación de cada fase.

El proyecto se lleva a cabo en grupos cuya conformación es definida al inicio del curso por parte del equipo docente. En los contenidos del curso podrán encontrar la asignación de grupos.

Objetivos del curso relacionados con el proyecto:

1. Identificar la estructura de una base de datos para el análisis de redes.
2. Reconocer y calcular medidas de centralidad y prestigio de redes sociales.
3. Interpretar los resultados de las medidas de centralidad y prestigio de redes sociales.
4. Reconocer las características de los algoritmos de visualización de redes.
5. Interpretar características de las redes a partir de su visualización.
6. Reconocer medidas de transitividad que permiten evaluar la presencia de clusters en la red.
7. Reconocer e implementar algoritmos para la detección de comunidades en redes sociales.
8. Evaluar asociaciones entre las comunidades detectadas en redes, los atributos y las medidas estructurales de los individuos de la red.
9. Reconocer e implementar el modelo estadístico QAP.
10. Interpretar los resultados del modelo estadístico QAP para inferir la relación entre las características de los individuos y la estructura de la red.
11. Reconocer e implementar los modelos estadísticos ERGM y STERGM.
12. Interpretar los resultados de los modelos ERGM y STERGM para establecer la presencia de homofilia en la red

## Contexto

Normalmente las compañías buscan fuentes de financiación y dedican esfuerzos a mejorar sus posibilidades de recibir inversiones (Liang & Yuan, 2016). Uno de los problemas que surgen en esta búsqueda de fuentes de financiación es que existen asimetrías de información. Esto ocurre porque las compañías que buscan financiación tienen información acerca de su potencial y de sus oportunidades que los inversionistas no tienen (Shane & Cable, 2002). En este contexto el análisis de redes sociales es una herramienta que permite analizar el comportamiento de los inversionistas teniendo en cuenta sus relaciones con otros. Esto permite conocer lo que los inversionistas están

buscando y los factores que determinan su comportamiento al momento de financiar empresas (Liang & Yuan, 2016).

## Datos

Para lograr el objetivo del proyecto ustedes cuentan con datos de Crunchbase que contienen la información de un conjunto de organizaciones y sus inversionistas. Los datos están organizados en los siguientes archivos de Excel de la siguiente manera:

- **Base inicial:** este archivo contiene todos los atributos de los nodos de la red que se va a analizar. Un nodo es una organización y cuenta con los siguientes atributos:
  - **Id:** Nombre de la empresa. ○ **Location:** Ubicación de la empresa.
  - **Operating Status:** Estado de operación de la empresa. Puede estar Activa (active) o Cerrada (closed).
  - **Company Type:** Indica si la organización es sin ánimo de lucro (non profit) o con ánimo de lucro (profit). ○ **Number of Investments:** Cantidad total de inversiones realizadas.
  - **Number of Lead Investments:** Cantidad de inversiones realizadas como inversor principal. ○ **Number of Diversity Investments:** Cantidad de inversiones diversificadas realizadas. Cuando se realiza una inversión diversificada se destina el dinero a invertir en diferentes portafolios. ○ **Number of Exits:** Cantidad de veces que en la compañía se han emitido acciones. ○ **Number of Lead Investors:** Cantidad de inversionistas principales en la compañía. ○ **Number of Investors:** Cantidad de inversionistas en la compañía.
  - **Industry Groups:** Grupos de industrias a los cuales pertenecen las organizaciones. La lista de grupos de industrias que podría haber y sus industrias puede [encontrarse aquí](#).
  - **Industries:** Industrias a las cuales pertenecen las organizaciones. La lista de grupos de industrias que podría haber y sus industrias puede [encontrarse aquí](#). ○ **Number of Funding Rounds:** Cantidad de rondas de financiación de la empresa.
  - **Funding Status:** Estado de financiación de la empresa. Puede ser Early Stage Venture, Late State Venture, IPO, M&A, Seed o Private Equity.
  - **Last Funding Date:** Última fecha de financiación.
  - **Last Funding Amount Currency (in USD):** Valor de la última financiación en dólares estadounidenses.

- **Last Funding Type:** Tipo de la última financiación. Los diferentes tipos de financiación pueden [encontrarse aquí](#).
- **Last Equity Funding Type:** Tipo de la última financiación excluyendo la deuda. Los diferentes tipos de financiación pueden [encontrarse aquí](#).
- **Last Equity Funding Amount Currency (in USD):** Valor de la última financiación en dólares estadounidenses excluyendo la deuda.
- **Total Funding Amount Currency (in USD):** Valor total de dinero recaudado en dólares estadounidenses durante todas las rondas de inversión. ○ **Number of Events:** Cantidad total de eventos en los que la organización ha aparecido.
- **SEMrush – Monthly Visits:** Cantidad de visitas al sitio web de la compañía en el último mes incluyendo visitas al sitio móvil y desktop.
- **SEMrush - Average Visits (6 months):** Cantidad promedio de visitas mensuales al sitio web de la compañía en los últimos 6 meses incluyendo visitas al sitio móvil y desktop.
- **SEMrush – Visit Duration:** Tiempo promedio que un usuario permanece en la página web de la compañía por visita. Se mide en segundos.
- **Aberdeen - IT Spend Currency (in USD):** Monto aproximado en dólares que la compañía invierte anualmente en tecnologías de información (TI).
- **Principales inversionistas:** Nombres de los principales inversionistas de la organización.

## Metodología y entregables

El proyecto es transversal al curso y se desarrollará a través de las siguientes fases y entregables durante las 8 semanas. En cada una de las fases el grupo tomará un rol diferente y deberá llevar a cabo diferentes actividades encaminadas a un objetivo en particular. Es recomendable que lea con detenimiento la descripción de la fase, la semana y la entrega.

| Detallado de Fases y Entregas  |  |  |
|--|--|--|
| Fase   | Semana de entrega  | Entregables  |
| <b>Fase 0: Cargue de datos y análisis visual de la red del proyecto en notebook o Gephi</b>      | <b>Semana 2</b><br>En esta fase ustedes serán los responsables en su organización de tomar los datos de Crunchbase y prepararlos para llevar a cabo un análisis visual y descriptivo utilizando el análisis de redes sociales. Se debe realizar la preparación y carga de los datos en un notebook o en Gephi y realizar un análisis visual de la red donde se identifiquen los actores relevantes de la red haciendo uso de las herramientas dadas en el curso hasta el momento: visualización de redes y medidas de centralidad y prestigio. | <b>Entregable:</b><br>En Padlet debe subirse una(s) imagen(es) y un análisis que contenga: <ol style="list-style-type: none"> <li>1. Una o varias imágenes de la red.</li> <li>2. Un párrafo de máximo 500 palabras que contenga un análisis con los <i>insights</i> más importantes obtenidos al visualizar la red y calcular medidas de centralidad y visualizarlas.</li> </ol>  |
| <b>Fase 1: Cargue de datos, análisis visual de la red e identificación de actores relevantes</b> | <b>Semana 3</b><br>Para esta fase, ustedes tendrán su rol original de estudiantes de un curso de análisis de redes sociales y deberán llevar a cabo una caracterización de la red social en la cual estamos trabajando. Para esto deben realizar la carga de los datos, analizarlos visualmente e identificar los actores relevantes de la red utilizando las herramientas dadas en el curso hasta el momento: visualización de datos, cálculo y análisis de medidas de centralidad y transitividad.   | <b>Entregable</b><br>Debe entregarse un informe de máximo 5 páginas donde: <ol style="list-style-type: none"> <li>1. Se documente el proceso de carga de los datos.</li> <li>2. Se realice un análisis visual de la red y se identifiquen los actores relevantes de la red. Para esto se espera que utilicen el trabajo realizado en la fase 0 y las medidas de centralidad y transitividad que consideren relevantes.</li> <li>3. Se realice una conclusión al respecto en el contexto del problema.</li> </ol> |

|  |   |  |
|--|---|--|
| <p><b>Fase 2: Identificación de atributos en común entre los miembros de las comunidades.</b></p>                                  | <p><b>Semana 4</b></p> <p>En esta fase ustedes actuarán bajo el rol de un fondo inversionista que busca comprender el comportamiento inversionista de diferentes organizaciones para mejorar la forma en la cual dispone de sus recursos en diferentes inversiones. En particular, el fondo quiere comprender cuáles son los atributos individuales de las organizaciones que son relevantes en la formación de relaciones de inversión. Para esto, se espera que implementen algoritmos de detección de comunidades y desarrollen un análisis descriptivo de los atributos de los nodos dentro y entre comunidades.</p>  | <p><b>Entregable</b></p> <p>Se debe entregar un notebook donde:</p> <ol style="list-style-type: none"> <li>1. Se realice un análisis de los atributos en común entre los nodos de las comunidades. Para esto, se espera que se utilicen los algoritmos de detección de comunidades que se consideren pertinentes.</li> <li>2. Se concluya al respecto para cada algoritmo seleccionado.</li> <li>3. Se presente una recomendación en el contexto del problema desde el rol de una organización que busca financiación.</li> </ol>                                      |
| <p><b>Fase 3: Estimación de cambios individuales y estructurales usando modelos de inferencia o influencia social en redes</b></p> | <p><b>Semana 7</b></p> <p>En esta fase ustedes actuarán bajo el rol de un fondo inversionista que busca comprender el comportamiento inversionista de diferentes organizaciones para mejorar la forma en la cual dispone de sus recursos en diferentes inversiones. En particular, el fondo quiere comprender cuáles son los atributos individuales de las organizaciones que son relevantes en la formación de relaciones de inversión. Para esto, se espera que se utilicen los modelos estadísticos de inferencia en redes que consideren relevantes (regresión logística QAP, ERGM, STERGM) para determinar los atributos individuales que son relevantes en la formación de lazos de financiación en la red.</p> | <p><b>Entregable</b></p> <p>Debe entregarse un informe de máximo 7 páginas donde:</p> <ol style="list-style-type: none"> <li>1. Se determinen los atributos individuales que son relevantes en la formación de lazos de financiación en la red. Se espera que se utilicen los modelos estadísticos de inferencia que consideren relevantes y justifiquen su elección (regresión logística QAP, ERGM, STERGM).</li> <li>2. Se concluya y se brinden recomendaciones a las empresas sobre qué factores deben tenerse en cuenta para financiar organizaciones.</li> </ol> |

## Insumos

Cada una de las fases cuenta con una guía metodológica detallada que se encuentra en la semana correspondiente a la entrega. Estas guías contienen los enlaces a las bases de datos para el desarrollo del proyecto.

## Evaluación

Cada fase del proyecto será evaluada con los siguientes pesos:

| Fase          | Porcentaje de la nota de cada entrega                        | Semana de entrega |
|---------------|--|-------------------|
| <b>Fase 0</b> | No tiene nota, pero es obligatoria para presentar la fase 1. | Semana 2          |
| <b>Fase 1</b> | 12%  | Semana 3          |
| <b>Fase 2</b> | 12%  | Semana 4          |
| <b>Fase 3</b> | 12%  | Semana 7          |

## Referencias

Liang, Y. E., & Yuan, S. T. D. (2016). Predicting investor funding behavior using crunchbase social network features. *Internet Research*, 26(1), 74–100. <https://doi.org/10.1108/INTR-09-2014-0231/FULL/XML>

Shane, S., & Cable, D. (2002). Network Ties, Reputation, and the Financing of New Ventures. *Management Science*, 48(3), 364–381. <http://www.jstor.org/stable/822571>