

# Análisis de la red de investigadores del IESTA, parte 1

---

Pablo Mones

Ramón Álvarez-Vaz

Serie Documentos de Trabajo

Nº7/21

Mayo, 2021

ISSN: 1688-6453

Universidad de la República.  
Facultad de Ciencias Económicas y de Administración,  
Instituto de Estadística (IESTA)

Montevideo, Uruguay.



Esta obra está bajo una Licencia Creative Commons Atribución - NoComercial - CompartirIgual 4.0 Internacional.

Forma de citación sugerida para este documento:

Pablo Mones, Ramón Álvarez-Vaz.(2021). *Análisis de la red de investigadores del IESTA, parte 1* (Serie Documentos de Trabajo; Nº7/21). Montevideo: Universidad de la República. Facultad de Ciencias Económicas y de Administración, Instituto de Estadística.

<https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/10518>

# Análisis de la red de investigadores del IESTA, parte 1

Pablo Mones <sup>1</sup>

Ramón Álvarez-Vaz <sup>2</sup>;

*Departamento de Métodos Cuantitativos, Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay.*

## Resumen

El presente documento de trabajo constituye un primer avance en el trabajo elaborado para obtener una primera visualización del cuerpo de investigadores del Instituto de Estadística (IESTA) de la Facultad de Ciencias Económicas y Administración (FCEA) de la Universidad de la República mediante el Análisis de Redes Sociales. El objetivo de este trabajo es generar un diagnóstico parcial del modo de trabajo de los investigadores del IESTA, generando un insumo muy importante para el actual Plan Estratégico del IESTA. Se buscarán construir distintas redes de acuerdo a diferentes características de los investigadores en cuanto a la visibilidad de sus diferentes perfiles académicos (Cvuy, ORCID, ResearchGate y Publons). Así como también parte de su producción bibliográfica bajo los formatos de artículos, libros, documentos de trabajo, proceedings, exposición en jornadas académicas de FCEA, exposición en jornadas académicas externas a FCEA, y tutorías de tesis o de becarios.

Finalmente, se deja a disposición de los lectores el código y los datos utilizados para llevar a cabo el análisis en un repositorio de dominio público con el fin de que el mismo sea reproducible (ver <https://gitlab.com/iesta.fcea.udelar/red-de-investigaci-n-del-iesta>).

**Palabras clave:** Análisis de Redes, Perfiles de investigación, Redes Sociales, Detección de comunidades.

**CÓDIGOS JEL:**D85, Y8

**Clasificación MSC2010:** 05C90, 05C85, 62P25, 62H30

---

<sup>1</sup> *email:* pablomones@gmail.com, ORCID: 0000-0002-7875-7148

<sup>2</sup> *email:* ramon.alvarez@fcea.edu.uy, ORCID: 0000-0002-2505-4238

## ABSTRACT

The present working paper constitutes a first advance in order to obtain a first visualization of the researchers working at the Institute of Statistics (IESTA) of the Faculty of Economics and Administration of the Universidad de la Republica, through Social Network Analysis. The objective of this work is to generate a partial diagnosis of the IESTA researchers, generating a very important input for the current Strategic Plan of the Institute. We will seek to build different networks according to the characteristics of the researchers. In terms of their academic visibility, the presence or absence of some academic profiles (Cvuy, ORCID, ResearchGate and Publons) was studied. Moreover, it was also studied their bibliographic production under the formats of articles, books, working papers, proceedings, presentations in FCEA academic conferences, exposition in external academic conferences, and thesis or fellowship tutoring.

Finally, the code and the data used to carry out the analysis are made available to the readers in a public domain repository in order to make it reproducible (see <https://gitlab.com/iesta.fcea.udelar/red-de-investigaci-n-del-iesta>).

**Key words:** Network Analysis, Research Profiles, Social Networks, Community Detection

**JEL CODES:** D85, Y8

**Mathematics Subject Classification MSC2010:** 05C90, 05C85, 62P25, 62H30

## 1. Introducción

En el presente trabajo se buscará estudiar la red de investigadores del Instituto de Estadística (IESTA) de la Facultad de Ciencias Económicas y Administración (FCEA) de la Universidad de la República. El IESTA está compuesto por investigadores del Departamento de Métodos Cuantitativos de la FCEA, a pesar de que varios investigadores trabajan simultáneamente en varias universidades. Cada investigador pertenece a uno o más grupos de investigación dentro del Instituto que evidentemente siguen líneas y objetivos de trabajo distintas. Los investigadores tienen asociados grados académicos que van entre 1 y 5, donde 1 es profesor ayudante y 5 profesor titular, que dan una idea de la trayectoria o importancia que tienen los investigadores dentro de la Universidad de la República.

Los estudios basados en el análisis de redes sociales han ganado relevancia en los últimos tiempos (Cárdenas, 2016). Esto se debe, entre otros factores, a la disponibilidad de mayor cantidad de datos y un aumento exponencial de la velocidad de cálculo de las computadoras, que permite realizar trabajos incorporando gran cantidad de agentes. Las redes sociales se utilizan con el fin de captar las relaciones entre agentes que conforman un mismo sistema. Estas relaciones son la clave del análisis y lo que diferencia esta metodología de otras. Se basa justamente en estudiar las relaciones entre agentes asumiendo que estas pueden ser complejas y no lineales, lo que da lugar a fenómenos de emergencia y retroalimentación no posible en sistemas lineales.

Existen diversos estudios que han utilizado análisis de redes para abordar distintos problemas, como pueden ser Organización Empresarial (Krackhardt y Hanson, 1993), Finanzas (Acemoglu *et al.* (2015) y Tumminello *et al.* (2010)), Sociología (Tindall y Wellman, 2001) y Ciencias Sociales (Jackson, 2010). La capacidad de captar relaciones entre agentes permite estudiar aspectos de un sistema que una mirada “macro” no logra entender, permitiendo también detectar problemas a nivel “micro” en el funcionamiento de un sistema. Asimismo, permite ver como evoluciona un determinado sistema en término de sus agentes (entrada o salida de nodos) y de las relaciones entre ellos (cambios en los aristas), por lo que provee información útil para la toma de decisiones y la evaluación de una determinada medida instaurada.

Este trabajo constituye una mirada estática a la red que permite comprender la situación de la cual se parte y evaluar la evolución de la red tras un lapso de tiempo, o la toma de ciertas decisiones por parte de la dirección y estudiar la eficiencia de las mismas.

De aquí en más, el documento se organiza de la siguiente manera: en la sección 2 se detalla la metodología a utilizar, en la sección 3 se describen las redes a estudiar, el procedimiento para la creación de las mismas y los resultados correspondientes; en la

sección 4 se presentan algunos comentarios, conclusiones y posibles extensiones.

## 2. Metodología

Se define una red o grafo  $G(V,E)$  como un conjunto de nodos ( $V$ ) relacionados entre sí a partir de aristas ( $E$ ), donde  $E \subseteq P(V)$ . En este caso, el conjunto de nodos estará compuesto por los  $N$  investigadores del Iesta siendo este conjunto  $V = \{v_1, v_2, \dots, v_N\}$ . Asimismo, el conjunto de aristas  $E = \{e_{1,1}, e_{1,2}, \dots, e_{N,N}\}$  puede ser pensado de diferentes formas de acuerdo a la característica de esta red que se busque estudiar, obteniéndose distintas redes con los mismos nodos. A modo de ejemplo, si se quisiera estudiar la relación de afinidad entre los diferentes investigadores, evidentemente esta red sería distinta a una que busque captar relaciones de co-autoría, ya que el compartir un trabajo de investigación no necesariamente implica una buena relación personal. En este trabajo, se tomarán dos formas distintas de obtener los aristas. En primera instancia, se buscará estudiar la visibilidad de los investigadores de acuerdo a si estos tienen o no cuatro características: CVuy, ORCID, ResearchGate y Publons. Por otra parte, se buscará estudiar qué tan similares son los perfiles de investigación de los integrantes del Instituto a partir de datos correspondientes a la cantidad de publicaciones de diversos tipos y otras instancias académicas.

Estas redes pueden ser representadas a partir de una matriz de adyacencia  $A_{N \times N}$  (ver Ecuación 1), donde el elemento  $A_{i,j}$  representa el enlace que va desde el nodo  $i$  hacia el nodo  $j$  y  $N$  la cantidad de nodos ( $N = |V|$ ). En caso de que no exista relación entre los nodos, se indica con un 0 en la matriz de adyacencia, por lo que  $A_{i,j} = 0 \Leftrightarrow e_{i,j} \notin E$ , los valores distintos de 0 implican que existe un enlace entre ambos nodos, y el elemento  $A_{i,j}$  de la matriz nos da la intensidad de ese vínculo. En este documento solamente trataremos el caso en el que  $A_{i,j} > 0$ , por lo que si se cumple que  $A_{i,j} < A_{i,k}$  diremos que la relación del nodo  $i$  con el nodo  $j$  es más débil que su relación con el nodo  $k$ .

Para este proyecto, se trabajarán con redes no dirigidas, esto quiere decir que la relación entre el nodo  $i$  y el nodo  $j$  es la misma que la relación entre el nodo  $j$  y el nodo  $i$ . Asimismo, las redes no tendrán bucles, esto quiere decir que no existe relación del nodo  $i$  consigo mismo. En términos de la matriz de adyacencia estas condiciones se pueden expresar como  $A_{i,j} = A_{j,i}$  y  $A_{i,i} = 0$  respectivamente, por lo que la matriz de adyacencia será una matriz simétrica con la diagonal principal nula.

$$A = \begin{pmatrix} 0 & A_{1,2} & \cdots & A_{1,N} \\ A_{2,1} & 0 & \cdots & A_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N,1} & A_{N,2} & \cdots & 0 \end{pmatrix} \quad (1)$$

Sin embargo, para poder construir una red es preciso transformar la información individual de cada investigador en relaciones entre individuos. Por lo tanto, aquí surge una pregunta trascendental que es la base de todo el análisis y conclusiones que puedan obtenerse de este trabajo, ¿cuándo se dice que dos investigadores están “conectados”? Obsérvese que esta pregunta no es trivial. Pensando nuevamente en una red de afinidad, ¿cuándo se puede decir que dos personas se llevan bien? Si consideramos la cantidad de conversaciones que mantienen durante la semana, ¿cuántas son suficientes para establecer si son amigos? Este tipo de preguntas requieren una reflexión sobre el caso a tratar y podría buscarse literatura previa que oriente a como solucionar este problema. En caso de que no exista literatura previa o que lo sugerido por esta no sea conveniente para el problema a estudiar, podría pensarse en establecer diferentes criterios o umbrales para establecer si dos nodos están conectados. Luego, analizar como quedaría la red para distintos rangos de valores, estudiar la evolución de diferentes métricas (entendidas como mediciones de diferentes características de una red) y elegir un valor que se considere conveniente por la forma de la estructura topológica de la red.

Para realizar este análisis se utilizarán métricas que se derivan directamente de la topología de la red (subsecciones 2.1 a 2.4). Asimismo, en la sección 2.5 se estudiará la estructura de comunidades para diferentes valores del umbral. Si bien las comunidades dentro de una red en particular no son una característica intrínseca de la misma (una misma red puede tener diferentes comunidades según el criterio o algoritmo que se utilice); pueden aportar a entender como se agrupan los nodos dentro de la misma. Por lo tanto, también se tendrá en cuenta la evolución de la cantidad de comunidades y la “calidad” de las mismas.

## 2.1. Densidad de la red

La densidad de la red (Ecuación 2) busca cuantificar que tan conectados están los nodos dentro de la red. Esta métrica toma valores en el intervalo  $[0, 1]$  donde valores próximos a 0 implican una red poco conexas y valores próximos a 1 indican que hay gran conexión entre los nodos de la red. Esta métrica se calcula como el cociente entre la cantidad de aristas y la cantidad máxima de aristas posibles, donde para el caso de una red no dirigida y sin auto-bucles es de  $N(N - 1)/2$ . Analíticamente:

$$den(G) = \frac{\text{Cantidad de aristas}}{\text{Cantidad de aristas posibles}} = \frac{|E(G)|}{N(N - 1)/2} \quad (2)$$

## 2.2. Transitividad

La transitividad mide la tendencia de los nodos a formar grupos cohesivos, es decir, grupos de nodos sumamente conectados entre sí mismos. Se calcula como el cociente entre la cantidad de caminos cerrados de largo tres sobre el total de caminos posibles de largo tres y toma valores en el intervalo  $[0, 1]$ . Sean tres nodos,  $i, j$  y  $k$  donde  $e_{i,j} \in E$  y  $e_{j,k} \in E$  una transitividad igual a 1 implica que necesariamente  $e_{i,k} \in E$ , mientras una transitividad igual a 0 implica que  $e_{i,k} \notin E$ . El valor que toma esta métrica puede ser interpretado como la probabilidad de que  $e_{i,k} \in E$  dado que  $e_{i,j} \in E$  y  $e_{j,k} \in E$  (Elsner y Jagger, 2013). Analíticamente:

$$transit = \frac{\# \text{ caminos cerrados de largo 3}}{\# \text{ posible de caminos cerrados de largo 3}} \approx P(e_{i,k} \in E / e_{i,j} \in E; e_{j,k} \in E) \quad (3)$$

Esta métrica fue calculada utilizando el software R (R Core Team, 2019) utilizando la definición inducida por Barrat *et al.* (2004) para grafo ponderados y no dirigidos.

## 2.3. Tamaño del componente gigante

Para poder definir el componente gigante se deben introducir algunos conceptos previos:

- **Grafo conexo:** Un grafo  $G$  se dice conexo si y solamente si para todo  $i, j \in V$  existe un camino que los une.
- **Subgrafo:** Se dice que  $H(V', E')$  es un subgrafo de  $G$  si  $V' \subseteq V$  y  $E' \subseteq E$ .
- **Subgrafo inducido:** Se dice que un grafo  $H(V', E')$  es un subgrafo de  $G$  inducido por  $V'$  y denotamos  $H = G[V']$  si  $H$  contiene a todas las aristas de  $G$ , restringido a los vértices de  $V'$ .
- **Componente conexa:** La componente conexa de un grafo  $G(V, E)$  para un nodo  $i$  denotada como  $G_i$  se define como el subgrafo inducido por el conjunto de vértices  $V_i = \{j : \text{hay un camino de } i \text{ a } j \text{ en } G\}$ . Nótese que un grafo conexo tendrá solamente una componente conexa ( $V = V_i \Rightarrow G = G_i \forall i$ ), mientras que si el grafo no es conexo ( $\exists i$  y  $j \in V /$  no existe camino entre  $i$  y  $j$ ) tendrá al menos dos componentes conexas  $G_i$  y  $G_j$  donde  $G_i \cap G_j = (\emptyset, \emptyset)$  y  $\bigcup_i G_i = G$ .

El componente gigante lo definiremos como el componente conexo con mayor cantidad de nodos.



## 2.4. Distancia promedio

Mide la cantidad de pasos que se deben dar, en promedio, para ir desde un nodo a otro dentro de la red. Esta métrica no toma en cuenta los pesos de los aristas en el momento del cálculo, por lo que supone que la distancia entre dos nodos adyacentes cualesquiera es de 1, mientras que la distancia entre dos nodos no adyacentes viene dada por la longitud del camino más cercano que los une. Esta métrica permite aproximar qué tan rápido fluye la información dentro de la red. En concreto, el valor que toma puede ser interpretado como la cantidad promedio de pasos que lleva ir desde un nodo  $i$  a un nodo  $j$  cualquiera.

## 2.5. Comunidades

Detectar comunidades permite identificar subgrupos de nodos con características similares dentro de la red a estudiar. Existen diversos algoritmos que permiten identificar comunidades dentro de una red y no existe una regla universalmente válida para determinar cual es el mejor. En ese trabajo se utilizarán 2 algoritmos: el algoritmo de *Louvain* (Blondel *et al.*, 2008) y el *fast greedy*. De este último se encuentra una explicación resumida en Álvarez-Vaz y Altmark (2019), para el lector que busque más información el artículo completo se encuentra en Clauset *et al.* (2004).

El algoritmo de Louvain es uno de los algoritmos más utilizados para la detección de comunidades en redes y tiene la bondad de ser muy eficiente en términos computacionales (si se lo compara con los algoritmos que le preceden). Es un algoritmo que busca encontrar comunidades de nodos con la mayor cantidad de enlaces entre ellos y pocos enlaces entre los nodos de diferentes comunidades. Para lograr esto, busca maximizar una métrica llamada modularidad (ver Ecuación 4) la cual toma valores en el intervalo  $[-1, 1]$ , donde valores próximos a 1 indican buena calidad de comunidades (muchos enlaces entre nodos de la misma comunidad y pocos enlaces entre comunidades) y valores próximos a -1 lo contrario.

$$Q = \frac{1}{2m} \sum_{i=1}^N \sum_{j=1}^N \left[ A_{i,j} - \frac{k_i k_j}{2m} \right] \delta(C_i, C_j) \quad (4)$$

donde  $N$  es la cantidad de nodos en la red,  $A_{i,j}$  representa el enlace que va desde el nodo  $i$  al nodo  $j$ ,  $k_i = \sum_{j=1}^N A_{i,j}$  la suma de los aristas relacionados al nodo  $i$ ,  $m = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N A_{i,j}$  (el valor total de los enlaces en la red),  $C_i$  la comunidad del nodo  $i$  y  $\delta(C_i, C_j)$  la delta de Kronecker: la cual vale 1 si  $C_i = C_j$  y 0 en caso contrario.

Sin embargo, optimizar dicha expresión es un problema del tipo NP-hard, por lo cual no puede ser resuelto en un tiempo de computación razonable. A pesar de esto, el algoritmo

propuesto llega a óptimos locales y ha logrado mejores resultados en términos de eficiencia computacional y modularidad que los algoritmos que le preceden.

A continuación, se describirá el procedimiento que sigue este algoritmo. Se parte de una red compuesta por  $N$  nodos, y se comienza asignando cada nodo a una comunidad distinta, por lo cual se tienen inicialmente  $N$  comunidades. Cada iteración del algoritmo está compuesta por dos fases:

### Fase 1

Para cada nodo  $i$  se consideran sus  $j$  vecinos y se evalúa la ganancia de modularidad  $\Delta Q$  (ver Ecuación 5) que surge de asignar el nodo  $i$  a la comunidad del nodo  $j$ . Se asigna  $i$  a la comunidad en la cual la ganancia de modularidad sea máxima, siempre y cuando esta sea mayor que 0. En caso contrario, el nodo  $i$  permanece en su comunidad inicial.

$$\Delta Q = \left[ \frac{\sum_{in} + 2k_{i,in}}{2m} - \left( \frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[ \frac{\sum_{in}}{2m} - \left( \frac{\sum_{tot}}{2m} \right)^2 - \left( \frac{k_i}{2m} \right)^2 \right] \quad (5)$$

donde  $\sum_{in}$  es la suma de los pesos de los enlaces dentro de la comunidad a la cual se está moviendo el nodo  $i$ ,  $\sum_{tot}$  es la suma de todos los pesos de los enlaces a los nodos de la comunidad a la cual se está moviendo  $i$ ,  $k_{i,in}$  es la suma de los pesos del nodo  $i$  con los miembros de su nueva comunidad, mientras que  $k_i$  y  $m$  tienen el mismo significado que el empleado en la Ecuación 4.

Este proceso se realiza recursivamente hasta que no se puedan lograr mejoras en la modularidad, llegando así a un óptimo local. Cabe aclarar que un nodo puede ser evaluado sucesivas veces y trasladado de comunidad en caso de que se produzca una ganancia de modularidad. Asimismo, también es importante notar que los resultados obtenidos en esta fase son sensibles al orden en el cual se tomen los nodos inicialmente, aunque de acuerdo a Blondel *et al.* (2008), los resultados no difieren demasiado al cambiar el orden.

### Fase 2

Una vez completada la etapa anterior, se tiene un conjunto de  $K$  comunidades  $C^1 = \{C_1^1, C_2^1, \dots, C_K^1\}$  las cuales satisfacen:

- $C_i^1 \cap C_j^1 = \emptyset \quad \forall i \neq j$
- $\cup_{k=1}^K C_k^1 = V$  el conjunto de nodos de la red

Luego, se construirá una nueva red tomando como nodos a estas  $K$  comunidades, donde el enlace entre dos comunidades estará dado por la suma de los pesos de los enlaces entre los

nodos que componen las correspondientes comunidades. Denotando por  $A_{C,C'}$  el enlace entre los nodos compuestos por las comunidades  $C$  y  $C'$  y  $n_C$  la cantidad de nodos en la comunidad  $C$  se tiene que:

$$A_{C,C'} = \sum_{i=1}^{n_C} \sum_{j=1}^{n_{C'}} A_{i,j} \quad (6)$$

Esta nueva red tendrá la presencia de auto-bucles, cuyo valor está dado por la sumatoria de los pesos de los enlaces dentro de la propia comunidad, es decir:

$$A_{C,C} = \sum_{i=1}^{n_C} \sum_{j=1}^{n_C} A_{i,j} \quad (7)$$

Luego de completada esta fase es posible pasar a la siguiente iteración, en la cual se repite para esta nueva red hallada el ya descrito procedimiento hasta arribar a la estructura final de comunidades. La Figura 1 presenta de manera esquemática los pasos que sigue este algoritmo:

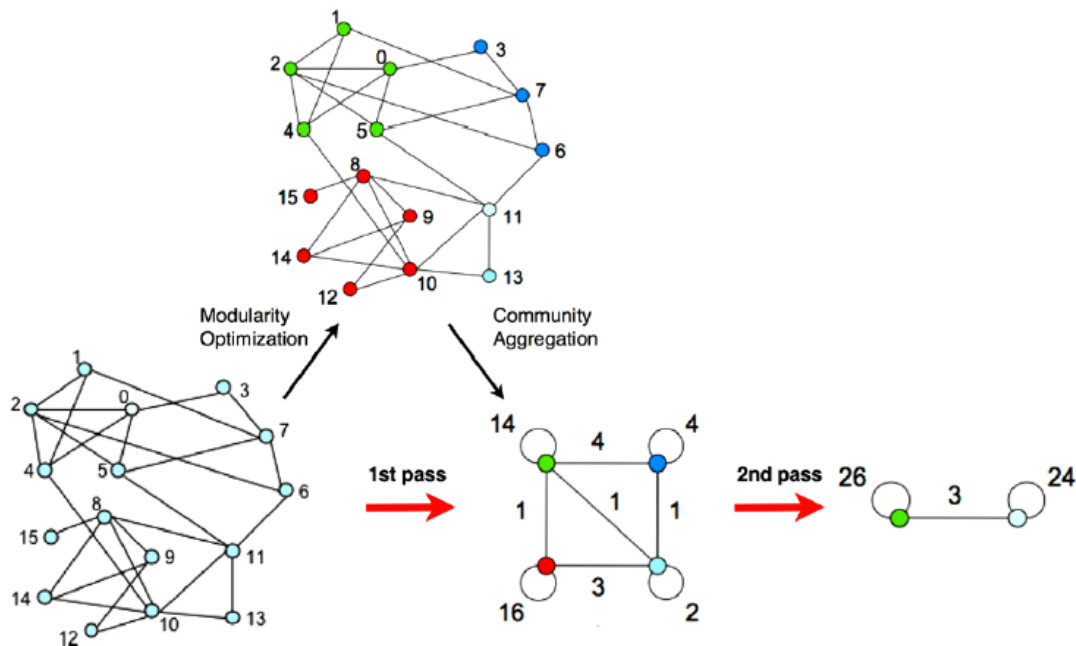


Figura 1: Esquema del algoritmo de Louvain. Fuente: Blondel *et al.* (2008)

Para el proceso de selección del umbral descrito al inicio de esta sección se tomarán la cantidad de comunidades y el valor de la modularidad. Luego, las comunidades obtenidas

tras el filtrado serán analizadas y caracterizadas con el fin de encontrar investigadores con perfiles similares.

### 3. Forma de definición de la red de investigación

Una vez definida la metodología para llevar a cabo el análisis es posible crear las redes a estudiar. Los datos fueron extraídos de la página web del Instituto de Estadística<sup>3</sup> hasta el 30 junio de 2020, por lo tanto, todos los datos incorporados después de dicha fecha no serán considerados. Se cuentan con 31 investigadores que serán los nodos de las redes a estudiar. Como ya se mencionó anteriormente, se realizarán dos redes de investigación distintas, cuyos procedimientos se detallan a continuación.

#### 3.1. Red de visibilidad de perfiles académicos

En este trabajo se entiende por visibilidad el tener perfiles de dominio público que permitan visualizar con facilidad las actividades de investigación o docencia que hayan realizado los integrantes del Instituto. Como ya se mencionó, para armar esta red se disponen de 4 perfiles distintos: CVuy, ORCID, ResearchGate y Publons, por lo cual las variables son del tipo binaria: “tiene” o “no tiene”. Para considerar que un investigador “tiene” un perfil de ORCID, este debe ser de dominio público (de nada sirve un perfil al cual solo puede acceder el propio investigador), mientras que para Publons o ResearchGate es necesario que el investigador aclare que pertenece a la Universidad de la República (o alguna otra institución en la cual trabaje) o en caso contrario que sea posible distinguir sin ambigüedad que se trata del investigador en cuestión. En este caso, es posible obtener la información de los 31 investigadores por lo cual se dispondrán de 31 nodos.

| Grado    | Carga horaria<br>media | Carga horaria<br>Total | Cantidad de<br>nodos | Tienen<br>CVuy u ORCID |
|----------|------------------------|------------------------|----------------------|------------------------|
| Becarios | 20                     | 40                     | 2                    | 2                      |
| 1        | 20.3                   | 122                    | 6                    | 2                      |
| 2        | 25.6                   | 128                    | 6                    | 6                      |
| 3        | 22.1                   | 221                    | 9                    | 6                      |
| 4        | 28.6                   | 86                     | 5                    | 3                      |
| 5        | 34                     | 68                     | 3                    | 2                      |
| Total    | 13.8                   | 665                    | 31                   | 21                     |

Tabla 1: Características de los nodos por grado

<sup>3</sup><http://www.iesta.edu.uy/docentes/>

Ahora bien, ¿cómo definir los enlaces? Para este caso, se tomará una métrica de similaridad entre dos nodos (investigadores) definida como el total de categorías en las cuales los nodos coinciden, es decir:

$$A_{i,j} = \text{Cantidad de categorías en común} \quad (8)$$

Se puede observar que en este caso si dos individuos no poseen ninguno de los perfiles de visibilidad estarán conectados entre si y tendrán el grado de conexión más fuerte posible. Esto es deseable debido a que se entiende que dos individuos son igualmente visibles si no tienen ningún perfil.

Sin embargo, al construirse los enlaces de esta forma basta que compartan una sola categoría para que dos investigadores estén conectados, daría lugar a la formación de redes excesivamente densas. Por esto se impondrá un umbral  $u$  con el fin de filtrar las relaciones de baja intensidad. El nuevo conjunto de enlaces  $E' = \{e'_{1,1}, e'_{1,2}, \dots, e'_{N,N}\}$  quedará definido de la siguiente forma:

$$\begin{cases} A'_{i,j} = \frac{A_{i,j}}{\max_{i,j \in V}(A_{i,j})} & \text{si } A_{i,j} \geq u \\ A'_{i,j} = 0 & \text{si } A_{i,j} < u \end{cases} \quad (9)$$

De esta forma se filtrarán las relaciones débiles y se obtendrá un nuevo conjunto de enlaces cuyos valores de encuentren entre 0 y 1. Los valores razonables para el umbral van entre 1 y 4, lo cual si se piensa en términos de perfiles es equivalente a exigir un determinado número de perfiles en común para decir que dos investigadores están conectados. Este proceso se realizó mediante funciones implementadas en R, a través de la función `filtrado`, la cual se encuentra disponible en el anexo A y en el repositorio <https://gitlab.com/iesta.fcea.udelar/red-de-investigaci-n-del-iesta>.

Para poder hacer comparables los análisis para las diferentes redes elaboradas: de visibilidad (Red1) y de variedad de producción bibliográfica (Red2, Red3, Red4). Los nodos se etiquetan al comienzo de modo que al trabajar con las redes que tienen menos nodos, los números que etiquetan cada nodo son los mismos.

En la Figura 2 se muestra como quedaría conformada la red de investigadores para distintos valores del umbral  $u$ . Cada nodo se corresponde con un investigador del Instituto y los colores de cada nodo indica la cantidad de perfiles que cada investigador posee. Puede verse claramente que a medida que se aumenta el valor del umbral la red se torna menos densa (lo cual ocurre siempre por definición) y para ciertos valores del umbral se forman diversos componentes inconexos. Parecería que en caso de ser muy exigentes con el umbral ( $u = 4$ ) se pierde demasiada información y se genera la presencia de componentes

inconexos del tipo *clique*.

La Figura 2 permite tener una visualización inicial de lo que sucede al filtrar ciertos enlaces de la red, sin embargo, la visualización de la red puede no ser una opción viable en caso de estudiar redes con gran cantidad de nodos. Es por esto que las métricas definidas anteriormente pueden ayudar a comprender de mejor forma lo que esta sucediendo en la red al tomar distintos valores del umbral, sin que el análisis dependa solamente de una percepción visual de la red para distintos valores de  $u$ . Para poder incorporar las ya mencionadas métricas al análisis es preciso calcularlas para las distintas redes que se forman a partir de distintos valores de  $u$ . El paquete *igraph* (Csardi y Nepusz, 2006) del sistema R cuenta con diversas funciones que permiten calcular las métricas propuestas para un grafo particular, luego, mediante una función creada para este proyecto (ver anexo A) se compila la información de dichas métricas para distintos valores del umbral, la cual se muestra en la Tabla 2.

| u | Densidad | Distancia Promedio | Comunidades Louvain | Modularidad Louvain | Comunidades Fast Greedy | Modularidad Fast Greedy | Trans. | Tamaño Compenente gigante |
|---|----------|--------------------|---------------------|---------------------|-------------------------|-------------------------|--------|---------------------------|
| 1 | 0.91     | 1.09               | 2                   | 0.17                | 2                       | 0.17                    | 0.93   | 31                        |
| 2 | 0.72     | 1.28               | 2                   | 0.23                | 2                       | 0.23                    | 0.82   | 31                        |
| 3 | 0.44     | 1.93               | 2                   | 0.36                | 4                       | 0.32                    | 0.72   | 31                        |
| 4 | 0.13     | 1.00               | 9                   | 0.75                | 9                       | 0.75                    | 1.00   | 7                         |

Tabla 2: Métricas de las redes formadas a partir de los perfiles académicos para distintos valores de  $u$

En la Tabla 2 puede verse que la estructura de la red difiere al cambiar el valor del umbral. La densidad decrece de forma monótona al aumentar el valor del umbral, lo cual es algo necesario debido a que aumentos en el valor del umbral eliminan enlaces y no agregan ninguno, aunque puede verse un salto abrupto al pasar de  $u = 2$  a  $u = 3$ . Por otra parte, la red permanece conexa para todos los valores del umbral distintos de 4, lo cual da una idea que ese valor “rompe” la estructura anterior y no sería una buena idea realizar el análisis con dicho valor. Asimismo, los dos algoritmos de detección de comunidades parecen arrojar resultados similares para el rango de valores del umbral, lo cual, en parte, puede deberse a la pequeña cantidad de nodos que se dispone; en este sentido, la cantidad de comunidades parecería ser estable hasta valores de  $u$  menores o iguales a 3.

Por otra parte, puede verse en la Figura 2 que si bien el nodo 16 tiene solo 3 de los 4 perfiles, siempre parecería estar más próximo a los nodos que poseen los 4 perfiles que el resto de quienes tienen solo 3 perfiles. Esto se debe a que el nodo 16 es el único de los nodos con 3 perfiles que tiene Publons, lo cual le permite destacarse de los demás nodos con similares características. Asimismo, si se toma  $u = 3$  los nodos 9 y 21 parecerían estar más próximos a los nodos que poseen tres de los cuatro perfiles que el resto de quienes

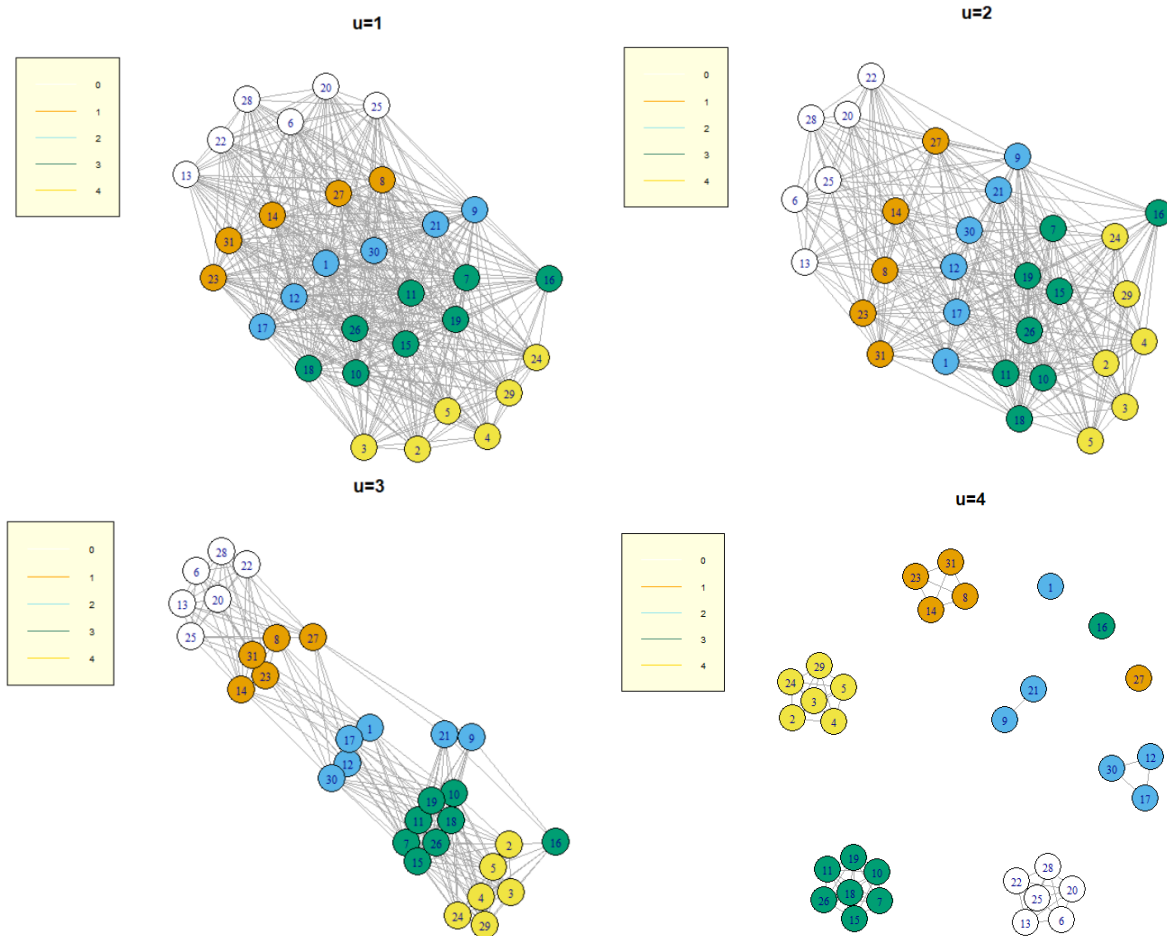


Figura 2: Red de visibilidad según perfiles públicos de investigación (CVuy, ORCID, ResearchGate y Publons) para distintos valores de  $u$

tienen solamente dos. Esto se debe a que poseen ORCID y CVuy en lugar de ResearGate y CVuy (lo que tienen los demás nodos con 2 perfiles) lo cual les permite distinguirse mejor.

Para poder interpretar de mejor manera los resultados con respecto a la transitividad, es útil comparar los resultados de esta red con respecto a los valores teóricos de una red aleatoria de tipo Erdős y Rényi (1959). A modo de ilustrar el fenómeno, la Figura 3 muestra como luce nuestra red para  $u = 3$  en comparación con una red E-R con la misma cantidad de aristas.

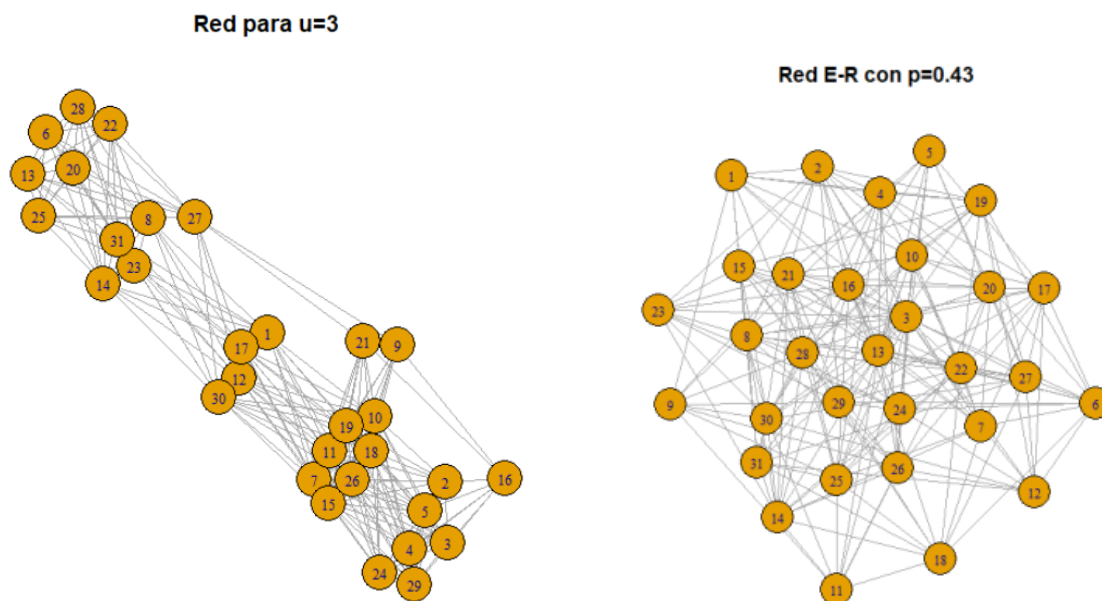


Figura 3: Comparación red empírica vs. red de E-R con la misma cantidad de aristas

Se analizarán las comunidades halladas para  $u = 3$ . En este caso, el algoritmo de Louvain arroja mejores resultados en términos de modularidad que el de *fast greedy* por lo cual se realizará el análisis para la estructura de comunidades halladas por el primero de ellos. Se obtuvieron dos comunidades que se encuentran en la Figura 4. Al analizar las características de los investigadores de cada comunidad, puede verse que la cantidad de perfiles están asociados con las comunidades halladas, todos los nodos con 3 o 4 perfiles están en una misma comunidad y lo mismo sucede con los nodos con 1 o ningún perfil, mientras que los investigadores con 2 perfiles se reparten entre ambas comunidades. Con respecto al grado académico, se da la peculiaridad que todos los grados 2 se encuentran en la comunidad 1, es decir, la de mayor visibilidad, mientras que el resto de los grados no parecerían estar asociados a ninguna de las comunidades.



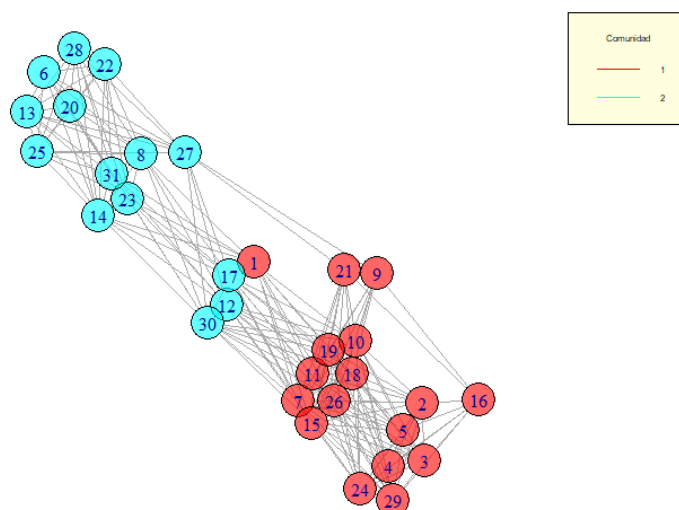


Figura 4: Red de CV para  $u = 3$  coloreada por comunidad de pertenencia de cada nodo

### 3.2. Red de perfiles de investigación

En esta sección buscaremos estudiar cuan similares son los perfiles de investigación de los distintos investigadores que conforman el Instituto de Estadística. Se entiende por perfiles de investigación la forma de producción académica de los diferentes investigadores del instituto. Para esto, se tomarán en cuenta 7 posibles tipos de producción académica: artículos, libros, documentos de trabajo, *proceedings*, exposición en jornadas académicas, exposición en las jornadas académicas de la Facultad de Ciencias Económicas y de Administración y tutorías, ya sea de tesis o becarios. Si bien estos items no tienen igual importancia, en este trabajo no se hará una valoración sobre la importancia de cada uno sino que se buscará ver que investigadores producen de forma similar a otros. En este caso, a diferencia de la situación anterior, no es deseable que situaciones donde ninguno de los investigadores posea un atributo sean consideradas como elementos en común entre ellos. Si se piensa por absurdo, en caso de tener un investigador que solo produce artículos y se busca la relación entre los perfiles de investigación con un investigador joven que aún no ha producido nada se llegaría a la absurda conclusión que estos tienen perfiles de investigación similares, lo cual es erróneo.

Al igual que en el caso anterior, también es posible realizar un proceso de filtrado para eliminar las relaciones débiles (ver anexo A) aunque en este caso hay que tener en cuenta que este proceso dejaría aislados de la red a investigadores que realicen solamente actividades de investigación muy específicas. Volviendo al ejemplo del investigador que solo

produce artículos, un umbral donde se exijan al menos dos categorías en común lo dejaría completamente aislado de la red.

Los datos sobre la producción de los distintos investigadores fueron extraídos de sus respectivos CVuy y ORCID, lo cual dará lugar a dos redes distintas, ya que los investigadores no suelen tener la misma información en ambos perfiles. Para ambos casos se consideró solamente la producción del año 2015 en adelante.

### 3.2.1. Red en base al CVuy

Este perfil suele ser en el que se encuentra la mayor cantidad de información disponible sobre los investigadores. La información extraída incluye datos hasta el 30 de junio del 2020 y se extrajo de los CVuy de los investigadores que lo tienen disponible. En el caso de que los investigadores presenten un CV pero no CVuy, los datos serán extraídos de este primero a efectos de poder disponer de mayor información para efectuar el análisis. Se dispone de información acerca de 23 de los 31 investigadores, por lo que la red a estudiar estará conformada por estos. La Tabla 3 muestra la estructura de la red para diferentes valores de  $u$ .

| $u$ | Densidad | Distancia Promedio | Comunidades Louvain | Modularidad Louvain | Comunidades Fast Greedy | Modularidad Fast Greedy | Trans. | Tamaño Componente gigante |
|-----|----------|--------------------|---------------------|---------------------|-------------------------|-------------------------|--------|---------------------------|
| 1   | 0.86     | 1.06               | 4                   | 0.00                | 3                       | 0.01                    | 0.95   | 22                        |
| 2   | 0.62     | 1.08               | 6                   | 0.01                | 6                       | 0.01                    | 0.94   | 19                        |
| 3   | 0.46     | 1.14               | 8                   | 0.01                | 8                       | 0.01                    | 0.89   | 17                        |
| 4   | 0.34     | 1.27               | 10                  | 0.03                | 9                       | 0.04                    | 0.86   | 16                        |
| 5   | 0.19     | 1.15               | 14                  | 0.03                | 14                      | 0.03                    | 0.89   | 11                        |
| 6   | 0.09     | 1.21               | 17                  | 0.04                | 17                      | 0.04                    | 0.83   | 8                         |
| 7   | 0.01     | 1.00               | 21                  | 0.00                | 21                      | 0.00                    | 1.00   | 3                         |

Tabla 3: Métricas de las redes formadas por información contenida en el CVuy para distintos valores de  $u$

Al igual que en el caso anterior, puede verse que la densidad decae de forma monótona al aumentar el valor de  $u$ . Tampoco se observan diferencias sustantivas entre el desempeño de ambos algoritmos de detección de comunidades, aunque al observar los valores de la modularidad puede verse que no parecería haber una estructura de grupos clara para ningún valor de  $u$  ya que esta métrica siempre toma valores muy próximos a 0. Del mismo modo, la red parece tener una mayor transitividad que una Erdos-Renyi para todos los valores de  $u$ , lo cual es algo esperable debido a la forma en que se construyeron los enlaces. Asimismo, puede verse que a medida que se aumenta el valor de  $u$  se producen quiebres en la red que va sacando nodos del componente principal al no poder estar lo suficientemente

vinculado con los demás.

Para analizar las peculiaridades de esta red más a fondo se trabajará de aquí en más con  $u = 4$ , aunque se prescindirá del análisis de comunidades debido a los pobres resultados que se obtienen en términos de modularidad. La red asociada a dicho umbral se encuentra en la Figura 5 y los colores de cada nodo corresponden al grado que los diferentes docentes tienen en UdelaR, aunque cabe aclarar que por simplicidad los becarios fueron coloreados como grado 1. Puede observarse la presencia de un componente principal donde se encuentran la mayoría de los nodos y otro conjunto de nodos que se encuentran aislados. Estos últimos se corresponden con docentes que han tenido una baja producción académica en los últimos 5 años (o una baja variedad de producción). Asimismo, si bien hay docentes de todos los grados en la periferia de la red, puede verse que la mayoría se corresponde con docentes grado 1 o becarios, lo cual es algo esperable debido a que, usualmente, han ingresado al Instituto recientemente. Los demás docentes que se encuentran desconectados y pertenecen a grados superiores se corresponden con aquellos docentes que han relegado actividades de investigación para dedicarse a la enseñanza y/o extensión. Sin embargo, el nodo 29 está integrado a la parte principal de la red a pesar de ser grado 1, lo que muestra que a pesar de tener un grado bajo ha tenido niveles de producción superiores a los de su mismo grado, lo cual también podría corresponderse al hecho de haber ingresado anteriormente.

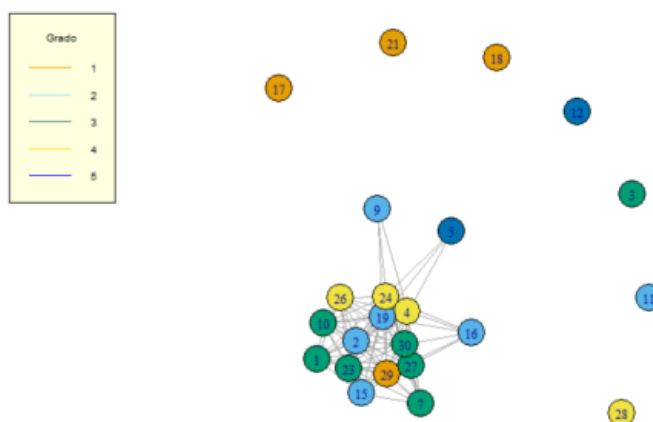


Figura 5: Red de CV para  $u = 4$  coloreada por grado

### 3.2.2. Red en base a ORCID

En esta sección, se muestra la construcción de la red de investigación de la misma forma que en la sección anterior con la única diferencia que los datos serán extraídos de los perfiles de ORCID de los diferentes investigadores en lugar del CVuy. De los 31 investigadores del Instituto, solamente 17 cuentan con perfil público de ORCID, por lo cuál la red se formará a partir de estos investigadores. La Tabla 4 muestra la estructura de la red para distintos valores de  $u$ . A diferencia de las redes de la sección anterior, en este caso el análisis deja de tener sentido para valores de  $u \geq 3$  debido a que se forma una red compuesta únicamente por nodos inconexos, lo cual quiere decir que en la información disponible en ORCID no hay dos investigadores que compartan más de tres categorías de producción. Por tanto, se constata un tendencia a incorporar menos información en el ORCID, especialmente la que es de menos “prestigio” académico, sin embargo, los investigadores suelen incluir sus artículos.

| $u$ | Densidad | Distancia Promedio | Comunidades Louvain | Modularidad Louvain | Comunidades Fast Greedy | Modularidad Fast Greedy | Trans. | Tamaño Compenente gigante |
|-----|----------|--------------------|---------------------|---------------------|-------------------------|-------------------------|--------|---------------------------|
| 1   | 0.55     | 1.30               | 5                   | 0.07                | 5                       | 0.07                    | 0.89   | 15                        |
| 2   | 0.04     | 1.93               | 13                  | 0.21                | 13                      | 0.21                    | 0.38   | 6                         |
| 3   | 0        | -                  | 17                  | 0                   | 17                      | 0                       |        | 1                         |
| 4   | 0        | -                  | 17                  | 0                   | 17                      | 0                       |        | 1                         |
| 5   | 0        | -                  | 17                  | 0                   | 17                      | 0                       |        | 1                         |
| 6   | 0        | -                  | 17                  | 0                   | 17                      | 0                       |        | 1                         |
| 7   | 0        | -                  | 17                  | 0                   | 17                      | 0                       |        | 1                         |

Tabla 4: Métricas de las redes formadas por información contenida en el ORCID para distintos valores de  $u$

### 3.2.3. Red considerando (CVuy+ORCID)

En esta sección se construyen los enlaces considerando conjuntamente la información disponible en el CVuy y el ORCID de los investigadores. La Tabla 5 muestra la estructura de la red para distintos valores de  $u$ . Las diferencias de esta Tabla (5) con la Tabla 3 son ínfimas, lo cual implica que la información en ambos perfiles no difiere casi con la información que se obtendría al considerar solo el CVuy.

## 4. Discusión y conclusiones

Este trabajo es una primera aproximación al análisis de la forma en que los investigadores gestionan su visibilidad académica y el tipo de producción académica en el que se centan.

| u | Densidad | Distancia Promedio | Comunidades Louvain | Modularidad Louvain | Comunidades Fast Greedy | Mod. FG Fast Greedy | Trans. | Compenentes gigantes | Tamaño |
|---|----------|--------------------|---------------------|---------------------|-------------------------|---------------------|--------|----------------------|--------|
| 1 | 0.87     | 1.04               | 4                   | 0.00                | 3                       | 0.01                | 0.96   |                      | 22     |
| 2 | 0.68     | 1.09               | 5                   | 0.01                | 5                       | 0.01                | 0.93   |                      | 20     |
| 3 | 0.46     | 1.14               | 8                   | 0.01                | 8                       | 0.01                | 0.89   |                      | 17     |
| 4 | 0.34     | 1.27               | 10                  | 0.03                | 9                       | 0.04                | 0.86   |                      | 16     |
| 5 | 0.19     | 1.15               | 14                  | 0.03                | 14                      | 0.03                | 0.89   |                      | 11     |
| 6 | 0.09     | 1.21               | 17                  | 0.04                | 17                      | 0.04                | 0.83   |                      | 8      |
| 7 | 0.01     | 1.00               | 21                  | 0.00                | 21                      | 0.00                | 1.00   |                      | 3      |

Tabla 5: Métricas de las redes formadas por información de CVuy y ORCID para distintos valores de  $u$

Respecto a la cantidad de plataformas, donde aparece el CVuy que es un currículum en línea (casi un estándar para el conjunto de investigadores), así como diferentes PID (Identificadores Digitales Persistentes) que a su vez cumplen diferentes funciones. Entre los PID utilizados se encuentran el ORCID (una especie de cédula de identidad a nivel mundial en investigación), el ResearchGate que es una red social académica no controlada y el Perfil Publons que está pensado esencialmente para mostrar la labor como evaluador de pares. La tenencia (o no) de estos perfiles por parte de los investigadores dan origen a la Red 1. Por otra parte, una segunda dimensión elaborada en términos del tipo de producción científica, con la cual se elaboraron otras 3 Redes.

Estas redes fueron construidas con datos de hasta junio del 2020, por lo cual sería factible realizar ejercicios de estática comparativa y comparar la evolución de las diferentes redes luego de un cierto período.

Asimismo, pudo comprobarse que hay distintos niveles de visibilidad dentro de los investigadores del instituto, aunque aún hay varios investigadores que no poseen ningún perfil académico de dominio público que permita ver su desempeño académico. Esto puede ser problemático debido a que pueden existir investigadores que hayan realizado diferentes actividades pero que estas no sean visibles por el resto, dificultándose la difusión de las mismas.

Finalmente, parecería que los investigadores suelen depositar distinta información en el CVuy y en el ORCID. El ORCID suele ser utilizado para algunos tipos de trabajos puntuales, especialmente los artículos, que suelen ser más prestigiosos. Es decir, que parecería verse al ORCID como un perfil donde publicar poca información, pero clave, un lugar donde se pueda ver rápidamente un resumen del trabajo del investigador. En cambio, en el CVuy parecería incorporarse la mayor cantidad de información posible (incluida la del ORCID) y es un perfil que permite ver más a fondo lo que hace cada investigador, pero que sin embargo es local.

A modo de síntesis, pueden establecerse diferentes objetivos a lograr en el corto y mediano plazo:

- Hay investigadores que aún no tienen a disposición ninguno de los 4 perfiles usados para generar el grafo de visibilidad de perfiles (algo a cambiar rápidamente, ya que se estipuló durante el primer semestre que cada investigador debe tener CVuy y ORCID).
- Hay casi un tercio de los investigadores que no tienen ORCID (en condiciones, es decir con contenido visible), lo que sugiere un aspecto a mejorar y lograr que se entienda la relevancia de este PID como de los más importantes
- La Red 2 para ORCID indica que la estrategia de promoción de la producción no está bien visibilizada. Muchos investigadores lo toman como forma de mostrar esencialmente los artículos, tal vez porque el ORCID permite en forma automática registrarlos, ya que muchas revistas indexadas tienen interoperabilidad con el ORCID. Esta conducta puede deberse a comodidad o desconocimiento de la cantidad de tipos de publicación que se pueden consignar en este sistema y no en CVuy; por lo tanto es otro aspecto a trabajar.

Se espera poder trabajar en este segundo semestre con datos obtenidos en un corte a junio de 2021, donde se pueda ver como han evolucionado las diferentes redes, esperando que se hayan actualizado los perfiles. Asimismo, podrían incorporarse las siguientes extensiones para poder obtener un análisis más completo:

- Considerar el perfil de Google Académico (Google Scholar)
- Estudiar métricas de homofilia (assortativity) que puedan mostrar asociación entre las comunidades y la membresía de cada investigador a los grupos de investigación que ya tiene el IESTA
- Considerar otro tipo de redes que incorporen dimensiones relativas a aspectos de participación y acceso a fondos concursables y de aspectos de formación (grado alcanzado y en qué área de los Métodos Cuantitativos)
- Evaluar cómo incorporar a la red de investigadores del IESTA otras redes académicas de las cuales algunos integrantes del Instituto son miembros, y viceversa (cómo considerar investigadores externos al IESTA que son parte de los grupos de investigación del IESTA).

## 5. Anexo

### A. Funciones

```

library(readxl)
library(igraph)
library(CINNA)
library(xtable)
library(dplyr)

#Importacion de datos
datos <- read_excel("datos_red_iesta.xlsx")
datos[c(18,21),2]=c('1','1') #A efectos de visibilidad, se agrupa a los becarios (filas 18 y 21) como grado 1
dim(datos) #31 individuos

###Grafo segn visibilidad---
datos_v=datos[,c(1,3:6)] #Docente y caractersticas a estudiar

#Hay que elegir un umbral en base a las coincidencias

#u es el valor del umbral y dat es un data.set
filtrado=function(u,dat){###Filtro de "relaciones" debiles
n=nrow(dat)
m=ncol(dat)
puntaje=matrix(0,nrow = n,ncol=m) #Adj Matrix(nxn)
for (i in 1:n) {for(j in 1:m){
coincidencias=sum(dat[i,c(2:m)]==dat[j,c(2:m)]) #Cantidad de categorias coincidentes
if((coincidencias>u)&(i!=j)){#Me fijo si las coincidencias superan el umbral (u) y que i sea dist. de j
puntaje[i,j]=sum(dat[i,c(2:m)]==dat[j,c(2:m)])
#El puntaje es la cantidad de coincidencias
}
}}
puntaje=puntaje/max(puntaje)} #El enlace mas fuerte valdra 1

#rango=vector con rango de valores para los cuales se quiere evaluar el umbral (ej: rango=c(1,10))
#f es una funcion que realice el filtrado (ej: filtrado o filtrado2)
#dat es un data.set

estructura=function(rango,f,dat){#Funcion para analizar redes para distintos umbrales
NDS=c() #Network density
DP=c() #distancia promedio
CC_L=c() #cantidad de comunidades con Louv
MOD_L=c() #Modularity con Louv
CC_FG=c() #cantidad de comunidades con FG
MOD_FG=c() #Modularity con FG
TRS=c() #Transitivity
NRP=c() #Nodos en el componenete gigante

for (u in min(rango):max(rango)) {#Itero para el rango de valores en u
adj_m=f(u,dat) #La matriz de adyacencia se obtiene tras realizar el filtrado
g=graph_from_adjacency_matrix(adj_m,weighted = T, mode = "undirected") #Construyo el grafo
louv=cluster_louvain(g) #Deteccion de comunidades con el algoritmo de Louvain
greed=fastgreedy.community(g) #Deteccion de comunidades con el algoritmo fast greedy

NDS=c(NDS,edge_density(g)) #Network density
DP=c(DP,mean_distance(g,directed=F)) #distancia promedio
CC_L=c(CC_L,length(louv)) #cantidad de comunidades Louv
MOD_L=c(MOD_L,modularity(louv)) #Modularity Louv
CC_FG=c(CC_FG,length(greed)) #cantidad de comunidades FG
MOD_FG=c(MOD_FG,modularity(greed)) #Modularity FG
TRS=c(TRS,transitivity(g)) #Transitivity
NRP=c(NRP,dim(get.adjacency(giant_component_extract(g, directed = FALSE)[[1]])))[1]
}
evol=cbind(NDS,DP,CC_L,MOD_L,CC_FG,MOD_FG,TRS,NRP)
}

sim_v=estructura(c(1,4),filtrado,datos_v) ###Valores de distintas metricas para distintos valores de u

#Agrego columna con cdad de perfiles que tienen los investigadores
datos_v2=datos_v
for (i in 1:31) {
datos_v2$sum[i]=sum(datos_v2[i,2:5])}

#Grafos de red de visibilidad de perfiles academicos para distintos
#valores de u coloreados por cantidad de perfiles disponibles

#u1
grafo1=graph_from_adjacency_matrix(filtrado(1,datos_v),mode="undirected",weighted = TRUE)%>%set.vertex.attribute("Docente",value=datos_v$Etiqueta)

```

```

%>%set.vertex.attribute("CVuy",value=datos_v$CVuy)%>%set.vertex.attribute("ORCID",value=datos_v$ORCID)%>%set.vertex.attribute("RN",value=datos_v$Researchergate_net)
%>%set.vertex.attribute("Publons",value=datos_v$Publons)%>%set.vertex.attribute("Suma",value=datos_v2$sum)
set.seed(20)
plot(grafo1, vertex.color=vertex.attributes(grafo1)$Suma,layout=layout_(grafo1, nicely()),main="u=1",vertex.label.cex=c(0.75))
legend("topleft", c("0","1","2","3","4"), col=c("white","orange","cadetblue2","aquamarine4","gold"),lty=1,bg="lightyellow", cex=0.6)
#u2
grafo2=graph_from_adjacency_matrix(filtrado(2,datos_v),mode="undirected",weighted = TRUE)%>%set.vertex.attribute("Docente",value=datos_v$Docente)
%>%set.vertex.attribute("CVuy",value=datos_v$CVuy)%>%set.vertex.attribute("ORCID",value=datos_v$ORCID)%>%set.vertex.attribute("RN",value=datos_v$Researchergate_net)
%>%set.vertex.attribute("Publons",value=datos_v$Publons)%>%set.vertex.attribute("Suma",value=datos_v2$sum)
set.seed(400)
plot(grafo2, vertex.color=vertex.attributes(grafo2)$Suma,layout=layout_(grafo2, nicely()),main="u=2",vertex.label.cex=c(0.75))
legend("topleft", c("0","1","2","3","4"), col=c("white","orange","cadetblue2","aquamarine4","gold"),lty=1,bg="lightyellow", cex=0.6)
#u3
grafo3=graph_from_adjacency_matrix(filtrado(3,datos_v),mode="undirected",weighted = TRUE)%>%set.vertex.attribute("Docente",value=datos_v$Docente)
%>%set.vertex.attribute("CVuy",value=datos_v$CVuy)%>%set.vertex.attribute("ORCID",value=datos_v$ORCID)%>%set.vertex.attribute("RN",value=datos_v$Researchergate_net)
%>%set.vertex.attribute("Publons",value=datos_v$Publons)%>%set.vertex.attribute("Suma",value=datos_v2$sum)
set.seed(91)
plot(grafo3, vertex.color=vertex.attributes(grafo3)$Suma,layout=layout_(grafo3, nicely()),main="u=3",vertex.label.cex=c(0.75))
legend("topleft", c("0","1","2","3","4"), col=c("white","orange","cadetblue2","aquamarine4","gold"),lty=1,bg="lightyellow", cex=0.6)

#u4
grafo4=graph_from_adjacency_matrix(filtrado(4,datos_v),mode="undirected",weighted = TRUE)%>%set.vertex.attribute("Docente",value=datos_v$Docente)
%>%set.vertex.attribute("CVuy",value=datos_v$CVuy)%>%set.vertex.attribute("ORCID",value=datos_v$ORCID)%>%set.vertex.attribute("RN",value=datos_v$Researchergate_net)
%>%set.vertex.attribute("Publons",value=datos_v$Publons)%>%set.vertex.attribute("Suma",value=datos_v2$sum)
set.seed(20)
plot(grafo4, vertex.color=vertex.attributes(grafo4)$Suma,layout=layout_(grafo4, nicely()),main="u=4",vertex.label.cex=c(0.75))
legend("topleft", c("0","1","2","3","4"), col=c("white","orange","cadetblue2","aquamarine4","gold"),lty=1,bg="lightyellow", cex=0.6)

#Comparacion con red Erdos-Renyi (ER)

##Nuestra red
set.seed(91)
plot(grafo3, layout=layout_(grafo3, nicely()),main="Red para u=3",vertex.label.cex=c(0.75))
##Red ER
set.seed(10)
ER=erdos.renyi.game(31,p=.43)
plot(ER,main="Red E-R con p=0.43",vertex.label.cex=c(0.75))

#Análisis de comunidades para u=3 - se usa Louv xq tiene + Modul

Clus_louv3=cluster_louvain(grafo3)
Perteneencia3=membership(Clus_louv3)
CantidadComunidades3=length(Clus_louv3)
set.seed(91)
plot(grafo3, vertex.color=rainbow(CantidadComunidades3, alpha=0.6)[Clus_louv3$membership])
legend("topright",c("1", "2"),title="Comunidad", col=c(2,5),lty=1,bg="lightyellow", cex=0.51)

Com=data.frame(cbind(datos[,1:2],as.vector(Perteneencia3))) #Data.frame con pertenencia de nodos a comunidades

###Grafo segn CVuy----
datos_c=datos[,c(1,7:13)] #Datos de CVuy
datos_c=datos_c[which(datos_c$'Libros - C'>-1),] #Forma "particular" de sacar los NA

filtrado2=function(u,dat){
n=nrow(dat)
m=ncol(dat)
puntaje=matrix(0,nrow = n,ncol=m) #Adj Matrix(nxn)
for (i in 1:n) {for(j in 1:m){
if((i!=j)){aux=sum((dat[i,2:m]>0)&(dat[j,2:m]>0)) #Cdad. de veces que i y j tienen 1 item en comun
if(aux>u){puntaje[i,j]=aux} #Si esa cantidad es mayor o igual a u lo incluyo
}}
if(max(puntaje)!=0){puntaje=puntaje/max(puntaje)} #El enlace mas fuerte valdra 1
return(puntaje)
}
}
sim_c=estructura(c(1:7),filtrado2,datos_c)

#Grafo CVuy para u=4 coloreado por grado academico
grafo2=graph_from_adjacency_matrix(filtrado2(4,datos_c),mode="undirected",weighted = TRUE)%>%set.vertex.attribute("Docente",value=datos_c$Etiqueta)
%>%set.vertex.attribute("Grado",value=datos_c[which(as.numeric(rownames(datos)))/in/unlist(datos_c[,1]),2]$GRADO)
set.seed(20)
plot(grafo2,vertex.label=(as.numeric(unlist(datos_c[,1]))),vertex.color=vertex.attributes(grafo2)$"Grado",vertex.label.cex=c(0.75))
legend("topleft", c("1","2","3","4","5"), col=c("orange","cadetblue2","aquamarine4","gold","blue"),lty=1,bg="lightyellow", cex=0.6,title="Grado")

###Grafo segn ORCID----
datos_o=datos[,c(1,14:20)] #Datos ORCID
datos_o=datos_o[which(datos_o$'Libros - O'>-1),] #me saco los NA

sim_o=estructura(c(1:7),filtrado2,datos_o) #Redes de ORCID para distintos valores de u

#ORCID+CVuy

```



```
datos2=datos[-c(6,8,13,14,20,22,25,31),] #datos a utilizar
datos2[c(10,13,17,20,21,23),c(14:20)]=0 #Los NA los toma como 0
datos2=mutate(datos2,Articulos_2=ifelse((datos2[,7]+datos2[,14])>0,1,0),Libros_2=ifelse((datos2[,8]+datos2[,15])>0,1,0),Proceedings_2=ifelse((datos2[,9]+datos2[,16])>0,1,0),
datos2=datos2[,c(1,21:27)] #Datos a utilizar

sim_2=estructura(c(1:7),filtrado2,datos2) #Incluyo la 20 xq el codigo empieza a contar desde la col. 2
#Muy similar a sim_c
```

## Referencias Bibliográficas

- Acemoglu, D., Ozdaglar, A., y Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.
- Álvarez-Vaz, R. y Altmark, S. (2019). Estudio del gasto en turistas de cruceros en Uruguay para la temporada 2010-2011 mediante el análisis de redes. *Cuadernos del CIMBAGE*, 1(21):27–64.
- Barrat, A., Barthelemy, M., Pastor-Satorras, R., y Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the national academy of sciences*, 101(11):3747–3752.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., y Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.
- Cárdenas, J. (2016). El análisis de redes: qué es, orígenes, crecimiento y futuro. *Pensando Psicología*, 12(19):5–10.
- Clauset, A., Newman, M. E., y Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70(6):066111.
- Csardi, G. y Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.
- Elsner, J. B. y Jagger, T. H. (2013). *Hurricane climatology: a modern statistical guide using R*. Oxford University Press.
- Erdős, P. y Rényi, A. (1959). On random graphs publ. *Math. debrecen*, 6:290–297.
- Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.
- Krackhardt, D. y Hanson, J. R. (1993). Informal networks. *Harvard business review*, 71(4):104–111.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Tindall, D. B. y Wellman, B. (2001). Canada as social structure: Social network analysis and Canadian sociology. *Canadian Journal of Sociology/Cahiers canadiens de sociologie*, pp. 265–308.
- Tumminello, M., Lillo, F., y Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of economic behavior & organization*, 75(1):40–58.

# Instituto de Estadística

---

## Serie Documentos de Trabajo



FACULTAD DE  
CIENCIAS ECONÓMICAS  
Y DE ADMINISTRACIÓN

**IESTA** INSTITUTO  
DE ESTADÍSTICA



UNIVERSIDAD  
DE LA REPÚBLICA  
URUGUAY

Gonzalo Ramirez 1926, Piso 1, Oficina 23 - C.P. 11200 -  
Montevideo, Uruguay  
Teléfono: (598) 2410 2564  
<https://iesta.fcea.udelar.edu.uy/>  
Área Publicaciones

Mayo, 2021

Nº7/21