

# Evaluación de las y los investigadores del IESTA, mediante el Análisis de Redes

## Parte 1

Pablo Mones y Ramón Álvarez-Vaz

SIESTA  
Universidad de la República

21/04/2021

1 **Introducción**

2 Metodología

3 Aplicación

4 Comentarios y trabajo a futuro

## Introducción

- Tratar de ver como I@s investigador@s del IESTA mostramos nuestra producción a través de algunos perfiles académicos
- Aplicación de análisis de redes para representar al conjunto de investigador@s del IESTA
- Dos tipos de redes distintas
  - Visibilidad de perfiles académicos
  - Variedad de producción bibliográfica
- Permitirá establecer diagnóstico basal para ayudar el equipo de dirección del Instituto para fomentar estos hábitos de divulgación

## Introducción

¿Por qué utilizar análisis de redes?

- Permite centrarse en las **relaciones entre agentes (investigador@s)**
- Se pueden detectar problemas o fallas en estas relaciones
- Es útil para estudiar la evolución de un sistema tanto en términos de sus componentes (nodos) como de las relaciones entre éstos (aristas)
- Permite hacer fácilmente estática comparativa. ¿Cómo cambia la red ante un *shock* o una nueva propuesta de la dirección?

1 Introducción

**2 Metodología**

3 Aplicación

4 Comentarios y trabajo a futuro

# Metodología (1)

Análisis de redes: (Newman, 2010)

- Grafo o red:  $G(V,E)$
- $V = \{v_1, v_2, \dots, v_N\}$  Conjunto de investigadores del IESTA
- $E = \{e_{1,1}, e_{1,2}, \dots, e_{N,N}\}$  Puede ser pensado de diferentes formas de acuerdo a las características del sistema que se quiera estudiar
- Estas redes pueden ser representadas a partir de una matriz de adyacencia  $A_{N \times N}$ 
  - $A_{i,j}$  es el enlace que va desde el nodo  $i$  al nodo  $j$
  - $A_{i,j} = 0 \Leftrightarrow e_{i,j} \notin E$
  - Para este estudio, las redes serán simétricas ( $A_{i,j} = A_{j,i}$ ) y sin auto-bucles ( $A_{i,i} = 0 \forall i \in \{1, 2, \dots, n\}$ )

## Metodología (2)

- Ahora bien, ¿cuándo diremos que dos investigador@s están conectad@s? → Los resultados del trabajo son sensibles a la respuesta a esta pregunta
- Responder esta pregunta requiera una reflexión sobre el problema a tratar. Tres posibles abordajes:
  - 1 Revisar literatura previa
  - 2 Elección *ad hoc* en base a algún criterio
  - 3 **Establecer diferentes criterios o umbrales para determinar si dos nodos están conectados**

## Metodología (3)

- Para cada una de las dimensiones que se quieren estudiar, se construirán varias redes con diferentes exigencias en términos de la relación entre nodos
- Para cada nivel de “exigencia” se calcularán ciertas métricas que permitan resumir aspectos relevantes acerca de la topología de la red
- Se elegirá la red con una estructura “deseable”



## Metodología (4)

### Métricas: 1) Densidad de la red

- Busca cuantificar qué tan conectados están los nodos dentro de la red
- Toma valores en el intervalo  $[0, 1]$

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

## Metodología (5)

### Indicadores: 2) Transitividad

- Tendencia de los nodos a formar grupos cohesivos
- El valor que toma esta métrica puede ser interpretado como la probabilidad de que  $e_{i,j} \in E$  dado que  $e_{i,k} \in E$  y  $e_{j,k} \in E$  (Elsner and Jagger, 2013)

$$Trans = P(e_{i,j} \in E / e_{j,k}, e_{i,k} \in E) \quad (2)$$

## Metodología (6)

### Indicadores: 3) Tamaño del componente gigante

- El componente gigante lo definiremos como la componente conexa con mayor cantidad de nodos.
- Un grafo es conexo si cada par de vértices está conectado por un camino; es decir, si para cualquier par de nodos  $(a, b)$ , existe un enlace desde  $a$  hacia  $b$ .

## Metodología (7)

### Indicadores: 4) Distancia promedio

- Distancia promedio entre dos nodos cualesquiera de la red
- No toma en cuenta el peso de las aristas

$$\bar{d} = \frac{1}{n(n-1)} \sum_{i,j} \sum_{i \neq j} d_{i,j} \quad (3)$$

## Metodología (8)

### Algoritmo de detección de comunidades

- Se utilizó (principalmente) el algoritmo de Louvain
- Se particiona al conjunto de nodos  $V = \{v_1, \dots, v_n\}$  en  $k$  comunidades que satisfacen:
  - $\bigcup_i^k C_i = V$
  - $C_i \cap C_j = \emptyset \quad \forall i \neq j \in \{1, 2, \dots, k\}$
- El algoritmo obtiene las comunidades de forma tal de maximizar la modularidad ( $Q$ ) de la partición

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

## Metodología (9)

### Algoritmo de detección de comunidades

Dos fases:

#### 1 Fase 1

- Para cada nodo  $i$  se consideran sus  $j$  vecinos y se evalúa la ganancia de modularidad ( $\Delta Q$ ) que surge de asignar el nodo  $i$  a la comunidad del nodo  $j$

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

- Se repite este proceso hasta que no se puedan obtener más ganancias de modularidad

#### 2 Fase 2

- Finalizada la fase 1 se tiene un conjunto de  $k_1$  comunidades
- Se construirá una nueva red tomando como nodos a estas  $k_1$  comunidades y como aristas:  $A_{c,c'} = \sum_{i=1}^{n_c} \sum_{j=1}^{n_{c'}} A_{i,j}$
- Se repite la Fase 1

## Metodología (10)

### Algoritmo de detección de comunidades

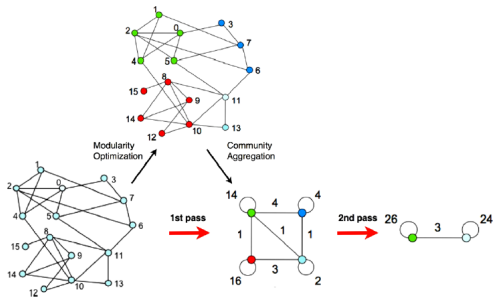


Figura 1: Esquema del algoritmo de Louvain. Fuente: (Blondel et al., 2008)

## Contra que comparar (Benchmark): redes Erdos-Renyi

- Modelo simple de formación de grafos aleatorios
- Dado un conjunto de nodos  $V = \{v_1, v_2, \dots, v_N\}$  la probabilidad de formación de enlaces viene dada por:

$$P(e_{i,j} \in E) = p \quad \forall i, j \in V$$

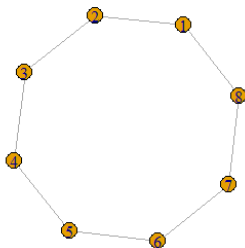
- Es fácil demostrar que para una red de tipo Erdos-Renyi:

$$Trans = P(e_{i,j} \in E / e_{j,k}, e_{i,k} \in E) = P(e_{i,j} \in E) = p \quad \forall i, j \in V$$



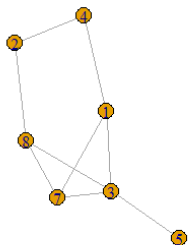
## Benchmark: redes Erdos-Renyi (Erdős and Rényi, 1959)

Red anillo



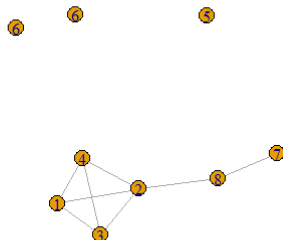
$p=0,29$  /  $trans=0$

Red Erdos-Renyi



$p=0,29$  /  $trans=0,35$

Red cohesiva



$p=0,29$  /  $trans=0,75$

Figura 2: Redes con la misma cantidad de nodos y enlaces pero distinta transitividad

1 Introducción

2 Metodología

**3 Aplicación**

4 Comentarios y trabajo a futuro

## Aplicación a la red de investigadores del instituto (1)

Red de visibilidad de perfiles académicos

Se trabaja en el sistema R (R Core Team, 2020), usando librería *igraph* (Csardi and Nepusz, 2006)

Docente	etiqueta	GRADO	Cvuy	Orcid	Researcher gate net	Publons	Artículos - C	Libros - C	Proceedings - C	DT - C	JIAA fact - C	Otras jornadas - C	Tutor - C
1		3	0	1	1	0	2	0	0	2	4	5	0
2		2	1	1	1	1	4	0	2	2	2	15	2
3		3	1	1	1	1	4	0	0	0	0	0	0
4		4	1	1	1	1	18	4	15	11	16	22	5
5		5	1	1	1	1	36	1	0	0	0	4	9
6		4	0	0	0	0							
7		3	1	1	1	0	4	0	2	2	0	7	0
8		3	0	0	1	0							
9		2	1	1	0	0	0	2	0	0	6	7	1
10		3	1	1	1	0	4	0	0	2	1	18	2
11		2	1	1	1	0	0	0	5	1	0	0	0
12		5	1	0	1	0	1	0	0	2	0	0	2
13		1	0	0	0	0							
14		1	0	0	1	0							
15		2	1	1	1	0	1	0	2	2	1	11	0
16		2	1	1	0	1	5	2	5	0	0	1	0
17		1	1	0	1	0	0	0	0	0	0	2	0
18	Becario		1	1	1	0	0	0	0	0	0	1	0
19		2	1	1	1	0	7	1	1	1	2	26	3
20		5	0	0	0	0							
21	Becario		1	1	0	0	1	0	0	1	0	0	0
22		1	0	0	0	0							0

Figura 3: Matriz de Datos

## Aplicación a la red de investigadores del instituto (2)

### 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”
- Se construirá una red (Red 1) basándonos en si los investigadores tienen o no 4 perfiles distintos de un total más amplio:
  - 1 CVuy -Curriculum Vitae de la ANII (casi un estándar hoy en día para CV)  
<https://cvuy.anii.org.uy/>
  - 2 Orcid (PID) que otorga cédula de identidad como investigad@r a nivel mundial  
<https://orcid.org/>
  - 3 ResearchGate - Red académica no controlada pero muy usada  
<https://www.researchgate.net/>
  - 4 Publons- perfil que entre otras características muestra actividad como revisor de artículos científicos <https://publons.com/about/home/>
- La información se extrajo de la página web del IESTA 06/2020

## Aplicación a la red de investigadores del instituto (3)

Red de visibilidad de perfiles académicos

Grado	Carga horaria media	Carga horaria Total	Frecuencia	Tienen 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

Cuadro 1: Características de los nodos por grado

## Aplicación a la red de investigadores del instituto (4a)

### Red de visibilidad de perfiles académicos

- En una primera instancia, diremos que dos investigadores estarán conectad@s de acuerdo a la cantidad de coincidencias en los perfiles académicos, es decir:  
 $A_{i,j} = \text{Categorías en común}$
- Sin embargo, esto daría lugar a redes excesivamente densas
- Se impondrá un umbral  $u$  con el fin de filtrar las relaciones de baja intensidad
- Los nuevos enlaces estarán dados por:

$$\begin{cases} A'_{i,j} = \frac{A_{i,j}}{\max_{i,j \in V}(A_{i,j})} & \text{si } A_{i,j} \geq u \\ A'_{i,j} = 0 & \text{si } A_{i,j} < u \end{cases} \quad (6)$$

- Para elegir el mejor valor de  $u$  se obtuvieron las redes para el conjunto de valores posibles  $u = 1, 2, 3, 4$

## Aplicación a la red de investigadores del instituto (4b)

Red de visibilidad de perfiles académicos: Función de filtrado

```
#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=n) #Adj Matrix(nxn)
for (i in 1:n) {for(j in 1:n){
if((sum(dat[i,c(2:m)]==dat[j,c(2:m)])>=u)&(i!=j)){#Si coinciden u veces o m
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 ms fuerte vale 1
```

## Aplicación a la red de investigadores del instituto (5)

Red de visibilidad de perfiles académicos

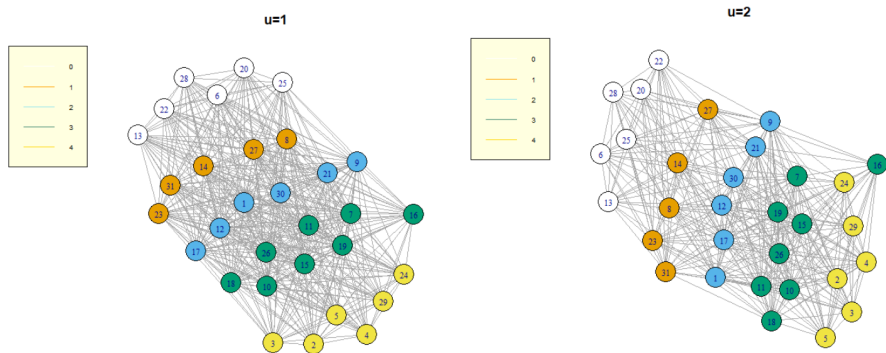


Figura 4: Red de visibilidad según perfiles públicos de investigación (CVuy, Orcid, ResearchGate y Publons) para distintos valores de  $u$  coloreados por cantidad de perfiles disponibles



## Aplicación a la red de investigadores del instituto (6)

Red de visibilidad de perfiles académicos

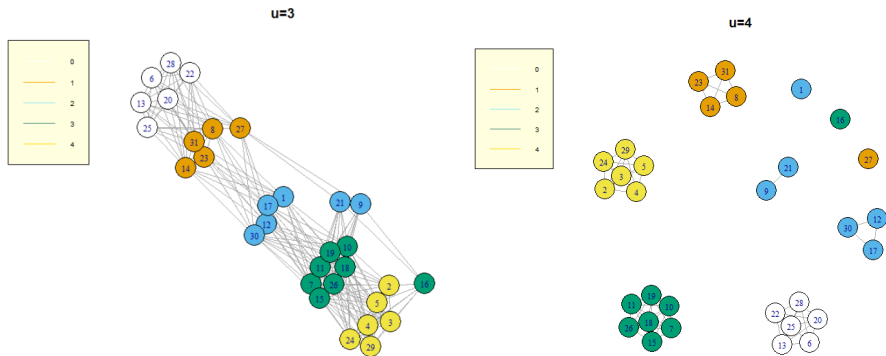


Figura 5: Red de visibilidad según perfiles públicos de investigación (CVuy, Orcid, ResearchGate y Publons) para distintos valores de  $u$  coloreados por cantidad de perfiles disponibles

## Aplicación a la red de investigadores del instituto (7)

Red de visibilidad de perfiles académicos

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

## Aplicación a la red de investigadores del instituto (8)

Red de visibilidad de perfiles académicos

Para  $u = 3$  se detectaron 2 comunidades:

$u=3$

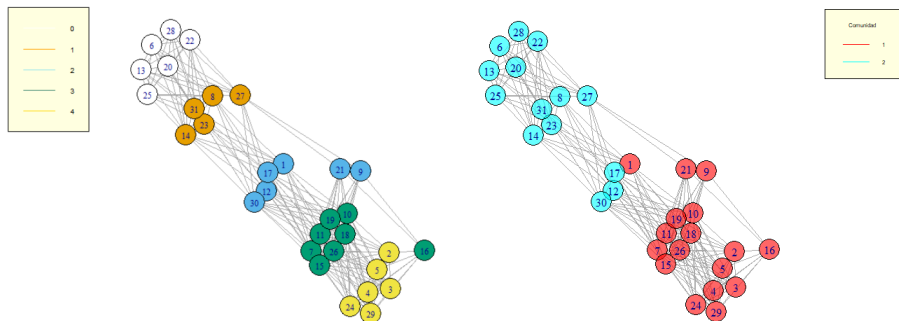


Figura 6: Red de visibilidad para  $u = 3$  coloreada por cantidad de perfiles y comunidad de pertenencia

¿Qué es lo que está segregando las comunidades? Orcid

## Red de perfiles de investigación en producción bibliográfica (1)

### Construcción

- “Se entiende por perfiles de investigación la forma de producción académica de los diferentes investigadores del instituto”
- Se analizarán 7 tipos de producciones distintas, aunque no se hará una valoración sobre la importancia de cada una, sino que buscaremos estudiar la similitud en la forma de actividad académica:
  - 1 Artículos
  - 2 Participación en libros
  - 3 Documentos de Trabajo
  - 4 *Proceedings*
  - 5 Exposición en jornadas académicas
  - 6 Exposición en las jornadas académicas de la Facultad de Ciencias Económicas y de Administración
  - 7 Tutorías de tesis o becarios
- Se tomaron datos desde 2015 hasta junio de 2020

## Red de perfiles de investigación en producción bibliográfica (2)

### Construcción

- A diferencia de la red anterior, entenderemos que dos investigadores están conectad@s si comparten algún (¿cuántos?) tipo de publicación en común
- Posible problema: ¿Qué pasa con l@s investigador@s que solo producen un tipo de documento?
- Estos datos fueron extraídos de:
  - ④ CVuy (Red 2)

## Red de perfiles de investigación en producción bibliográfica (2)

### Construcción

- A diferencia de la red anterior, entenderemos que dos investigadores están conectad@s si comparten algún (¿cuántos?) tipo de publicación en común
- Posible problema: ¿Qué pasa con l@s investigador@s que solo producen un tipo de documento?
- Estos datos fueron extraídos de:
  - 1 CVuy (Red 2)
  - 2 Orcid (Red 3)
- ¿Qué pasa con los investigadores que no tienen ninguno de estos perfiles o los tienen desactualizados?

## Red de perfiles de investigación en producción bibliográfica (3)

Red 2: CVuy

- L@s investigador@s del IESTA suelen depositar aquí la mayor cantidad de información acerca de sus actividades de investigación
- Se dispone información acerca de 23 de l@s 31 investigador@s del Instituto

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

**Cuadro 2:** Métricas de las redes formadas por información contenida en el CVuy para distintos valores de  $u$

## Red de perfiles de investigación en producción bibliográfica (4)

Red 2: CVuy



Figura 7: Red de CV para  $u = 4$  coloreada por grado



## Red de perfiles de investigación en producción bibliográfica (5)

Red 2: CVuy

- Presencia de nodos segregados ¿por qué?
- Un grupo central de investigador@s más activ@s
- Solo hay un grado 1 dentro de ese grupo

## Red de perfiles de investigación en producción bibliográfica (6)

### Red 3: Orcid

- Solo 17 de los 31 investigador@s tenían perfil de Orcid público
- Las redes dejan de tener sentido para  $u > 2$

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

**Cuadro 3:** Métricas de las redes formadas por información contenida en el Orcid para distintos valores de  $u$

1 Introducción

2 Metodología

3 Aplicación

4 Comentarios y trabajo a futuro

## Comentarios finales

- Al momento de realizar el estudio, habían investigador@s que no tenían ninguno de los 4 perfiles o no estaban en “condiciones mínimas”
- L@s investigador@s suelen depositar distinta información en el CVuy y en el Orcid
- Posibles problemas de difusión/visibilidad de trabajos

## Trabajo a futuro

- Repetir el ejercicio y observar posibles cambios en las redes luego de casi 10 meses, con una nueva página web, más actualizada y con grupos de investigación ya más conformados
- Impartir un taller donde explicar claramente las ventajas del uso de los 4 perfiles (su potencial y utilidad) y la necesidad de tenerlos actualizados
- Definir los grupos de investigación  
<https://iesta.fcea.udelar.edu.uy/grupos-investigacion/>, como unidad de análisis ya que hay multiplicidad
- Estudiar redes de coautorías o cooperación

## Referencias Bibliográficas

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.
- Elsner, J. B. and Jagger, T. H. (2013). *Hurricane climatology: a modern statistical guide using R*. Oxford University Press.
- Erdős, P. and Rényi, A. (1959). On random graphs publ. *Math. debrecen*, 6:290–297.
- Newman, M. (2010). *Networks: an introduction*. Oxford University Press.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.