



# **Inferencia de Redes ARACNE**

**Alejandra Yamamoto**



# Outline

## 1. Introducción

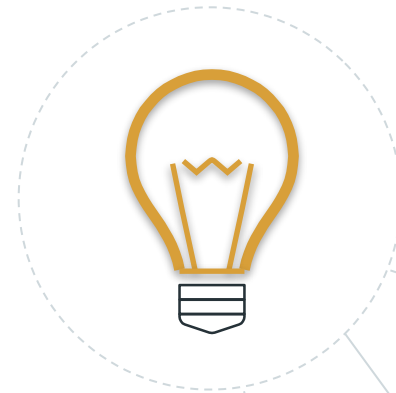
- Motivación
- Objetivo
- Métodos

## 2. RN & ARACNE

- Fundamentos
- Hipótesis
- Estimaciones
- Filtrados

## 3. Resumen

## 4. Práctica



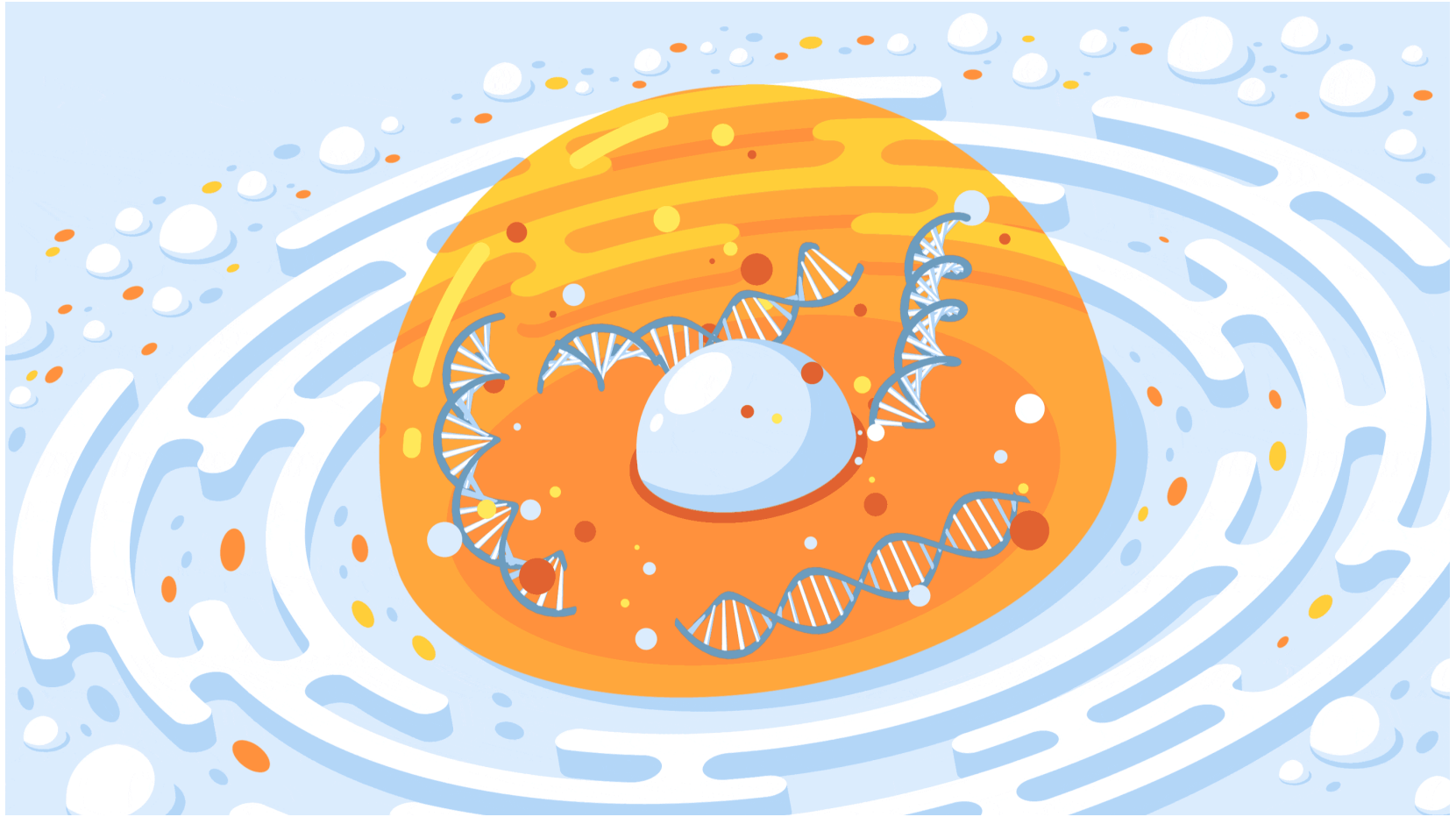


1.

# Introducción

Inferencia de redes

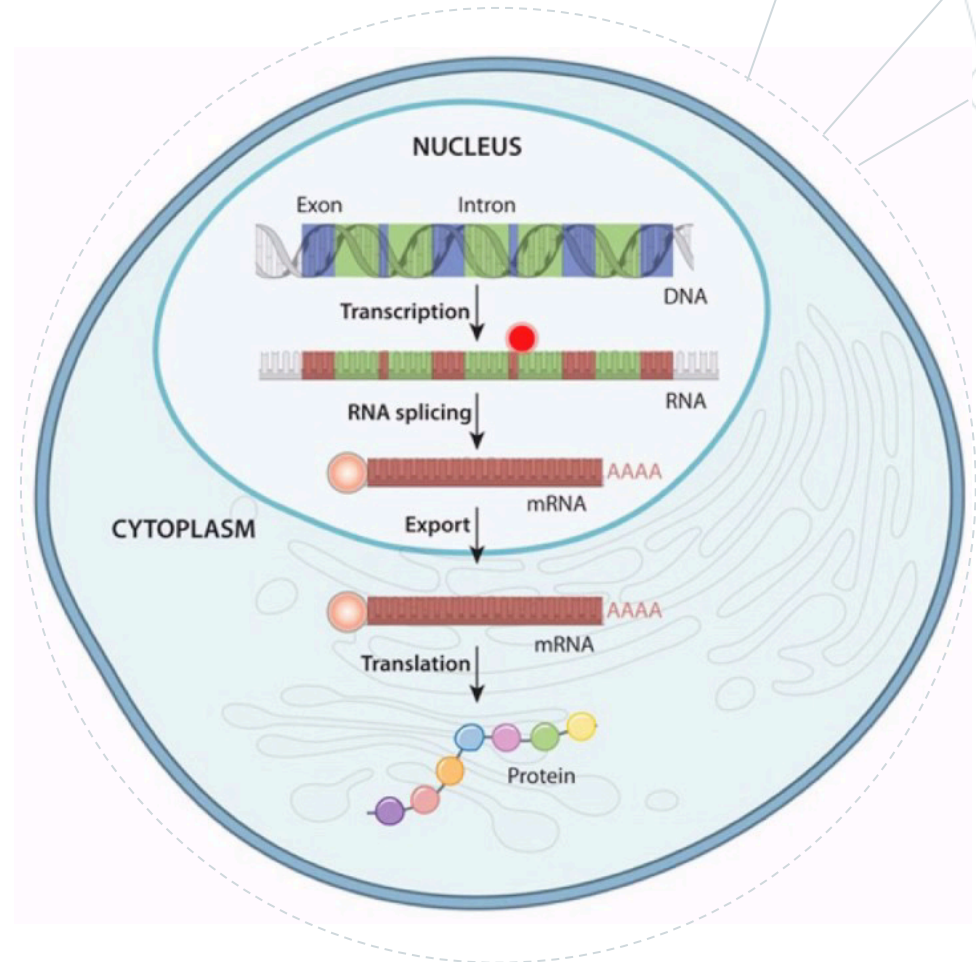
## Motivación



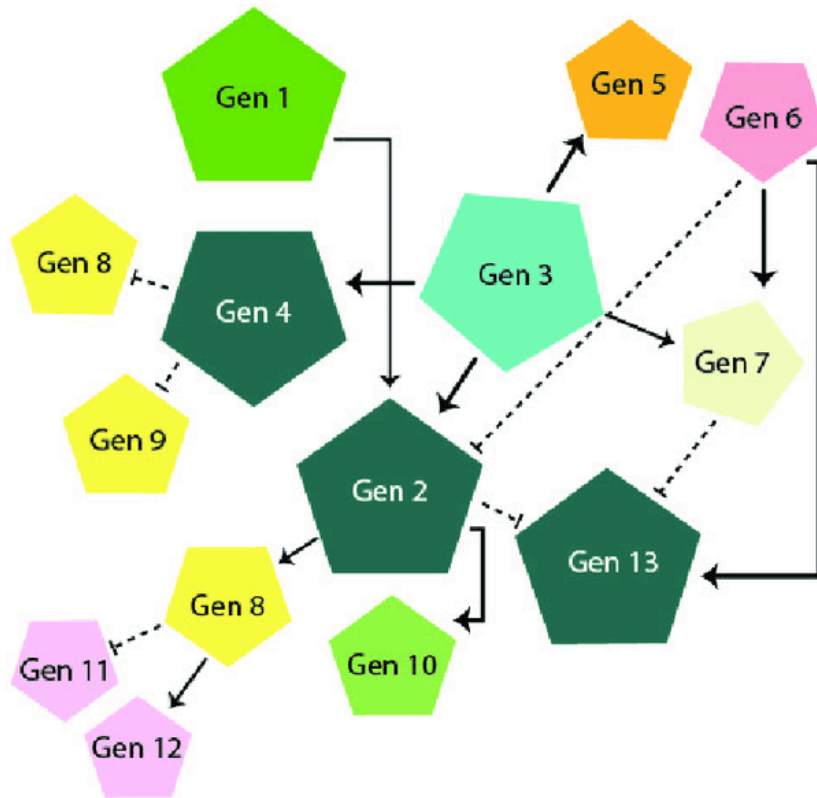
Organismos celulares

## Motivación

- Proceso de regulación
- Influencias de genes
- Construcción de modelos
- Redes de regulación génica (GRN)



## ¿Por qué redes?



- Interacción
- Tasa de producción de proteínas
- Estudio de factores de transcripción
- Nodos → Genes
- Enlaces → Relación entre TF y genes
  - Activación
  - Inhibición
- Medidas de centralidad



# Objetivo

¿Cuál es la forma de la red de regulación transcripcional?

¿Tipos de datos a estudiar?

¿Qué métodos existen para la inferencia de redes?

¿Características?

## Métodos Disponibles

### Supervisado

- Conocimiento a priori
- Datos conocidos
- Función capaz de predecir resultados.
- **SIRENE** (Supervised Inference of Regulatory Networks)

### No supervisado

- Parte de propiedades de los datos
- Búsqueda de la caracterización de datos
- Datos de entrada se consideran variables aleatorias



## Métodos No Supervisados

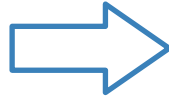
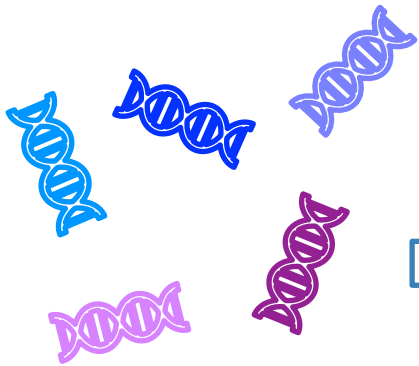
### Ejemplos

Basado en MI

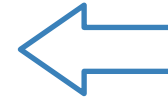
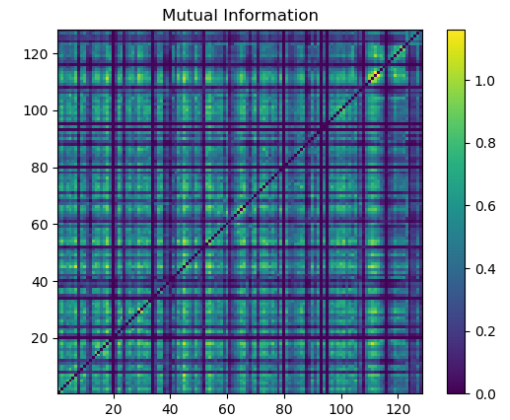
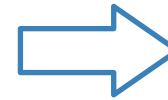
- **RN** (Relevance Networks)
- **ARACNE** (Algorithm for the Reconstruction of Accurate Cellular Networks)
- **MRNET** (Minimum Redundancy/Maximum Relevance Networks)
- **CLR** (Context Likelihood Relatedness)

Basado en correlación de Pearson

- **CORRELATIONS**
- **PCIT** (Partial Correlation and Information Theory)
- **WGCNA** (Weighted Gene Co-expression Network Analysis)
- **GENIE3** (Gene Network Inference with Ensemble of Trees)



	S1	S2	S2	S3	S4
G1					
G2					
G3					
G4					





2.

# Relevance Networks

Butte & Kohane (2000)

# RN & ARACNE

## Fundamentos

Reconstrucción de redes  
de co-expresión génica

Basado en teoría de la  
información

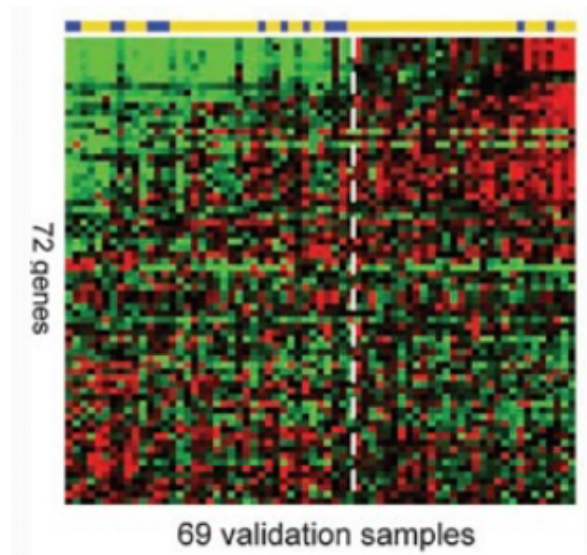
Datos de microarrays / RNA-Seq

Relaciones derivadas por  
dependencias estadísticas



## Datos de entrada

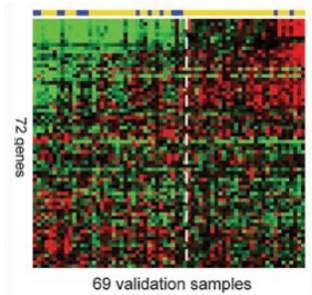
- Microarreglos o RNA-Seq
- Matrices



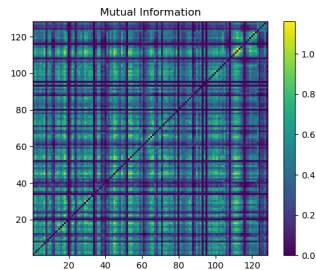
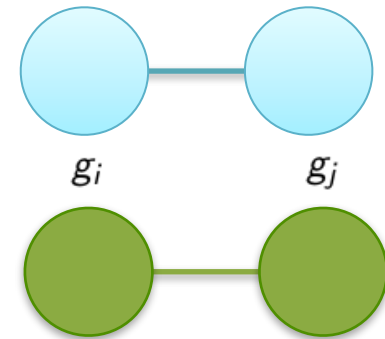
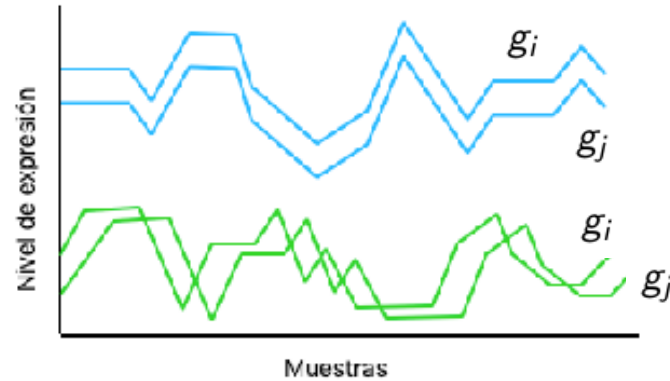
	S1	S2	S2	S3	S4
G1					
G2					
G3					
G4					

# Dependencia estadística

## Dependientes

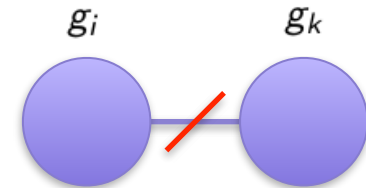
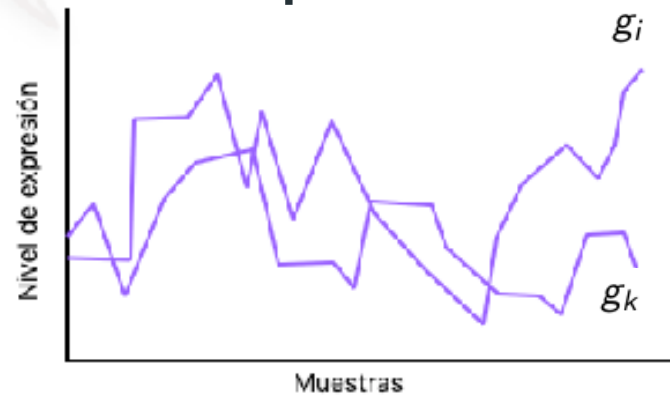


Nivel de expresión de genes

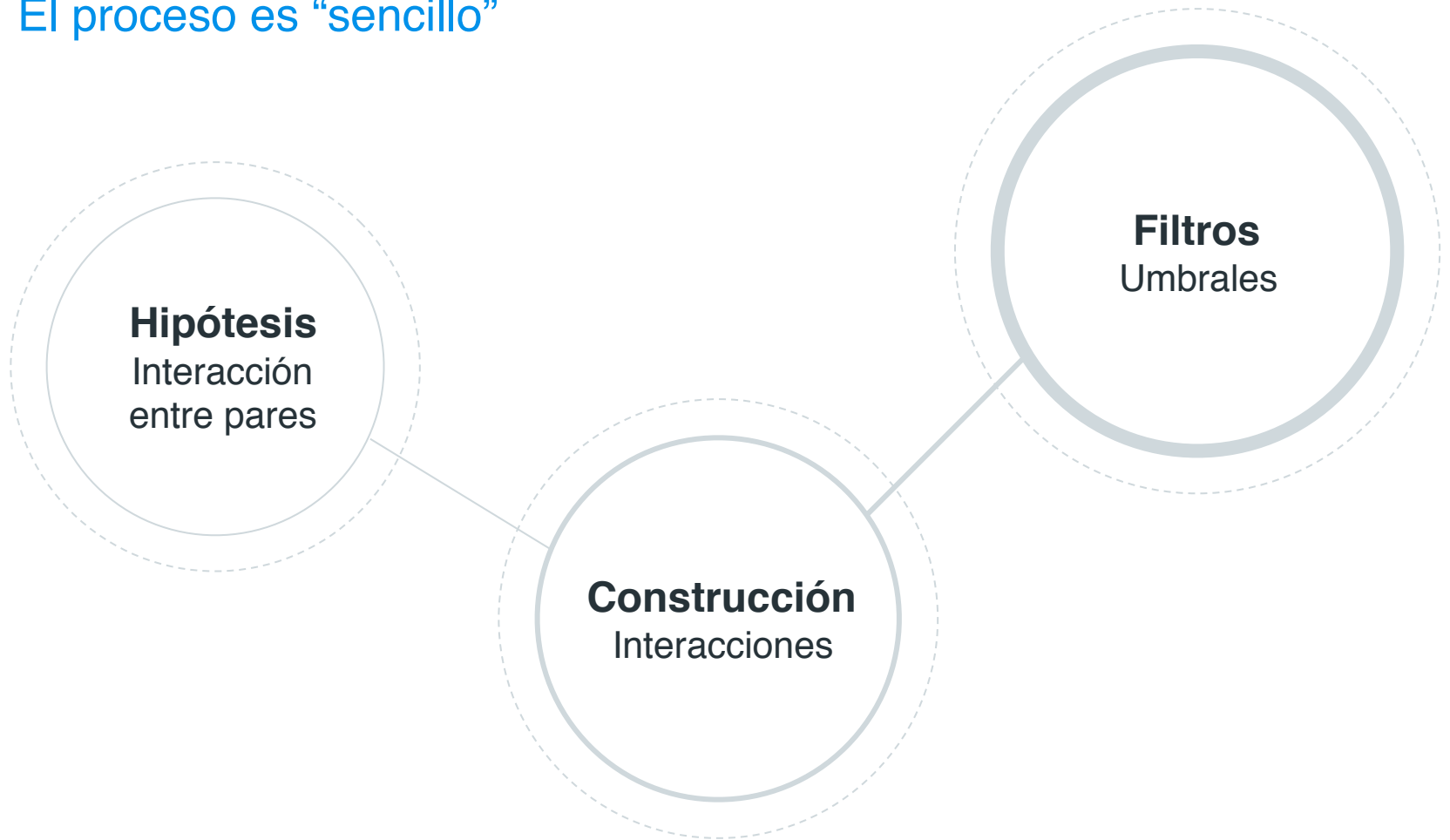


Información mutua

## Independientes



El proceso es “sencillo”



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are solid grey and others are hollow with a grey outline. The lines connecting them are thin and grey, creating a dense, organic structure.

# Conceptos

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It consists of a cluster of nodes (solid grey circles and hollow circles with grey outlines) connected by thin grey lines, forming a complex, interconnected web.



## Procesos estadísticos

### **Proceso determinista**

Bajo mismas condiciones,  
mismos resultados.

### **Proceso aleatorio**

Bajo mismas condiciones,  
resultados distintos.



## Conceptos

### Variable aleatoria

Función definida en el espacio de probabilidad asociado a un experimento aleatorio.

$$X : \Omega \longrightarrow \mathbb{R}$$

### Espacio probabilístico

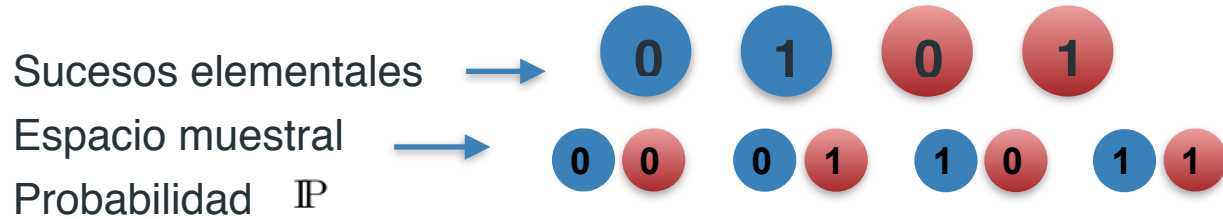
Sucesos elementales  $\longrightarrow$  





Espacio muestral  $\longrightarrow$  

Probabilidad  $\mathbb{P}$

# Conceptos

## Espacio probabilístico



		
	$(\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$	$(\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$
	$(\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$	$(\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$

# Hipótesis

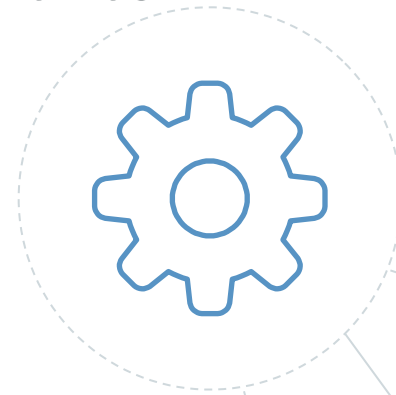
- Nivel de expresión  $\longrightarrow$  Variable aleatoria ( $g$ )

## Hipótesis Nula

- Genes estadísticamente independientes

$$\mathbb{P}(g_i, g_j) = \mathbb{P}(g_i)\mathbb{P}(g_j)$$

- Interacción entre pares de genes



# Información Mutua

$$I(g_i, g_j) \equiv I_{ij}$$

$$I(g_i, g_j) = \sum_{g_i, g_j} p(g_i, g_j) \ln \frac{p(g_i, g_j)}{p(g_i)p(g_j)}$$

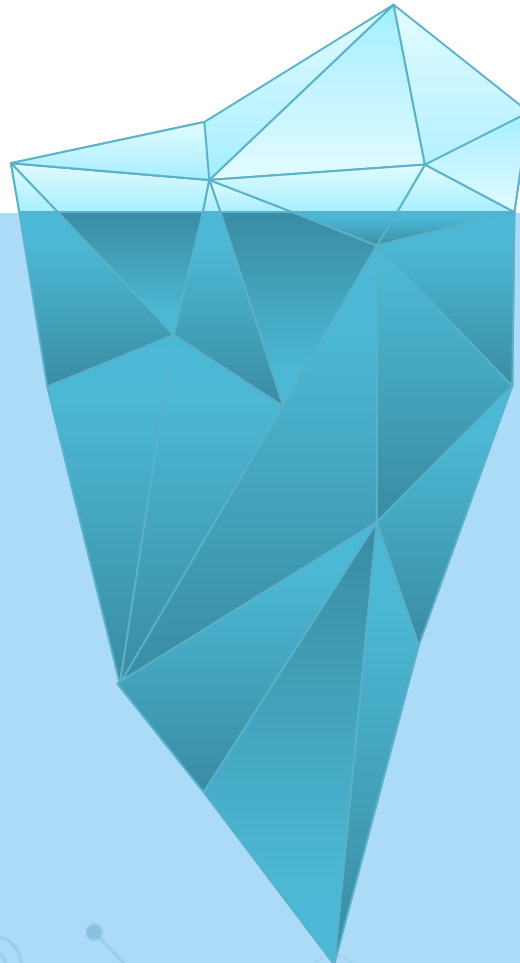
- Nivel de dependencia estadística entre 2 v.a.
- Basado en entropía de Shannon
- Genes independientes:

$$I_{ij} = 0$$

## Estimaciones (MI y JPD)

Distribución de  
Probabilidad  
Conjunta  
 $\mathbb{P}(\vec{g})$

- Aproximaciones de máxima entropía
- Factorización en potenciales

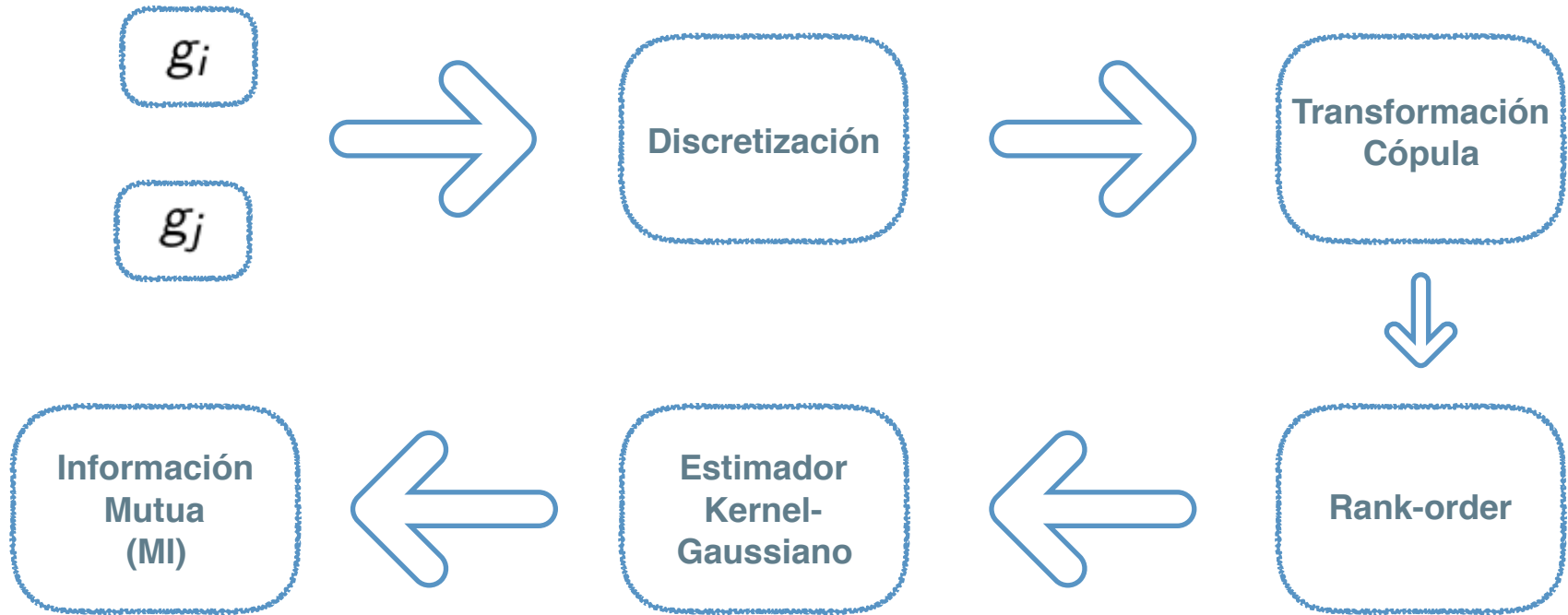


Información  
Mutua

$$I(g_i, g_j) \equiv I_{ij}$$

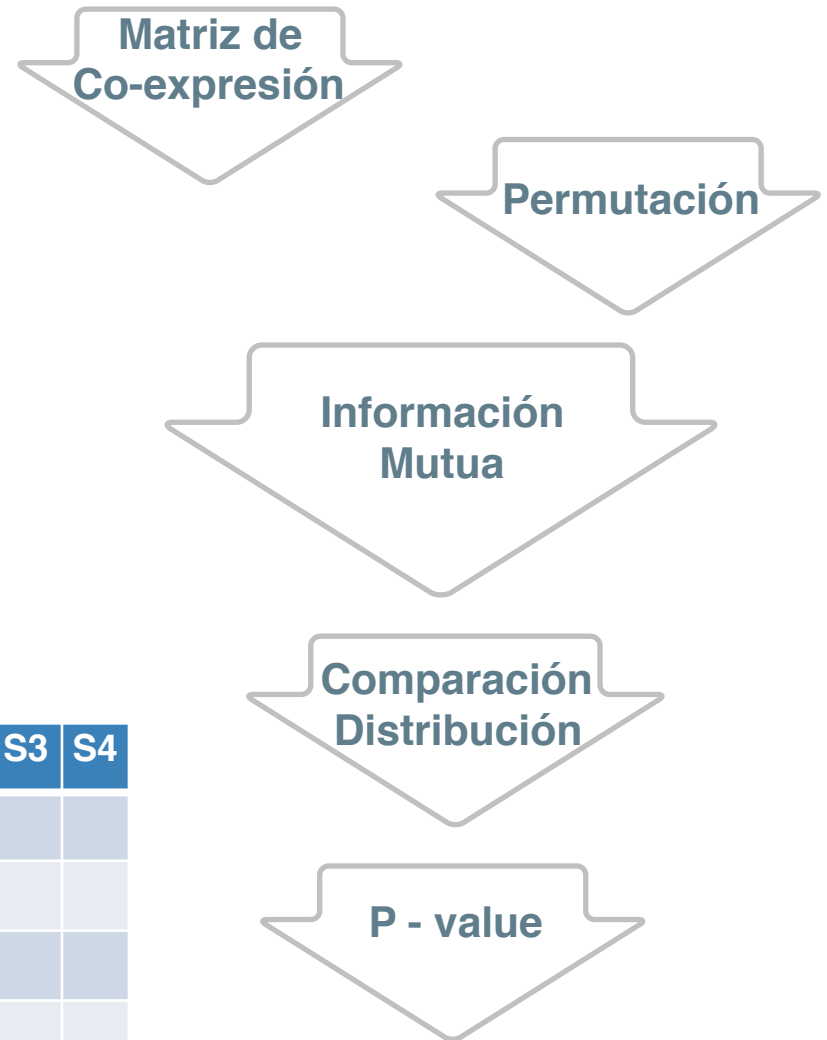
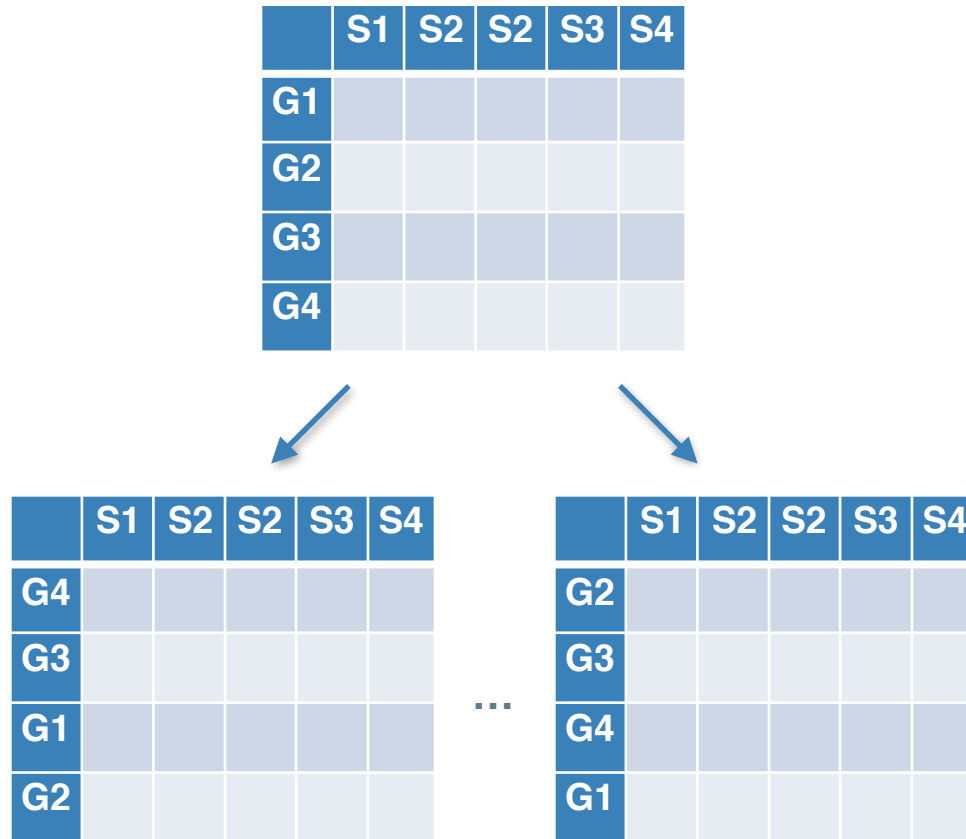
- Discretización
- Transformación cópula
- Rank-order
- Estimador Kernel-Gaussiano
  - F. P. Marginales
  - F. P. Conjuntas
  - Ancho de kernel ( $h$ )

## Estimación MI



$$I(g_i, g_j) = \sum_{g_i, g_j} p^*(g_i, g_j) \ln \frac{p^*(g_i, g_j)}{p^*(g_i)p^*(g_j)}$$

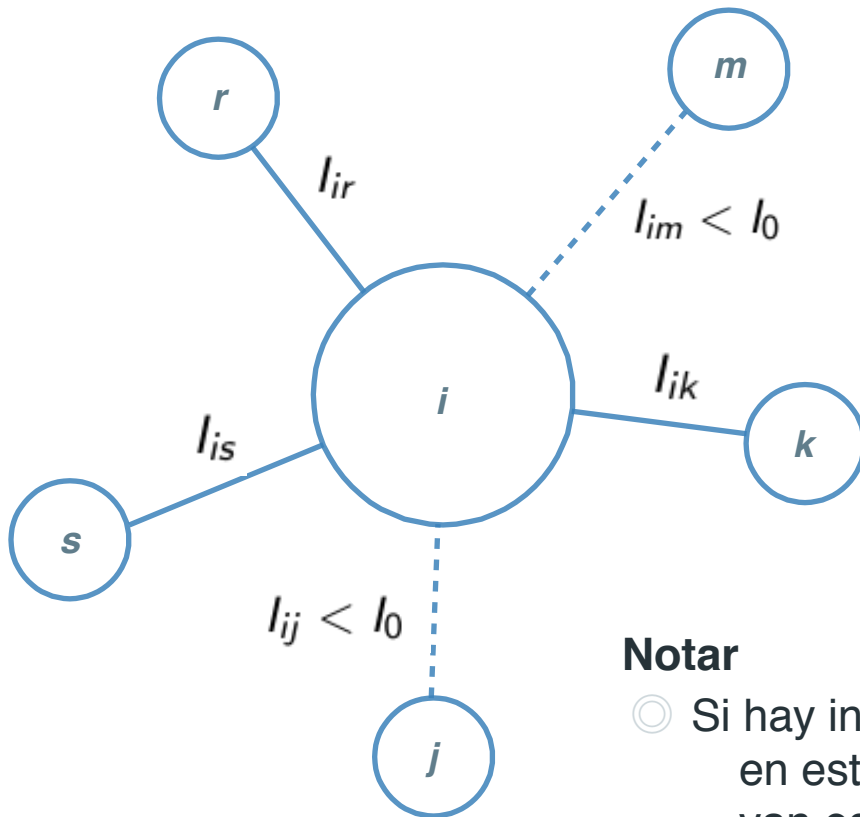
## I. Filtro Modelo Nulo





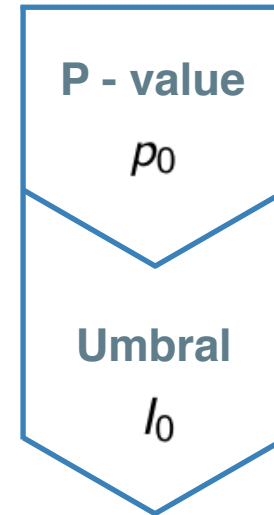
## I. Filtro

### Umbral por Información Mutua



#### Notar

- Si hay independencia, en este momento se van esos enlaces.

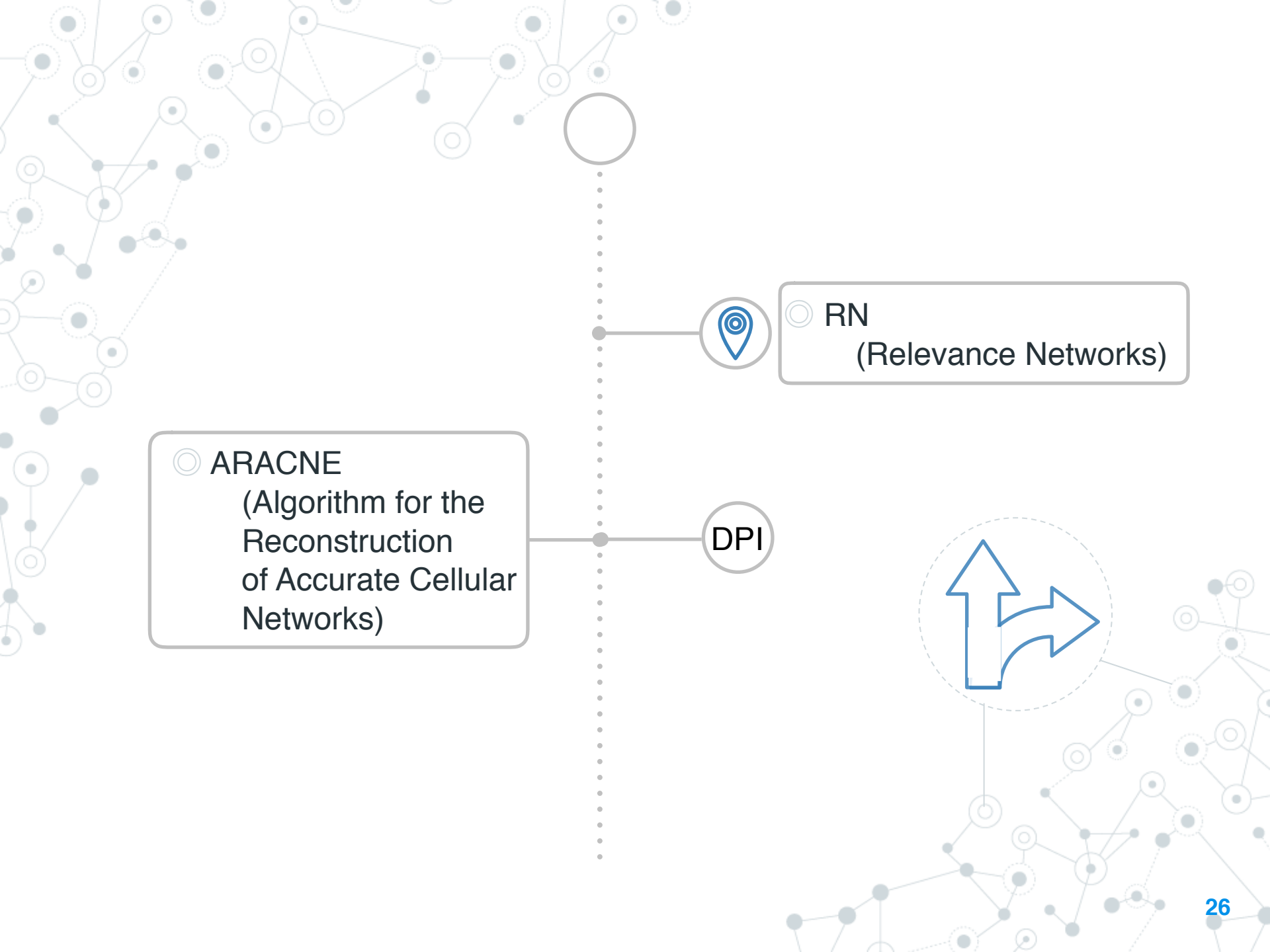


$$I_{ij} < I_0$$



#### Limitaciones

- Falsos positivos
- Genes intermediarios

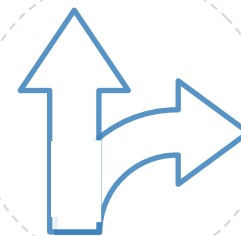


ARACNE  
(Algorithm for the  
Reconstruction  
of Accurate Cellular  
Networks)

DPI



RN  
(Relevance Networks)



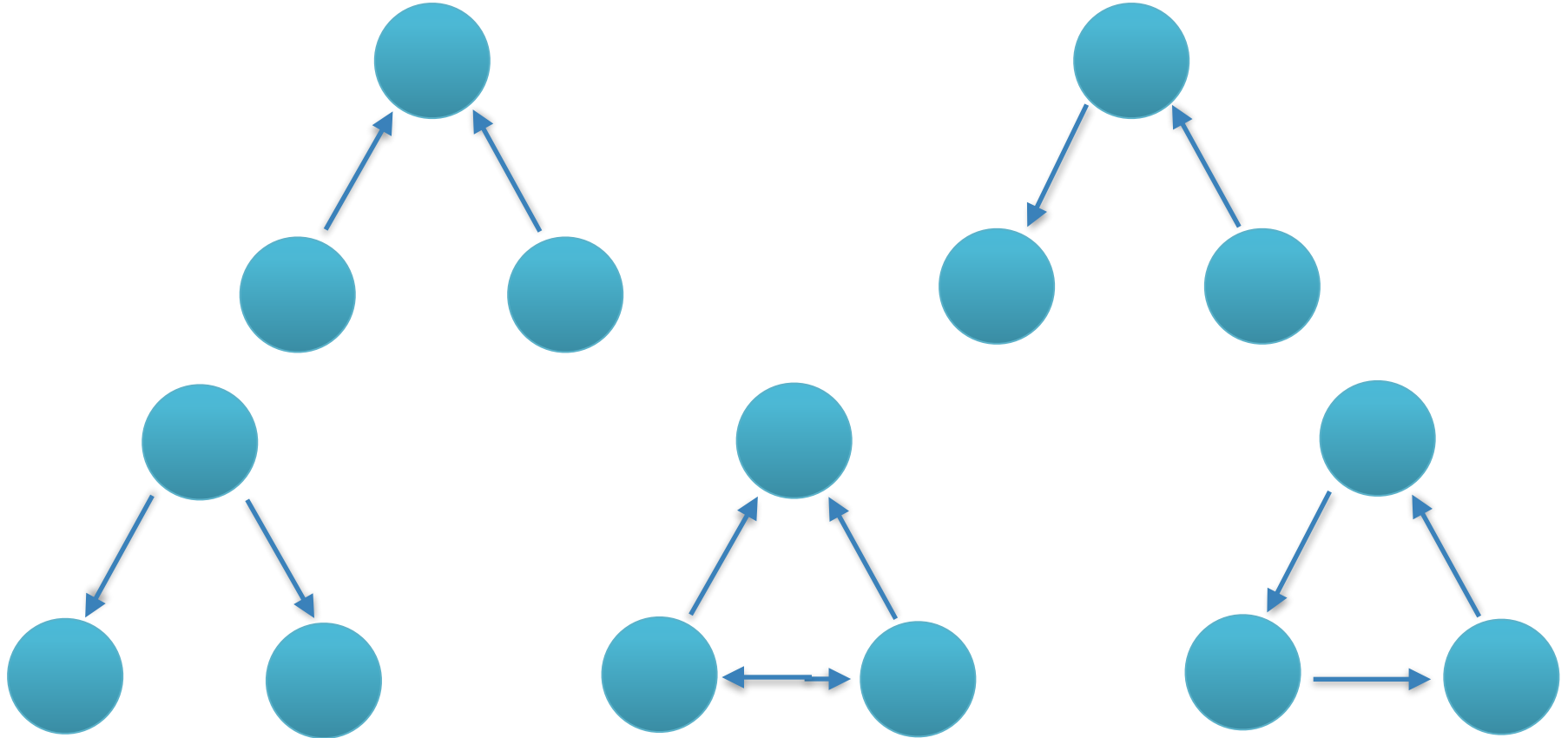


3.

# ARACNE

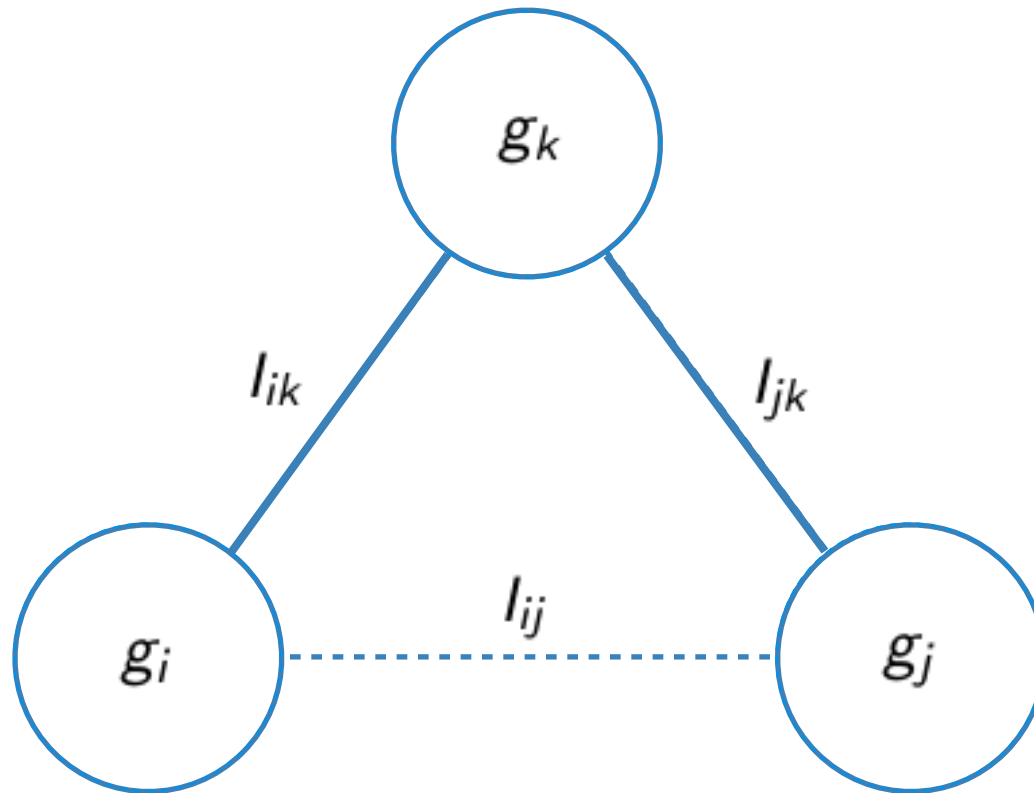
Margolin et al. (2006)

## Interacciones indirectas



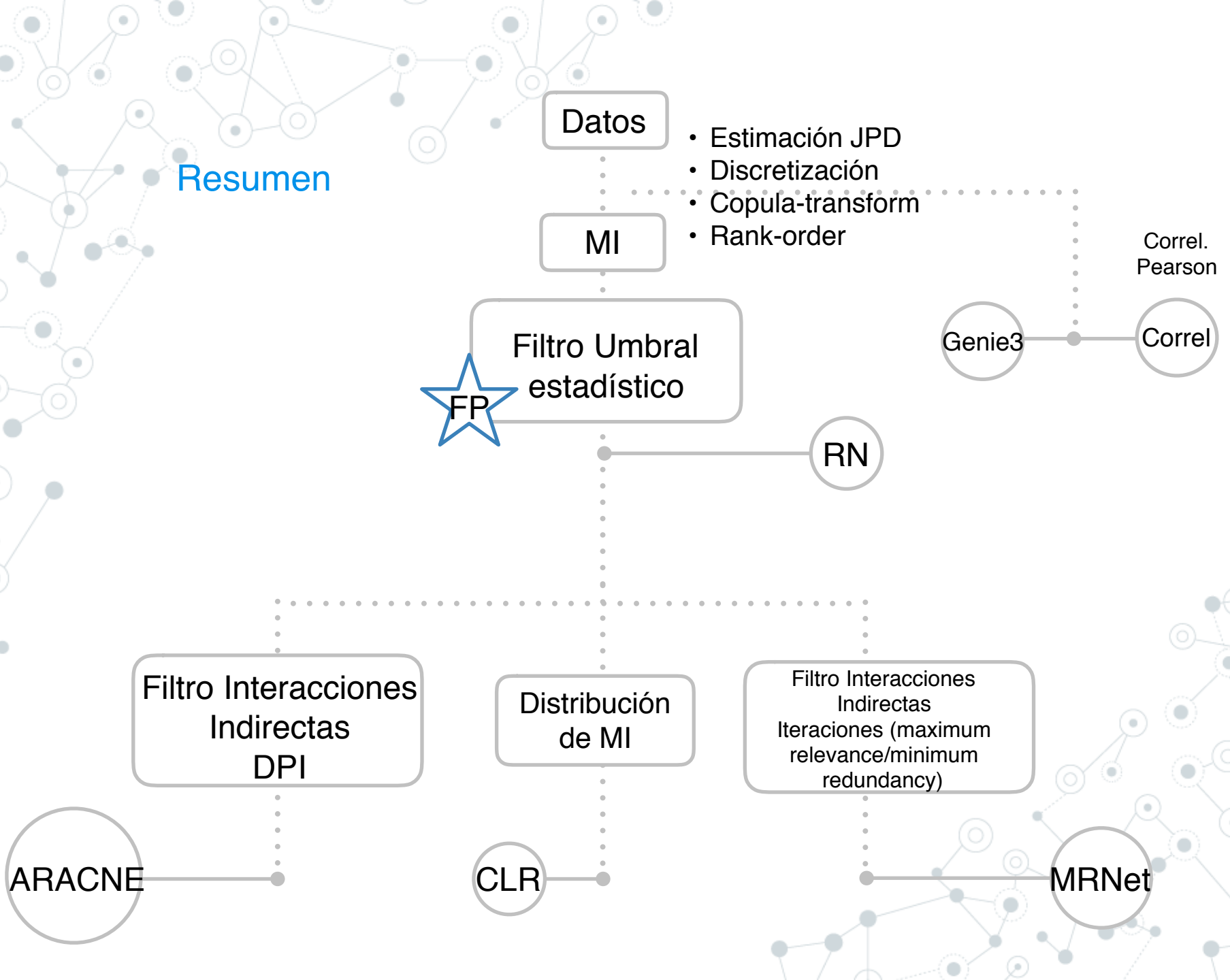
## II. Filtro

### Data Processing Inequality (DPI)



$$I(g_i, g_j) \leq \min[I(g_i, g_k); I(g_k, g_j)] \quad \text{Y} \quad \text{Umbral } \mathcal{T}$$

Resumen

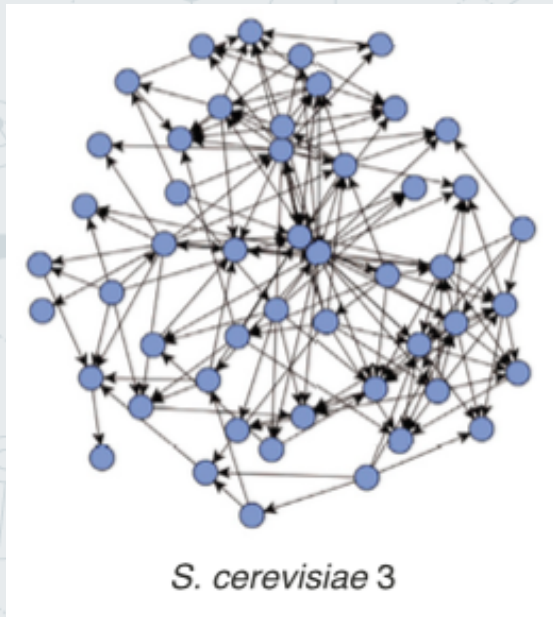


The background of the slide is a complex network diagram. It consists of numerous nodes, represented by small circles, some of which are solid grey and others are hollow with a grey outline. These nodes are interconnected by a web of thin, light-grey lines, creating a dense, interconnected pattern that fills the entire slide. The overall aesthetic is technical and scientific, suggesting a focus on data networks or systems biology.

# Práctica

Reconstruyamos una red de co-expresión

## I. Red de levadura



Nodos - genes (50)

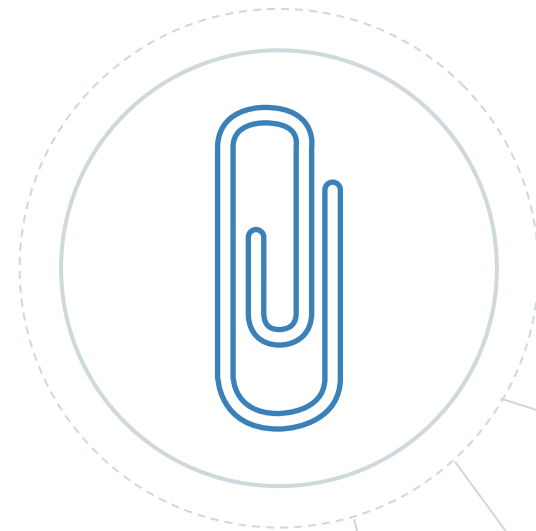


Enlaces - regulación



# 3. **Apéndice**

Cálculos



## Ecuaciones

- © Distribución de probabilidad conjunta (JPD)

$$\mathbb{P}(\{g_i\}) = \frac{1}{Z} \exp\left[-\sum_i^N \phi_i(g_i) - \sum_{i,j}^N \phi_{ij}(g_i, g_j) - \sum_{i,j,k}^N \phi_{ijk}(g_i, g_j, g_k) - \dots\right]$$

- © Condición de interacción entre genes

$$\phi_i \neq 0$$

- © Interacción entre pares de genes

$$H(\{g_i\}) = \sum_i^N \phi_i(g_i) + \sum_{i,j}^N \phi_{ij}(g_i, g_j)$$

## Ecuaciones

- Condición de independencia entre genes (MI)

$$I_{ij} = 0 \iff \mathbb{P}(g_i, g_j) = \mathbb{P}(g_i)\mathbb{P}(g_j)$$

- Información mutua en términos de la entropía

$$I(X, Y) = S(X) + S(Y) - S(X, Y)$$

- Entropía en términos de funciones de probabilidad

$$S(X) = - \sum_{x \in X} \mathbb{P}(X = x) \ln \mathbb{P}(X = x)$$

## Ecuaciones

- Relación entre MI y entropía en términos de función de probabilidad

$$I(X, Y) = \sum_{x,y} p(x, y) \ln \frac{p(x, y)}{p(x)p(y)}$$

- Notación:

$$p(x_i) = \mathbb{P}(X = x_i)$$

## Ecuaciones del estimador

### © Estimador Kernel-Gaussiano

$$f(z) = \frac{1}{M} \sum_i \frac{G|z - z_i|}{h^3}$$

### © Información mutua dado el estimador

$$I(x_i, y_i) = \frac{1}{M} \sum_i \log \frac{f(x_i, y_i)}{f(x_i)f(y_i)}$$



**Gracias por su  
atención**

**¿Preguntas?**