

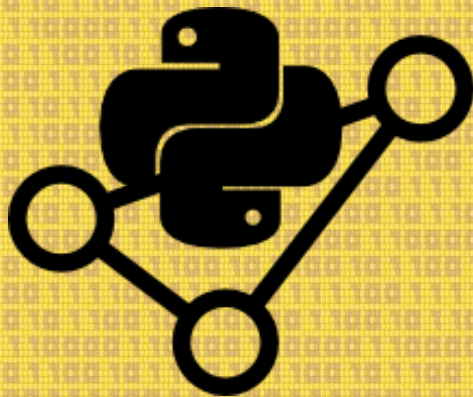
# yatel



Artwork Based on: <http://ipadwallpapergallery.com/calnity-by-pr09studio/> - cc-by-nc-sa



# Caracterización del Sistema *Mal de Río Cuarto del Maíz* mediante Minería de Datos y Análisis de Redes





# Integrantes



García, Mario Alejandro



Cabral, Juan Bautista



Gimenez Pecci, María de la Paz



Vera, Carlos



Laguna, Irma Graciela



Bisonard, Eduardo Matías



Maurino, Fernanda



Vankeirsbilck, Inés



Cucco, Noelia del Valle



Nieto Castillo, Adrián L.



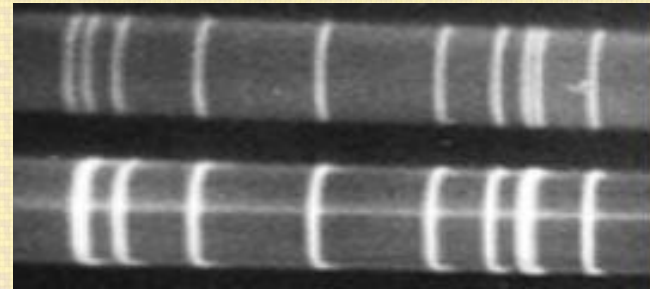
# Knowledge Discovery in Database (KDD)

- Es un proceso no trivial de identificación de información útil y desconocida que permanece oculta en una base de datos [Fayyad, 1996]
- Es un proceso centrado en la persona (*human-centered*) [Brachman, 1996]



## *Mal de Río Cuarto virus*

- Análisis electroforético:



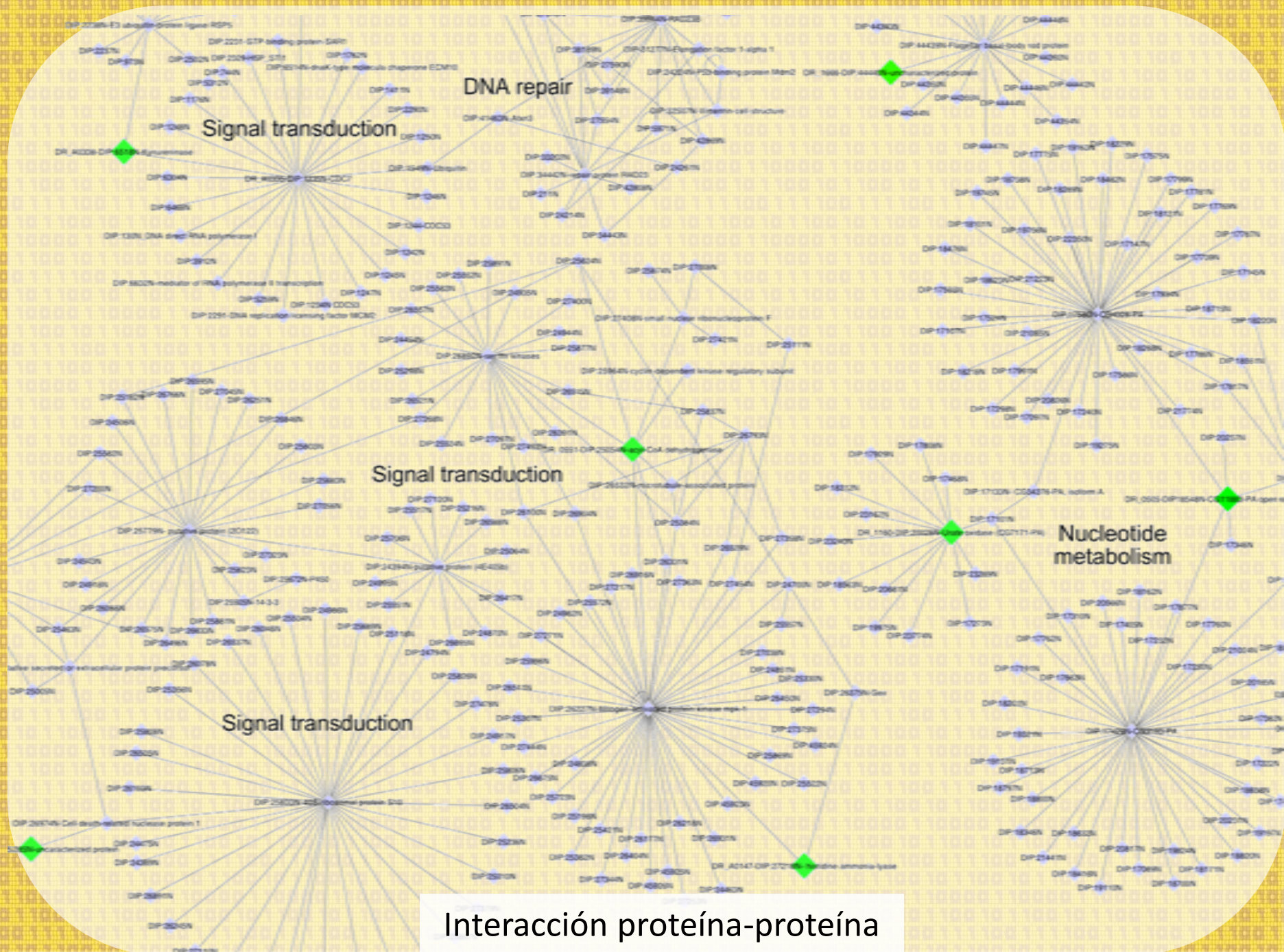
- Base de datos formada por perfiles electroforéticos + atributos que definen el ambiente de la planta
- Resultados de estudios anteriores: Algunos segmentos electroforéticos dependen de otros



# Network Science

- Es el estudio de las redes que representan fenómenos físicos, biológicos y sociales conduciendo a modelos predictivos de estos fenómenos.
- Topologías
- Características comunes

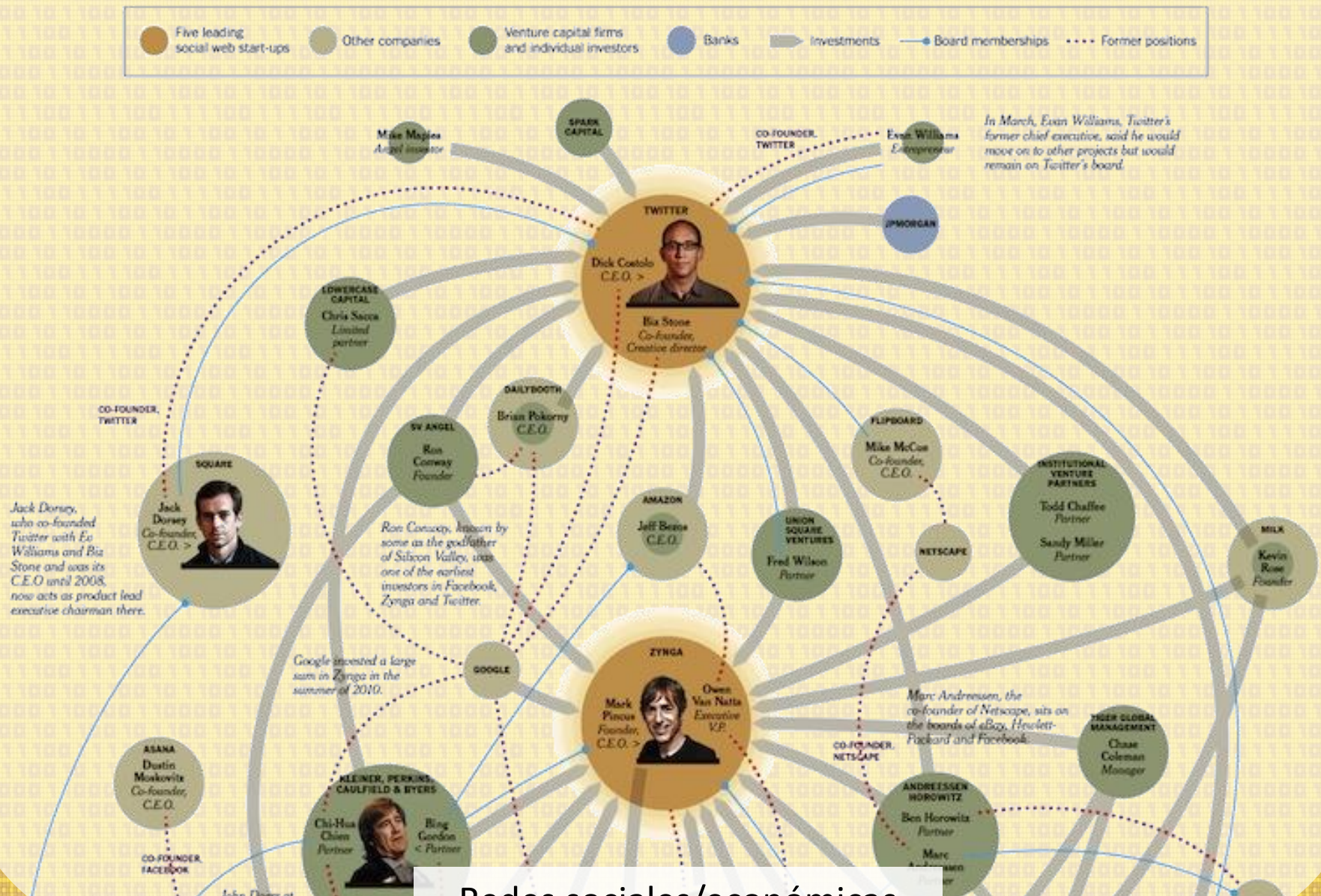




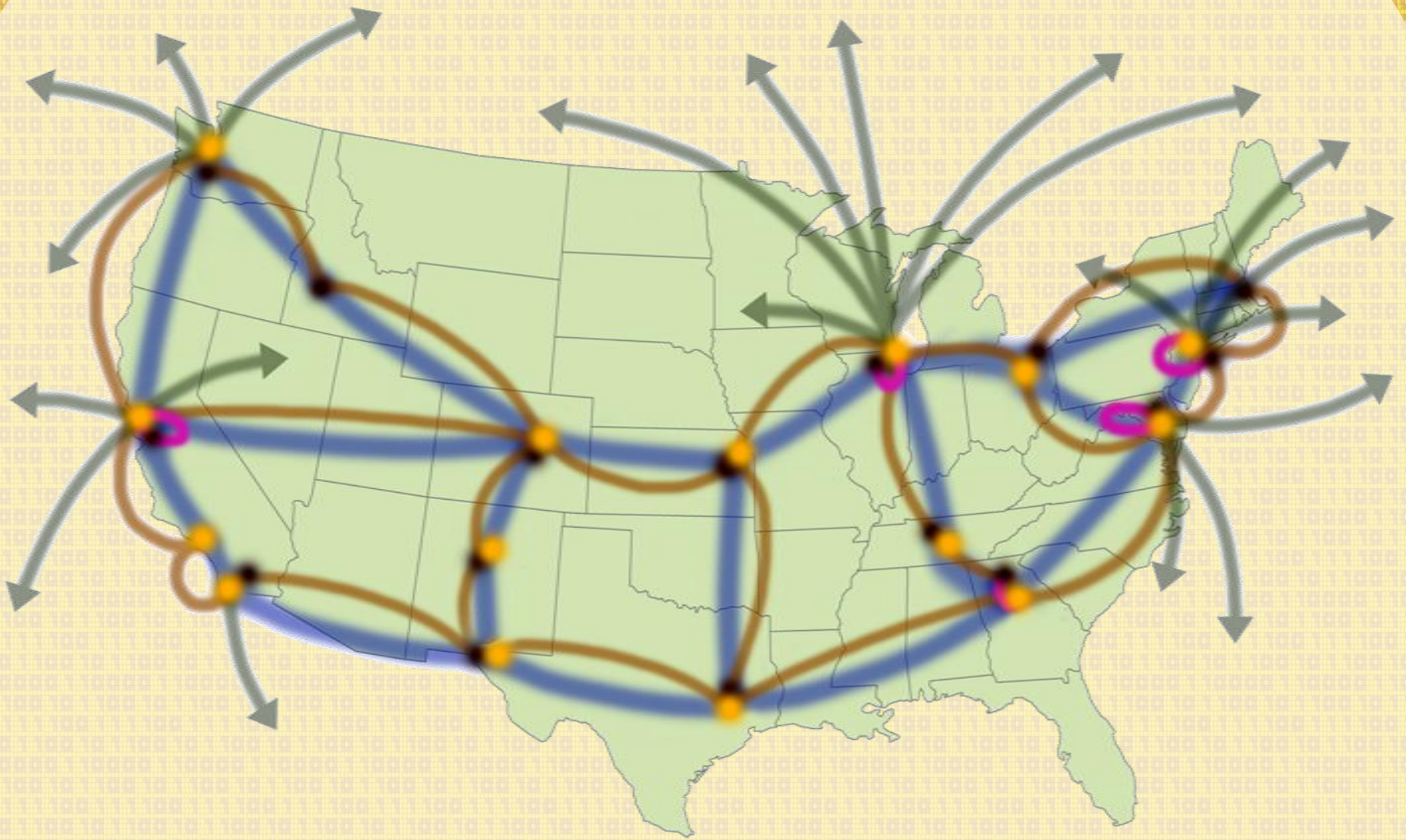
Interacción proteína-proteína



## Redes sociales/económicas







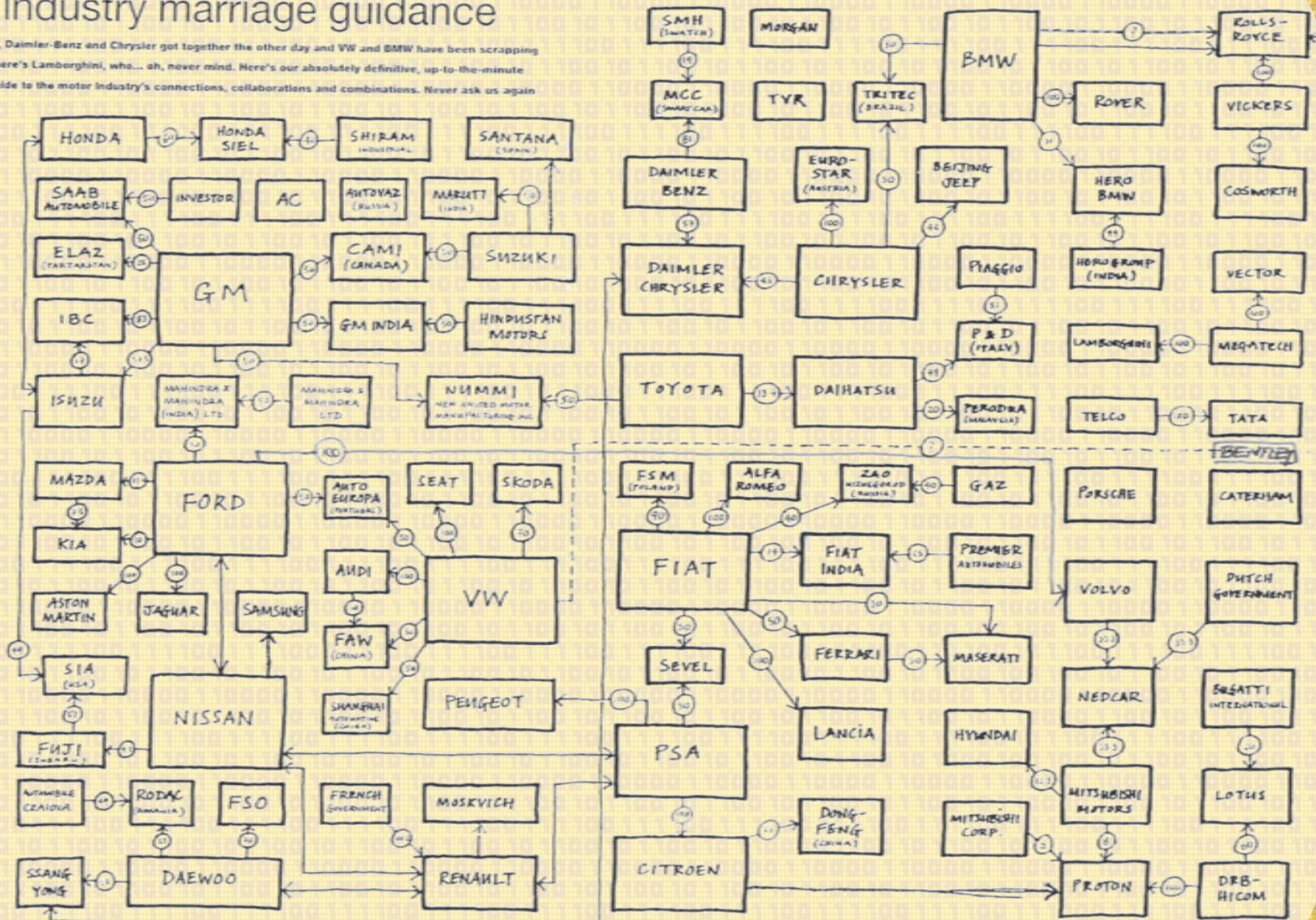
Red de distribución de energía



# motor industry marriage guidance

Who really owns who? Well, Daimler-Benz and Chrysler got together the other day and VW and BMW have been scrapping over Rolls-Royce, then there's Lamborghini, who... ah, never mind. Here's our absolutely definitive, up-to-the-minute [well, this week anyway] guide to the motor industry's connections, collaborations and combinations. Never ask us again

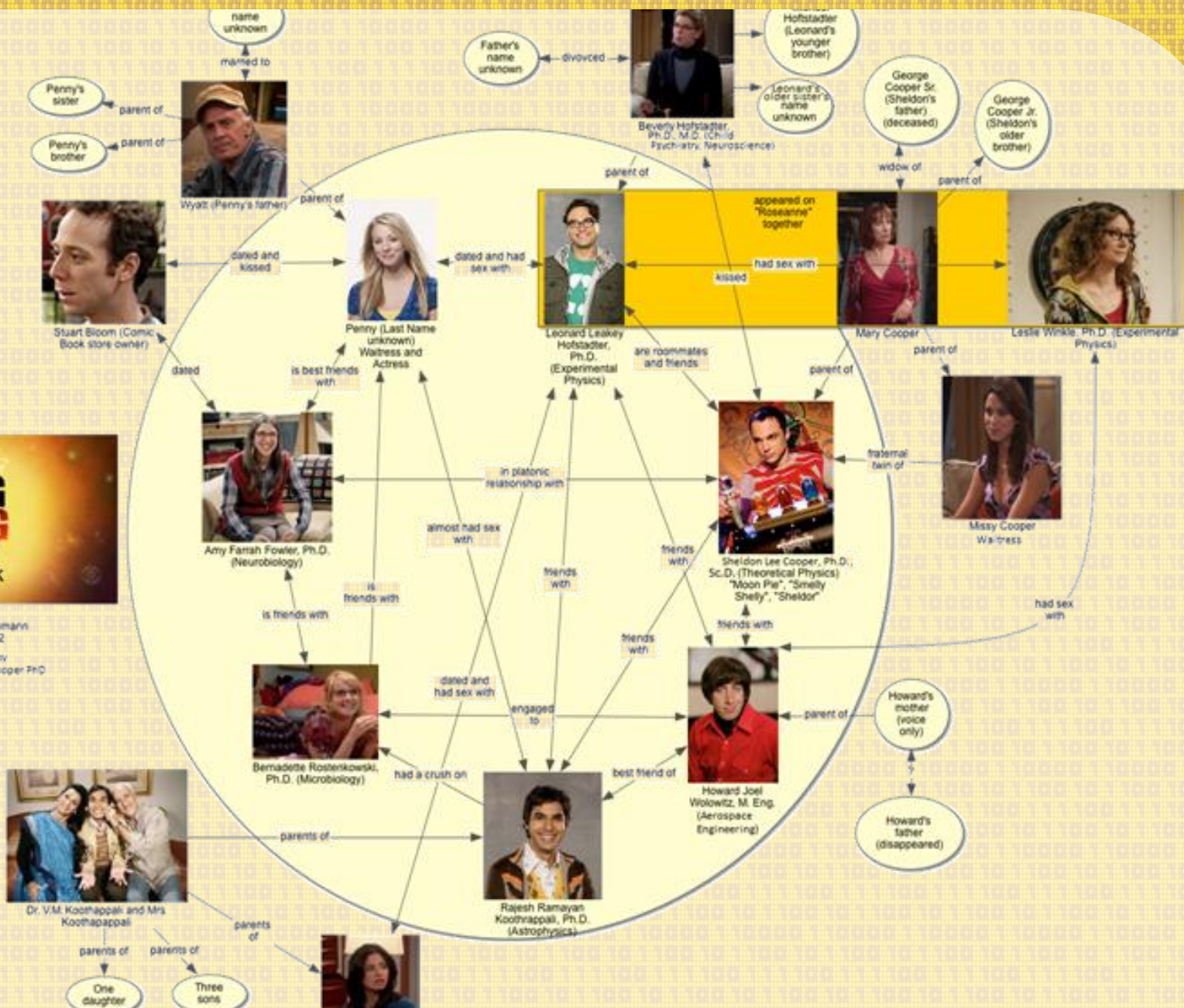
News of the mega merger between Daimler-Benz and Chrysler and the squalid battle for the hand of Rolls-Royce's famous Silver Lady have hogged the headlines recently, but a glance at these pages shows that similar things have been going on for years. The number of marriages may not be going down these days, but the careers of decision makers are getting fewer as the quest for cost reductions goes on. Today BMW owns Rover. A few years ago Honda and British Aerospace did. Before that, Saab and Volvo. Along the way, Maserati and Austin, the cornerstones of the original Rover group, were dumped in the scrapyard. It was survival of the fittest. These days companies collaborate because it cuts investment in product development, component buying and manufacturing. They get together because they are weak or need to collaborate for specific models - Volvo 40 series/Peugeot 405, for example. What is driving this trend - apart from ambition, greed and ego - is that car buyers are demanding wider model variety, more equipment and lower prices and producers are having to meet more and more safety and emissions regulation. Heavily saddled with the ability to produce millions more cars than they can sell, car makers must economise or go bust. If that means sharing engines for the new Mini in South Korea, wages, etc. of government grants, along with another company taking the cost, so be it. The decision to merge by two successful companies - Daimler and Chrysler - has profound implications for those which are not so lucky. The weak will get weaker, and one day will replace merger. The Rolls-Royce sale is pending. The futures of Fiat and Lamborghini remain fuzzy. But what of the rest? What of Japan's current favour cases, Renault and Mitsubishi? Big business, it's not the people who want for it, it's a week into a just nation period. It is time to think the unthinkable. Richard Pearson







created by  
Beat A. Schwendmann  
January 2012  
modified by  
UserSheldon Cooper PhD



Red semántica TBBT

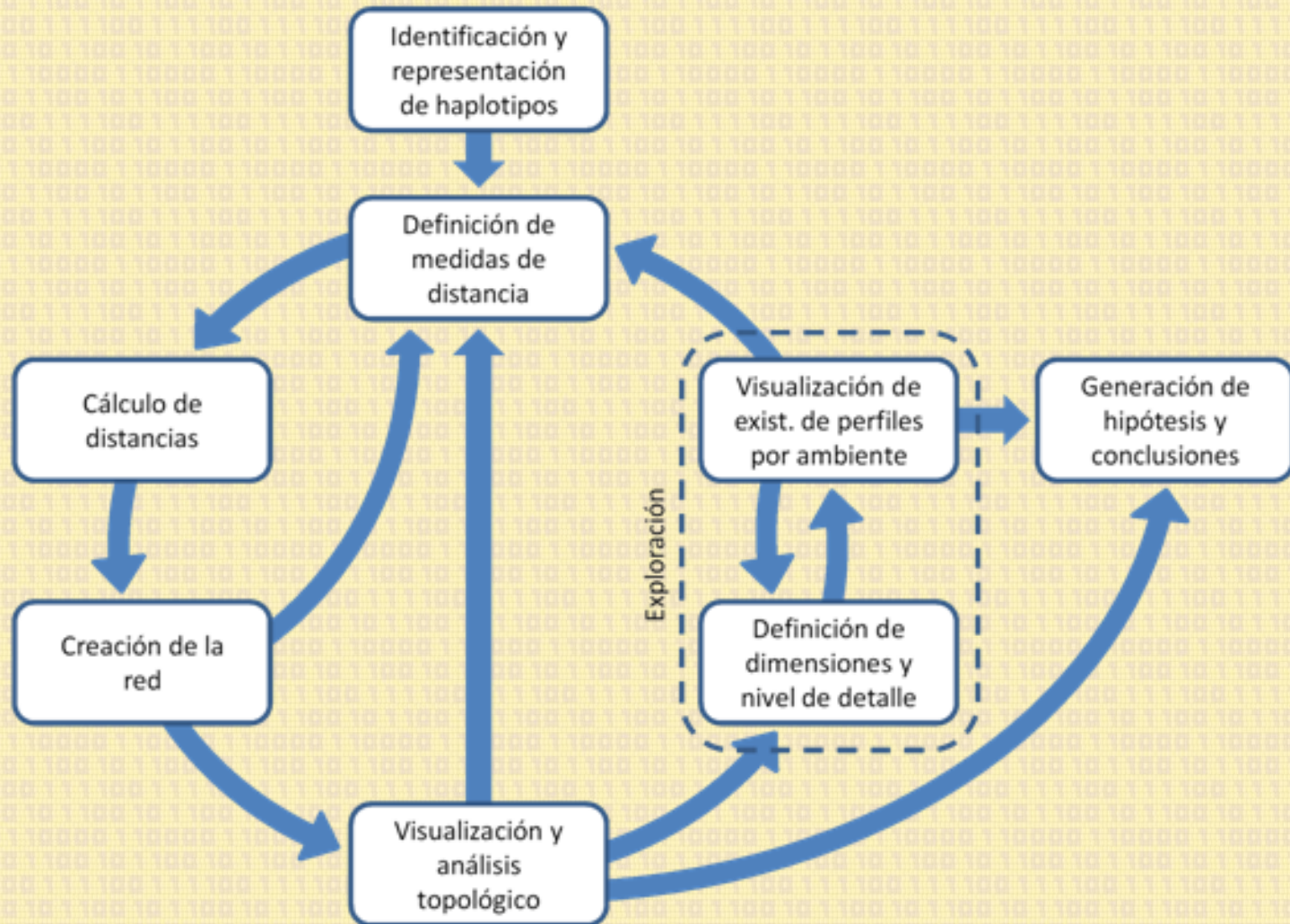


# Network Science

- Es el estudio de las redes que representan fenómenos físicos, biológicos y sociales conduciendo a modelos predictivos de estos fenómenos.
- Topologías
- Características comunes

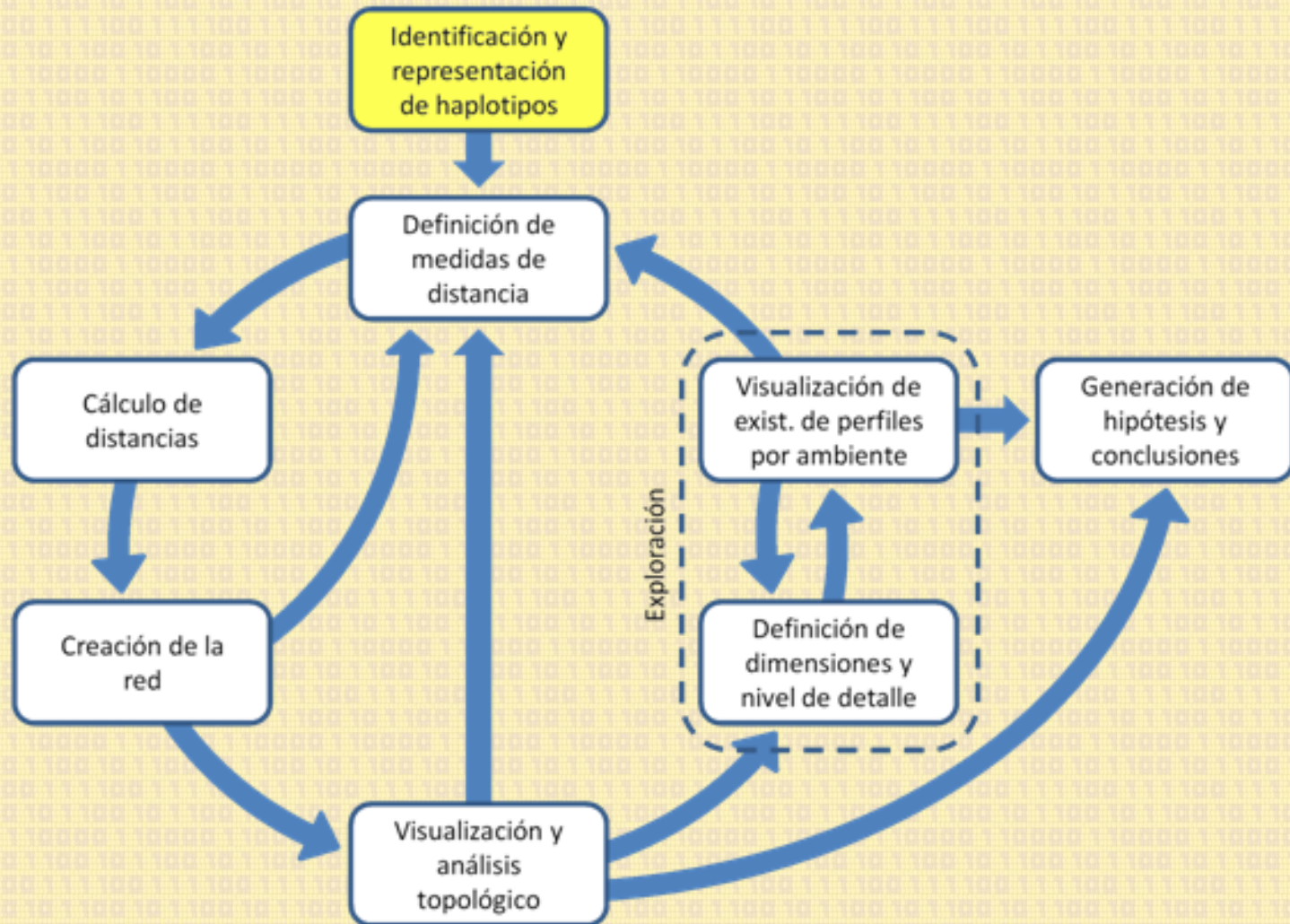


# Proceso de análisis



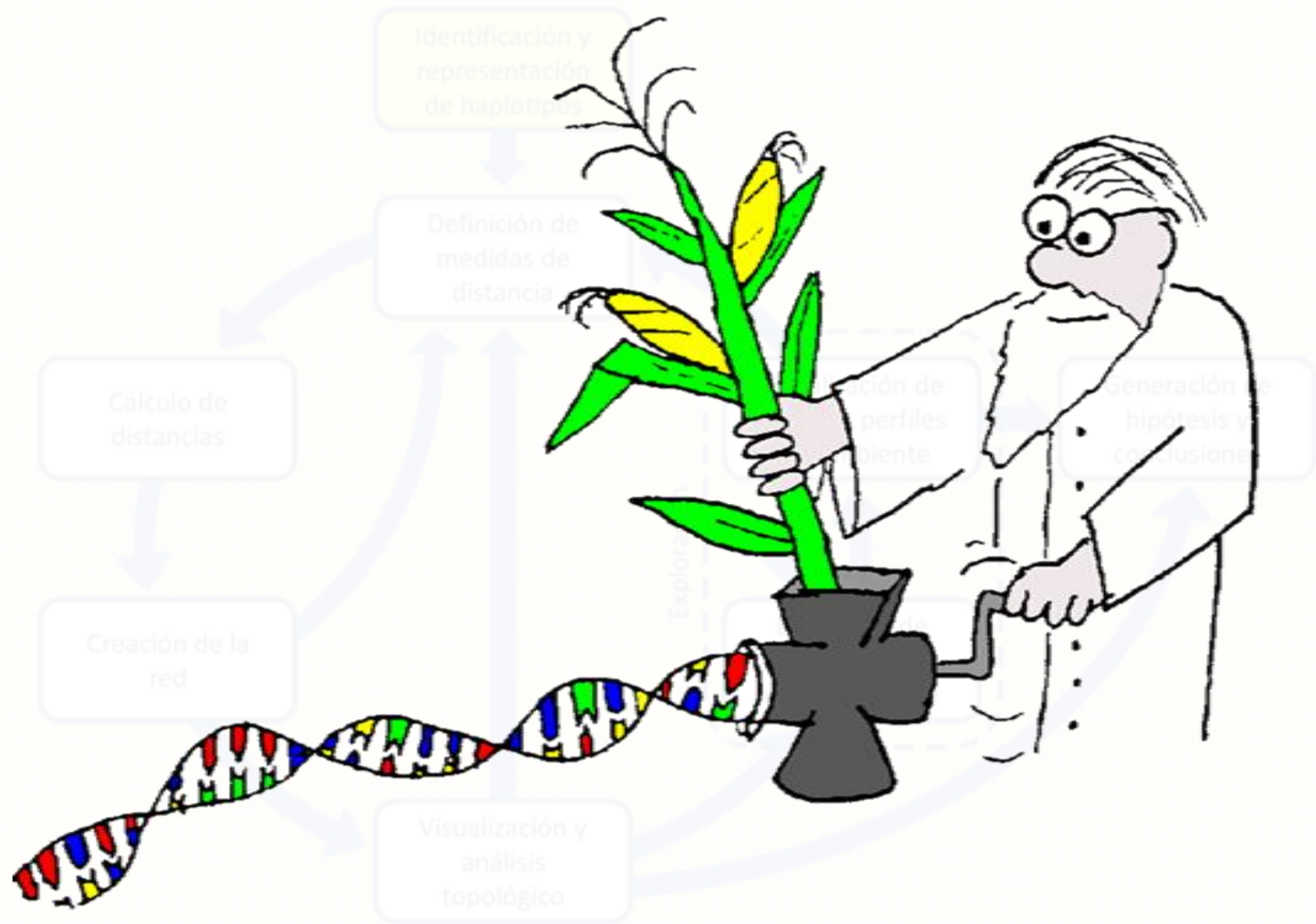


# Identificación y representación de haplotipos



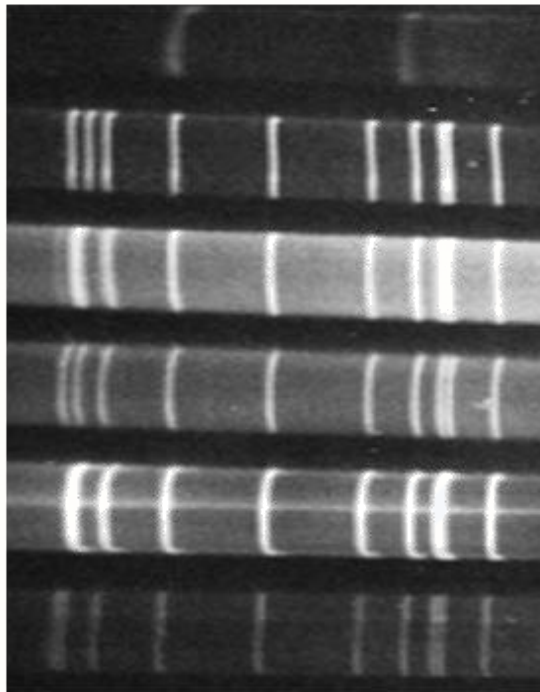


# Identificación y representación de haplotipos





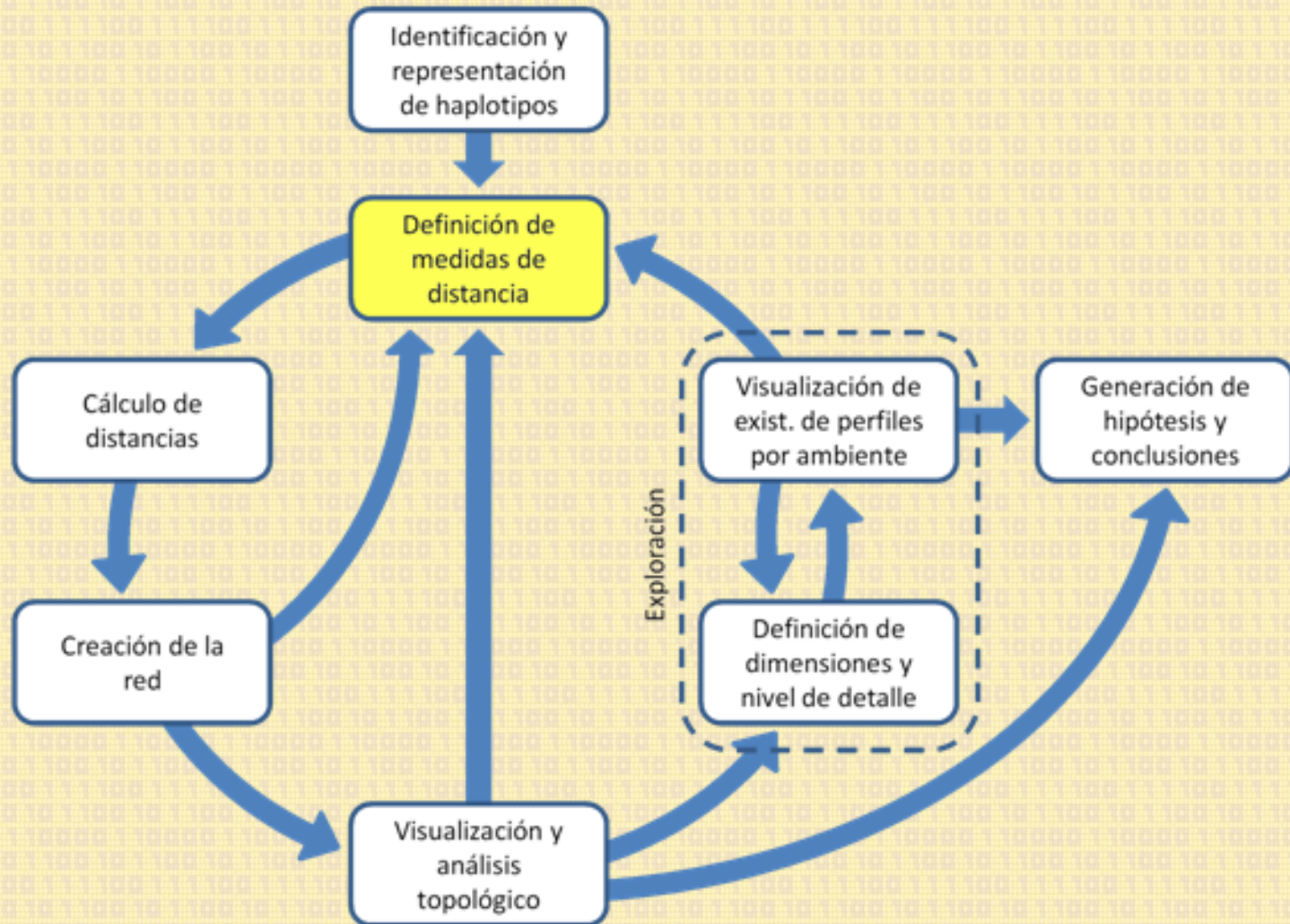
# Identificación y representación de haplotipos



| Hapl. | B3a | B3b | B5 | B8 | B9a | B9b | B9c | B10a | B10b | E5 | E10 |
|-------|-----|-----|----|----|-----|-----|-----|------|------|----|-----|
| 1     | 1   | 0   | 1  | 1  | 1   | 1   | 0   | 0    | 0    | 0  | 0   |
| 2     | 1   | 0   | 1  | 1  | 1   | 0   | 1   | 1    | 0    | 0  | 0   |
| 3     | 1   | 0   | 1  | 1  | 1   | 0   | 0   | 1    | 1    | 0  | 0   |
| 4     | 1   | 0   | 1  | 1  | 1   | 0   | 0   | 1    | 0    | 0  | 0   |
| 5     | 1   | 0   | 1  | 1  | 0   | 1   | 0   | 1    | 0    | 0  | 0   |
| 6     | 1   | 0   | 1  | 1  | 0   | 1   | 0   | 0    | 0    | 0  | 0   |
| 7     | 1   | 0   | 1  | 1  | 0   | 0   | 1   | 1    | 0    | 1  | 1   |
| 8     | 1   | 0   | 1  | 1  | 0   | 0   | 1   | 1    | 0    | 1  | 0   |
| 9     | 1   | 0   | 1  | 1  | 0   | 0   | 1   | 1    | 0    | 0  | 0   |
| 10    | 1   | 0   | 1  | 1  | 0   | 0   | 1   | 0    | 1    | 0  | 0   |
| 15    | 1   | 0   | 1  | 1  | 0   | 0   | 1   | 0    | 0    | 0  | 1   |
| 16    | 1   | 0   | 1  | 1  | 0   | 0   | 1   | 0    | 0    | 0  | 0   |
| 11    | 1   | 0   | 1  | 1  | 0   | 0   | 0   | 1    | 1    | 0  | 0   |
| 17    | 1   | 0   | 0  | 1  | 1   | 0   | 1   | 1    | 0    | 0  | 0   |
| 18    | 1   | 0   | 0  | 0  | 0   | 1   | 0   | 1    | 0    | 0  | 0   |
| 19    | 0   | 1   | 1  | 1  | 1   | 1   | 0   | 1    | 1    | 0  | 0   |
| 20    | 0   | 1   | 1  | 1  | 0   | 1   | 0   | 1    | 0    | 1  | 1   |
| 12    | 0   | 1   | 1  | 1  | 0   | 1   | 0   | 1    | 0    | 0  | 0   |
| 13    | 0   | 1   | 1  | 1  | 0   | 0   | 1   | 1    | 0    | 1  | 1   |
| 14    | 0   | 1   | 1  | 1  | 0   | 0   | 1   | 1    | 0    | 0  | 0   |
| 21    | 0   | 1   | 1  | 1  | 0   | 0   | 1   | 0    | 0    | 0  | 0   |



## Definición de medidas de distancia





# Definición de medidas de distancia

$$d_{ij} = dB3_{ij} + dB5_{ij} + dB8_{ij} + dB9_{ij} + dB10_{ij} + dBE5_{ij} + dBE10_{ij}$$

donde:

$$dB3_{ij} = (|B3a_i - B3a_j| + |B3b_i - B3b_j| + |B3a_i - B3a_j + B3b_i - B3b_j|)/2 \quad (\text{excepción 1})$$

$$dB5_{ij} = |B5_i - B5_j| \quad (\text{dist. de Hamming})$$

$$dB8_{ij} = |B8_i - B8_j| \quad (\text{dist. de Hamming})$$

$$dB9_{ij} = (|B9a_i - B9a_j| + |B9b_i - B9b_j| + |B9c_i - B9c_j| + |B9a_i - B9a_j + B9b_i - B9b_j + B9c_i - B9c_j|)/2 \quad (\text{excepción 1})$$

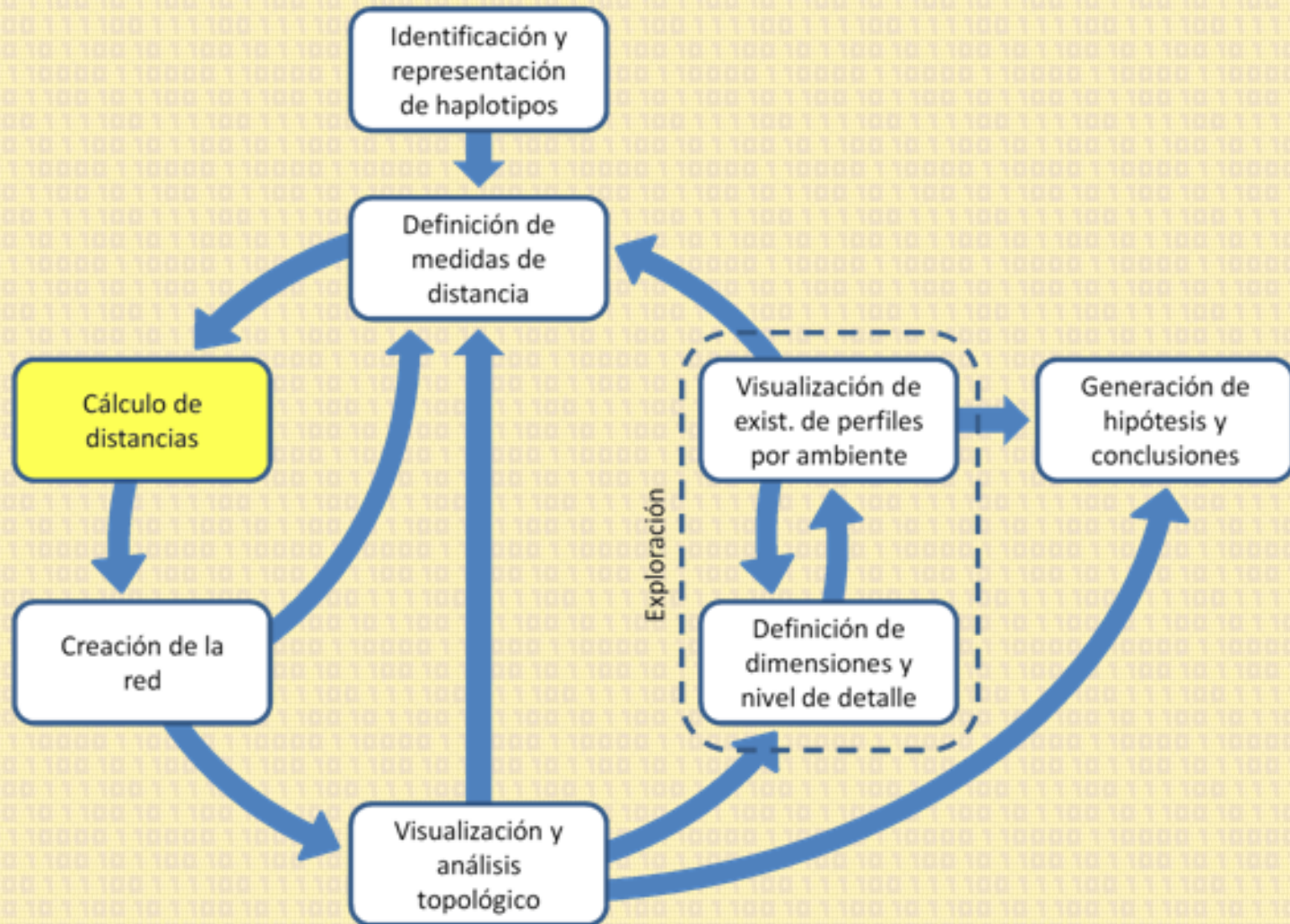
$$dB10_{ij} = (|B10a_i - B10a_j| + |B10b_i - B10b_j| + |B10a_i - B10a_j + B10b_i - B10b_j|)/2 \quad (\text{excepción 1})$$

$$dBE5_{ij} = |BE5_i - BE5_j| (1 - |B5_i - B5_j|) \quad (\text{excepción 3})$$

$$dBE10_{ij} = |BE10_i - BE10_j| (1 - (|B3a_i - B3a_j| \text{ OR } |B3b_i - B3b_j|)) \quad (\text{excepción 2})$$



# Cálculo de distancias

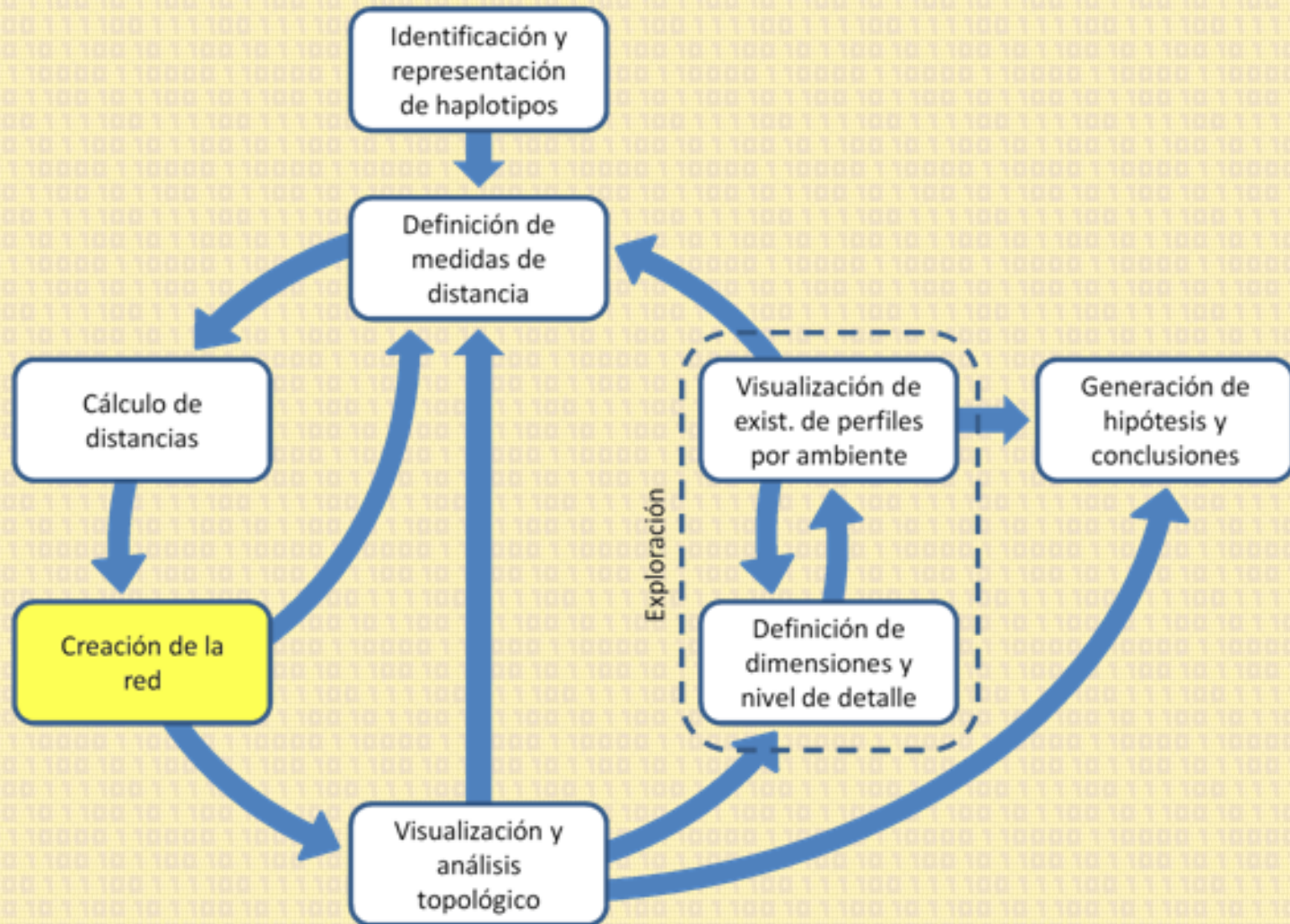




[illegible]

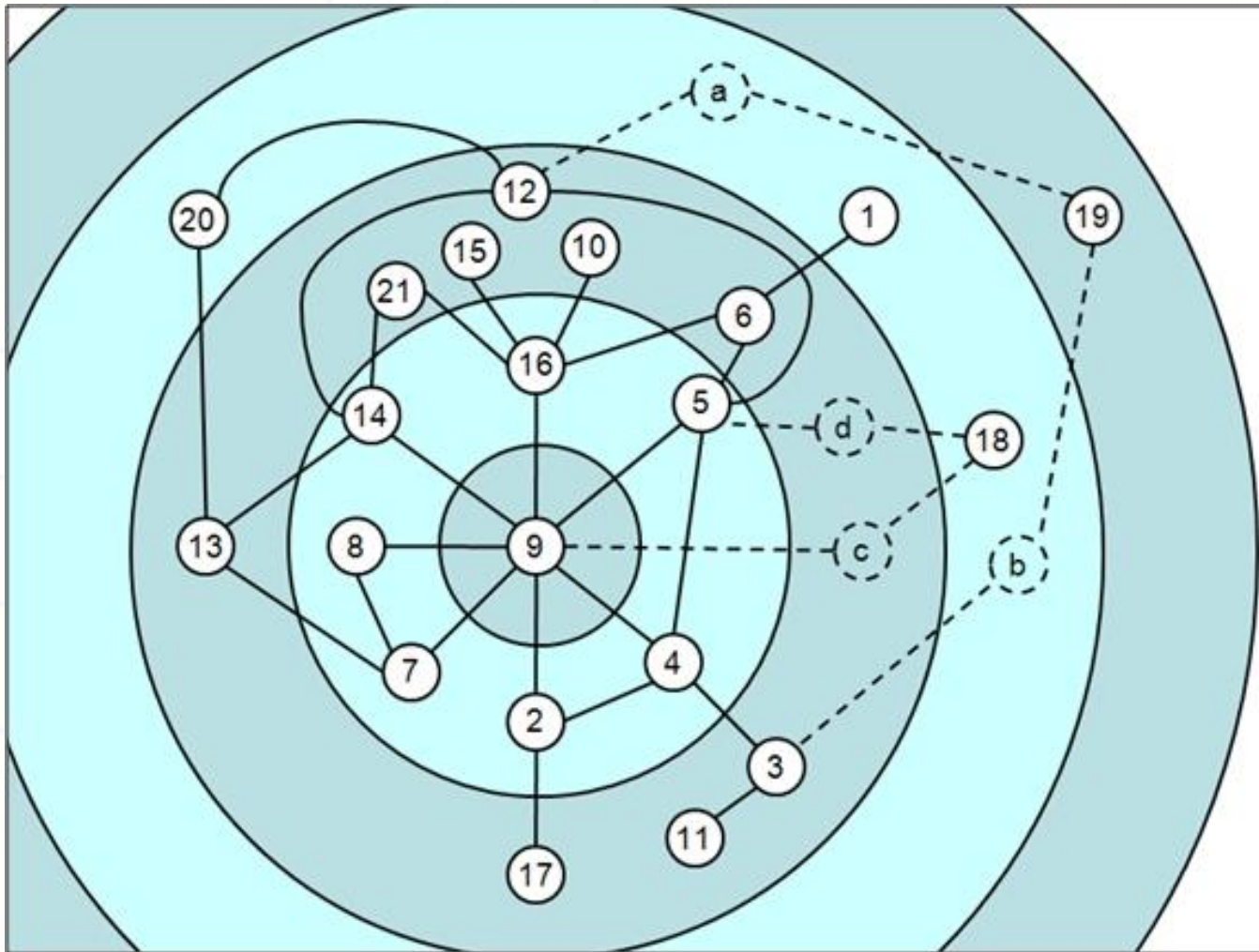


# Creación de la red



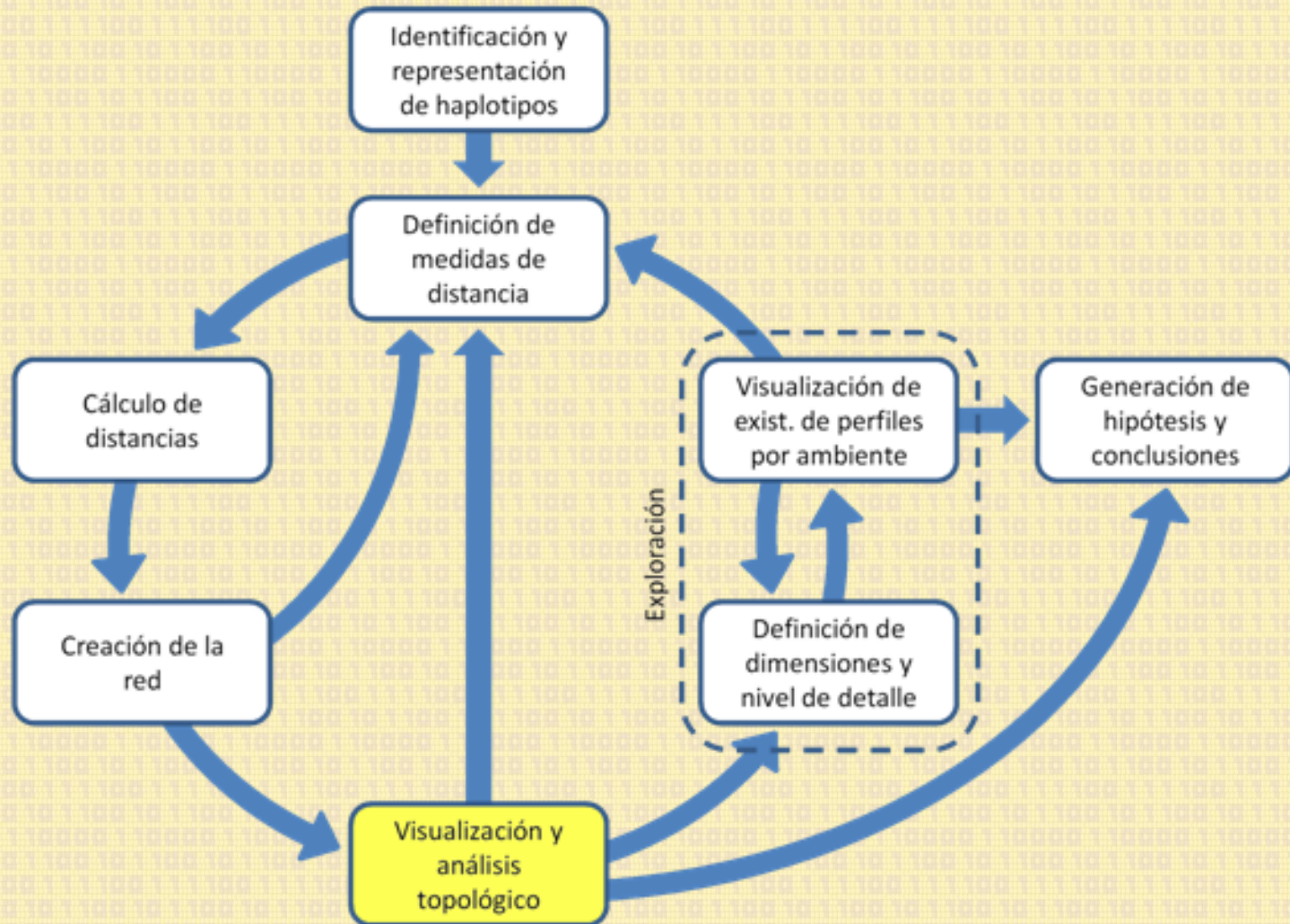


# Creación de la red





# Visualización y análisis topológico





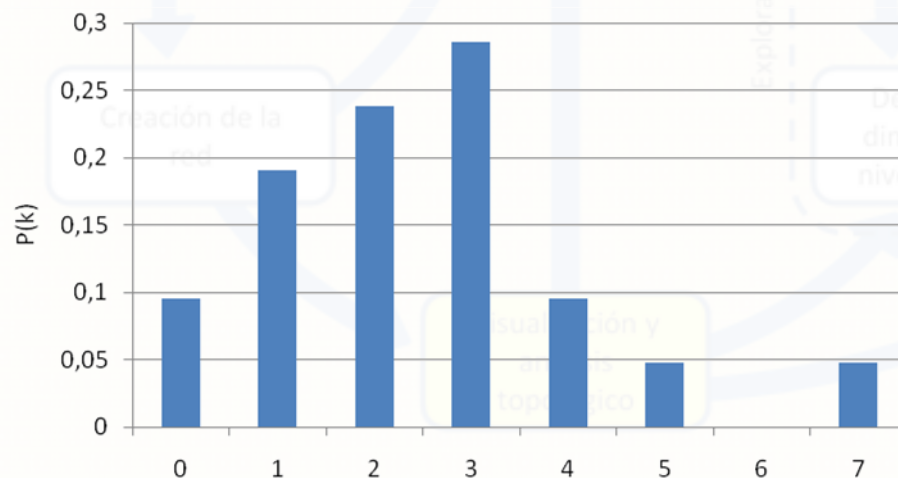
# Visualización y análisis topológico

Clustering coefficient:  $CC_i = \frac{2c_i}{k_i(k_i - 1)}$   $CC = \frac{\sum_{i=1}^n CC_i}{n} = 0,246$

Diámetro = 5

Distancia promedio = 2,767

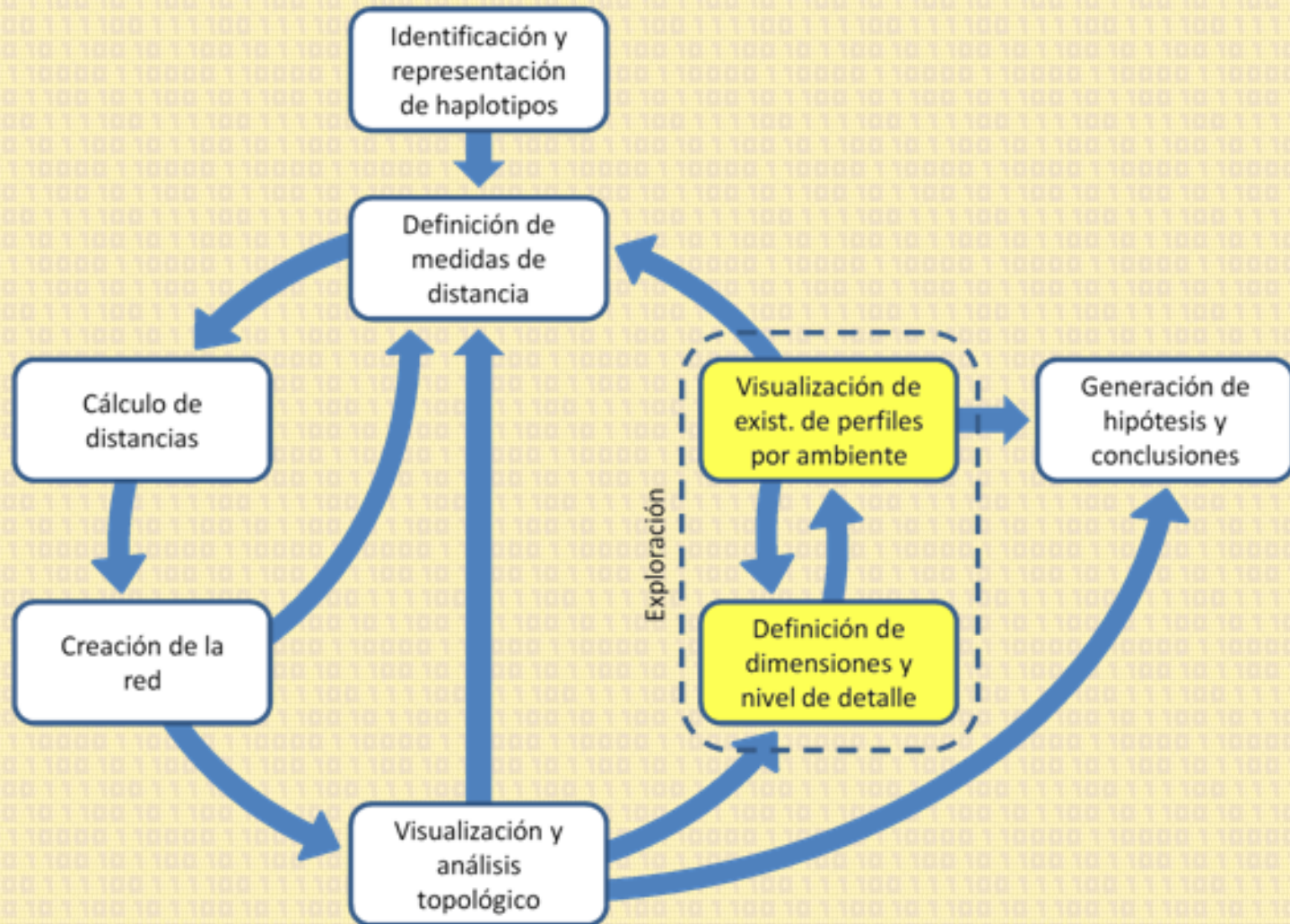
Distribución de grado de conectividad:



$$k_i = \sum_{j=1}^N d_{ij} \mid d_{ij} = 1$$

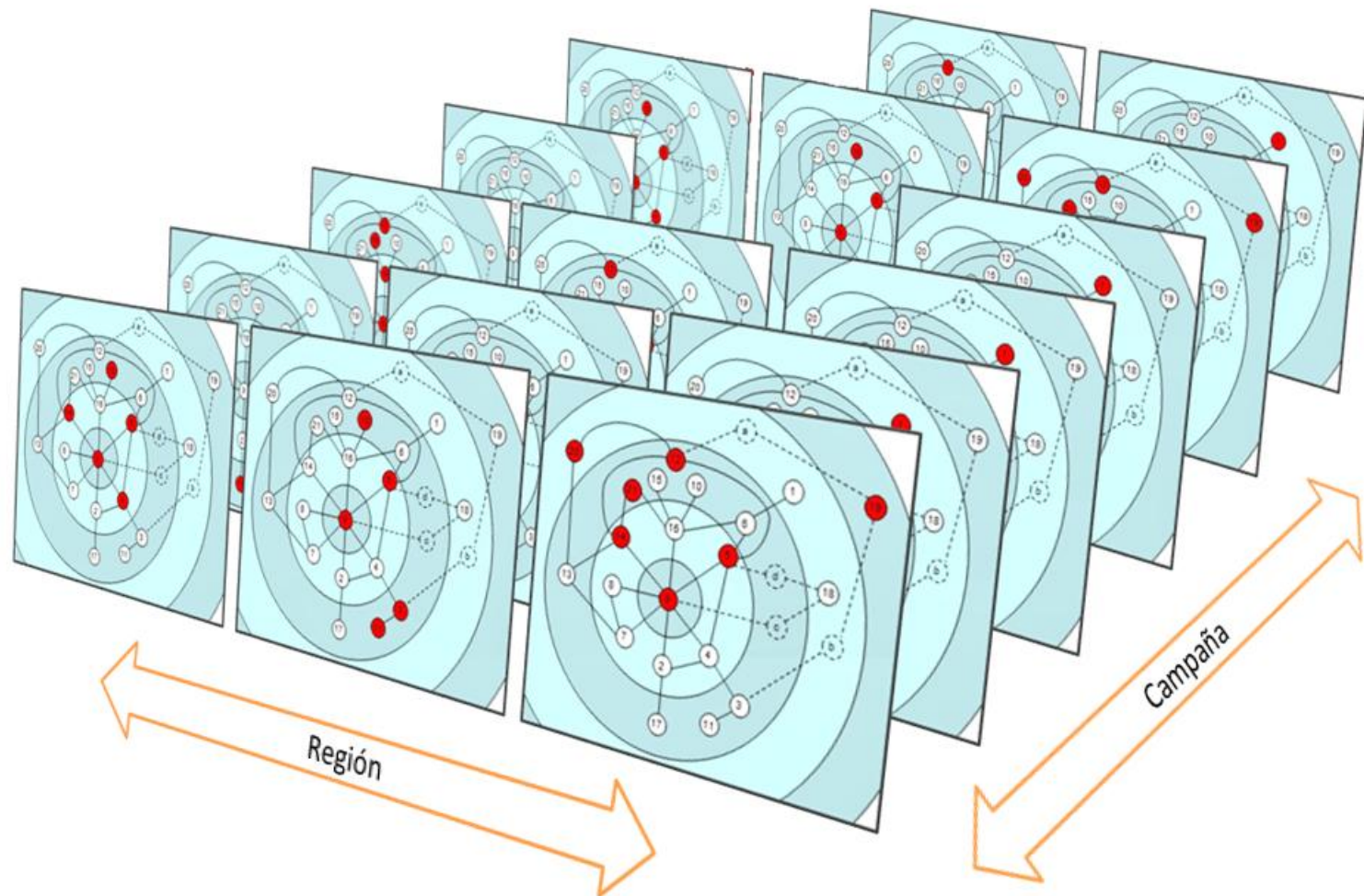


# Exploración



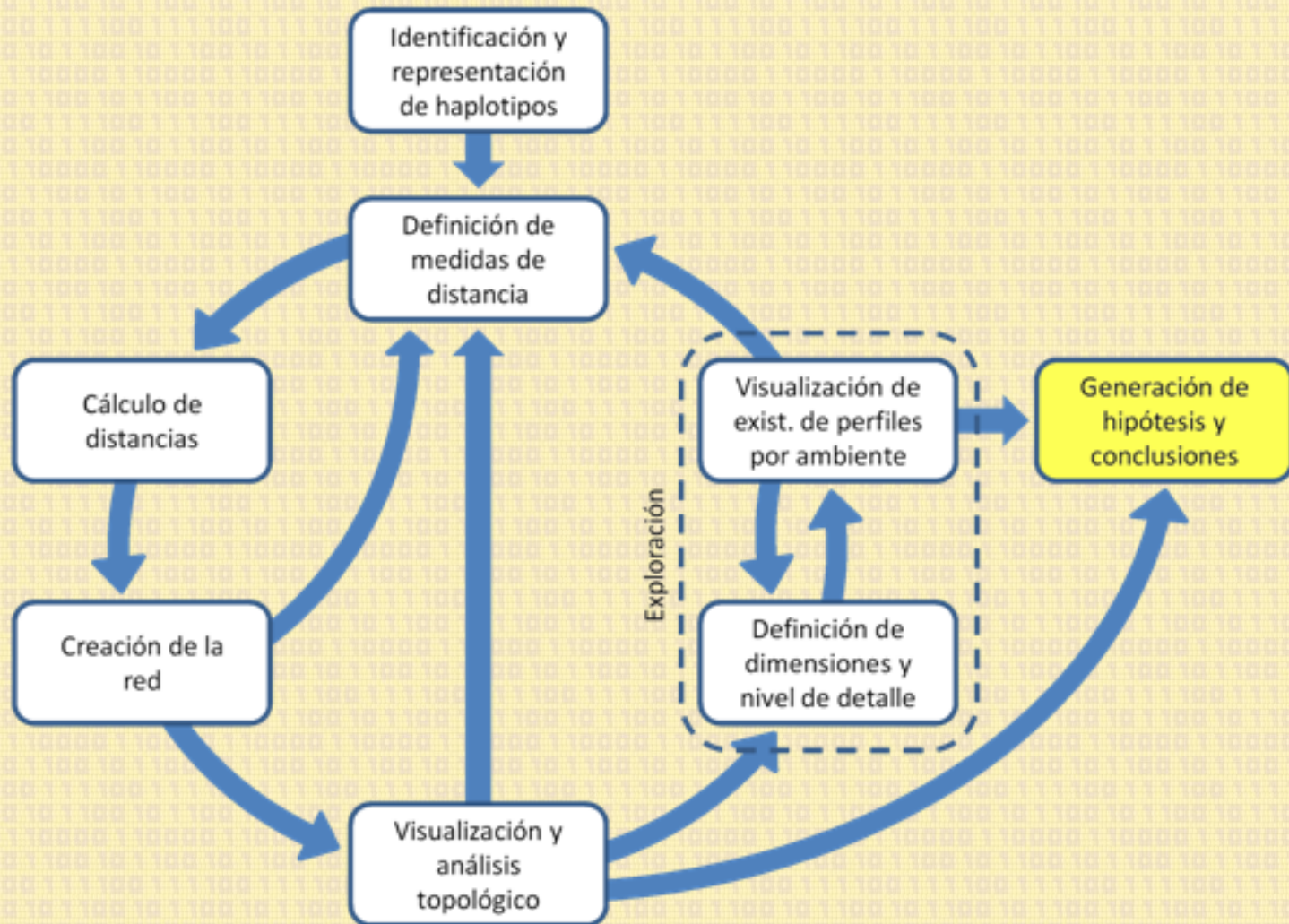


# Exploración



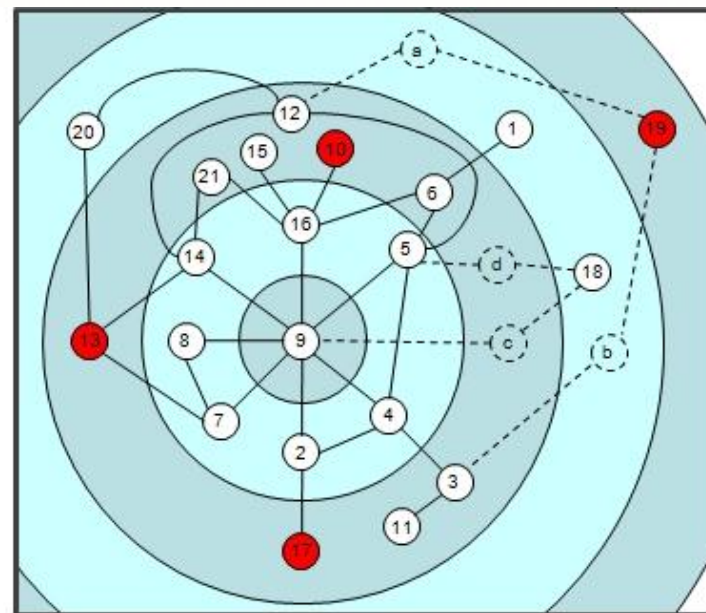
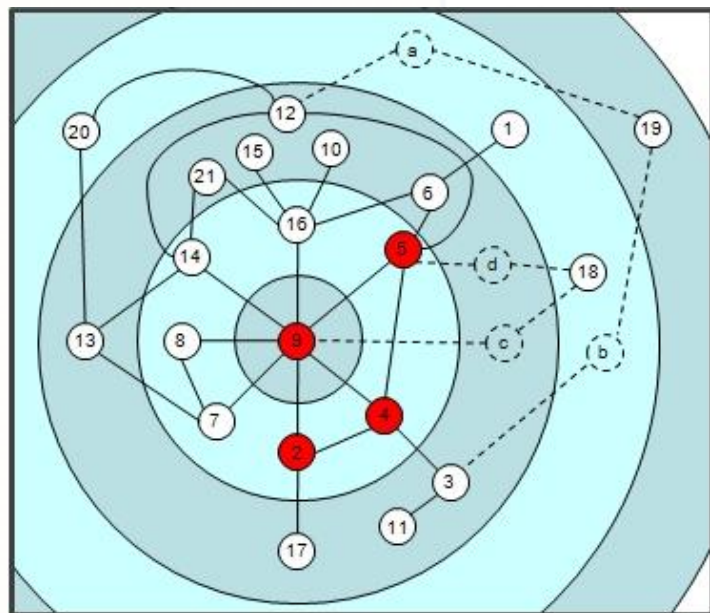


# Generación de hipótesis y conclusiones





# Generación de hipótesis y conclusiones



$$SDH_A = \sum_{i=1}^{n_A-1} \sum_{j=i+1}^{n_A} d_{ij}$$

donde:

$SDH_A$ : suma de distancias entre los haplotipos del ambiente A

$n_A$ : cantidad de haplotipos del ambiente A

$d_{ij}$ : distancia entre el haplotipo i y el haplotipo j



# Generación de hipótesis y conclusiones

$$E(SDH_A) = \sum_{i=1}^{n_A-1} \sum_{j=i+1}^{n_A} (1 - (1 - P(h_i))^{n_A}) (1 - (1 - P(h_j))^{n_A}) d_{ij}$$

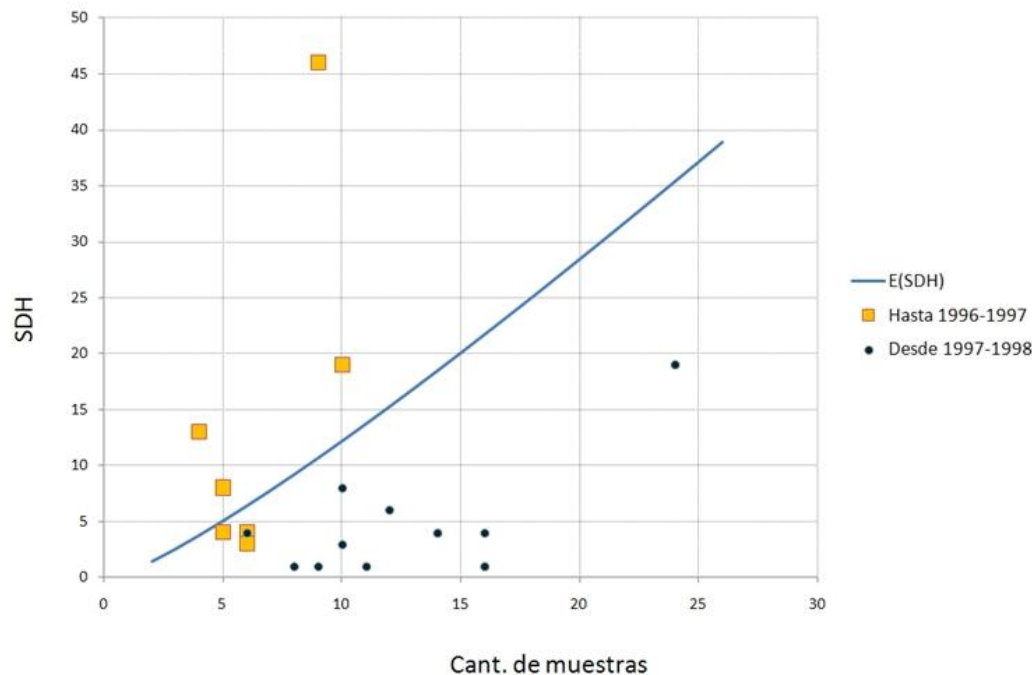
**donde:**

$E(SDH_A)$ : valor esperado de SDH del ambiente A

$n_A$ : cantidad de haplotipos del ambiente A

$d_{ij}$ : distancia entre el haplotipo i y el haplotipo j

$P(h_i)$ : Probabilidad de existencia del haplotipo i



Generación de  
hipótesis y  
conclusiones

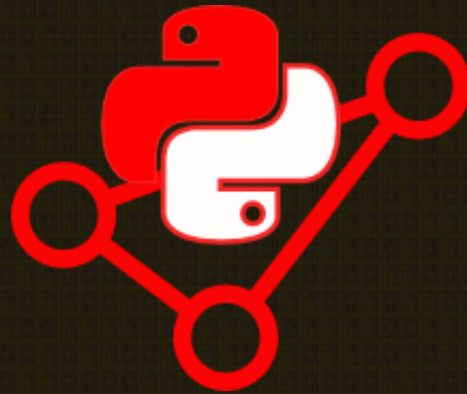


# Conclusiones del proyecto

- Según el índice calculado, la variabilidad del Mal de Río Cuarto virus, ha disminuido con el tiempo, habiendo una clara división del indicador en la campaña posterior a la epidemia de la campaña 1996/97.
- La utilización de redes en el proceso de KDD resultó muy satisfactoria y logró resaltar un comportamiento del objeto de estudio que no había sido evidente hasta el momento.
- En un proceso centrado en la persona (*human-centered*), donde la creatividad y experiencia del analista juega un rol fundamental, la herramienta propuesta es capaz de ofrecer una perspectiva novedosa y complementaria con las demás técnicas del proceso de KDD



# yatel





# Arquitectura

yatel.gui [pkg]

yatel.conversors [pkg]

yatel.db

yatel.weight

lpython

Pilas

QScintilla

csvcool  
graphtool

yatel.dom

numpy

Qt

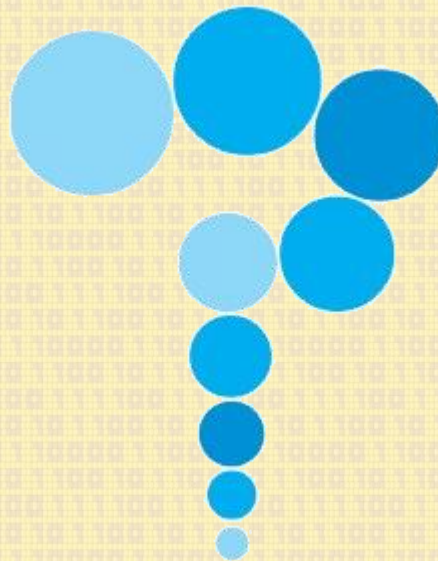


# Como librería

```
1 from yatel import dom, weight, db
2
3 haps = [dom.HaploType("hap0", b1=0, b2=1, b3=1),
4         dom.HaploType("hap1", b1=0, b2=1, b3=0),
5         dom.HaploType("hap2", b1=1, b2=1, b3=1)]
6
7 facts = [dom.Fact(haps[0].hap_id,
8                  date="25/5/10", place="Rio Cuarto", clima="lluvioso"),
9         dom.Fact(haps[1].hap_id,
10                  date="25/6/10", place="Rio Tercero",
11                  clima="soleado", estado_maiz="muerto"),
12         dom.Fact(haps[2].hap_id,
13                  date="25/5/11", place="Rio Cuarto", clima="nublado"),
14         dom.Fact(haps[0].hap_id,
15                  date="25/5/10", clima="soleado", estado_maiz="aislado")]
16
17 hamm = weight.Hamming()
18 edges = [dom.Edge(hamm.weight(haps[0], haps[1]),
19                  haps[0].hap_id, haps[1].hap_id),
20         dom.Edge(hamm.weight(haps[1], haps[2]),
21                  haps[1].hap_id, haps[2].hap_id)]
22
23 conn = db.YatelConnection("sqlite", "/home/juan/slides.db")
24 conn.init_with_values(haps, facts, edges)
25
26 # SECOND TIME!
27 conn.init_yatel_database()
```



Preguntas...





# Gracias!!

