

**XI**  
**JORNADAS**  
**STIC**  
**CCN-CERT**

**Ciberamenazas\_**  
**El reto de compartir**

**#XIJornadasCCNCERT**

**MADRID.**  
**13 Y 14 DE DICIEMBRE**  
**2017**

# **NEBULARITY REPORT**

## **Anticipándonos al Análisis de Malware**



- Juan José Torres
- Accenture
- [juan.j.torres.garcia@accenture.com](mailto:juan.j.torres.garcia@accenture.com)



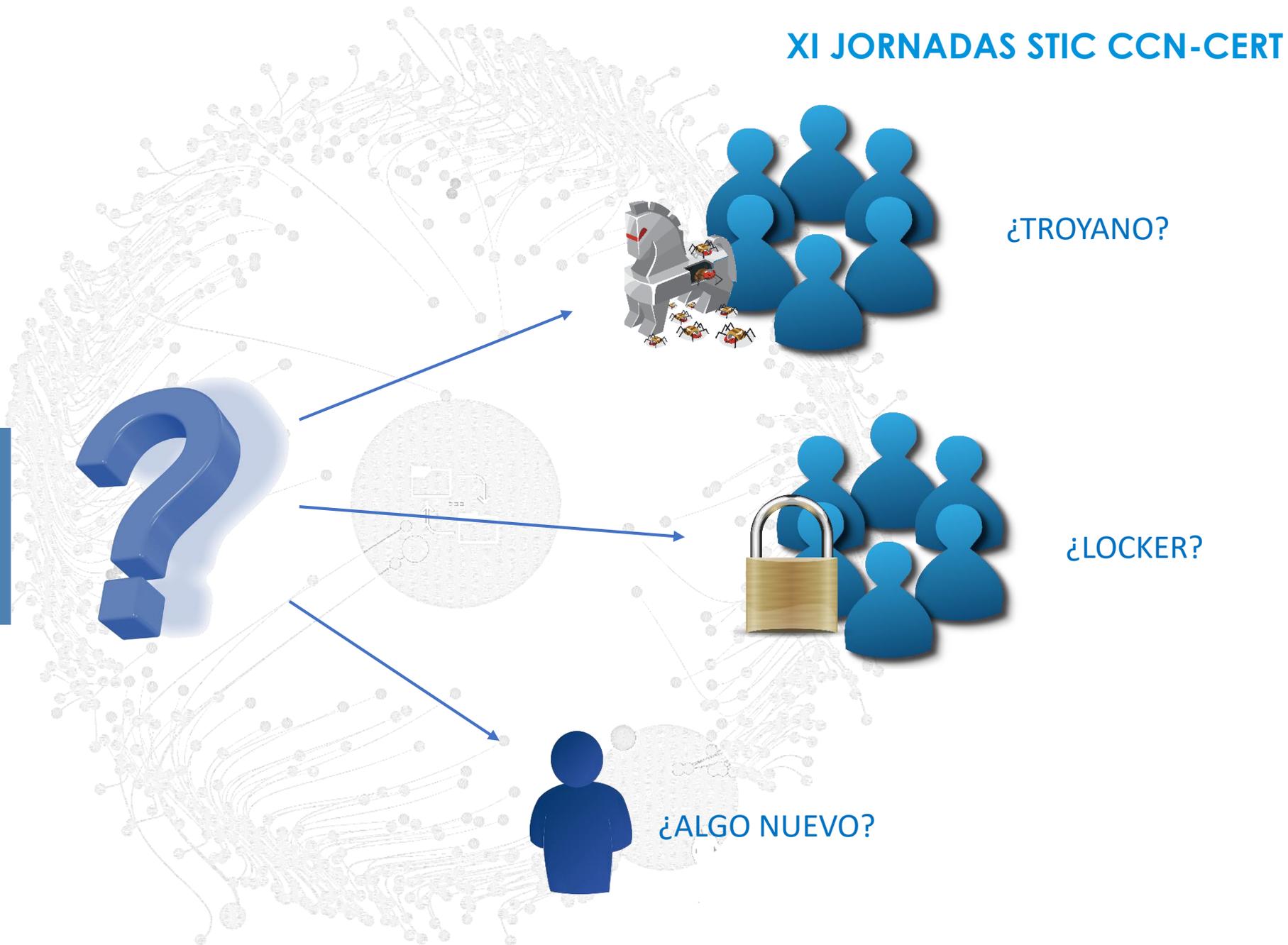
- Francisco Javier Sucunza
- Innotec Entelgy
- [francisco\\_sucunza@innotecsystem.com](mailto:francisco_sucunza@innotecsystem.com)



b93dc895e947e320b432b6e3631f51d1 fcc4915a48e0e0bc68bb803cae5a014a

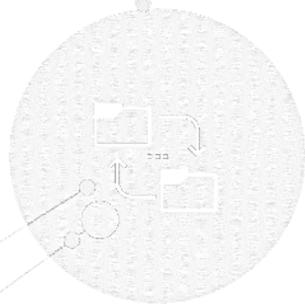


MUESTRA NUEVA



# Hashes Tradicionales

- MD5, SHA1, SHA256...
- Efecto Avalancha.



# Fuzzy Hashing

10101101  
00101110  
01010111  
00110110

MD5: b93dc895e947e320b432b6e3631f51d1  
Fuzzy Hash: 3a2b78f3c5

10101101  
00101110  
01011001  
00110110

MD5: fcc4915a48e0e0bc68bb803cae5a014a  
Fuzzy Hash: 3a2b78f325

# SSDeep

96:EQOJvOI4ab3hhiNFXc4wwcweomr0cNJDBoqXjmAHKX8dEt001nfEhVluX0dDcs:3mzpAsZpprbshfu3oujldENdp21



# Fuzzy Hashing

10011100
10011010
01010110
01010101
10010111
10101011
01001011
01010101
00010110
101011011
11000111
00101101
10101011
00101010
01001010
10101010

3a2b7

10011100
10011010
01010110
01010101
10010111
10101011
01001011
01010101
00010110
101011011
11000111
00101101
10101011
<b>10100111</b>
01001010
10101010

3a2c7

# SSDeep Deficiencias

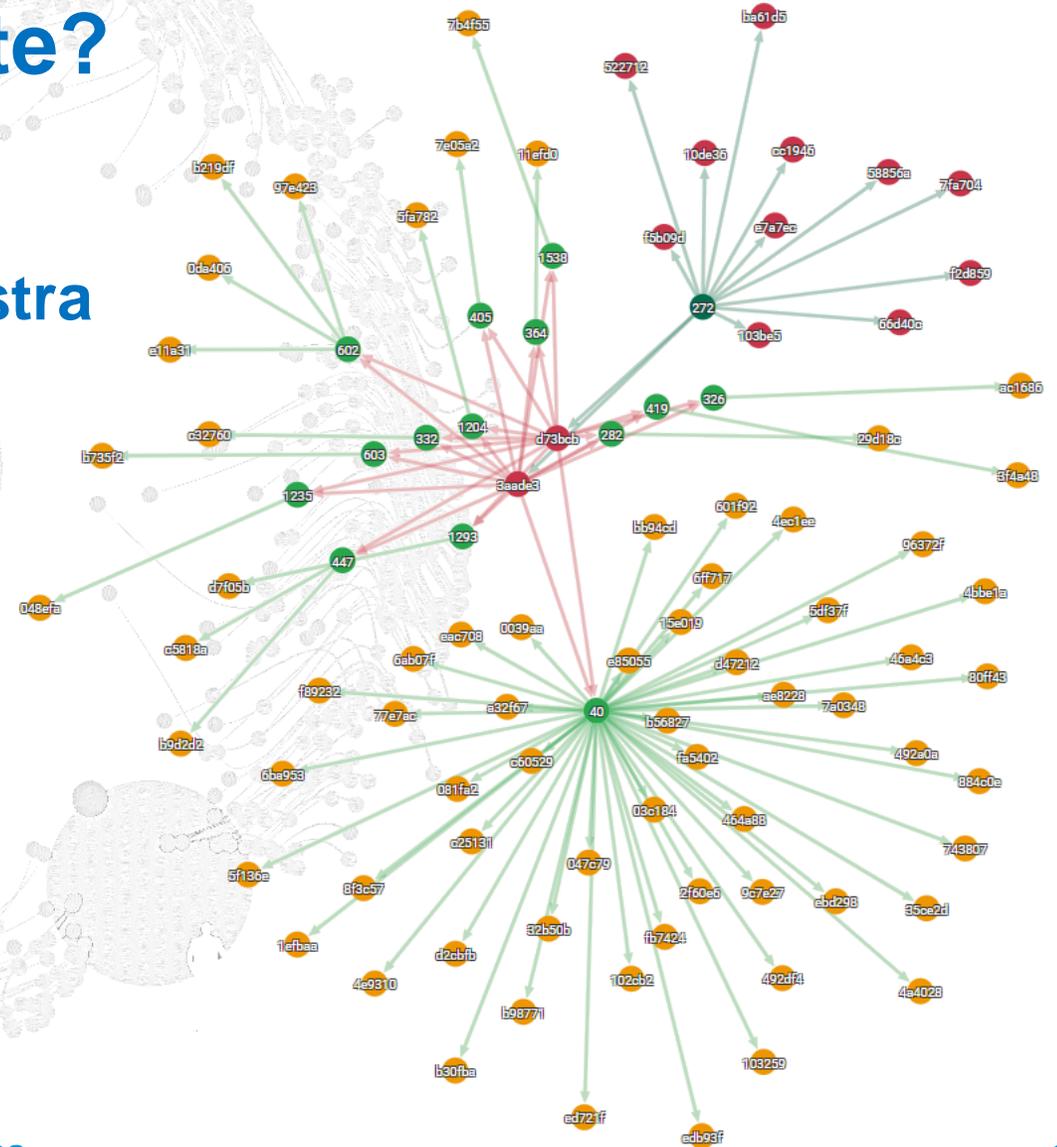
- Colisiones en los hashes de los bloques.
- Incapacidad de comparar muestras de tamaños muy diferentes.
- No tiene en cuenta el tamaño de los bloques al calcular la similitud.
- Ineficiencia al calcular la similitud entre dos hashes.

# Similarity Uniform Fuzzy Hash

- Decidimos implementar nuestro propio algoritmo de Fuzzy Hashing.
  - Preciso.
  - Eficiente.
  - Puede comparar muestras de tamaños muy diferentes.

## ¿Suficiente?

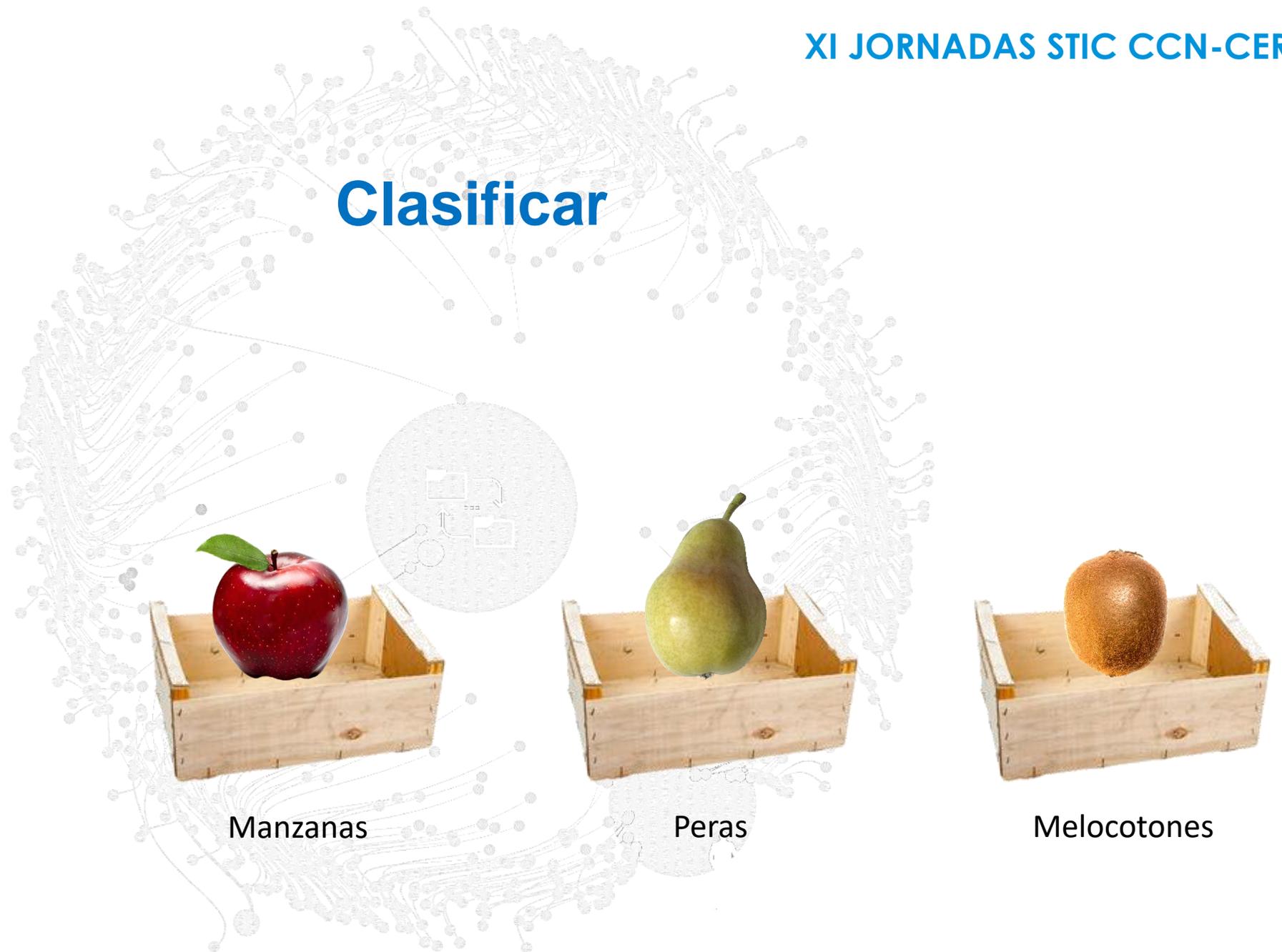
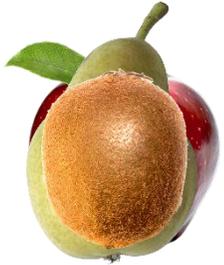
- Ya podemos determinar si una nueva muestra es similar a algunas de las que ya teníamos.
- Pero no es suficiente para clasificarla.
- Es necesario un algoritmo de **clustering / agrupamiento**.



## Clasificar vs Agrupar

- Los algoritmos de **clasificación** tienen definidas las posibles distintas clasificaciones antes de empezar a analizar las muestras.
- Los algoritmos de **agrupamiento**, en cambio, crean los grupos al vuelo, conforme las van analizando.

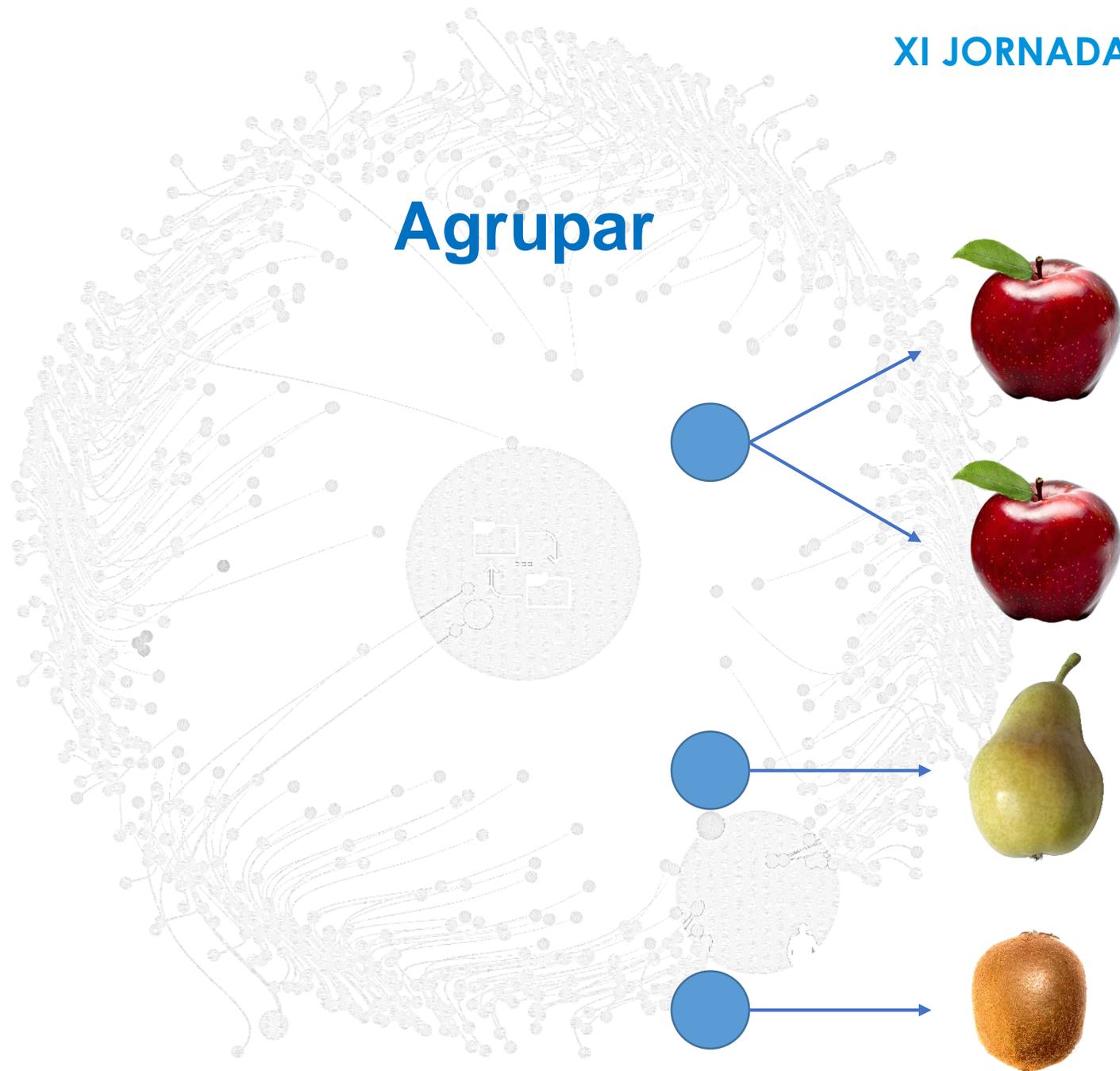
# Clasificar



Manzanas

Peras

Melocotones



# Clustering de Malware por Similitud

- Buscamos investigaciones ya realizadas de clustering de malware por similitud.



# Clustering de Malware por Similitud

- Los algoritmos de agrupamiento funcionan mejor con muchas muestras.
- Pero cuantas más muestras manejan, más pesados se vuelven.

# Machine Learning vs Deep Learning

- Los algoritmos de **Machine Learning** tienen predefinidas las características: tamaño, color, longitud del rabo, etc.
- Los algoritmos de **Deep Learning** las definen según van analizando muestras.

# Nebularity

- Necesitamos un algoritmo de clusterizado **escalable** y basado en **Deep Learning**.
- Decidimos crear nuestro propio algoritmo de agrupamiento: **Nebularity**.
- Está basado en centroides.
- Es preciso, eficiente, escalable y puede ir recibiendo muestras poco a poco.



# Nebularity

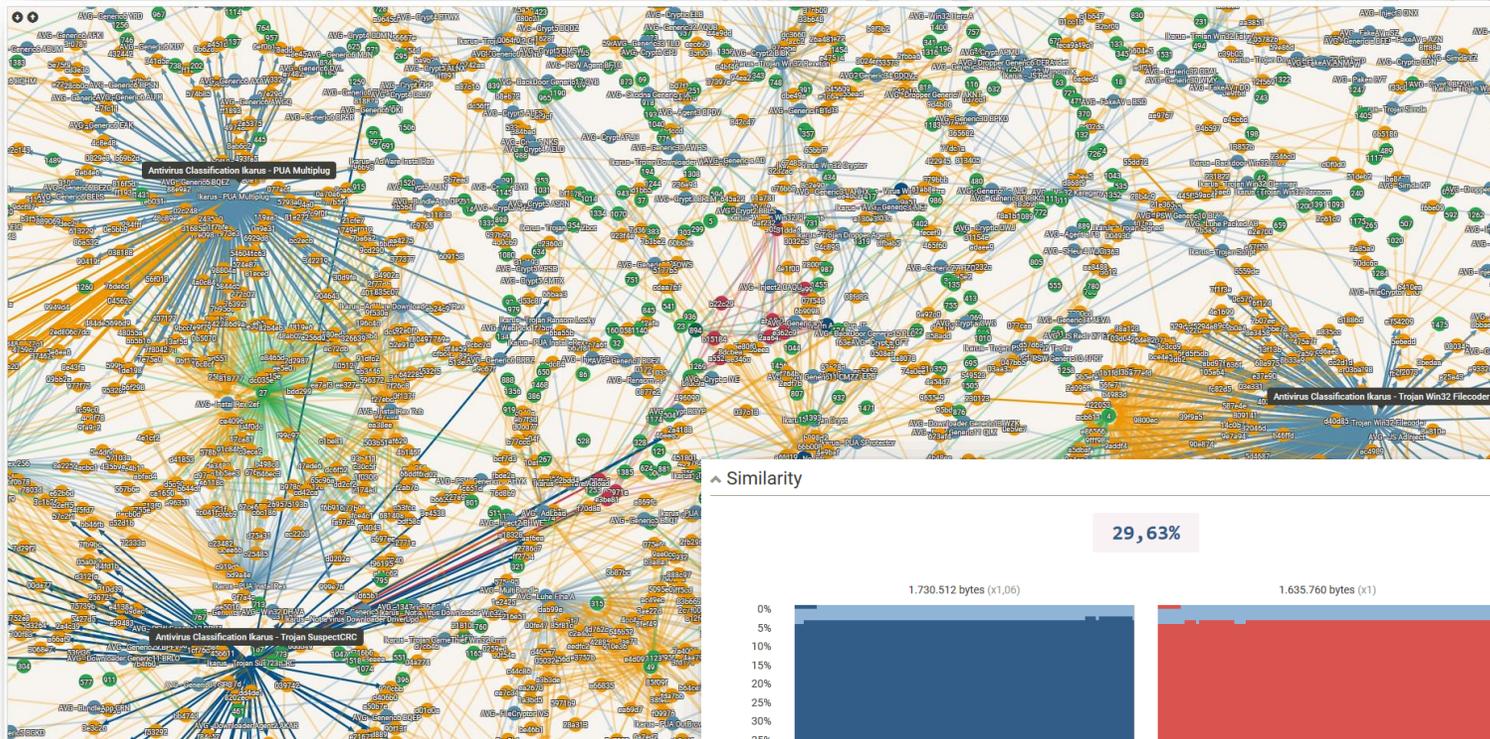
- Hacemos las primeras pruebas. Parece que funciona.
- Pero: sólo agrupa, no clasifica.
- Vemos las muestras agrupadas, pero ¿qué son?

# Antivirus

- Integramos Nebularity con antivirus.
- La clasificación de los antivirus de la muestra más representativa de cada grupo da una idea de lo que es el grupo.
- Podemos clasificar una muestra nueva en función de la clasificación que dan los antivirus de sus muestras similares... **¡Objetivo conseguido!**
- O no...

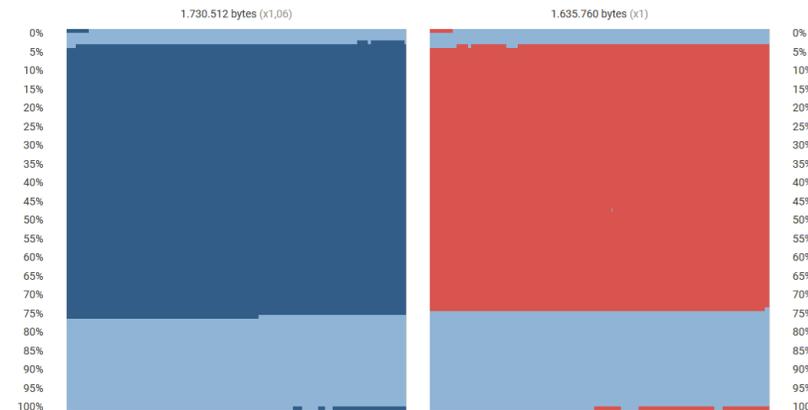


# Demostración



Similarity

29, 63%



Exclude Singleton Samples Page: 1 of 14 Size: 25

MRS Name	MRS MD5	MRS Value	N° Samples
VirusShare_26957575d3931b4201ed64454f0adb44	26957575d3931b4201ed64454f0adb44	54,66%	198
Sample Name	Sample MD5	Sim. MRS	Groups IDs
Share_26957575d3931b4201ed64454f0adb44	26957575d3931b4201ed64454f0adb44	100%	27
Share_bcb128fa23d4daf6806d5ed2b14547e0	bcb128fa23d4daf6806d5ed2b14547e0	99,44%	27
Share_0a9e31ead2486e7aecdf8561095028bf	0a9e31ead2486e7aecdf8561095028bf	99,44%	27
Share_cd42ca5ef7506f6667cddb5370a2065	cd42ca5ef7506f6667cddb5370a2065	99,14%	27
Share_a9c67fd1235f05c0a5c01cfd7b747ce	a9c67fd1235f05c0a5c01cfd7b747ce	99,13%	27
Share_eaf629018495725c3a5f92748f79a1e4	eaf629018495725c3a5f92748f79a1e4	99,13%	27
dfdfde8c43f4c77eb67b516	1f0306754dffde8c43f4c77eb67b516	99,13%	27
2e2c3d5e0b631c99cba473	2c9f0f1dc9de2c3d5e0b631c99cba473	99,13%	27
f9b5ab9be93b4f93d8907	f04043e1419895ab9be93b4f93d8907	99,13%	27
8d345f33492da25e01bca6b	4819e6d4c8d345f33492da25e01bca6b	99,13%	27
5a8545e34e4ae8ed5795623	b77b01ad65a854e34e4ae8ed5795623	99,13%	27
ec2208205ec2692d2dc552e6c7dd9b45	ec2208205ec2692d2dc552e6c7dd9b45	99,13%	27
cce7085ba23216e232493a2	f99c97d3dce7085ba23216e232493a2	99,13%	27
8654df5272e7ca015e9efdd	81e27250e8654df5272e7ca015e9efdd	99,13%	27
18667da8d6628516075963d	c697eedf618667da8d6628516075963d	99,13%	27
0d99180faa492e365a68755	eeb361715d099180faa492e365a68755	99,13%	27
9d9cb3f2b93545a273cfe50	119aa1c7cd9cb3f2b93545a273cfe50	99,13%	27
5f616f6a5131969c810ccc0	2e30963625f616f6a5131969c810ccc0	99,13%	27
41a9a9c8aea61e928ee699c	93b41128641a9a9c8aea61e928ee699c	99,13%	27
046abdf8f7e533a2590eef9	47ede6ede046abdf8f7e533a2590eef9	99,13%	27
cb85e7b65ab4bf837838b8	405127a56cb85e7b65ab4bf837838b8	99,13%	27
86297a09cabcf81774401c60	f9619549786397a09cabcf81774401c60	99,13%	27

**¡Muchas Gracias!**



## ➤ E-Mails

- [info@ccn-cert.cni.es](mailto:info@ccn-cert.cni.es)
- [ccn@cni.es](mailto:ccn@cni.es)
- [sat-inet@ccn-cert.cni.es](mailto:sat-inet@ccn-cert.cni.es)
- [sat-sara@ccn-cert.cni.es](mailto:sat-sara@ccn-cert.cni.es)
- [organismo.certificacion@cni.es](mailto:organismo.certificacion@cni.es)

## ➤ Websites

- [www.ccn.cni.es](http://www.ccn.cni.es)
- [www.ccn-cert.cni.es](http://www.ccn-cert.cni.es)
- [www.oc.ccn.cni.es](http://www.oc.ccn.cni.es)

## ➤ Síguenos en

