



**Big
Data**



**Big
Networks**

Albert Díaz Guilera
<http://diaz-guilera.net>



complexitat.CAT

Física Estadística

- Principis del segle XX

Mecànica

10^{23} eqs diferencials
del moviment

Termodinàmica

variables
macroscopiques

$$S(E, N, V) = k_B \log(\Omega)$$

Meitat Segle XX

- Aplicaciones a la materia: física i química
- Transicions de fase: concepte d'universalitat

Darrer quart del segle XX

- Aplicacions a la biologia
- Sorgeix el concepte de complexitat
- Com ho havien fet abans: caos, sinergia,

Complexitat

- El tot no és la suma de les parts
- Comportaments emergents a partir dels comportaments individuals de les unitats que el formen
- Aplicacions a ciències socials



Complexitat

I think the next [21st] century will be the century of complexity. We have already discovered the basic laws that govern matter and understand all the normal situations. We don't know how the laws fit together, and what happens under extreme conditions. But I expect we will find a complete unified theory sometime this century. There is no limit to the complexity that we can build using those basic laws.

[Answer to question: Some say that while the twentieth century was the century of physics, we are now entering the century of biology. What do you think of this?]

— Stephen W. Hawking



Interdisciplinarietat

- Física
- Química
- Geologia
- Biologia
- Matemàtiques
- Computació
- Ciències socials

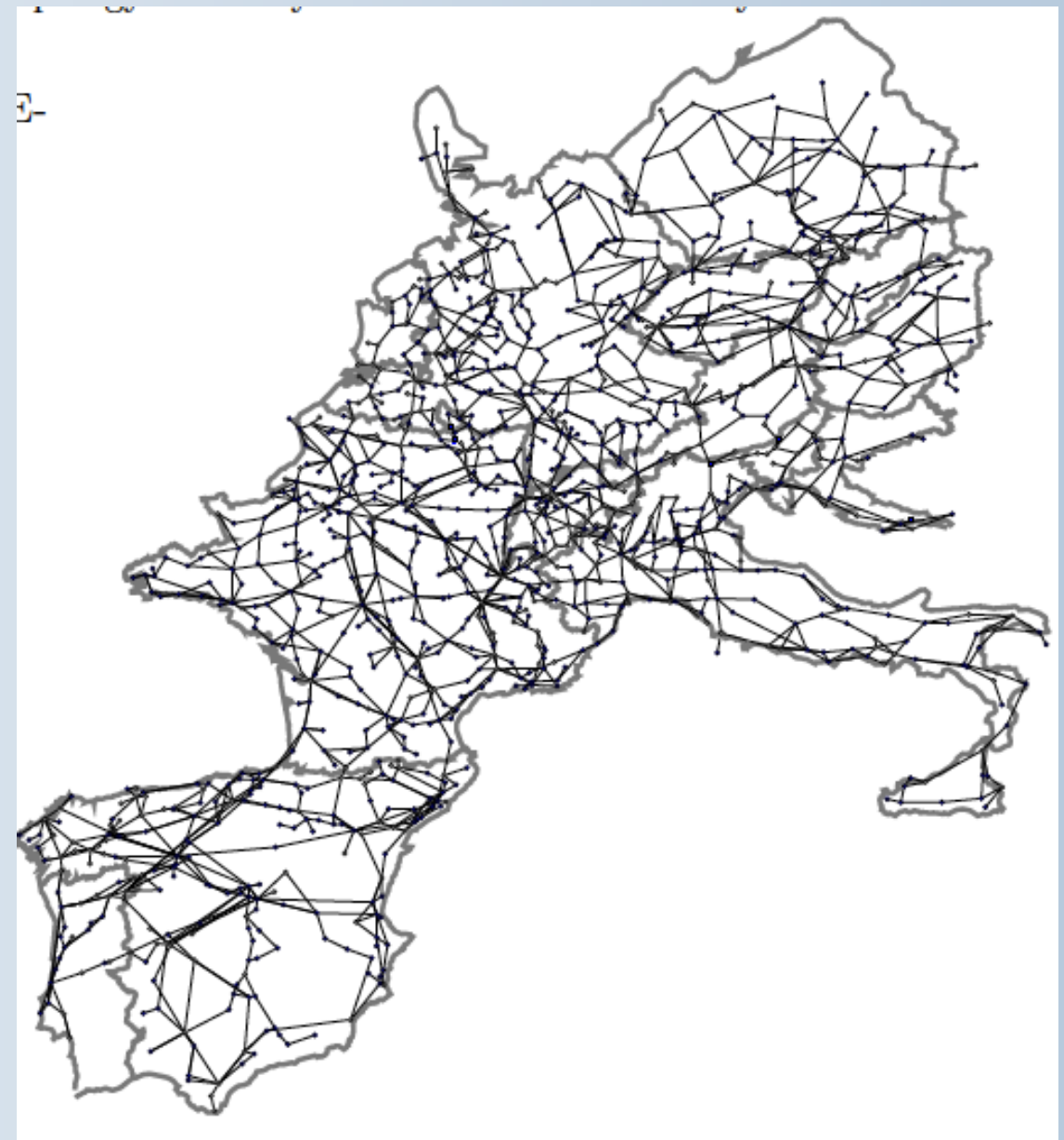
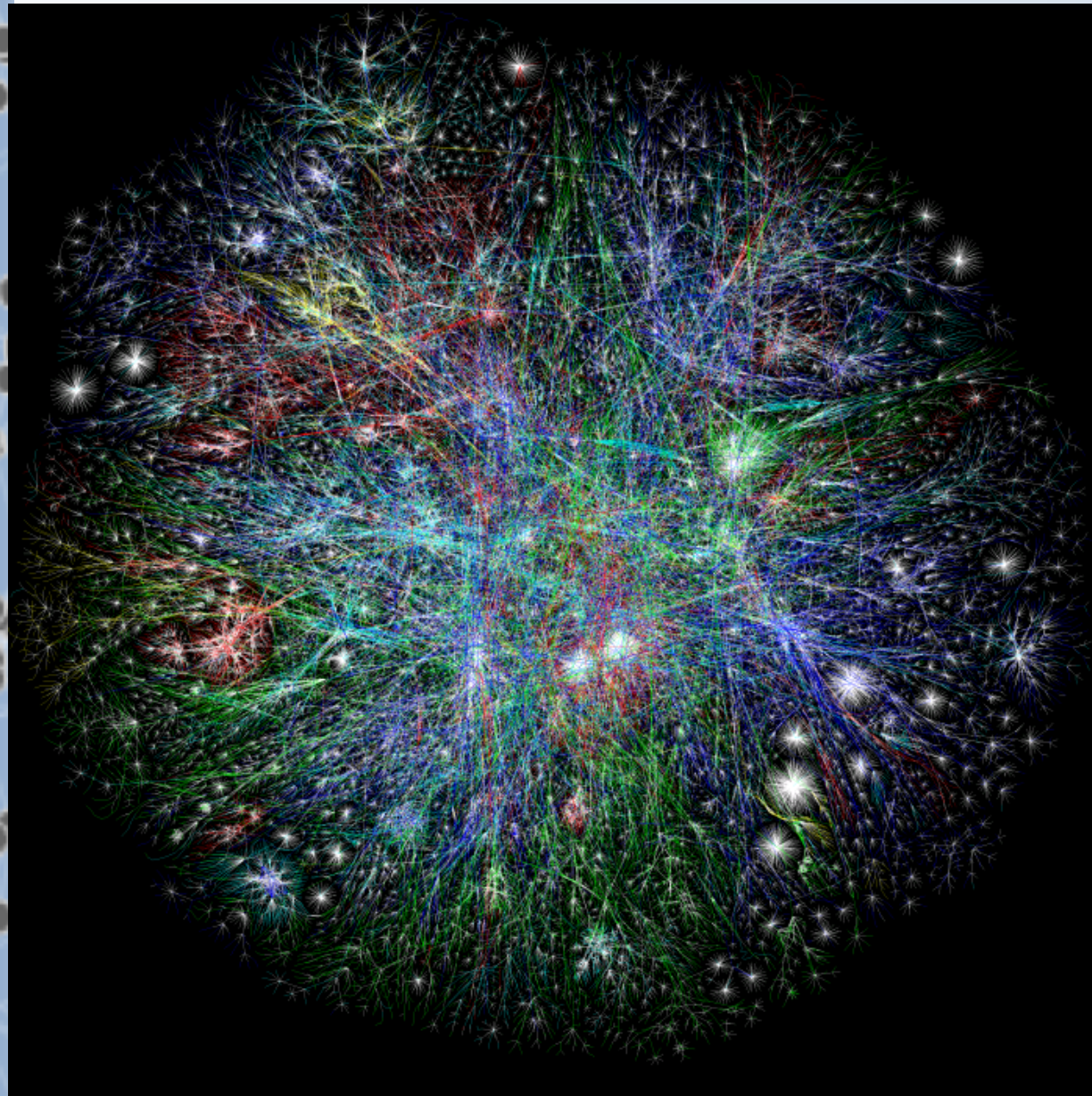
Principis del segle XXI

- Xarxes complexes
- Com ens afecta tot el que hem dit abans?

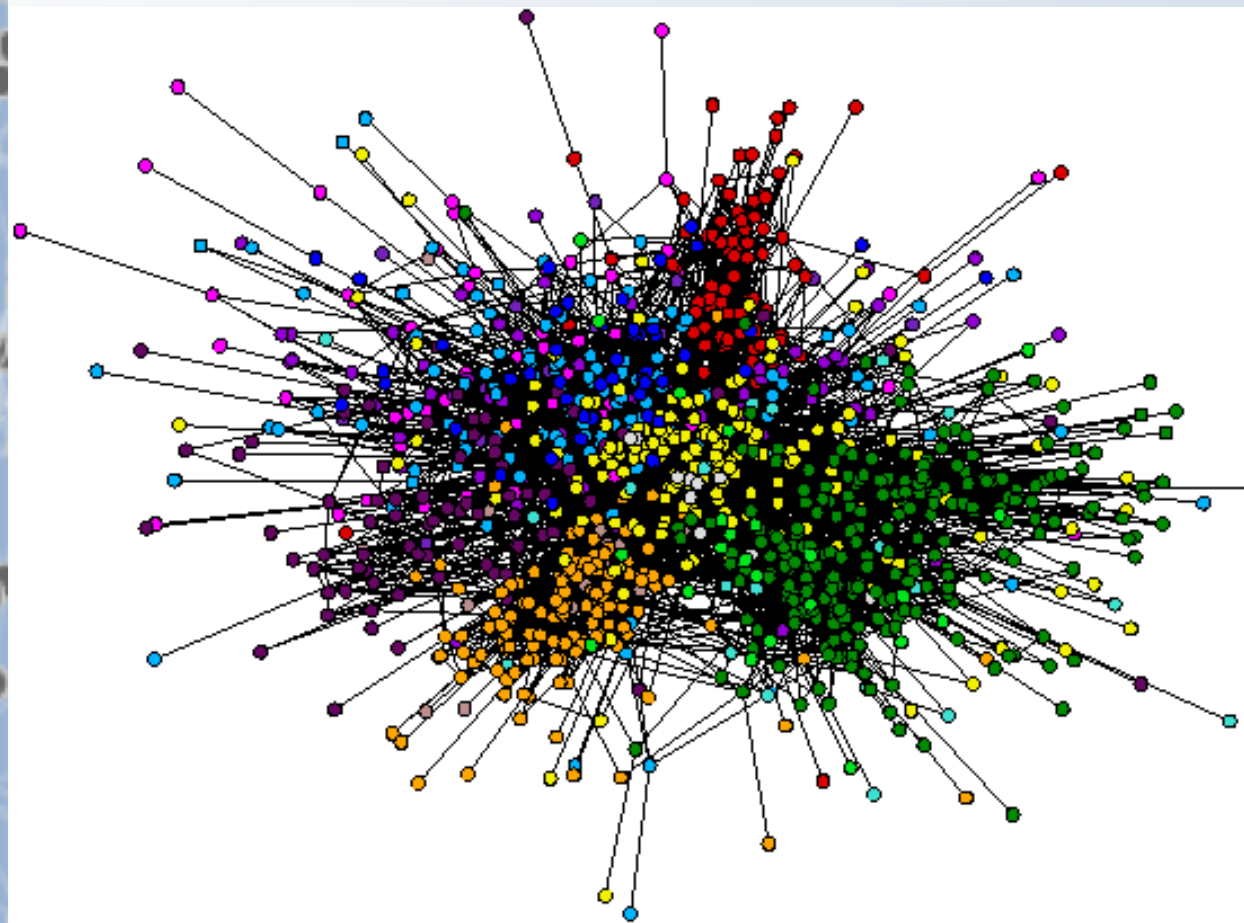
Xarxes complexes

- Mooolts exemples

Exemples

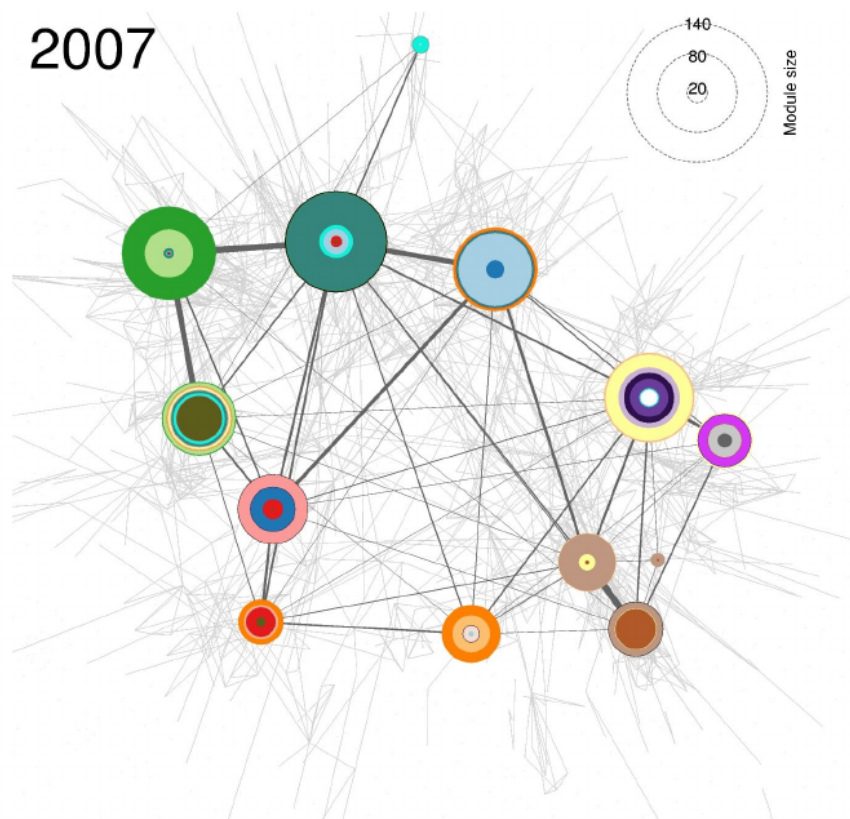


Examples

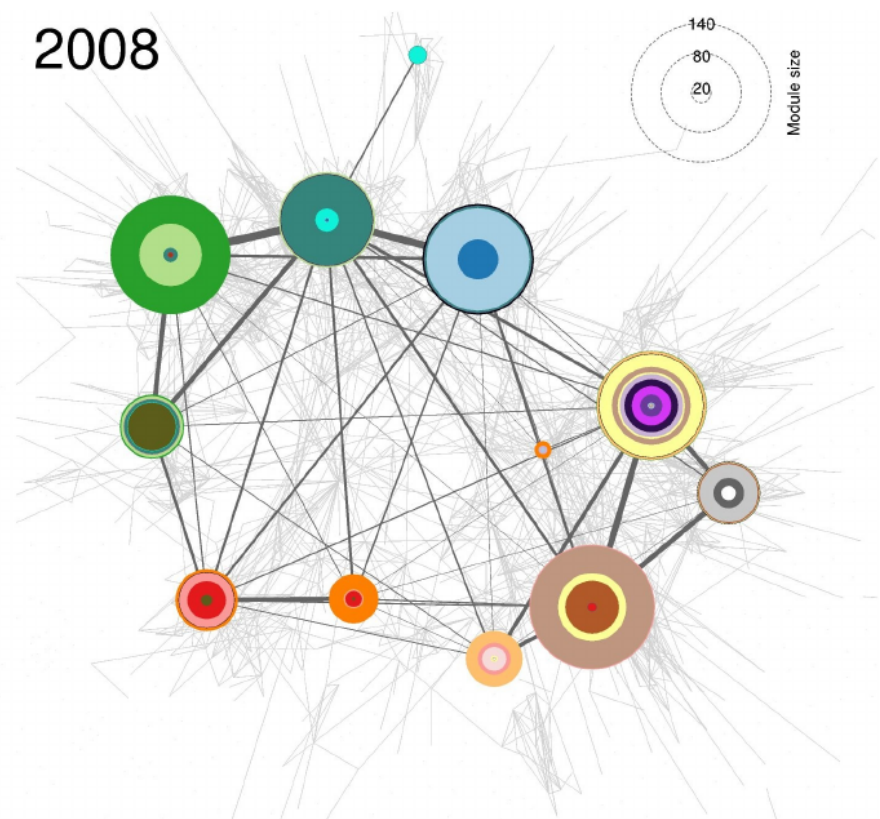


Exemples

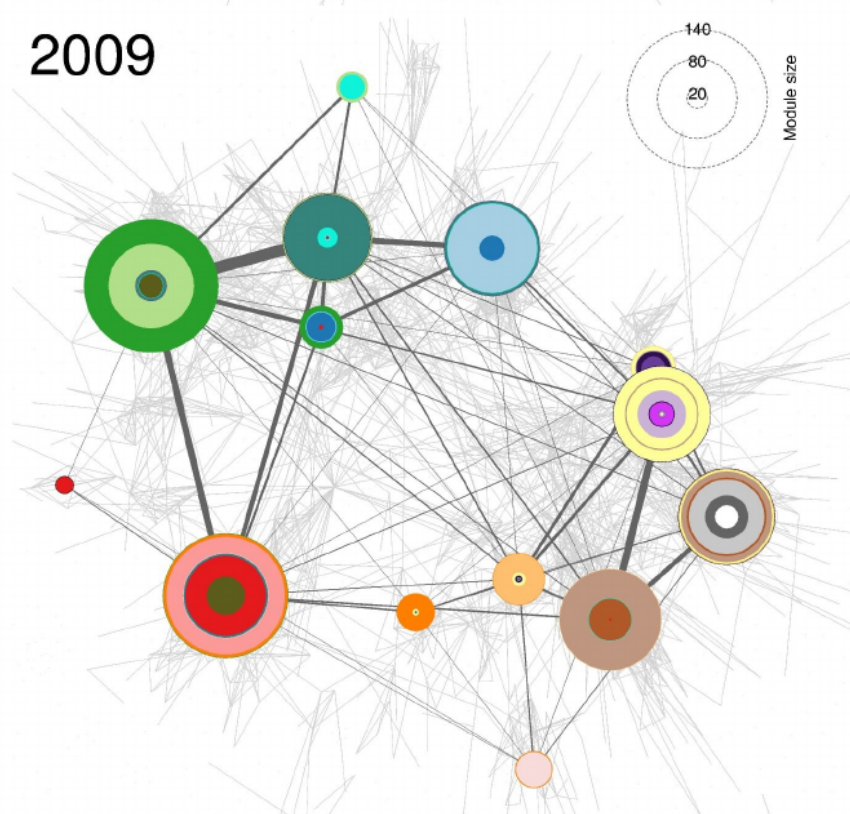
2007



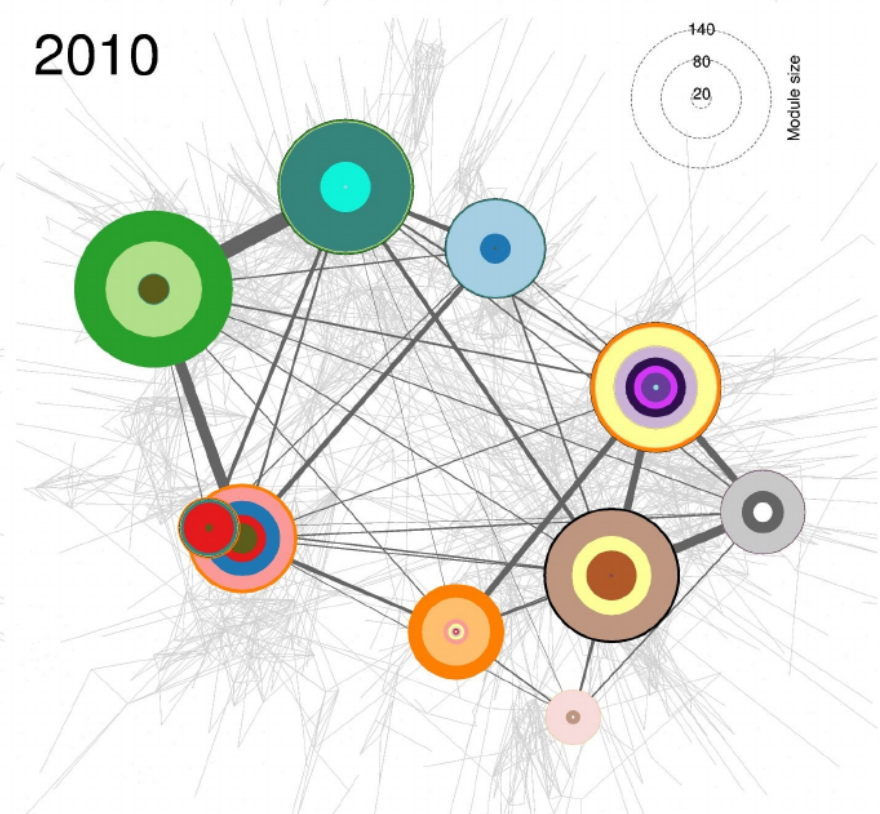
2008



2009



2010



Què tenen en comú?

- Nodes: entitats
- Enllaços: connexions reals (físiques)

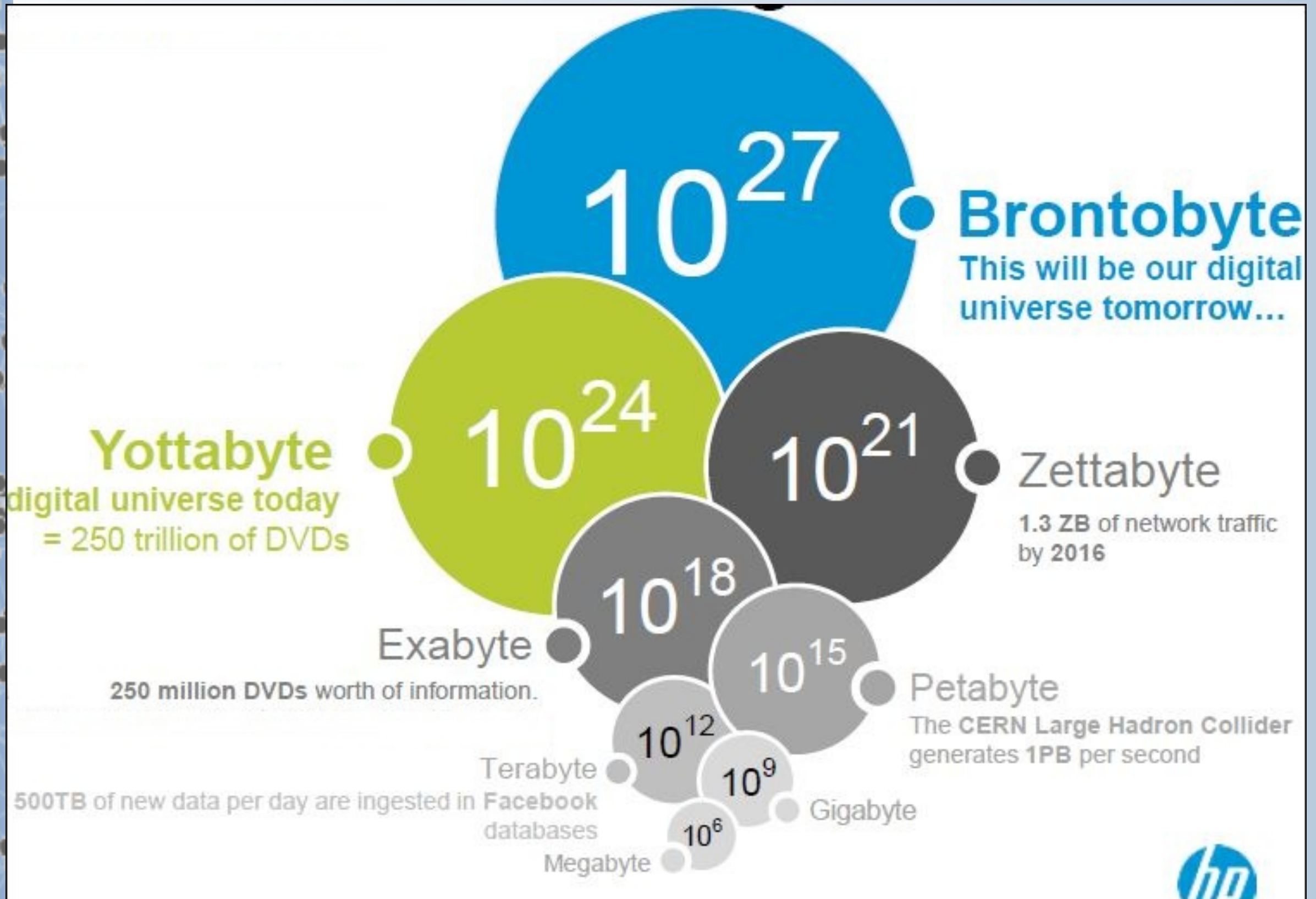
Dades

- No sempre les dades estan estructurades en forma de xarxa
- Creem estructures relacionals o funcionals
- Estructura vs. funcionalidad

Big Data: origen

- Ciències de la vida: -òmica
- Ciències socials:
 - mobilitat
 - comunicació, amb les màquines i amb altres humans
 - transaccions econòmiques
- Som els protagonistes de l'experiment. Intervenim en el propi resultat?

Ordres de magnitud



El ciclo de les dades

- Recollida
- Enmagatzament
- Pre-processat
- Anàlisi exploratori
- De les dades als models

Mètode científic

- hipotetitzar
- recollir
- mesurar
- caracteritzar
- modelitzar
- predir

Com aconseguim les dades

- Dades de les grans xarxes:
 - Facebook
 - Twitter
 - Google



<https://dev.twitter.com/>
<https://developers.google.com/maps/>
<https://developers.google.com/+api/>
<http://developers.facebook.com/>

Enmagatzemar?

- Bases de dades
- Relacions
- Optimizar cerques i accessos
- Problema computacional

Preprocessament

- Dades són incompletes
- Sorolloses
- Inconsistents

Que fer quan falten dades?

- Simplement les eliminem?
- Recorrim a mètodes estadístics d'inferència
- “Maximum Likelihood Estimation”
- Mostrejos incomplets

En xarxes

- No sempre coneixem els enllaços reals
- Pot haver-hi tant falsos negatius com falsos positius
- Inferència

Xarxes complexes

- No sempre els enllaços corresponen a connexions reals
- Xarxes estructurals vs. funcionals
 - En el cervell

Big Networks => Big Data

- Recollida dels grans proveïdors de serveis
- Les grans xarxes

Big Data => Big Networks

- Recollim dades i les convertim en xarxes, seguim el procés invers

Xarxa transport vs.

xarxa de mobilitat

- Dades estàtiques extretes dels mapes
- Dades dinàmiques extretes de registres:
 - GPS
 - enquestes origen - destí
 - trucades telefòniques

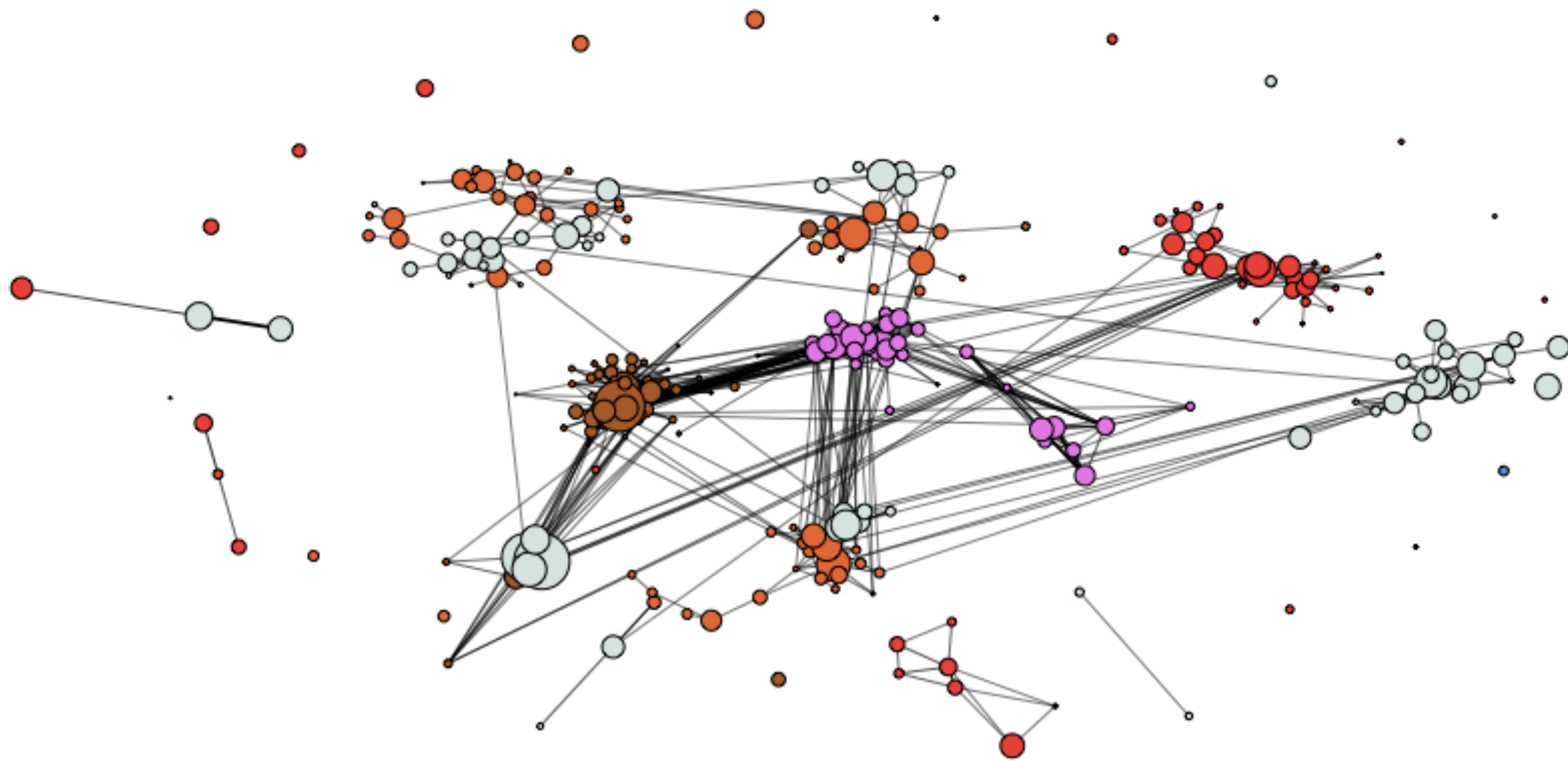
Xarxa de mobilitat

- Taula amb origens i destins?
- Procurem anar més lluny?

Importància de les xarxes

- Representació
- Taula versus xarxa

Xarxa ajuda a visualitzar



També a caracteritzar

- Mesures de centralitat
 - local: amb qui col.labora cadscú
 - global: qui està més “aprop” de la resta?

Rols dels diferents actors

- Individu per individu
- Qui és qui a l'estructura i/o a la funcionalitat

A nivell global

- Com caracteritzem una xarxa
- Oblidem els rols individuals
- Estadística

Estadística

- Distribuciones
- Moments de les distribucions

Complexitat

- Estadístiques “normals” poques i no serien gaire interessants
- Cues llargues: lleis potencials i tot el que això implica

Escalas

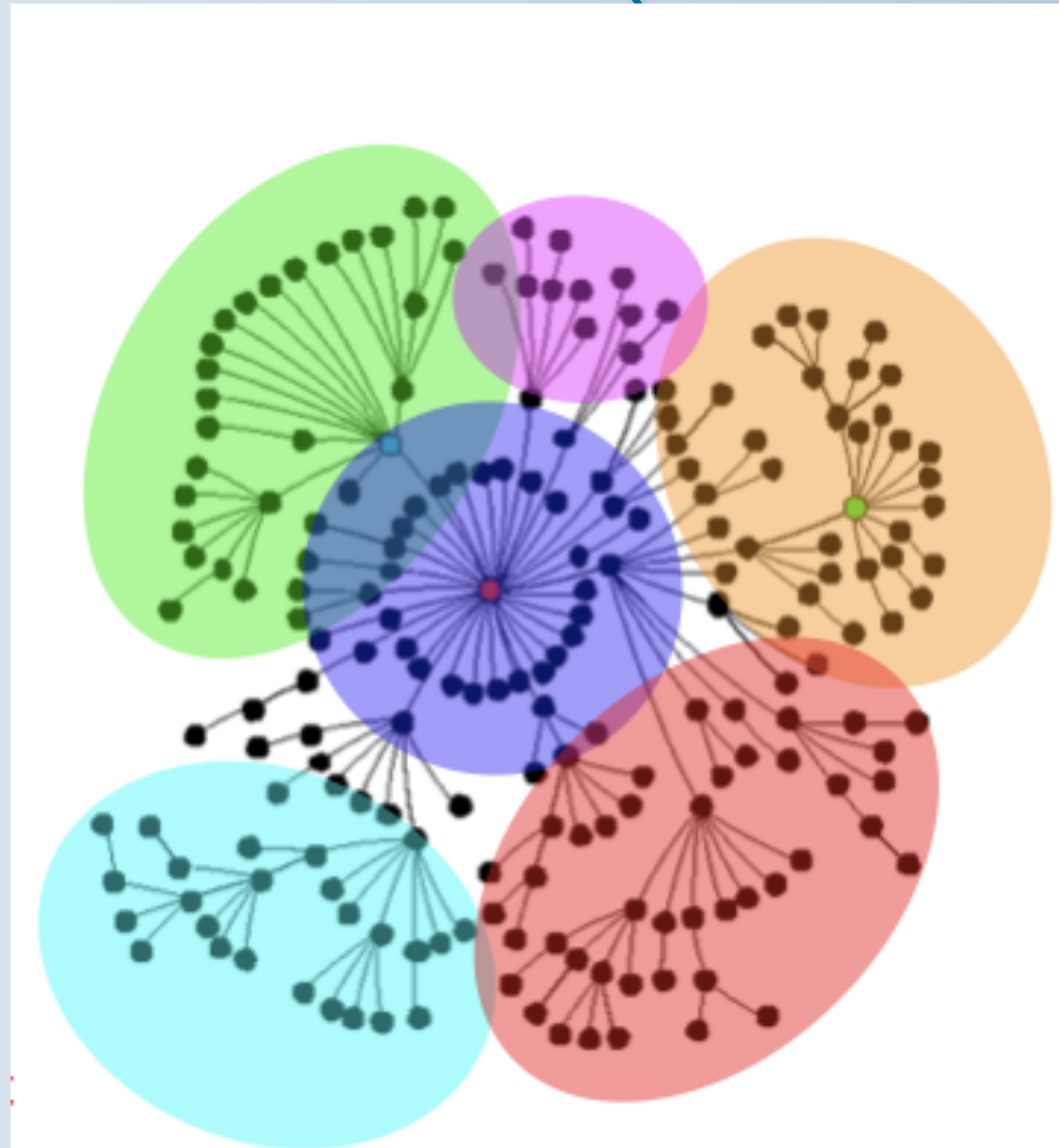


Totes les escales són importants (simultàniament)

- Local: microescala
- Global: macroescala
- Mitja: mesoescala

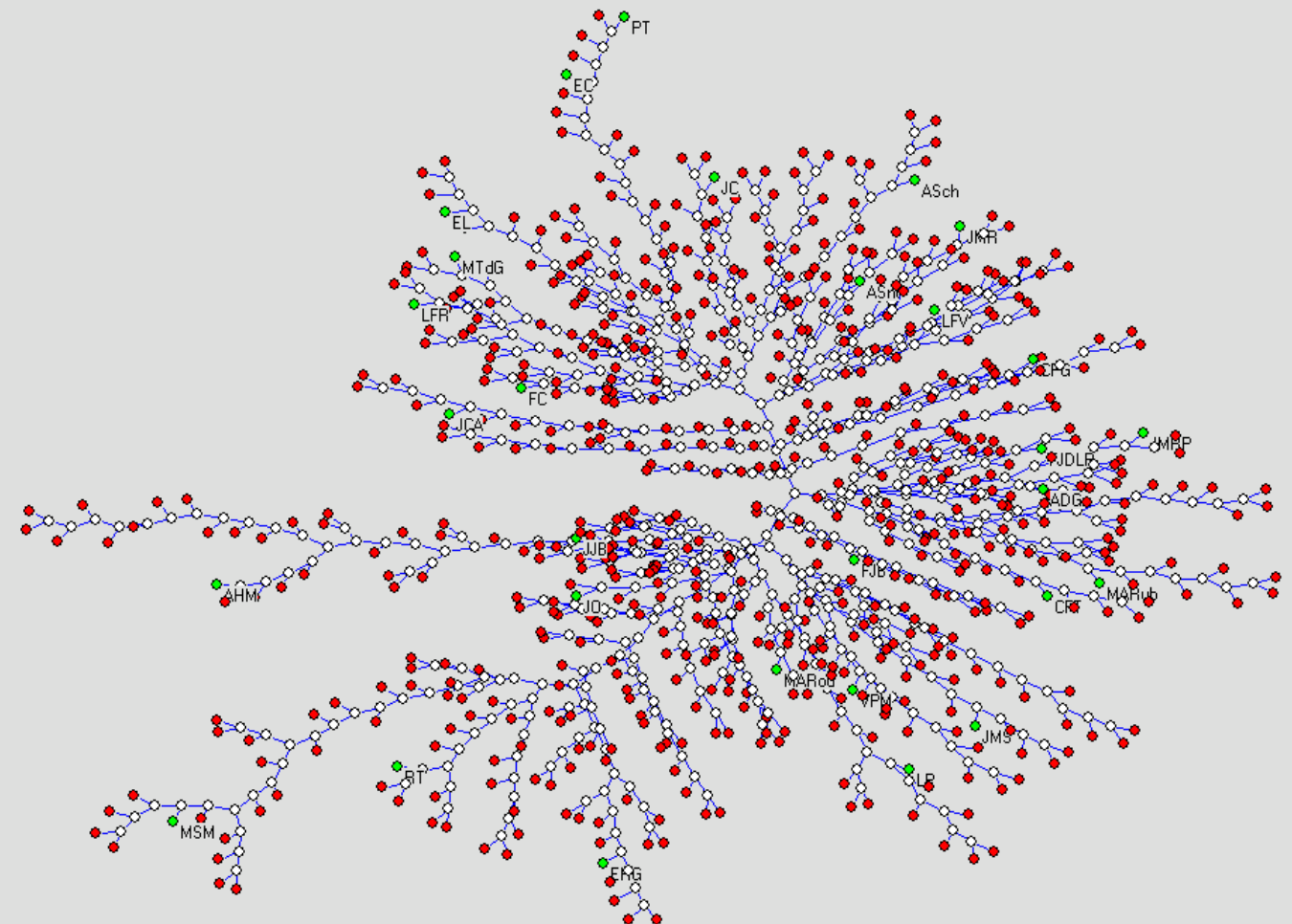
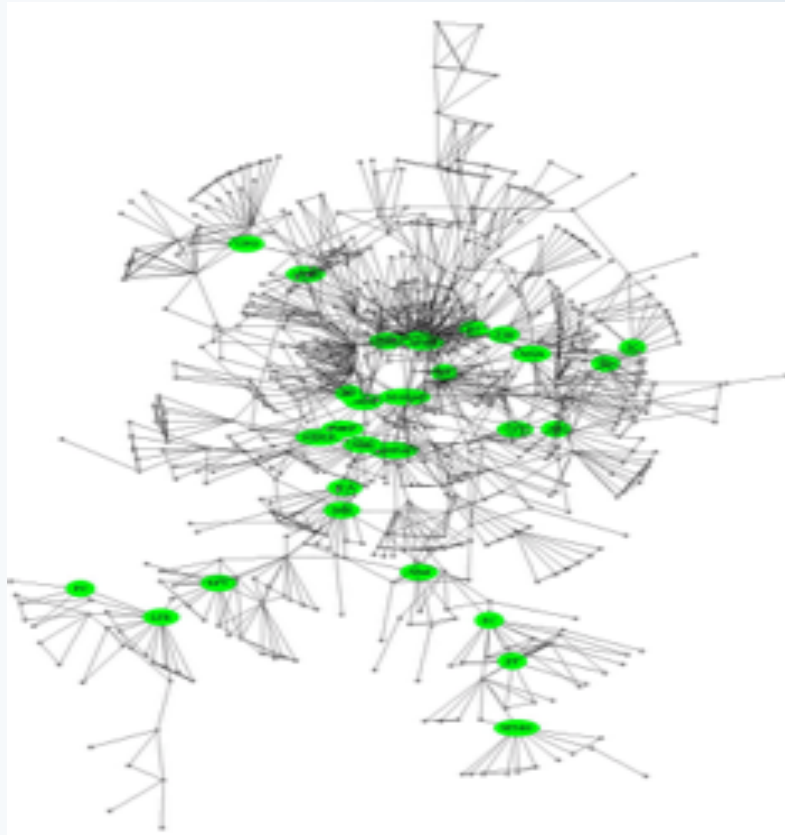
Comunitats

- Grups que tenen més enllaços entre ells que amb la resta (en promig)

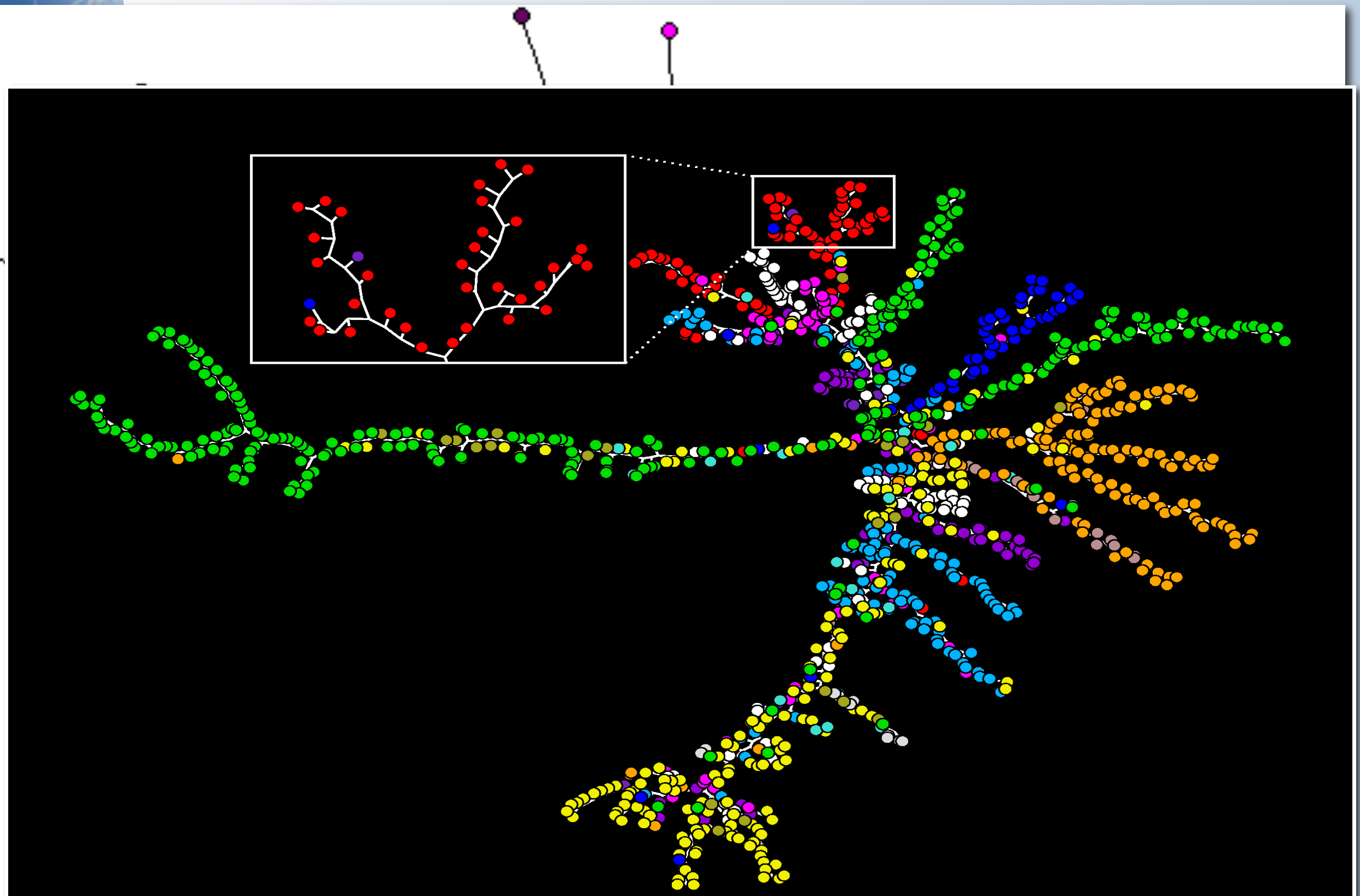


Comunitats: FisEs

- Regularidades a todas las escalas



Email network



Detecció de comunitats

- Community detection: modularity optimization

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j).$$

From data to models

- Models estadístics
- Models físics

Models de xarxes

- Erdős-Reny: Random graphs
- Watts-Strogatz: small worlds
- Barabasi-Albert: scale free networks

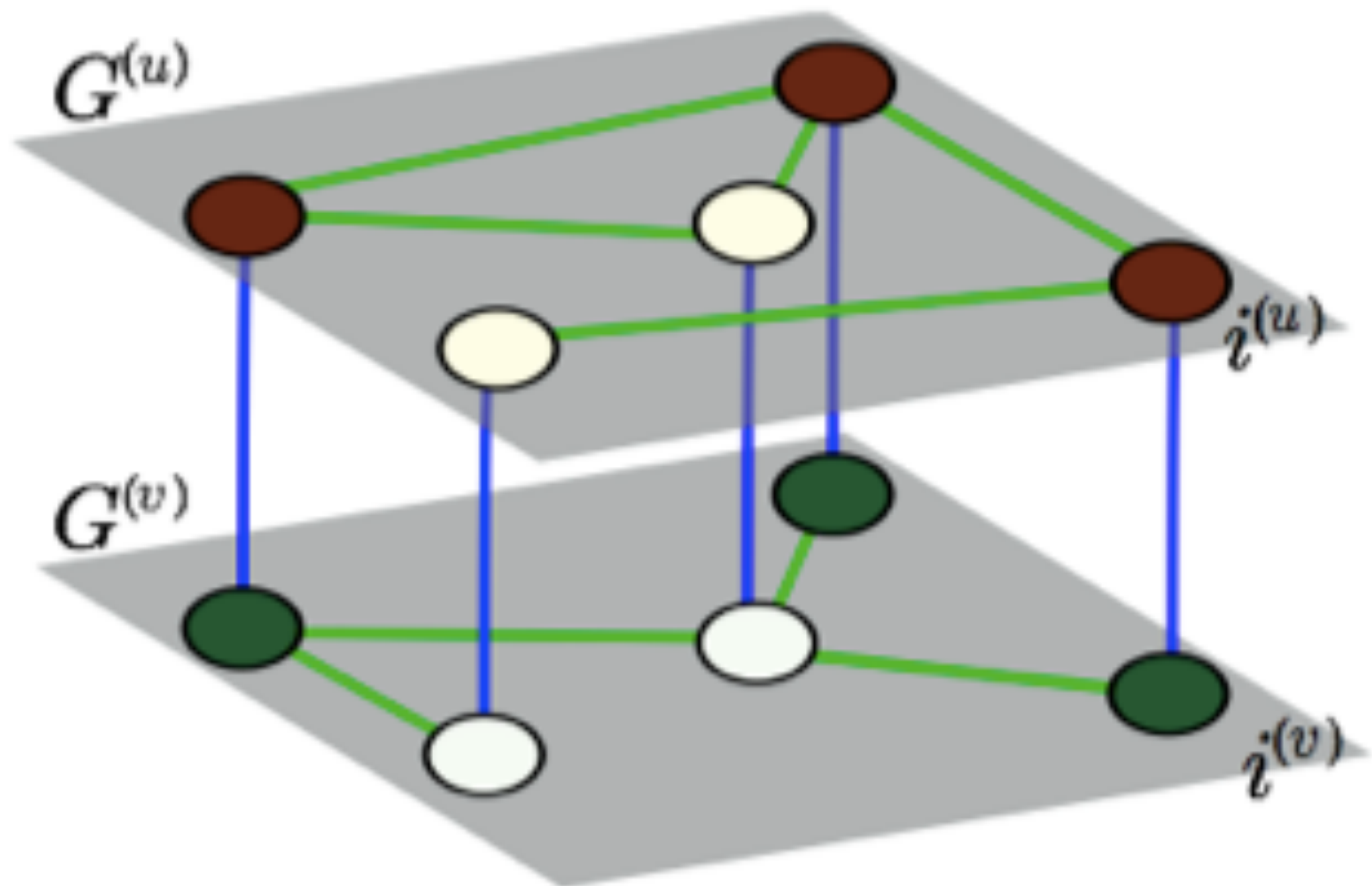
Universalitat

- Què tenen en comú xarxes en “mons” tant diferents
- Minimitzar les variables rellevants
- Proposem mecanismes
- No importen casos particulars

Big data => more complex networks

- Nous paradigmes:
 - xarxes de xarxes
 - xarxes que depenen del temps
 - xarxes interconnectades
 - xarxes en múltiples capes (multiplex)

Multiplex



Big networks



La plataforma N°1 del mundo en relaciones sociales

Hootsuite te ayuda a gestionar, analizar y entender las redes sociales para los negocios

- ✓ Gestiona hasta 100 perfiles sociales
- ✓ Programa cientos de publicaciones por anticipado
- ✓ Usa datos reales para optimizar tus campañas sociales
- ✓ Colabora con tu equipo y ahorra tiempo

Comienza ya tu prueba gratuita

Sin riesgos ni obligaciones.



“Take home message”

- Big Data $\Leftrightarrow$ Big Networks
- Diferents passes en el procés de les dades
- Importància de l'Estadística



complexitat.CAT

XARXA CATALANA PER A L'ESTUDI DELS SISTEMES COMPLEXOS
CATALAN NETWORK FOR THE STUDY OF COMPLEX SYSTEMS

[Learn more](#) [Follow us on Twitter](#)



Workshop: Data Visualization from different universes. Some applications from the scientific, artistic and business world

Dins el cicle *Més enllà de la física: reptes multidisciplinaris*

complexitat.cat (complexitat.cat)

Location: Sala de Graus Eduard Fontseré. Facultat de Física. Martí i Franquès 1, 08028 Barcelona

Start time: May 11, 2015, 9:30 a.m.

Abstract

“Data is the new soil...”

... but it needs to be mapped!