

L'intelligenza artificiale per l'elaborazione dei dati nello scenario di Industria 4.0

Roberto Bellotti
Dipartimento Interateneo di Fisica "M. Merlin"
Istituto Nazionale di Fisica Nucleare

Indice

- ✓ *I Big Data come carburante per i sistemi di Intelligenza Artificiale (AI)*
- ✓ *AI & Reti Complesse*
 - ✓ *Propaganda politica*
 - ✓ *Studio delle malattie neurodegenerative con tecniche di AI*
 - ✓ *Valutazione degli investimenti*
- ✓ **Conclusioni**

Intelligenza artificiale



Intelligenza Artificiale (IA) Disciplina che studia se e in che modo si possano riprodurre i processi mentali più complessi mediante l'uso di un computer. Tale ricerca si sviluppa secondo due percorsi complementari: da un lato l'i. artificiale cerca di avvicinare il funzionamento dei computer alle capacità dell'intelligenza umana, dall'altro usa le simulazioni informatiche per fare ipotesi sui meccanismi utilizzati dalla mente umana (Treccani)

Chi svilupperà la migliore intelligenza artificiale, diventerà il padrone del mondo (Putin, 2017)



Numeri ed Esempi

1 bit = 0/1 → Una lettera = 1 byte (= 8 bit).



Un libro = una foto di buona qualità = circa 1 Megabyte.



1 Gibabyte = 1.000 libri



1 Terabyte = 1.000.000 di libri



Facebook:

500 Terabyte di dati al giorno, tra cui circa 3 miliardi di “like” e 300 milioni di foto.

Stima dei dati posseduti da FB: 100.000 Terabyte.

Google e Amazon → oltre un milione di Terabyte.



Walmart registra più di 1 milione di “operazioni” all’ora!

Un Boeing 737 genera, in un viaggio attraverso gli Stati Uniti circa 240 Terabytes di dati.



Cosa sono i Big Data?

Le tre V:

- ✓ **Volume**
- ✓ **Varietà**
- ✓ **Velocità**

✓ Big Data:

- ✓ Introdotto nel 2013 nell'Oxford English Dictionary
- ✓ Introdotto nel 2014 Merriam-Webster's Collegiate

✓ Historically, most decisions — **political**, military, business, and personal — have been made by brains [that] have unpredictable logic and operate on subjective experiential evidence. “Big data” represents a **cultural shift** in which more and more decisions are made by algorithms with transparent logic, operating on documented immutable evidence. I think “big” refers more to the **pervasive nature of this change** than to any particular amount of data.

[datascience.berkeley.edu/what-is-big-data/]

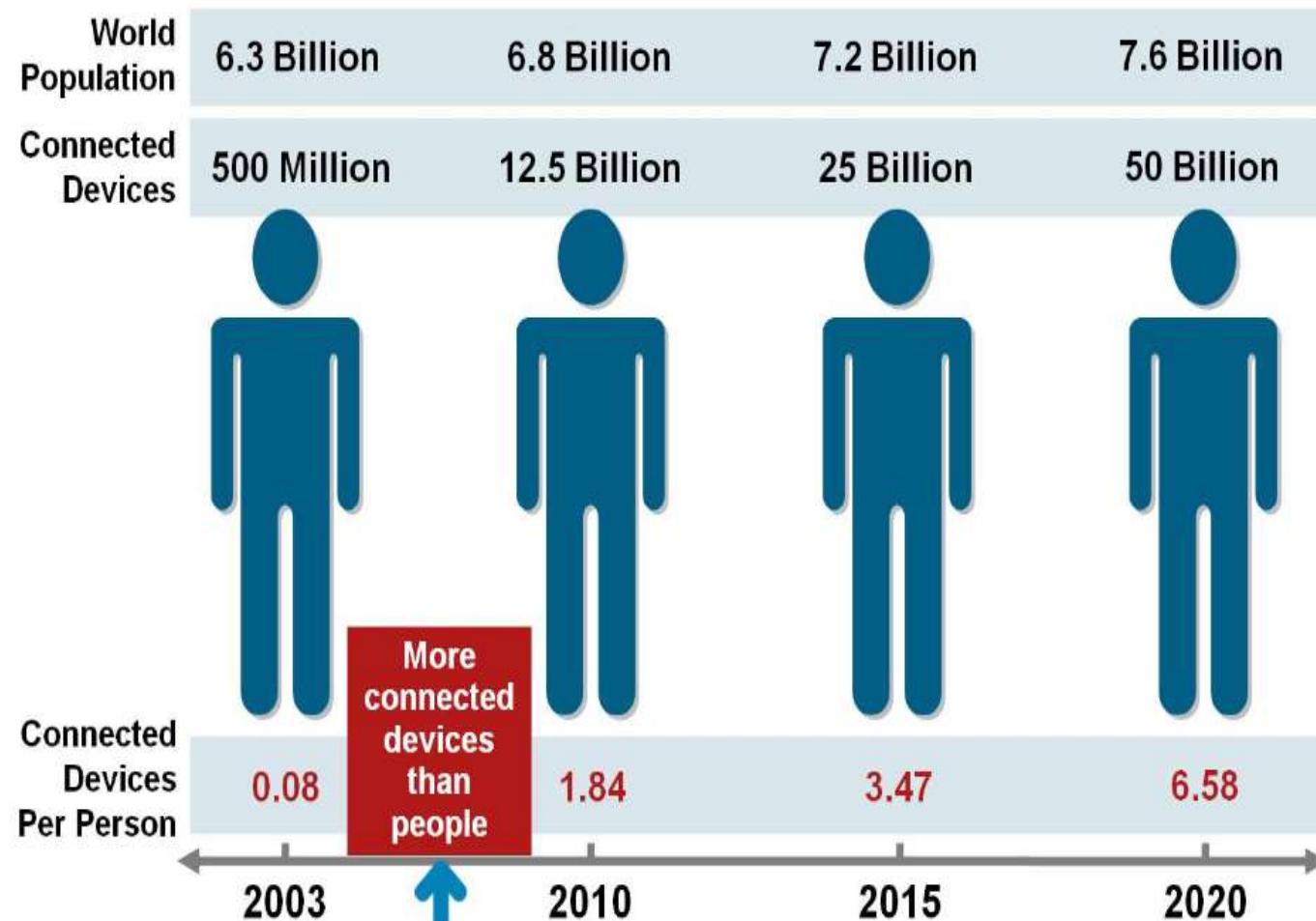
✓ Glossario Gartner:
“Big data is high-**Volume**, high-**Velocity** and/or high-**Variety** information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making and process automation”.

Popolazione vs Dispositivi

Quanti dispositivi
“connessi” possiede
ognuno di voi?

Il McKinsey Global Institute stima una crescita del volume dei dati prodotti pari al 40% per anno e un fattore moltiplicativo di 44 nel periodo 2009-2020.

Figure 1. The Internet of Things Was “Born” Between 2008 and 2009



Source: Cisco IBSG, April 2011

N.B. $6.58/0.08 > 8000$ (in 17 anni!) 6

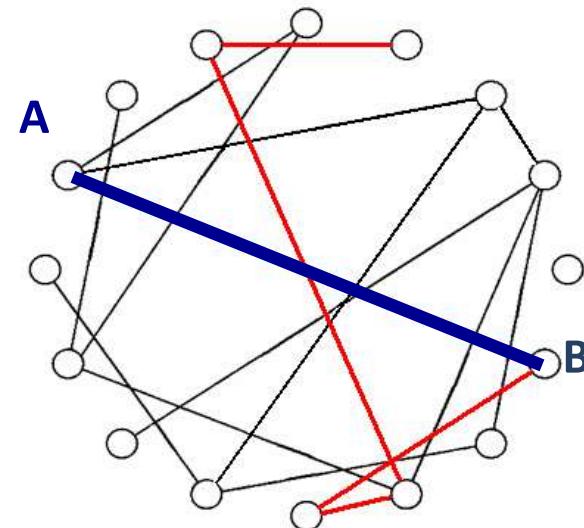


WAL-MART
ALWAYS LOW PRICES.
Always.



Big data tra loro connessi → reti complesse

- **Trasporti**
- **Reti elettriche**
- **Mercati finanziari**
- **Sistemi biologici**



8

e.g. La nazione **A** è collegata alla nazione **B** se **A** compra/vende un prodotto da **B**.

Cosa serve per vincere le elezioni?

Messaggio su LinkedIn del Data Scientist Rayid Ghani:

“Siamo in cerca di esperti **data scientist** che vogliono fare la differenza.

La campagna di Obama vedrà un ampliamento del team “analitico” per risolvere problemi di data mining su vasta scala e a forte impatto.

Si presentano diverse opportunità di inserimento professionale a tutti i livelli di esperienza.

Cerchiamo esperti di statistica, di apprendimento automatico, di text analytics e di analisi predittiva per lavorare su grandi volumi di dati e **contribuire ad orientare la strategia elettorale**”

Joe McGinnis, **Come si vende un Presidente**, Mondadori, Milano, 1970
(sulla Elezione di Nixon)

61-million-experiment in Social Influence and Political Mobilization - Nature, 2012

E' possibile aumentare l'affluenza alle urne attraverso gli *online social network*?

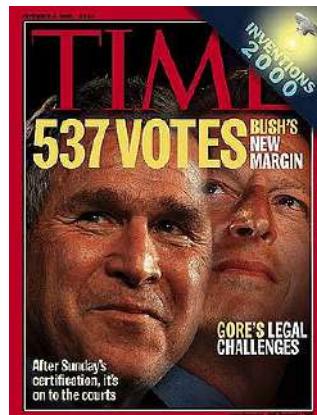
Possono, gli *online social network*, generare un "contagio sociale"?

Se si, quanto "vale" questo effetto?

Un effetto, anche piccolo, con i collegi uninominali può essere rilevante?



Nelle elezioni USA del 2000 George Bush ha battuto Al Gore in Florida per 537 voti (meno del 0.01% degli iscritti alle liste elettorali in Florida).



L'esperimento si è svolto il 2 Novembre 2010, giorno delle elezioni presidenziali USA, sugli utenti facebook che in quel giorno hanno utilizzato facebook.

L'esperimento è stato condotto a insaputa degli utenti → In Italia è stato condotto il 4 marzo 2018!

L'esperimento “megafono dell'elettore” di Facebook

a

Informational message



Social message



Tre gruppi disgiunti.

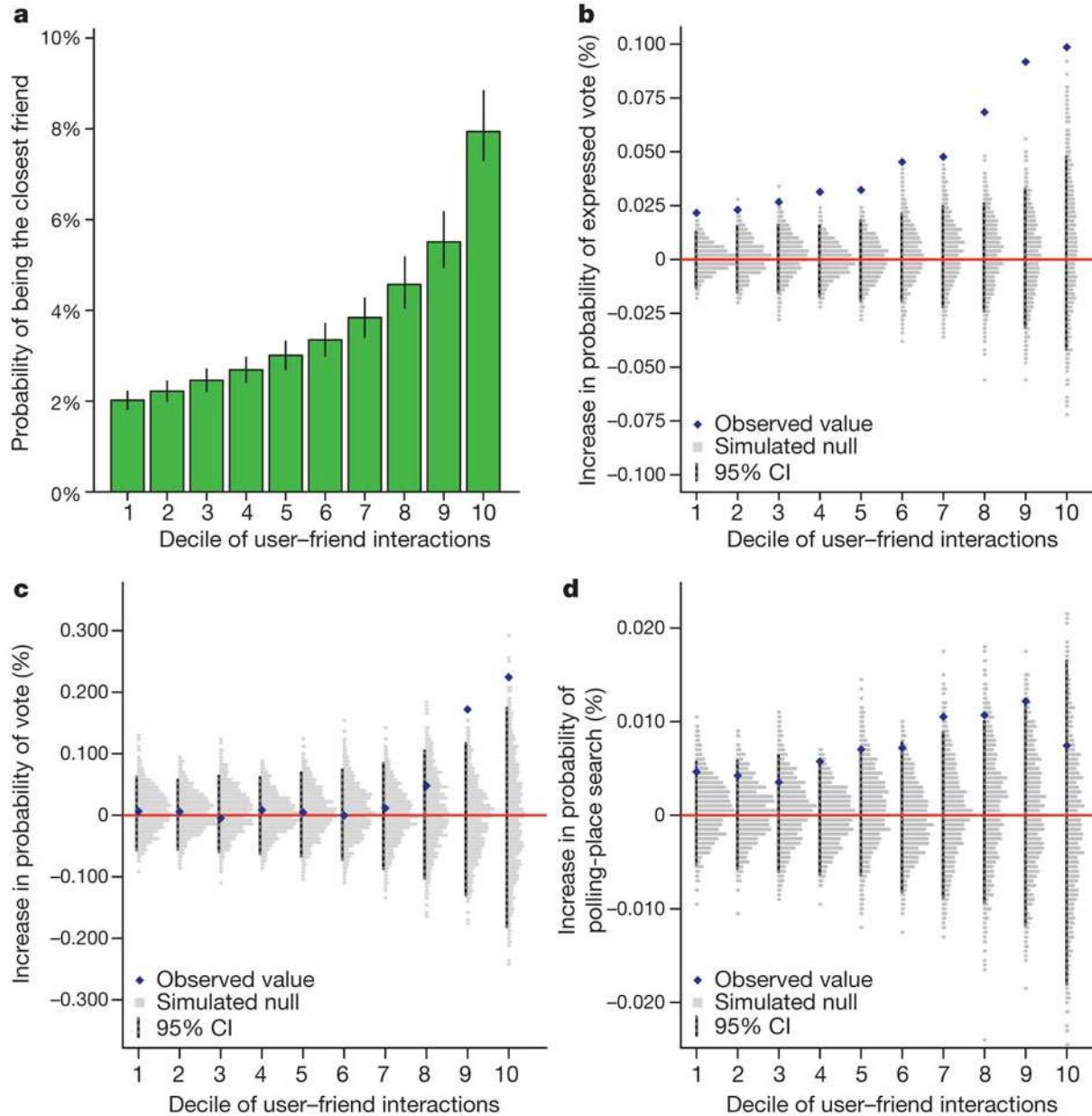
Gruppo di “controllo”
(=613,096)

Gruppo “informato”
(=611.044)

The social message group
(= 60.055.176)

I dati “social” sono stati incrociati con i dati reali per un sottocampione di circa 6 milioni di elettori.

Inoltre: gli amici non sono tutti uguali



Polarization, Partisanship and Junk News Consumption over Social Media in the US

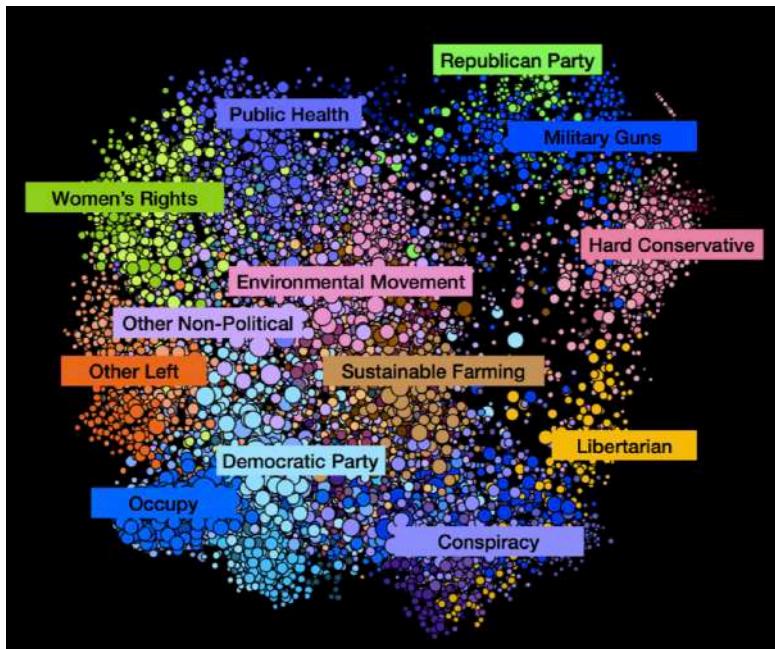
Studio effettuato da:

Università di Oxford,
Dipartimento “Oxford Internet Institute”
(oi.ox.ac.uk)

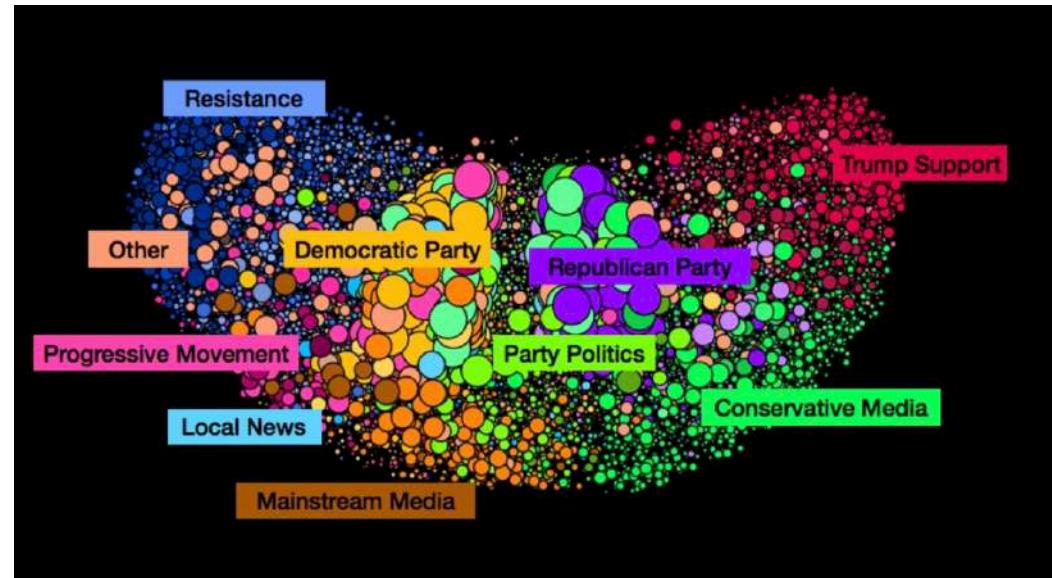
nell’ambito del progetto:

COMPUTATIONAL PROPAGANDA

Full Illustration of US Audience Groups on Facebook



Full Illustration of US Audience Groups on Twitter



Analisi di circa 22 milioni di Tweets raccolti nel periodo 1-11 novembre 2016 (le elezioni si sono svolte l'8 novembre 2016) da siti di “propaganda”

<https://www.eticaeconomia.it/propaganda-e-manipolazione-nelle-elezioni-politiche-il-ruolo-dei-social-network-e-degli-algoritmi-basati-sulla-intelligenza-artificiale/>



Istituto Nazionale di Fisica Nucleare

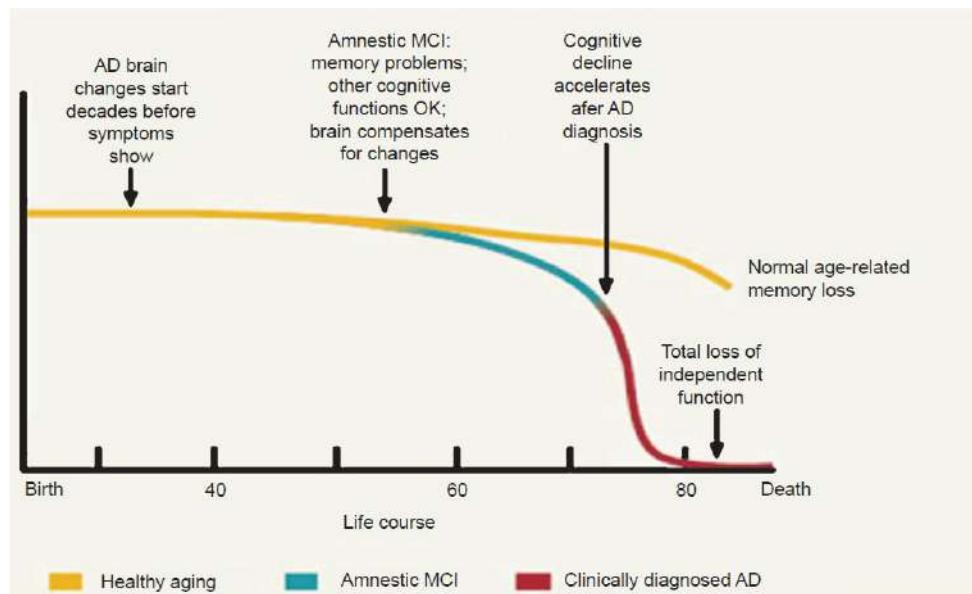


Studio delle malattie neurodegenerative con tecniche di AI

Mild Cognitive Impairment (MCI): a potential precursor to Alzheimer's Disease

MCI is a condition in which an individual has mild but measurable changes in thinking abilities that are noticeable to the person affected and to family members and friends, but do not affect the individual's ability to carry out everyday activities.

People with MCI, especially MCI involving memory problems, are more likely to develop Alzheimer's or other dementias than people without MCI.

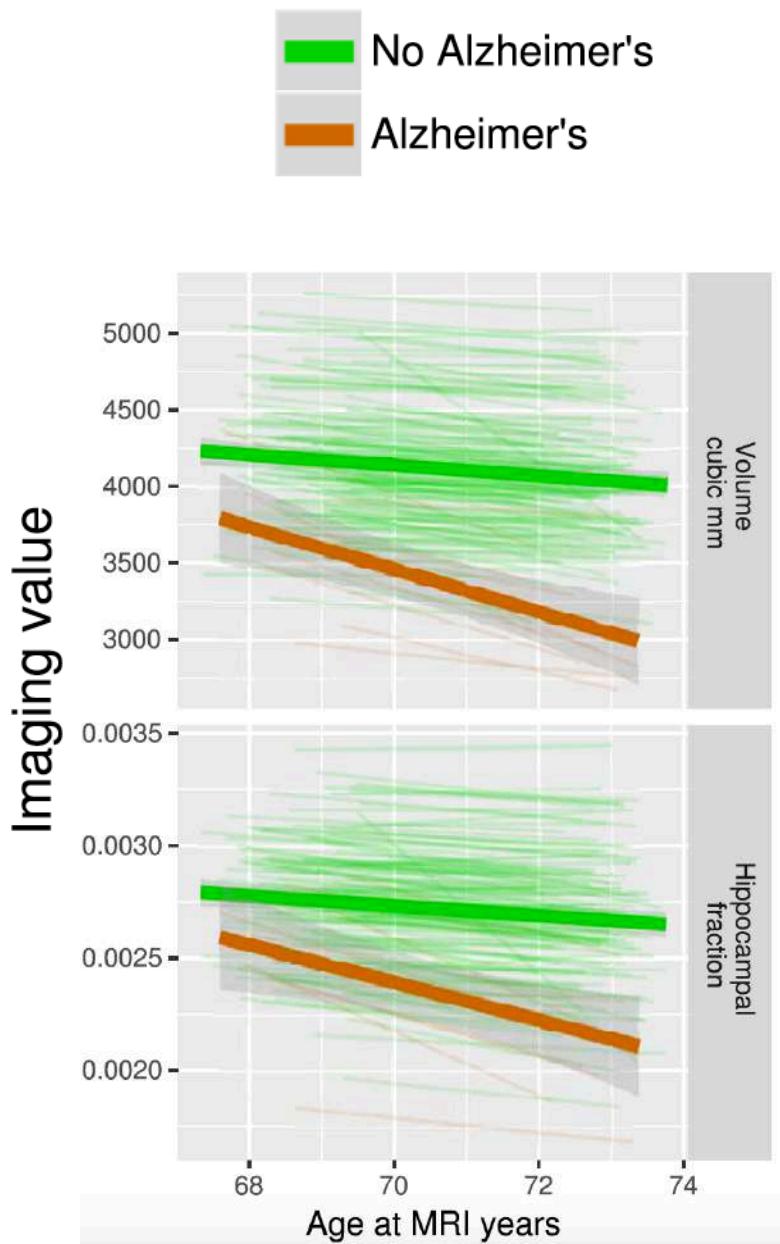


An average of 32 percent of individuals with MCI developed Alzheimer's dementia in 5 years.

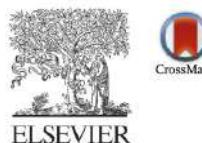
Identifying which individuals with MCI are more likely to develop Alzheimer's or other dementias is a major goal of current research.

MCI can develop for reasons other than Alzheimer's, and MCI does not always lead to dementia.

Hippocampal volumetry as a Biomarker for AD



Automated volumetry measuring hippocampal size at age 69 years and subsequent rate of change predicts Alzheimer's dementia development



Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 6 (2017) 31-39

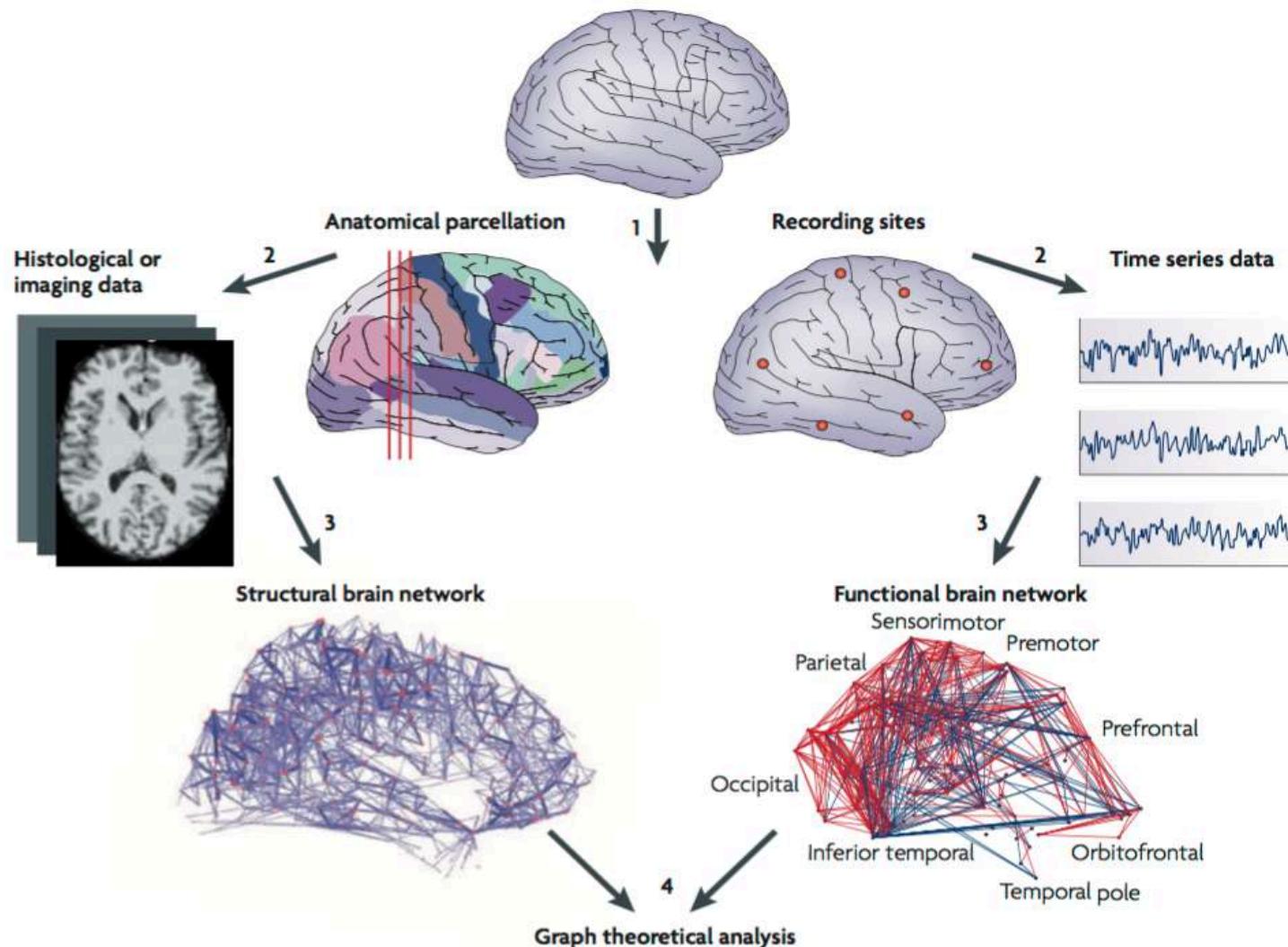
**Alzheimer's
&
Dementia**

A comparison of measurement methods of hippocampal atrophy rate for predicting Alzheimer's dementia in the Aberdeen Birth Cohort of 1936

Arnab K. Rana^{a,*}, Anca-Larisa Sandu^a, Kenna L. Robertson^a, Christopher J. McNeil^a, Lawrence J. Whalley^b, Roger T. Staff^c, Alison D. Murray^a

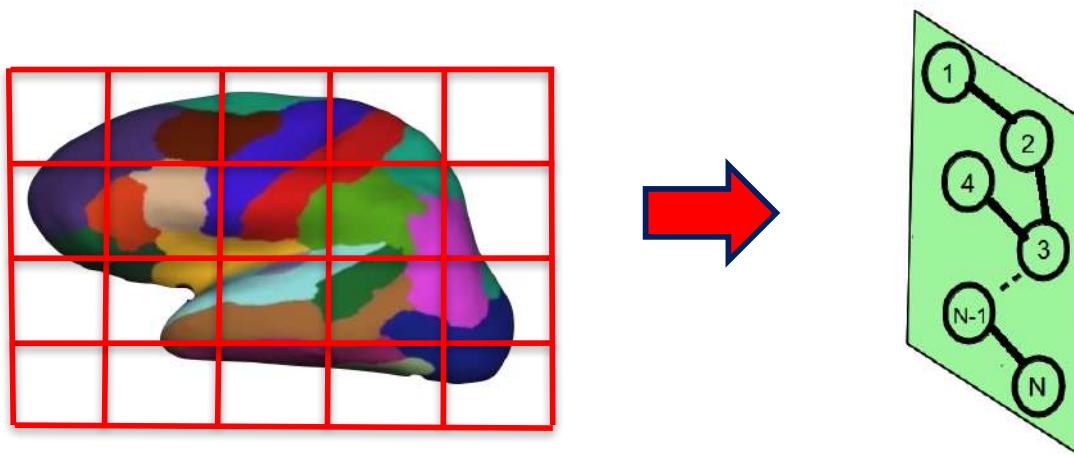
Hippocampal size over time. Each thin line represents one of the 149 participants. Participants who developed AD are marked with red lines and the other participants are marked with green lines.

Structural and Functional Networks

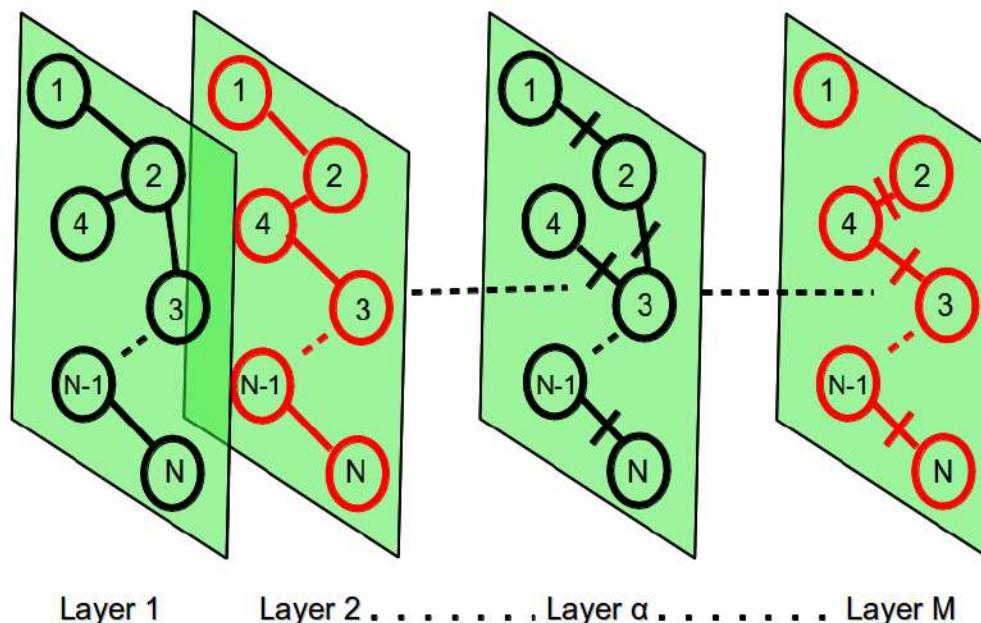


Bullmore and Sporns, Nature, 2009

A novel connectivity model



For each image a weighted graph was built upon a similarity measurement given by pairwise **Pearson's correlation** among the nodes represented by the patches of each subject.



A multilayer network $G=(G_1, G_2, \dots, G_\alpha, \dots, G_M)$ is a set of M graphs $G_\alpha=(N_\alpha, E_\alpha)$ with $\alpha=(1, \dots, M)$, each of one representing a **layer**. If the set of nodes N_α is fixed then we call it by definition a **multiplex**.

Independent test accuracy

Classification Accuracy	
Control - Alzheimer	Control - MCI (converter)
0.86 ± 0.01	0.84 ± 0.01

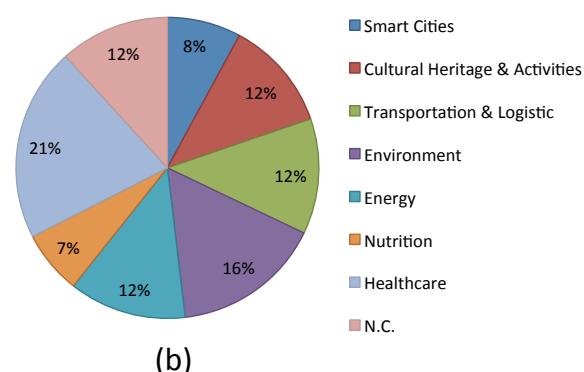
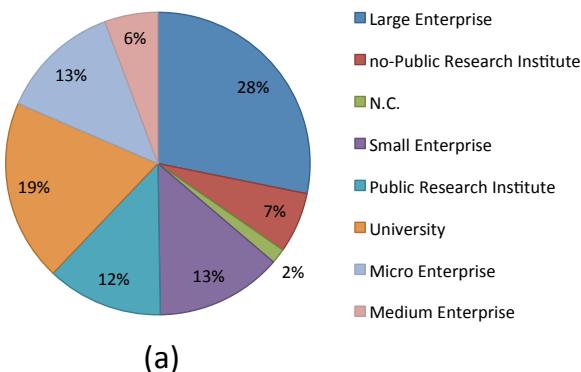
The proposed methodology is intrinsically data-driven and, as a consequence, it could suffer from typical **over-training** issues which on turn could mine the reliability of the findings. As a further assessment we performed a binary classification (NC-AD and NC-cMCI) on the independent test set.

The Italian Public Funding Program (2007-2013)

- ✓ Italian Program for the Convergence objective regions (less developed regions in Southern Italy)
- ✓ (Program: National Operative Program (PON) for Project in research & development)
- ✓ Goal: Evaluate the impact of public funding at regional level

Total cost of the PON Projects → 2500 Million of Euros

About 300 Different R&D Projects 769 distinct partners



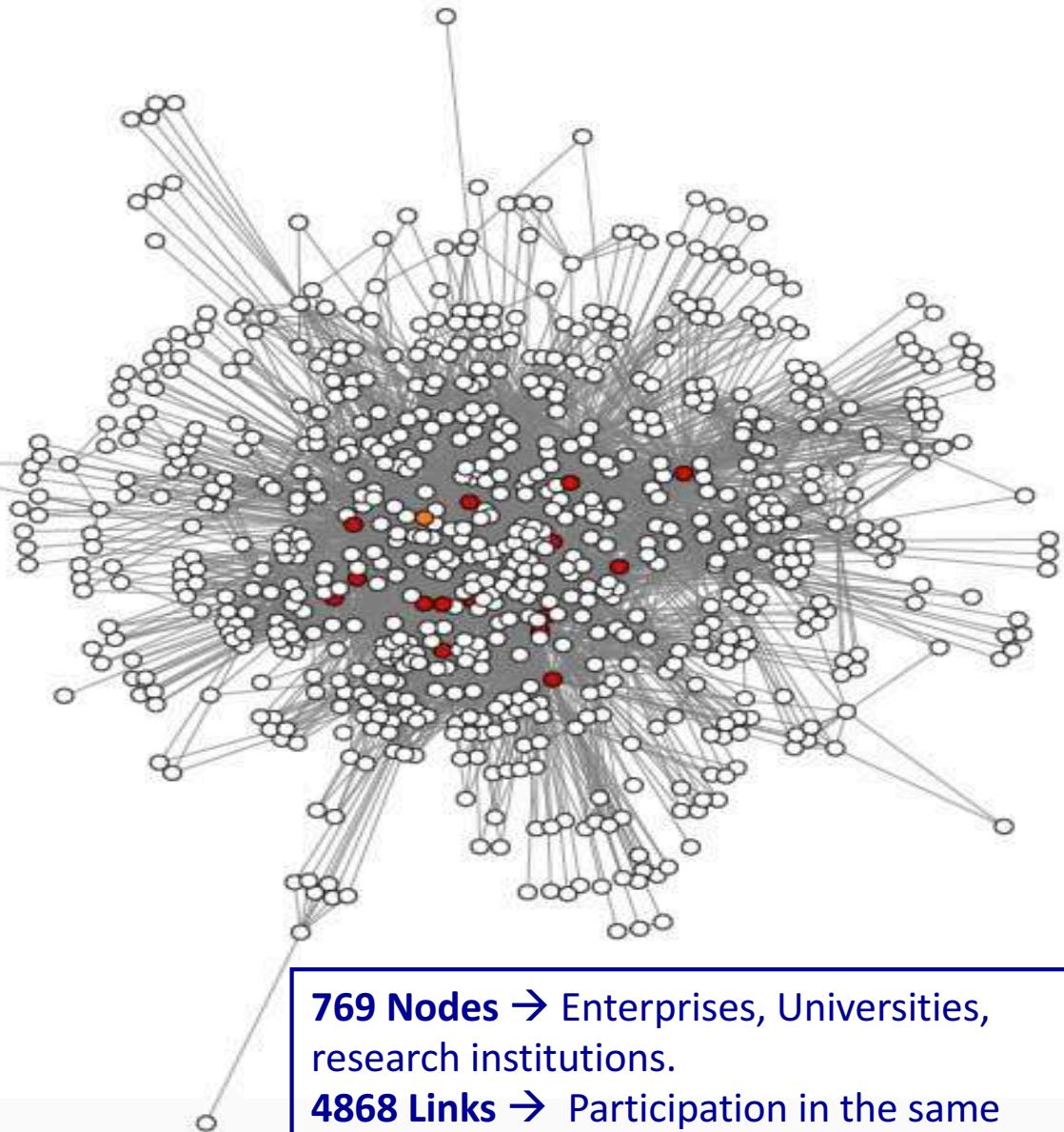
- Available information: Calls and funding measures, projects, proponents and participants, funding, geographical information, etc.
- Data format: **open data** (xls, XLM, CSV)
- Source: : <http://www.dati.puglia.it>, <http://opencoesione.gov.it>

2007-2013 Italian Public Funding Program: from dataset to data models.

Projects → 10104 entries with 52 attributes describing project information about program references, activities, textual description of project scope and objectives, detail about partners and so on.

Locations → 11390 entries with 8 attributes describing details about geographical localization of project partners.

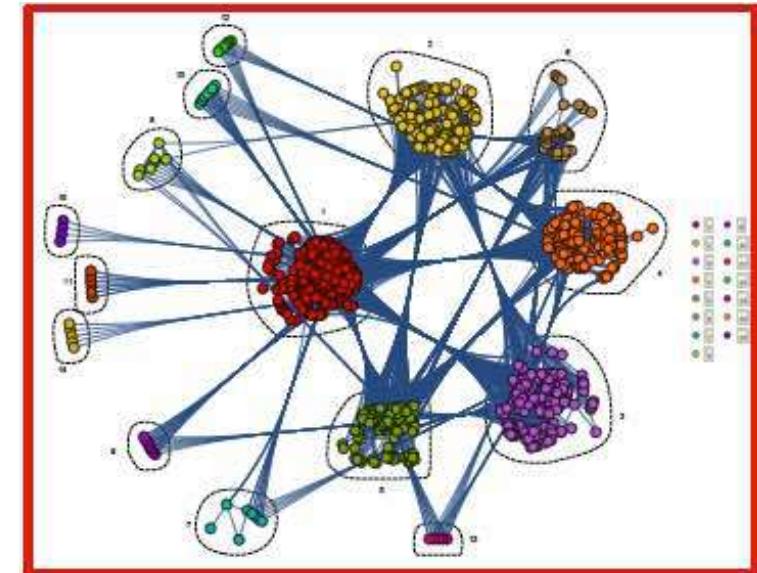
Budgets → 5670 entries with 13 attributes describing details about amount and state of project funding.



Result #1: community detection

We found 15 main Communities →

- provides a deep understanding of how the fund allocation criteria are able to influence the economic development of a Region;
- discovering the existence of groups within a certain network of relationships;
- highlighting such groups can be very important for the analysis of a productive system;
- The PON R&D network shows strongly heterogeneous communities, with hugely populated groups and very small ones.
- when communities grow in size, they tend to include important nodes. For example, the largest community includes the National Research Council (CNR, next slide)

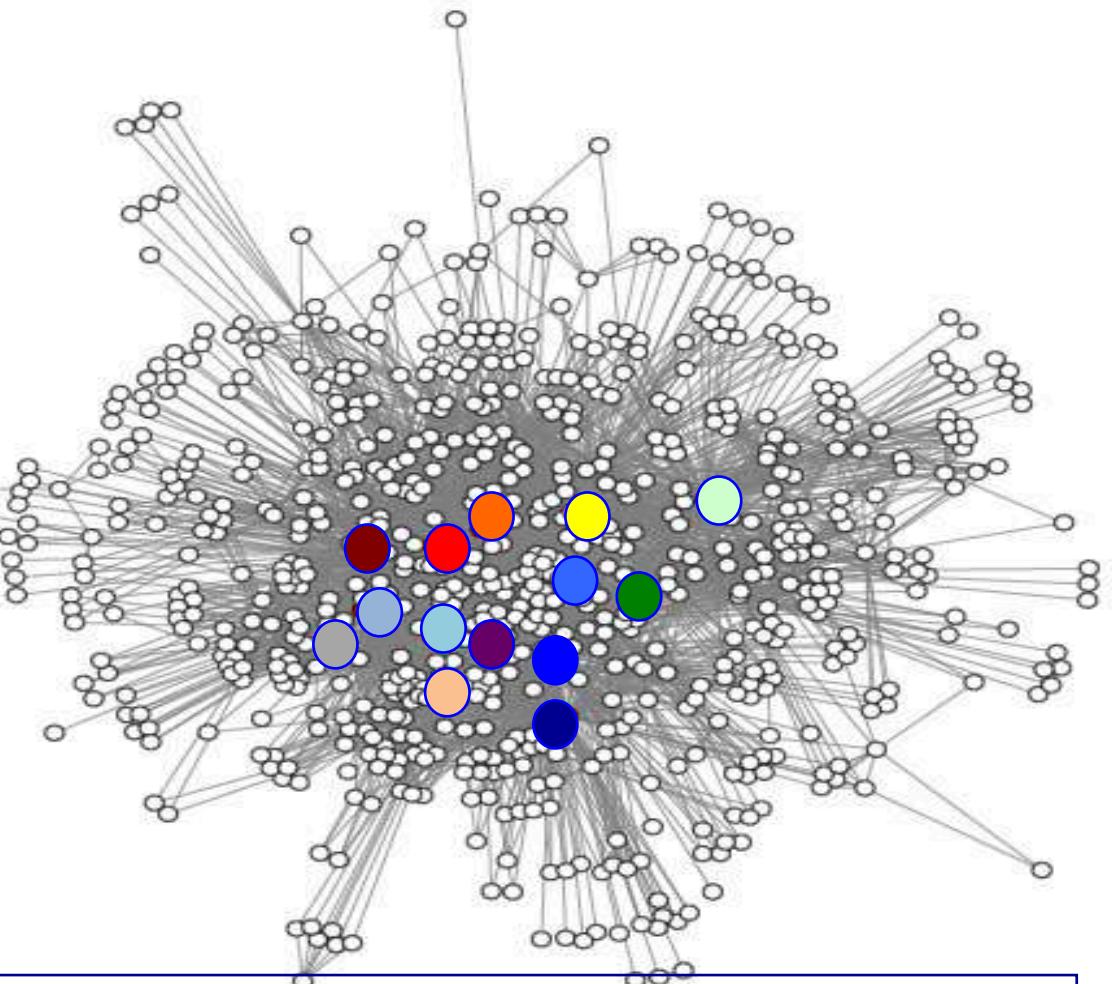


The community structure of the (giant component of the) PON R&D network. 15 communities are highlighted, found with the Newman-Girvan algorithm.

Result #2: it is a network with Hubs

Scale free network →

- Inhomogeneous degree distribution, with many nodes having more connections than the average (hubs)
- Resistance to “random failures”, indeed the removal of a random node would not systematically affect the main hubs
- Policymakers are interested in generating a solid network of relationships between productive actors on the territory



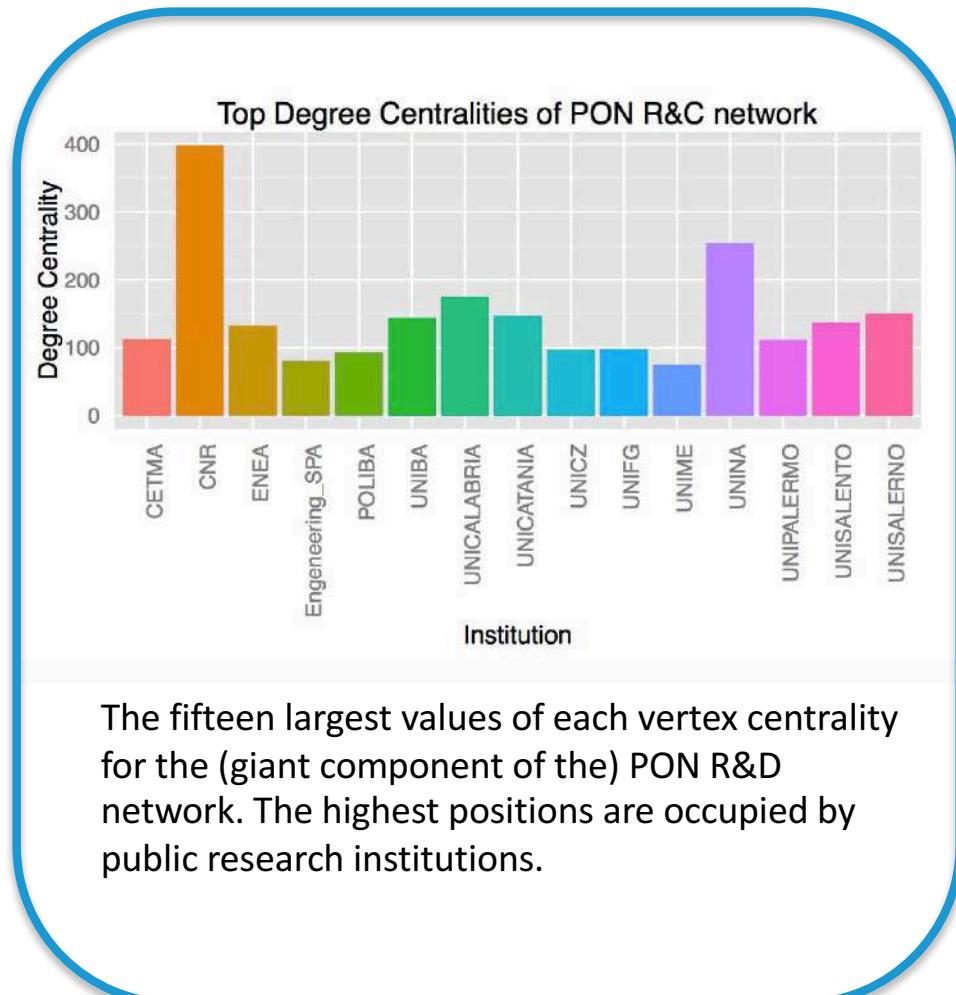
Strong indication that the network of funded project gravitates around large poles involving research centers

Result #3: who are the hubs?

Centrality of nodes →

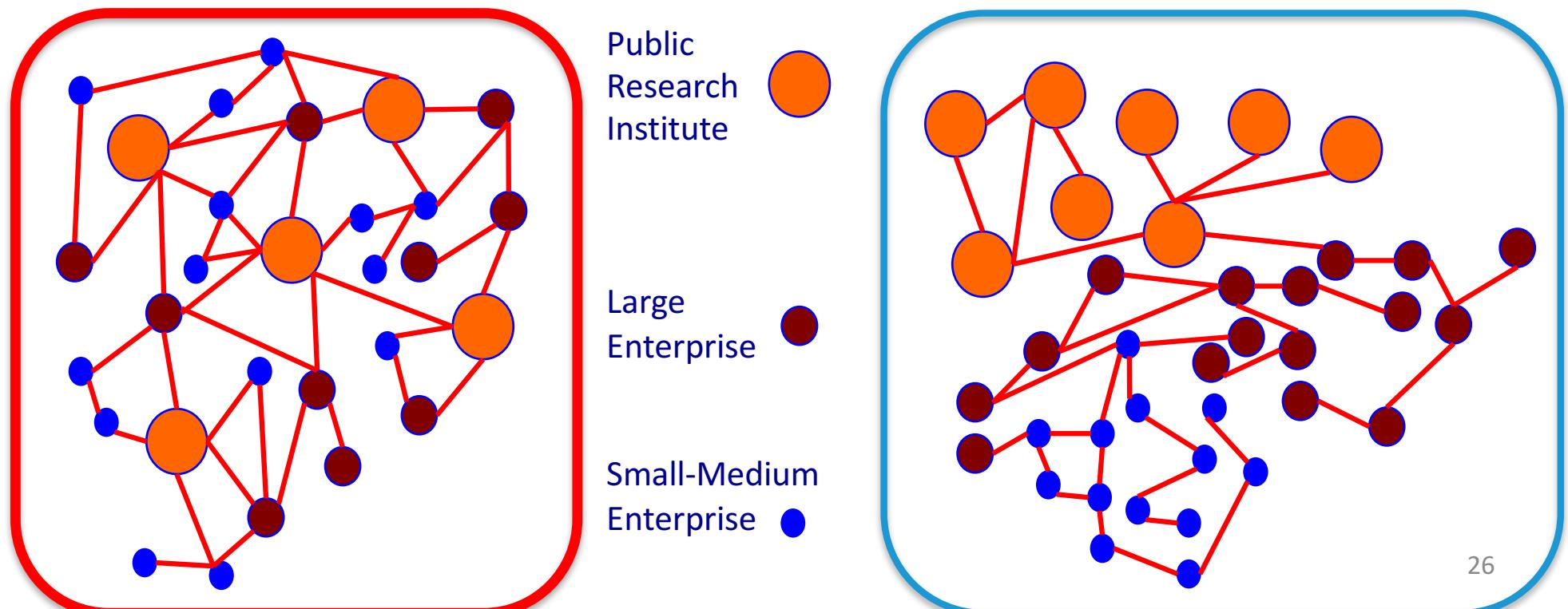
identifies the most important nodes within a network

- Dominant role of public research
- Universities and research centers play the role of the “glue” i.e. they are responsible of the connectedness of the network
- Ex-post indicator.



Result #4: the network is anti-assortative

- Low tendency to form “groups of interest” or “lobbies” among important actors.
- Hubs are strongly connected to smaller and less connected enterprises/institutions.
- It is an interesting result, since most social networks show assortative behavior.
- Anti-assortative networks are more sensitive to the removal of high-degree nodes, which is an indication for the policymaker of the importance that public research has in the productive system.





I sistemi informativi aziendali producono già molti dati.

I sistemi di business intelligence utilizzano al meglio i dati?

Quale valore aggiunto possono dare sistemi sofisticati di gestione e analisi dei dati?

“Fra il forte e il debole è la libertà che opprime e la legge che libera”

J. H. Lacordaire