CIPA Workshop 2019

Centro Carlo Azeglio Ciampi, December 17th, 2019

Knowledge Graphs e Ragionamento Automatico

applicazioni in ambito finanziario

Luigi Bellomarini

Divisione **Ricerca** sulle **Tecnologie Avanzate**
Dipartimento **Informatica**
A R T
www.bankit.art

BANCA D'ITALIA
EUROSISTEMA

Knowledge Graphs as | Large "World" KBs

Cyc [Lenat & Guha 1989]
comprehensive ontology and knowledge base of everyday common sense knowledge.

Freebase [Bollacker et al. 2007]
online collection of structured data harvested from many sources, including user-submitted wiki contributions.

Google KG [Singhal 2012] + **K.Vault** [Dong et al. 2014]
KB used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources.

Knowledge Graphs as Large "World" KBs

DBpedia [Auer et al. 2007] * **Yago** [Suchanek et al 2007] both generate structured ontologies from Wikipedia.

Wikidata [Vrandečić 2012, Krötzsch+V. 2014] open Knowledge Bases that can be read and edited by both humans and machines.

More Specialized Knowledge Graphs

Facebook KG Social graph with people, places, things

Amazon PG knowledge graph of all products

Factual Businesses & places

Wolfram KB World facts + mathematics

RIT People, skills, recruiting

Commercial Banks Customers, Companies, Risks, …

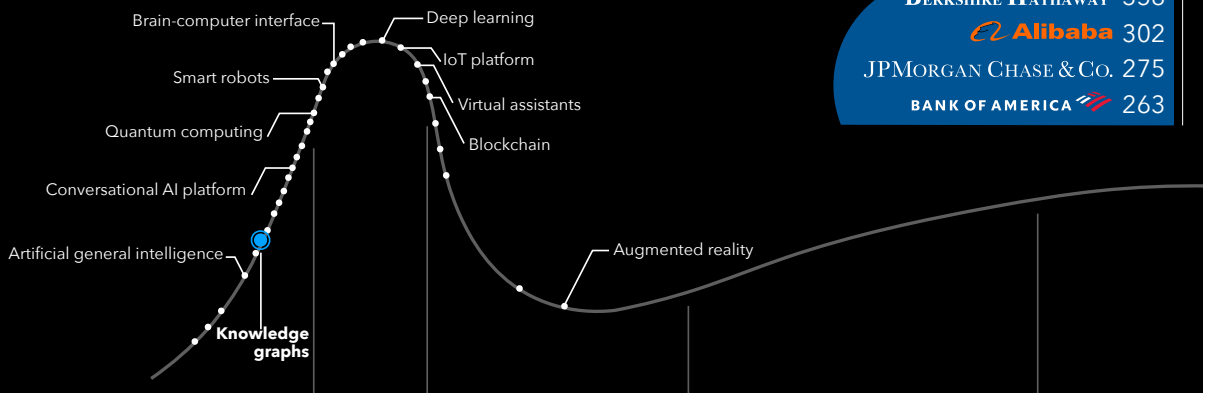
Rating Agencies Companies, Evaluations, Risks, …

Knowledge Graphs in the world: an *innovation trigger*

change in market cap 2009-2018 ($bn)

Apple	757
amazon	670
Google	609
Microsoft	540
Tencent 腾讯	483
facebook	383
Berkshire Hathaway	358
Alibaba	302
JPMorgan Chase & Co.	275
BANK OF AMERICA	263

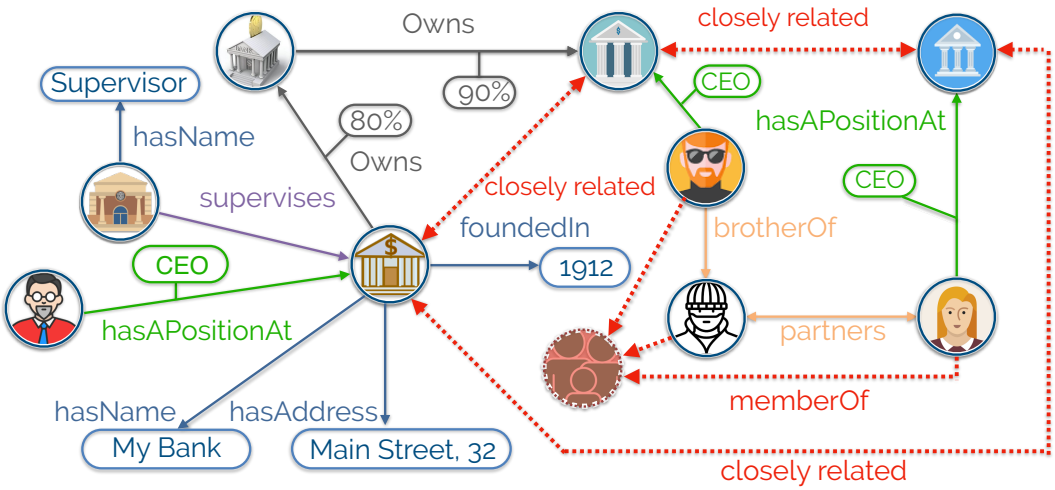
extensive adoption in the financial sector



Thousands of medium to large size companies now want their own **corporate knowledge graph**.

This not just for semantic indexing and search, but for **advanced reasoning tasks** on top of **machine learning**.

Knowledge Graph by Example



So what is a Knowledge Graph?

The Knowledge Graph is a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources.

W *Knowledge Graph*

"Knowledge graph" redirects here. For the Google knowledge base, see Knowledge Graph. For other uses, see Knowledge engine (disambiguation).

W *Ontology (information science)*

So what is a Knowledge Graph?

A possible (buzzwordy) definition

KG = Knowledge Base + Reasoning

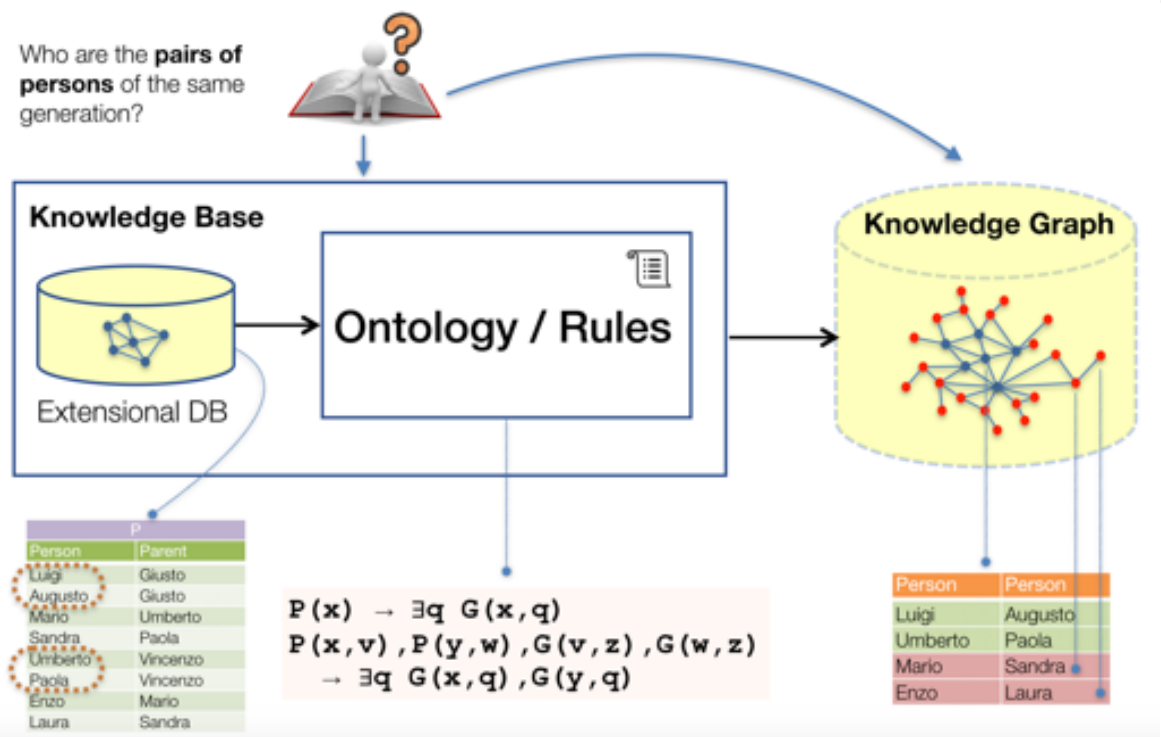
So, what is a Knowledge Base (KB)?

KB = DB + Ontology

So, what is an Ontology?

Formal representation of the **entities**, **naming**, **definition** of the properties, categories and relations between such entities, for one, many or all domains

Reasoning on KGs



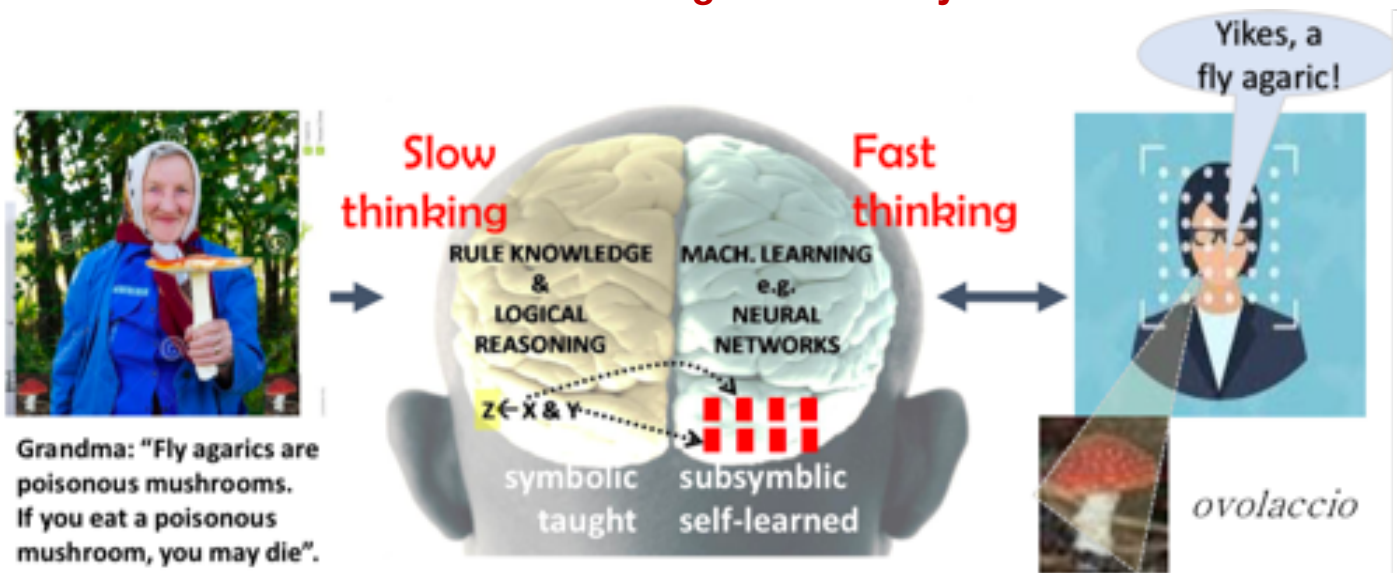
Traditional technologies (**graph databases, description logics**, **relational databases**) do not suffice.

Reasoning tasks are required that **cannot be expressed** by description logics, and cannot be reasonably managed by relational DBMS, nor by graph DBMS.

Knowledge Graph Management Systems

KGMS combine the power of rule-based reasoning with machine learning over Big Data:

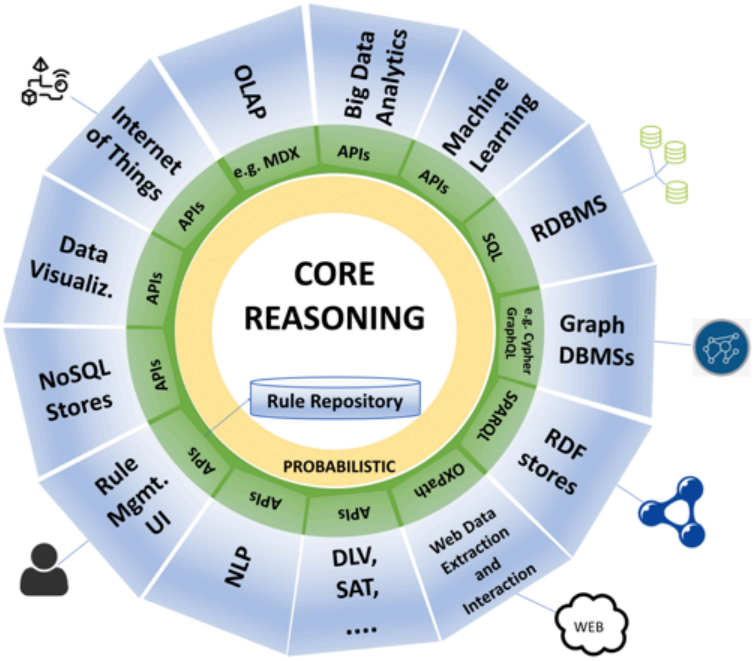
KGMS = KBMS + Big Data + Analytics

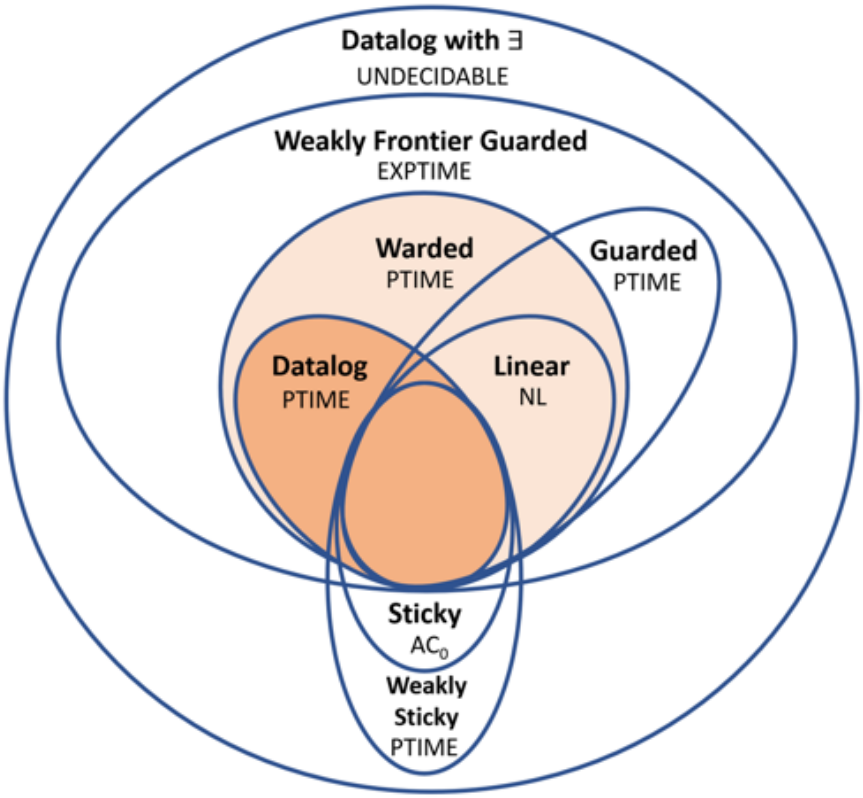


Vadalog KGMS

BANCA D'ITALIA
EUROSISTEMA

UNIVERSITY OF OXFORD





Scientific Innovation

Publishing at **top** AI and database **conferences** and **journals**:

It's not a fantasy!!

- **In many many sectors**: commercial banks (current projects), Big Social Networks (recent project), Health (current project), Energy & Utilities (current projects), Telco, Security Companies (current project), …
- **Fintech**: financial intelligence decisions, anti-money laundering, **company network analysis**, support for creditworthiness decisions, know your customer, customer profiling, data quality enhancement, intelligent business processes (e.g., intelligent BPM), statistical metadata dictionary, register data matching, micro-data data lakes (e.g., research data centers), data confidentiality, payment network analytics, financial shock propagation, P2P lending network analysis, market stability, risk assessment, regtech, suptech, …and many many more.

Many KGs in the financial realm!!

Company graphs

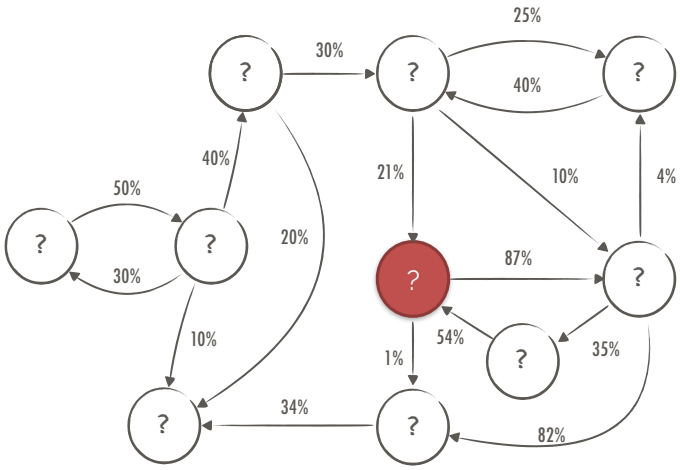
We build graphs of company networks, to:

1. reveal **power**
   1. finding **controllers**
   2. studying the **structure** of Italian market
   3. studying **dispersion** of control
   4. global **shareholding** analysis

2. detect **collusion** and do **forensics**
   1. support **AML**
   2. detecting **ultimate beneficial owners**

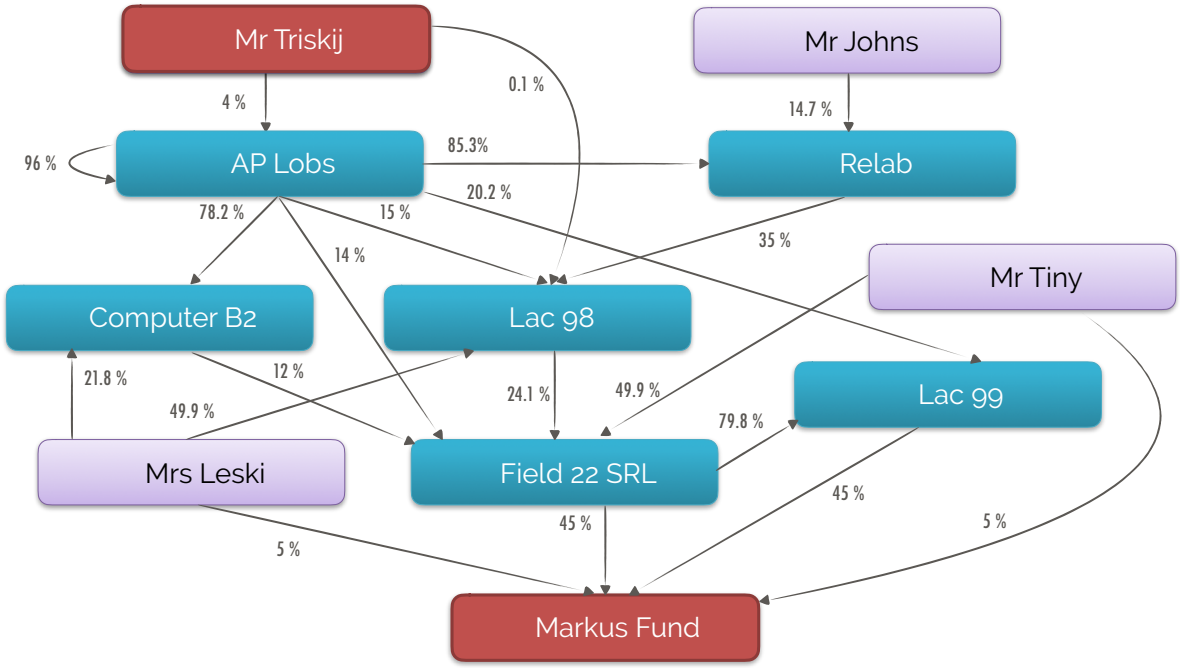
3. evaluate **risks**
4. model **propagations** (e.g., of shocks)
5. guarantee **compliance**
6. perform enhanced **due diligence**
7. understand complex **foreign shareholder structures**
8. know real **cash flows**

The setting



- Who takes **decisions**?
- Who's the ultimate **beneficial owner**?
- Is there **collusion**?
- How does **risk** propagate?
- What are the **real cash flows**?

The setting



Ownership and Control

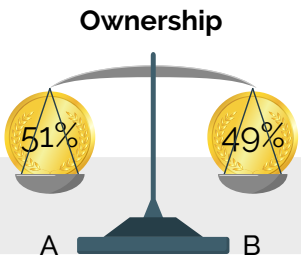
- **Integrated ownership** is about direct and indirect, owners of a company
  - it can be seen in terms of cash flow rights

- **Control** is about voting power
  - of any direct and indirect owner of a company

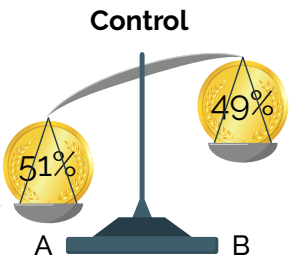
Ownership vs Control

Companies A and B have both **direct** (shares) and **indirect** ownership of company C.

A has 51% of C
B has 49% of C

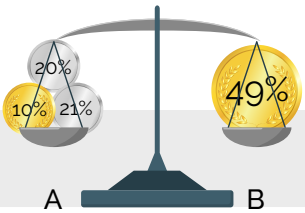


Ownership of C is comparable

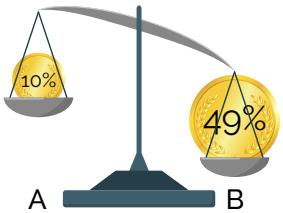


A has 100% decision power
A controls C

A has 10% and 41% of C
B has 49% of C



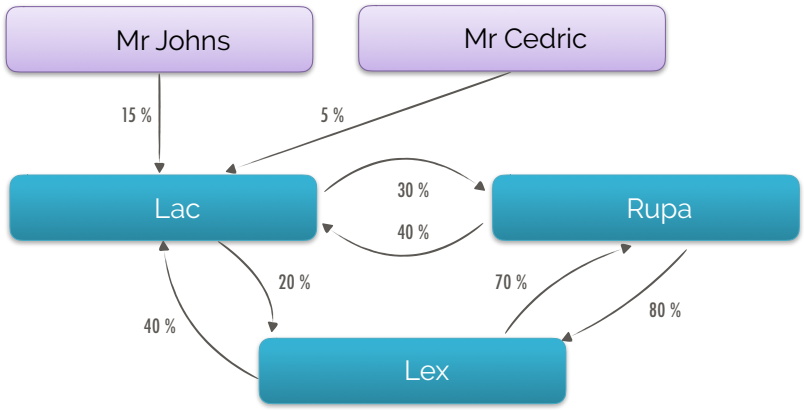
Ownership of C is comparable



A has 10% decision power
A does not control C

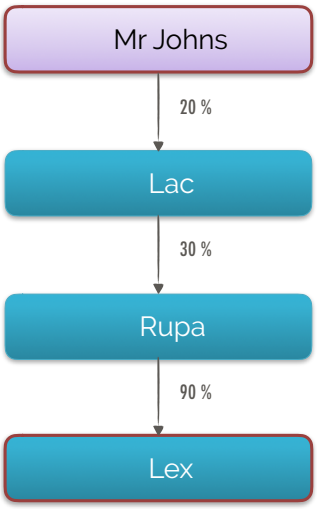
Integrated Ownership

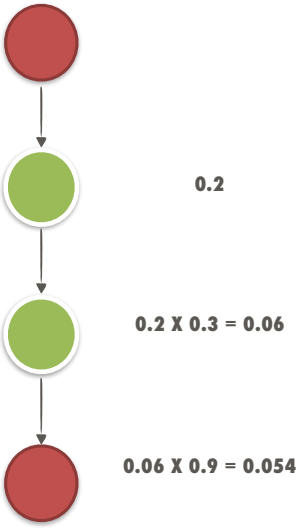
- Traditional systems only store the first-level (the closest) shareholders for a specific company



Integrated Ownership: the basic math

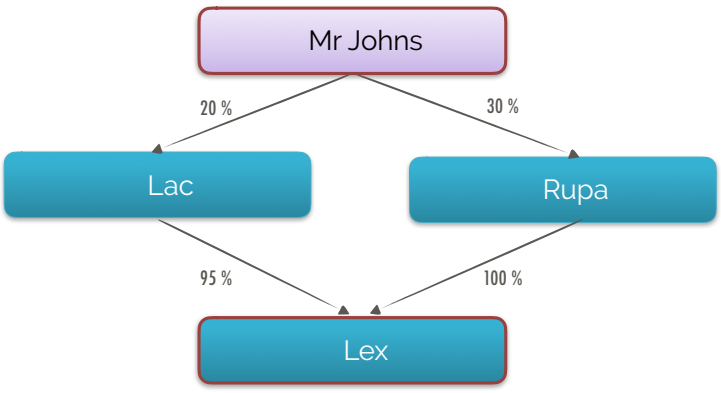
- Indirect ownership

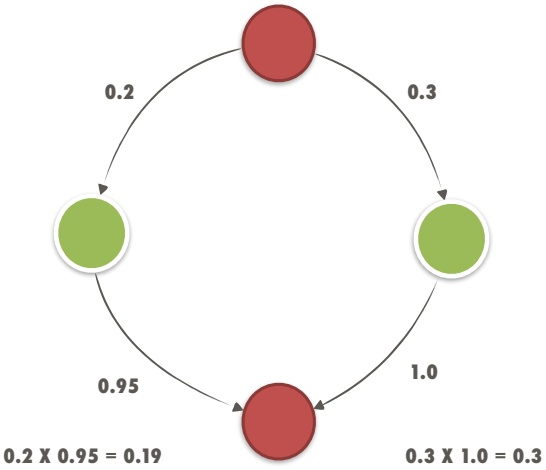




Integrated Ownership: the basic math

- Parallel ownership

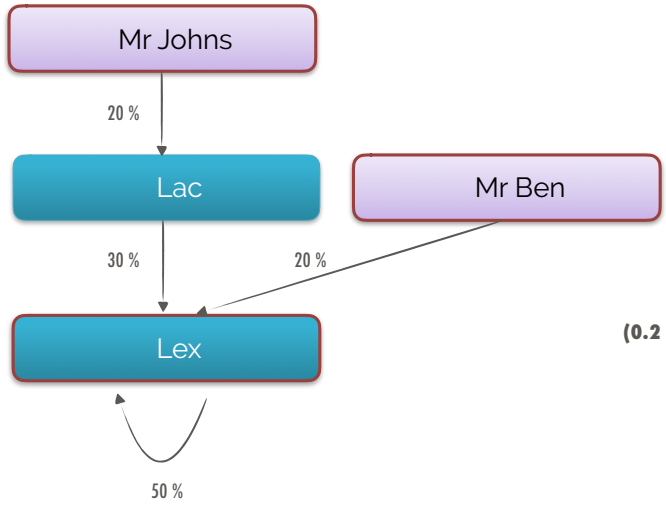


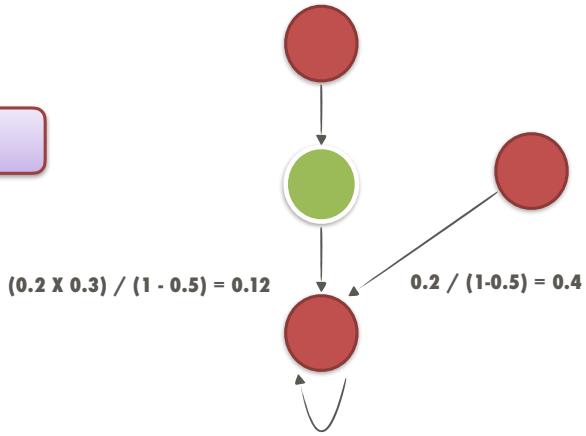


0.19 + 0.3 = 0.49

Integrated Ownership: the basic math

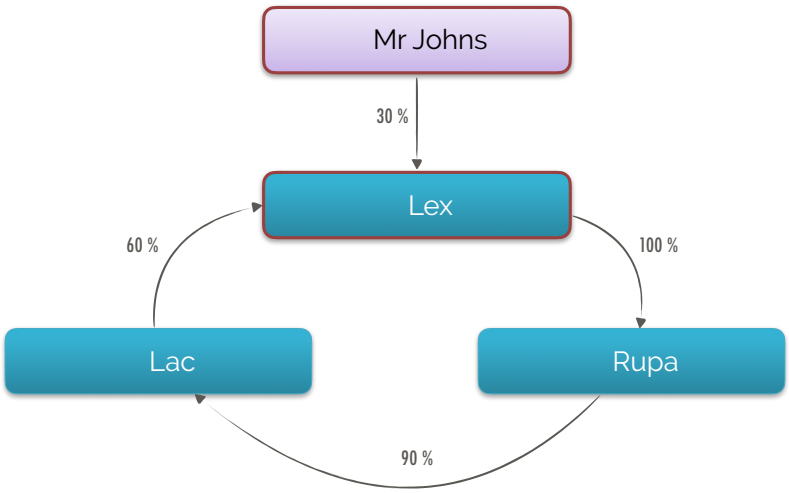
- Cycles (direct)

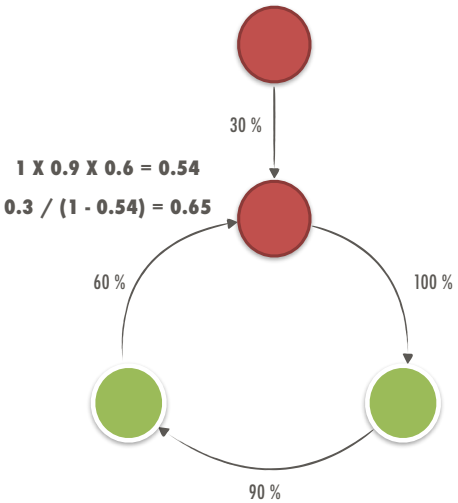




Integrated Ownership: the basic math

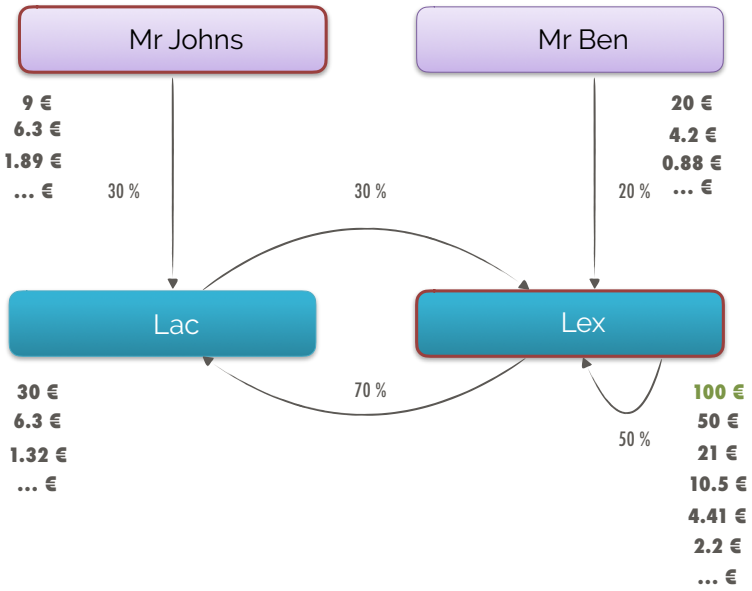
- Cycles (indirect)





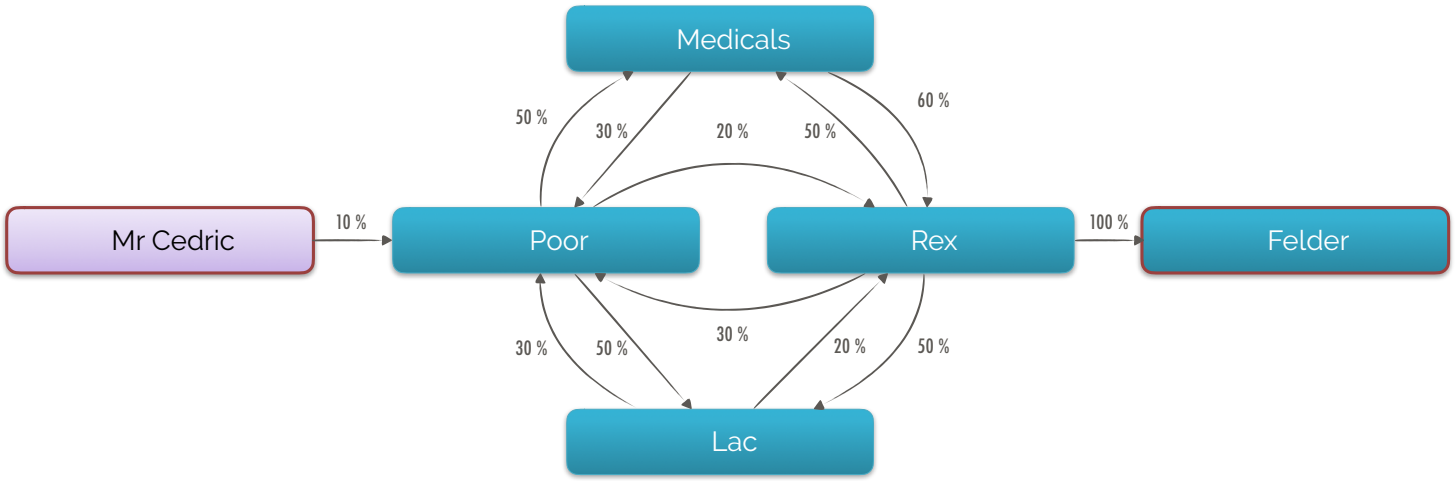
Integrated Ownership: the basic math

- Cycles (nested)



Integrated Ownership: real cases

- Cycles (nested)



Integrated Ownership: computation methods

- **Problem complexity**

  Computing "all-to-all" integrated (Baldone) ownerships can be solved in polynomial time in the number of companies. Conjecture: $n^y, y \in [2,3]$

- Our approaches:
  - **Closed-form expression** (exact solution, but unfeasible with large –real– graphs)
  - **Pure KG reasoning** (approximate solution, efficient results)
  - **Ad-hoc algorithm** (exact solution, top-level performance, ad-hoc projects)
  - **Closed-form** + **KG reasoning** (exact and efficient results, reusable and explainable)

In summary

- **Ownership problem characterization**
  - Theoretical study (e.g., complexity analysis)
  - Novel techniques to compute **all-to-all (Baldone) ownerships**
  - Efficient and fully transparent ownership model
- **Construction of the Italian company graph**
  - all Italian companies, all links, all shareholders
    - ~6.9M **nodes**, ~6.2M **edges**, ~6.9M **SCC**, ~1.3M **WCC**
  - **family links** between shareholders
- **Data + AI tools for many applications …**

Thank you!

Luigi Bellomarini
bellomarini@yahoo.it

Divisione **Ricerca** sulle **Tecnologie Avanzate**
Dipartimento **Informatica**
A R T
www.bankit.art

BANCA D'ITALIA
EUROSISTEMA