

Massive Graph Management & Analytics

GENERAL INTRODUCTION

Nacéra Seghouani

Computer Science Department, CentraleSupélec
Laboratoire Interdisciplinaire des Sciences du Numérique, LISN
nacera.seghouani@centralesupelec.fr

2024-2025

The Bridges of Königsberg

- ☞ Königsberg (Kaliningrad), the capital of Eastern Prussia, supported the trade by its fleet of ships by building seven bridges across the river Pregel that surrounded the town.
- ☞ The question is can one walk across all seven bridges and never cross the same one twice?
- ☞ The problem remained unsolved until 1735, when Leonard Euler, a Swiss born mathematician: Euler's proof was the first time someone solved a mathematical problem using a graph.



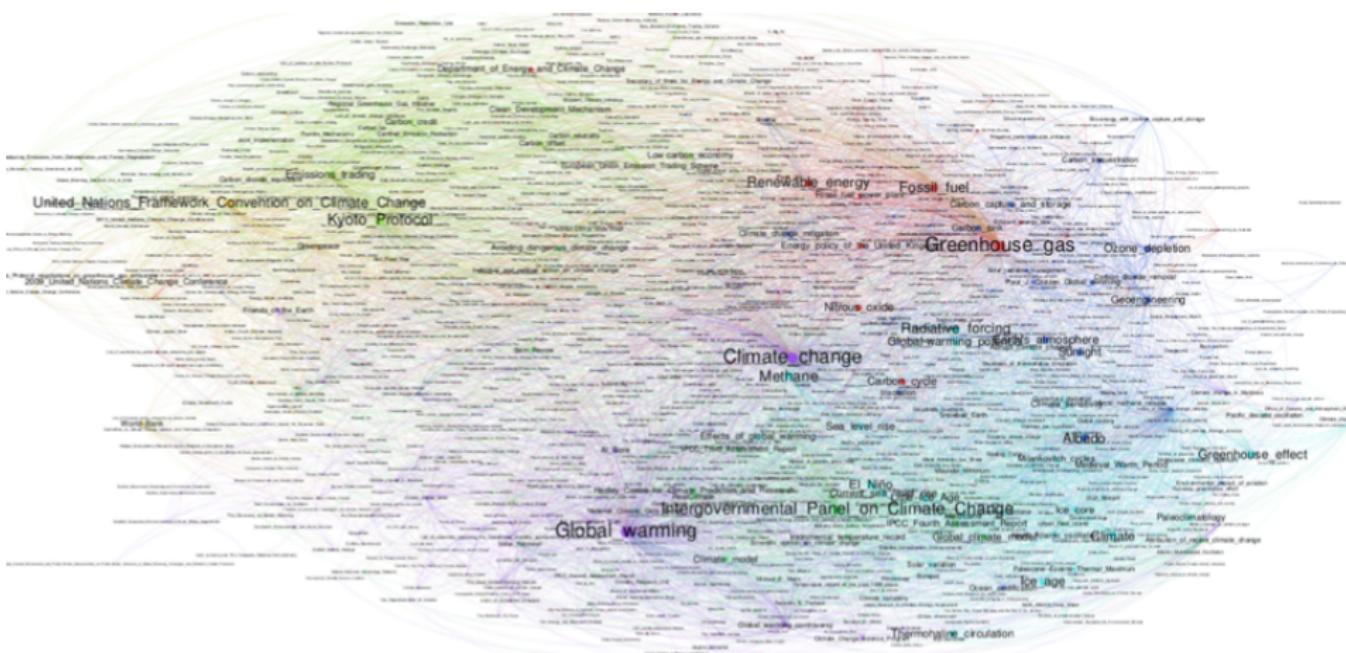
7 Bridges problem ↗

Graph-structured data is everywhere

- ☞ Graph-structured data at the heart of complex systems, play a major role in our daily life, in science and economy
 - ✓ Cooperation between billions of individuals, or communication infrastructures - billions of cell phones with computers and satellites -, interactions between thousands of genes and metabolites within our cells, billions of neurons in our brain, internet of things ...
- ☞ Understanding mathematical foundations, description, prediction, and eventually control this kind of systems is one of the major scientific challenges of the 21st century.
 - ✓ The Bridges of Königsberg (18th century with Euler), Paul Erdős and Alfréd Rényi (1959), Mark Granovetter (1973)
 - ✓ Emergence of Network Science
 - ✓ Also the most revolutionary technologies, empowering everything from Google to Facebook, CISCO, Twitter, LinkedIn, ...
 - ✓ And neural networks ...

Web Graphs

- Vertices: Web pages
 - Edges: Links
 - Questions: the most referenced pages, hidden/isolated pages, ...



Social Networks

- Vertices: Users, Posts (Videos, Music, Comments ...)
- Edges: Social Relationships (directed: Twitter, undirected: Facebook), Likes, Texts ...
- Questions: the influencers, which kind communities and profiles, ...

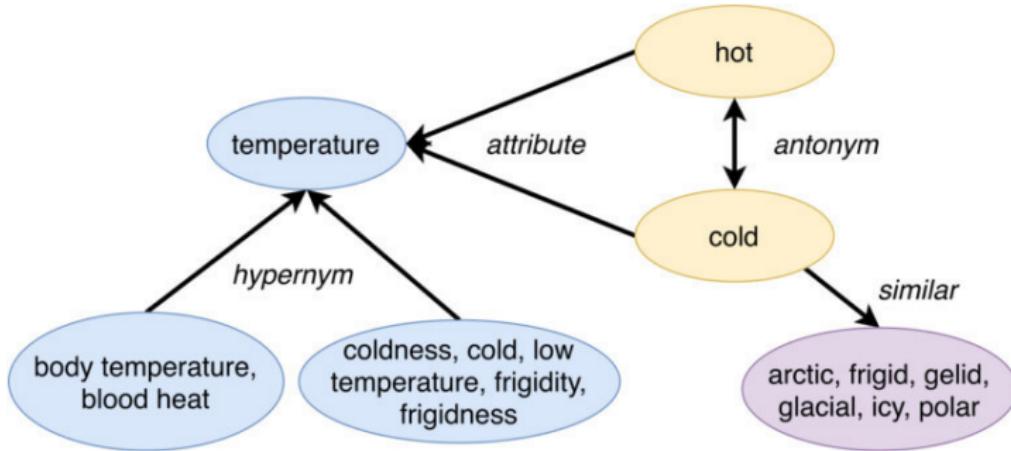


SOCIAL MEDIA
In Business

Encyclopedia, Semantic & Knowledge Networks

Wordnet: network of English words (extended to other languages) linked by lexical and semantic relations. It can thus be seen as a combination and extension of a dictionary and thesaurus.

- ☒ nouns, verbs, adjectives, adverbs, ...
- ☒ synonymy or the opposite antonymy, meronymy (denotes a constituent part of) or the opposite holonymy, hyponymy (denotes a kind of) or hypernymy ...



Encyclopedia, Semantic & Knowledge Networks

Typical NLP problems

- Information extraction at sentence or document level (across sentences)
- Word sense disambiguation (neighbouring words)
- Entity Resolution, Entity Linking, Relation Extraction

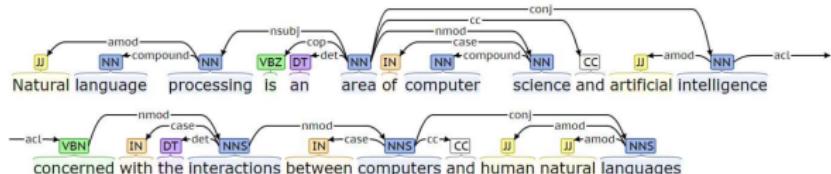
Part-of-speech tagging (POS tagging)

Natural language processing Is an area of computer science and artificial Intelligence concerned with
the interactions between computers and human natural languages

Named entity recognition

Thomas Cruise, born at Syracuse, New York, U.S., July 3, 1962, is an American actor and producer.
He has been nominated for three Academy Awards and has won three Golden Globe Awards.
He started his career at age 19 in the film Endless Love.

Dependency Parsing



Encyclopedia, Semantic & Knowledge Networks

- Vertices: Articles
- Edges: cross-language links, internal links, category links, redirect links ...

WIKIPEDIA

The Free Encyclopedia

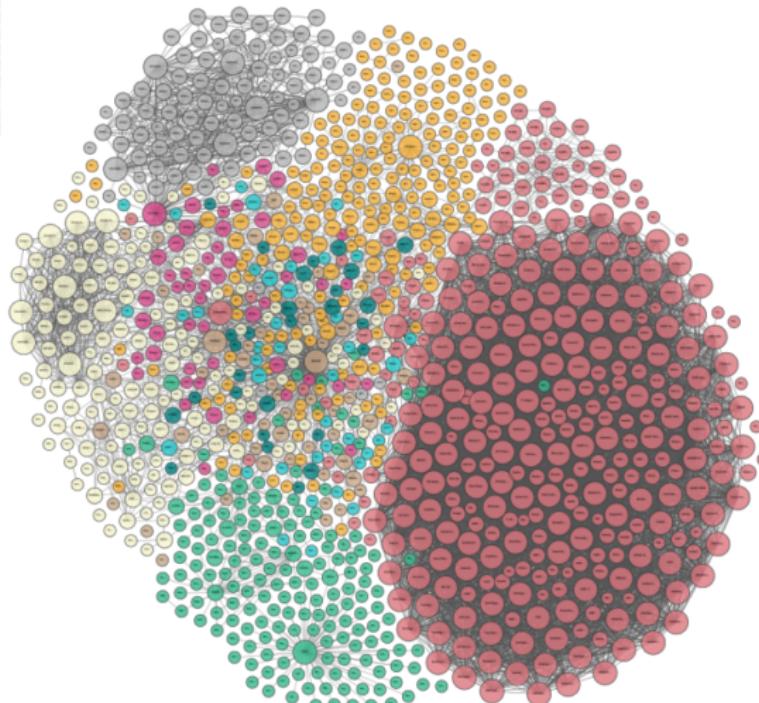


Encyclopedia, Semantic & Knowledge Networks

DBpedia: Open data project built from wikipedia and other thesaurus based on ontologies (Yago, wordnet, DBLP, Freebase, Wikidata...). Used by Google, IBM Watson's, ... Popular: Google and Satori Knowledge Graphs ↗

- According to Gruber, *ontology is an explicit specification of a conceptualization*, which makes it possible to specify in a formal language the concepts of a domain and their relationships.

IBM Watson's Jeopardy!

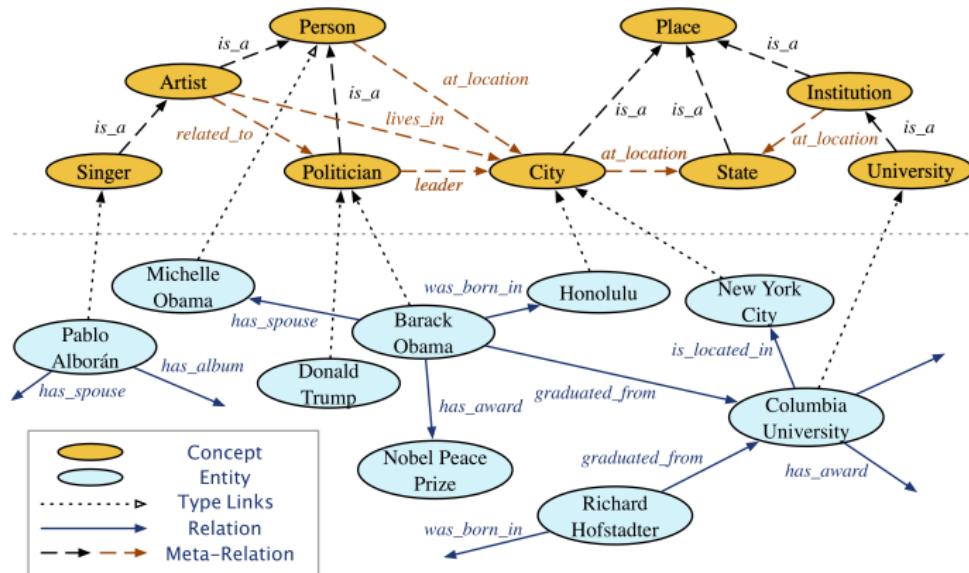


Encyclopedia, Semantic & Knowledge Networks

Vertices: Subject s , Object o - Edges: Semantic relationships r

KB is a set of triples $\langle s, r, o \rangle$: T-box (is-a, subclassOf, equivalentClass, properties, domain, range, ...) and A-box (class instances and their relationships)

Ontology-view Knowledge Graph



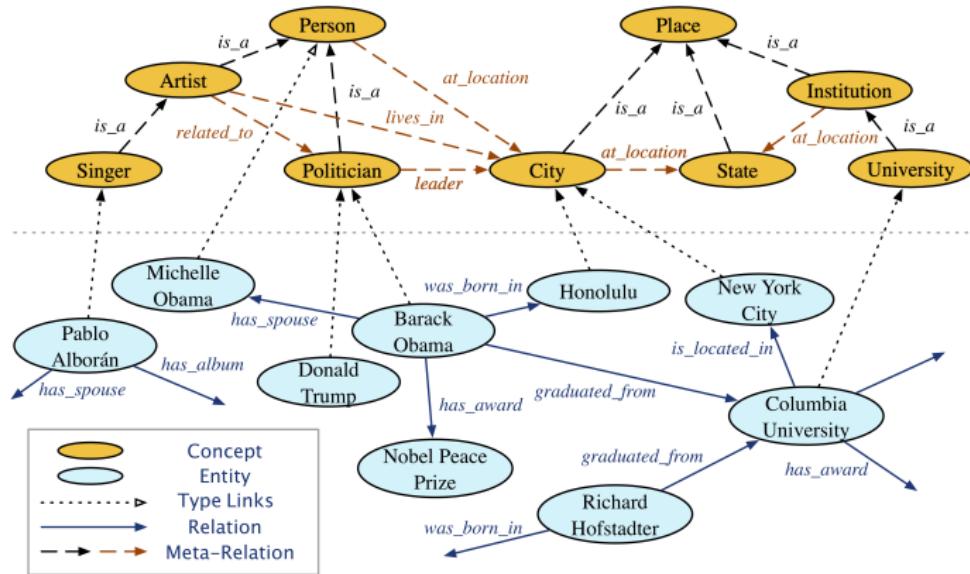
Instance-view Knowledge Graph

Encyclopedia, Semantic & Knowledge Networks

A knowledge graph:

- Inference engine base on description logics Challenges: Incomplete, Inconsistent, Ambiguous
- W3C languages: RDF, OWL, SPARQL, ...

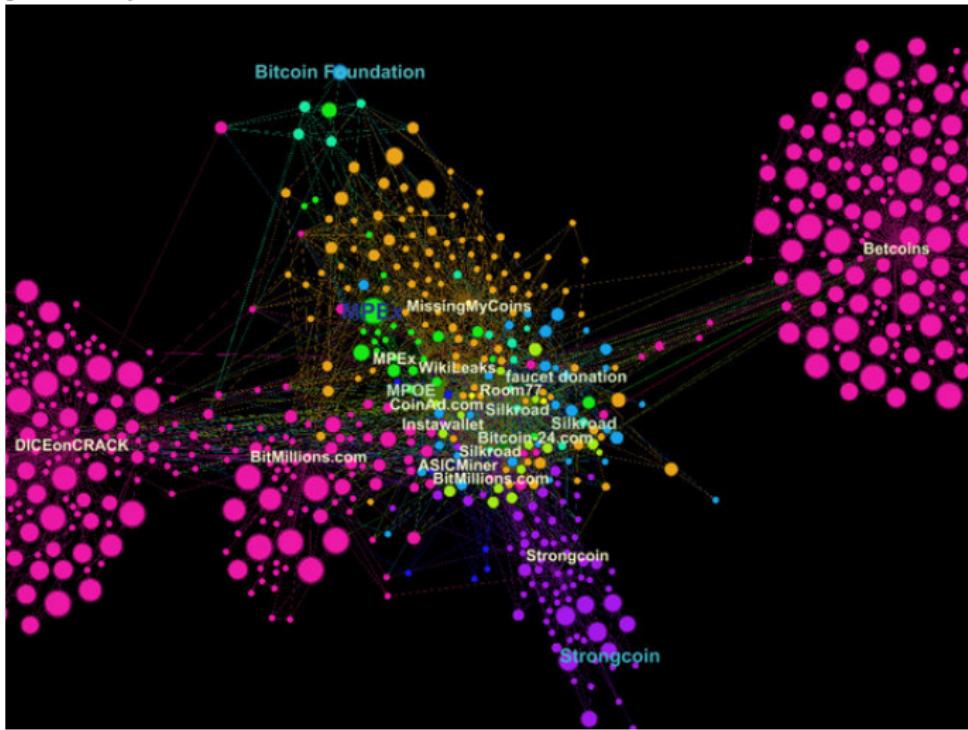
Ontology-view Knowledge Graph



Instance-view Knowledge Graph

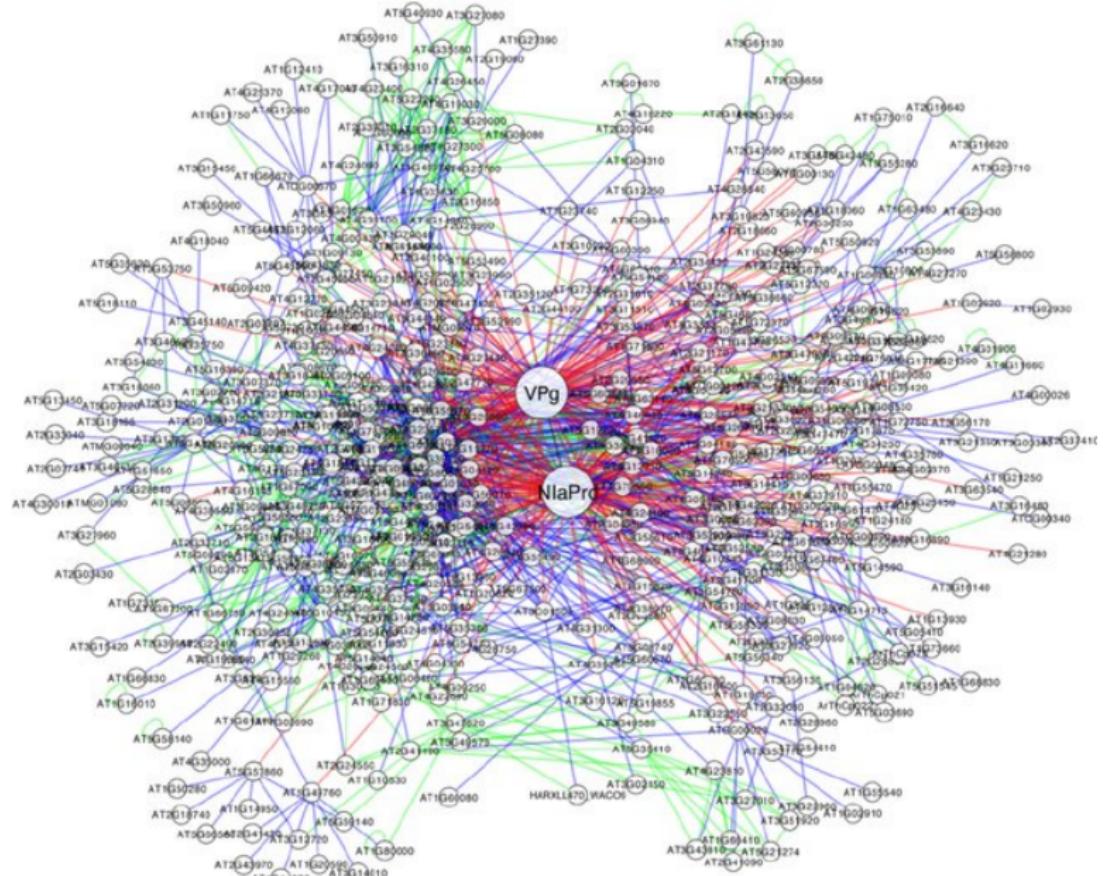
Transaction Networks (e.g., Bitcoin)

- Vertices: Users
- Edges: Exchange of Currency



Biological Networks

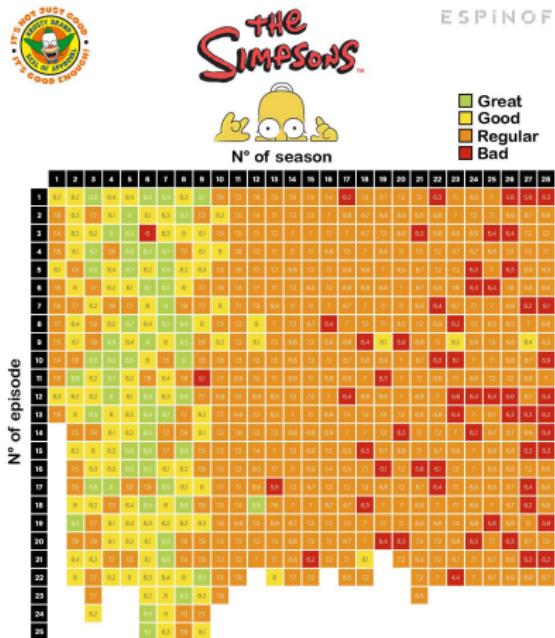
Protein-protein interaction, Metabolic networks, Gene regulatory networks



Recommender Systems

IMDb: Database related to films, television programs, videos, games, streaming content, ...

- Vertices: Users and Items (content-based) or Users and Users (collaboration-based)
- Edges: ratings or similarities



Outline

☞ Main Topics:

PART 1

- ✓ Preliminaries on graph theory & linear algebra
- ✓ Random Walk & PageRank algorithms
- ✓ Graph centrality measures
- ✓ Community detection, Influence maximisation, Label propagation problems

PART 2

- ✓ Embedding graphs: Spectral approaches, Learning-based graph embedding approaches.

☞ Assessment: different kind of grading during the classes

- ✓ Individual assessment: 2 main quizzes for each part 70%
- ✓ Assessment per pair for each part: 2 main mini-project (presentation + delivers) 30%
- ✓ No resit exam.