

# IA<sup>3</sup> 2020

10th Workshop on Irregular  
Applications: Architectures and  
Algorithms

November 11, 2020  
Virtual Workshop  
In conjunction with SC20



## SC20

Everywhere | more

# The Future is Big Graphs!

Angela Bonifati (Lyon 1 University & INRIA, France)

# New challenges for Next-generation Graph Processing Systems

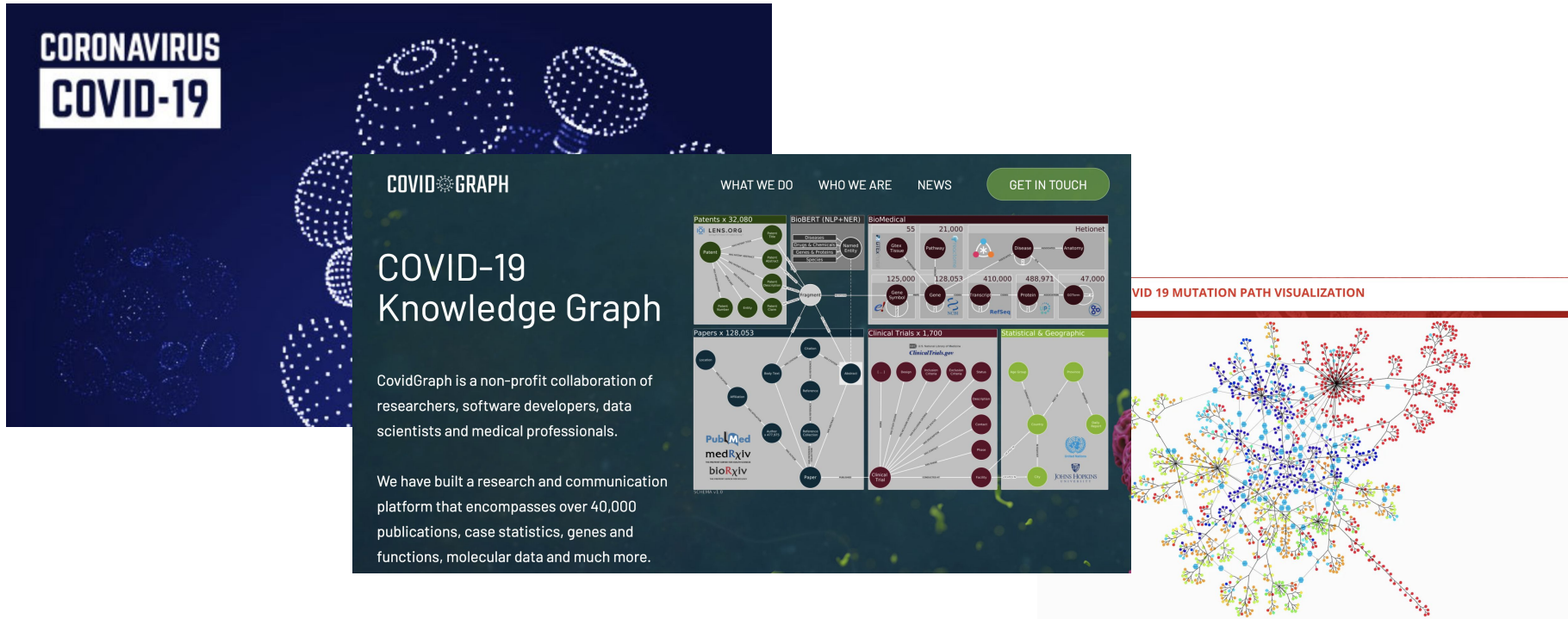
**Ch1. A lattice of graph data models and graph algebras**

**Ch2. Complex data management ecosystems**

**Ch3. Performance and benchmarking**

S. Sakr, A. Bonifati, H. Voigt, A. Iosup et al. "The Future is Big Graphs: A Community View on Graph Processing Systems" To appear in Communications of the ACM

# Graphs are ubiquitous across diverse applications: an example is the covidgraph.org



**CORONAVIRUS  
COVID-19**

**COVID GRAPH**

**COVID-19  
Knowledge Graph**

CovidGraph is a non-profit collaboration of researchers, software developers, data scientists and medical professionals.

We have built a research and communication platform that encompasses over 40,000 publications, case statistics, genes and functions, molecular data and much more.

WHAT WE DO WHO WE ARE NEWS GET IN TOUCH

Patents x 32,080  
LENS.ORG

BioBERT (NLP+NER)  
Genes & Compounds  
Genes & Proteins  
Protein  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein

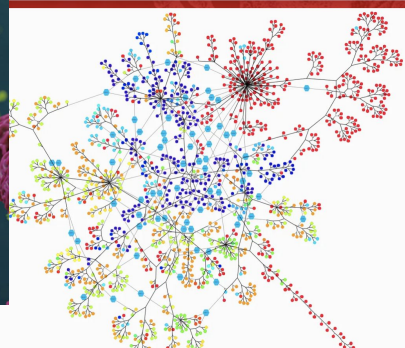
BioMedical  
55 21,000  
Hetionet  
125,000 128,053 410,000 488,971 47,000  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein

Papers x 128,053  
PubMed  
medRxiv  
bioRxiv  
Paper  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein

Clinical Trials x 1,700  
ClinicalTrials.gov  
Clinical Trial  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein

Statistical & Geographic  
Statistical  
Geographic  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein  
Gene  
Protein

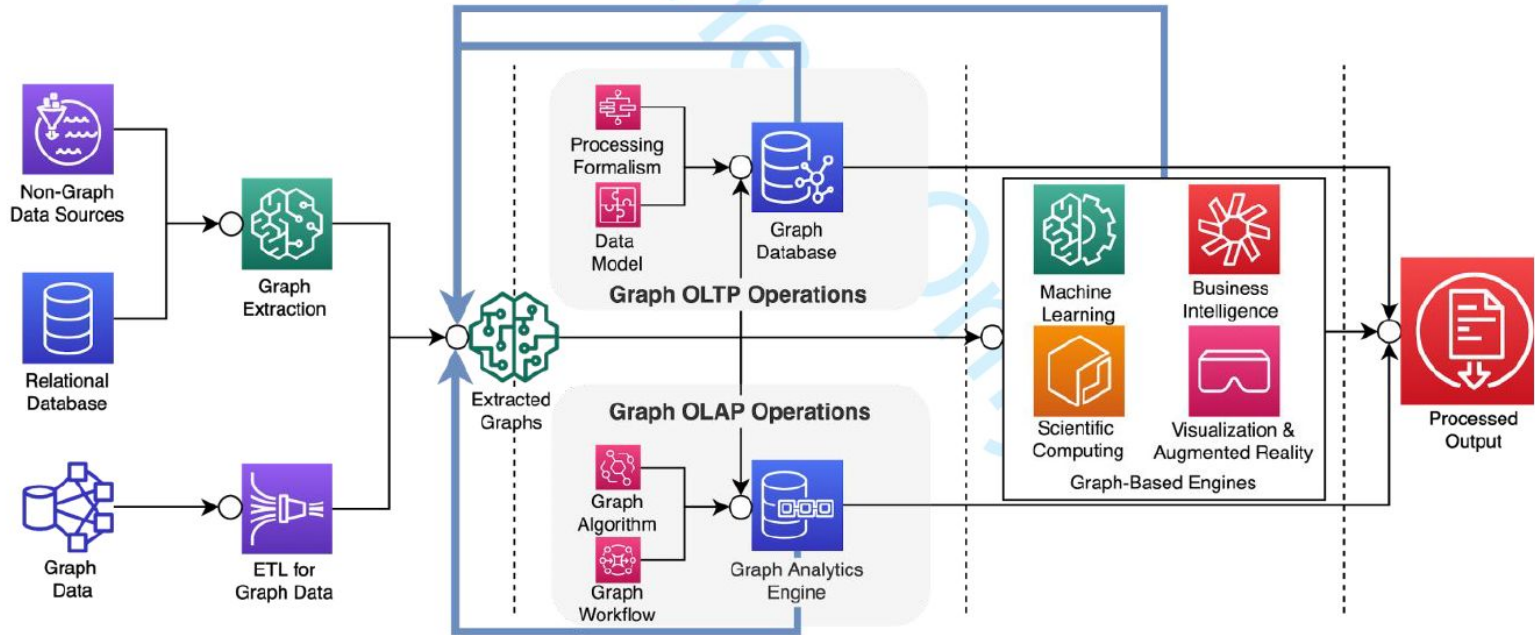
**COVID 19 MUTATION PATH VISUALIZATION**



# Ch1. Expressivity of the graphs/queries

- Dependence on the chosen data model
  - How do **humans** conceptualize graphs?
  - The **interoperability** issues (due to multiple heterogeneous data sources) are to be taken into account
  - A **data model lattice** to navigate across data models, balancing understandability and expressive power
  - A new **algebra** for the variety of graph workloads

# Ch2. A complex data management ecosystem



## Ch3. Performance and benchmarking

- The need for new, **reproducible experimental** methodologies to facilitate quick yet meaningful performance-testing?
- How to define more faithful **metrics** for executing a graph algorithm, query, program, or workflow?
- How to **generate workloads** with combined operations, covering temporal, spatial, and streaming aspects?
- How to **benchmark** pipelines including machine learning and simulation?