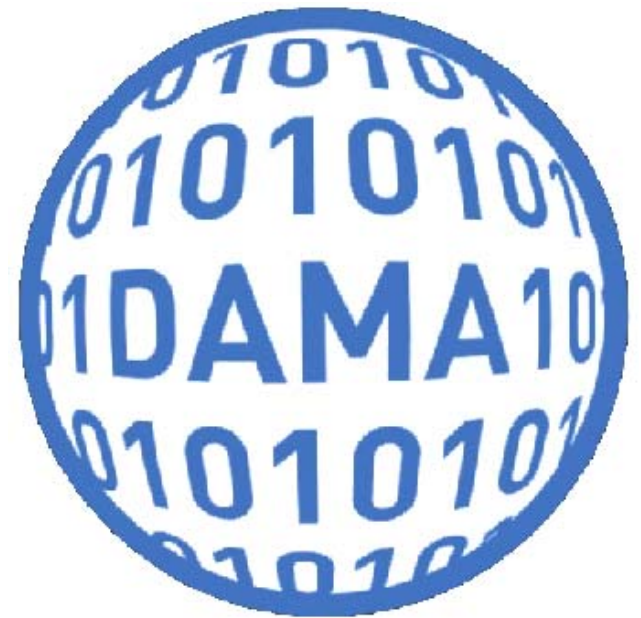




DAMA-UPC. DATA MANAGEMENT
UNIVERSITAT POLITÈCNICA DE CATALUNYA



Graphs for the analysis of large amounts of internet data

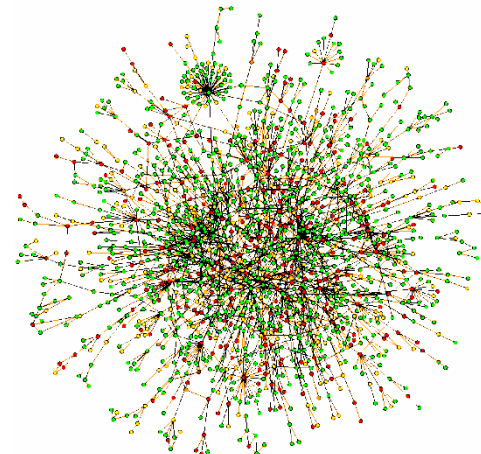
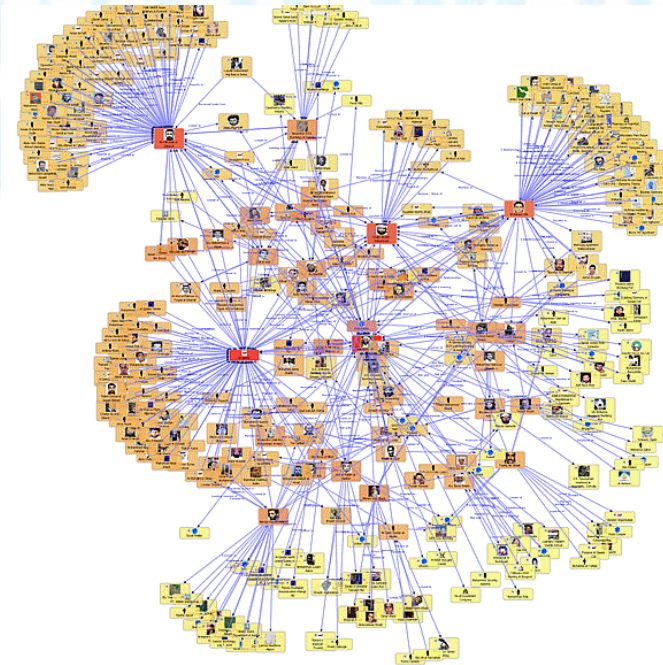
Josep Lluís Larriba Pey

larri@ac.upc.edu

www.dama.upc.edu

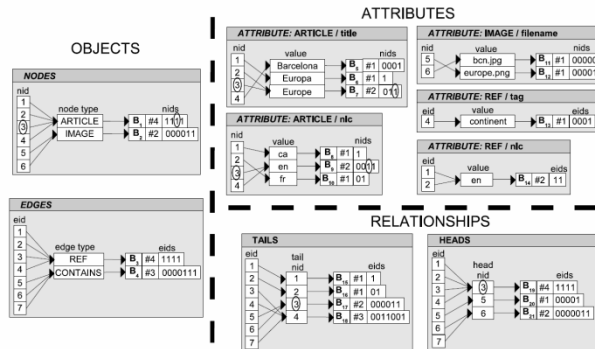
Graph Databases (GDBs)

- The amount of applications calling for **efficient** management of **massive** graphs is **increasing** everyday. Specially in internet.
- Interest for the **structural analysis** of entity relationship: patterns, communities, recommendation based on history, etc.
- However, graphs are growing so rapidly that
 - Graph size prevents the system from accommodating it entirely in-memory
 - Operations are very complex and time-consuming

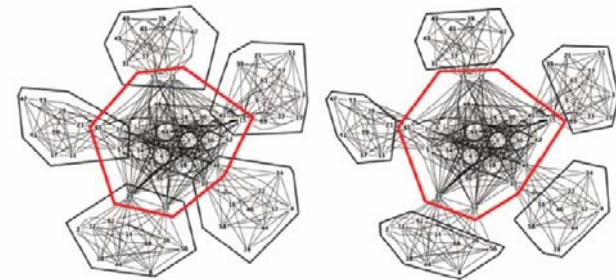


Challenges in GDBs

- Efficient storage

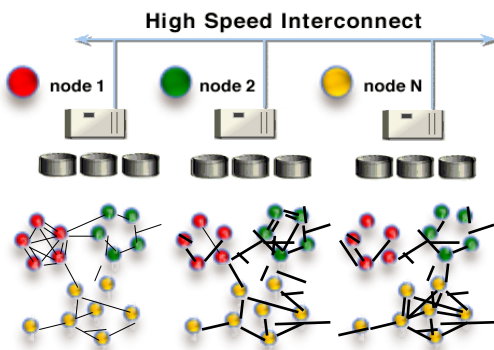


- Community search in very large graphs

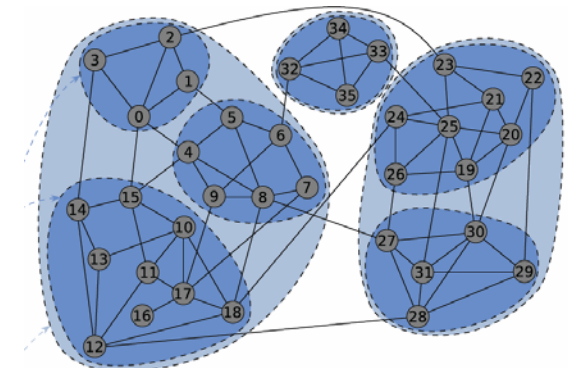
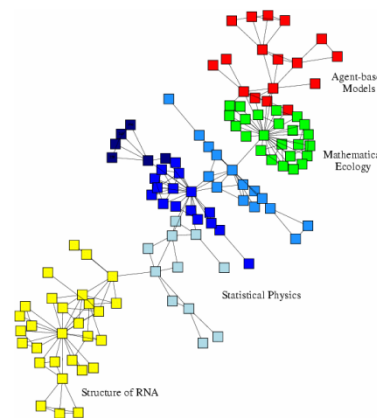


- Graph layouts and performance

- Distributed Graph Databases



- Graph Integration



- Recommendation based on graphs

Technology infrastructure

DEX: graph database management system

- Manages large graphs with more than 1 billion objects
- Already tested in real applications: registrars, cancer, PPI, bibliographic exploration, publicist apps, etc.
- Can integrate several data sources
- Fast response times for complex queries
- A support team with engineers specializing in the kernel and applications
- Needed a lot of research, but use this as a platform for research and development of prototypes

DAMA-UPC, 25 people with knowledge and experience in:

- Graph databases,
- Data cleansing,
- Management of large datasets

