ESIGMA

Efficiency and Structure in Graph Mining Applications

Michalis Vazirgiannis
LIX (Laboratoire d'Informatique de l'École polytechnique)

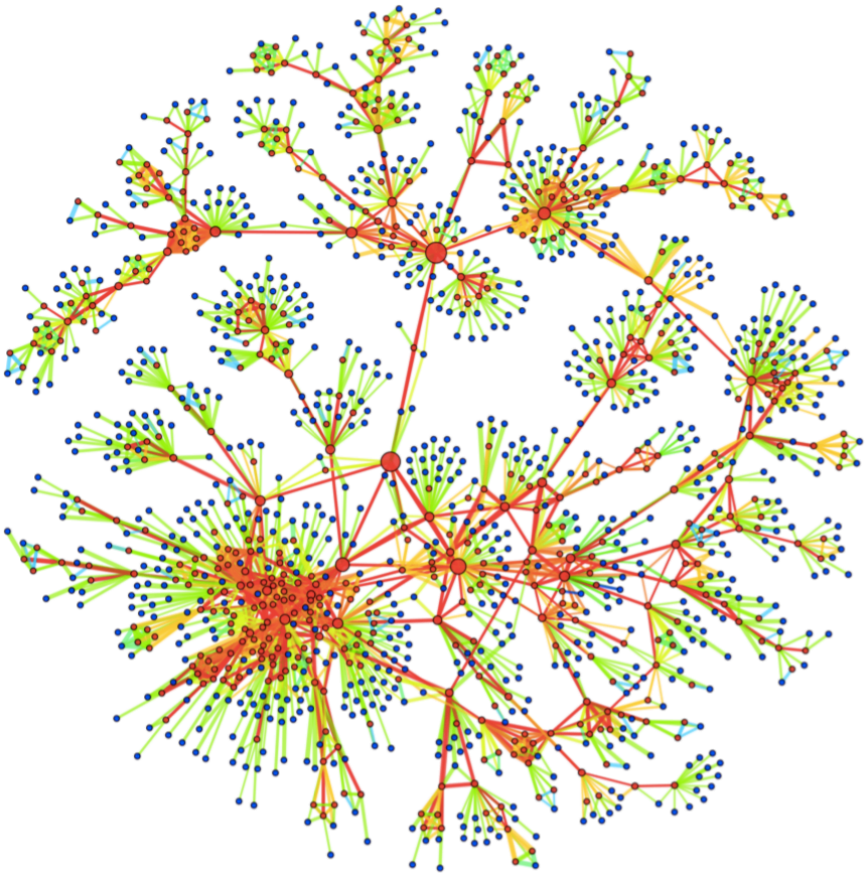
Kickoff meeting - ANR ESIGMA project, May 31 2018

Participating institutions:

- LAMSADE (CNRS/Univ. Paris-Dauphine), Eunjung Kim
- LIRMM (CNRS/Univ. Montpellier), Dimitrios Thilikos Touloupas
- LIX (École polytechnique), Michalis Vazirgiannis

Data mining/ Machine Learning: identify, characterise, & analyze structure of data sets

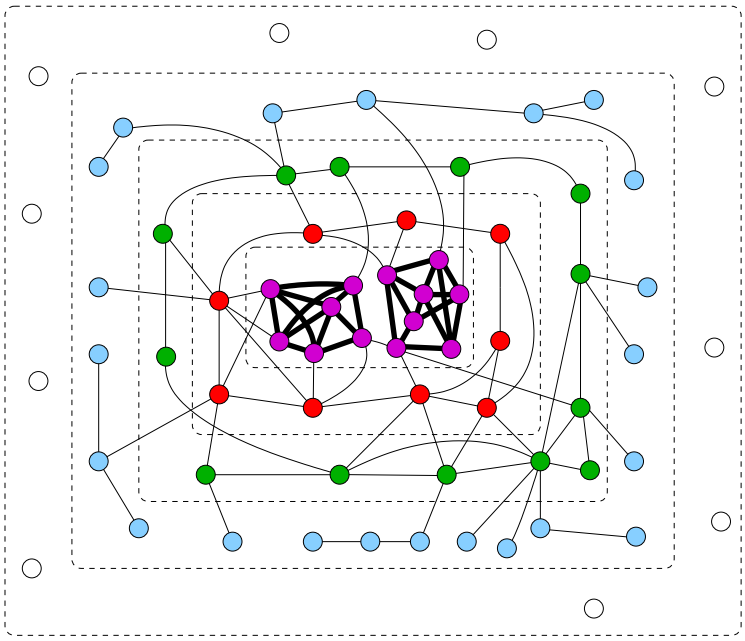
Graph mining: data is organized/represented in the form of **graphs**



Graph data illustrate diverse **structural** characteristics:

Example:

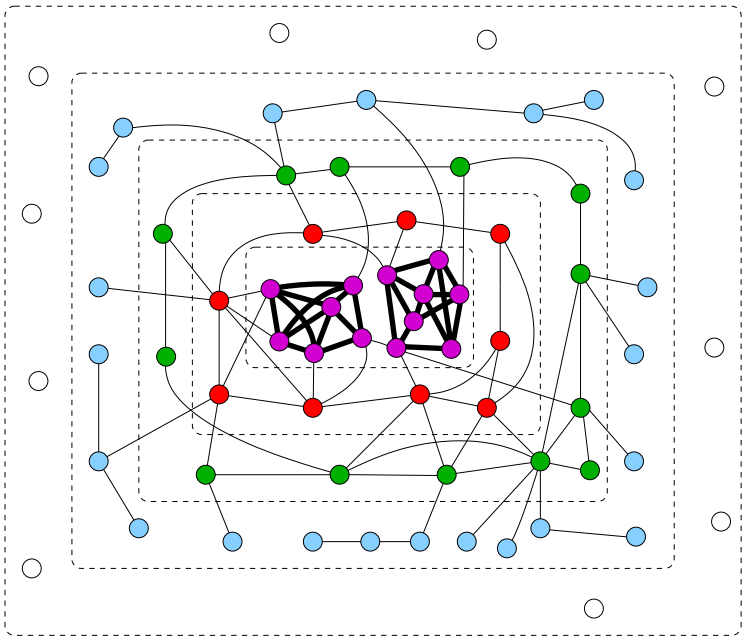
Graph **core** hierarchies as approximation of the densest subgraph. Applications: clustering, supervised learning, text mining, graph classification ...



Graph data illustrate diverse **structural** characteristics:

Example:

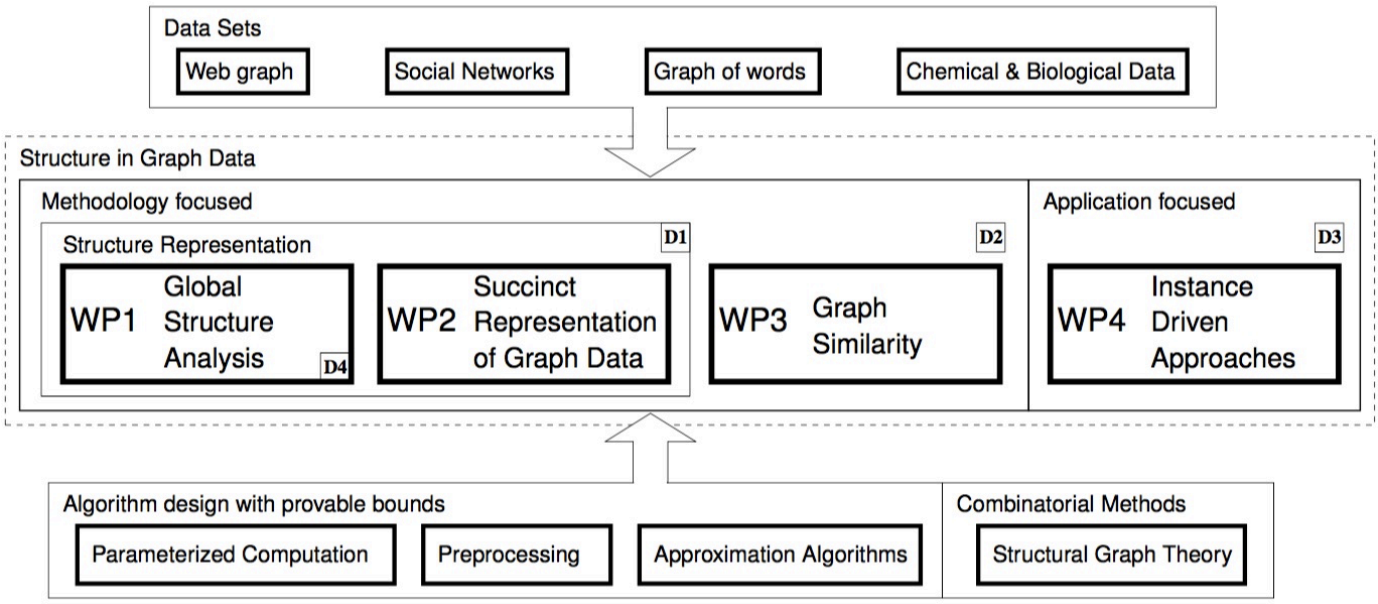
Graph **core** hierarchies as approximation of the densest subgraph. Applications: clustering, supervised learning, text mining, graph classification ...



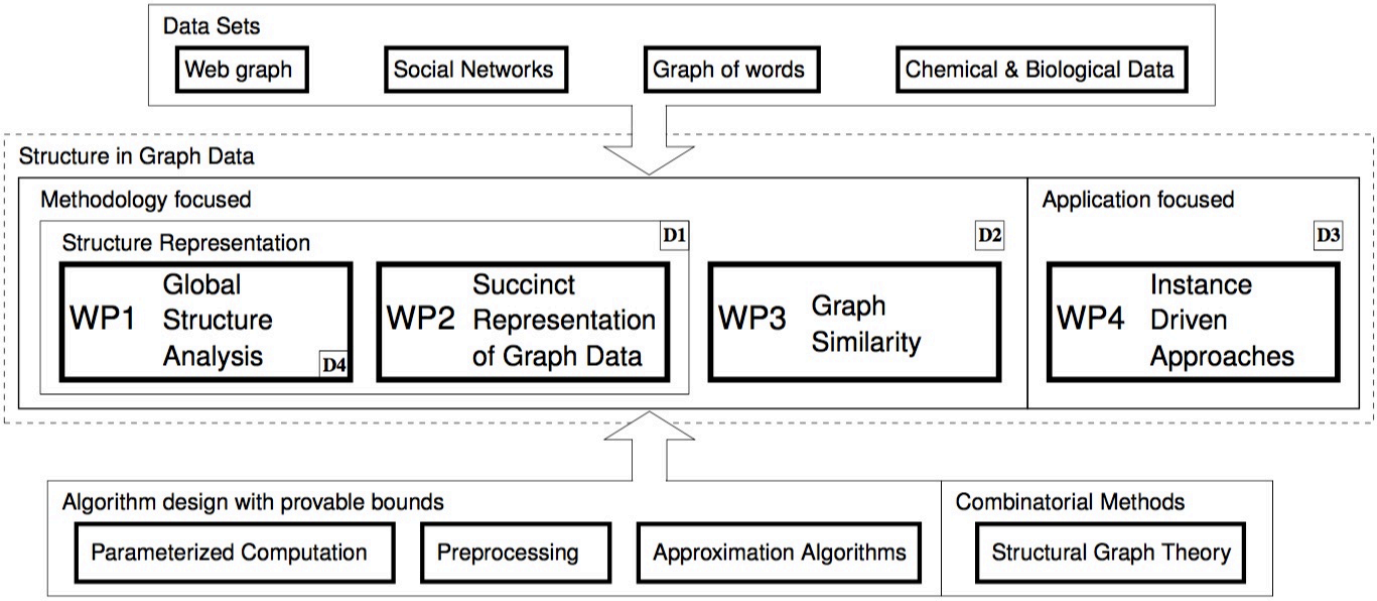
Theoretical Context:

- **Graph theory**
- **Discrete algorithms**

ESIGMA: study algorithmic/combinatorial facets of structure in graph data

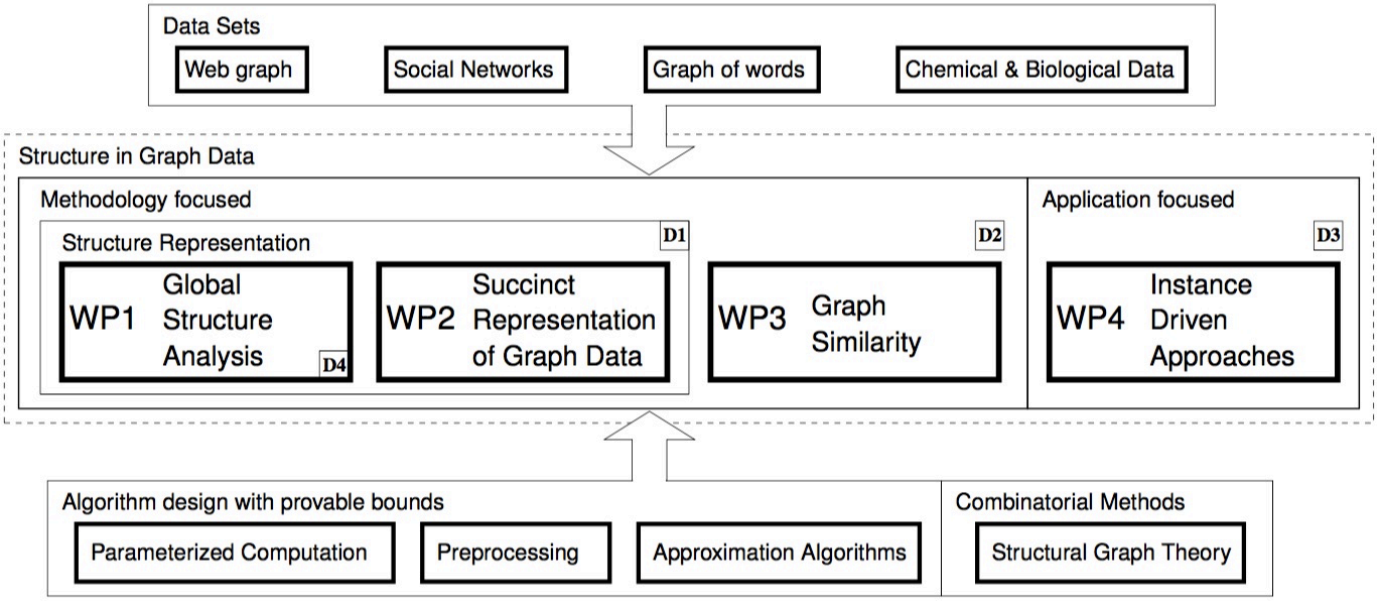


ESIGMA: study algorithmic/combinatorial facets of structure in graph data



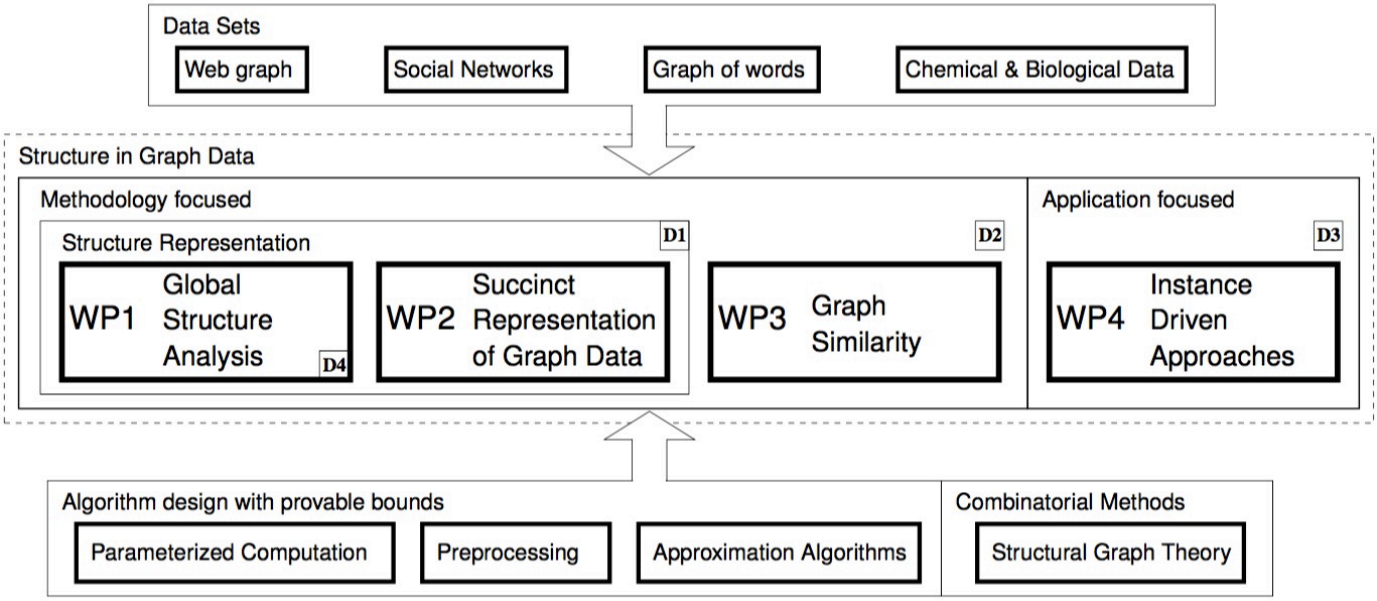
- **WP1:** Global Structure Analysis
  - Core decompositions.
  - Alternative criteria for global structure analysis
  - Theoretical tools for global structure analysis

ESIGMA: study algorithmic/combinatorial facets of structure in graph data



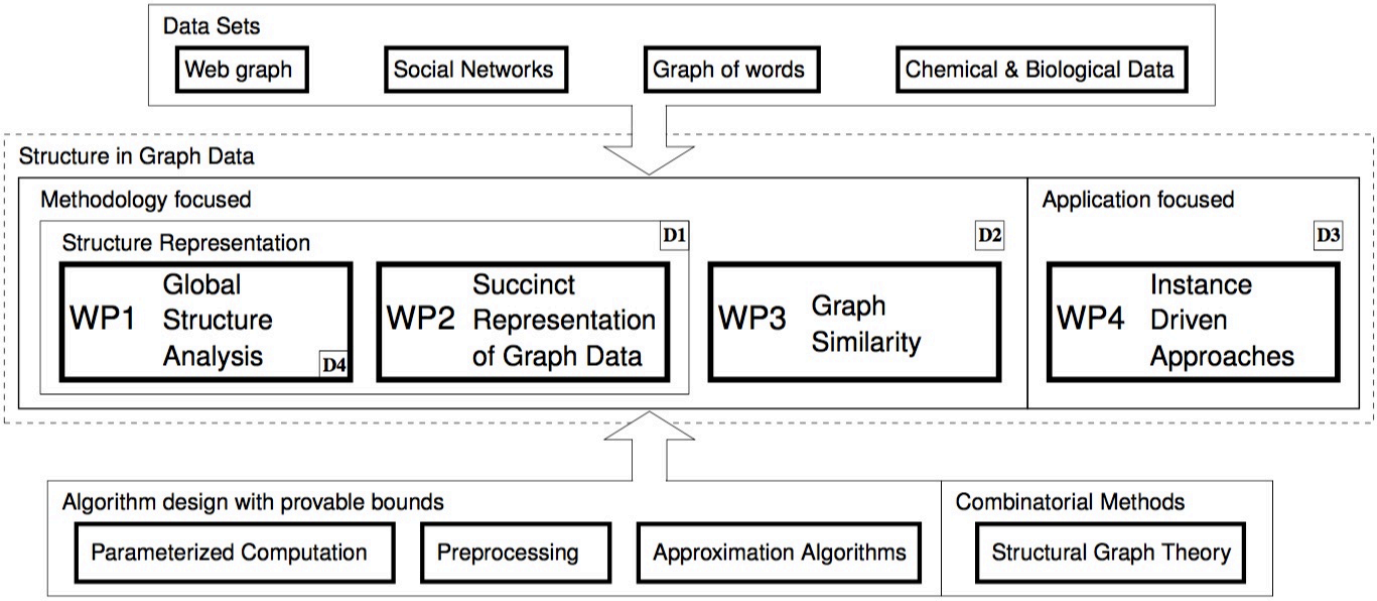
- **WP2:** Succinct Representation of Graph Data:
  - Sparse coding
  - Graph dictionary learning

ESIGMA: study algorithmic/combinatorial facets of structure in graph data



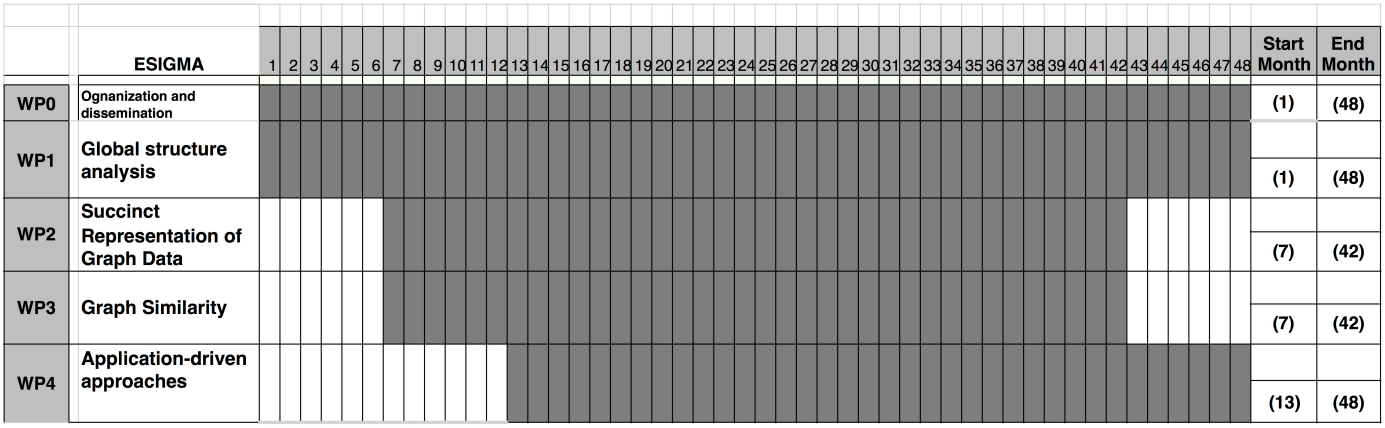
- **WP3:** Graph Similarity:
  - Graph matching using the structure of graph data
  - Graph similarity measured by edit distance
  - Graph Kernels
  - Frequent subgraph mining in graphs

ESIGMA: study algorithmic/combinatorial facets of structure in graph data



- **WP4:** Use the above methodologies to study concrete applications:
  - Community detection and evaluation for social and academic data
  - Identification of influential spreaders
  - Graph based event detection in social media streams
  - Grpah of words for text mining
  - Privacy and Anonymization

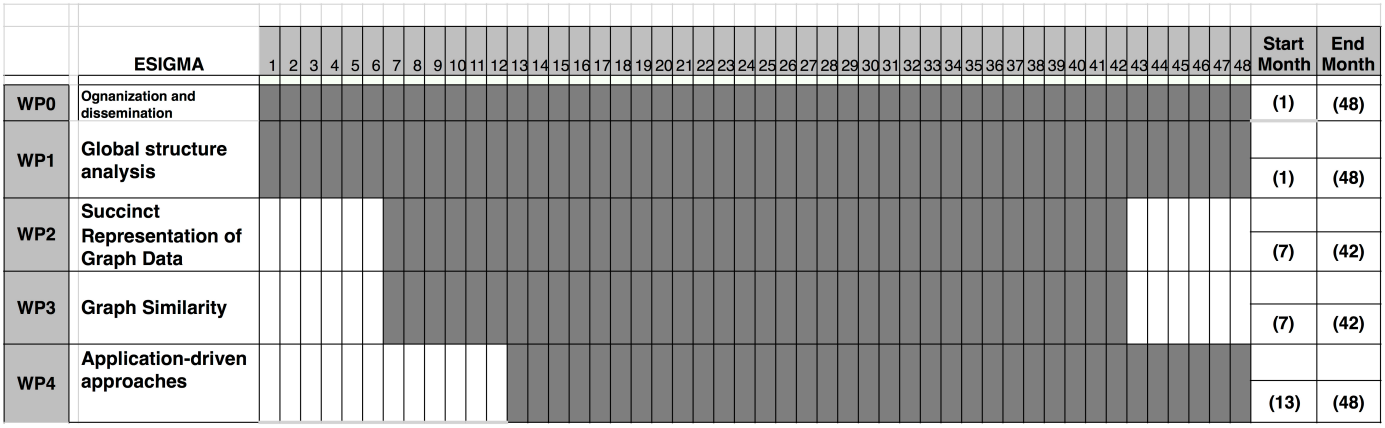
Schedule



	ESIGMA	Start Month	End Month
WP0	Ognanization and dissemination	(1)	(48)
WP1	Global structure analysis	(1)	(48)
WP2	Succinct Representation of Graph Data	(7)	(42)
WP3	Graph Similarity	(7)	(42)
WP4	Application-driven approaches	(13)	(48)

Duration: 48 months

Schedule



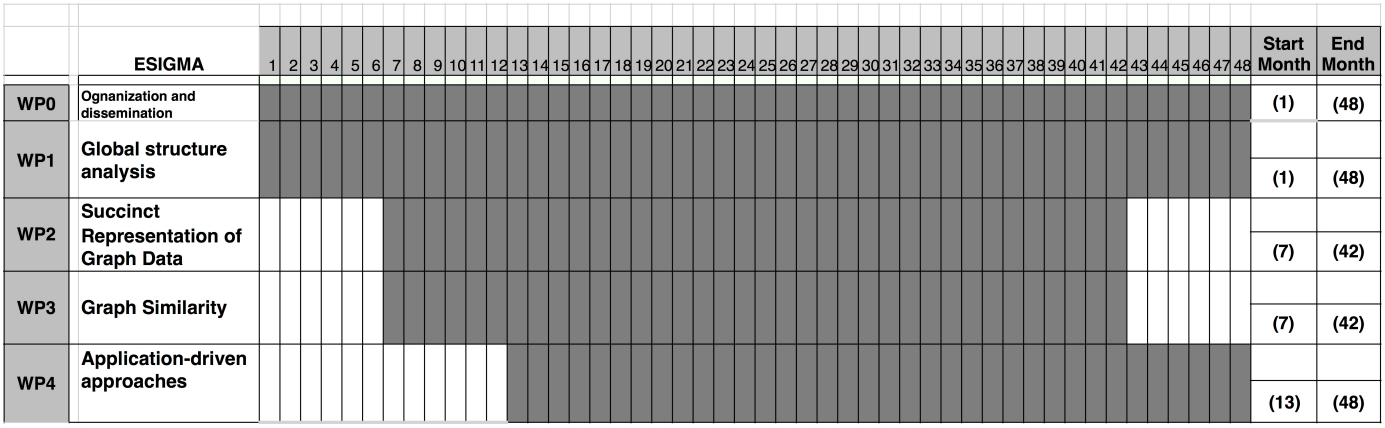
	ESIGMA	Start Month	End Month
WP0	Ognanization and dissemination	(1)	(48)
WP1	Global structure analysis	(1)	(48)
WP2	Succinct Representation of Graph Data	(7)	(42)
WP3	Graph Similarity	(7)	(42)
WP4	Application-driven approaches	(13)	(48)

Duration: 48 months

- **Objective of ESIGMA**:

To bridge the gap between graph mining and graph theory, focusing on the combinatorial aspects of graph-data structure.

Schedule



	ESIGMA	Start Month	End Month
WP0	Ognanization and dissemination	(1)	(48)
WP1	Global structure analysis	(1)	(48)
WP2	Succinct Representation of Graph Data	(7)	(42)
WP3	Graph Similarity	(7)	(42)
WP4	Application-driven approaches	(13)	(48)

Duration: 48 months

- **Objective of ESIGMA**:

To bridge the gap between graph mining and graph theory, focusing on the combinatorial aspects of graph-data structure.

Consortium:

• LIX: C. Giatsidis, C. Palamidessi, J. Read, and *M. Vazirgiannis* (scientific coordinator).

• LAMSADE: *E.J. Kim* (local coordinator), M. Lampis, B. Negrevergne, F. Sikora, and F. Yger.

• LIRMM: C. Paul, I. Sau, and *D. Thilikos* (local coordinator).

Consortium:

• LIX: C. Giatsidis, C. Palamidessi, J. Read, and *M. Vazirgiannis* (scientific coordinator).

• LAMSADE: *E.J. Kim* (local coordinator), M. Lampis, B. Negrevergne, F. Sikora, and F. Yger.

• LIRMM: C. Paul, I. Sau, and *D. Thilikos* (local coordinator).

Combined expertise from Machine Learning /Data Mining and Theoretical Computer Science:

- Structural graph theory (Sau, Thilikos),
- Approximate & parameterized algorithms (Kim, Lampis, Paul, Sikora, Thilikos),
- Graph Similarity, deep learning for classification - Network Embeddings, Community Detection (Giatsidis, Read, Thilikos, Vazirgiannis),
- Influence Maximization, Applications in Text Mining (keyword extraction, summarization) (Giatsidis, Nikolentzos, Vazirgiannis),
- Privacy in Graph Mining (Palamidessi, Vazirgiannis),
- Representation learning (Yger),
- High performance computing (Negrevergne),
- Automated pattern mining, Constraint programming/satisfaction (Negrevergne, Kim, Lampis, Sikora)

TRL: 2-5

Budget of ESIGMA

	LIX	LAMSADE	LIRMM	SUM
Post-docs	102.690 €	96.098 €	0€	153.720 €
Internships	0 €	4.400 €	0 €	14.400 €
PhD students	0 €	0 €	94.320 €	125.760 €
Total salaries	**102.690 €**	**100.498 €**	**94.320 €**	**297.508 €**
Project meetings, event	8.000 €	5.000 €	3.000€	16.000 €
Visits, workshops, conferences	18.000 €	22.653 €	14.794 €	55.447 €
Total travels	**26.000 €**	**27.653 €**	**17.794 €**	**71.447 €**
Equipments (Computers)	**3.000 €**	**3.000 €**	**3.000 €**	**9.000 €**
Sum	**131.690 €**	**131.151 €**	**115.114 €**	**377.955 €**
Structure cost	10.535,20 €	10.492,08 €	9.209,12 €	
Total	**142.225,20 €**	**141.643,08 €**	**124.323,12 €**	**408.191,4 €**

Budget of ESIGMA

	LIX	LAMSADE	LIRMM	SUM
Post-docs	102.690 €	96.098 €	0€	153.720 €
Internships	0 €	4.400 €	0 €	14.400 €
PhD students	0 €	0 €	94.320 €	125.760 €
Total salaries	**102.690 €**	**100.498 €**	**94.320 €**	**297.508 €**
Project meetings, event	8.000 €	5.000 €	3.000€	16.000 €
Visits, workshops, conferences	18.000 €	22.653 €	14.794 €	55.447 €
Total travels	**26.000 €**	**27.653 €**	**17.794 €**	**71.447 €**
Equipments (Computers)	**3.000 €**	**3.000 €**	**3.000 €**	**9.000 €**
Sum	**131.690 €**	**131.151 €**	**115.114 €**	**377.955 €**
Structure cost	10.535,20 €	10.492,08 €	9.209,12 €	
Total	**142.225,20 €**	**141.643,08 €**	**124.323,12 €**	**408.191,4 €**

Integration of the project:

Défi 7: Société de l'information et de la communication.

Axe 5: Données, Connaissances, Big Data - Contenus multimédias.

Orientation 28: Exploitation des grandes masses de données

Merci!