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

・ロト・日本・日本・日本・日本

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

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



ロト 4 母 ト 4 目 ト 4 目 ・ の へ (?)

Graph data illustrate diverse structural characteristics:

Example:

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



・ロト・西ト・ヨト・ヨト つへぐ

Graph data illustrate diverse structural characteristics:

Example:

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



 $\bullet \square \bullet$

▲ □ ▶ ▲ 三 ▶ ▲ 三 ▶

E

SQ (~

Theoretical Context:

- Graph theory
- Discrete algorithms

Data Sets Web graph Social Networks Graph of words Chemical & Biological Data Structure in Graph Data Methodology focused Application focused D2 D1 D3 Structure Representation Instance Global Succinct Graph WP2 Representation WP3 WP1 WP4 Structure Driven Similarity of Graph Data Analysis Approaches D4 Algorithm design with provable bounds **Combinatorial Methods** Parameterized Computation Preprocessing Approximation Algorithms Structural Graph Theory

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

▲□▶▲□▶▲□▶▲□▶ ▲□ ◆ � ◆

Data Sets Web graph	Social Networks	Graph of words	Chemical &	Biological Data
tructure in Graph Data				
Methodology focused				Application focused
Structure Representation	~~~~	D1	D2	I
Global WP1 Structure Analysis _{D4}	Succinct WP2 Representation of Graph Data	WP3 Graph Simila	n arity	Instance WP4 Driven Approaches
Algorithm design with provable	e bounds		Combi	natorial Methods
Parameterized Computatio	n Preprocessing	Approximation Algorith	ms Str	uctural Graph Theory

WP2: Succinct Representation of Graph Data:

- Sparse coding
- Graph dictionary learning

◆□ ▶ ◆□ ▶ ◆ 三 ▶ ◆ 三 ● のへぐ

Data Sets Web graph	Social Networks	Araph of words	Chemical &	Biological Data
tructure in Graph Data	7	7		
Methodology focused	· · · · · · · · · · · · · · · · · · ·			Application focused
Structure Representation		D1	D2	D3
Global WP1 Structure Analysis _{D4}	WP2 Succinct WP2 Representation of Graph Data	WP3 Graph Similarity	/	Instance WP4 Driven Approaches
Algorithm design with provable	bounds		Combi	natorial Methods
Parameterized Computation	Preprocessing	approximation Algorithms	Str	uctural Graph Theory

WP3: Graph Similarity:

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

・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・
・

lethodology focused		<u></u>		Application focused
Structure Representation		D1	D2	
Global WP1 Structure Analysis _{D4}	Succinct WP2 Representatio of Graph Data	n WP3 Graph Similarity	/	Instance WP4 Driven Approaches
Analysis D4	of Graph Data			Approaches

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

◆□▶ ◆□▶ ◆ ≧▶ ◆ ≧▶ ≧ のへ⊙

Schedule

	ESIGMA	1	2	2 3	3	4	5	6	7	8	9	10	2 11	1 12	2 1:	3 1	41	15 [.]	16	17	18	19	20	02	21 2	22	23	24	2	52	62	72	28 2	29 3	30	31	32	33	34	35	36	37	38	39	94	04-	14	24	34	44	15 4	46 -	47	48	Start Month	i Me	ind onth
WP0	Ognanization and dissemination																																																						(1)	(48)
WP1	Global structure analysis																																																					-	(1)	(48)
WP2	Succinct Representation of Graph Data				T																																																		(7)	(42)
WP3	Graph Similarity																																																						(7)	(42)
WP4	Application-driven																																																								
																T																																							(13)	6	48)

Duration: 48 months

<ロ> < 団> < 団> < 三> < 三> < 三> < 三</p>

Schedule

	ESIGMA	1	2	2 3	3	4 ;	5	6	7	8	9	10	11	11	21	31	14	15	16	1	71	81	19:	20	21	22	22	32	24 2	25	26	27	28	82	93	03	31	32	33	34	35	36	37	38	33	940	04	14	24	13 ²	14 4	15	464	47	48	Start Month	Er Mo	nd nth
WP0	Ognanization and dissemination																																																							(1)	(4	8)
WP1	Global structure analysis																																																							(1)	(4	8)
WP2	Succinct Representation of Graph Data																																																							(7)	(4	2)
WP3	Graph Similarity																																																							(7)	(4	2)
WP4	Application-driven																																																									
																																																								(13)	(4	8)

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	1	2	2 3	3	4 ;	5	6	7	8	9	10	11	11	21	31	14	15	16	1	71	81	19:	20	21	22	22	32	24 2	25	26	27	28	82	93	03	31	32	33	34	35	36	37	38	33	940	04	14	24	13 ²	14 4	15	464	47	48	Start Month	Er Mo	nd nth
WP0	Ognanization and dissemination																																																							(1)	(4	8)
WP1	Global structure analysis																																																							(1)	(4	8)
WP2	Succinct Representation of Graph Data																																																							(7)	(4	2)
WP3	Graph Similarity																																																							(7)	(4	2)
WP4	Application-driven																																																									
																																																								(13)	(4	8)

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.

◆□▶ ◆□▶ ◆ 三▶ ◆ 三▶ ○ 三 ○ のへぐ

• 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

▲□▶▲□▶▲□▶▲□▶ □ 少へ⊙

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 €

<□ > < @ > < E > < E > E のQ @

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

◆□ ▶ ◆□ ▶ ◆ 三 ▶ ◆ 三 ● のへで

Merci!

ロ > < 団 > < 三 > < 三 > < 三 < の < ?