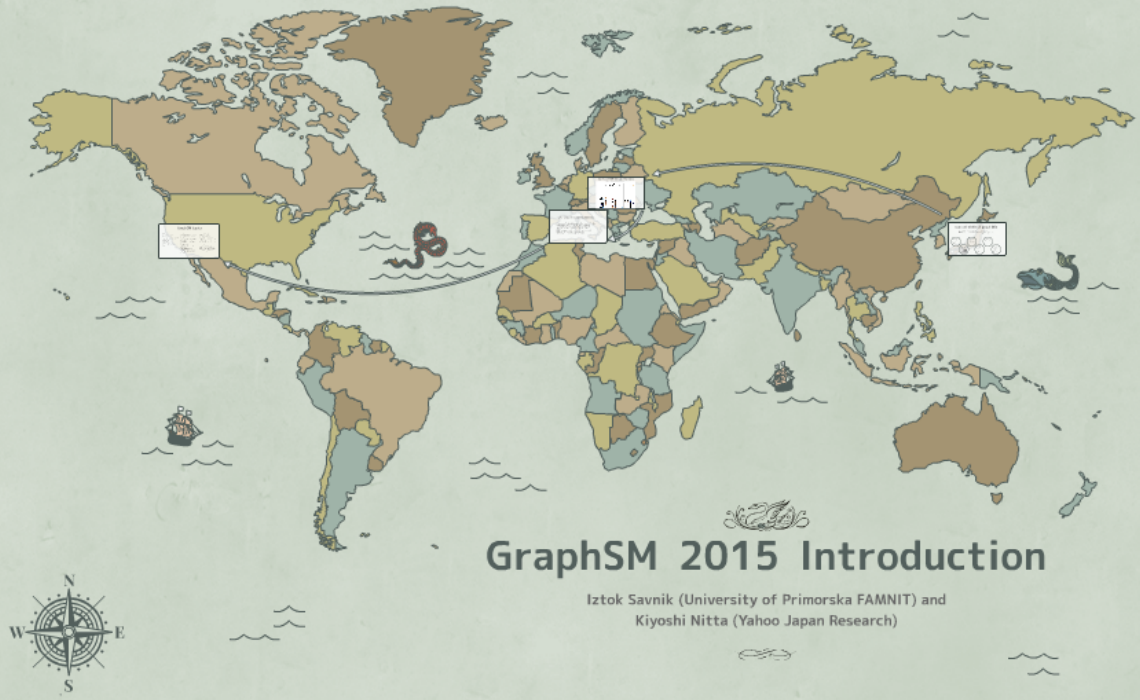
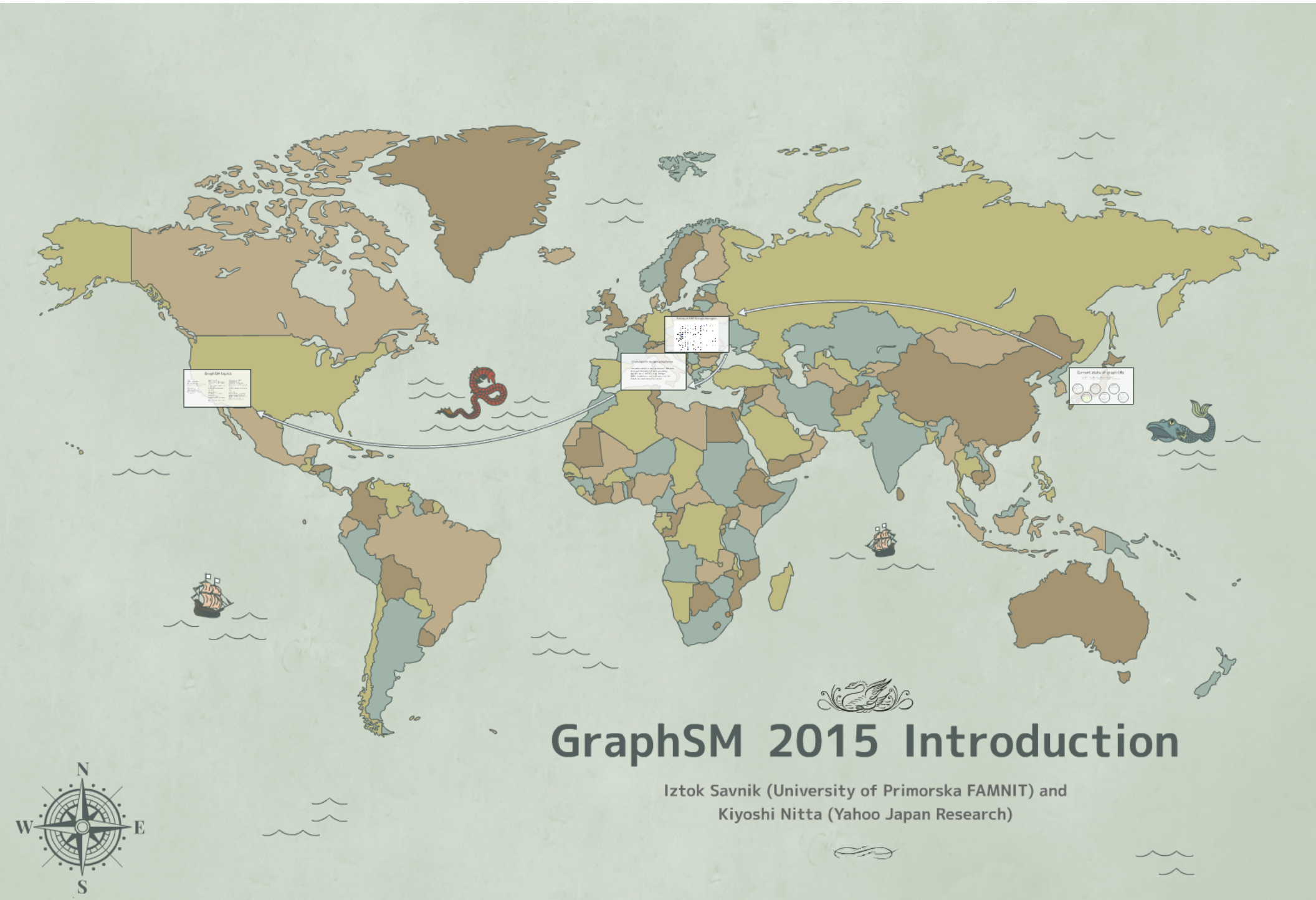


GraphSM 2015 Introduction

Iztok Savnik (University of Primorska FAMNIT) and
Kiyoshi Nitta (Yahoo Japan Research)



GraphSM 2015 Introduction

Iztok Savnik (University of Primorska FAMNIT) and
Kiyoshi Nitta (Yahoo Japan Research)

Current state of graph DBs

- There exist a variety of different dictionaries, properties, concepts, ...
- There exist a variety of formats and models for knowledge and data representation

Wordnet

- Princeton's large lexical database of English.
- Cognitve synonims: synsets = concepts
- 117,000 synsets
- Synsets are linked by:
  - conceptual-semantic relationships, and
  - lexical relationships.
  - Include definitions of synsets.

Freebase

- Free, knowledge graph:
  - people, places and things,
  - 2,478,168,612 facts, 43,459,442 topics

Wikidata

- Free knowledge base with 14,550,852 items
- Properties of person, organization, works, events, etc.

Linked Open Data

- Number of triples: 33 Giga (109) (2011)
- the number of datasets has grown by 271% (2011 -> 2014)

YAGO

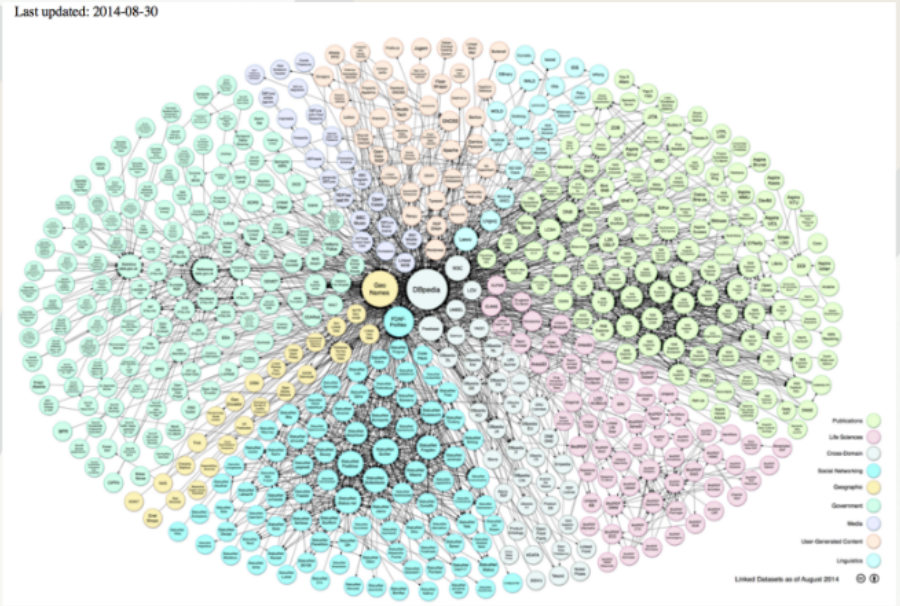
- 10 Mega (106) concepts
  - Max Planc Institute, Informatik
  - Accuracy of 95%
- Includes:
  - Wikipedia, WordNet, GeoNames
  - Links Wordnet to Wikipedia taxonomy (350K concepts)
  - Anchored in time and space

Cyc

- Knowledge base
  - Doug Lenat
  - Conceptual networks (ontologies)
  - Higher ontology, basic theories, specific theories
  - Predefined semantic relationships
- Common sense reasoner
  - Based on predicate calculus
  - Rule-based reasoning

Linked Open Data

- Number of triples: 33 Giga (109) (2011)
- the number of datasets has grown by 271% (2011 -> 2014)



Survey of RDF Storage Managers

TABLE I. PROPERTIES OF RDF STORAGE MANAGERS

	$\mathcal{S}$								$\mathcal{M}$			
	T_s	I_s	Q_s	S_s	J_s	C_s	D_s	F_s	D_m	Q_m	S_m	A_m
3store	v		S	U	R		R	T	n	n	n	
4store	v		S	U	o		R		h	p		n
Virtuoso	v	G	S	Ulo	R		R	TA	n	n	n	
RDF-3X	v	6	S	Ul	o		R		n	n	n	
Hexastore	v	6	o	Ul		n			n	n	n	
Apache Jena	p		S	Ulo	R	m	R		n	n	n	
SW-Store	h			Uo	c	m	c		n	n	n	
BitMat	v	m	S	Ul	p		c				p	
AllegroGraph			S				c		h	p		m
Hadoop/HBase	h				c		c			p		m
H_2RDF+	v	6	S	Ulo	c		c	n		p	n	m
TriAD	v	6	S		p		R		h	p	n	n
TripleBit	p	6	o	U	o		c		n	n	n	
Trinity.RDF	g		S	l	g		c		c	p	n	m
gStore	g	a	S	Ul	g	n	c		n	n	n	
Neo4j	g	a	o	Ulo	g	f	c			p	n	
big3store	v	3	S	n	R	l	R	n	d	r	p	n

gStore	g
Neo4j	g
big3store	v

Challenges in designing big3store

- Automatic distribution and replication of RDF data
- Intelligent distribution of query processing
- Dynamic updates in RDF storage manager
- Multi-threaded architecture of query executor
- Distributed cache for query executor

GraphSM topics

THEORY
Search in graph databases
Algebra and logic of graphs
Expressive power of graph query languages
Formalizations of graph databases

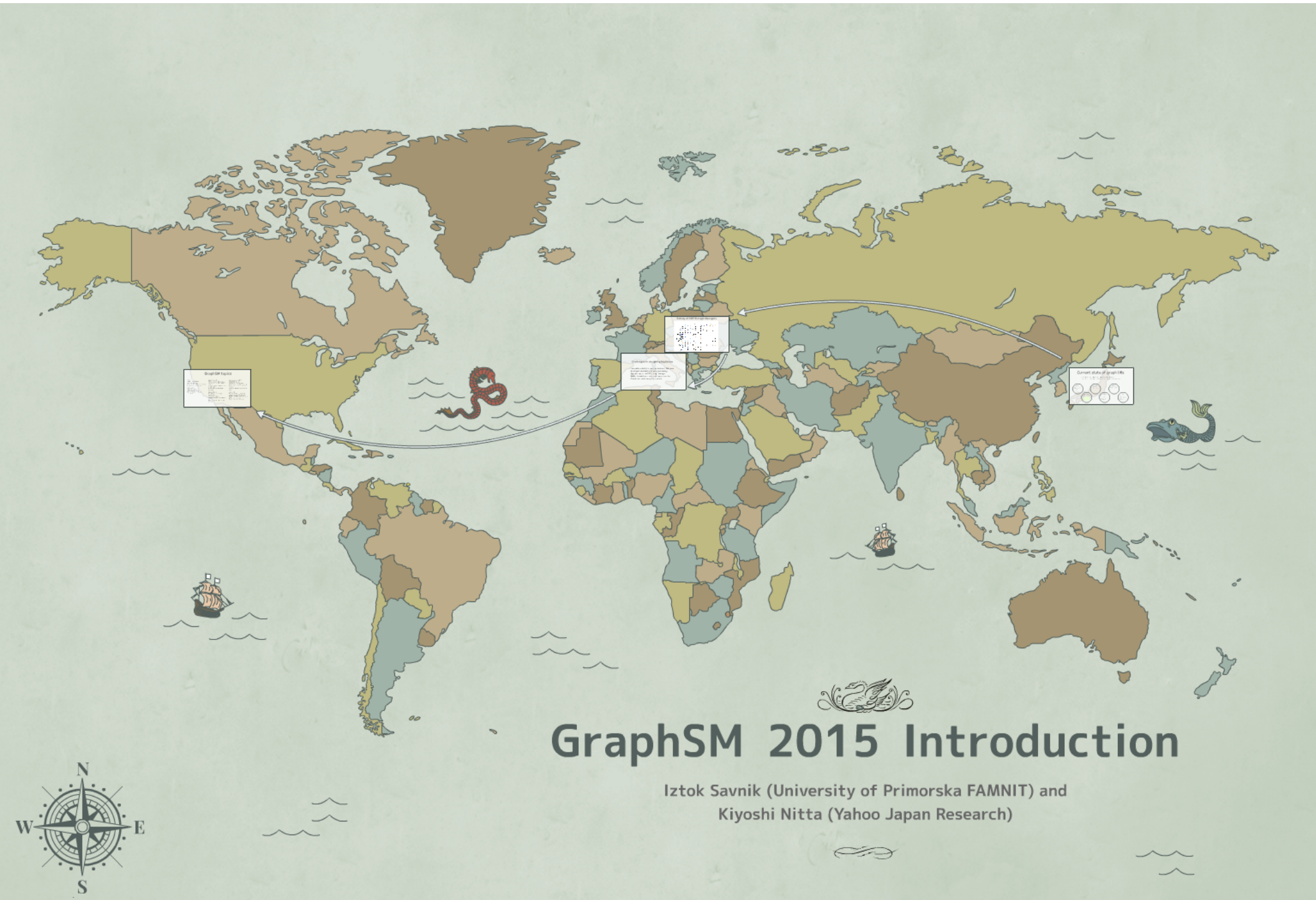
DATA MODELLING
Graph data modelling
Advanced graph data models
Data modelling for specific graph applications

STORAGE MANAGERS
Indexing methods for graph processing
Storage systems for large-scale graph databases
Automatic distribution and replication of graph databases
Storage managers for specific graph applications

QUERY PROCESSING
Flexible query answering on graph-structured data
SPARQL query processing
Intelligent distribution of SPARQL query processing
Using map-reduce operations for graph processing

ANALYSIS AND MANAGEMENT
Graph pattern matching
Knowledge discovery from graphs
Algorithms for graph database processing
Analysis of graph databases from specific domains
Information retrieval on graph-structured data

APPLICATIONS
Biological and medical graph databases
Graph processing for Social Networks
Visualizing, browsing and navigating graph data
User-interfaces for graph databases



GraphSM 2015 Introduction

Iztok Savnik (University of Primorska FAMNIT) and
Kiyoshi Nitta (Yahoo Japan Research)