

From Graph to Knowledge Graph

Mining Large-scale Heterogeneous Networks Using Spark

KDD 2019 Anchorage

Aug 8th, 2019

Microsoft Research – Microsoft Academic Graph team

Module I: Welcome and Intro

9:30 am - 10:30 am

- What would be covered
- Who are the targeted audiences
- Who are we

WHAT & WHY

Morning (9:30am - 12:30pm) **Graph**

- *Module 1*: Welcome and intro (environment setup + dataset)
- *Module 2*: Graph basics
- *Module 3*: Graph representation learning

----Lunch break----

Afternoon (1:00pm – 3:50pm) **Knowledge Graph (KG)**

- *Module 4*: KG fundamentals and construction
- *Module 5*: KG inference and application
- *Module 6*: Summary and looking forward

Tutorial Overview

More comprehensive algorithms and theories, see our edX course:

<https://www.edx.org/course/from-graph-to-knowledge-graph-algorithms-and-applications>

Basics

- Spark and Databricks
- Graph and knowledge graph (KG)

Environment Setup

- Create free Azure account
- Setup Azure Databricks
- Get the data

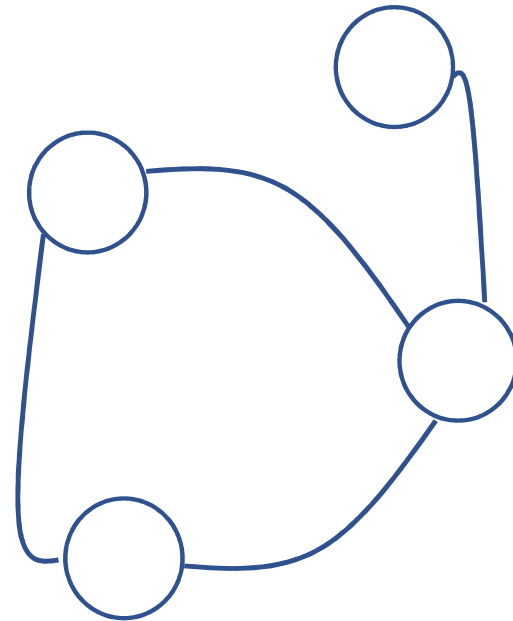
Labs – dataset understanding

- *Microsoft Academic Graph (MAG)* schema
- Top CS conference sub-graph
- Sub-graph entity and relation stats

Module I Overview

basics + environment setup + dataset

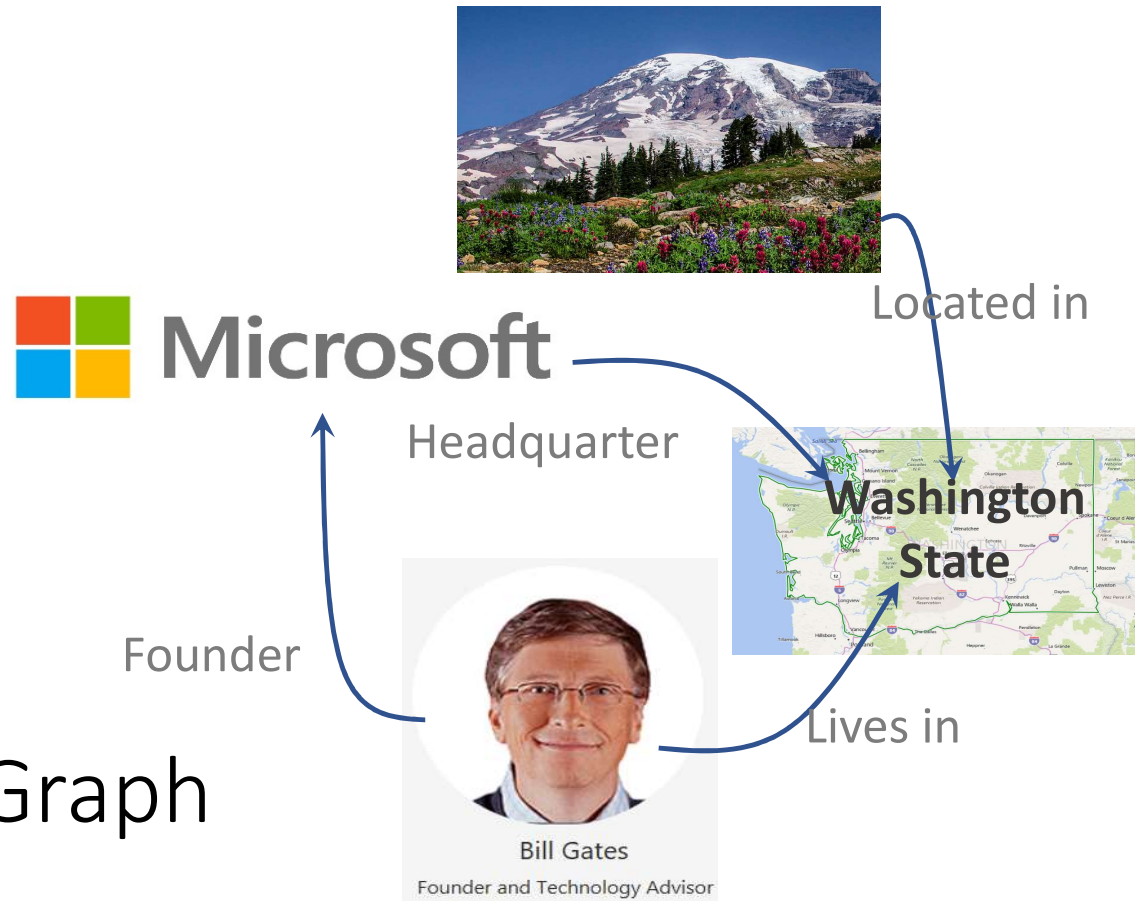
- Node
- Edge
- Structure



Basics – Graph

- Node - entity
- Edge - relation
- Structure - semantics

Basics – Knowledge Graph



- Spark
- Databricks

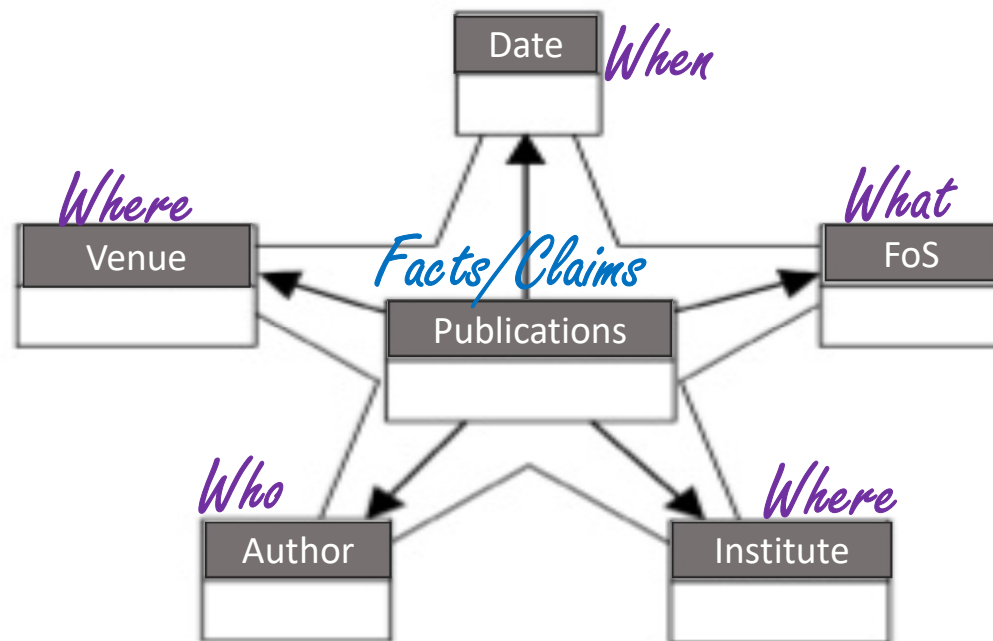
Basics – Spark and databricks

- Create Free Azure Account
- Setup Azure Databricks
- Get the data
 - Full set (**free**), see instruction below:
<https://docs.microsoft.com/en-us/academic-services/graph/get-started-setup-provisioning>
 - Tutorial sub-graph
 - Azure storage key or SHS

Environment Setup

- MAG Schema
- Top CS conference sub-graph
- Understand MAG entities and relations

Lab 1: Get MAG & basic Stats



	223,145,529 Papers
	241,823,525 Authors
	664,719 Topics
	4,397 Conferences
	48,757 Journals
	25,546 Institutions

MAG Schema

Full schema documentation: <https://docs.microsoft.com/en-us/academic-services/graph/reference-data-schema>

- ▶ 103 selected top tier CS conferences
- ▶ Direct linked Papers / Authors / Affiliations
- ▶ All fields-of-study and the taxonomy

MAG CS TOP CONFERENCES SUBGRAPH