Graphs in Machine Learning

Spectral Clustering: Background

Motivation and Problem Setting

Michal Valko

Inria & ENS Paris-Saclay, MVA
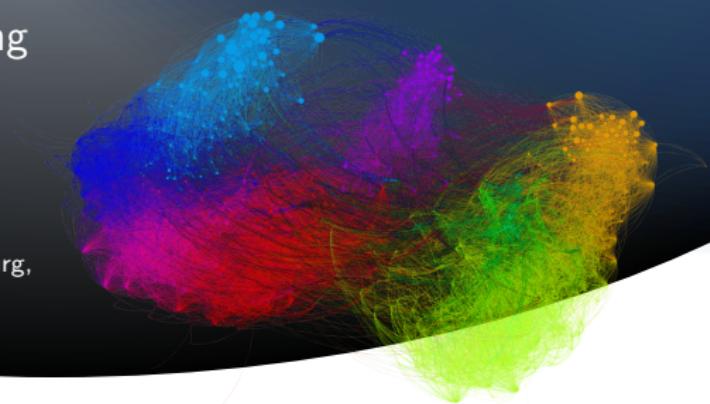
How to rule the world: "AI" is here
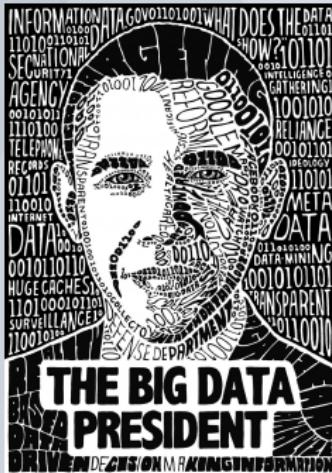
Application of Graphs for ML: Clustering
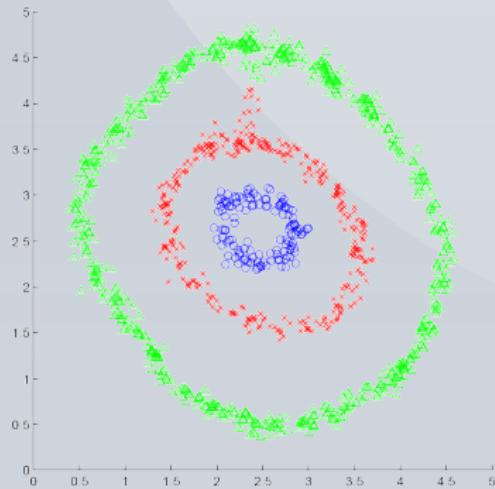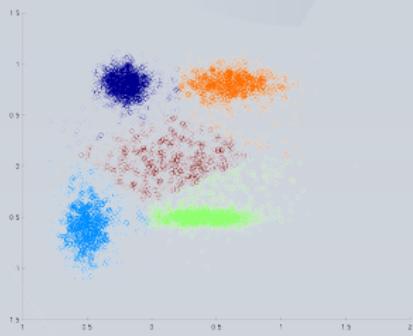
Background: Clustering

- What do we know about the **clustering** in general?

Background: Clustering

- What do we know about the **clustering** in general?
  – ill defined problem (different tasks → different paradigms)

Background: Clustering

- What do we know about the **clustering** in general?
  - ill defined problem (different tasks $\rightarrow$ different paradigms)
  - "I know it when I see it"

Background: Clustering

- What do we know about the **clustering** in general?
  - ill defined problem (different tasks $\rightarrow$ different paradigms)
  - "I know it when I see it"
  - inconsistent (wrt. Kleinberg's axioms)
    - ▶ scale-invariance, richness, consistency

Background: Clustering

- What do we know about the **clustering** in general?
    - ill defined problem (different tasks $\rightarrow$ different paradigms)
    - "I know it when I see it"
    - inconsistent (wrt. Kleinberg's axioms)
        - scale-invariance, richness, consistency
    - number of clusters k need often be known

Background: Clustering

- What do we know about the **clustering** in general?
    - ill defined problem (different tasks $\rightarrow$ different paradigms)
    - "I know it when I see it"
    - inconsistent (wrt. Kleinberg's axioms)
        - scale-invariance, richness, consistency
    - number of clusters k need often be known
    - difficult to evaluate

Background: Clustering

- What do we know about the **clustering** in general?
  - ill defined problem (different tasks $\rightarrow$ different paradigms)
  - "I know it when I see it"
  - inconsistent (wrt. Kleinberg's axioms)
    - scale-invariance, richness, consistency
  - number of clusters k need often be known
  - difficult to evaluate

- What do we know about k-**means**?

Background: Clustering

- What do we know about the **clustering** in general?
    - ill defined problem (different tasks $\rightarrow$ different paradigms)
    - "I know it when I see it"
    - inconsistent (wrt. Kleinberg's axioms)
        - scale-invariance, richness, consistency
    - number of clusters k need often be known
    - difficult to evaluate

- What do we know about k-**means**?
    - "hard" version of EM clustering

Background: Clustering

- What do we know about the **clustering** in general?
    - ill defined problem (different tasks $\rightarrow$ different paradigms)
    - "I know it when I see it"
    - inconsistent (wrt. Kleinberg's axioms)
        - scale-invariance, richness, consistency
    - number of clusters k need often be known
    - difficult to evaluate

- What do we know about k-**means**?
    - "hard" version of EM clustering
    - sensitive to initialization

Background: Clustering

- What do we know about the **clustering** in general?
    - ill defined problem (different tasks $\rightarrow$ different paradigms)
    - "I know it when I see it"
    - inconsistent (wrt. Kleinberg's axioms)
        - ▶ scale-invariance, richness, consistency
    - number of clusters k need often be known
    - difficult to evaluate

- What do we know about k-**means**?
    - "hard" version of EM clustering
    - sensitive to initialization
    - optimizes for **compactness**

Background: Clustering

- What do we know about the **clustering** in general?
    - ill defined problem (different tasks $\rightarrow$ different paradigms)
    - "I know it when I see it"
    - inconsistent (wrt. Kleinberg's axioms)
        - scale-invariance, richness, consistency
    - number of clusters k need often be known
    - difficult to evaluate

- What do we know about k-**means**?
    - "hard" version of EM clustering
    - sensitive to initialization
    - optimizes for **compactness**
    - yet: algorithm-to-go

Michal Valko

michal.valko@inria.fr

Inria & ENS Paris-Saclay, MVA

https://misovalko.github.io/mva-ml-graphs.html