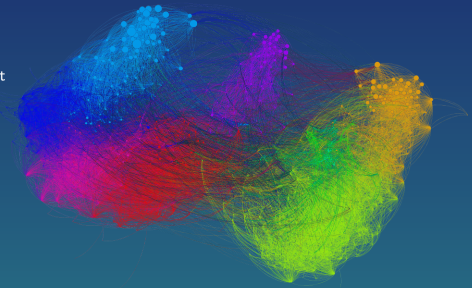
Graphs in Machine Learning

Michal Valko

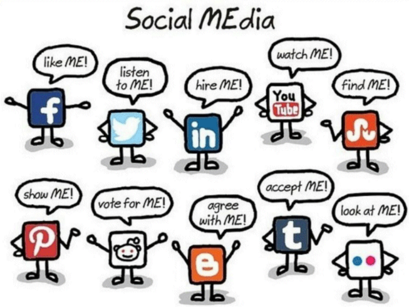
DeepMind Paris and Inria Lille

TA: Omar Darwiche Domingues with the help of Pierre Perrault



Graphs from social networks

- people and their interactions
- directed (Twitter) and undirected (Facebook)
- structure is rather a *phenomena*
- typical ML tasks
  - advertising
  - product placement
  - link prediction (PYMK)



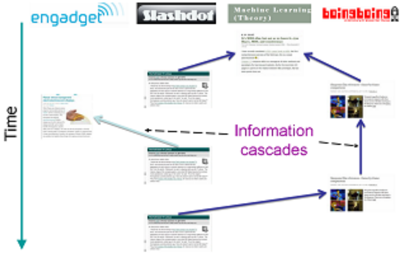
Graphs from utility and technology networks

- link services
- power grids, roads, Internet, sensor networks
- structure is either *hand designed* or not
- typical ML tasks
  - best routing under unknown or variable costs
  - identify the node of interest

Berkeley's Floating Sensor Network

Graphs from information networks

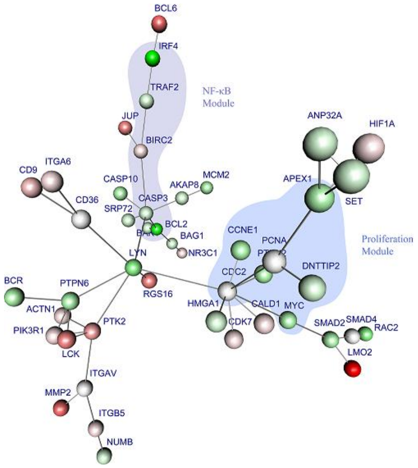
- web
- blogs
- wikipedia
- typical ML tasks
  - find influential sources
  - search (pagerank)



Blog cascades (ETH) - *submodularity*

Graphs from biological networks

- protein-protein interactions
- gene regulatory networks
- typical ML tasks
  - discover unexplored interactions
  - learn or reconstruct the structure



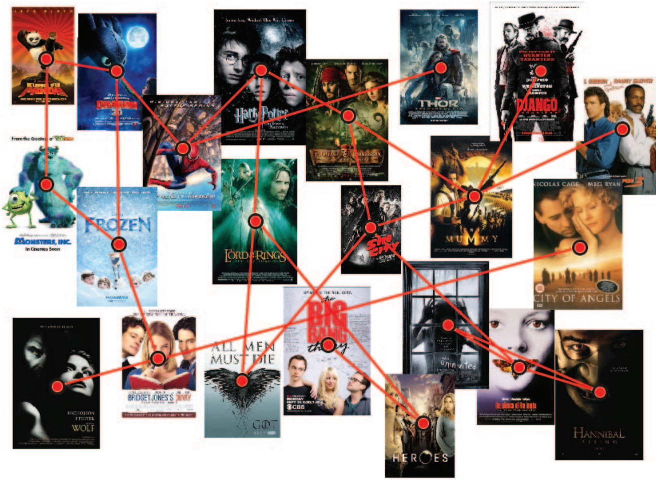
Diffuse large B-cell lymphomas - Dittrich et al. (2008)

Graphs from similarity networks

graph is not naturally given

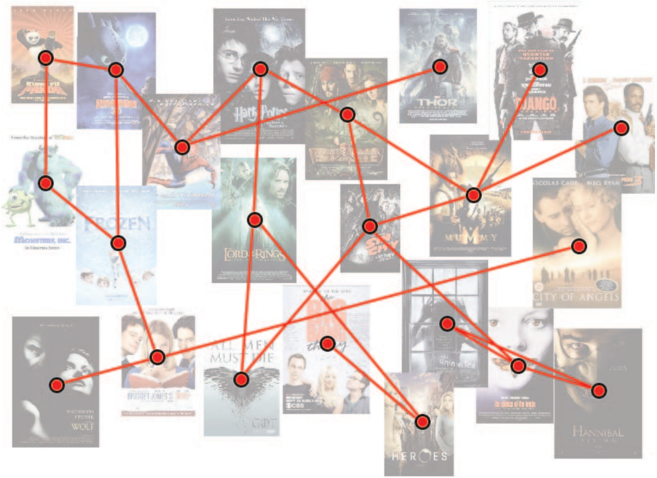
Graphs from similarity networks

but we can construct it



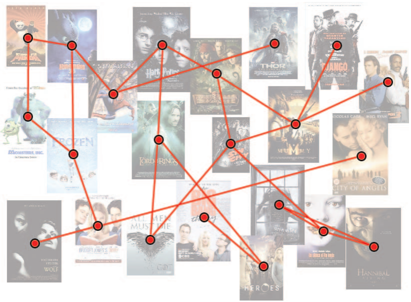
Graphs from similarity networks

and use it as an abstraction



Graphs from similarity networks

- vision
- audio
- text
- typical ML tasks
  - semi-supervised learning
  - spectral clustering
  - manifold learning



Movie similarity

Two sources of graphs in ML

Graph as models for networks

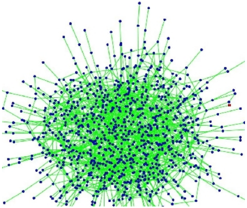
- given as an input
- discover interesting properties of the structure
- represent useful information (viral marketing)
- is the object of study (anomaly detection)

Graph as nonparametric basis

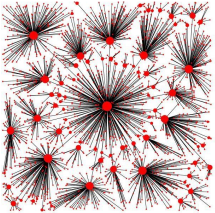
- we create (learn) the structure (it's a tool)
- flat vectorial data $\rightarrow$ similarity graph
- nonparametric regularizer
- encode structural properties: smoothness, independence, ...

Random Graph Models

Erdős-Rényi
independent edges



Barabási-Albert
preferential attachment



Stochastic Blocks
modeling communities

Watts-Strogatz, Chung-Lu, Fiedler, ...

Deep learning :O

Why not deep?

Why not!

Graph neural networks *are so 2019!*

What for: Physics prediction and everything else!

New addition to the curriculum: including TD

What will you learn in the Graphs in ML course?

Concepts, tools, and methods to work with graphs in ML.

Theoretical toolbox to analyze graph-based algorithms.

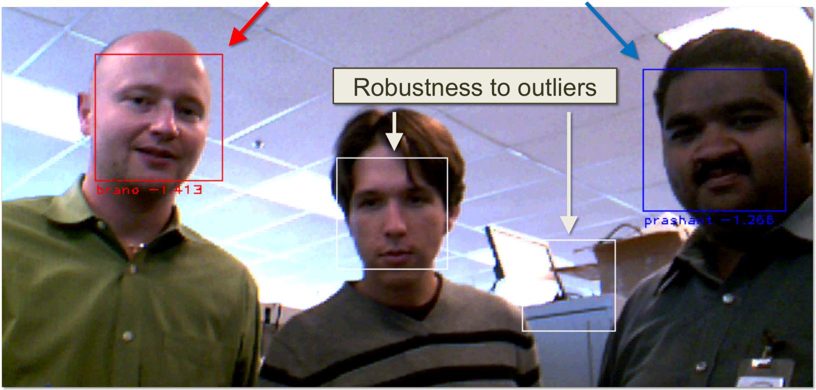
Specific applications of graphs in ML.

How to tackle: *large graphs, online setting, graph construction* ...

One example: Online Semi-Supervised Face Recognition

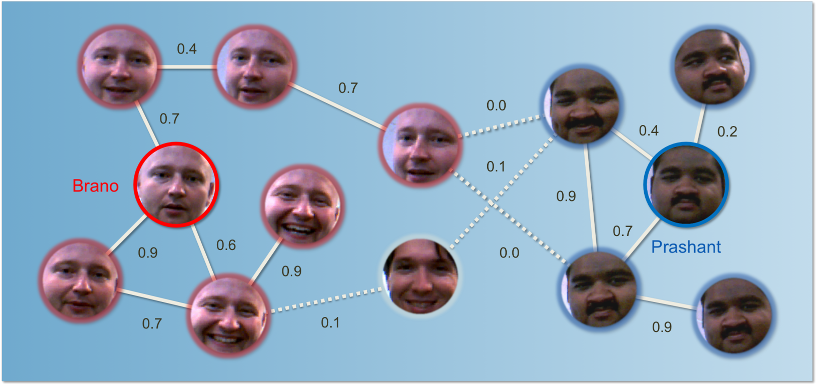
Online Semi-Supervised Face Recognition

graph is not given



Online Semi-Supervised Face Recognition

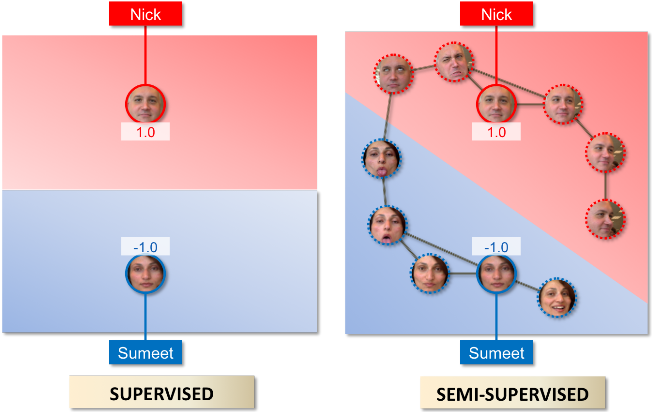
we will construct it!



An example of a similarity graph over faces. The faces are vertices of the graph. The edges of the graph connect similar faces. Labeled faces are outlined by thick solid lines.

Online Semi-Supervised Face Recognition

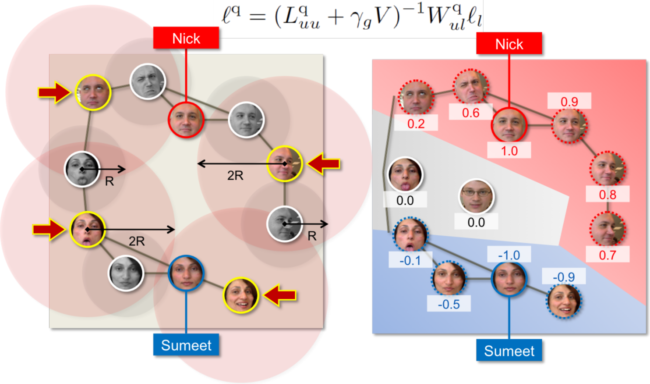
graph-based semi-supervised learning



Online Semi-Supervised Face Recognition

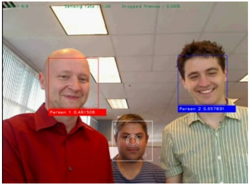
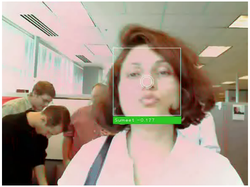
online learning - graph sparsification

$$\ell^{\mathrm{q}} = (L^{\mathrm{q}}_{uu} + \gamma_g V)^{-1} W^{\mathrm{q}}_{ul} \ell_l$$



DEMO

second TD

see the demo: `http://researchers.lille.inria.fr/~valko/hp/serve.php?what=publications/kveton2009nipsdemo.officespace.mov`

OSS FaceReco: Analysis

$$\frac{1}{n}\sum_t (\ell_t^{\mathrm{q}}[t]-y_t)^2 \le \frac{3}{n}\sum_t (\ell_t^*-y_t)^2 + \frac{3}{n}\sum_t (\ell_t^{\mathrm{o}}[t]-\ell_t^*)^2 + \frac{3}{n}\sum_t (\ell_t^{\mathrm{q}}[t]-\ell_t^{\mathrm{o}}[t])^2$$

- Error of our solution
- Offline learning error
- Online learning error
- Quantization error

Claim: When the regularization parameter is set as $\gamma_g = \Omega(n_l^{3/2})$, the difference between the risks on labeled and all vertices decreases at the rate of $O(n_l^{-1/2})$ (with a high probability)

$$\frac{1}{n}\sum_t (\ell_t^*-y_t)^2 \le \frac{1}{n_l}\sum_{i\in l} (\ell_i^*-y_i)^2 + \beta + \sqrt{\frac{2\ln(2/\delta)}{n_l}}(n_l\beta+4)$$

$$\beta \le \left[\frac{\sqrt{2}}{\gamma_g+1} + \sqrt{2n_l}\,\frac{1-\sqrt{c_u}}{\sqrt{c_u}}\,\frac{\lambda_M(L)+\gamma_g}{\gamma_g^2+1}\right]$$

OSS FaceReco: Analysis

$$\frac{1}{n}\sum_t (\ell_t^{q}[t]-y_t)^2 \le \frac{3}{n}\sum_t (\ell_t^{*}-y_t)^2 + \frac{3}{n}\sum_t (\ell_t^{o}[t]-\ell_t^{*})^2 + \frac{3}{n}\sum_t (\ell_t^{q}[t]-\ell_t^{o}[t])^2$$

Error of our solution	Offline learning error	Online learning error	Quantization error

Claim: When the regularization parameter is set as $\gamma_g = \Omega(n^{1/4})$, the average error between the offline and online HFS predictions decreases at the rate of $O(n^{-1/2})$

$$\frac{1}{n}\sum_t (\ell_t^{o}[t]-\ell_t^{*})^2 \le \frac{1}{n}\sum_t \left\|\ell^{o}[t]-\ell^{*}\right\|_2^2 \le \frac{4n_l}{(\gamma_g+1)^2}$$

$$\|\ell\|_2 \le \frac{\|y\|_2}{\lambda_m(C^{-1}K+I)} = \frac{\|y\|_2}{\lambda_m(K)\lambda_M^{-1}(C)+1} \le \frac{\sqrt{n_l}}{\gamma_g+1}$$

OSS FaceReco: Analysis

$$\frac{1}{n}\sum_t (\ell_t^{\mathrm{q}}[t]-y_t)^2 \le \frac{3}{n}\sum_t (\ell_t^* - y_t)^2 + \frac{3}{n}\sum_t (\ell_t^{\mathrm{o}}[t]-\ell_t^*)^2 + \frac{3}{n}\sum_t (\ell_t^{\mathrm{q}}[t]-\ell_t^{\mathrm{o}}[t])^2$$

Error of our solution	Offline learning error	Online learning error	Quantization error

Claim: When the regularization parameter is set as $\gamma_g = \Omega(n^{1/8})$, and the Laplacians L^{q} and L^{o} and normalized, the average error between the online and online quantized HFS predictions decreases at the rate of $O(n^{-1/2})$

$$\frac{1}{n}\sum_t (\ell_t^{\mathrm{q}}[t]-\ell_t^{\mathrm{o}}[t])^2 \le \frac{1}{n}\sum_t \left\|\ell^{\mathrm{q}}[t]-\ell^{\mathrm{o}}[t]\right\|_2^2 \le \frac{n_l}{c_u^2\gamma_g^4}\left\|L^{\mathrm{q}}-L^{\mathrm{o}}\right\|_F^2$$

$$\left\|L^{\mathrm{q}}-L^{\mathrm{o}}\right\|_F^2 \propto O(k^{-2/d})$$

The distortion rate of online k-center clustering is $O(k^{-1/d})$, where d is dimension of the manifold and k is the number of representative vertices

Some of the other topics

- spectral graph theory, graph Laplacians, spectral clustering
- semi-supervised learning and manifold learning
- learnability on graphs - transductive learning
- online decision-making on graphs, graph bandits
- submodularity on graphs
- real-world graphs scalability and approximations
- spectral sparsification
- social network and recommender systems applications
- link prediction/link clasification
- signed networks (eOpinions)
- generalization bounds by perturbation analysis

MVA and Graphs: 2 courses

The two MVA graph courses offer complementary material.

1: **Graphs in ML**

this class

- focus on learning
- spectral clustering
- random walks
- graph Laplacian
- semi-supervised learning
- manifold learning
- theoretical analyses
- online learning
- recommender systems

2: **ALTeGraD**

by Michalis Vazirgiannis

- dimensionality reduction
- feature selection
- text mining
- graph mining
- community mining
- graph generators
- graph-evaluation measures
- privacy in graph mining
- big data

Administrivia

Time: Tuesday afternoons, next week at 13:30

Place: ENS Cachan, next week at Salle Condorcet

7 lectures + 3 recitations (TDs)

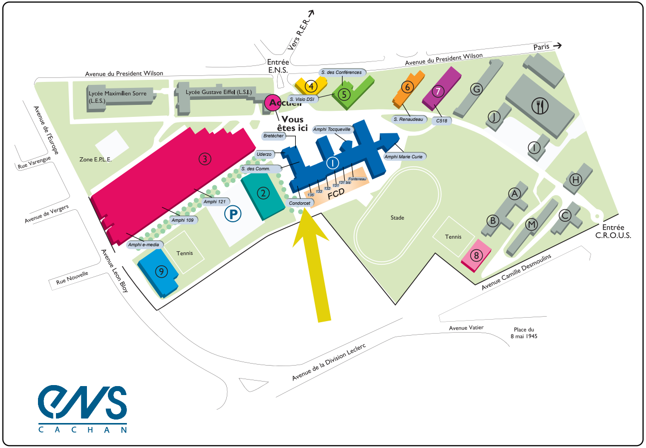
Validation: grades from TDs (40%) + class project (60%)

Research: contact me for *internships*, *Ph.D. theses*, *projects*, etc.

Course website:
`http://researchers.lille.inria.fr/~valko/hp/mva-ml-graphs`

Contact, online class discussions, and announcements:
`https://piazza.com/ens_cachan/fall2019/mvagraphsml`
class code given during the class

First class on Tuesday, October 1st at 13:30!



Michal Valko
contact via Piazza