The Most Persistent Soft-Clique in a Set of Sampled Graphs

Novi Quadrianto

Joint work with Chao Chen and Christoph Lampert
Cambridge Machine Learning Group
ICML Edinburgh, 29th June 2012

Motivation

- Setting: Many people (including the Simpsons) are visiting a theme park during the weekends;
- Question: Can we identify the Simpsons?

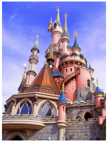
time $t = 0$: at entrance

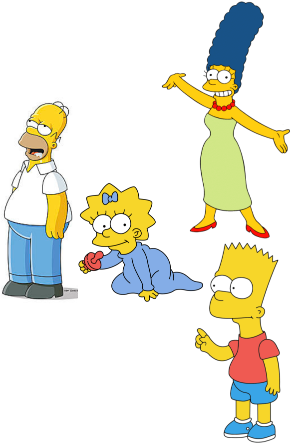
Motivation

- Setting: Many people (including the Simpsons) are visiting a theme park during the weekends;
- Question: Can we identify the Simpsons?

time $t = 0$: at entrance

Motivation

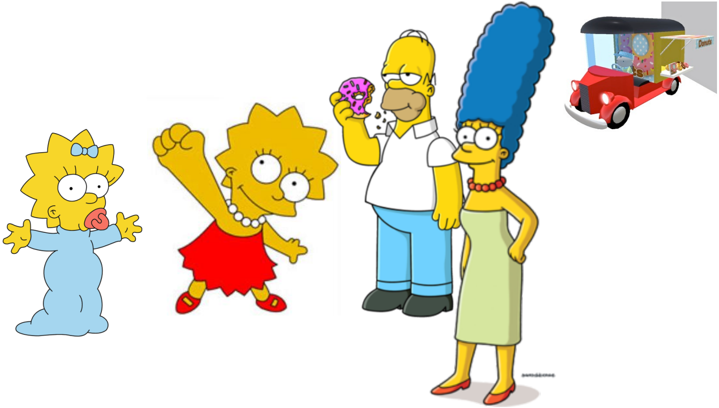




Fox

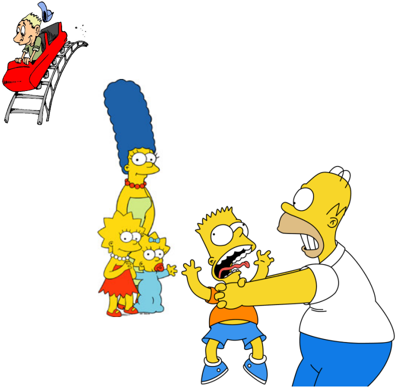
time t $=$ 1: at sleeping beauty castle

Motivation



time $t = 2$: at donuts stand

Motivation



time t = 3: at roller coaster

Motivation

time t = 4: at rocky pool

Morale of the Story

- Dense subgraphs that appear in all of the samples of the graphs are the groups of friends or families that we would like to identify;

- However, in each individual observation of the graphs not every person will be observed within the subgraph: he or she could have left the group temporarily due to other excitements.

Regularized Risk Functionals

- Dreams:
  We want to find a subset of vertices, that 1) is almost fully or at least densely connected, 2) occurs in all or almost all graph instances, and 3) has the maximum weight;
- Turning Dreams into a Regularized Risk Functional:

$$\min_{x \in \{0,1\}^{|\mathcal{V}|}} \min_{\beta \in \mathbb{R}_+^T} \underbrace{\eta \|\beta\|_{L_p}^p}_{\text{The Cost of Losing Dreams}} - \underbrace{\sum_{1 \le i < j \le n} x_i x_j k(v_i, v_j)}_{\text{The Regularizer}}$$

$$\text{subject to} \quad \underbrace{\sum_{1 \le i < j \le n} x_i x_j \mathbb{I}[(i,j) \notin \mathcal{E}_t] \le \beta_t \quad \forall t = 1, \ldots, T\,.}_{\text{The Cost of Losing Dreams}}$$

Regularized Risk Functionals

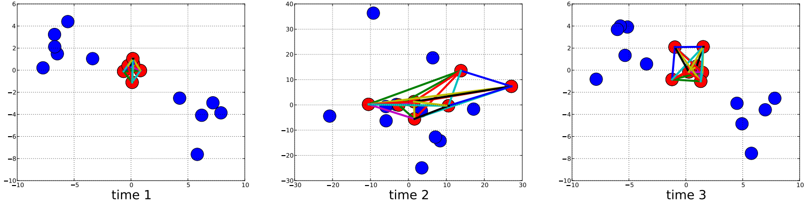
- Dreams:
  We want to find a subset of vertices, that 1) is almost fully or at least densely connected, 2) occurs in all or almost all graph instances, and 3) has the maximum weight;
- Turning Dreams into a Regularized Risk Functional:

$$\min_{x \in \{0,1\}^{|\mathcal{V}|}} \min_{\beta \in \mathbb{R}_+^T} \underbrace{\eta \|\beta\|_{L_p}^p}_{\text{The Cost of Losing Dreams}} - \underbrace{\sum_{1 \le i < j \le n} x_i x_j k(v_i, v_j)}_{\text{The Regularizer}}$$

subject to $$\underbrace{\sum_{1 \le i < j \le n} x_i x_j \mathbb{I}[(i,j) \notin \mathcal{E}_t] \le \beta_t \quad \forall t = 1, \ldots, T\,.}_{\text{The Cost of Losing Dreams}}$$

Results

Data: at time 1 are drawn from a Gaussian mixture with 3 components. At time 2 and 3, the data are corrupted with a random Gaussian noise.



Thank You

For more information, please refer to my poster:
#24 at Informatics Forum.