See the Tree through the Lines:

The Shazoo Algorithm

F.Vitale* N.Cesa-Bianchi* C.Gentile° G.Zappella*

* → Università degli studi di Milano, Italy ° → Università dell'Insubria, Italy

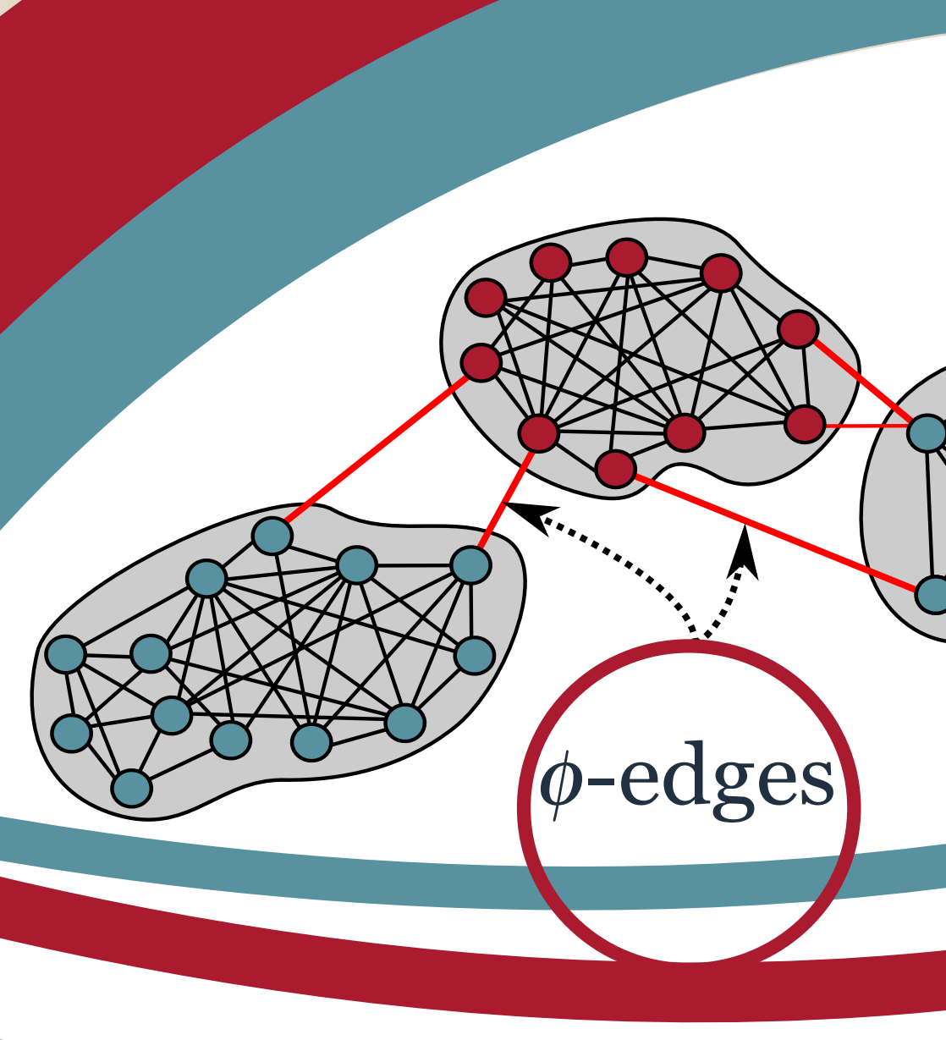
T82

Our learning problem: Node classification of weighted and undirected graphs based only on the network topology. This is useful in several domains, including document networks, social networks, biological networks.

Protocol: online learning. Vertices are issued one by one in an arbitrary (possibly adversarial) order $v_1, v_2, ..., v_n$. At each time t: the learner predicts the binary label of v_t and then observes its true value.

Transductive classification: the entire unlabeled graph is known in advance.

Goals: few prediction mistakes and scalability



small weigthed cutsize

Bias

strongly connected nodes have the same label

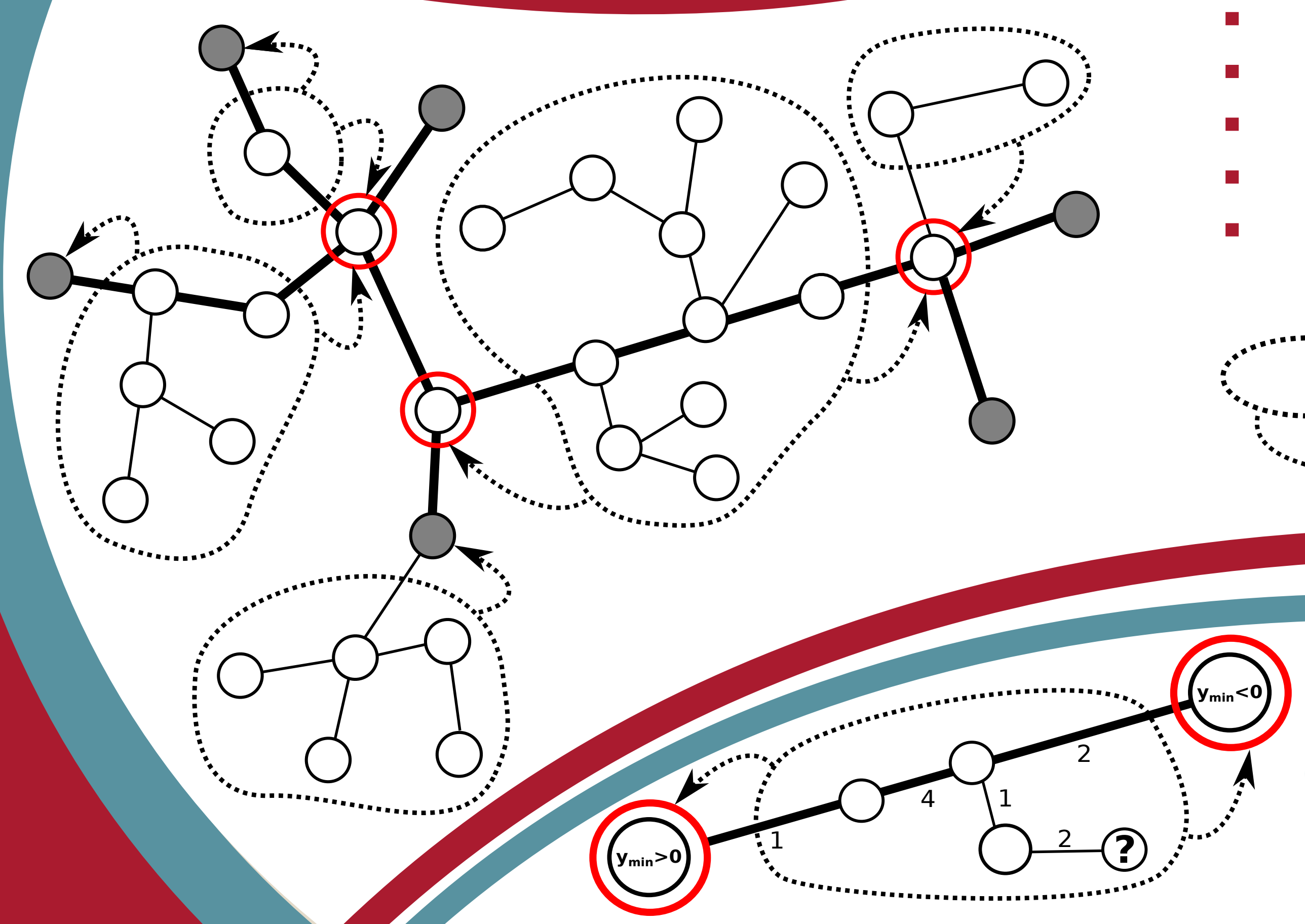
● **Labeled nodes**

○ **Fork nodes**: connected to labeled nodes via at least three edge-disjoint paths

A fork node gets assigned the label y_{min}, minimizing the weighted cutsize

Hinge trees

○ **Connection nodes**



The labels of hinge tree nodes are set using nearest neighbor

Predicted label: -1

Resistance distance on a tree: sum up the inverse of the weigths of the edges

(1/2+1)+1/2 < (1/2+1)+1/4+1

Algo

OMV: majority vote over the labeled neighbors

WTA: nearest neighbor on a linearized tree (see [1]), it requires amortized constant time per prediction

LABPROP: Label Propagation is an iterative algorithm that solves a sparse linear system in time O(|E|x|V|) (see [2])

"n*algo" means a committee of n istances of 'algo', where predictions are combined with a majority vote

Data

USPS: handwritten characters (9,298 nodes)

RCV1: a subset of 10,000 news from RCV1 (10,000 nodes)

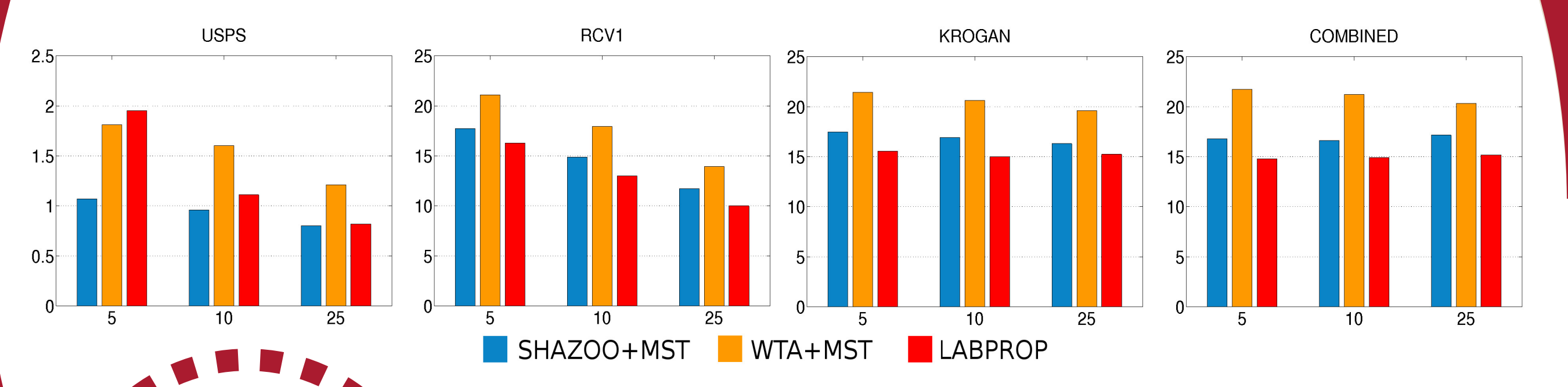
KROGAN: biological graph (2,169 nodes)

COMBINED: biological graph (2,871 nodes)

WEBSPAM: hosts graph created for Webspam Challenge 2008 (110,900 nodes)

USPS and RCV1 graphs have been created using k-NN (k=10) and Gaussian weighting of the edges

Results



Datasets / Predictors	USPS 5%	USPS 10%	USPS 25%	RCV1 5%	RCV1 10%	RCV1 25%	KROGAN 5%	KROGAN 10%	KROGAN 25%	COMBINED 5%	COMBINED 10%	COMBINED 25%
SHAZOO+RST	3.62	2.82	2.02	21.72	18.70	15.68	18.11	17.68	17.10	17.77	17.24	17.34
SHAZOO+NWRST	3.88	3.03	2.18	21.97	19.21	15.95	18.11	18.14	17.32	17.22	17.21	17.53
SHAZOO+MST	1.07	0.96	0.80	17.71	14.87	11.73	17.46	16.92	16.30	16.79	16.64	17.15
WTA+RST	5.34	4.23	3.02	25.53	22.66	19.05	21.82	21.05	20.08	21.76	21.38	20.26
WTA+NWRST	5.74	4.45	3.26	25.50	22.70	19.24	21.90	21.28	20.18	21.58	21.42	20.64
WTA+MST	1.81	1.60	1.21	21.07	17.94	13.92	21.41	20.63	19.61	21.74	21.20	20.32
7*SHAZOO+RST	1.68	1.28	0.97	16.33	13.52	11.07	15.54	15.58	15.46	15.12	15.24	15.84
7*SHAZOO+NWRST	1.89	1.38	1.06	16.49	13.98	11.37	15.61	15.62	15.50	15.02	15.12	15.80
7*WTA+RST	2.10	1.56	1.14	17.44	14.74	12.15	16.75	16.64	15.88	16.42	16.09	15.72
7*WTA+NWRST	2.33	1.73	1.24	17.69	15.18	12.53	16.71	16.60	16.00	16.24	16.13	15.79
11*SHAZOO+RST	1.52	1.17	0.89	15.82	13.04	10.59	15.36	15.40	15.29	14.91	15.06	15.61
11*SHAZOO+NWRST	1.70	1.27	0.98	15.95	13.42	10.93	15.40	15.33	15.32	14.87	14.99	15.67
11*WTA+RST	1.84	1.36	1.01	16.40	13.95	11.42	16.20	16.15	15.53	15.90	15.58	15.30
11*WTA+NWRST	2.04	1.51	1.12	16.70	14.28	11.68	16.22	16.05	15.50	15.74	15.57	15.33
OMV	24.79	12.34	2.10	31.65	22.35	11.79	43.13	38.75	29.84	44.72	40.86	33.24
LABPROP	1.95	1.11	0.82	16.28	12.99	10.00	15.56	14.98	15.23	14.79	14.93	15.18

Webspam (F-Measure)

SHAZOO	WTA	OMV	LABPROP	3*WTA	3*SHAZOO	7*WTA	7*SHAZOO
0.954	0.947	0.706	0.931	0.967	0.964	0.968	0.968

Experiments

We ran our experiments in **batch mode**, using differents sizes of randomly selected training sets (5%, 10%, 25%).

The results are **averaged over 10 runs** for each combination of train set size and algorithm.

We tested WTA and SHAZOO on **different trees**:
- **MST**: the minimum spanning tree, generated in time O(|E| log|V|).
- **RST**: random spanning tree, generated in time O(|E|) for most weighted graphs.
- **NWRST**: random spanning tree of the unweighted graph, generated in time O(|V|) for most graphs.

Batch case: **Shazoo+NWRST** takes constant time per prediction on most graphs.

Shazoo is suitable to large scale networks

The Algorithm

Shazoo operates on **weighted trees** with a combination of **mincut** and **nearest neighbor** strategies: mincut is used to assign the label y_{min} to forks; nearest neighbor with resistance distance is used to label all nodes of a hinge tree.

Multiclass: in this paper we focus on binary classification, but it is possible to extend Shazoo to multiclass just by adding a quasilinear factor to its time complexity.

Pseudocode

```
for t=1...|V|

v(t) is next node to be predicted;

if (v(t) is a fork)
   predict with $y_{min}(v(t))$;
else
   $c_{v(t)}$ <- connection node
               closest to v(t);
   predict with $y_{min}(c_{v(t)})$;
```

Performances

Accuracy: **Optimal mistake bound** (up to log factors)

Time complexity:
- On-line setting: the worst case time per prediction is O(|V|) (rarely encountered in practice)
- Batch setting: the worst case time for predicting all labels of the test set is O(|V|)

Space complexity: O(|V|) if the input is a spanning tree, O(|E|) otherwise

Results

1. **SHAZOO outperforms WTA** irrespective of the type of spanning tree being used
2. The predictive performance of SHAZOO+MST is comparable to, and sometimes better than, that of LABPROP (which is slower)
3. Committees of SHAZOO are effective: **they outperform LABPROP, when the training set is small**
4. **NWRST is extremely fast** to generate and in our experiments is only slightly inferior to RST

Essential Bibliography

[1] N. Cesa-Bianchi, C. Gentile, F. Vitale, and G. Zappella.
Random spanning trees and the prediction of weighted graphs
ICML 2010

[2] X. Zhu, Z. Ghahramani, and J. Lafferty.
Semi-supervised learning using gaussian fields and harmonic functions
ICML 2003

[3] M. Herbster, M. Pontil, and S. Rojas-Galeano.
Fast prediction on a tree.
NIPS 2009