Interaction Prediction in Dynamic Networks exploiting Community Discovery

Giulio Rossetti[1,2] Riccardo Guidotti[1,2] Diego Pennacchioli[2] Dino Pedreschi[1] Fosca Giannotti[2]

[1] University of Pisa {*rossetti,guidotti,pedre*}*@di.unipi.it* [2] ISTI-CNR {*rossetti,guidotti,pennacchioli,giannotti*}*@isti.cnr.it*

Introduction

Interactions between people are becoming more and more easy to establish and track due to the growing availability of online social services. Nowadays human activities generate digital footprints, which describe complex, rapidly evolving, dynamic networks. In such scenario one of the most challenging task is the prediction of future interactions between couples of actors.

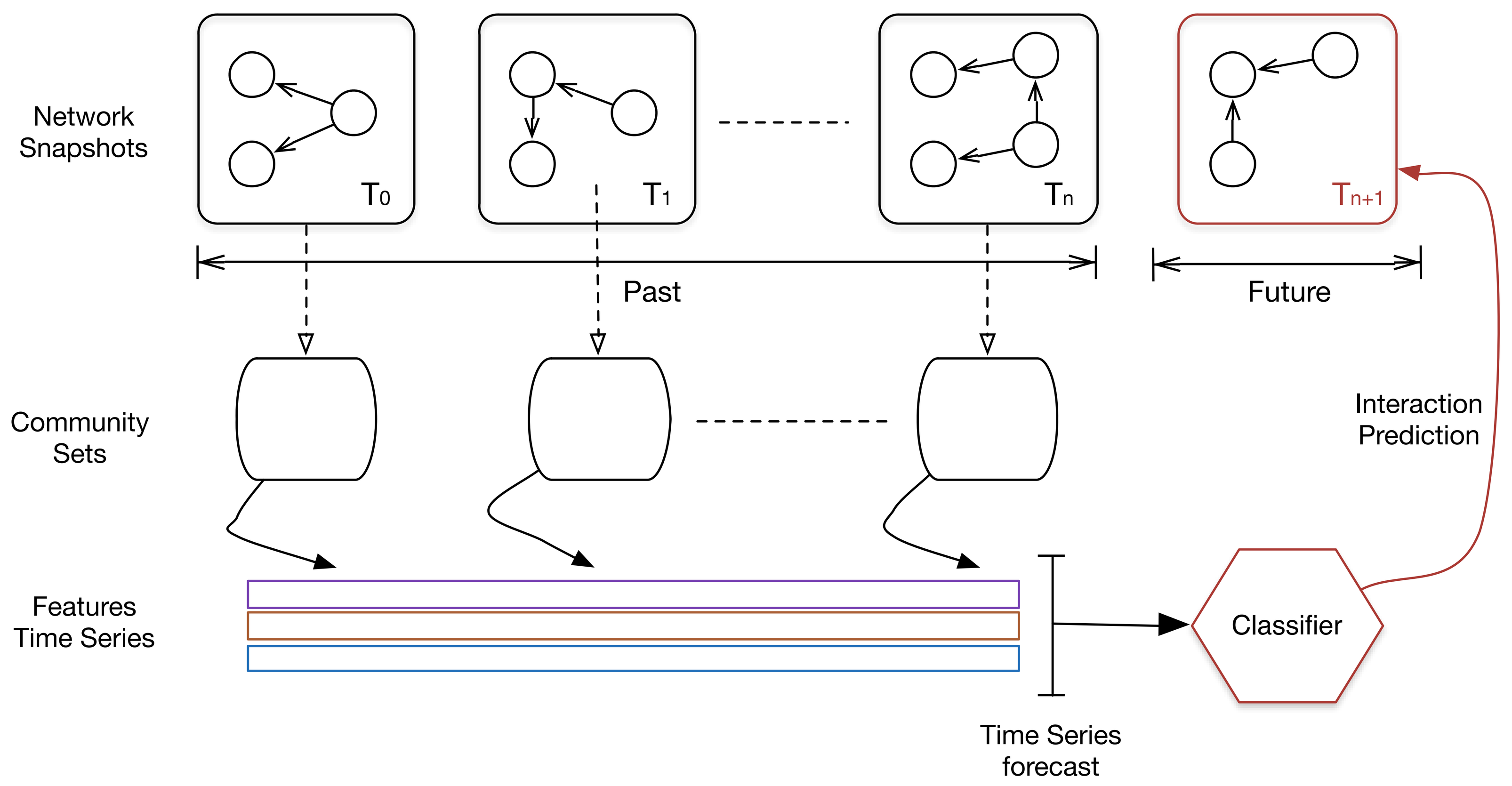
Interaction Prediction: Problem Definition

Given a set $\mathcal{G} = \{G_0, \dots G_t, \dots G_\tau\}$ of ordered network observations, with $t \in T = \{0 \dots \tau\}$, predict the interactions that are most likely to took place at time $\tau + 1$ thus composing $G_{\tau+1}$.

Case Study

- **DBLP**: co-author graph built upon the decade 2001-2010:
  9 years used as training and 1 as test.
- **Social**: an online social interaction network:
  5 months of messages are used as training and one as test.

Analytical Workflow



i. **Community Discovery**:
for each network snapshot we extract communities *(i)* to restrict the prediction to the edges among users belonging to the same social context, and *(ii)* to overcome the computational drawbacks caused by the sparseness of dynamic social structures;

ii. **Features design and Extraction**:
for each node pairs of each community we compute pairwise topological features, global features and community features;

iii. **Features Time-series Forecast**:
for each feature we build a time-series and we forecast its next value;

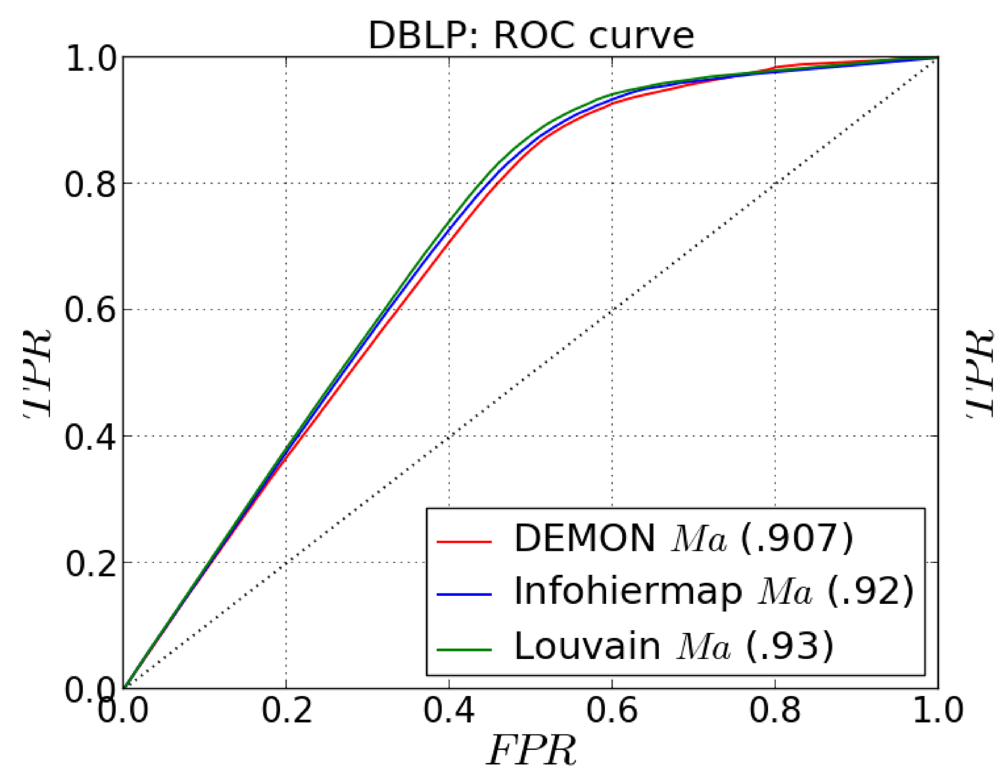
iv. **Supervised Interaction Prediction**:
using the forecasted features we build a supervised classifier to predict the presence of future interactions.

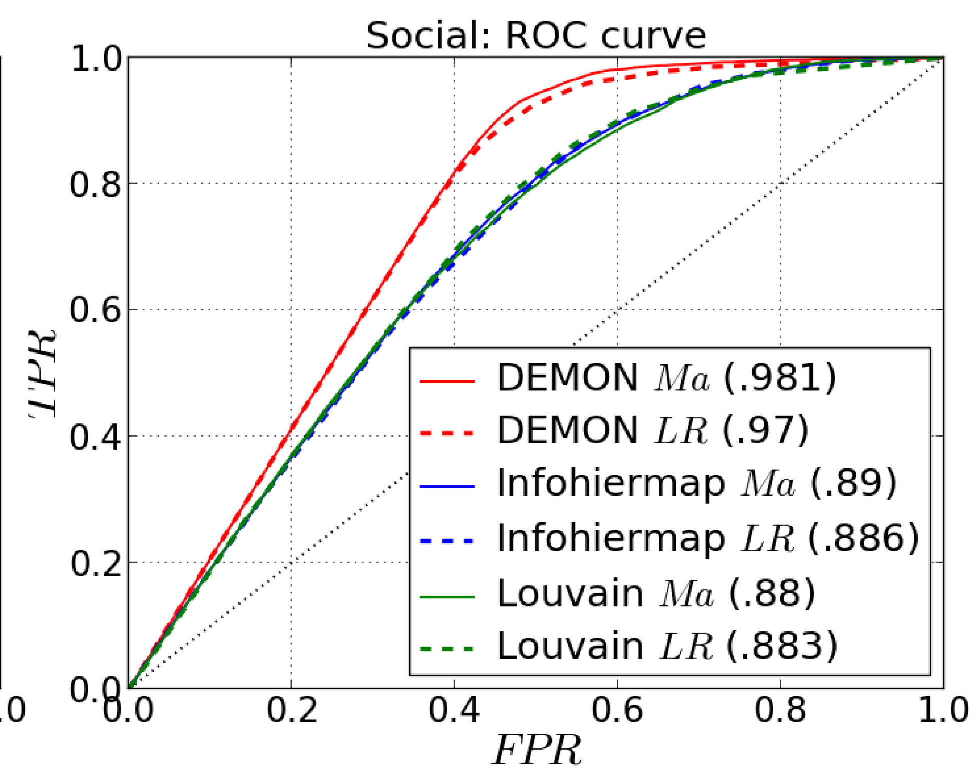
Experiments: Balanced Scenario

Class balancing reached through downsampling
(i.e., equal number of positive and negative instances)

Network	*DBLP*		*Social*	
Algorithm	AUC	ACC	AUC	ACC
DEMON *Ma*	0.907	85.58%	**0.981**	93.55%
DEMON *LR*	0.901	84.35%	0.970	91.87%
LOUVAIN *Ma*	**0.930**	87.72%	0.880	80.27%
LOUVAIN *LR*	0.926	87.48%	0.883	81.37%
INFOHIERMAP *Ma*	0.920	86.69%	0.890	81.34%
INFOHIERMAP *LR*	0.917	86.18%	0.886	80.89%

Accuracy (ACC) and Area under ROC (AUC) vary w.r.t. the community discovery and time series forecasting method used.

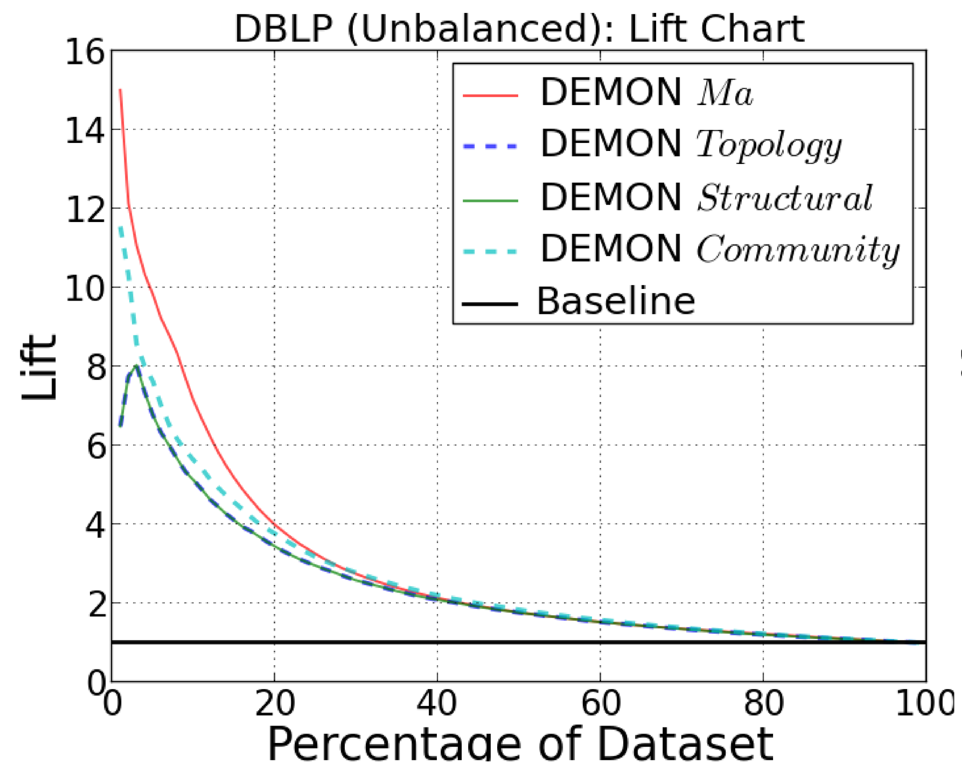


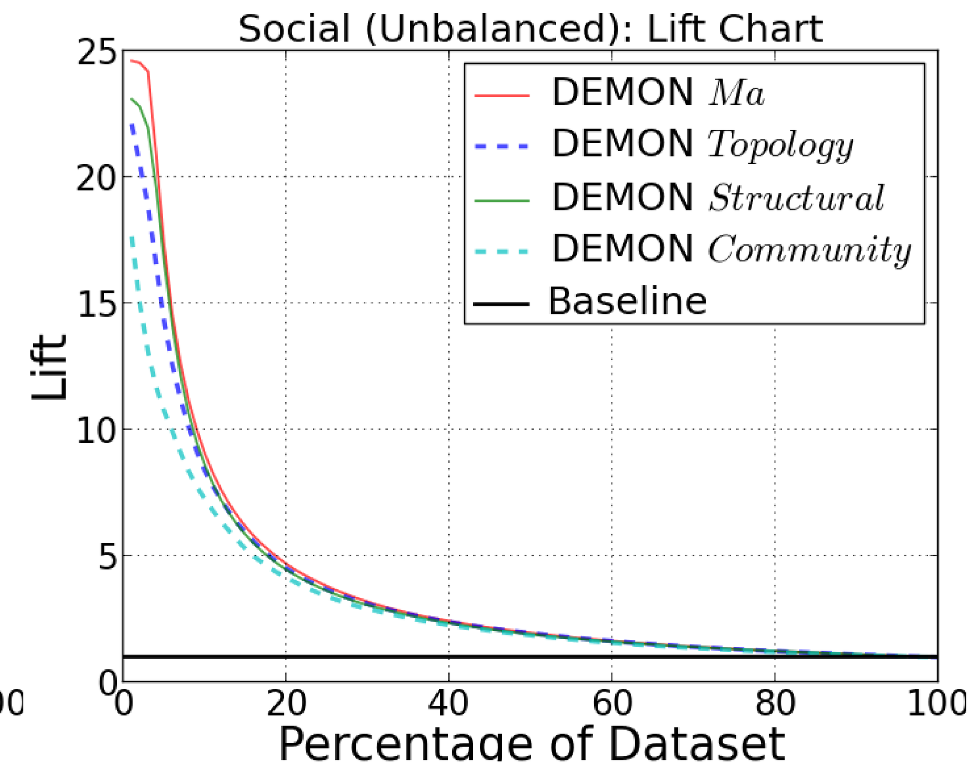


Experiments: Unbalanced Scenario

Real class distributions
(i.e., original ratio of node pairs)

- **DBLP**: Positive class 1%, Negative class 99%.
  Precision achieved w.r.t. Positive class 96%
- **Social**: Positive class 4%, Negative class 96%.
  Precision achieved w.r.t. Positive class 45%





Conclusions

The proposed methodology achieves high performances both in balanced and unbalanced class distribution scenarios:

- Communities are able to introduce a valid filter able to reduce False Positive Predictions
- Time series forecasting produces a topological description of the expected future network status

References

- S. Soundarajan , J. Hopcroft, *Using community information to improve the precision of link prediction methods*, In ACM WWW, 2012.
- P. R. da Silva Soares, et al. *Time series based link prediction.*, In IEEE IJCNN, 2012.
- M. Rosvall, et al. *Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems*, PloS one, 2011.
- M. Coscia, et al. *Demon: a local-first discovery method for overlapping communities*, In ACM SIGKDD, 2012.
- V. D. Blondel, et al. *Fast unfolding of communities in large networks*, JSTAT, 2008.