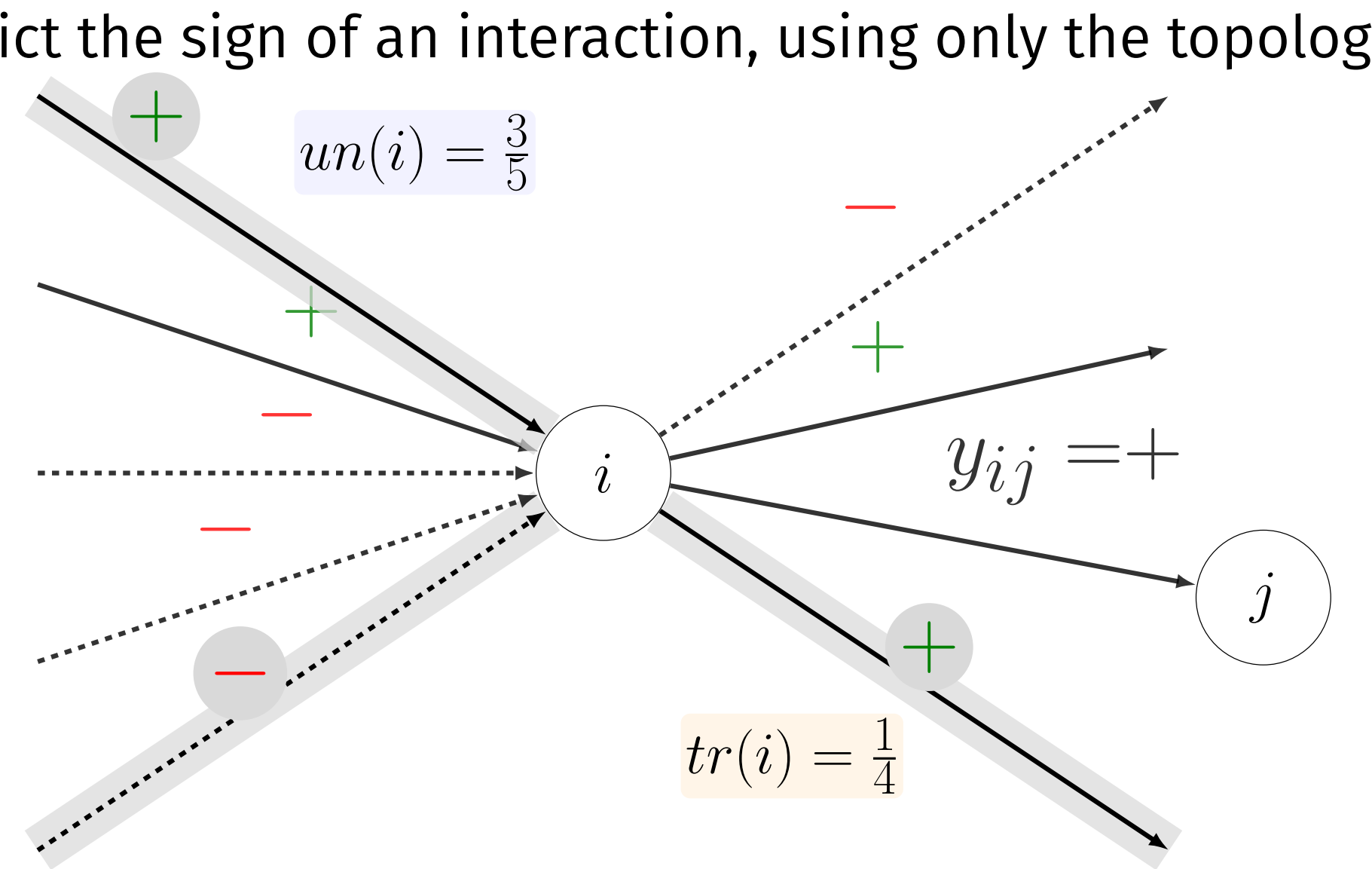
On the Troll-Trust Model for Edge Sign Prediction in Social Networks

Géraud Le Falher[1], Nicolò Cesa-Bianchi[2], Claudio Gentile[3] and Fabio Vitale[1,4]

[1] Inria, Univ. Lille, CNRS UMR 9189 – CRIStAL, France [2] Università degli Studi di Milano, Italy [3] University of Insubria, Italy [4] Department of Computer Science, Aalto University, Finland

Problem setup and notations

Given a directed graph of binary interactions between users, we want a **scalable** method to predict the sign of an interaction, using only the topology and the sign of some edges in $\mathbf{E}_0$.



The **trollness** $tr(i)$ of i is its fraction of negative outgoing edges, its **untrustworthiness** $un(i)$ is its fraction of negative incoming edges.

Contributions

1. A simple **generative model**, justifying existing heuristics and providing a **new principled predictor** (BLC(tr, un))
2. A maximum likelihood approximation by a **label propagation algorithm** (L. PROP.), leveraging a reduction from **edge to node classification**
3. Extensive comparative **experiments on real data** against state of the art methods

Generative model

Each node i has 2 parameters, drawn from an arbitrary distribution μ over pairs in $[0,1]^2$:

- its tendency p_i to send positive edges (i.e. niceness) and
- its tendency q_i to receive positive edges (i.e. pleasantness)

$$\begin{tikzcd} i \arrow[r, "\mathbb{P}(y_{ij}=+1)=\frac{1}{2}(p_i+q_j)"] & j \end{tikzcd}$$

$$(p_i, q_i) \sim \mu(p,q) \qquad (p_j, q_j) \sim \mu(p,q)$$

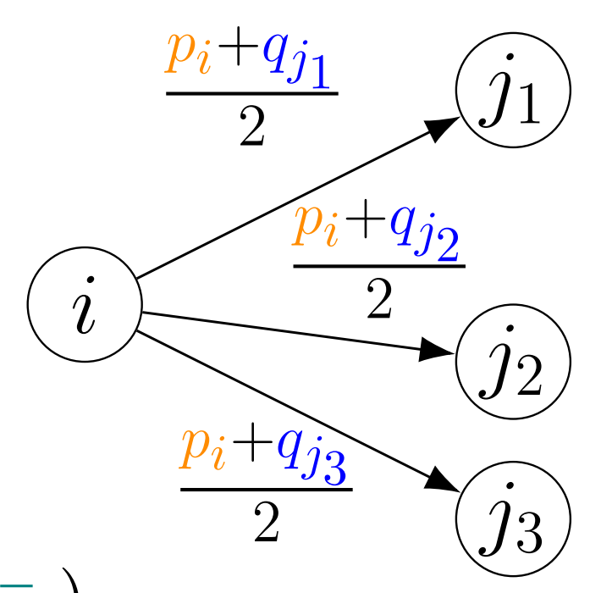
The Bayes optimal prediction for $y_{i,j}$ is thus

$$y^*(i,j) = \text{SGN}\left(\mathbb{P}\big(y_{i,j}=+1\big) - \tfrac{1}{2}\right)$$

Active algorithm: BLC$(\mathrm{tr}, \mathrm{un})$

We use the complementary to 1 of trollness and untrustworthiness (estimated on the training set $\mathbf{E}_0$) as proxy for p_i and q_j and predict

$$\widehat{y}(i,j) = \text{SGN}\Big(\underbrace{\big(1-\widehat{tr}(i)\big) + \big(1-\widehat{un}(j)\big) - \tau}_{\approx\frac{1}{2}(p_i+q_j)=\mathbb{P}(y_{ij}=+1)} - \tfrac{1}{2}\Big) \qquad (1)$$



$1-\widehat{tr}(i)$ is an indirect observation of p_i. Indeed, letting

$$\overline{q}_i = \frac{1}{d_{\text{out}}(i)} \sum_{j \text{ s.t.}(i,j)\in E} q_j$$

we have

$$1-\widehat{tr}(i) \approx \frac{1}{2}\left(p_i + \overline{q}_i\right) \quad \text{and likewise} \quad 1-\widehat{un}(j) \approx \frac{1}{2}\left(q_j + \overline{p}_j\right)$$

Thus we need to subtract

$$\tau = \frac{1}{2}\left(\mu_p + \mu_q\right)$$

as $\overline{p}_j$ and $\overline{q}_i$ concentrate around their mean μ_p and μ_q.
We sample Q outgoing and incoming edges for each node, estimate empirically $\widehat{tr}(i)$, $\widehat{un}(j)$ and τ as the overall fraction of positive edges then finally predict remaining edges according to (1).

Setting $Q = \frac{1}{\varepsilon^2}\ln\frac{|V|}{\delta}$ we query $\Theta\big(|V|\ln|V|\big)$ edges.

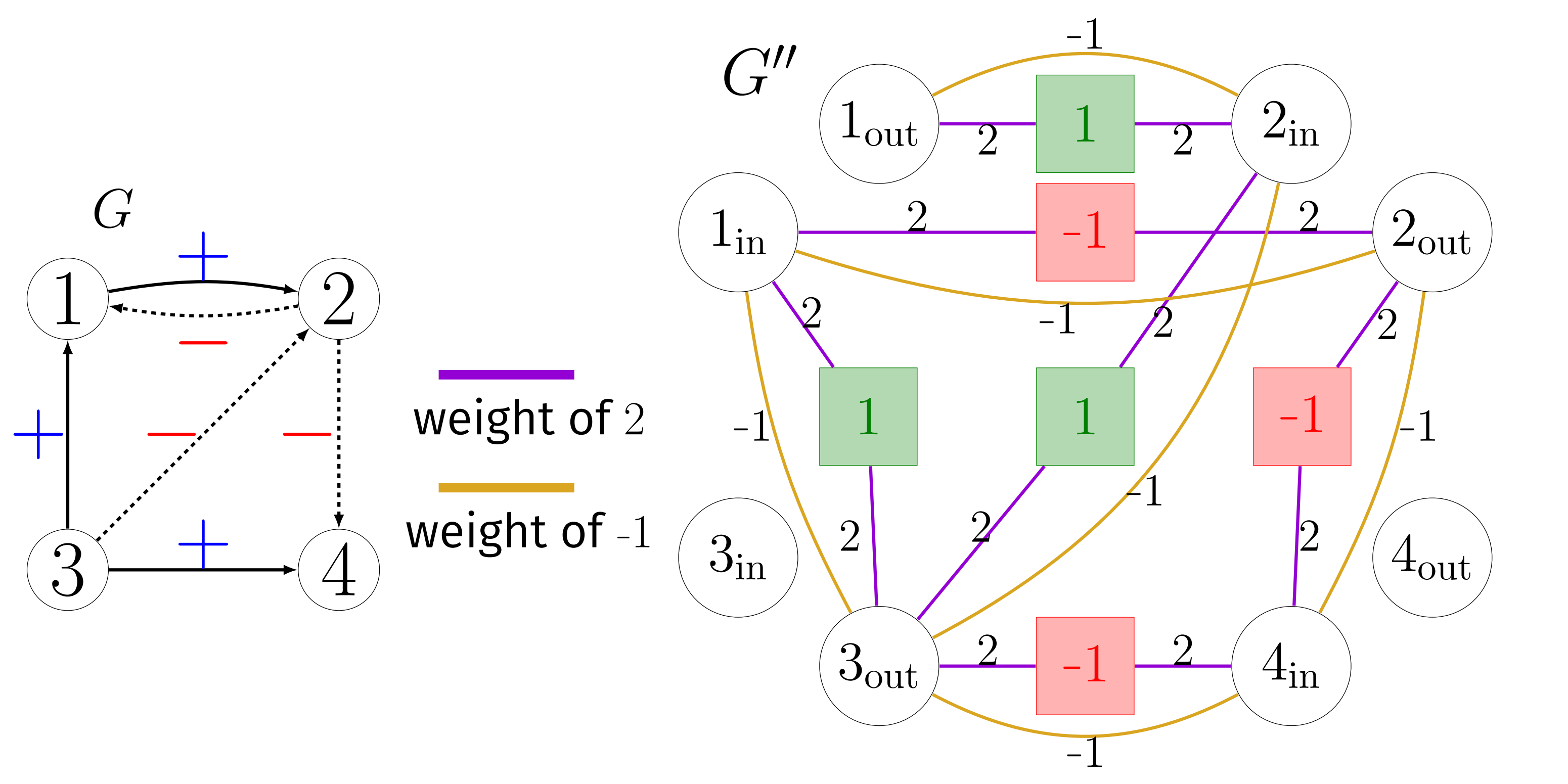
This is enough to guarantee that

$$\left|\left[\big(1-\widehat{tr}(i)\big) + \big(1-\widehat{un}(j)\big) - \widehat{\tau}\right] - \left[\frac{p_i+q_j}{2}\right]\right| \le 8\epsilon$$

holds with probability at least $1-10\delta$ simultaneously for all non-queried edges $(i,j)\in E$ such that $d_{\text{out}}(i), d_{\text{in}}(j) \ge Q$, therefore providing correct prediction as long as $\mathbb{P}(y_{i,j}=+1)$ is bounded away from $\frac{1}{2}$.
The BLC(tr, un) algorithm is inspired by this fact and also predicts according to (1), except there is no sampling but rather it uses the provided batch training set directly.

Batch algorithm: L. PROP.

Unfortunately, a good sampling requires a rather dense graph which is not usually the case in social media. In order to address this issue, first we transform the problem from **edge to node classification**.



Then we **approximate** $y^*(i,j)$ by resorting to a maximum likelihood estimator of the parameters $\{p_i,q_i\}_{i=1}^{|V|}\,\mathbb{P}\left(E_0 \,|\, \{p_i,q_i\}_{i=1}^{|V|}\right)$. Because the gradients of the log-likelihood function are not linear (e.g. w.r.t. p_ℓ)

$$\sum_{\ell,j\in E_0; y_{\ell j}=+1} \frac{1}{p_\ell + q_j} \quad - \sum_{\ell,j\in E_0; y_{\ell j}=-1} \frac{1}{2-p_\ell-q_j}$$

we approximate them, which is equivalent to setting to zero the gradient w.r.t. $(\boldsymbol{p},\boldsymbol{q}) = \{p_i,q_i\}_{i=1}^{|V|}$ of the quadratic function

$$f_{E_0}(\boldsymbol{p},\boldsymbol{q}) = \sum_{(i,j)\in E_0}\Big(\underbrace{\frac{p_i+q_j}{2}}_{\in[0,1]} - \underbrace{\frac{1+y_{i,j}}{2}}_{\in[0,1]}\Big)^2$$

We follow a label propagation approach by **making the test labels appear** and minimizing $f_{E_0}(\boldsymbol{p},\boldsymbol{q}) + f_{E\setminus E_0}(\boldsymbol{p},\boldsymbol{q},y_{i,j})$ w.r.t. both $(\boldsymbol{p},\boldsymbol{q})$ and all $y_{i,j}\in[-1,+1]$, for $(i,j)\in E\setminus E_0$.
Setting the edge weights as in G'' introduces an extra regularization term.

$$\underbrace{\widehat{f}\big(\boldsymbol{p},\boldsymbol{q},y_{i,j\,(i,j)\in E\setminus E_0}\big)}_{\text{energy function on } G''} = f_{E_0}(\boldsymbol{p},\boldsymbol{q}) + f_{E\setminus E_0}(\boldsymbol{p},\boldsymbol{q},y_{i,j}) + \widetilde{\mathrm{regul}}$$

We run $\mathrm{diameter}(G)$ iterations of label propagation and use a binary threshold over the estimated $y_{i,j}$ to predict signs.

Online setting

We also study an online setting where the signs are **adversarial** rather than generated by our model.

Letting Y be the vector of all labels, $\Psi_{\text{out}}(i,Y)$ is the number of least used label outgoing from i, and $\Psi_{\text{out}}(Y) = \sum_{i\in V}\Psi_{\text{out}}(i,Y)$. Likewise for incoming edges, $\Psi_{\text{in}}(Y) = \sum_{j\in V}\Psi_{\text{in}}(j,Y)$ and finally **the regularity of a labeling** Y is $\Psi_G(Y) = \min\left\{\Psi_{\text{in}}(Y), \Psi_{\text{out}}(Y)\right\}$.

We devise an algorithm consisting of a combination of Randomized Weighted Majority instances built on top of each other that on average makes $\Psi_G(Y) + O\left(\sqrt{|V|\Psi_G(Y)} + |V|\right)$ mistakes.

On the lower side, for any directed graph G and any integer K, there exists a labeling Y forcing at least $\frac{K}{2}$ mistakes to any online algorithms, while $\Psi_G(Y) \le K$.

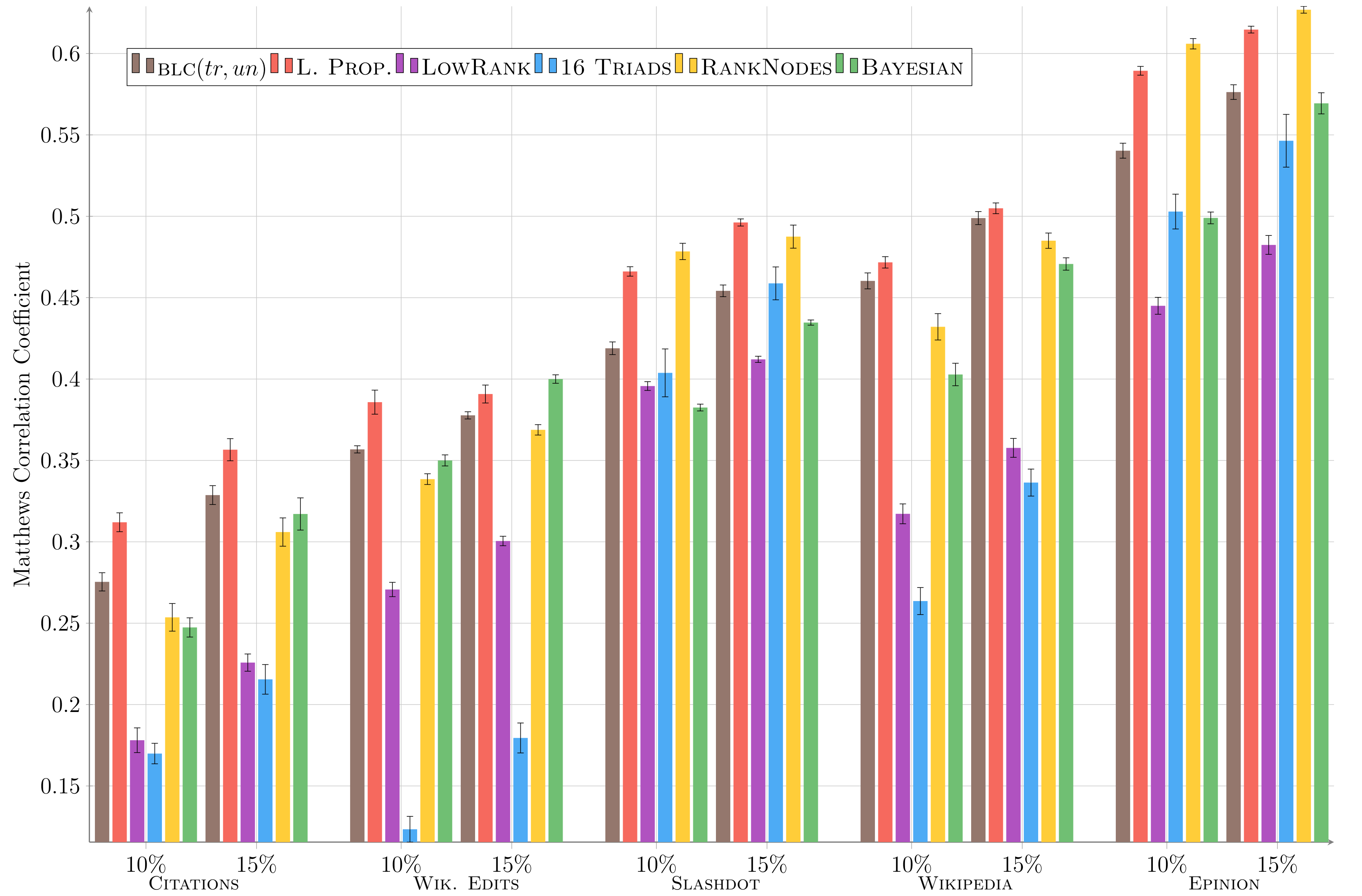
Experiments

5 datasets of different size from different domains:
CITATIONS i the work of j to praise it or criticise it.
WIKIPEDIA i vote for or against j promotion to adminship.
SLASHDOT i consider j as a friend or foe.
EPINION i trust or not the reviews made by j.
WIK. EDITS i reacted to a Wikipedia edit made by j, to enhance it or revert it.
Because of imbalance, we evaluate using the Matthews Correlation Coefficient (MCC):

$$\mathrm{MCC} = \frac{tp\times tn - fp\times fn}{\sqrt{(tp+fp)(tp+fn)(tn+fp)(tn+fn)}}\begin{cases}1 & \text{all predictions correct}\\ 0 & \text{random predictions}\\ -1 & \text{all predictions incorrect}\end{cases}$$

Competitors are representative of 2 classes of methods: LOWRANK **matrix completion** [1] and using **logistic regression on edge features** built:

- on 16 TRIADS features, as signed graphs exhibit specific triangle patterns according to the status theory [2]
- on a high number of so-called "BAYESIAN" features [3]
- on RANKNODES scores computed with a PageRank-inspired algorithm tailored to directed graphs with negative edges [4]



time (ms)	BLC(tr, un)	L. PROP.	LOWRANK	16 TRIADS	RANKNODES	BAYESIAN
CITATIONS	0.6	19.6	3,279	6.2	155	4,813
WIK. EDITS	10.1	1,329	127,654	209	3,174	104,305
SLASHDOT	8.3	677	69,988	131	2,441	68,085
WIKIPEDIA	9.6	927	129,460	177	3,890	92,719
EPINION	1.6	41.9	8,523	14.8	249	12,507

Discussion

We presented two batch algorithms derived from our generative model. One is local (BLC(tr, un)) and therefore extremely scalable while being performant both in theory and in practice. The other (L. PROP.) propagates sign information along the graph which provides more accurate results while remaining faster than previous approaches.
Later, we would like to extend our results to **weighted graphs** and **incorporate side information**.

References

[1] K.-y. Chiang *et al.*, "Prediction and Clustering in Signed Networks: A Local to Global Perspective," *JMLR*, vol. 15, pp. 1177–1213, 2014.

[2] J. Leskovec *et al.*, "Predicting positive and negative links in online social networks," in *Proc. of the 19th international conference on World wide web*, 2010, p. 641.

[3] D. Song *et al.*, "Link sign prediction and ranking in signed directed social networks," *Social Network Analysis and Mining*, vol. 5, no. 1, 2015.

[4] Z. Wu *et al.*, "The troll-trust model for ranking in signed networks," in *WSDM*, 2016, pp. 447–456.

LaTeX TikZposter