

Introduction

We discuss an inspiring use of the links between random processes on graphs and Laplacian-based numerical algebra:

- Based on *random spanning forests*, we propose two novel and efficient Monte Carlo estimators for smoothing graph signals.
- Moreover, we provide a theoretical analysis on bias and variance of our estimators.
- On the empirical side, these estimators are illustrated in two well-known applications.

Problem Definition

Regularized Regression on Graphs

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{z} \in \mathbb{R}^n} q \underbrace{\|\mathbf{y} - \mathbf{z}\|^2}_{\text{Fidelity}} + \underbrace{\mathbf{z}^T \mathbf{L} \mathbf{z}}_{\text{Regularization}}, \quad q > 0$$

where $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is a graph signal and y_i is the measurement from node i . \mathbf{L} denotes the graph Laplacian of the given graph and q sets the trade-off between data-fidelity and regularization.

The explicit solution to this problem is:

$$\hat{\mathbf{x}} = \mathbf{K} \mathbf{y} \text{ with } \mathbf{K} = (\mathbf{L} + q\mathbf{I})^{-1} q \mathbf{I}$$

where \mathbf{I} is the identity matrix.

- Direct computation of \mathbf{K} requires $\mathcal{O}(n^3)$ elementary operations due to the inverse.
- For large n , iterative methods and polynomial approximations are the state-of-the-art. Both compute $\hat{\mathbf{x}}$ in linear time in the number of edges $|\mathcal{E}|$.

Random Spanning Forests on Graphs

For an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$:

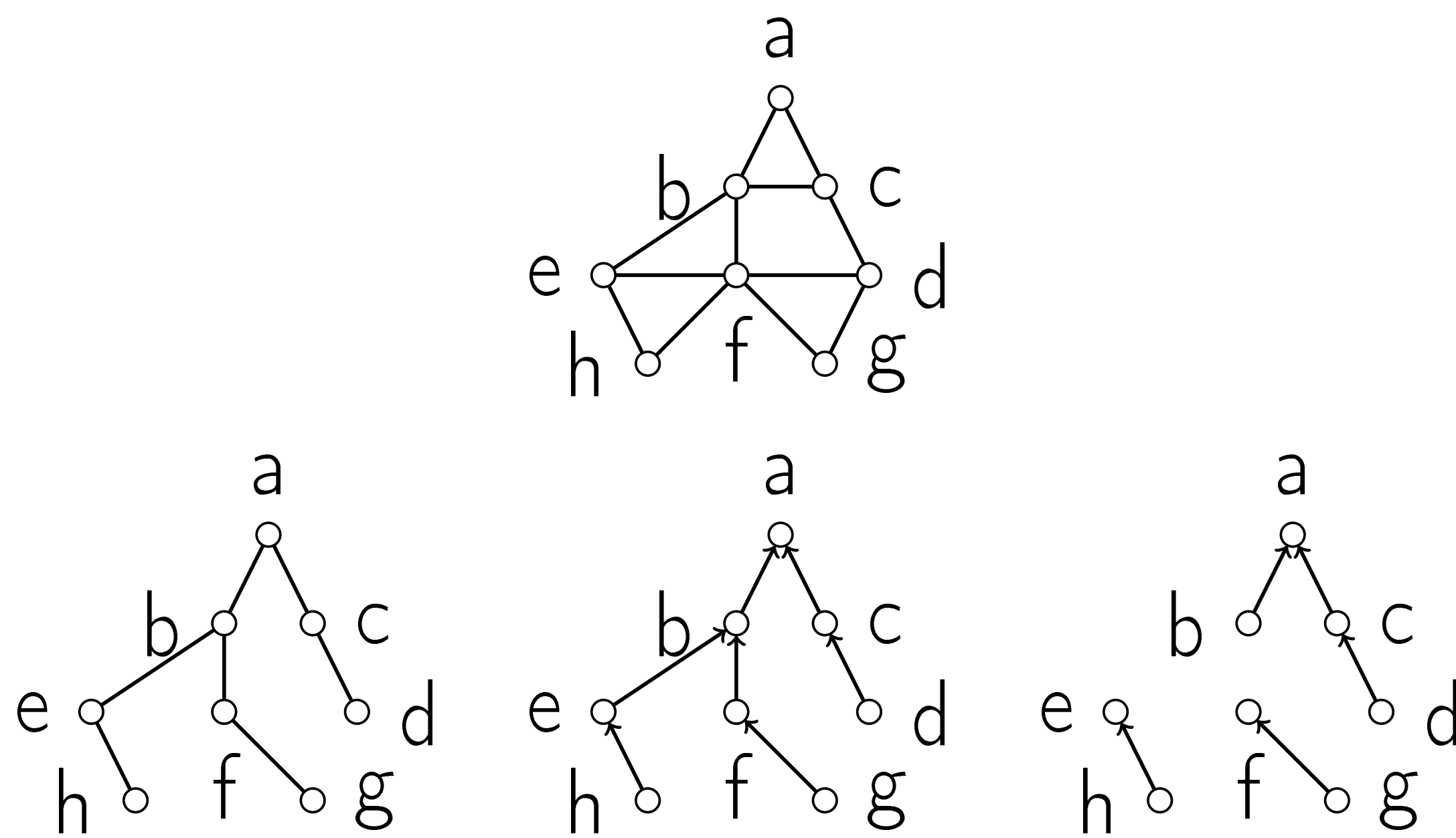


Fig. 1: Original graph, a spanning tree, a rooted spanning tree and a rooted spanning forest

Random Spanning Forests (RSF)

Consider the following parametric distribution over rooted spanning forests:

$$P(\Phi_q = \phi) \propto q^{|\rho(\phi)|} \prod_{\tau \in \phi} \prod_{(i,j) \in \tau} W_{i,j}$$

where q is a parameter and $\rho(\phi)$ denotes the set of roots in the forest ϕ . One can sample from this distribution by a variant of Wilson's algorithm in time $\mathcal{O}(|\mathcal{E}|/q)$ [1].

A very surprising connection between RSFs and linear algebra appears in the following simple equation:

$$P(r_{\Phi_q}(i) = j) = \mathbf{K}_{i,j} \text{ with } \mathbf{K} = (\mathbf{L} + q\mathbf{I})^{-1} q \mathbf{I}$$

where $r_{\Phi_q}(\cdot)$ returns the root of node i in the spanning forest Φ_q .

RSF based Estimators

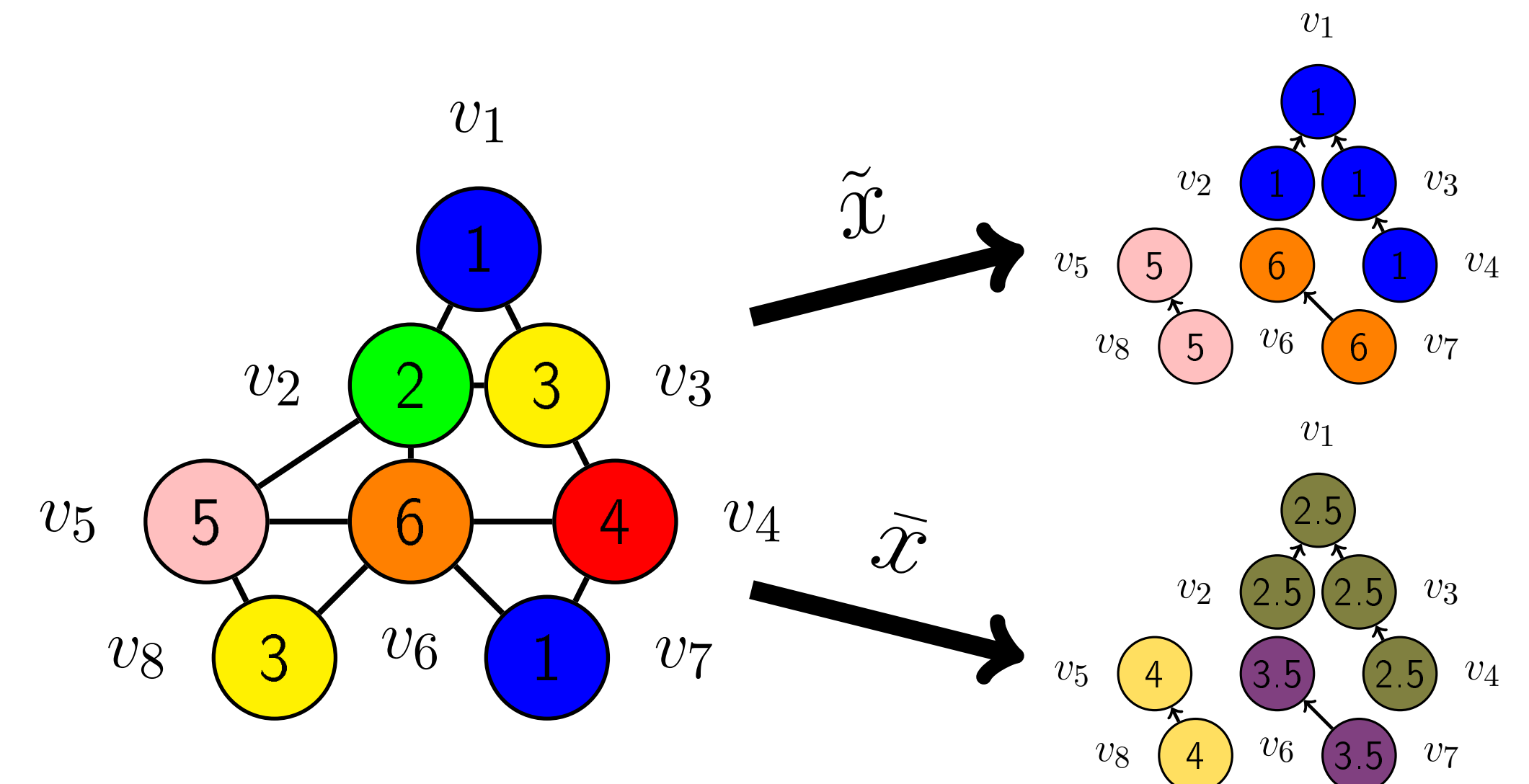


Fig. 2: An illustration for the estimators

RSF Estimator	Formula	Expectation	Variance
$\tilde{x}(i)$	$y[r_{\Phi_q}(i)]$	$\mathbb{E}[\tilde{x}(i)] = \delta_i \mathbf{K} \mathbf{y}$	$\sum_{i \in \mathcal{V}} \text{Var}(\tilde{x}(i)) = \mathbf{y}^T (\mathbf{I} - \mathbf{K}^2) \mathbf{y}$
$\bar{x}(i)$	$\frac{1}{ \mathcal{V}_{i(i)} } \sum_{j \in \mathcal{V}_{i(i)}} y_j$	$\mathbb{E}[\bar{x}(i)] = \delta_i \mathbf{K} \mathbf{y}$	$\sum_{i \in \mathcal{V}} \text{Var}(\bar{x}(i)) = \mathbf{y}^T (\mathbf{K} - \mathbf{K}^2) \mathbf{y}$

Experiments

Tikhonov denoising of graph signals. We consider an image as a noisy graph signal where the underlying graph is 2-D grid:

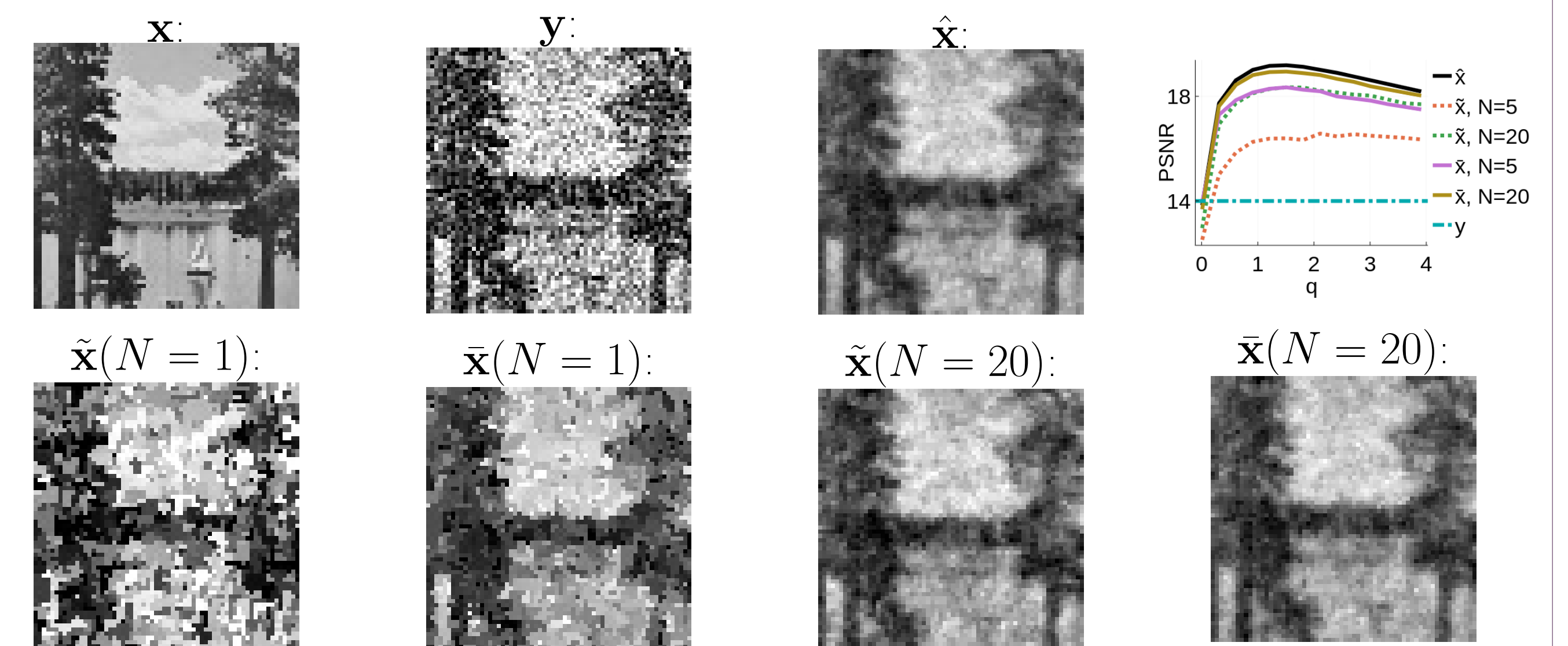


Fig. 3: Illustration on an image.

SSL for node classification on graphs aims at recovering labels of all nodes with a few available labels as *a priori* knowledge. Given labels in \mathbf{y}_l for class l , the solution [2] under a smoothness prior regulated by $\mu > 0$ is the classification function $\mathbf{f}_l = \frac{\mu}{2+\mu} \left(\mathbf{I} - \frac{2}{2+\mu} \mathbf{D}^{-\sigma} \mathbf{W} \mathbf{D}^{\sigma-1} \right)^{-1} \mathbf{y}_l = \mathbf{D}^{1-\sigma} \mathbf{K} \mathbf{D}^{\sigma-1} \mathbf{y}_l$ where $\mathbf{K} = (\mathbf{Q} + \mathbf{L})^{-1} \mathbf{Q}$ and $\mathbf{Q} = \frac{\mu}{2} \mathbf{D}$. We run our methods to estimate \mathbf{f}_l 's on a graph with $|\mathcal{V}| = 3000$ and two communities.

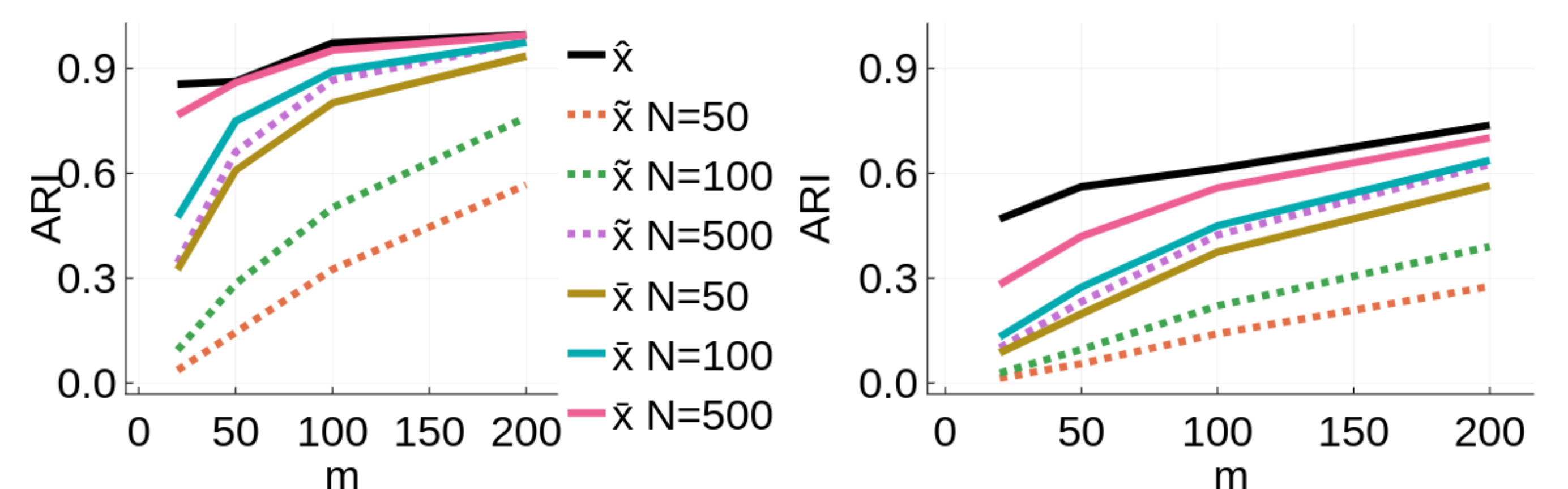


Fig. 4: Adjusted Rand Index (ARI) compared on strong (left) and fuzzy (right) communities

References

- [1] L. Avena and A. Gaudilière. Random spanning forests, Markov matrix spectra and well distributed points. *arXiv:1310.1723 [math]*, Oct. 2013. arXiv: 1310.1723.
[2] K. Avrachenkov, A. Mishenin, P. Gonçalves, and M. Sokol. Generalized Optimization Framework for Graph-based Semi-supervised Learning. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, pages 966–974.