# Graphical Nadaraya-Watson Estimator on Latent Position Models

Martin Gjorgjevski[1], Nicolas Keriven[1], Simon Barthelme[1], Yohann De Castro[2]

[1]CNRS Gipsa-Lab, Grenoble, [2]CNRS Ecole Centrale Lyon, Lyon

## Summary

The ***Graphical Nadaraya Watson*** Estimator $\hat{f}_{GNW}$ is a signal averaging estimator on graphs, inspired by the ***Nadaraya-Watson Estimator*** $\hat{f}_{NW}$ in nonparametric estimation. We study concentration properties and risk decay rates of $\hat{f}_{GNW}$ in terms of the growth of the degree of a vertex. We show that under mild assumptions, the estimator concentrates with a rate that decreases *exponentially* in the degree of a vertex. We also show that for smooth signals $\hat{f}_{GNW}$ and $\hat{f}_{NW}$ achieve similar risk rates.

## Framework: *Latent Position Models*

- $X_1, ...X_n, X$ i.i.d. $\sim p$, $p$ a density on $\mathbb{R}^d$ *not observed*
- $k_n : \mathbb{R}^d \to [0,1]$ *probability kernel*
- $a(X_i, X_j) = bern(k_n(X_i, X_j))$ *edge between nodes i and j*
- $Y_i - f(X_i) + \epsilon_i$, $\epsilon = (\epsilon_i)_{i=1}^n$ *noise independent from* $(X_i)_{i=1}^n$, with $\mathbb{E}\epsilon_i = 0$, $\mathbb{E}\epsilon_i^2 = \sigma^2 < \infty$
- $d_n(x) = \mathbb{E}(\sum_{i=1}^n a(X, X_i) | X = x)$ *local expected degree at* $x$

*Goal:* ***Estimate*** $f(X)$

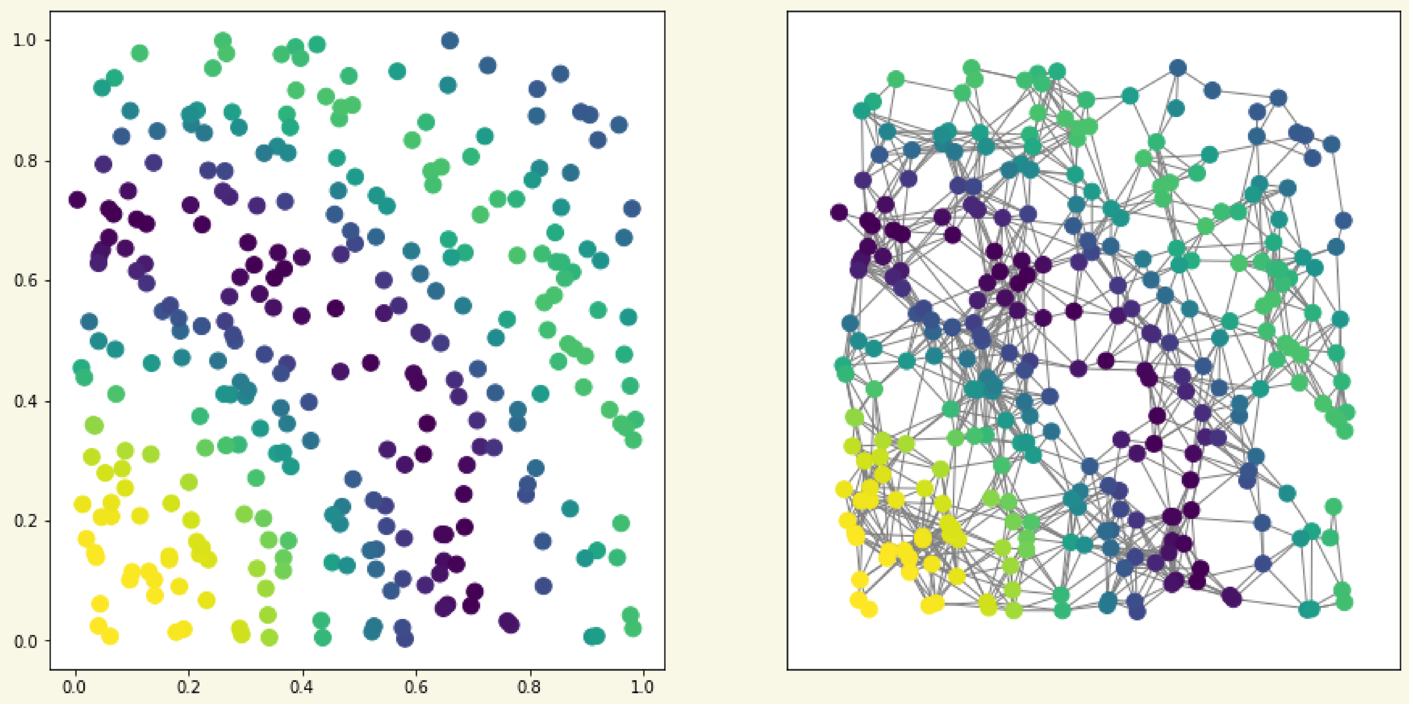

**Figure:** Left- latent positions, Right - Latent Position Random Graph

## Main Result

LLN heuristics: by setting $b_n(f,x) = \dfrac{\int f(z)k_n(x,z)p(z)dz}{\int k_n(x,z)p(z)dz}$ we have

$$\hat{f}_{GNW}(x) \sim b_n(f,x)$$

Suprisingly, we can compute

$$\mathbb{E}(\hat{f}_{GNW}(x)) = b_n(f,x)(1 - (1 - \frac{d_n(x)}{n})^n)$$

**Theorem**

If $f\colon \mathbb{R}^d \to \mathbb{R}$ is s.t. $\|f\|_\infty \leqslant B$ and $\mathbb{E}(\epsilon_1^2) = \sigma^2$. Then

$$\frac{\sigma^2(1 - e^{-d_n(x)})}{d_n(x)} \leqslant \mathbb{E}(\hat{f}_{GNW}(x) - b_n(f,x))^2 \leqslant \frac{C(B, \sigma^2)}{d_n(x)}$$

- As soon as $d_n(x) \to \infty$, the *variance term* tends to $0$
- Left to bound $|b_n(f,x) - f(x)|$ (the **bias** term)
- $d_n(x) \sim nh_n^d p(x)$ (*Lebesgue Density theorem*)

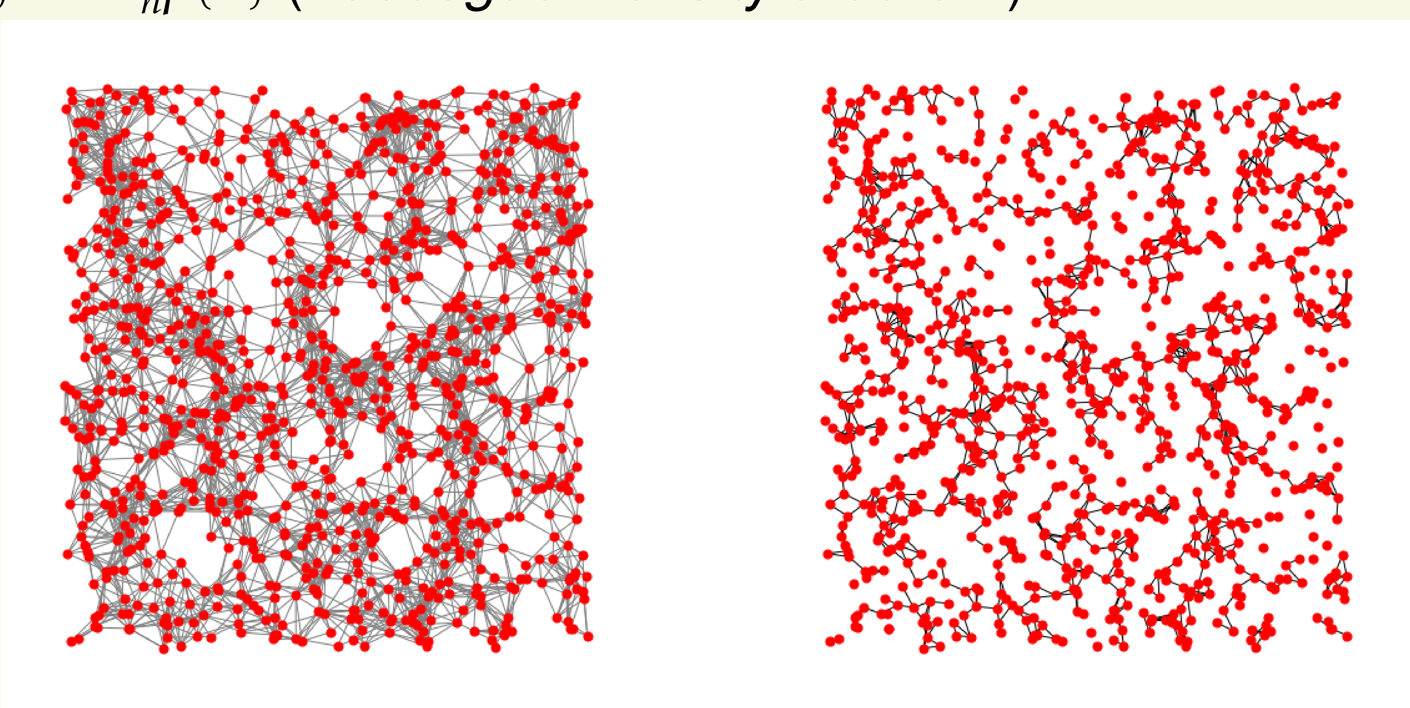

**Figure:** Sparse random graphs. Left: $d_n(x) \sim \log(n)$, Right: $d_n(x) \sim \log(\log(n))$

## The NW and GNW estimators

When the positions $X_1, ....X_n$ are known, a popular approach is **Nadaraya-Watson** Estimator

$$\hat{f}_{NW}(X) = \frac{\sum_{i=1}^n Y_i K(\frac{X - X_i}{h_n})}{\sum_{i=1}^n K(\frac{X - X_i}{h_n})}$$

In the LPM setting, we consider **Graphical Nadaraya-Watson** Estimator

$$\hat{f}_{GNW}(X) = \frac{\sum_{i=1}^n Y_i a(X, X_i)}{\sum_{i=1}^n a(X, X_i)}$$

The $L^2$ risk of the **NW** estimator admits the bias-variance decomposition

$$\mathbb{E}(\hat{f}_{NW}(x) - f(x))^2 = \mathbb{V}(\hat{f}_{NW}(x)) + (\mathbb{E}(\hat{f}_{NW}(x)) - f(x))^2$$

**Questions**

1. How does the quality of $\hat{f}_{GNW}$ depend on the degree ?
2. How does the $L^2$ risk of $\hat{f}_{GNW}$ compare to that of $\hat{f}_{NW}$?

## Proof Sketch - the Decoupling trick

For $I \subseteq [n]$. Define[a]

$$R_I(x) = \frac{1}{|I| + \sum_{j \notin I} a(x, X_j)}$$

For all pairs of **disjoint** subsets $I, J \subseteq$[n] we have

$$R_J(x) \prod_{i \in I} a(x, X_i) = R_{I \cup J}(x) \prod_{i \in I} a(x, X_i)$$

and $R_{I \cup J}(x)$ is ***independent*** from $\{a(x, X_i) | i \in I\}$.

- "Linearized" representation $\hat{f}_{GNW}(x) = \sum_{i=1}^n Y_i a(x, X_i) R_i(x)$

## MISE bound for convolutional kernels

Convolutional kernels $k_n(x,z) = K(\frac{x-z}{h_n})$ with $K\colon \mathbb{R}^d \to [0,1]$, $h_n > 0$

**Theorem**

- $K$ compactly supported
- $p(x) \geqslant p_0 > 0$ on $Q$ and $Q$ *satisfies interior cone condition*
- f is $\alpha$ *Hölder continnous on* supp $p$

then for sufficiently small bandwiths $h_n$ we have

$$\mathbb{E}(\hat{f}_{GNW}(X) - f(X))^2 \leqslant C_1(\alpha)h_n^\alpha + \frac{C(B, \sigma)}{nh_n^d}$$

[a]with the convention that $1/0 = 0$

## Simulations

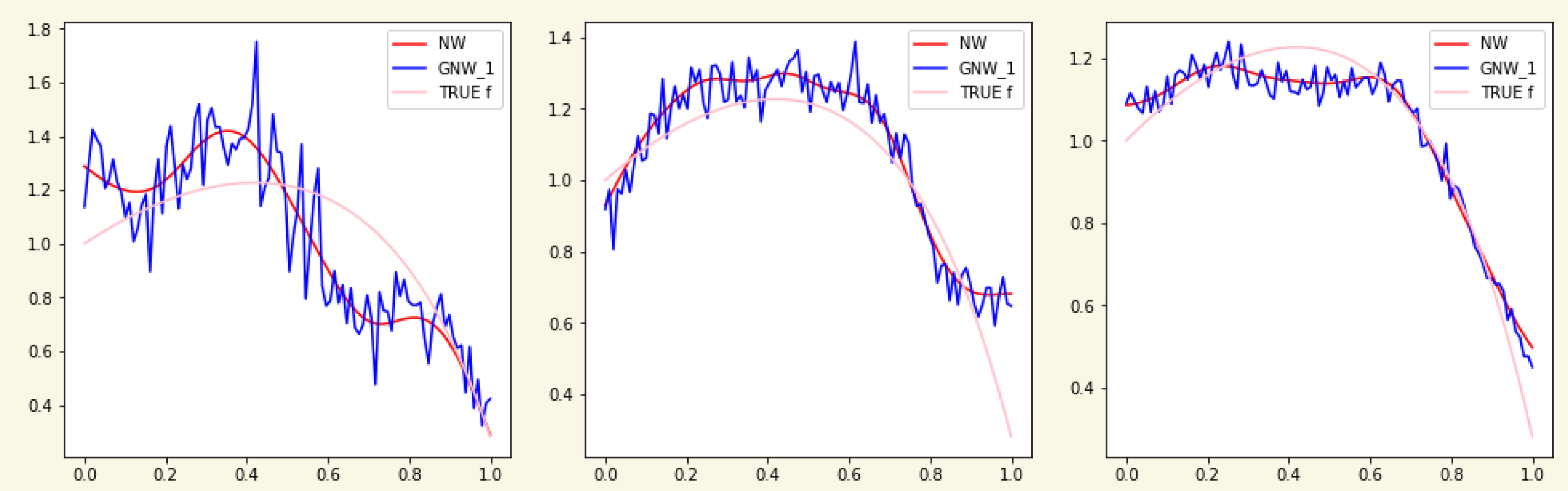


**Figure:** Comparison of **NW** and **GNW** estimators: Left - sample size of $n = 100$, center - sample size of $n = 500$, right -sample size of $n = 2000$



This work was supported in part by blah blah