Comparison between Z and W_1 distances

In Ding et al. (2021, Section 5), the authors say "*... In all numerical experiments, we simply use degree profiles defined through the usual vertex degrees. Moreover, instead of using the Z distance (28) defined as the total variation distance between discretized degree profiles, we directly use the 1-Wasserstein W_1-distance between degree profiles ...*". We now propose a simulation to try to verify the equivalence between Z distance and W_1 distance in practice.

Definitions

Suppose we have symmetric matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$ from the correlated Erdős-Rényi model $\mathcal{G}(n, q; s)$ with permutation $\pi^* : [n] \mapsto [n]$; i.e., for all $i < j$, $\boldsymbol{A}_{ij} \sim \text{Ber}(q)$ and conditional on $\boldsymbol{A}$

$$\boldsymbol{B}_{\pi^*(i)\pi^*(j)} \sim \begin{cases} \text{Ber}(s) & \text{if } \boldsymbol{A}_{ij} = 1 \\ \text{Ber}\left(\frac{q(1-s)}{1-q}\right) & \text{if } \boldsymbol{A}_{ij} = 0 \end{cases}.$$

Let $N_{\boldsymbol{A}}(i) = \{j \in [n] : \boldsymbol{A}_{ij} = 1, j \neq i\}$ denote the open neighborhood of node i in graph $\boldsymbol{A}$ for all $i \in [n]$; and define similar $N_{\boldsymbol{B}}(k)$ for graph $\boldsymbol{B}$ for all $k \in [n]$. Consider the standard degree of node j in graphs $\boldsymbol{A}$ and $\boldsymbol{B}$

$$a_j = \frac{1}{\sqrt{nq(1-q)}} \sum_{l \in [n]} (\boldsymbol{A}_{lj} - q), \quad b_j = \frac{1}{\sqrt{nq(1-q)}} \sum_{l \in [n]} (\boldsymbol{B}_{lj} - q),$$

and the corresponding empirical distributions

$$\mu_i = \frac{1}{|N_{\boldsymbol{A}}(i)|} \sum_{j \in N_{\boldsymbol{A}}(i)} \delta_{a_j}, \quad \nu_k = \frac{1}{|N_{\boldsymbol{B}}(k)|} \sum_{j \in N_{\boldsymbol{B}}(k)} \delta_{b_j}.$$

Consider the L-uniform partition over $[-1/2, 1/2]$ denoted as $\{I_l\}_{l \in [L]}$. The Z distance of pair $(i, k) \in [n]^2$ is defined as

$$Z_{ik} \coloneqq \sum_{l \in [L]} |\mu_i(I_l) - \nu_k(I_l)|.$$

The W_1 distance of pair $(i, k) \in [n]^2$ is defined as

$$W_{ik} \coloneqq \int_{t \in \mathbb{R}} |F_i(t) - G_k(t)| dt = \int_{u \in [0,1]} |F_i^{-1}(u) - G_k^{-1}(u)| du,$$

where $F_i(t) = \mu_i((-\infty, t))$ and $G_k(t) = \nu_i((-\infty, t))$ are the empirical CDF corresponding to μ_i, ν_k for all $t \in \mathbb{R}$.

Simulation results

We run the Algorithm 1 in Ding et al. (2021) with Z_{ik} and W_{ik} respectively. We consider the simulation with $n = 50, q = 1/2$ and $\sigma = \sqrt{1-s} \in \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$. We repeat the experiment under the same setting for 10 times and report average accuracy defined in Ding et al. (2021), $\mathrm{acc}(\hat{\pi}) = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}\{\pi^*(i) = \hat{\pi}(i)\}$ with the standard deviation across the replications. For Z distance, we exhaustively search the optimal L from 2 to n with gap 2 (i.e., try $L = 2, 4, 6, \ldots, n$) and report the best result.

Figure 1 indicates that the performance of W_1 distance is much better than the Z distance proposed in Ding et al. (2021, Equation (28)).

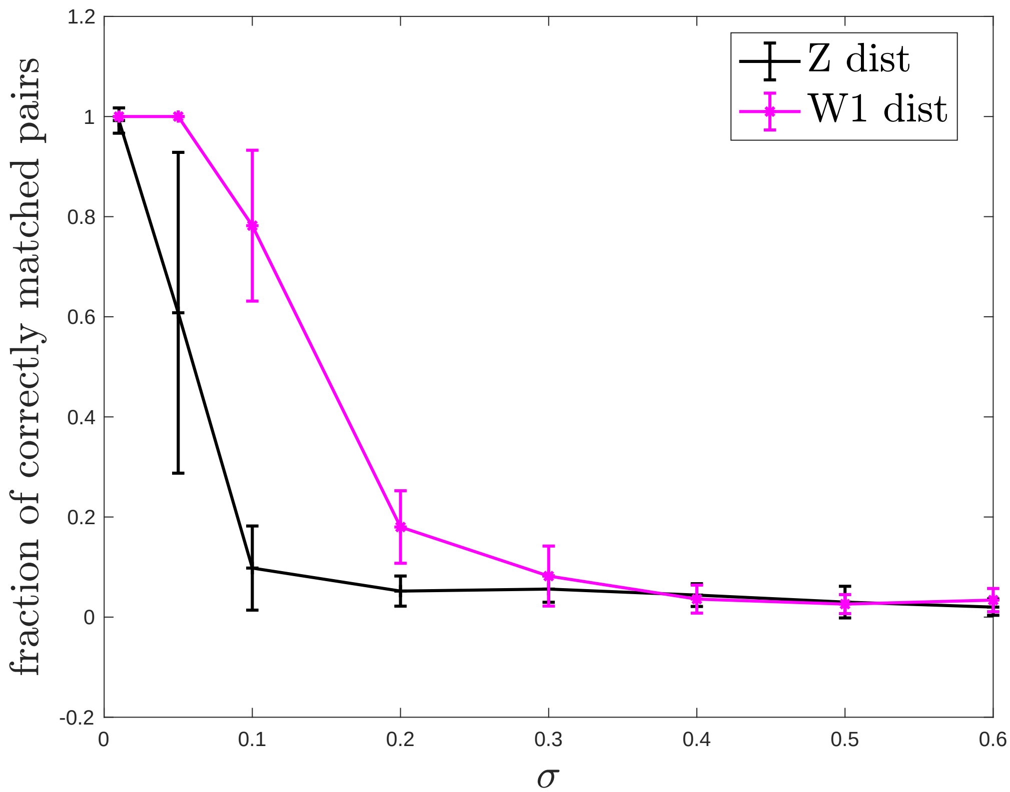

Figure 1: Accuracy comparison between Z and W_1 distances with varying noise level σ.

References

Ding, J., Ma, Z., Wu, Y., and Xu, J. (2021). Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179(1):29–115.