On Graph-Valued Stochastic Processes, Perturbations and Applications

Alejandro Morera Alvarez

Contents

Prelude

Chapter 1

Canonical Graph Process

1.1 Setting the Ground

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ the set of all undirected graphs $G = (V, E)$, where V denotes the set of vertices and E the set of edges of the graph G. We endow $\mathcal{G}$ with a metric $d : \mathcal{G} \times \mathcal{G} \to [0, \infty)$, which induces the Borel σ-field $\mathcal{B}(\mathcal{G}) := \sigma(\mathcal{O}(d))$, that is, the smallest σ-field containing all the open sets $U \in \mathcal{O}(d)$. Finally, let $P = P_n : (\Omega, \mathcal{F}) \to ([0,1]^{n\times n}, \mathcal{B}([0,1]^{n\times n}))$ be measurable. If $p_{ij}(\omega) = 0$ $\mathbb{P}$-a.s. for $i \geq j$, then, we call such a matrix-valued map a **clustering**.

For a graph-valued random variable $X : (\Omega, \mathcal{F}) \to (\mathcal{G}, \mathcal{B}(\mathcal{G}))$, we denote $V(X)$ its set of vertices and $E(X)$ its edges.

The central object of study of this chapter is the following process:

Definition 1.1.1 (Canonical Graph Process)
Let P be a clustering. The **canonical graph process** (CGP) $X = (X_n)_{n\in\mathbb{N}}$ is defined as the graph-valued stochastic process

$$X_n : (\Omega, \mathcal{F}) \to (\mathcal{G}, \mathcal{B}(\mathcal{G}))$$

described as follows:

$$X_0(\omega) := (\emptyset, \emptyset), \qquad X_1(\omega) := (\{v_1\}, \emptyset) \ \mathbb{P}\text{-a.s.},$$

and for $n \geq 2$,

$$\begin{aligned} V(X_n)(\omega) &:= V(X_{n-1})(\omega) \cup \{v_n\}, \\ E(X_n)(\omega) &:= E(X_{n-1})(\omega) \cup F(X_n)(\omega), \end{aligned}$$

where $F(X_n) \subseteq \mathcal{P}(V(X_n))$ is a random family consisting of two-element sets with

$$\mathbb{P}_P(\{v_i, v_n\} \in F(X_n)) := p_{in} \text{ for } i \in [n-1].$$

Remark 1.1.2 Note that P itself is allowed to be random. For instance, one can consider a sequence of iid

$(U_n) \overset{\mathbb{Q}}{\sim} \mathrm{Unif}(0,1)$ for $\mathbb{Q} \in \mathcal{M}_1(\widetilde{\Omega})$ where $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ is some measurable space, and

$$p_{in} := (1 - U_n) p_{i,n-1} \quad , \quad n \geq 2.$$

That is, $\mathbb{P}_P$ itself becomes a random measure.

We briefly recall the following definition for the result that follows:

Definition 1.1.3 (Poisson Binomial Distribution) Let $n \in \mathbb{N}$ and let $p = (p_1, \ldots, p_n) \in [0,1]^n$ be a vector of probabilities. The **Poisson Binomial distribution** is the probability distribution of the random variable

$$S : (\Omega, \mathcal{F}) \to \{0, \ldots, n\}$$

defined as the sum of independent indicators

$$S(\omega) := \sum_{i=1}^{n} Z_i(\omega),$$

where $Z_1, \ldots, Z_n$ are mutually independent Bernoulli variables with

$$\mathbb{P}(Z_i = 1) = p_i \quad \text{for } i \in [n].$$

One of the several properties one can consider to study about the CGP is the degrees of the vertices. To this end, we define

$$\delta_i^{(n)} := |\, \{\{v_i, v_j\} \in E(X_n) : j \in [n]\} \,|$$

the degree of the i-th vertex at time n. We aim to calculate its distribution.

Proposition 1.1.4 The distribution of $\delta_i^{(n)}$ is the **Poisson Binomial Distribution** with parameter vector:

$$\mathbf{p} = (p_{1,i}, \ldots, p_{i-1,i}, p_{i,i+1}, \ldots, p_{i,n})$$

Proof: Note that for the transition from step $n-1$ to n, the new node v_n connects to v_i with probability $p_{i,n}$ and does not connect with probability $1 - p_{i,n}$. Therefore, the degree $\delta_i^{(n)}$ relates to $\delta_i^{(n-1)}$ as follows:

$$\begin{aligned}
\mathbb{P}_P(\delta_i^{(n)} = k) &= \sum_{j=0}^{n-2} \mathbb{P}_P(\delta_i^{(n)} = k \mid \delta_i^{(n-1)} = j) \mathbb{P}_P(\delta_i^{(n-1)} = j) \\
&= \mathbb{P}_P(\delta_i^{(n)} = k \mid \delta_i^{(n-1)} = k-1) \mathbb{P}_P(\delta_i^{(n-1)} = k-1) + \mathbb{P}_P(\delta_i^{(n)} = k \mid \delta_i^{(n-1)} = k) \mathbb{P}_P(\delta_i^{(n-1)} = k) \\
&= \mathbb{P}(\{v_i, v_n\} \in F(X_n)) \mathbb{P}_P(\delta_i^{(n-1)} = k-1) + \mathbb{P}(\{v_i, v_n\} \notin F(X_n)) \mathbb{P}_P(\delta_i^{(n-1)} = k) \\
&= p_{i,n} \cdot \mathbb{P}_P(\delta_i^{(n-1)} = k-1) + (1 - p_{i,n}) \cdot \mathbb{P}_P(\delta_i^{(n-1)} = k)
\end{aligned}$$

This recursive relationship indicates that $\delta_i^{(n)}$ does not collapse into a simple product formula. Instead, $\delta_i^{(n)}$ is the sum of independent Bernoulli trials representing the edges formed at each step.

Specifically, we can decompose $\delta_i^{(n)}$ into two components:

1. **Initial Degree** (Edges formed when v_i arrived, connecting to older nodes $v_1, \ldots, v_{i-1}$):

$$D_{\text{init}} = \sum_{l=1}^{i-1} \mathbb{1}_{(\{v_l, v_i\} \in E(X_i))} \quad \text{where } \mathbb{1}_{(\{v_l, v_i\} \in E(X_i))} \sim \text{Ber}(p_{l,i})$$

2. **Accumulated Degree** (Edges formed as newer nodes $v_{i+1}, \ldots, v_n$ arrived and connected to v_i):

$$D_{\text{acc}} = \sum_{j=i+1}^{n} \mathbb{1}_{(\{v_i, v_j\} \in E(X_j))} \quad \text{where } \mathbb{1}_{(\{v_i, v_j\} \in E(X_j))} \sim \text{Ber}(p_{i,j})$$

Thus, the total degree is a sum of independent, non-identically distributed Bernoulli trials:

$$\delta_i^{(n)} = \sum_{l=1}^{i-1} \text{Ber}(p_{l,i}) + \sum_{j=i+1}^{n} \text{Ber}(p_{i,j}),$$

which yields the desired claim. □

Remark 1.1.5 A closed-form "product formula" like the one in the original notes exists only if we look for the probability of a specific path (e.g., node i connects to *all* subsequent nodes), which would be:

$$\mathbb{P}(\delta_i^{(n)} = \delta_i^{(i)} + (n-i)) = \prod_{j=i+1}^{n} p_{i,j}$$

However, for the general probability $\mathbb{P}(\delta_i^{(n)} = k)$, one must compute the coefficient of x^k in the generating function:

$$G(x) = \left(\prod_{l=1}^{i-1} ((1 - p_{l,i}) + p_{l,i} x) \right) \left(\prod_{j=i+1}^{n} ((1 - p_{i,j}) + p_{i,j} x) \right)$$

Corollary 1.1.6 (Isolation Probability) Let $\mathcal{I}_n \subseteq V(X_n)$ be the set of isolated vertices at time n. The probability that vertex v_i is isolated is given by:

$$\mathbb{P}(v_i \in \mathcal{I}_n) = \mathbb{P}(\delta_i^{(n)} = 0) = \prod_{l=1}^{i-1} (1 - p_{li}) \prod_{j=i+1}^{n} (1 - p_{ij}).$$

1.1.1 Gaussian Limit of the Degree Distribution

We now consider the specific case where the attachment probabilities are homogeneous. That is, we assume that the clustering P is deterministic and uniform such that for a fixed parameter $p \in (0, 1)$,

$$p_{in} \equiv p \quad \text{for all } n \geq 2,\ i < n.$$

This corresponds to the sequential construction of the Erdos-Renyi random graph $G(n, p)$. Under this assumption, the Poisson Binomial distribution simplifies significantly, allowing us to derive an asymptotic limit.

Proposition 1.1.7 (Homogeneous Degree Distribution) Let $p \in (0, 1)$. If the connection probabilities satisfy

$p_{in} = p$ for all i, n, then the degree of the i-th vertex at time n, denoted $\delta_i^{(n)}$, follows a Binomial distribution:

$$\delta_i^{(n)} \sim \mathrm{Bin}(n-1, p).$$

Proof: Recall the decomposition of the degree $\delta_i^{(n)}$ into independent Bernoulli trials from the previous proposition:

$$\delta_i^{(n)} = \sum_{l=1}^{i-1} \mathbb{1}_{(\{v_l, v_i\} \in E(X_i))} + \sum_{j=i+1}^{n} \mathbb{1}_{(\{v_i, v_j\} \in E(X_j))}.$$

Under the homogeneity assumption, every indicator variable in these sums is a Bernoulli random variable with success probability p. Furthermore, by the definition of the CGP, these edge formations are mutually independent.

The total number of trials is $(i-1) + (n-i) = n-1$. The sum of $n-1$ independent and identically distributed (i.i.d.) Bernoulli(p) variables is, by definition, a Binomial($n-1, p$) variable. $\square$

Theorem 1.1.8 (Gaussian Approximation)
As $n \to \infty$, the standardized degree distribution converges in distribution to the standard normal distribution. Specifically,

$$\frac{\delta_i^{(n)} - (n-1)p}{\sqrt{(n-1)p(1-p)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof: Since $\delta_i^{(n)}$ is the sum of $n-1$ i.i.d. random variables $Z_k \sim \mathrm{Ber}(p)$ with finite expectation $\mathbb{E}[Z_k] = p$ and finite variance $\mathrm{Var}(Z_k) = p(1-p) > 0$ (since $p \in (0,1)$), we may apply the De Moivre-Laplace Theorem, a special case of the Central Limit Theorem.

Let $\mu_n = \mathbb{E}[\delta_i^{(n)}] = (n-1)p$ and $\sigma_n^2 = \mathrm{Var}(\delta_i^{(n)}) = (n-1)p(1-p)$. The Cumulative Distribution Function (CDF) of the standardized variable $S_n := (\delta_i^{(n)} - \mu_n)/\sigma_n$ satisfies:

$$\lim_{n\to\infty} \mathbb{P}(S_n \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt, \quad \text{for all } x \in \mathbb{R}.$$

Thus, for large n, the distribution of the degree $\delta_i^{(n)}$ is well-approximated by the Gaussian distribution $\mathcal{N}(\mu_n, \sigma_n^2)$. $\square$

Perhaps a deeper question is whether we can choose $n \mapsto p_{i,n}$ to decay fast enough for all $i \in [n-1]$ such that we are guaranteed to have more than one connected component. We shall investigate this in the next three subsections.

1.1.2 Slow Decay

Theorem 1.1.9 (Connectivity in the Slow Decay Regime)
Let P be a clustering and $X = (X_n)_{n\in\mathbb{N}}$ be the associated canonical graph process. Suppose there exists a constant $\epsilon > 0$ such that for all sufficiently large n, the entries of P satisfy

$$p_{in} \geq (1+\epsilon)\frac{\ln n}{n} \quad \mathbb{P}\text{-a.s.} \quad \text{for all } i \in [n-1].$$

Then, the graph becomes almost surely connected as $n \to \infty$. Specifically,

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : X_n(\omega) \text{ consists of a single connected component}\}) = 1.$$

Proof: We proceed by showing that the probability of any vertex being isolated tends to zero. Since the process starts with $X_1 = (\{v_1\}, \emptyset)$, the graph X_n is connected if and only if every vertex v_k for $2 \leq k \leq n$ connects to at least one predecessor in the set $\{v_1, \ldots, v_{k-1}\}$.

Let D_n denote the event that the vertex v_n fails to connect to any vertex in $V(X_{n-1})$. That is,

$$D_n := \{\omega \in \Omega : \forall i \in [n-1], \{v_i, v_n\} \notin F(X_n)(\omega)\}.$$

Conditioned on the clustering P, the edge formations are independent. Thus, the probability of v_n being isolated is:

$$\mathbb{P}(D_n \mid P) = \prod_{i=1}^{n-1}(1-p_{in}),$$

where we denote $\mathbb{P}(\cdot \mid P) := \mathbb{P}_P(\cdot)$ Using the inequality $1 - x \leq e^{-x}$ for $x \geq 0$, we have

$$\mathbb{P}(D_n \mid P) \leq \exp\left(-\sum_{i=1}^{n-1} p_{in}\right).$$

By the hypothesis, $p_{in} \geq (1+\epsilon)\frac{\ln n}{n}$ $\mathbb{P}$-a.s.. Substituting this lower bound into the summation:

$$\begin{aligned}\sum_{i=1}^{n-1} p_{in} &\geq (n-1)(1+\epsilon)\frac{\ln n}{n} \\ &= (1+\epsilon)\ln n\left(\frac{n-1}{n}\right).\end{aligned}$$

For sufficiently large n, the factor $\frac{n-1}{n}$ can be absorbed into the constant ϵ. Let $\alpha = 1 + \epsilon/2$. Then, for large n:

$$\mathbb{P}(D_n \mid P) \leq \exp(-\alpha \ln n) = n^{-\alpha}.$$

We now consider the event that *any* node v_n (for n larger than some seed size N_0) is isolated. By the union bound:

$$\mathbb{P}\left(\bigcup_{n>N_0} D_n\right) \leq \sum_{n>N_0} \mathbb{P}(D_n) \leq \sum_{n>N_0} n^{-\alpha}.$$

Since $\alpha > 1$, the series $\sum n^{-\alpha}$ converges. By the Borel-Cantelli lemma, the probability that infinitely many vertices are isolated is zero. Consequently, there exists almost surely a finite time after which no new vertex is isolated.

Since the non-isolated vertices join the existing component of X_{n-1}, and the process starts from a single vertex, X_n eventually forms a single connected component $\mathbb{P}$-a.s. □

1.1.3 Fast Decay

Theorem 1.1.10 (Disconnectedness in the Fast Decay Regime)
Let P be a clustering and $X = (X_n)_{n\in\mathbb{N}}$ be the associated canonical graph process. Suppose there exists a constant $\epsilon > 0$ such that for all sufficiently large n, the entries of P satisfy the upper bound

$$p_{in} \leq (1-\epsilon)\frac{\ln n}{n} \quad \mathbb{P}\text{-a.s.} \quad \text{for all } i \in [n-1].$$

Then, the graph fails to be connected almost surely as $n \to \infty$. Specifically,

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \text{ contains isolated vertices infinitely often}\}) = 1,$$

and thus

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : X_n(\omega) \text{ consists of a single connected component}\}) = 0.$$

Proof: We prove this by demonstrating that infinitely many vertices v_n fail to connect to the existing graph upon their arrival. Let D_n be the event that vertex v_n is isolated at time n:

$$D_n := \{\omega \in \Omega : \forall i \in [n-1], \{v_i, v_n\} \notin F(X_n)(\omega)\}.$$

The edge sets $F(X_n)$ for different n are formed independently (conditioned on P). Therefore, the sequence of events $(D_n)_{n\in\mathbb{N}}$ consists of independent events. To apply the Second Borel-Cantelli lemma, we must show that the sum of probabilities $\sum_n \mathbb{P}(D_n \mid P)$ diverges.

The probability of D_n is given by:

$$\mathbb{P}(D_n \mid P) = \prod_{i=1}^{n-1}(1-p_{in}).$$

Using the inequality $1-x \geq e^{-x-x^2}$ (valid for sufficiently small x), or simply noting that for small p_{in}, $1-p_{in} \approx e^{-p_{in}}$, we can bound this from below. However, a standard bound suffices: if $0 \leq p_{in} < 1$, then $\ln(1-p_{in}) \geq -p_{in} - cp_{in}^2$ for some constant. For the sake of the scaling limit where $p_{in} \to 0$, the dominant term is linear. More simply, we use the bound $1-x \geq \exp(\frac{-x}{1-x})$ or asymptotic equivalence. Given the upper bound hypothesis $p_{in} \leq (1-\epsilon)\frac{\ln n}{n}$:

$$\mathbb{P}(D_n \mid P) \geq \prod_{i=1}^{n-1}\left(1-(1-\epsilon)\frac{\ln n}{n}\right).$$

Approximating the product for large n:

$$\begin{aligned}\mathbb{P}(D_n \mid P) &\approx \exp\left(-\sum_{i=1}^{n-1}(1-\epsilon)\frac{\ln n}{n}\right)\\ &= \exp\left(-(n-1)(1-\epsilon)\frac{\ln n}{n}\right)\\ &\approx \exp\left(-(1-\epsilon)\ln n\right)\\ &= n^{-(1-\epsilon)}.\end{aligned}$$

Since $\epsilon > 0$, the exponent $-(1-\epsilon) > -1$. Therefore, the series diverges:

$$\sum_{n=2}^{\infty} \mathbb{P}(D_n \mid P) \approx \sum_{n=2}^{\infty} \frac{1}{n^{1-\epsilon}} = \infty.$$

Since the events D_n are independent and the sum of their probabilities diverges, the Second Borel-Cantelli lemma implies that:

$$\mathbb{P}(D_n \text{ occurs infinitely often}) = 1.$$

This means that as the process continues, new isolated vertices will constantly be added to the system. Since an isolated vertex forms a connected component of size 1 disjoint from the rest of the graph, X_n cannot be a single connected component. □

1.1.4 Critical Regime

Theorem 1.1.11 (Phase Transition and the Giant Component)
Let P be a clustering and $X = (X_n)_{n\in\mathbb{N}}$ the associated canonical graph process. Suppose that for large n, the connection probability scales as:

$$p_{in} \sim \frac{\lambda}{n} \quad \mathbb{P}\text{-a.s.}$$

where $\lambda > 0$ is a constant. The structure of X_n exhibits a phase transition at $\lambda = 1$:

1. **Subcritical Phase ($\lambda < 1$):** If $\lambda < 1$, the graph consists of small, fragmented components. Almost surely, the size of the largest connected component $|C_{\max}(X_n)|$ is sublinear:

$$\frac{|C_{\max}(X_n)|}{n} \xrightarrow{n\to\infty} 0 \quad \mathbb{P}\text{-a.s.}.$$

2. **Supercritical Phase ($\lambda > 1$):** If $\lambda > 1$, a giant component emerges. There exists a constant $\zeta > 0$ such that:

$$\underline{\lim}_{n\to\infty} \frac{|C_{\max}(X_n)|}{n} \geq \zeta \quad \mathbb{P}\text{-a.s.}.$$

Proof: The proof relies on coupling the exploration of a connected component with a Galton-Watson branching process.

Consider a vertex $v \in V(X_n)$ and explore its connected component $C(v)$. For large n, the local neighborhood of v resembles a tree. The formation of edges can be viewed as an offspring production process:

- **Backward connections:** When v_k arrives, it connects to previous nodes with probability $\approx \lambda/k$.
- **Forward connections:** Older nodes receive connections from incoming nodes v_t ($t > k$) with probability $\approx \lambda/t$.

In the asymptotic limit, the expected number of neighbors (degree) for a typical vertex converges to 2λ (assuming symmetric growth) or λ depending on the directed nature. However, for the emergence of the giant component in uniform attachment processes, the critical parameter is the expected number of edges a new node introduces, which is $\sum_{i=1}^{n-1} \frac{\lambda}{n} \approx \lambda$.

Let Z_k denote the number of vertices at distance k from a starting vertex v in a breadth-first search. This sequence behaves asymptotically like a Galton-Watson process with offspring mean $\mu \approx \lambda$.

Case 1: $\lambda < 1$ (Subcritical). The associated branching process is subcritical. The probability that the process survives forever (implying an infinite component in the limit) is 0. The total progeny of a subcritical branching process has a finite expectation. Thus, the components remain small (typically of order $O(\ln n)$), and no component grows linearly with n.

Case 2: $\lambda > 1$ (Supercritical). The branching process is supercritical. There is a strictly positive probability $\rho > 0$ that the branching process does not go extinct (i.e., it grows infinitely). In the context of the finite graph X_n, this corresponds to the component reaching a size comparable to n.

Specifically, the "survival" of the local branching process translates to the vertex v belonging to the Giant Component. Since a non-zero fraction of vertices start such "surviving" trees, these successful explorations merge to form a single giant cluster of size $\Theta(n)$. The fraction of nodes in the giant component converges to the survival probability of the branching process. □

1.1.5 Universal Approximation

In this section, we concern ourselves with a result regarding approximating deterministic graphs with random ones from samples of a CGP. In simple words, it states that if we have a graph G with enough nodes, we can be sure (in a measure-theoretic sense), that G and a sample of a CGP are not that far apart.

Graphons and the Cut Metric

To facilitate the approximation of graphs of varying sizes, we embed the set $\mathcal{G}$ into a larger, continuous space of objects called graphons.

Definition 1.1.12 (Graphon)
A **graphon** is a symmetric, Borel-measurable function $W : [0, 1]^2 \to [0, 1]$. We denote the set of all graphons by $\mathcal{W}$.

To compare a discrete graph $G \in \mathcal{G}$ with a graphon W, we represent G as a step function.

Definition 1.1.13 (Empirical Graphon) Let $G \in \mathcal{G}$ with $|V(G)| = n$. The **empirical graphon** $W_G \in \mathcal{W}$ is defined by partitioning $[0,1]$ into n intervals $I_1, \dots, I_n$ of the form $I_k = [\frac{k-1}{n}, \frac{k}{n})$, and setting:

$$W_G(x,y) := \begin{cases} 1 & \text{if } \{v_{\lceil nx \rceil}, v_{\lceil ny \rceil}\} \in E(G) \\ 0 & \text{otherwise.} \end{cases}$$

Definition 1.1.14 (Cut Metric)
For any $W_1, W_2 \in \mathcal{W}$, the **cut distance** is defined as:

$$d_\square(W_1, W_2) := \sup_{S,T \in \mathcal{B}([0,1])} \left| \int_S \int_T (W_1(x,y) - W_2(x,y))\, dx\, dy \right|.$$

For two graphs $G_1, G_2 \in \mathcal{G}$, we define $d(G_1, G_2) := d_\square(W_{G_1}, W_{G_2})$.

Now that we can compare graphs of different and possibly growing sizes, we are ready to tackle the universal approximation property of deterministic graphs via a random process.

Theorem 1.1.15 (Universal Approximation Theorem)
Let $G \in \mathcal{G}$ be a graph. For every $\epsilon > 0$ and $\eta > 0$, there exists a clustering P and $N \in \mathbb{N}$ such that for all $n \geq N$:

$$\mathbb{P}_P\left(d(G, X_n) < \epsilon\right) > 1 - \eta,$$

where X is a CGP with clustering P.

Proof. Let $k = |V(G)|$. We proceed in four steps.

Step 1: Construction of the Clustering. Define the $k \times k$ adjacency matrix A^G such that $A^G_{ij} = 1$ if $\{v_i, v_j\} \in E(G)$ and 0 otherwise. We define the infinite clustering P by tiling A^G over $\mathbb{N} \times \mathbb{N}$. Specifically, let $p_{ij} := A^G_{(i \pmod k),(j \pmod k)}$ for $i < j$, and $p_{ij} = 0$ for $i \geq j$. This ensures that for any n that is a multiple of k, the expected graphon $\mathbb{E}[W_{X_n}]$ is exactly W_G.

Step 2: Stochastic Representation. Let X_n be the n-th step of the CGP. The edges $e_{ij}(\omega) = \mathbb{1}_{\{\{v_i,v_j\} \in E(X_n)(\omega)\}}$ are independent Bernoulli random variables with parameters p_{ij}. The empirical graphon is $W_{X_n}(x,y) = \sum_{i,j=1}^n e_{ij} \mathbb{1}_{I_i \times I_j}(x,y)$. We note that $\mathbb{E}_P[W_{X_n}(x,y)] = \sum_{i,j=1}^n p_{ij} \mathbb{1}_{I_i \times I_j}(x,y)$. Let this expectation be denoted $W_P^{(n)}$. Note that $d_\square(W_P^{(n)}, W_G) \to 0$ as $n \to \infty$ by the convergence of Riemann sums of step functions.

Step 3: Concentration of the Cut Norm. We must show $d_\square(W_{X_n}, W_P^{(n)}) \to 0$ in probability. For any fixed Borel sets S, T, let $Z_{S,T} = \int_S \int_T (W_{X_n} - W_P^{(n)})\, dxdy$. This is a sum of n^2 independent bounded random variables $\xi_{ij} = \int_{S \cap I_i} \int_{T \cap I_j} (e_{ij} - p_{ij}) dxdy$. By Hoeffding's inequality, $\mathbb{P}(|Z_{S,T}| > \epsilon) \leq 2\exp(-2\epsilon^2 n^2)$. However, the cut metric involves a supremum over an uncountable family of sets. We employ a discretization (net) argument. There are at most 2^n possible unions of intervals I_i. For any S, T, there exist S', T' which are unions of intervals such that $|\int_S \int_T (W_{X_n} - W_P^{(n)}) - \int_{S'} \int_{T'} (W_{X_n} - W_P^{(n)})| < \frac{2}{n}$. Taking a union bound over all $2^n \times 2^n$ such pairs (S', T'):

$$\mathbb{P}_P(\sup_{S,T} |Z_{S,T}| > \epsilon) \leq 2^{2n} \cdot 2\exp(-c\epsilon^2 n^2).$$

For sufficiently large n, the quadratic term in the exponent dominates the linear $2n \ln 2$ term.

Step 4: Conclusion. By the triangle inequality:

$$d_\square(W_G, W_{X_n}) \le d_\square(W_G, W_P^{(n)}) + d_\square(W_P^{(n)}, W_{X_n}).$$

The first term is deterministic and vanishes as $n \to \infty$. The second term vanishes in probability by Step 3. Thus, there exists N such that for $n \ge N$, $\mathbb{P}_P(d_\square(W_G, W_{X_n}) \ge \epsilon) < \eta$, which is the desired claim. $\square$