Gulf Journal of Mathematics ISSN: 2309-4966
Vol 16, Issue 2 (2024) 1-11 https://doi.org/10.56947/gjom.v16i2.1866

WEIGHTED EULERIAN EXTENSIONS OF RANDOM GRAPHS

GHURUMURUHAN GANESAN[1]

ABSTRACT. The Eulerian extension number of any graph H (i.e. the minimum number of edges needed to be added to make H Eulerian) is at least $t(H)$, half the number of odd degree vertices of H. In this paper we consider weighted Eulerian extensions of a random graph G where we add edges of bounded weights and use an iterative probabilistic method to obtain sufficient conditions for the weighted Eulerian extension number of G to grow *linearly* with $t(G)$. We derive our conditions in terms of the average edge probabilities and edge density and also show that bounded extensions are rare by estimating the skewness of a fixed weighted extension. Finally, we briefly describe a decomposition involving Eulerian extensions of G to convert a large dataset into small dissimilar batches.

1. INTRODUCTION

For a deterministic graph H (i.e., for graphs whose edge states are deterministic), [2] studied conditions under which H is a subeulerian graph; i.e. occurs as a spanning subgraph of an Eulerian graph and also classified all subeulerian graphs. Later [9] presented a detailed survey on subgraphs and supergraphs of Eulerian graphs and multigraphs. For applications of Eulerian extensions, we refer to [8] and [5].

Recently, the paper [6] has studied Eulerian edit numbers of random graphs (i.e., for graphs where edge states are random) with uniform edge probabilities, where we show that roughly $\frac{n}{4}$ operations suffices to obtain an Eulerian extension with high probability, i.e. with probability converging to one as $n \to \infty$. The constructed extension is in fact linear, since the number of odd degree vertices is roughly $\frac{n}{2}$ with high probability (for details, we refer to [6]).

In this paper, we equip edges of the complete graph with *random weights* and construct Eulerian extensions of random graphs with edge weights bounded above and below. We obtain sufficient conditions for such "bounded" extendability in terms of the average vertex degrees and the edge density and use an iterative probabilistic method to construct the desired extension. We remark that an arxiv version of our paper is available at [7].

The paper is organized as follows: In Section 2, we study Eulerian extensions of random graphs with bounded edge weights and use the probabilistic method

Date: Received: Feb 6, 2024; Accepted: Feb 26, 2024.
* Corresponding author.
2020 *Mathematics Subject Classification.* Primary 60C05; Secondary 60J35.
Key words and phrases. random graphs, weighted Eulerian extensions, bounded edge weights, skewness, data decomposition.

to obtain the extension. Next in Section 3, we consider a partial converse of bounded extensions and show that any given linear extension has unbounded skewness with high probability, i.e. with probability converging to one as the number of vertices $n \to \infty$. Finally, in Section 4, we use Eulerian extensions to decompose large datasets into small dissimilar batches.

2. Weighted Eulerian Extensions

Let K_n denote the complete graph with vertex set $\{1, 2, \ldots, n\}$ and let $\{X(h)\}_{h \in K_n}$ be independent random variables with indices over the edge set of K_n and satisfying

$$\mathbb{P}(X(h) = 1) = p(h) = 1 - \mathbb{P}(X(h) = 0) \tag{2.1}$$

for every edge h, where $0 \leq p(h) \leq 1$ is a real number that possibly depends on the number of vertices n as well. If h has endvertices u and v, then we also write $p(h) = p(u, v) = p(v, u)$. Let $\{w(h)\}_{h \in K_n}$ be independent and identically distributed (i.i.d.) positive random variables indexed by the edge set of K_n, that are also independent of $\{X(h)\}$. We define $w(h)$ to be the *weight* of edge h.

Let G be the random graph formed by the set of all edges h satisfying $X(h) = 1$ and let $m(G) = \#E$ be the number of edges in G. For a vertex v define $d_G(v)$ to be the degree of vertex v, i.e., the number of vertices adjacent to v in G. Let $\Delta(G)$ denote the maximum vertex degree in G and set $\overline{G}$ to be the complement of G, i.e., an edge $(x, y) \in \overline{G}$ if and only if $(x, y) \notin G$. A walk in G is a sequence of vertices $P := (v_1, v_2, \ldots, v_t)$ such that each v_i is adjacent to v_{i+1}. If in addition v_l is also adjacent to v_1, then P is said to be a *circuit.* If each edge G occurs exactly once in P, then P is an *Eulerian* circuit and G is said to be Eulerian [10]. If P has no repeated vertices, then P is said to be a *path.* Similarly if P is a circuit and no vertex in P is repeated, then P is said to be a *cycle.* We say that G is *connected* if any two vertices of G can be connected by a path and *disconnected* otherwise.

We say that G is *Eulerian extendable* or simply extendable if we can convert G into an Eulerian graph by simply adding edges from the complement. The resulting graph is then said to be an *Eulerian extension* of G. In [6] it is shown that if G_{hom} is the homogenous random graph with edge probability p, then with high probability G_{hom} is extendable and we also obtain estimates on the minimum number of edges needed for such an extension.

We are interested in obtaining Eulerian extensions of G by adding edges whose weight lies in a predetermined interval. Formally, for constants $0 < x \leq y$, we say that G is $(x, y)-$extendable if G can be converted into an Eulerian graph by adding edges each of whose weight lies between x and y. We call the resulting Eulerian extension as an $(x, y)-$extension of G. If G is extendable, then any extension of G has only even degree vertices and so if $N_{ext} = N_{ext}(x, y)$ is the minimum number of edges needed to obtain an $(x, y)-$extension of G, then $N_{ext} \geq t(G)$, where $t(G)$ is *half* the number of odd degree vertices in G.

Letting E_{disc} denote the event that G is disconnected, the following result obtains upper bounds for N_{ext} in terms of the average edge probabilities

$$\alpha_{low} := \min_u \frac{1}{n-1}\sum_{v\neq u} p(u,v), \quad \alpha_{up} := \max_u \frac{1}{n-1}\sum_{v\neq u} p(u,v)$$

and the average edge density $\alpha_e := \binom{n}{2}^{-1}\sum_{1\leq u<v\leq n} p(u,v)$, all of which may depend on n, i.e. $\alpha_{low} = \alpha_{low}(n), \alpha_{up} = \alpha_{up}(n)$ and $\alpha_e = \alpha_e(n)$.

Theorem 2.1. *If*

$$\frac{1}{n^{\beta}} \leq \alpha_{low} \leq \alpha_{up} \leq \max\left(\frac{1}{2}, 1-\sqrt{\frac{\alpha_e(n)}{2}}\right) - \gamma, \tag{2.2}$$

for some constants $0 < \beta < \frac{1}{2}$ *and* $0 < \gamma < \frac{1}{2}$, *then there are constants* $0 < x \leq y < \infty$ *and* $D > 0$ *such that*

$$\mathbb{P}\left(N_{ext}(x,y) \leq 3t(G)\right) \geq 1 - e^{-D(\log n)^2} - \mathbb{P}(E_{disc}). \tag{2.3}$$

Thus if G is connected, then with high probability the number of edges that need to be added to G for obtaining an extension, grows linearly in $t(G)$. The condition $\alpha_{low} \leq \alpha_e \leq \alpha_{up}$ is obtained using the fact that the sum of the vertex degrees of a graph is twice the number of edges and so from (2.2), we see that $\alpha_e \leq 1 - \sqrt{\frac{\alpha_e}{2}} - \gamma$ or equivalently all values of $\alpha_e \leq \frac{1}{2} - O(\gamma)$ are permissible.

Before proving Theorem 2.1, we illustrate the condition (2.2) and the event E_{disc} with an example. Let $0 < b < a < 1$ be two constants, that do not depend on n and set $p(u,v) = 1$ if $1 \leq u < v \leq \frac{n}{\log n}$ or the edge (u,v) belongs to the cycle $(1,2,\ldots,n,1)$. Similarly, set $p(u,v) = 0$ if $\frac{n}{\log n} + 1 \leq u < v \leq \frac{2n}{\log n}$ and for any other edge (n,v) containing n as an endvertex, we set $p(n,v) = a$. For any other edge e, we set $p(e) = b$. For the vertex n, we have that

$$\sum_{1\leq v\leq n-1} p(n,v) = a\,(n-3) + 2 \tag{2.4}$$

and so $\alpha_{up}(n) = a(1+o(1))$, where $o(1) \longrightarrow 0$ as $n \to \infty$. Similarly, by considering the vertex 1 we get that

$$\sum_{2\leq v\leq n} p(1,v) \geq b\left(n - 1 - \frac{n}{\log n}\right) \tag{2.5}$$

and so $\alpha_{low}(n) = b(1+o(1))$.

The number of edges having an edge probability $\geq a$ is at most $\left(\frac{n}{\log n}\right)^2 + 2n$ and so

$$\sum_{1\leq u<v\leq n} p(u,v) \leq \left(\frac{n}{\log n}\right)^2 + 2n + b\binom{n}{2}.$$

Similarly the number of edges for which the edge probability is zero equals $\left(\frac{n}{\log n}\right)^2$ and so

$$\sum_{1 \leq u < v \leq n} p(u,v) \geq b\left(\binom{n}{2} - \left(\frac{n}{\log n}\right)^2\right).$$

Combining the above estimates, we get that $\alpha_e(n) = b \cdot (1 + o(1))$. Finally, by definition, the cycle $(1, 2, \ldots, n, 1)$ on n vertices is a subgraph of G almost surely and so G is connected with probability one.

From the discussions following Theorem 2.1, we see that for every (a, b) satisfying $b < \frac{1}{2}$ and $b < a < 1 - \sqrt{\frac{b}{2}}$, the random graph G has a bounded linear extension in the sense of (2.3).

Below, we use the iterative probabilistic method and the following standard deviation estimate, to prove Theorem 2.1. Let $X_1, \ldots, X_m$ be independent Bernoulli random variables satisfying $\mathbb{P}(X_j = 1) = 1 - \mathbb{P}(X_j = 0) > 0$. If $T_m := \sum_{j=1}^m X_j$ and $\mu_m = \mathbb{E}T_m$, then for any $0 < \epsilon \leq \frac{1}{2}$ we have that

$$\mathbb{P}\left(|T_m - \mu_m| \geq \epsilon \mu_m\right) \leq \exp\left(-\frac{\epsilon^2}{4}\mu_m\right). \tag{2.6}$$

For a proof of (2.6), we refer to Corollary A.1.14, pp. 312 of [1].

Proof of Theorem 2.1: Define an edge (u, v) to be *good* if (u, v) belongs to the complement graph $\overline{G}$ and $a \leq w(u, v) \leq b$. For a good edge (u, v), we say that u is a good neighbour of v and vice versa. Our goal is to obtain an extension of G using good edges and we first consider the case $\alpha_{up} \leq \frac{1}{2} - \gamma$. For vertices $u \neq v$ let

$$Y(u,v) := \sum_{\substack{1 \leq z \leq n \\ z \neq u,v}} (1 - X(u,z))(1 - X(v,z))p_0^2$$

be the number of vertices z such that both the edges (u, z) and (v, z) are good. We have that

$$\begin{aligned}
\mathbb{E}Y(u,v) &= \sum_{\substack{1 \leq z \leq n \\ z \neq u,v}} p_0^2(1 - p(u,z))(1 - p(v,z)) \\
&\geq p_0^2\left(n - \sum_{z \neq u,v} p(u,z) - \sum_{z \neq u,v} p(v,z)\right) \\
&\geq p_0^2\left(n - \sum_{z \neq u} p(u,z) - \sum_{z \neq v} p(v,z) - 2\right) \\
&\geq np_0^2(1 - 2\alpha_{up}) - 2
\end{aligned}$$

Using $\alpha_{up} \leq \frac{1}{2} - \gamma$ we get that $\mathbb{E}Y(u,v) \geq 2np_0^2\gamma - 2$ and so applying the standard deviation estimate (2.6) with $\epsilon := \frac{1}{n^\zeta}$, we get that

$$\mathbb{P}\left(Y(u,v) \geq \gamma p_0^2 n\right) \geq 1 - e^{-Cn^{1-2\zeta}} \longrightarrow 1 \tag{2.7}$$

for some constant $C > 0$, provided $\zeta < \frac{1}{2}$ strictly. We fix such a ζ. Defining $E_{all} := \bigcap_{u \neq v} \{Y(u,v) \geq \gamma p_0^2 n\}$ we get from (2.7) and the union bound that

$$\mathbb{P}(E_{all}) \geq 1 - n^2 \cdot e^{-Cn^{1-2\zeta}} \tag{2.8}$$

for all n large.

Let E_{con} denote the event that G is connected and suppose that the joint event $E_{all} \cap E_{con}$ occurs. Let $T = \{v_1, \ldots, v_{2t}\}, t = t(G)$, be the set of all odd degree vertices of G. For each pair $\{v_{2i-1}, v_{2i}\}, 1 \leq i \leq t$, pick a vertex q_i adjacent to neither v_{2i-1} nor v_{2i} in G. Adding the edges $\{(q_i, v_{2i-1}), (q_i, v_{2i})\}_{1 \leq i \leq t}$ to G, we then get an Eulerian graph H. Each path we have added has exactly two edges and so using (2.8) and the union bound we then get that

$$\mathbb{P}(N_{ext} \geq 2t(G)) \leq \mathbb{P}(E_{all}^c \cup E_{con}^c) \leq n^2 e^{-Cn^{1-\gamma-2\zeta}} + \mathbb{P}(E_{con}^c).$$

Since $E_{con}^c = E_{disc}$, this proves (2.3) for the case $\alpha_e \leq \frac{1}{2} - \frac{1}{n^\gamma}$ in (2.2).

We now consider the remaining condition in (2.2) and for notational simplicity, write $\alpha_{low}(n), \alpha_{up}(n)$ and $\alpha_e(n)$ simply as $\alpha_{low}, \alpha_{up}$ and α_e, respectively. The proof outline is as follows. We first use the conditions in (2.2) to estimate the maximum vertex degree and the number of edges of G. Next, we use the probabilistic method to construct the extension of G *iteratively*.

As before we set $\epsilon = \frac{1}{n^\zeta}$, where ζ is such that $0 < \zeta < \frac{1-\beta}{2}$. Recalling that $d(u)$ is the degree of u we use the deviation estimate (2.6) and (2.2) to get that

$$\mathbb{P}\left(d(u) \leq \alpha_{up}(1+\epsilon)(n-1)\right) \geq 1 - \exp\left(-\frac{\epsilon^2}{4}(n-1)\alpha_{low}\right). \tag{2.9}$$

Since $\alpha_{low} \geq \frac{1}{n^\beta}$ (see (2.2)), we get that $\frac{\epsilon^2}{4}(n-1)\alpha_{low} \geq \frac{1}{8} n^{1-\beta-2\zeta}$ for all n large, by choice of ζ. By the union bound we get that the maximum vertex degree $\Delta(G)$ satisfies

$$\mathbb{P}\left(\Delta(G) \leq \alpha_{up}(1+\epsilon)(n-1)\right) \geq 1 - n \cdot \exp\left(-\frac{1}{8} n^{1-\beta-2\zeta}\right). \tag{2.10}$$

Next we obtain deviation estimates for the number of edges in G. Using (2.2) and the lower bound for α_{low} in (2.2), the expected number of edges in G is

$$\mathbb{E}m(G) = \alpha_e \cdot \binom{n}{2} \geq \alpha_{low} \cdot \binom{n}{2} \geq \frac{n^{2-\beta}}{4}$$

for all n large and so again using the deviation estimate (2.6) we get that

$$\begin{aligned} \mathbb{P}\left(m(G) \leq \alpha_e(1+\epsilon) \cdot \binom{n}{2}\right) &\geq 1 - \exp\left(-C\epsilon^2 \cdot n^{2-\beta}\right) \\ &\geq 1 - \exp\left(-Cn^{2-\beta-2\zeta}\right) \end{aligned} \tag{2.11}$$

which converges to one as $n \to \infty$, again by choice of ϵ.

Defining the event

$$E_{comb} := \{\Delta(G) \leq \alpha_{up}(1+\epsilon)(n-1)\} \bigcap \left\{m(G) \leq \alpha_e(1+\epsilon) \cdot \binom{n}{2}\right\}, \tag{2.12}$$

we get from (2.10), (2.11) and the union bound that

$$\mathbb{P}(E_{comb}) \geq 1 - e^{-n^a} \tag{2.13}$$

for some constant $a > 0$ and all n large.

In our final preliminary computation, we obtain an estimate for the overall number of good edges and the number of good edges per vertex. Given $m(G)$, the overall number of good edges N_{good} is Binomially distributed with parameters p_0 and $\binom{n}{2} - m(G)$ and if E_{comb} occurs, the expected number of good edges is at least $\binom{n}{2}p_0(1-\alpha_e(1+\epsilon))$. Thus using (2.6), we get that

$$\begin{aligned}\mathbb{P}\left(N_{good} \geq \theta\binom{n}{2} \mid E_{comb}\right) &\geq 1 - \exp\left(-C\epsilon^2 n^2 p_0\right)\\ &\geq 1 - \exp\left(-Cp_0 n^{2-2\zeta}\right)\end{aligned} \tag{2.14}$$

for some constant $C > 0$, where $\theta_e := p_0(1-\alpha_e(1+\epsilon))(1-\epsilon)$.

Let $N_{good}(v)$ be the number of good edges containing v as an endvertex. If $d(v)$ is the degree of vertex v in G, then given G, the term $N_{good}(v)$ is Binomially distributed with parameters p_0 and $n-1-d(v)$. Therefore given that E_{comb} occurs, the expected number of good neighbours of v is at least $(n-1)p_0(1-\alpha_{up}(1+\epsilon))$ and so setting $\theta_v := p_0(1-\alpha_{up}(1+\epsilon))(1-\epsilon)$ and arguing as in (2.14), we get

$$\begin{aligned}\mathbb{P}\left(N_{good}(v) \geq \theta_v(n-1) \mid E_{comb}\right) &\geq 1 - \exp\left(-2C\epsilon^2(n-1)p_0\right)\\ &\geq 1 - \exp\left(-Cp_0 n^{1-2\zeta}\right)\end{aligned} \tag{2.15}$$

for some constant $C > 0$, which as before, converges to one as $n \to \infty$ by choice of ζ.

Letting

$$E_{tot} := E_{comb} \bigcap \left\{N_{good} \geq \theta_e \binom{n}{2}\right\} \bigcap_v \{N_{good}(v) \geq \theta_v(n-1)\}$$

we get from (2.13), (2.14), (2.15) and the union bound that

$$\mathbb{P}(E_{tot}) \geq \left(1 - e^{-Cp_0 n^{2-2\zeta}} - ne^{-Cp_0 n^{1-2\zeta}}\right)(1 - e^{-n^a}). \tag{2.16}$$

Letting E_{con} denote the event that G is connected, we assume henceforth that the event $E_{tot} \cap E_{con}$ occurs. Collect all odd degree vertices of G and call it T. Initially all vertices of T are set to be unmarked. Let M be the largest set of vertex disjoint good edges whose endvertices are all in T. Mark all endvertices of edges in M.

Let $\{a_i\}_{1\leq i\leq 2s} \subset T$ be the remaining set of unmarked vertices and set $P := \emptyset$. If there are two vertices a_i, a_j with a common good neighbour $z \notin \bigcup_i \{a_i\}$, then mark a_i and a_j and add the edges (a_i, z) and (a_j, z) to P. Continue this iteratively until we are left with a set $U := \{u_i, v_i\}_{1\leq i\leq t} \subset T$ of unmarked vertices such that any two vertices in U have disjoint good neighbourhoods. Add all the edges of $M \cup P$ to G and call the resultant graph G_{mod}. By construction

$$\Delta(G_{mod}) \leq \Delta(G) + 1 \text{ and } m(G_{mod}) \leq m(G) + 2n, \tag{2.17}$$

where we recall that $m(G)$ is the number of edges of G.

Setting $H_0 := G_{mod}$, we now iteratively convert the odd degree vertices in G_{mod} into even degree vertices by constructing a increasing sequence of graphs $H_0 \subseteq H_1 \subseteq \ldots \subseteq H_t$. In the graph H_i, all the vertices $\{u_j, v_j\}_{1\leq j\leq i} \subseteq U$ would have even degrees and so the graph H_t obtained at the end would then be the desired extension.

Let $\{(Y_i, Z_i)\}_{1\leq i\leq t}$ be independent and identically distributed (i.i.d.) vertex pairs in $\{1, 2, \ldots, n\}^2$ that are also independent of G. Let $X_i := (Y_i, Z_i)$ and let $\mathbb{P}_X$ denote the distribution of $(X_1, \ldots, X_t)$. Also define F_i to be the event that either (u_i, Y_i, Z_i, v_i) or (u_i, Z_i, Y_i, v_i) is a path of good edges in the complement graph $\overline{H}_{i-1}$. If F_i occurs, then we pick exactly one such path and add it to H_{i-1} to obtain the new graph H_i. Otherwise we simply set $H_i = H_{i-1}$.

If $F_{tot} := \bigcap_{i=1}^t F_i$ occurs then the graph H_t obtained at the end of t iterations, is the desired weighted Eulerian extension. Our aim is to show that $\mathbb{P}_X(F_{tot}) > 0$ and we begin by writing $F_i = F_{i,1} \cap F_{i,2}$ where $F_{i,1}$ is the event that (u_i, Y_i) and (Z_i, v_i) are good edges in the complement graph $\overline{H}_{i-1}$ or (u_i, Z_i) and (Y_i, v_i) are good edges in $\overline{H}_{i-1}$ and $F_{i,2}$ is the event that (Y_i, Z_i) is a good edge.

By construction, the degree of the vertices u_i and v_i in H_{i-1} is the *same* as in G and therefore at most $n\alpha_{up}(1+\epsilon)$, since E_{tot} occurs. Thus the $\mathbb{P}_X-$probability that Y_i is a good neighbour of u_i is at least $\theta_v\left(1-\frac{1}{n}\right)$. Since u_i and v_i have disjoint good neighbourhoods in the complement graph $\overline{G}$, we see that

$$\mathbb{P}_X(F_{i,1} \mid X_1, \ldots, X_{i-1}) \geq 2\theta_v^2 \left(1-\frac{1}{n}\right)^2. \tag{2.18}$$

Next, by construction, the number of edges in H_{i-1} is at most

$$m(G_{mod}) + 3i \leq m(G_{mod}) + 3n \leq m(G) + 4n,$$

by (2.17) and the number of edges that belong to $\overline{H}_{i-1}$ but are not good is at most

$$\binom{n}{2} - m(G) - \binom{n}{2}\theta_e = \binom{n}{2}(1-\theta_e) - m(G).$$

Since the probability that $Y_i = Z_i$ equals $\frac{1}{n}$, we get that

$$\begin{aligned}
&\mathbb{P}_X(F_{i,2}^c \mid X_1, \ldots, X_{i-1}) \\
&\quad \leq \frac{1}{n} + \frac{2(m(G)+4n)}{n^2} + \frac{2}{n^2}\left(\binom{n}{2}(1-\theta_e) - m(G)\right) \\
&\quad = \frac{9}{n} + (1-\theta_e)\left(1-\frac{1}{n}\right) \\
&\quad \leq \frac{9}{n} + 1 - \theta_e.
\end{aligned} \tag{2.19}$$

Combining (2.18) and (2.19), we then get that

$$\begin{aligned}\mathbb{P}_X\left(F_i \mid X_1,\ldots,X_{i-1}\right) &\geq 2\theta_v^2\left(1-\frac{1}{n}\right)^2 - 1 + \theta_e - \frac{9}{n} \\ &\geq \theta_{tot} - C\epsilon - \frac{C}{n} \\ &\geq \theta_{tot} - 2C\epsilon, \end{aligned} \tag{2.20}$$

for some constant $C > 0$, where $\theta_{tot} := 2p_0^2(1-\alpha_{up})^2 - 1 + p_0(1-\alpha_e)$ and the final relation in (2.20) is true since $\epsilon = \frac{1}{n^\zeta}$ for some $0 < \zeta < 1$ strictly. Proceeding iteratively with (2.20), we get

$$\mathbb{P}_X(F_{tot}) \geq \left(\theta_{tot} - 2C\epsilon\right)^t. \tag{2.21}$$

From the condition (2.2) in the statement of Theorem 2.1, we see that $2(1-\alpha_{up})^2 - \alpha_e \geq \gamma_0$ for some small constant $\gamma_0 > 0$ and so choosing a small and b large so that p_0 is sufficiently close to 1, we get that $\theta_{tot} \geq \gamma_1$ for some constant $\gamma_1 > 0$. Since $\epsilon = \frac{1}{n^\zeta} \longrightarrow 0$, we get that $\mathbb{P}_X(F_{tot}) > 0$ for all n large. Summarizing, we see that if the event $E_{tot} \cap E_{con}$ occurs, then $N_{ext}(a,b) \leq 3t(G)$. From the estimate (2.16) for the event E_{tot} and the union bound, we obtain (2.3) and this completes the proof of the Theorem. ∎

3. Skewness of Weighted Extensions

The Eulerian extension Γ of the random graph G constructed in the proof of Theorem 2.1 has the following properties:
(i) The number of edges used to obtain the extension of G is at most $r \cdot t(G)$ for some integer constant $r \geq 1$.
(ii) If $\{u_i, v_i\}_{1\leq i\leq t}$ are the odd degree vertices of G, then there are edge disjoint paths P_i with endvertices u_i and v_i, for each $1 \leq i \leq t$.
(iii) Each edge in P_i has weight between a and b and there is at least one edge and at most 3 edges in each P_i. Therefore if we let $W(P_i) := \sum_{e\in P_i} w(e)$ be the total weight of edges in P_i, then the *skewness* defined as

$$\eta(G,\Gamma) := \frac{\max_{1\leq i\leq t} W(P_i)}{\min_{1\leq i\leq t} W(P_i)} \tag{3.1}$$

is bounded above by $\frac{3b}{a}$, with probability converging to one as $n \to \infty$.

The following result shows that extensions like Γ are rare in the sense that if the edge weights are unbounded either from above or from below, then the skewness of any *particular* linear extension of G *grows* with the number of odd degree vertices. Let $E = E(G)$ be the set of all extensions of G satisfying properties $(i) - (ii)$ above.

Theorem 3.1. *Let G be the random graph as in (2.1) and let $H \in E$ be an extension containing at most $r \cdot t$ edges of the complement graph $\overline{G}$, for some constant $r > 0$. Also suppose the cumulative distribution function (cdf) F of the edge weights satisfies*

$$F(0) = 0 \text{ and } F(x) > 0 \text{ for all } x > 0 \quad \text{or} \quad F(x) < 1 \text{ for all } x > 0 \tag{3.2}$$

and the mean $\mu := \mathbb{E}w(e)$ of the edge weights is a constant. There are functions $a(t) \longrightarrow \infty$ and $b(t) \longrightarrow 0$ as $t \to \infty$ such that

$$\mathbb{P}\left(\eta(G,H) \geq a(t)\right) \geq 1 - b(t) \tag{3.3}$$

Thus the skewness $\eta(G,H)$ of any particular extension becomes arbitrarily large with probability converging to one as $t \to \infty$. We remark that the first condition in (3.2) relates to weights that are not bounded from above while the second condition in (3.2) concerns weights that are not bounded from below. In the proof of Theorem 3.1 below, we find explicit expressions for $a(t)$ and $b(t)$.

Proof of Theorem 3.1: Suppose first that the weights are unbounded from above, i.e. $F(x) < 1$ for all $x > 0$. We let $x_t \longrightarrow \infty$ be any sequence such that

$$t(1 - F(x_t)) \longrightarrow \infty \tag{3.4}$$

as $t \to \infty$.

We recall that the walk P_i contains the odd degree vertices u_{2i-1} and u_{2i} as endvertices. Let e_i be an edge of P_i that contains u_{2i} as an endvertex. Using

$$\mathbb{P}(W(P_i) \leq x_t) \leq \mathbb{P}(w(e_i) \leq x_t) = F(x_t)$$

we then get that

$$\mathbb{P}\left(\max_{1\leq i\leq t} W(P)_i \leq x_t\right) \leq F^t(x_t) = (1-(1-F(x_t)))^t \leq e^{-t(1-F(x_t))}. \tag{3.5}$$

To upper bound the minimum weight of the paths $\{P_i\}_{1\leq i\leq t}$, we use the fact that H is a linear extension that contains at most rt edges more than G. Therefore by the pigeonhole principle, there exists a walk P_j that contains at most r edges. The expected weight of P_j is at most μr and so by the Markov inequality

$$\mathbb{P}\left(W(P_j) \geq \sqrt{x_t}\right) \leq \frac{\mu r}{\sqrt{x_t}}. \tag{3.6}$$

Combining (3.5) and (3.6) and using the union bound, we get that

$$\mathbb{P}\left(\zeta(G,H) \geq \sqrt{x_t}\right) \geq 1 - e^{-t(1-F(x_t))} - \frac{\mu r}{\sqrt{x_t}} =: b(t). \tag{3.7}$$

Setting $a(t) := \sqrt{x_t}$, we then get (3.3).

Suppose now that $F(0) = 0$ and $F(x) > 0$ for all $x > 0$. By the right continuity of F, we know that $F(x) \longrightarrow 0$ as $x \downarrow 0$ and so we let $y_t \longrightarrow \infty$ be any sequence such that

$$t \cdot F^{2r}\left(\frac{1}{2ry_t^2}\right) \longrightarrow \infty \tag{3.8}$$

as $t \to \infty$.

Since the total number of edges in the paths $P_i, 1 \leq i \leq t$ is at most rt, at least $\frac{t}{2}$ of these paths have length at most $2r$. For simplicity, we assume that t is even and that each path $P_i, 1 \leq i \leq \frac{t}{2}$ has length at most $2r$. Suppose $Y_1, \ldots, Y_k$ are i.i.d. with cdf F. Using

$$\mathbb{P}\left(\sum_{i=1}^{k} Y_i \leq x\right) \geq \mathbb{P}\left(\bigcap_{1\leq i\leq k}\left\{Y_i \leq \frac{x}{k}\right\}\right) \geq F^k\left(\frac{x}{k}\right) \geq F^{2r}\left(\frac{x}{2r}\right) \tag{3.9}$$

with $x = \frac{1}{y_t^2}$, we get that

$$\mathbb{P}\left(W(P_i) \leq \frac{1}{y_t^2}\right) \geq F^{2r}\left(\frac{1}{2ry_t^2}\right) =: z_t$$

for each $1 \leq i \leq \frac{t}{2}$. Therefore

$$\mathbb{P}\left(\bigcap_{1\leq i\leq t/2}\left\{W(P_i) > \frac{1}{y_t^2}\right\}\right) \leq (1-z_t)^{\frac{t}{2}} \leq \exp\left(-\frac{tz_t}{2}\right) \longrightarrow 0$$

as $t \to \infty$ by (3.8) and consequently,

$$\mathbb{P}\left(\min_{1\leq i\leq t} W(P_i) \leq \frac{1}{y_t^2}\right) \geq 1 - \exp\left(-\frac{tz_t}{2}\right) \tag{3.10}$$

By definition we have for any i that

$$\mathbb{P}\left(W(P_i) \leq \frac{1}{y_t}\right) \leq F\left(\frac{1}{y_t}\right) \longrightarrow 0$$

as $t \to \infty$ and so

$$\mathbb{P}\left(\max_{1\leq i\leq t} W(P_i) > \frac{1}{y_t}\right) \geq 1 - F^t\left(\frac{1}{y_t}\right). \tag{3.11}$$

Combining (3.10) and (3.11) and using the union bound, we get

$$\mathbb{P}\left(\zeta(G,H) \geq y_t\right) \geq 1 - F^t\left(\frac{1}{y_t}\right) - \exp\left(-\frac{tz_t}{2}\right) =: b(t) \tag{3.12}$$

Setting $a(t) := y_t$, we then get (3.3). ∎

4. Data Decomposition

In this section, we briefly discuss data decomposition using Eulerian paths. In many predictive modeling methods particularly involving neural networks, a large dataset is split into small batches that are then fed sequentially to the model as part of the training process (see for e.g. Chapters $5, 9$ [3]). Feeding similar batches successively therefore affects the overall performance of the model and it is of interest to avoid such situations.

Suppose $W(i), 1 \leq i \leq n$, are i.i.d. random elements chosen from some arbitrary space W. We refer to $W(i)$ as the i^{th} data point and further suppose that there is a function $d(W(i), W(j))$ that measures the "distance" between data points $W(i)$ and $W(j)$. Given $\lambda > 0$ let G be the random graph with vertex set $\{1, 2, \ldots, n\}$ where vertices i and j are joined by an edge if $d(W(i), W(j)) > \lambda$.

If we assume that the edges of G are independently present with a common edge probability p, then from [6], we know that with high probability G can be converted into an Eulerian graph H by adding roughly $z = \frac{n}{4}(1+o(1))$ edges $f_1, \ldots, f_z$. Let $(P_1, f_1, P_2, f_2, P_z, f_z, P_{z+1})$ be an Eulerian path in H where each P_i is a walk in G. Say that P_i is a *good* walk if the number of edges in P_i is at least $\frac{np}{8}$. Since the total number of edges in G is roughly $\frac{n^2p}{2}(1+o(1))$ with high probability, the number of good walks is at least $x = \frac{n}{16}$ with high probability.

If $\{P_i\}_{1\leq i\leq x}$ are the good walks, then each P_i contains at least $\frac{np}{8}$ edges (and therefore at least $\frac{\sqrt{np}}{4}$ distinct vertices). Let $Q_i = (b_{i,0}, b_{i,1}, \ldots, b_{i,y})$ be the subpath of P_i consisting of the first $y = \frac{np}{8}$ edges and for $0 \leq j \leq y$, define $R_j := \{b_{i,j}\}_{1\leq i\leq x}$ to be the j^{th} batch of data point indices. By construction, successive batches are dissimilar in the sense that for each i and j, the data points $W(b_{i,j})$ and $W(b_{i,j+1})$ are at a distance of at least λ apart.

Alternately, we could simply set the vertices in Q_i to be the indices of the i^{th} data batch. By definition, any two successive data points within a batch are at least λ apart. Moreover, because of the Eulerian path condition, if (u, v) is an edge in P_i, then the edge (u, v) is not present in any other P_j. Therefore the resulting batches of data are again dissimilar in the sense that if we feed data points W_u and W_v successively in one batch, we would not repeat the same for any other batch.

Acknowledgement. I thank Professors Rahul Roy, Thomas Mountford, Federico Camia, Alberto Gandolfi and C. R. Subramanian for crucial comments and also thank IMSc and IISER Bhopal for my fellowships.

References

1. Alon N. and Spencer, J. (2008). The Probabilistic Method. Wiley. https://doi.org/10.1002/0471722154
2. Boesch, F. T., Suffel, C. and Tindell, R. (1977). The Spanning Subgraph of Eulerian Graphs. Journal of Graph Theory, 1, 79–84. https://doi.org/10.1002/jgt.3190010115
3. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer. https://doi.org/10.1117/1.2819119
4. Bollobás, B. (2001). Random Graphs. Cambridge University Press. https://doi.org/10.1017/CBO9780511814068
5. Dorn, F., Moser, H., Niedermeier, R. and Weller, M. (2013). Efficient Algorithms for extension and Rural Postman Problem. SIAM Journal on Discrete Mathematics, 27, 75–94. https://doi.org/10.1137/110834810
6. Ganesan, G. (2020). Deviation Estimates for Eulerian Edit Numbers of Random Graphs. Statistics and Probability Letters, 171, 1–7. https://doi.org/10.1016/j.spl.2020.109025
7. Ganesan, G. (2023). Linear Eulerian Extensions of Inhomogenous Random Graphs. Arxiv Link: https://arxiv.org/abs/2305.07137
8. Höhn, W., Jacobs, T. and Megow, N. (2012). On Eulerian extensions and Their Application to No-Wait Flowshop Scheduling. Journal of Scheduling, 15, pp. 295–309. https://doi.org/10.1007/s10951-011-0241-1
9. Lesniak, L. and Oellermann, O. R., (1986). An Eulerian Exposition. Journal of Graph Theory, 10, 277–297. https://doi.org/10.1002/jgt.3190100306
10. West, D. B. (2001). Introduction to Graph Theory. *Prentice Hall*.

[1] Department of Mathematics, IISER Bhopal.
Email address: gganesan82@gmail.com