Additional Proofs for Uncertain Butterfly Counting

Alexander Zhou, Yue Wang and Lei Chen

1 UBS is Unbiased [Section 6.1]

We will prove that the vertex-centric variation is unbiased. A similar proof follows for the edge-centric variation.

Recall that C_t is the number of uncertain butterflies in our network. Also recall that $C_t(u)$ is the number of uncertain butterflies containing a sampled node u and that $C_t^e(u) = \frac{C_t(u)|V|}{4}$ is the extrapolated (i.e. estimated) uncertain butterfly count of the network based on u.

Suppose we have random variables $X_1, \dots, X_{C_t}$ where $X_i = 1$ if the ith uncertain butterfly contains u and $X_i = 0$ otherwise. Naturally, $\sum_{i=1}^{C_t} X_i = C_t(u)$. Next, we can trivially see that $\mathbb{E}[X_i] = \frac{4}{|V|}$ given that each uncertain butterfly contains four nodes. Let $\hat{C}_t(u)$ be the estimated uncertain butterfly count for any u. We may derive $\mathbb{E}[\hat{C}_t(u)] = \sum_{i=1}^{C_t} \mathbb{E}[X_i] = \frac{4C_t}{|V|}$. The estimated extrapolated uncertain butterfly count of UBS is $\mathbb{E}[\hat{C}_t^e(u)] = C_t$ and thus our method is unbiased.

2 PES is Unbiased [Section 6.2]

Firstly, the deterministic butterfly counting methods we utilise in PES is known to be unbiased (assuming the proportion we provide is unbiased) [1]. Thus we only need to show that our proportion $\hat{\alpha}$ is unbiased in respect to α.

Let α be the proportion of the $E4$-PDF and $\hat{\alpha}$ be the proportion of any sampled distribution of the $E4$-PDF (including the B-PDF). Let Z_{E4} (Z_S) be the number of all four edge instances in the $E4$-PDF (sampled) distribution. Let X_i be a random variable where $X_i = 1$ if the product of the four edge probabilities represented by the ith four edge instance is greater than t and $X_i = 0$ otherwise. Trivially, $\alpha = \sum_{i=1}^{Z_{E4}} \frac{X_i}{Z_{E4}}$ and $\hat{\alpha} = \sum_{i=1}^{Z_S} \frac{X_i}{Z_S}$. This set-up is effectively a Monte-Carlo sampling environment which is well-known to be unbiased.

3 References

[1] S.-V. Sanei-Mehri, A. E. Sariyuce, and S. Tirthapura. Butterfly counting in bipartite networks. *SIGKDD*, 24:2150–2160, 2018.