1 Statistical Uncertainty

In contrast to deterministic differentially private data generation algorithms such as TraVaS, our GAN- and diffusion-based generators comprise intrinsic randomness. This randomness results from two main structural components: a) Both networks generate synthetic event data from random noise during inference, and b) the network training uses random weight initialization and random noise insertion. To reduce the dependence of our experimental results on a), we first generate large event logs (oversampling with a factor of 100) and then average the final log down to its original size (downsampling). On the contrary, the uncertainty of b) is part of the training procedure and thus cannot be scaled during inference. As a result, to interpret and validate the quality of generated synthetic event data and our quantitative comparison in the main paper, it is important to estimate how this randomness translates into data utility variability.
Subsequently, we show a brief analysis of the variance of *relative log similarity* and *absolute log difference* caused by different random network training runs.

1.1 Experiments

To determine the training-related uncertainty of our data utility measures, we train both TraVaG and the diffusion network 10 times with different noise initialization for all investigated ϵ privacy settings and the BPIC12App-, BPIC13-, and Sepsis event log.

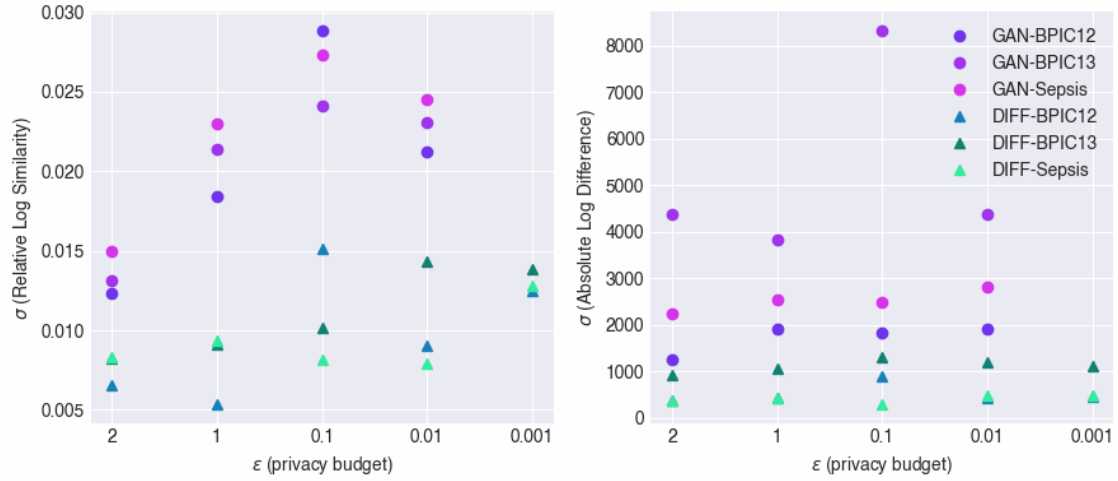

Figure 1.1: Measurements of the standard deviation of both the *relative log similarity* (left) and *absolute log difference* (right) evaluations on trained GAN (GAN-x) and Diffusion (DIFF-x) networks for different ϵ and all three event logs under consideration (see main paper). All data points represent standard deviations from 10 synthetic event logs generated by 10 separately trained networks. δ is kept constant at the median of the available δ parameter set.

Since the major data utility variation appears for different ϵ rather than for different δ budgets (see paper), all training cycles are conducted at the median δ of the available δ range at a given ϵ. As an example, at $\epsilon = 1$ for the BPIC12App log, the centric $\delta = 0.0001$ is used.[1] After training, we utilize the resulting different generators to create synthetic

[1] According to first tests, the uncertainty for different δ settings at a fixed ϵ remained rather constant, supporting our hypothesis and experimental setup.

event logs and compute the standard deviation of *relative log similarity* and *absolute log difference*.

The experimental results are shown in Figure 1.1. For both metrics, the GAN of TraVaG generates more significant randomness than the diffusion algorithm, which could be explained by the generally more stable training procedure. Moreover, particularly for *relative log similarity*, the uncertainty seems to increase with stronger differential privacy. We relate this effect to the greater importance of introduced noise during training. Finally, when comparing the standard deviation estimates with the data utility results (see paper), even the largest contributions are small enough to still validate the major systematic trends in privacy-related utility deterioration.

1.2 Note of Authorship

This document is part of the supplementary material created for the paper *Releasing Differentially Private Event Logs Using Generative Models*, written by Frederik Wangelik, Majid Rafiei, Mahsa Pourbafrani and Wil M.P. Van der Aalst. Please contact *frederik.wangelik@rwth-aachen.de* for further information.