A PROOF OF PROPOSITION 4.1

Since $f_{sd} \sim \mathbb{P}(f;\lambda_{sd})$ where $\mathbb{P}(f;\lambda_{sd})$ is a Gamma distribution that follows $\mathbb{P}(f;\lambda) = \frac{1}{\lambda^\theta \Gamma(\theta)} f^{\theta-1} \exp(-\frac{f}{\lambda})$, $\mathbb{E}[f_{uv}]$ can be represented as $\mathbb{E}[f_{uv}] = \theta\lambda_{uv}$. Recall that we parameterize $\lambda_{sd} = \exp(h_s^T Q h_d)$ for all $(s.d)$ pairs. We suppose that h_v lies in the convex hull of the embeddings of a group of nodes $C = \{v_1, v_2, ..., v_k\}$, i.e., $h_v \sum_{i=1}^k a_i h_{v_i}$ for non-negative and ℓ-1 normalized $\{a_i\}_{i=1}^k$. Then, the upper bound of $\mathbb{E}[f_{uv}]$ can be derived as follows:

$$
\begin{aligned}
\frac{\mathbb{E}[f_{uv}]}{\theta} &= \lambda_{uv} = \exp(h_u^T Q h_v) = \exp\left(h_u^T Q(\sum_{i=1}^k a_i h_{v_i})\right) \\
&\leq \sum_{i=1}^k a_i \exp(h_u^T Q h_{v_i}) = \sum_{i=1}^k a_i \lambda_{uv_i} = \sum_{i=1}^k a_i \frac{\mathbb{E}[f_{uv_i}]}{\theta},
\end{aligned}
$$

where the inequality is due to the convexity of the exponential function. Regarding the lower bound of $\mathbb{E}[f_{uv}]$, we can derive it as follows:

$$
\begin{aligned}
&\frac{\mathbb{E}[f_{uv}]}{\theta} = \lambda_{uv} = \exp(h_s^T Q h_d) = \exp\left(h_u^T Q(\sum_{i=1}^k a_i h_{v_i})\right) \\
&= \exp\left(\sum_{i=1}^k a_i (h_u^T Q h_{v_i})\right) \geq \exp\left(\sum_{i=1}^k a_i \min_{1\leq i\leq k} h_u^T Q h_{v_i}\right) \\
&= \exp\left(\min_{1\leq i\leq k} h_u^T Q h_{v_i}\right) = \min_{1\leq i\leq k} \exp(h_u^T Q h_{v_i}) = \min_{1\leq i\leq k} \lambda_{uv_i} \\
&= \min_{1\leq i\leq k} \frac{\mathbb{E}[f_{uv_i}]}{\theta},
\end{aligned}
$$

where the inequality is due to non-negativity of $\{a_i\}_{i=1}^k$ and we also used $\sum_{i=1}^k a_i = 1$.

By combining the upper bound and the lower bound of $\mathbb{E}[f_{uv}]$, we prove that $\mathbb{E}[f_{uv}]$ is controlled by $\min_{1\leq i\leq k} \mathbb{E}[f_{uv_i}] \leq \mathbb{E}[f_{uv}] \leq \sum_{i=1}^k a_i \mathbb{E}[f_{uv_i}]$. The inequality of $\mathbb{E}[f_{vu}]$ can be derived similarly in the same way.

B PROOF OF PROPOSITION 4.3

As for the assumption, $\mathbb{P}(f;\lambda)$ (Eq. (2)) matches the distribution of the regular frequency. Then, for an interaction that is not anomaly, its observed frequency should follow $f^{(o)} \sim \mathbb{P}(f;\lambda)$. Our method will detect it as anomaly if, according to Eq. (4),

$$
-\frac{f^{(o)}}{\lambda} \leq \tau.
$$

Then, the false positive rate is nothing but the probability such that the above inequality is satisfied. That is, when $f^{(o)} \sim \mathbb{P}(f;\lambda)$,

$$
\mathbb{P}\left(-\frac{f^{(o)}}{\lambda} \leq \tau\right) = \mathbb{P}\left(f^{(o)} \geq -\lambda\tau\right) = \exp\left(-\frac{-\lambda\tau}{\lambda}\right) = \exp(\tau),
$$

which concludes the proof.

C TRAINING CONFIGURATION

We performed hyper parameter search for best performance for our method and all the baselines and used the following hyper-parameters to obtain the reported results:

For RTM graph, DARPA, and BARRA1-4, we setup all the models based on the data in the first 10% total time. Table 4 lists the hyperparamters and their values. The unit of W_{upd} is year for DBLP and minute for others, and the unit of initial f_{th} is 10^{-3}minute^{-1}.

For baselines, we used the implementations provided by their authors and we report the range of configurations used for baselines here:

SedanSpot: *numwalks* = $\{5, 10, 20, 50, 100\}$, *restart_prob* = $\{0.05, 0.1, 0.15, 0.2, 0.5\}$, *sample_size* = $\{50, 100, 200, 500, 1000\}$ on synthetic graphs, and *sample_size* = $\{10K, 20K, 50K\}$ on DARPA and BARRACUDA, and following hyper-parameter settings as suggested in the original paper on ENRON and DBLP.

AnomRank: *aggregation_timesteps* = $\{10, 30, 60, 180, 360, 720, 1440\}$ minutes on synthetic graphs, DARPA, and BARRACUDA.

NetWalk: *representation_size* = $\{20, 50, 100\}$, *num_walks* = $\{2, 3, 5\}$, *walk_length* = $\{3, 5, 10\}$, $\rho = \{0.1, 0.2, 0.3\}$, $k = \{5, 10, 20\}$, $\lambda = \{0.0005, 0.001, 0.005\}$, $\beta = \{0.1, 0.2, 0.5\}$, $\gamma = \{1, 5, 10\}$, $\alpha = \{0.3, 0.5, 0.7\}$, *snap_size* = $\{500, 1000, 2000\}$ for synthetic graphs. *embedding_size* = $\{20, 50, 100\}$, $\alpha = \{0.3, 0.5, 0.7\}$, $k = \{5, 10, 20\}$, *snap_size* = $\{250K, 500K\}$, and following the other hyper parameter settings as suggested in the original paper for real-word graphs on DARPA. *learning_rate* = 0.01 for adam optimizer as suggested in the public source code.

Midas: *decay_factor* = $\{0.3, 0.5, 0.7\}$, *num_hash* = $\{2, 5, 10\}$, *num_buckets* = $\{500, 1000, 2000, 5000\}$ on RTM graph, DARPA, and BARRACUDA. *decay_factor* = 0.5, *num_hash* = 10, and *num_buckets* = 5000 for ENRON and DBLP.

D EVENTS DETECTION IN ENRON

In this section we demonstrate the effectiveness of F-FADE on ENRON dataset in the main paper 5.3. The anomalies detected by F-FADE coincide with major events in the ENRON timeline [5] as follows:

(1) 05/22/2000: The California ISO (Independent System Operator), the organization in charge of California's electricity supply and demand, declares a Stage One Emergency, warning of low power reserves.
(2) 06/12/2000: Skilling makes joke at Las Vegas conference, comparing California to the *Titanic*.
(3) 11/01/2000: FERC investigation exonerates Enron for any wrongdoing in California.
(4) 03/2001: Enron transfers large portions of EES business into wholesale to hide EES losses.
(5) 07/13/2001: Skilling announces desire to resign to Lay. Lay asks Skilling to take the weekend and think it over.
(6) 10/17/2001: Wall Street Journal article reveals the details of Fastow's partnerships and shows the precarious nature of Enron's business.
(7) 11/08/2001: Enron files documents with SEC revising its financial statements for past five years to account for 586 million in losses. The company starts negotiations to sell itself to Dynegy, a smaller rival, to head off bankruptcy,
(8) 01/25/2002: Cliff Baxter, former Enron vice chairman, commits suicide.

[5] https://www.agsm.edu.au/bobm/teaching/BE/Enron/timeline.html

Dataset	W_{upd}	α	M	m	initial f_{th}
RTM-InjectionS	5, 10, 20, 40, 80, 160	0.999	100, 300, 500, 700, 1000	25, 50, 100	0.77
RTM-InjectionW	5, 10, 20, 40, 80, 160	0.999	100, 300, 500, 700, 1000	25, 50, 100	3.13
DARPA	60, 120, 360, 720, 1440	0.999	100, 200, 500, 1000, 2000	100, 150, 200	16.7
ENRON	10080	0.999	∞	100	0
DBLP	1	0.999	∞	100	0
BARRA1	10080, 21600, 43200	0.999	100, 200, 400, 800, 1600, 3200	100, 150, 200	2.6
BARRA2	10080, 21600, 43200	0.999	100, 200, 400, 800, 1600, 3200	100, 150, 200	0.93
BARRA3	10080, 21600, 43200	0.999	100, 200, 400, 800, 1600, 3200	100, 150, 200	1.2
BARRA4	10080, 21600, 43200	0.999	100, 200, 400, 800, 1600, 3200	100, 150, 200	1.1

Table 4: Hyperparameters and their value for F-FADE on different dataset

(9) 02/02/2002: The Powers Report, a 218-page summary of an internal investigation into Enron's collapse led by University of Texas School of Law Dean William Powers, spreads blame among self-dealing executives and negligent directors.
(10) 03/14/2002: Former Enron auditor Arthur Andersen LLP indicted for obstruction of justice for destroying tons of Enron-related documents as the SEC began investigating the energy company's finances in October 2001.

E EVENTS DETECTION IN DBLP

In this section we demonstrate the effectiveness of F-FADE on DBLP dataset in the main paper 5.3. We expect anomalous edges to represent unlikely collaborations. We verify anomalies using the public profiles of the authors as follows:

(1) 1988: G. M. Lathrop and J. M. Lalouel have 15 coauthor papers, but they don't have any coauthor paper before.
(2) 1998: Raj Jain has 63 papers this year, and he has 32 coauthor papers with Rohit Goyal and Sonia Fahmy. But Raj Jain has only 5 papers in 1997, and 4 of them are coauthor papers with Rohit Goyal and Sonia Fahmy.
(3) 2005: Elizabeth Dykstra-Erickson and Jonothan Arnowitz have 25 coauthor papers in this year. But before 2005, they have only 1 coauthor paper in 2003.
(4) 2007: Damien Chablat and Philippe Wenger have 61 coauthor papers, but they only have 1 coauthor paper in 2006.
(5) 2010: Alan Dearle and Graham N. C. Kirby have 27 coauthor papers. But before 2010, they have most 3 coauthor papers in 2003.