A Proofs

A.1 The following lemma states that the information retrieved via ParentTransitions monitoring operations decrease the expected uncertainty about the positioning of items after the transition step.

Lemma A.1.

$$\mathbf{F}_{PT}(S) \leq F_0$$

Proof. Indeed, using Equations (3.4) and (3.5) to expand Equation (3.3) we have

$$\begin{aligned}
\mathbf{F}_{PT}(S) &= \sum_{u \in V} \mathbf{x}'(u) \sum_{v \in V \setminus S} \mathbf{P}'(u,v)\,(1 - \mathbf{P}'(u,v)) \\
&= \sum_{u \in V} \mathbf{x}(u) \sum_{v \in V \setminus S} \mathbf{P}(u,v) \left(1 - \frac{\mathbf{P}(u,v)}{1 - \rho(u,S)}\right) \\
&\leq \sum_{u \in V} \mathbf{x}(u) \sum_{v \in V \setminus S} \mathbf{P}(u,v)\,(1 - \mathbf{P}(u,v)) \\
&= \mathbf{F}_{PT}(\emptyset).
\end{aligned}$$

A.2 The following theorem states that, for the same set of monitored nodes, ParentTransitions and NodeItems lead to the same value of the objective function.

Theorem A.1. $\mathbf{F}_{NI}(S) = \mathbf{F}_{PT}(S)$

Proof. We express $F(A_{NI})$ in terms of $F(A_{PT})$ as follows. (We write 'c.w.' for 'consistent with').

$$F(A_{NI}) = \sum_{A_{PT} \text{ c.w. } A_{NI}} F(A_{PT})\mathbf{Pr}(A_{PT}|A_{NI}) \tag{A.1}$$

We can now use the above equation to express the expected uncertainty $\mathbf{F}_{NI}(S)$ in terms of $\mathbf{F}_{PT}(S)$ as:

$$\begin{aligned}
\mathbf{F}_{NI}(S) &= E[F(A_{NI})] \\
&= \sum_{A_{NI}} F(A_{NI})\mathbf{Pr}(A_{NI}) = \\
&= \sum_{A_{NI}} \sum_{A_{PT} \text{ c.w. } A_{NI}} F(A_{PT})\mathbf{Pr}(A_{PT}|A_{NI})\mathbf{Pr}(A_{NI}) \\
&= \sum_{A_{NI}} \sum_{A_{PT} \text{ c.w. } A_{NI}} F(A_{PT})\mathbf{Pr}(A_{PT}, A_{NI}) \\
&= \sum_{A_{NI}} \sum_{A_{PT} \text{ c.w. } A_{NI}} F(A_{PT})\mathbf{Pr}(A_{PT}) \\
&= \sum_{A_{PT}} F(A_{PT}) \cdot \mathbf{Pr}(A_{PT}) \\
&= \mathbf{F}_{PT}(S),
\end{aligned} \tag{A.2}$$

which concludes the proof.

A.3 The following lemma states that, with choice restricted among the outgoing edges of a node, the optimal objective value in the EdgeTransitions setting is obtained for the edges of highest transition probability.

Lemma A.2.

$$ISOL_i(m) = \mathbf{F}_i(D_i^m) \tag{A.3}$$

Proof. The optimization function is proportional to the following quantity:

$$f(E) \propto \Big(\sum_{i \in D_u(E)} p_i\Big) - \sum_{i \in D_u(E)} p_i^2 \Big/ \Big(\sum_{i \in D_u(E)} p_i\Big) \tag{A.4}$$

where $D_u(E)$ are the remaining (i.e., non-queried) outgoing edges of parent-node u.

Consider two sets of edges $E_0,\ E_1 \subseteq O(u)$ of the same size, all outgoing from a single parent-node u, that differ only at one element. The probabilities of the corresponding sets of **remaining** edges are:

$$D_u(E_0) : \{p_0\} \cup C; \quad D_u(E_1) : \{p_1\} \cup C \tag{A.5}$$

where $p_0, p_1 \notin C$, $p_0 \leq p_1$.

Let $S = \sum_{i \in C} p_i$ and $SS = \sum_{i \in C} p_i^2$. We take the difference of the optimization functions for the two sets E_0 and E_1.

$$\begin{aligned}
f(E_0) - f(E_1) &\propto p_0 - p_1 - \frac{\sum_{i \in D_u(E_0)} p_i^2}{\sum_{i \in D_u(E_0)} p_i} + \frac{\sum_{i \in D_u(E_1)} p_i^2}{\sum_{i \in D_u(E_1)} p_i} \\
&= -(p_1 - p_0)\frac{SS + S^2}{(S + p_0)(S + p_1)} \leq 0.
\end{aligned}$$

The above shows that selecting the set of edges so that the remaining edges are associated with smaller probabilities leads to lower (better) values of the optimization function.

A.4 The following theorem concerns the optimality of algorithm `EdgeDP`.

Theorem A.2. *`EdgeDP` is optimal for the* EdgeTransitions *variant of the* Markov Chain Monitoring *problem.*

Proof. The proof follows from Lemma A.2 and by construction of the dynamic programming algorithm (Equation (6.17)).

B Additional Results

Figure 3 shows the performance of the `NodeGreedy` algorithm for the the `Geo` graphs, with each plot corresponding to a different item distribution schemes.

Observe that NodeGreedy significantly outperforms all other baselines, which capture different semantics of centrality. In particular, we observe that NodeGreedy achieves zero or near-zero expected uncertainty with a small fraction of selected nodes compared to baselines. Among the baselines, Closeness performs second-best in many cases, while In-Degree performs as well as Closeness for small k.

Similarly, Figure 4 shows the performance of the different algorithms for the EDGE-MONITORING and the Geo graphs, for all possible item-distribution schemes. As before, we observe that EdgeGreedy outperforms the baselines in all cases. We notice also that the pattern of performance differs somewhat for the case of Ego item distribution. With the exception of one baseline (Probability), all algorithms achieve steep decline in expected uncertainty for small value of k - EdgeGreedy performs best, but baselines are competitive. However, for larger k, the performance of baselines does not keep up with that of EdgeGreedy. We believe that this is can be explained as follows: the first edges selected by baselines are either central in terms of graph structure – and therefore near the part of the graph with high concentration of items (Edge-Betweenness) – or directly in the area of the graph with many items (Edge-NumItems). In terms of reducing expected uncertainty, this is beneficial at first. However, these baselines as they do not optimize our objective are not able to continue reducing the expected uncertainty with their subsequent selections.

Figure 5 and Figure 6 show the performance of the greedy algorithms on the NODE-MONITORING and the EDGE-MONITORING problems respectively. We observe that both the NodeGreedy and the EdgeGreedy algorithms are consistently the best when compared to the baselines. However, $k = 50$ represents about 1% of the total edges in the graph, hence their monitoring does not decrease the uncertainty significantly. While experiments with larger values of k are prohibitive due to time complexity of the EdgeGreedy algorithm, we postulate that the greedy algorithm will still continue outperforming the baselines.

Figures 7 and 8 provide a similar comparison for the different configurations of the BA graph. The greedy algorithms provide marginal benefits or perform on par with competitive baselines. On the BA graphs, for Direct, Uniform and Inverse item distributions, some baselines perform exactly the same as the greedy algorithms for relatively small number of monitoring operations i.e., $k = 50$. Lastly, we observe similar trends in case of the Grid graphs as evident in Figures 9 and 10. It should be noted that there is no baseline method that provides a consistently competitive performance with the greedy algorithms across all different configurations described above.

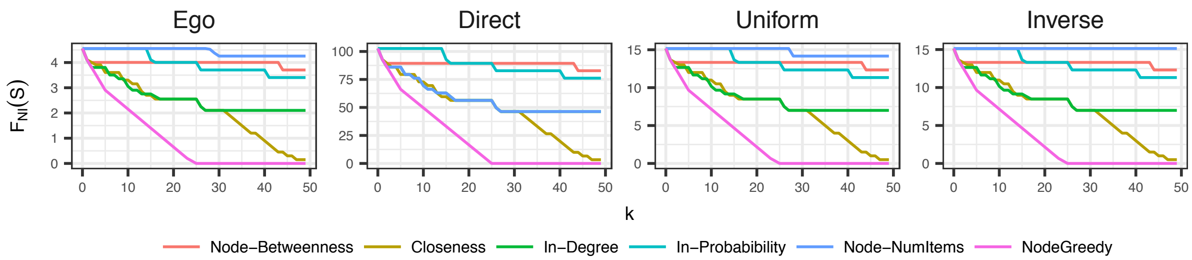

Figure 3: NODE-MONITORING `Geo` dataset; y-axis expected uncertainty, x-axis: number of monitored nodes (k).

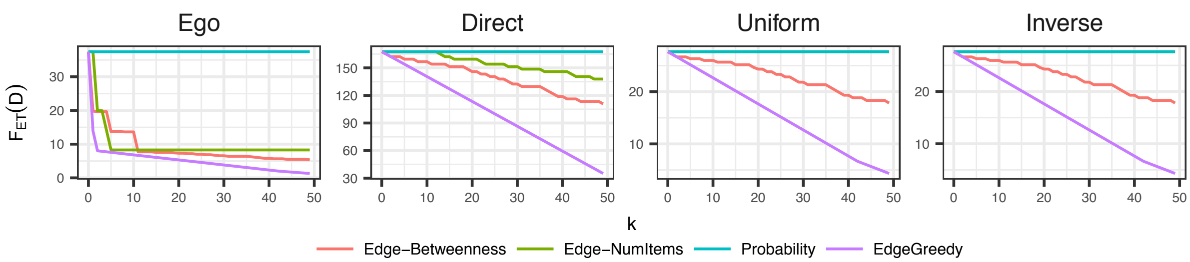

Figure 4: EDGE-MONITORING `Geo` dataset; y-axis expected uncertainty, x-axis: number of monitored edges (k).

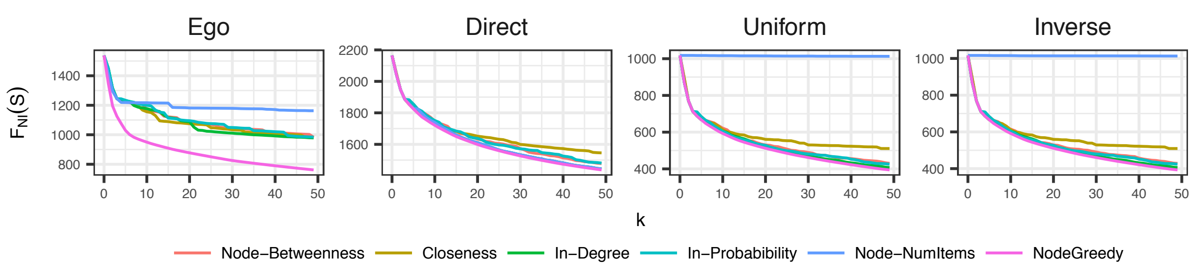

Figure 5: NODE-MONITORING `AS` dataset; y-axis expected uncertainty, x-axis: number of monitored nodes (k).

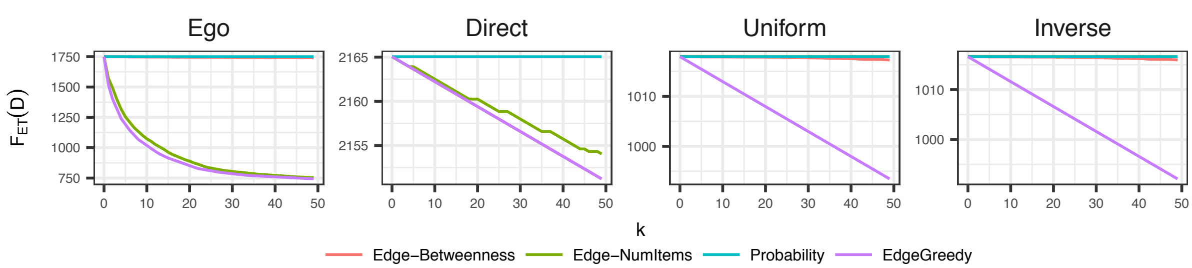

Figure 6: EDGE-MONITORING `AS` dataset; y-axis expected uncertainty, x-axis: number of monitored edges (k).

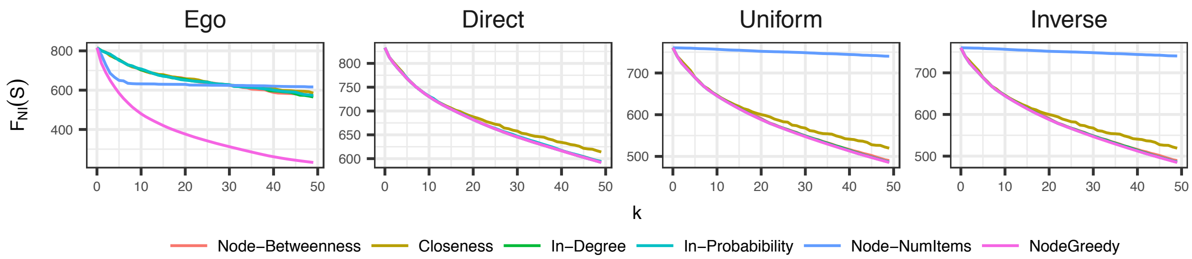

Figure 7: NODE-MONITORING `BA` dataset; y-axis expected uncertainty, x-axis: number of monitored nodes (k).

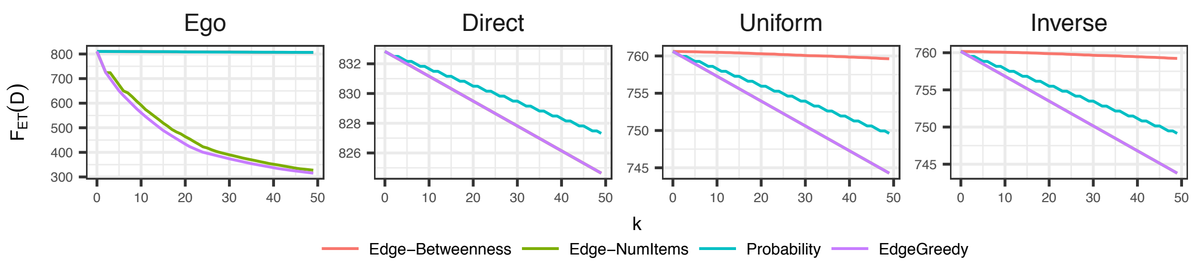

Figure 8: EDGE-MONITORING `BA` dataset; y-axis expected uncertainty, x-axis: number of monitored edges (k).

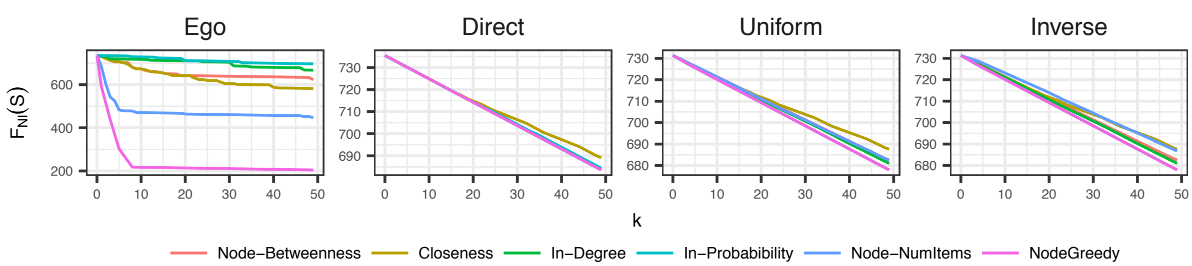

Figure 9: NODE-MONITORING `Grid` dataset; y-axis expected uncertainty, x-axis: number of monitored nodes (k).

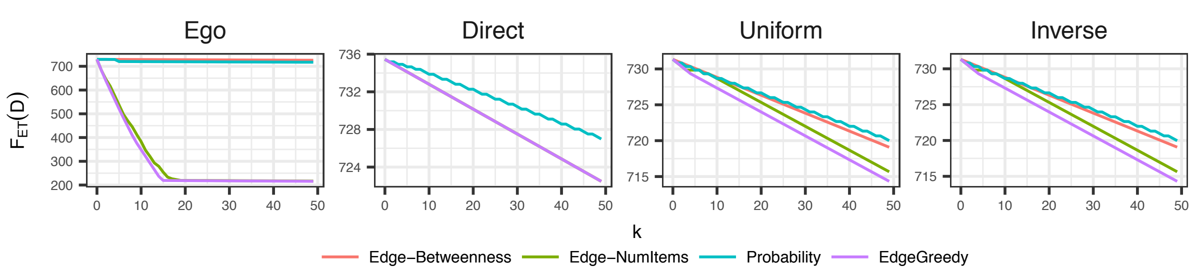

Figure 10: EDGE-MONITORING `Grid` dataset; y-axis expected uncertainty, x-axis: number of monitored edges (k).