

Interconnection: from Dynamic Graphs to Social Networks

Project: Information Persistence in Dynamic Graphs

Anna Kelm, Piotr Plebanski
110455, 121309

January 16, 2023

Contents

1	Introduction	2
2	Experimental	2
2.1	Graph parameter ranges	2
2.2	The transmission of information	2
2.3	Experimental procedure	3
2.4	Recorded graph characteristics	3
3	Results and Discussion	3
3.1	The graphs characteristics	3
3.2	The propagation and persistence of information	4

1 Introduction

This project aims at investigation the conditions influencing the information persistence within a dynamic graph. Three dynamic graph models were selected: Edge Markovian, Random Waypoint (RWP) and Manhattan model. While a set of nodes in each graph remained constant, the set of edges evolved according to the graph model which influenced the transmission of the information with a fixed lifetime.

2 Experimental

2.1 Graph parameter ranges

Environment settings and parameter ranges common for all models:

- number of experiments repetitions: 15
- maximum number of iterations: 1000
- environment size: 1000x1000
- number of nodes, n: {50, 100, 200, 300}

Parameters for geometric graph models (Manhattan and RWP)

- distance threshold, d: {20, 30, 50, 70, 80, 100, 120}

Parameters for Manhattan model:

- number of parallel streets is in each direction: 10

Parameters for Edge Markovian model:

- parameters for Edge Markovian model, p and q , are probabilities of nodes staying connected or unconnected, respectively, as shown in Fig 1. A $p+q = 1$, this Edge Markovian model corresponds to Erdos-Renyi graph model. The selected values of parameters p correspond to mean graph density in RWP model with distance threshold d (also known as excluded volume). That is:[1]:

$$p_i = \frac{4\pi \cdot d_i^2}{l^2} \quad (1)$$

where $l=1000$ denotes the size of the environment. Calculated values of p_i are { 0.001, 0.003, 0.008, 0.0015, 0.02, 0.031, 0.045 }.

2.2 The transmission of information

In each graph, at the start of each experiment one was provided with the information of a fixed time-to-live (TTL). The possible values of TTL were: {1, 2, 3, 4, 5} iterations. In each iteration, each node X with the information broadcasted it to the nodes $\{Y_i\}$ that had a common edge with X. If a node $\{Y_i\}$ does not have the information, it receives its copy with a full lifetime (equal to the starting lifetime). After every iteration, the current information TTLs in the nodes decrease by 1. A node with information TTL=0 does not carry the information. Each experiment was terminated when either the information was no longer present in the whole graph or the number of iterations reach 1000.

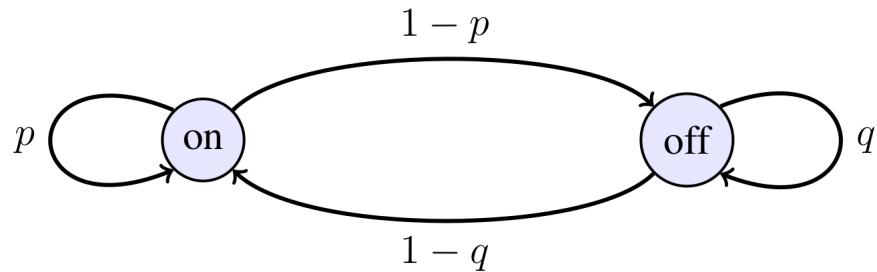


Figure 1: Automata regulating the state of an edge in an edge-markovian dynamic graph.[source: project description by F. Guinand]

2.3 Experimental procedure

Experiments were performed with the use of GraphStream library[2], version 1.3. While the graph simulation was performed in Java with GraphStream, the experiment parameter settings, collection of the results and plotting were performed in Python 3.10. The *.jar package containing the simulations was invoked from Python with the use of pyjnius Python library. The details of the experimental setup and source code can be found at: github.com/akelm/interconnection_final_project.

For each type of graph model, there were $4 \cdot 7 \cdot 5 = 140$ possible parameter settings and each experiment with given experimental settings was run 15 times for the statistics. In total, for all graphs, 6300 experiments were run.

Notably, it needs to be mentioned that experiments for Edge Markovian graph were performed NOT with a GraphStream library, but with a simple representation of graph using connectivity matrix. The reason behind this decision was the poor speed of the simulation, most probably due high edge nervousness.

2.4 Recorded graph characteristics

Graph characteristics collected during experiments in each iteration:

- number of edges,
- number of nodes with information,
- edge nervousness, defined as a Jaccard distance between node sets in i th and $i = 1$ iterations,
- number of connected components.

Calculated after the experiments:

- success rate at preserving information to the 1000th iteration.

3 Results and Discussion

3.1 The graphs characteristics

Fig 2-4 present the graphs characteristics in all three graph models for different n and d/p , averaged over 15 repeated experiments.

The graph density, as shown in Fig 2, is almost constant for the Edge-Markovian graph, while it changes by up to 100% in the course of 1000 iterations. The graph density grows

approximately quadratically with the distance threshold (or p in Edge Markovian graph) and seems independent of the node density. However, the considerable noise in the graph density, decreasing with increasing n , seems to indicate higher the temporal graph density oscillations for lower number of nodes.

The number of connected components divided by n (Fig 3) spans from $1/n$ in the connected graph to 1 for a graph with 0 edges. This measure is in a close relationship with graph density – the denser the graph, the lower number of connected components it owns. Contrary to the graph density, the number of connected components (divided by n) depends also on n – it decreases with n .

Fig 4 shows the time traces of edge nervousness in all three graph models for different n and d or p . Edge nervousness is very high in Edge Markovian graph and close to 0 for geometric graph models. This probably originates from the fact, that in an Edge Markovian graphs, edges disappear stochastically with high probability – meaning that the edge formed in $i - 1$ th iteration will almost certainly disappear in the i th iteration. Meanwhile, in geometric graphs, nodes need to move away from each other in order to for the edge to be removed, which takes more time. In all graph models, the nervousness lowers with d/p and is very noisy in graph with low n .

3.2 The propagation and persistence of information

In all of the experiments, the fraction of nodes with information in the Edge Markovian model (Fig 8) grows rapidly in ≤ 50 iterations, after it reaches an equilibrium value which stays constant (modulo some noise) until the 100th iterations. If the equilibrium value is comparable or lower to its temporal variations, the information has a high chance to be lost due to the high probability of the fraction of nodes with information reaching 0. These short-term temporal variations in Edge Markovian graph are considerably stronger than in the case of geometric graph, most probably due to very high nervousness. Reaching the equilibrium value of fraction of nodes with information means that equal number of nodes loses and gains information. The values of equilibrium fraction of nodes with information, p_{info} can be estimated by balancing the probability of receiving information by a node:

$$p_{rec} = (1 - p_{info}) \cdot (1 - (1 - p + p \cdot (1 - p_{info}))^{n-1}) \approx (1 - p_{info}) \cdot p \cdot p_{info} \cdot (n - 1) \quad (2)$$

with a probability that a node looses the information:

$$p_{loss} = p_{info}/TTL \quad (3)$$

which gives:

$$p_{info} = \frac{TTL \cdot p \cdot (n - 1) - 1}{TTL \cdot p \cdot (n - 1)} \quad (4)$$

The plot of the values (Fig 5) calculated the aforementioned way yields values close to those obtained in the experiments, further supporting the explanations above.

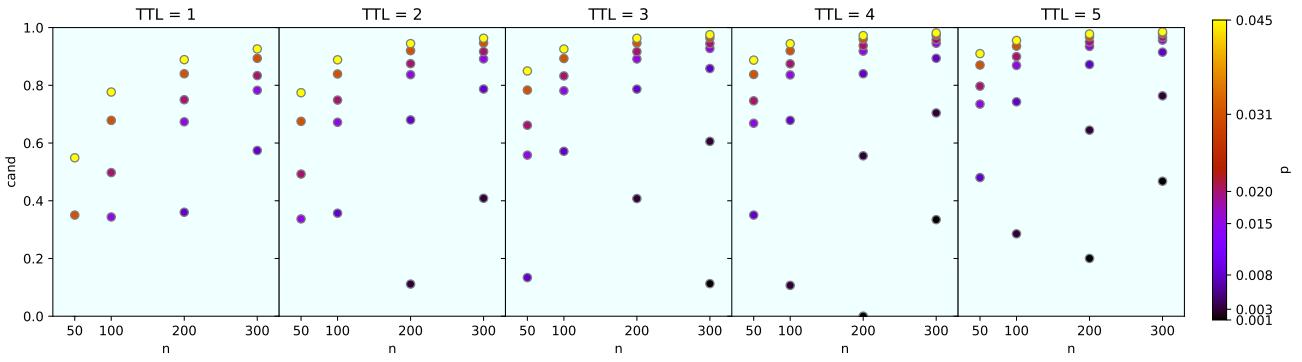


Figure 5: Fraction of the nodes with information for Edge Markovian graph estimated using Eq 4.

In the case of the geometric graphs (Fig 6-7), such equilibrium value of nodes with information cannot be clearly observed at the same extent, however one can clearly see plateaus in fraction of nodes with information after much longer (approx. 200-500 iterations) rising phase. The values in the plateaus fluctuate over time due especially for low n due to evolving in time uneven random surface node density in the graph, which controls the edges creation and destruction. It is also the origin of low edge nervousness, and low edge nervousness explains the slower spread of the information in comparison to Edge Markovian graph model. Furthermore, contrary to Edge Markovian graph model, the results clearly depict the correlation of the fraction of nodes with information rising time with the distance threshold (negative correlation) and the number of nodes (positive correlation).

The Manhattan graph model (6) has significant higher information fraction rising times for low d values (20, 30, 50) compared to RWP model (7). This probably is the results of a constraining the nodes movement to 10 streets in each direction, with the distance between neighboring streets of 100. Therefore for the d values of 50 and lower, the nodes cannot form edges with nodes from a parallel streets and the transfer of information takes place with the help of nodes from perpendicular streets if they are close enough. d higher than 50 makes the direct information transfer between the neighboring streets possible.

Clearly, with higher n and higher d , the information has more chances to persist up to 100th iteration, as shown in Fig 9. The Manhattan model seems to have the best performance in preserving information in unfavorable conditions – low TTL and low d . However, when condition for the information persistence are relatively good – $TTL \geq 4$ and d – the Edge Markovian graph preserves the message with 100% yield. Additionally, the figure depicts the gradual transition between the conditions not preserving the message and providing 100% message persistence. In the case of the Edge Markovian model, the dependence of the success of preserving the message on n and p is very sharp for all values of TTL.

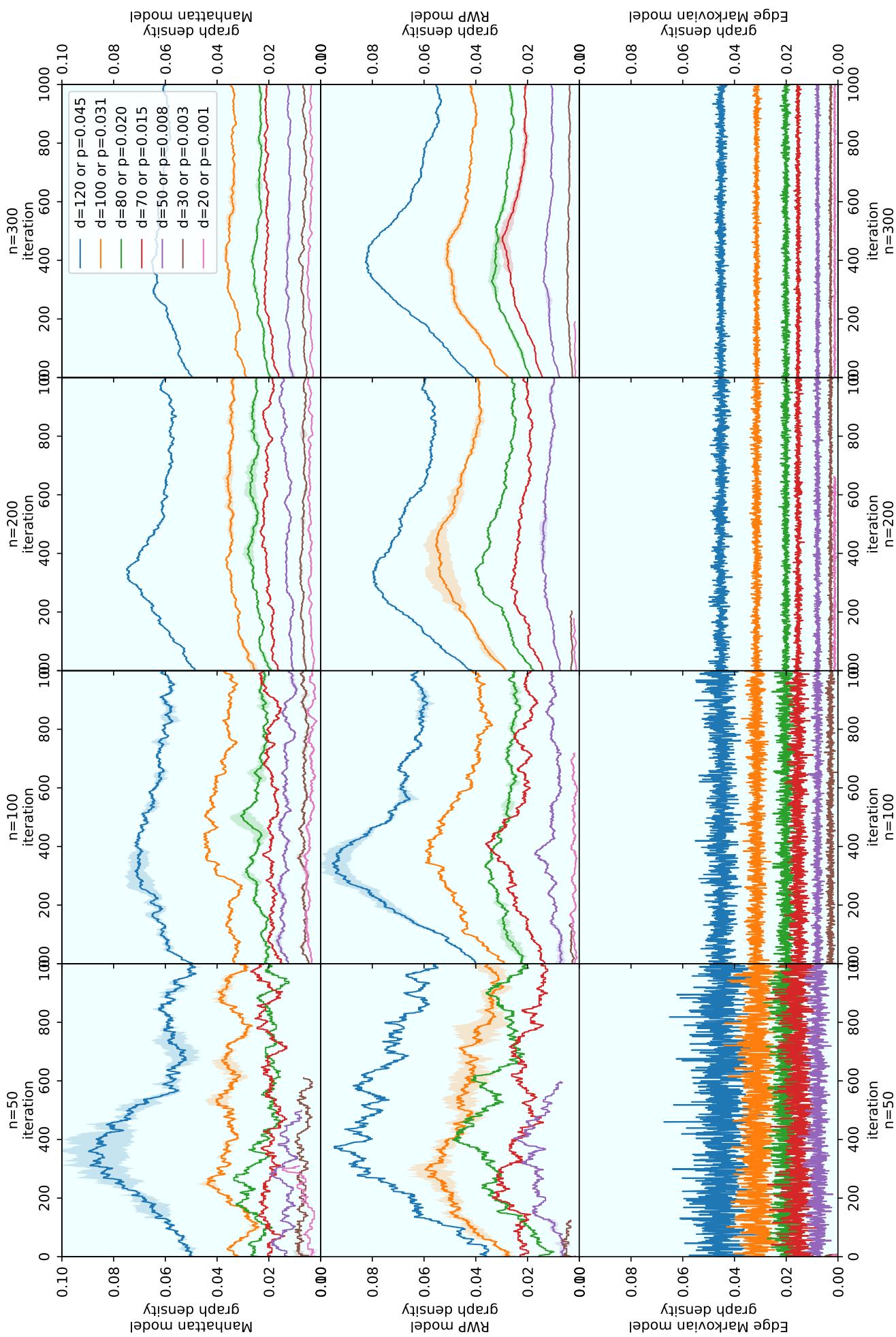


Figure 2: Graph density for all models. Time traces were averaged over 15 experiments. The shaded area around curves is their respective standard deviation.

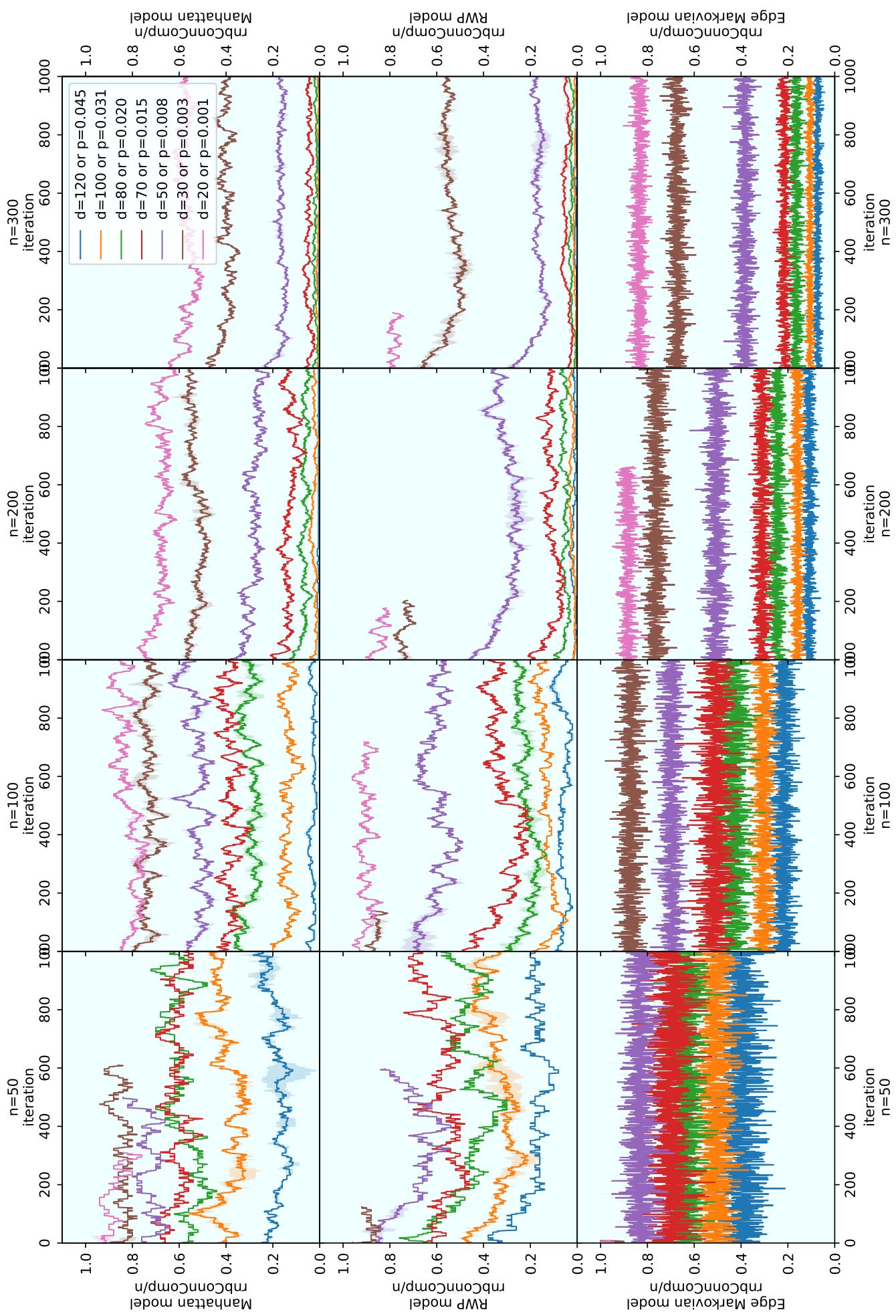


Figure 3: Number of connected components normalized to the number of nodes for all models. Time traces were averaged over 15 experiments. The shaded area around curves is their respective standard deviation.

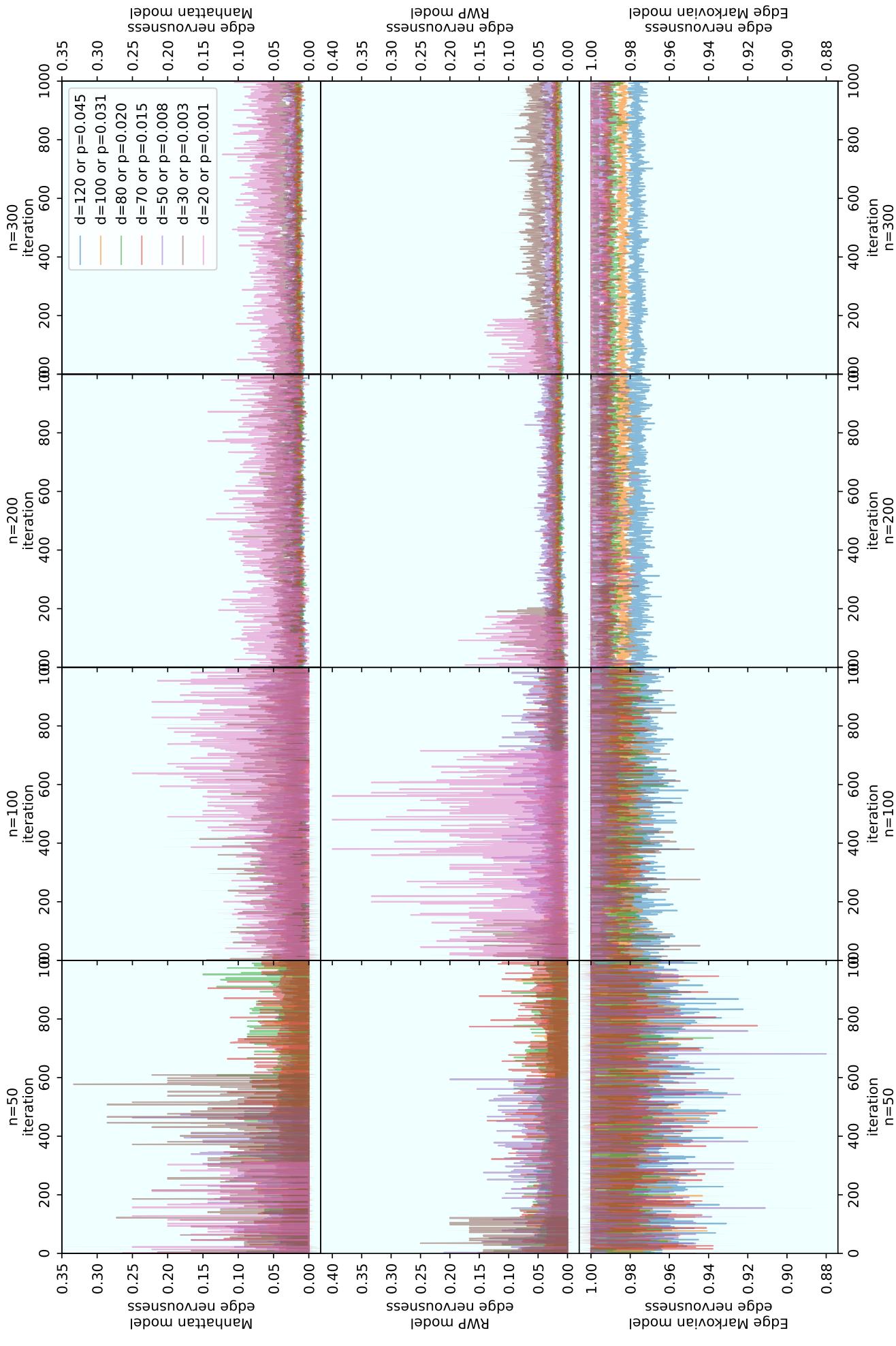


Figure 4: Edge nervousness for all models. Time traces were averaged over 15 experiments. The shaded area around curves is their respective standard deviation.

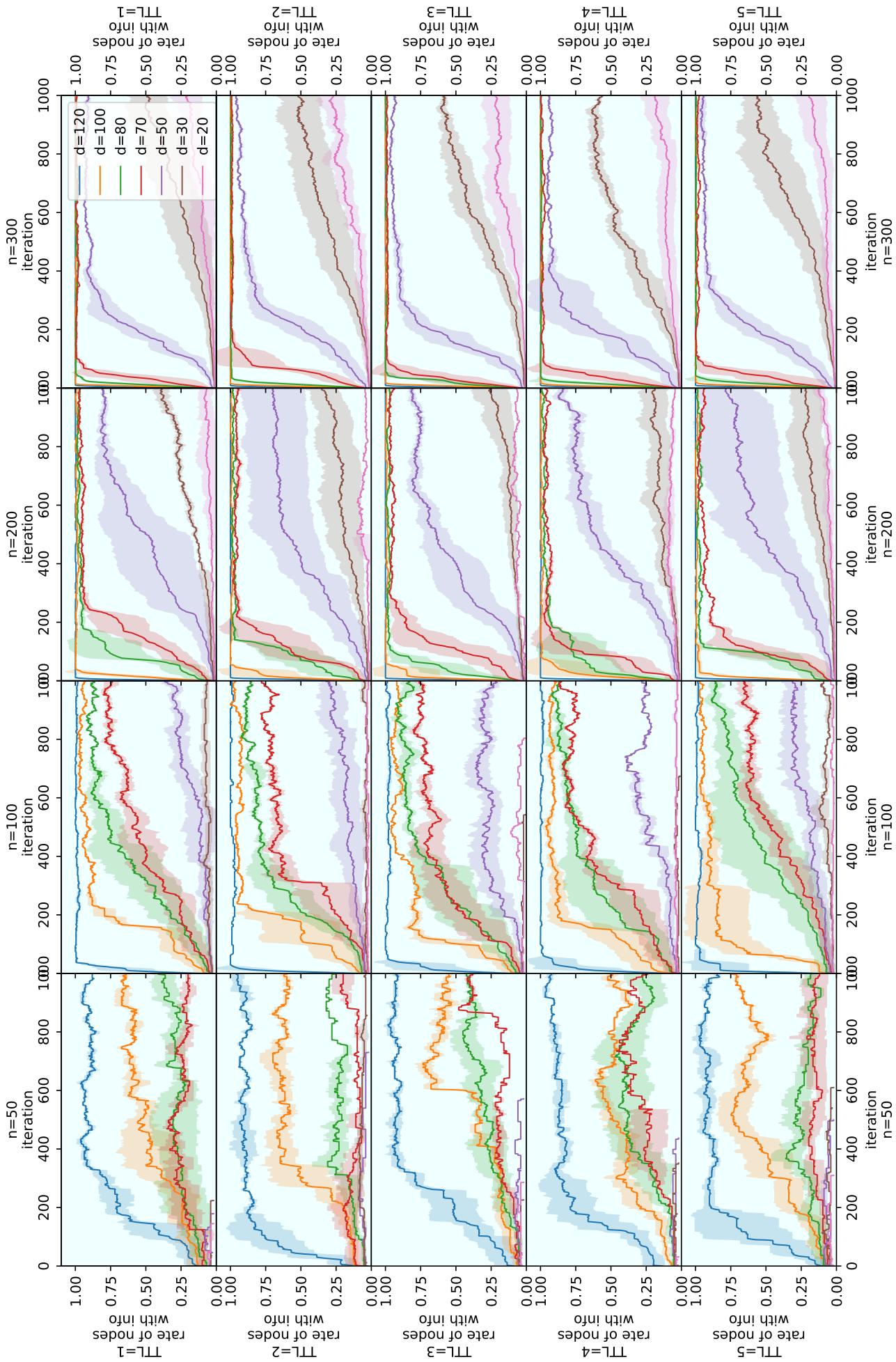


Figure 6: Fraction of the nodes with information in Manhattan model dynamic graph. Time traces were averaged over 15 experiments. The color area around curves is their respective standard deviation.

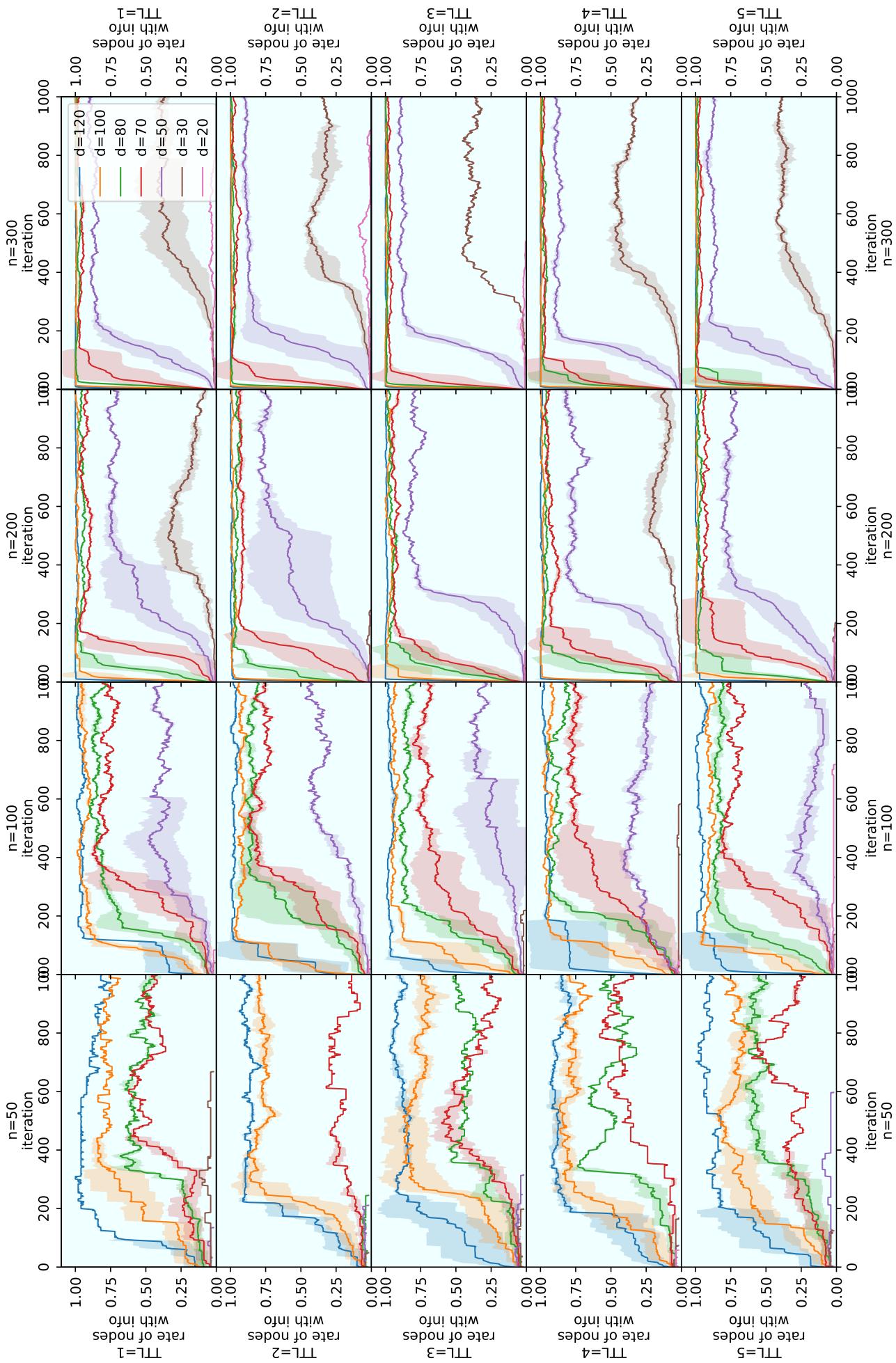


Figure 7: Fraction of the nodes with information in RWP model dynamic graph. Time traces were averaged over 15 experiments. The color area around curves is their respective standard deviation.

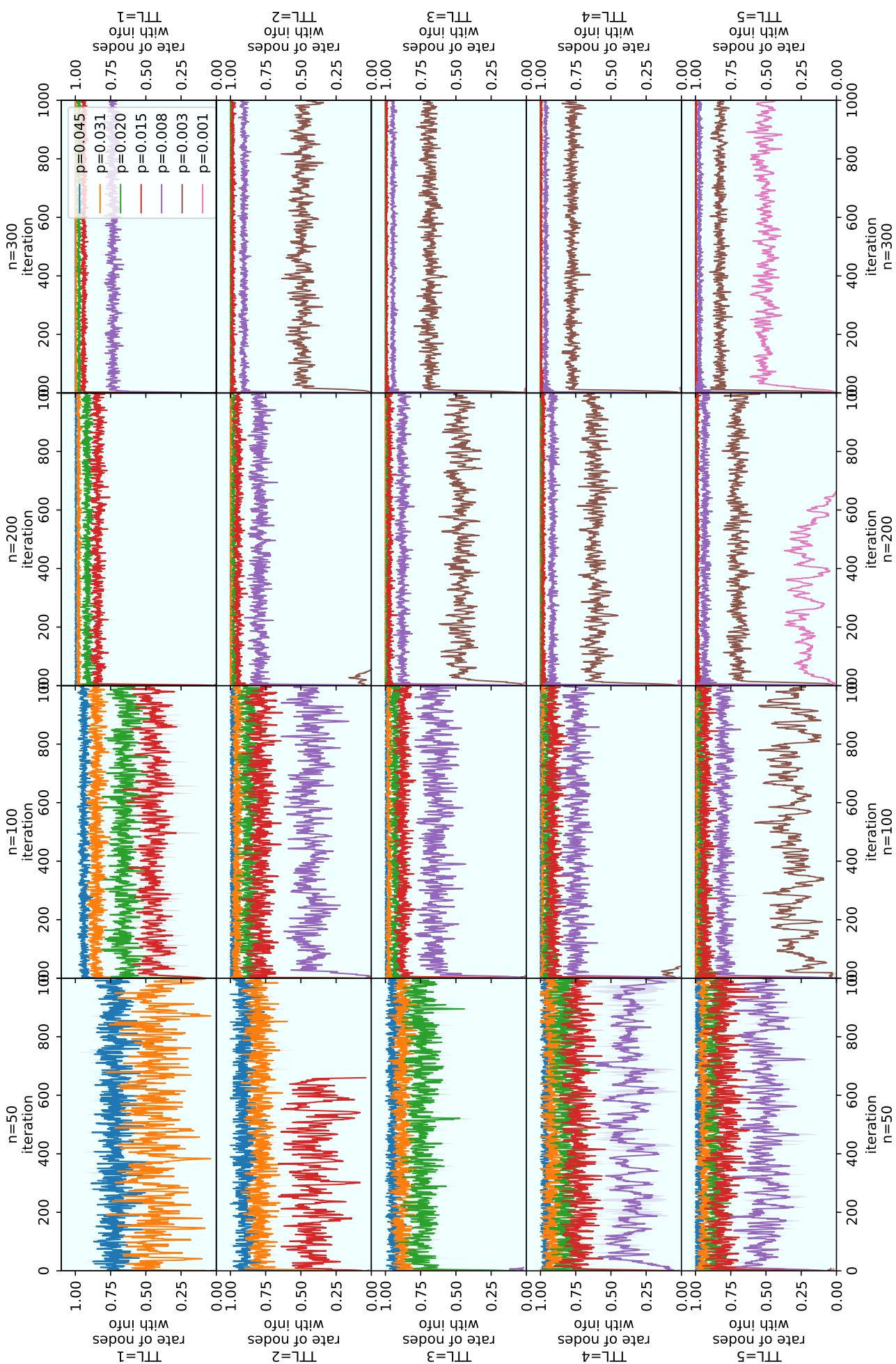


Figure 8: Fraction of the nodes with information in Edge Markovian model dynamic graph. Time traces were averaged over 15 experiments. The color area around curves is their respective standard deviation.

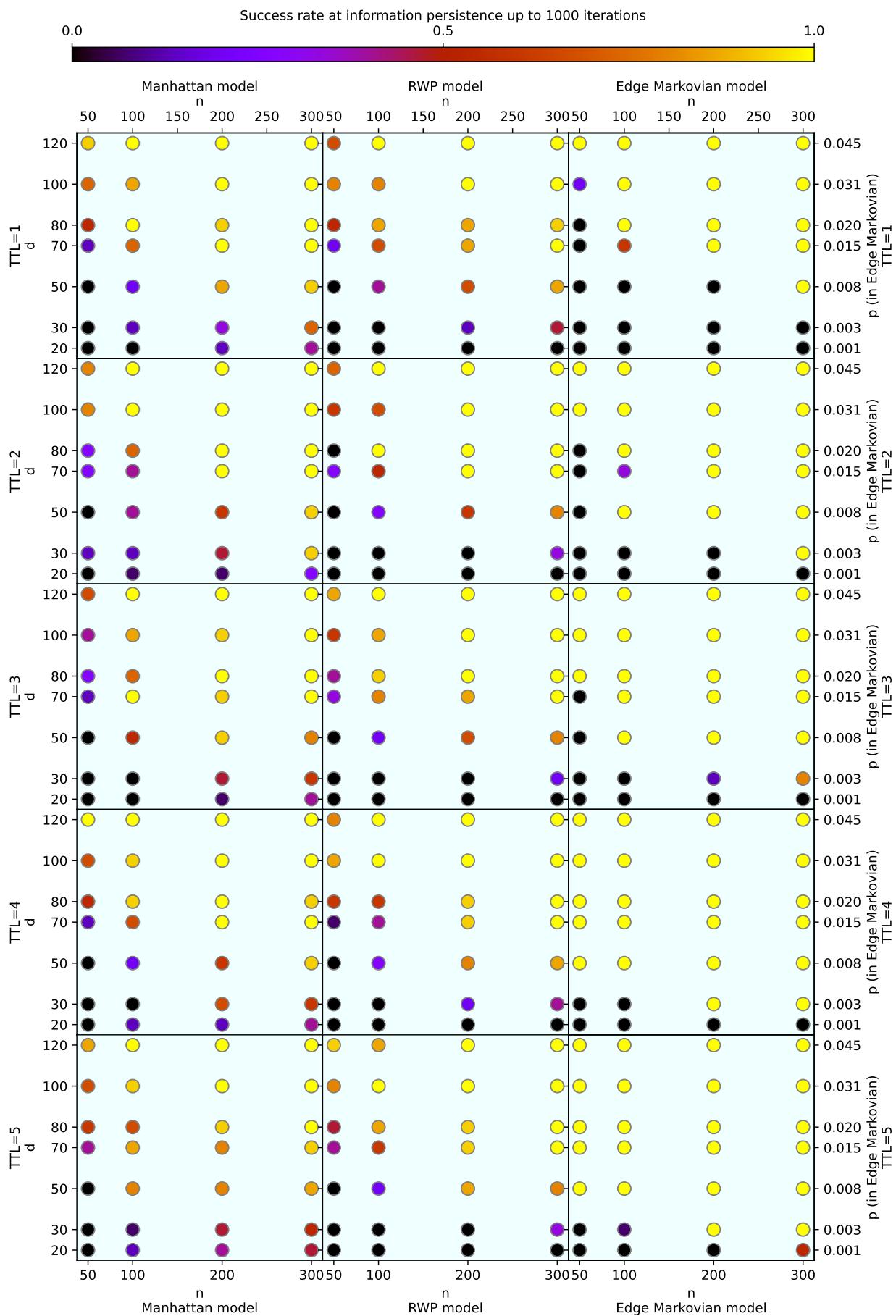


Figure 9: Fraction of the experiments (out of 15 total) where information survived up to 1000th iteration.

References

- [1] Dall, J.; Christensen, M. *Physical review E* **2002**, *66*, 016121.
- [2] Dutot, A.; Guinand, F.; Olivier, D.; Pigné, Y.; *Graphstream: A tool for bridging the gap between complex systems and dynamic graphs* W *Emergent Properties in Natural and Artificial Complex Systems. Satellite Conference within the 4th European Conference on Complex Systems (ECCS'2007)*.