Higher-order Temporal Network Prediction

Thesis for the MSc. Computer Science, Data Science and Technology Track

Mathieu Jung-Muller, 5520495

Thesis committee:
Huijuan Wang (thesis advisor), Associate professor at TU Delft
Robbert Fokkink (second member), Assistant professor at TU Delft

Defence: 20 July 2023

Delft University of Technology

Faculty of Electrical Engineering, Mathematics & Computer Science
Department of Intelligent Systems, Multimedia Computing Group

Higher-order Temporal Network Prediction

Mathieu Jung-Muller

July 2023

Abstract

Temporal networks, like physical contact networks, are networks whose topology changes over time. However, this representation does not account for group interactions, when people gather in groups of more than two people, that can be represented as higher-order events of a temporal network. The prediction of these higher-order events is usually overlooked in traditional temporal network prediction methods, where each higher-order event of a group is regarded as a set of pairwise interactions between every pair of individuals within the group. However, pairwise interactions only allow to partially capture interactions among constituents of a system. Therefore, we want to be able to predict the occurrence of such higher-order interactions one step ahead, based on the higher-order topology observed in the past, and to understand which types of interactions are the most influential for the prediction. We find that the similarity in network topology is relatively high at two time steps with a small time lag between them and that this similarity decreases when the time lag increases. This motivates us to propose a memory-based model that can predict a higher-order temporal network at the next time step based on the network observed in the past. In particular, the occurrence of a group event will be predicted based on the past activity of this target group and of other groups that form a subset or a superset of the target group. Our model is network-based, so it has a relatively low computational cost and allows for a good interpretation of its underlying mechanisms. We propose as a baseline the memory-based method for the traditional pairwise network prediction problem. In this baseline model, the predicted higher-order events at a prediction time step are then deduced from the predicted pairwise network at the same prediction time step. We evaluate the prediction quality of all models in eight real-world physical contact networks and find that our model outperforms the baseline model. We also analyze the contribution of group events of different sizes to the prediction quality. We find that the past activity of the target group is the most important factor for the prediction. Moreover, the past activity of groups of a larger size has, in general, a lower impact on the prediction of events of an arbitrary size than groups of a smaller size.

1 Introduction

Real-world complex systems are often represented as networks. In this representation, the nodes are the components of the system, and the links between nodes denote an interaction or a relation between the components. In many cases, the interactions between the components are not continuously active. Individuals may have a face-to-face interaction, send an email, or have a phone call, at a specific time. Temporal networks represent these systems more realistically with time-varying network topology, where each link between two nodes is activated only when the node pair interacts [1][2][3].

Studies in physical contact temporal networks usually focus on pairwise interactions. However, pairwise interactions do not allow to capture all the information about some specific behaviors of the network [4][5]. In the case of physical contacts, individuals may tend to gather in groups [6], which can not be seen directly from pairwise interactions. For another example, collaboration networks often have more than two authors for a scientific paper. Interactions that involve an arbitrary number of nodes are called higher-order interactions.

The classic temporal network prediction problem consists of predicting temporal contacts one step ahead based on the temporal network topology observed in a past window. It enables better forecast and mitigation of the spread of epidemics or misinformation on the network. The temporal network prediction problem is also equivalent to problems in recommender systems. For instance, it can be predicting which user will purchase which product, or which individuals will become acquaintances [7][8].

Different methods have been proposed for pairwise link prediction on temporal networks. Some methods rely on network embeddings: nodes are represented as points in a dimensional space where connected nodes are supposed to be closer in this embedding space [9][10][11]. Alternatively, deep learning methods have also been proposed [12], for instance, using LSTM methods [13] or adversarial networks [14]. However, these methods are at the expense of high computational costs and are limited in providing insights regarding which network mechanisms enable network prediction.

Finally, a few papers have recently proposed methods to predict the occurrence of higher-order interactions in the future. For example, network metrics [15][16] and embeddings [17] have been used to predict the chance of a set of nodes to be part of an event involving exactly that set of nodes, the so-called simplicial closure problem. However, this work does not correspond to the definition we gave of the classic temporal network prediction problem since they rather focus on predicting whether one interaction will occur at some point in the future.

Therefore, we want to be able to predict the occurrence of higher-order interactions one step ahead, based on the higher-order topology observed in the past, and to understand which types of interactions are the most influential for the prediction. In this paper, we propose a model that predicts which events will happen at a given time step based on the network observed in the past. For this, we establish a memory-based model that is a generalization to higher-order

of the pairwise model proposed by Zou et al. [18]. Our model assumes that the occurrence or not of a group event at the next time step is influenced by the past activity of this target group and of other groups that form a subset or a superset of the target group. Furthermore, the influence of recent events is considered more impactful than the influence of older events. Finally, we make two other assumptions to simplify the problem: the first one is that the number of group events of a given size at the prediction time step is known, and the second one is that the model has the knowledge of which group events occur at least once in the network. We propose as a baseline the memory-based method on traditional pairwise networks, adapted such that the predicted higher-order events at a prediction time step are then deduced from the predicted pairwise network at the same prediction time step. The predictions from our model consistently outperform this baseline, as evaluated in eight real-world physical contact datasets. We find that the past activity of the target group is the most important factor for the prediction. Moreover, the past activity of groups of a large size has, in general, a lower impact on the prediction of events of an arbitrary size than groups of a smaller size.

The rest of the paper is organized as follows. We introduce in Section 2 the representation of higher-order temporal networks and in Section 3 the datasets we use to design and evaluate our prediction method. Key temporal network properties are explored in Section 4 to motivate the design of our model, which is explained in Section 5. In Section 6, our model is evaluated and compared to our baseline before we discuss further analysis and interpretation of the results in Section 7.

2 Higher-order temporal network representation

We define a pairwise temporal network in strictly the same way as what was done by Zou et al. [18]. A pairwise temporal network G measured at discrete times can be represented as a sequence of network snapshots $G = \{G_1, G_2, ..., G_T\}$, where T is the duration of the observation window, $G_t = (V, E_t)$ is the snapshot at time step t with V and E_t being the set of nodes and contacts, respectively. If nodes a and b have a contact at time step t, then $(a, b) \in E_t$. Here, we assume all snapshots share the same set of nodes V. The set of links in the time-aggregated network is defined as $E = \bigcup_{t=1}^{t=T} E_t$. A pair of nodes is connected with a link in the aggregated network if at least one contact occurs between them in the temporal network. The total number of links is $M_E = |E|$. We give each link in the aggregated network an index i, where $i \in [1, M_E]$. The temporal connection or activity of link i over time can then be represented by a T-dimension vector x_i whose element is $x_i(t)$, where $t \in [1, T]$, such that $x_i(t) = 1$ when link i has a contact at time step t and $x_i(t) = 0$ if no contact occurs at t. A temporal network can thus be equivalently represented by its aggregated network, where each link i is further associated with its activity time series x_i.

This traditional temporal network representation records social contacts as a set of pairwise interactions. However, individuals may gather in larger groups,

where more than two people interact with one another at the same time. For instance, an interaction between three nodes a, b, c at time step t is usually measured and recorded as three pairwise interactions $(a, b), (b, c), (a, b)$. Physical interactions can be more accurately represented as a higher-order temporal network H (or temporal hypergraph, following the definition from Cencetti et al. [19]), which is a sequence of network snapshots $H = \{H_1, ..., H_T\}$, where T is the duration of the observation window, $H_T = (V, \mathcal{E}_t)$ is the snapshot at time step t with V being the set of nodes shared by all snapshots and $\mathcal{E}_t$ the set of hyperlinks of arbitrary order that are activated at time step t. The activation of a hyperlink $(u_1, ..., u_d)$ at time step t corresponds to a group interaction between nodes u_1, ..., u_d at time step t. The size or order of an interaction is the size of the group, i.e., the set of nodes involved in the interaction. The hyperlink $(u_1, ..., u_d)$ active at time step t is called an event, its size is d. We call a hyperlink of size d a d-hyperlink and an event of size d a d-event. For a hyperlink $h_1 = (u_1, ..., u_{d_1})$ and another hyperlink $h_2 = (u_1, ..., u_{d_1}, ..., u_{d_2})$ such that $d_1 \leq d_2$, which means that h_1 is included in h_2, we call h_2 a super-hyperlink of h_1 and h_1 a sub-hyperlink of h_2. The set of hyperlinks in the higher-order time-aggregated network is defined as $\mathcal{E} = \bigcup_{t=1}^{t=T} \mathcal{E}_t$. A hyperlink belongs to $\mathcal{E}$ if it is activated at least once in the temporal network. The total number of hyperlinks is $M_{\mathcal{E}} = |\mathcal{E}|$. We give each hyperlink in the higher-order aggregated network an index i, where $i \in [1, M_{\mathcal{E}}]$. The temporal connection or activity of hyperlink i over time can then be represented by a T-dimension vector x_i whose element is $x_i(t)$, where $t \in [1, T]$, such that $x_i(t) = 1$ when hyperlink i is activated at time step t and $x_i(t) = 0$ if no group interaction occurs at t. A higher-order temporal network can thus be equivalently represented by its higher-order aggregated network, where each hyperlink i is further associated with its activity time series x_i.

3 Datasets

To design and evaluate our temporal network prediction methods, we consider eight empirical physical contact networks from the SocioPatterns project[1]. They are collections of face-to-face interactions at a distance smaller than 2 m, in several social contexts such as study places (Highschool2012, Highschool2013, Primaryschool), conferences (SFHH, Hypertext2009), workplaces (Hospital, Workplace) or an art gallery (Science Gallery). These face-to-face interactions are recorded as a set of pairwise interactions. Based on them, group interactions are deduced by promoting every fully-connected clique of $\binom{d}{2}$ contacts happening at the same time step to an event of size d occurring at this time step. Since a clique of order d contains all its sub-cliques of order $d' < d$, only the maximal clique is promoted to a higher-order event, whereas sub-cliques are ignored. This method has been used by Cencetti et al. to deduce higher-order interactions from datasets of human face-to-face interactions [19]. The datasets are also preprocessed by removing nodes not connected to the largest connected

[1] http://www.sociopatterns.org/

Dataset	Order 2	Order 3	Order 4	Order 5	Order 6+
Science Gallery	12770	1421	77	7	0
Hospital	25487	2265	81	2	0
Highschool 2012	40671	1339	91	4	0
Highschool 2013	163973	7475	576	7	0
Primaryschool	97132	9262	471	12	0
Workplace	71529	2277	14	0	0
Hypertext 2009	18120	874	31	12	4
SFHH Conference	48175	5057	617	457	199

Table 1: Number of events of every order for every dataset after preprocessing.

component in the pairwise time-aggregated network and long periods of inactivity, when no event occurs in the network. Such periods usually correspond, e.g., to night and weekends, and are recognized as outliers. This corresponds to the preprocessing done by Ceria et al. [20][21], and we are using the same preprocessed datasets for our work. The distribution of events in our datasets, after preprocessing, is shown in Table 1.

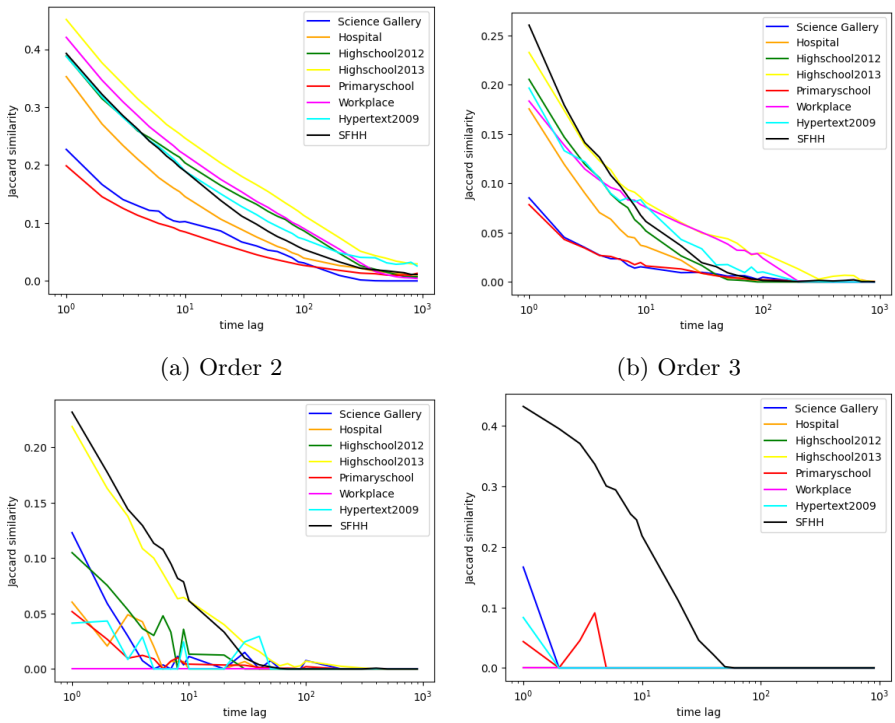
4 Network memory property

Zou et al. observed properties of time-decaying memory in pairwise temporal networks. This means that different snapshots of the network share certain similarities. These properties were used to better predict pairwise interactions [18]. Inspired by this, we also want to know whether higher-order temporal networks have similar properties at different orders and whether this property can be used to predict higher-order events in temporal networks.

Therefore, we examine the Jaccard similarity of the network at two different time steps and for every order. The Jaccard similarity measures how similar two given sets are by taking the ratio of the size of the intersection set over the size of the union set. In our case, we compute the Jaccard similarity, for every order n, between the set of n-hyperlinks active at a time step t_1, called $\mathcal{E}^n_{t_1}$, and the set of n-hyperlinks active at a time step t_2, called $\mathcal{E}^n_{t_2}$. Therefore, we compute the Jaccard similarity: $\frac{|\mathcal{E}^n_{t_1} \cap \mathcal{E}^n_{t_2}|}{|\mathcal{E}^n_{t_1} \cup \mathcal{E}^n_{t_2}|}$. The difference $t_2 - t_1$ is called the time lag.

As shown in Figure 1, the similarity decays as the time lag increases for orders 2, 3, and 4. This behavior can be observed for all datasets. The time-decaying memory at order 5 has not been observed except for SFHH. This is because the number of 5-events is very small in all networks except for SFHH, as shown in Table 1.

Finally, we check if these properties are also present in the case of collaboration networks, whose topological-temporal properties of events were shown to be substantially different than physical contact networks [21]. We thus compute the Jaccard similarity for every order on collaboration networks (see Figure A1



(a) Order 2

(b) Order 3

(c) Order 4

(d) Order 5

Figure 1: Jaccard similarities as a function of the time lag, for orders 2, 3, 4, 5, in eight real-world physical contact networks.

in the Appendix), and we find that collaboration networks do not show a significant pattern of decay with regard to the time lag. Therefore, we focus in this paper on physical contact networks, which possess the memory property that we will utilize to make the prediction.

5 Models

5.1 Baseline

Based on the observed time-decaying memory of the physical contact networks, Zou et al. created a memory-based model called SD (for Self-Driven model) that predicts a link's activity at the next time step based on its past activity [18]. The SD model makes a prediction of the links active at a time step $t+1$ based on the metric of the activation tendency. The tendency $w_j(t+1)$ of a link j to be active at time $t+1$ is defined as:

$$w_j(t+1) = \sum_{k=t-L+1}^{k=t} x_j(k) \exp^{-\tau(t-k)},$$

where $t+1$ is the time of interest for the prediction, j is the link of interest, L is the length of the window to use for the prediction, τ is the exponential decay factor, and $x_j(k)$ is the activation state of link j at time k: $x_j(k) = 1$ if the link j is active at time k and $x_j(k) = 0$ otherwise.

The model has the knowledge of the aggregated network, so it only computes the activation tendency of links that belong to E. We also give the model the number of contacts n occurring at time $t+1$. Therefore, the model will select the n links with the highest activation tendency value at $t+1$. These will be the n predicted contacts at time $t+1$.

We propose a baseline model that deduces a higher-order prediction from the pairwise prediction. At every time step $t+1$, there is a set of links that are predicted as active by the model. This set of pairwise links is promoted to a set of higher-order hyperlinks of arbitrary order using the same method from Cencetti et al. [19] as the one we explained in Section 3. This set of higher-order hyperlinks is considered the prediction made by the baseline of which hyperlinks will be active at time step $t+1$.

5.2 Generalized model

Since the time-decaying memory is also observed at different orders on the higher-order temporal networks, this inspires us to create a model that generalizes the memory-based model to higher-order networks. The essence of the model is that the future activity of a hyperlink should be dependent on its past activity. Furthermore, since Ceria et al. [21] showed that events of different orders that occur close in time tend to overlap in component nodes, we also consider that the activity of a hyperlink should be dependent on the past activity of its sub-hyperlinks and super-hyperlinks. Finally, recent events should have more influence than older events, based on the observed memory property.

Therefore, similarly to what was done for the SD model, we propose a generalized version of the tendency that applies to hyperlinks of arbitrary order. The activation tendency of a higher-order hyperlink j at $t+1$ is defined as:

$$w_j(t+1) = \sum_{k=t-L+1}^{k=t} \sum_{i \in S_j} c_{d_i d_j} x_i(k) \exp^{-\tau(t-k)},$$

where $t+1$ is the time of interest for the prediction, j is the hyperlink of interest, L is the length of the window to use for the prediction, τ is the exponential decay factor, S_j is the set of hyperlinks that are either sub-hyperlinks of j or that have j as a sub-hyperlink, and $x_i(k)$ is the activation state of hyperlink i at time k: $x_i(k) = 1$ if the hyperlink i is active at time k and $x_i(k) = 0$ otherwise. $c_{d_i d_j}$ is the coefficient of cross-order influence, for which d_i is the size of hyperlink i and d_j is the size of hyperlink j. For instance, c_{32} is the coefficient associated with the influence of the activation of a 3-hyperlink on the activation of one of its sub-hyperlinks of size 2. We put $c_{dd} = 1$ for any arbitrary hyperlink size d. The different sub-models of our general model are obtained by varying the values of the cross-order coefficients $c_{d_1 d_2}$ for $d_1 \neq d_2$.

The model has the knowledge of the aggregated network, so it only computes the activation tendency of hyperlinks that belong to $\mathcal{E}$. For every order o considered, we give to the model the number n_o of events of size o occurring at time $t+1$. Therefore, for every order o, the model will select the n_o hyperlinks with the highest activation tendency value at $t+1$.

Thus, our model does not require the assumption used in Section 3 for the reconstruction that two events of different orders can not occur at the same time step if one is included in the other.

6 Model evaluation

6.1 Network prediction quality

We aim to predict the activation of hyperlinks at time step $t+1$ based on the higher-order temporal network observed between $t-L+1$ and t. Therefore, for the evaluation of the model, we make a prediction for every time step $t+1$ that is between $L+1$ and T, where T is the global lifespan of the network. At every time step, we count the number of true positives for every order. In the end, the prediction quality for an arbitrary order is the ratio between the count of true positives for that order and the total number of events of that order that occurred between $L+1$ and T. Our prediction is deterministic and will give the same result for every run with the same parameters since it does not sample randomly but considers all time steps once. As the model knows how many events of every order to predict, the count of false positives is the same as the count of false negatives, and therefore they do not offer any extra information.

6.2 Choice of model parameters

Since the events of orders higher than 4 are very few in number in the real-world physical contact networks considered, we focus on the prediction of events of orders 2, 3, and 4 based on the previous activities of events of orders 2, 3, and 4. Therefore, events of order higher than 4 are ignored. For every order, we make its associated pair of cross-order coefficients take all possible values in $\{0.0, 0.1, ..., 1.0\} \times \{0.0, 0.1, ..., 1.0\}$.

We choose the value $L = 30$, which is equivalent to 600s in our real-world physical contact networks (20s interval time between two time steps). This also corresponds to 10 min. The value of $L = 30$ was found by comparing the accuracy of the prediction for different values of L between 1 and $T/2$ and for different values of τ between 0.25 and 1. Choosing $L = 30$ instead of the value of $L = T/2$ that was chosen by Zou et al. [18] does not result in a drop in accuracy and allows for much faster computations.

We also compared the prediction performance for different values of the decay factor τ. We tried different values $\tau \in \{0, 0.25, ..., 1.5\}$. $\tau = 0$ gave the worst results in every order for all datasets. At orders 2 and 3, the higher the value of $\tau \in [0.25, 1.5]$, the lower the prediction performance. However,

Dataset	Order 2	Order 3	Order 4
Science Gallery	0.33 (0.33)	0.23 (0.16)	0.57 (0.22)
Hospital	0.53 (0.52)	0.50 (0.32)	0.72 (0.17)
Highschool2012	0.56 (0.55)	0.50 (0.38)	0.67 (0.19)
Highschool2013	0.61 (0.61)	0.40 (0.34)	0.61 (0.36)
Primaryschool	0.32 (0.32)	0.19 (0.16)	0.33 (0.09)
Workplace	0.57 (0.56)	0.49 (0.30)	0.50 (0.07)
Hypertext2009	0.50 (0.50)	0.52 (0.34)	0.68 (0.10)
SFHH Conference	0.53 (0.52)	0.45 (0.38)	0.58 (0.39)

Table 2: Highest prediction accuracy obtained for our model per order for every dataset, compared to the baseline in parentheses.

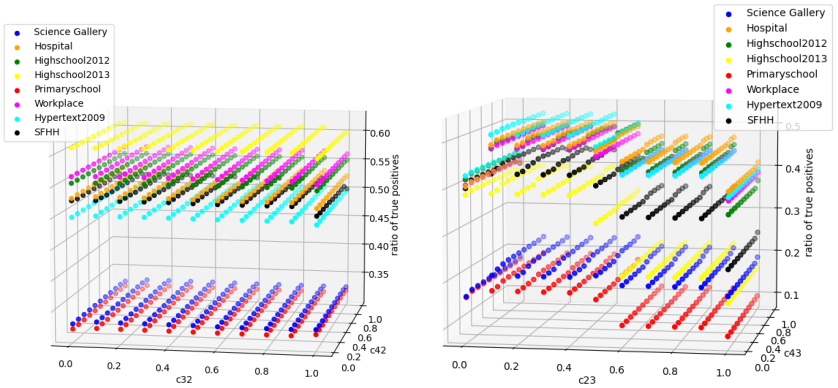
the difference in the accuracy was always very small, around 1.5% difference between the prediction accuracy for $\tau = 0.25$ and the prediction accuracy for $\tau = 1$, for instance. At order 4, no value for τ consistently gave better results across all datasets, with a difference in the prediction accuracy of the order of 1%. Substantially, we concluded that using any value $\tau \in [0.25, 1.5]$ did not significantly change the quality of the prediction. Since the highest prediction accuracy was obtained with the value $\tau = 0.25$, our further results analysis is done with this value $\tau = 0.25$.

6.3 Prediction performance

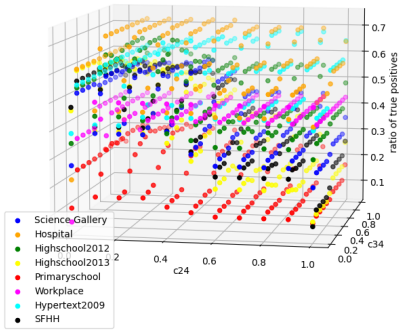
For every order (2, 3, and 4), we compute the prediction accuracy obtained at this order with any pair of cross-order coefficients in $\{0.0, 0.1, ..., 1.0\} \times \{0.0, 0.1, ..., 1.0\}$, which we compare with the prediction accuracy that we obtain from the baseline. Both the computations for the baseline and for our model are done with the value $\tau = 0.25$. The profiles are similar for other values of τ. We show in Figure 2 the prediction accuracy per order of our model, as a function of the coefficients, for all datasets together.

The highest prediction performance obtained per order for every dataset can be found in Table 2. At order 2, the best results of our model perform between 0.5% and 1.5% higher than the baseline, depending on the datasets. At order 3, the improvement compared to the baseline is between 2% and 25%, depending on the datasets. At order 4, it is between 18% and 60% of increase, depending on the datasets. Therefore, we find that our highest accuracy obtained consistently outperforms the baseline.

From our results, we find that the prediction accuracy changes significantly depending on the pairs of coefficients, especially at orders 3 and 4. We show in Figure 3 the ratio of true positives for every order and for two datasets, compared to the value achieved by the baseline. Our model beats the baseline for the majority of the pairs of coefficients. The same graphs for all datasets can be found in the Appendix in Figure A2. We summarize in Table 3 the percentage of pairs of coefficients for which the prediction accuracy is worse



(a) Order 2 (b) Order 3



(c) Order 4

Figure 2: Ratio of true positives as a function of the coefficients, for orders 2, 3, 4, in eight real-world physical contact networks.

than the baseline. For extra information, Figure A3 in the Appendix shows which pairs of coefficients lead to a worse prediction accuracy than the baseline.

Our model predicts events of orders 2, 3, and 4 independently. This means that the predicted network does not correspond to the type of reconstructed networks that we gave in Section 3. Indeed, our model can predict without any restriction that a 2-hyperlink (a, b) and a 3-hyperlink (a, b, c) will be active at the same time step. Since the baseline reconstructs the higher-order network with the same process as the one described in Section 3, this represents a limitation in the comparison between the baseline and our model.

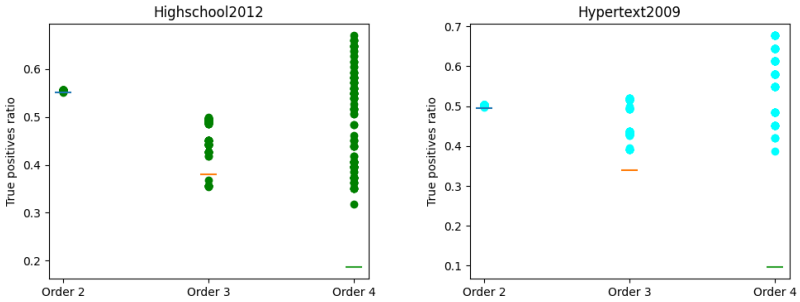


Figure 3: Ratio of true positives for every order, for Highschool2012 and Hypertext2009. One data point represents the ratio of true positives obtained with one pair of coefficients. The horizontal bar represents the value achieved by the baseline.

Dataset	Order 2	Order 3	Order 4
Science Gallery	0.83%	0	0
Hospital	0	0	0
Highschool2012	0.83%	9.09%	0
Highschool2013	17.36%	54.55%	43.80%
Primaryschool	0	45.45%	0
Workplace	0	0	0
Hypertext2009	0	0	0
SFHH Conference	17.36%	45.45%	47.93%

Table 3: Percentage of the pairs of coefficients for which the prediction accuracy of the model is worse than the baseline (out of a total of 121 pairs of coefficients for every cell).

7 Analysis and interpretation of the results for different values of the coefficients

7.1 Coefficients analysis

As we saw in 6.3, the accuracy of the prediction that we obtain is dependent on the coefficients chosen. Therefore, we want to understand how different elements contribute to the prediction performance. At first, we find which coefficients perform best across all datasets.

The best results for order 2 (order 3) are generally obtained for small values of c_{32} and c_{42} (c_{23} and c_{43}). Typically, $c_{32}, c_{42}, c_{23} \in \{0.1, ..., 0.4\}$ and $c_{43} \in \{0.1, ..., 0.5\}$. The most notable exception is at order 3 for both Primaryschool and Science Gallery, which have the highest results for $c_{43} = 1$. This may be due to the relatively low time-decaying memory pattern exhibited by these two

networks at order 3, as shown in Figure 1b in Section 4. For order 4, there is not a clear pattern to achieve the highest result across all datasets. The main thing we can notice is that for the four datasets Highschool2012, Highschool2013, Primaryschool, and SFHH, the highest result is obtained for values clustered in this area: $c_{24} \in \{0.1, ..., 0.3\}, c_{34} \in \{0.3, ..., 0.6\}$.

Therefore, in most cases, except for c_{34}, smaller values for the coefficients (except for the values 0 and 1 which are outliers) tend to lead to better prediction accuracy in general.

7.2 Influence of different orders

In order to better understand the influence of every order on the prediction performance, we draw the prediction accuracy at a given order as a function of one coefficient, with the other coefficient fixed.

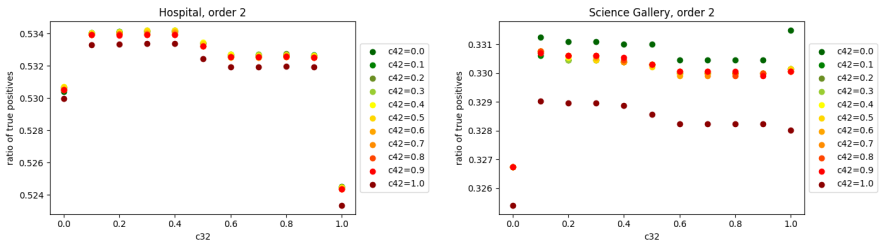


Figure 4: Ratio of true positives at order 2 as a function of c_{32}, with c_{42} fixed, for two datasets.

Excluding values $c_{42} = 0$ and $c_{42} = 1$ which are outliers, c_{42} has barely any impact at all on the ratio of true positives at order 2 for datasets Hospital, Hypertext2009, Science Gallery and Workplace. This is shown in Figure 4 for two datasets. Moreover, in general, neither c_{32} nor c_{42} have a big impact on the prediction accuracy at order 2. All results can be found in the Appendix in Figures A4 and A5.

At order 3, we see in Figure 5 that there is a higher plateau for $c_{23} \in \{0.1, ..., 0.4\}$ and a lower plateau for $c_{23} \in \{0.6, ..., 0.9\}$, with a drop in the accuracy occurring between 0.4 and 0.6. The prediction accuracy for the outliers $c_{23} = 0$ and $c_{23} = 1$ is generally lower and does not fit into any of the two plateaus. Similar results can be found in the other datasets in Figures A6 and A7 in the Appendix.

At order 4, we can not deduce a general pattern of influence from our results. Both c_{24} and c_{34} are impactful. All results at order 4 can be found in Figures A8 and A9 in the Appendix.

From our results, it seems that, in summary, the influence of c_{43} and, to a lower extent, c_{42} does not impact much the ratio of true positives at orders 3 and 2, which can indicate a lower influence of 4-events for the prediction of 3-events and 2-events. On the contrary, the influence of c_{23} and c_{24} is always

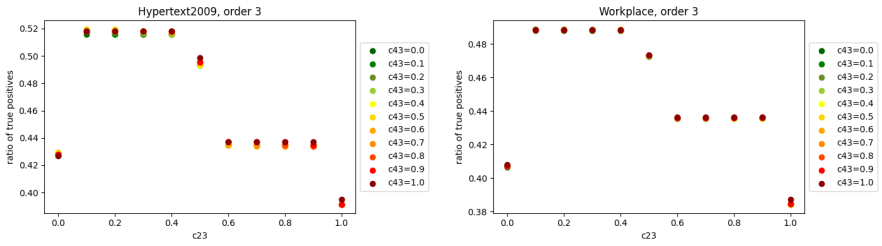


Figure 5: Ratio of true positives at order 3 as a function of c_{23}, with c_{43} fixed, for two datasets.

significant, which suggests a higher importance of 2-events to predict 3-events and 4-events.

Varying the coefficients does not change much in the absolute value of events correctly predicted at order 2. We believe this to be due to the fact that the networks are by far mostly containing events of order 2. Since events of orders 3 and 4 are very scarce compared to events of order 2, it makes sense that their relative influence in predicting events of order 2 is lower.

7.3 Comparison with pairwise results

Our results are consistent with the pairwise results obtained by Zou et al. [18]. In their work, the prediction quality of a link's future activity was mostly correlated to the contribution of the link itself. Furthermore, a slightly better prediction quality was achieved when considering the influence of links that formed a triangle with the target link. In our case, we also find that the prediction quality is mostly influenced by the past activity of the hyperlink of interest itself. Moreover, we find that small cross-order coefficients improve the prediction quality. As our higher-order temporal networks are formed by promoting fully-connected cliques of pairwise interactions to higher-order events, we can draw a parallel between hyperlinks in the higher-order network and links that form a triangle with the target link in the pairwise network. Therefore, both results are very consistent with each other, and they are viewed from a different perspective.

8 Conclusion

In this paper, we proposed a network-based temporal network prediction model that makes predictions on the activity of higher-order hyperlinks at the next time step based on the past activity of this hyperlink and of its sub- and super-hyperlinks. The contributions of the different hyperlinks are weighted with an exponential decay depending on how far in the past the events occurred. Our model was shown to perform consistently better than the baseline directly derived from a pairwise prediction method. We also found a range of values

for the coefficients for which the prediction accuracy consistently achieves the highest scores in all datasets.

Furthermore, we found that taking the influence of sub- and super-hyperlinks into account improves the prediction quality at every order. However, in general, assigning values for the cross-order influence which are relatively small, compared to the influence of the hyperlink itself, results in better prediction. This means that the past activity of the hyperlink itself is still the most important factor for predicting its future state.

References

[1] P. Holme and J. Saramäki, "Temporal networks," *Physics reports*, vol. 519, no. 3, pp. 97–125, 2012.

[2] P. Holme, "Modern temporal network theory: a colloquium," *The European Physical Journal B*, vol. 88, pp. 1–30, 2015.

[3] N. Masuda and R. Lambiotte, *A guide to temporal networks*. World Scientific, 2016.

[4] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, "Networks beyond pairwise interactions: structure and dynamics," *Physics Reports*, vol. 874, pp. 1–92, 2020.

[5] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, *et al.*, "The physics of higher-order interactions in complex systems," *Nature Physics*, vol. 17, no. 10, pp. 1093–1098, 2021.

[6] V. Sekara, A. Stopczynski, and S. Lehmann, "Fundamental structures of dynamic social networks," *Proceedings of the national academy of sciences*, vol. 113, no. 36, pp. 9977–9982, 2016.

[7] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Physics reports*, vol. 519, no. 1, pp. 1–49, 2012.

[8] A. Aleta, M. Tuninetti, D. Paolotti, Y. Moreno, and M. Starnini, "Link prediction in multiplex networks via triadic closure," *Physical Review Research*, vol. 2, no. 4, p. 042029, 2020.

[9] X. Wu, J. Wu, Y. Li, and Q. Zhang, "Link prediction of time-evolving network based on node ranking," *Knowledge-Based Systems*, vol. 195, p. 105740, 2020.

[10] X. Li, W. Liang, X. Zhang, X. Liu, and W. Wu, "A universal method based on structure subgraph feature for link prediction over dynamic networks," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pp. 1210–1220, IEEE, 2019.

[11] Y. Zhou, Y. Pei, Y. He, J. Mo, J. Wang, and N. Gao, "Dynamic graph link prediction by semantic evolution," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2019.

[12] X. Li, N. Du, H. Li, K. Li, J. Gao, and A. Zhang, "A deep learning approach to link prediction in dynamic networks," in *Proceedings of the 2014 SIAM International conference on data mining*, pp. 289–297, SIAM, 2014.

[13] J. Chen, J. Zhang, X. Xu, C. Fu, D. Zhang, Q. Zhang, and Q. Xuan, "E-lstm-d: A deep learning framework for dynamic network link prediction," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3699–3712, 2019.

[14] J. Chen, X. Lin, C. Jia, Y. Li, Y. Wu, H. Zheng, and Y. Liu, "Generative dynamic link prediction," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 29, no. 12, p. 123111, 2019.

[15] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, "Simplicial closure and higher-order link prediction," *Proceedings of the National Academy of Sciences*, vol. 115, no. 48, pp. E11221–E11230, 2018.

[16] H. Nassar, A. R. Benson, and D. F. Gleich, "Neighborhood and pagerank methods for pairwise link prediction," *Social Network Analysis and Mining*, vol. 10, pp. 1–13, 2020.

[17] S. Piaggesi, A. Panisson, and G. Petri, "Effective higher-order link prediction and reconstruction from simplicial complex embeddings," in *Learning on Graphs Conference*, pp. 55–1, PMLR, 2022.

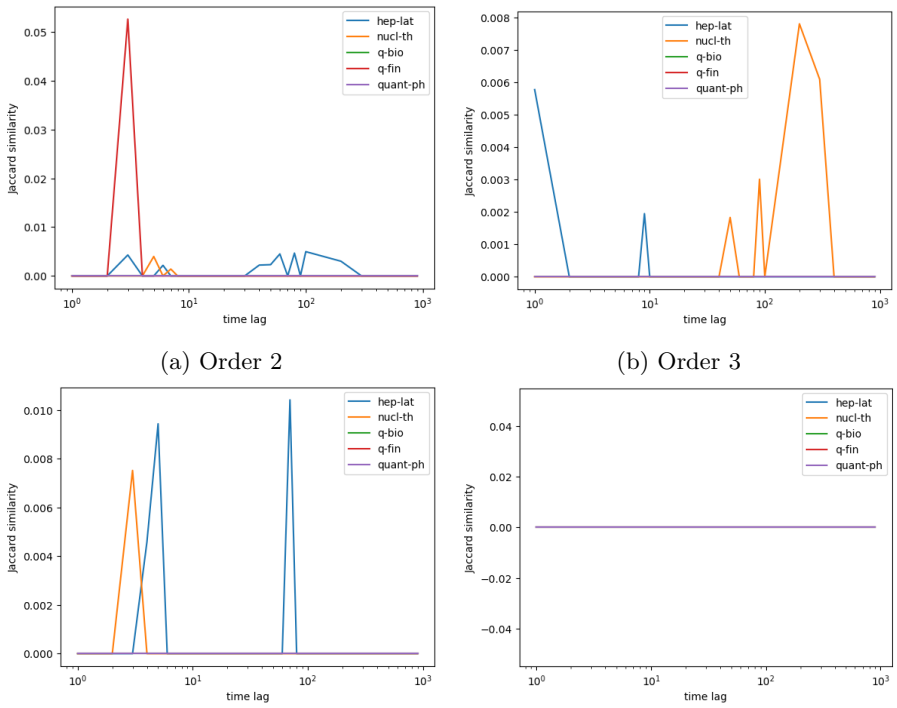
[18] L. Zou, A. Wang, and H. Wang, "Memory based temporal network prediction," in *Complex Networks and Their Applications XI: Proceedings of The Eleventh International Conference on Complex Networks and their Applications: COMPLEX NETWORKS 2022—Volume 2*, pp. 661–673, Springer, 2023.

[19] G. Cencetti, F. Battiston, B. Lepri, and M. Karsai, "Temporal properties of higher-order interactions in social networks," *Scientific reports*, vol. 11, no. 1, p. 7028, 2021.

[20] A. Ceria, S. Havlin, A. Hanjalic, and H. Wang, "Topological–temporal properties of evolving networks," *Journal of Complex Networks*, vol. 10, no. 5, p. cnac041, 2022.

[21] A. Ceria and H. Wang, "Temporal-topological properties of higher-order evolving networks," *Scientific Reports*, vol. 13, no. 1, p. 5885, 2023.

Appendix



(a) Order 2

(b) Order 3

(c) Order 4

(d) Order 5

Figure A1: Jaccard similarities with regard to the time lag, for orders 2, 3, 4, 5, on collaboration networks.

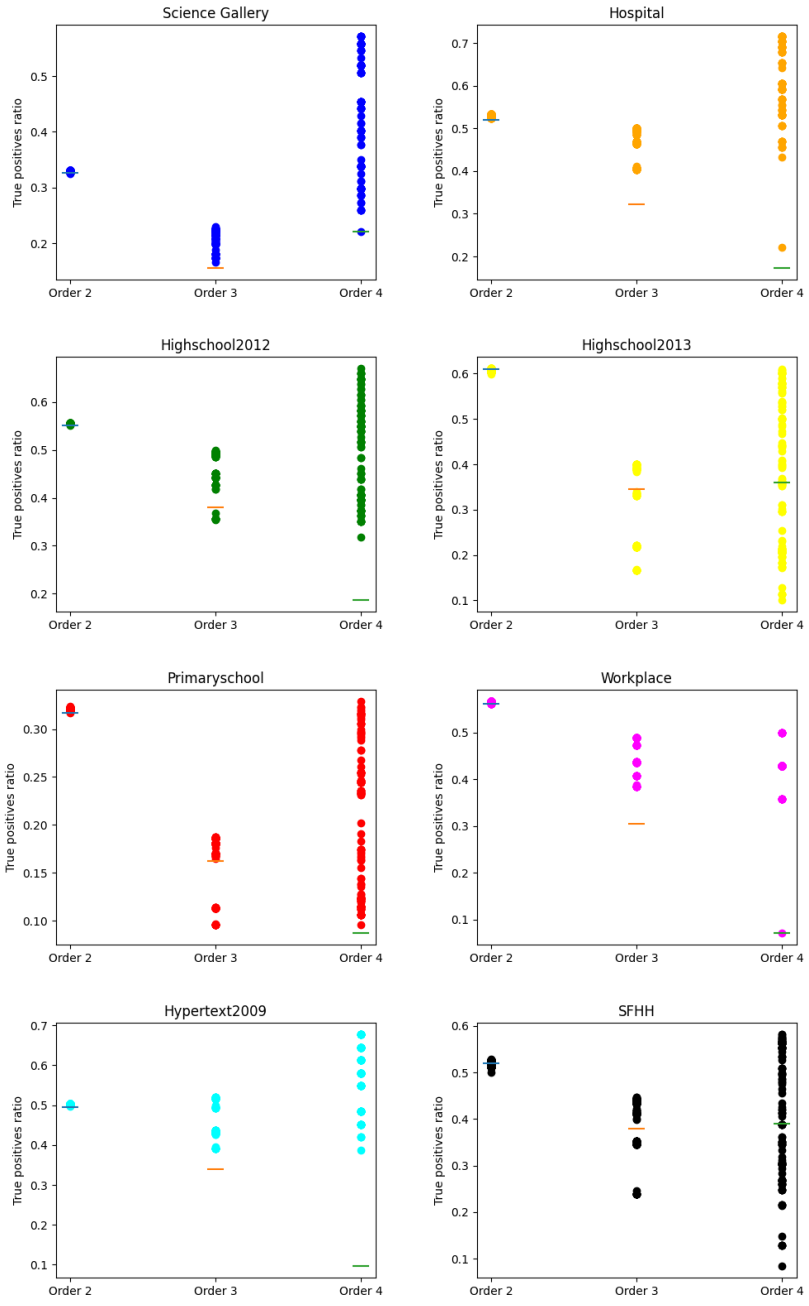


Figure A2: Ratio of true positives for every order, for all datasets, compared to the baseline (horizontal bar).

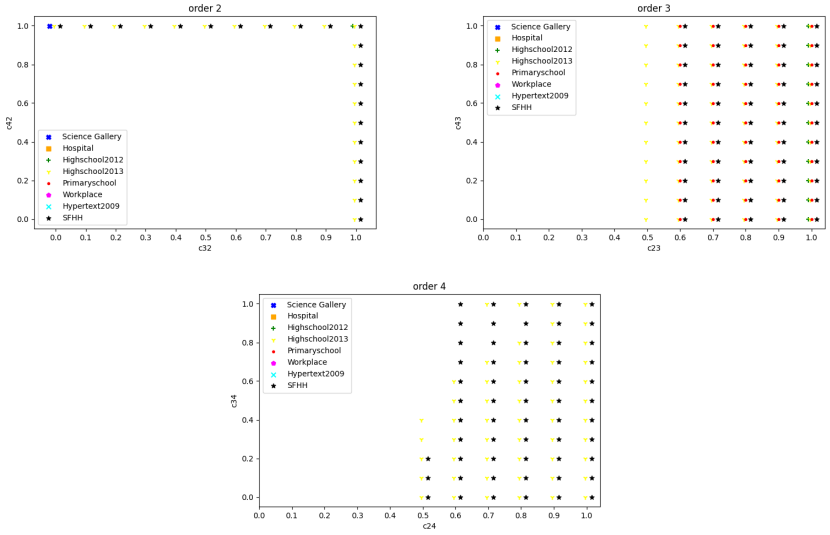


Figure A3: Pairs of coefficients for which the prediction accuracy is worse than the baseline, for orders 2, 3, and 4.

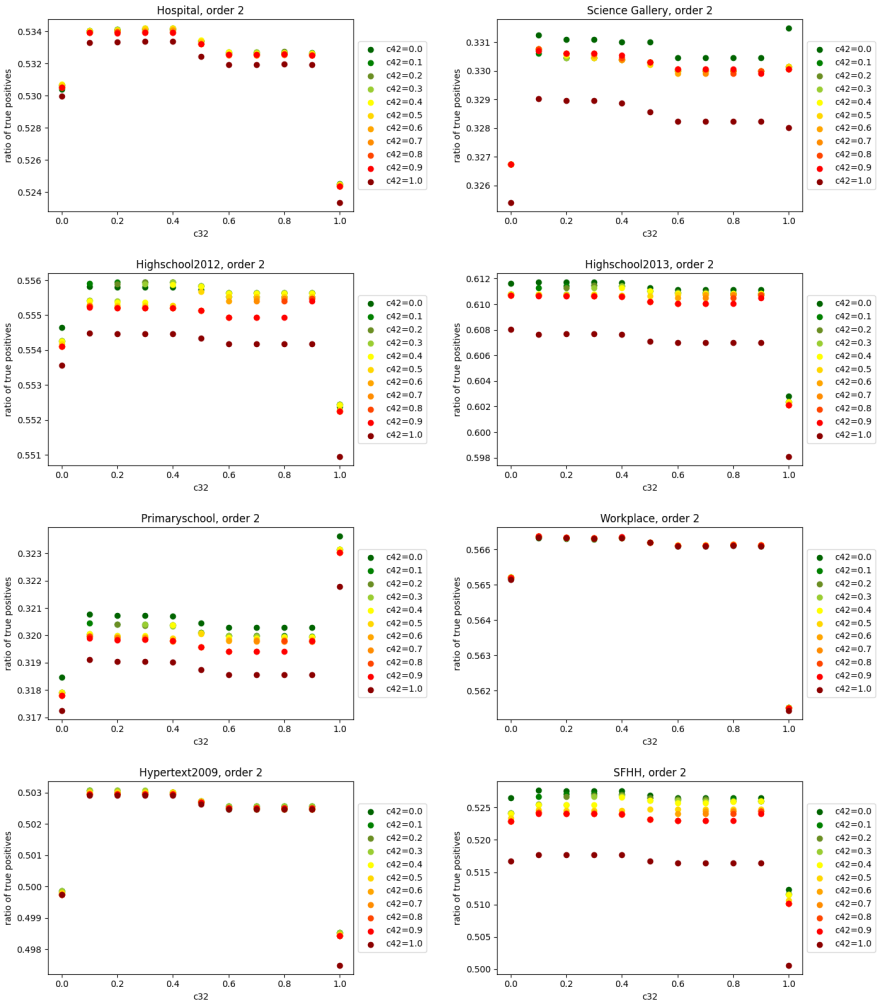


Figure A4: Ratio of true positives at order 2 as a function of c_{32}, with c_{42} fixed, for all datasets.

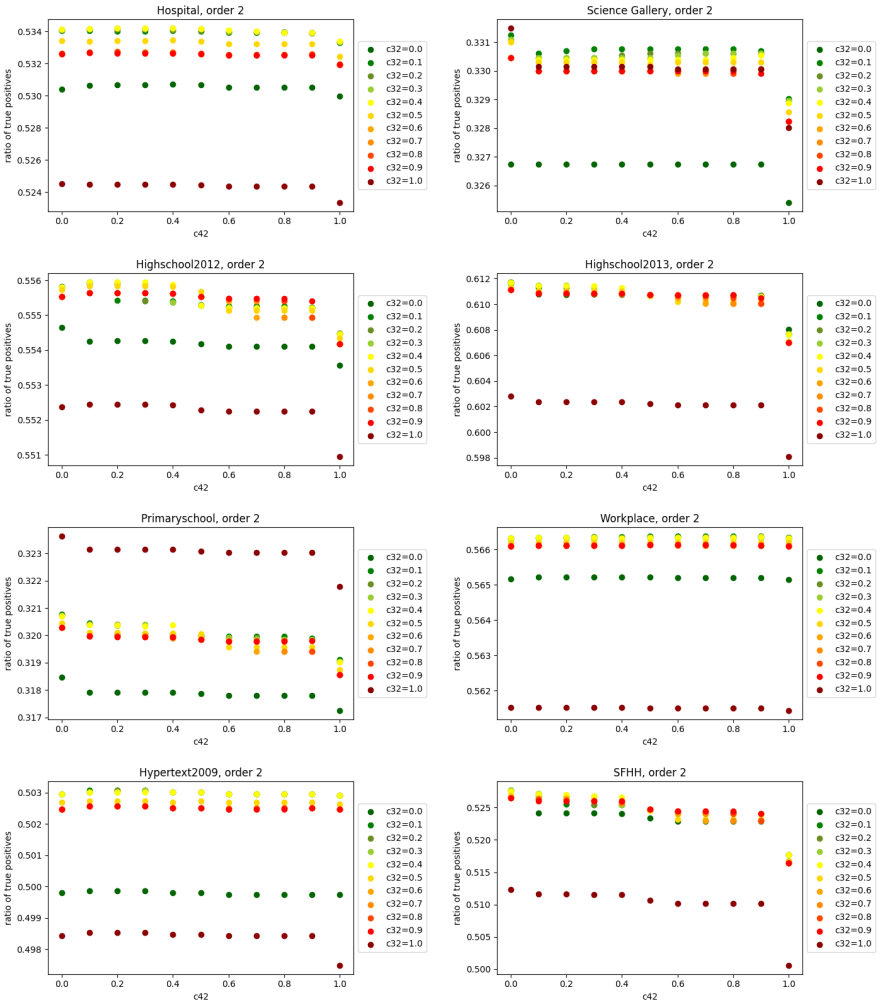


Figure A5: Ratio of true positives at order 2 as a function of c_{42}, with c_{32} fixed, for all datasets.

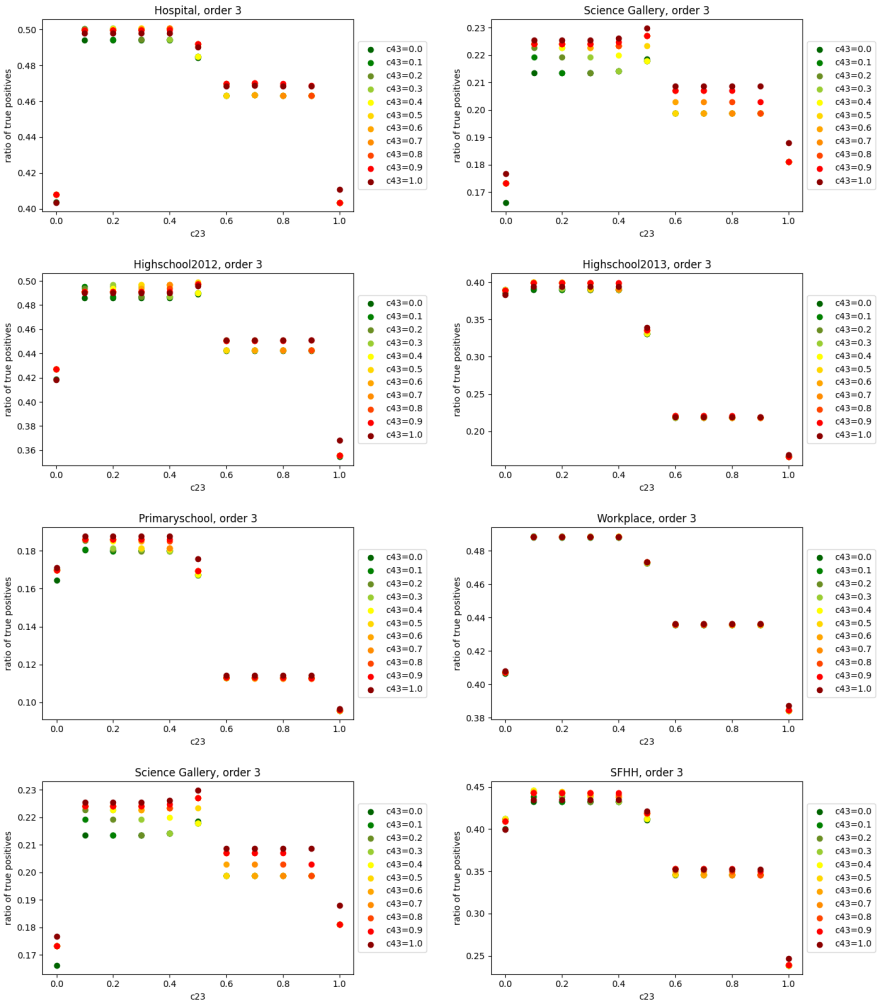


Figure A6: Ratio of true positives at order 3 as a function of c_{23}, with c_{43} fixed, for all datasets.

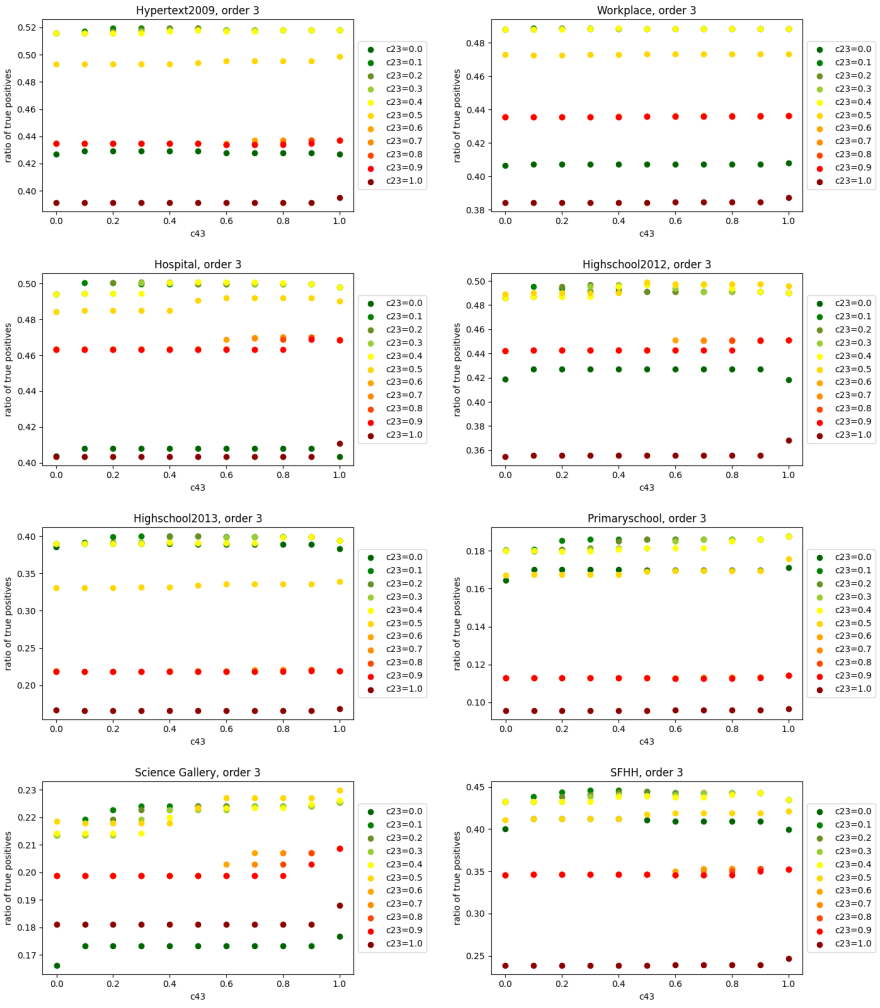


Figure A7: Ratio of true positives at order 3 as a function of c_{43}, with c_{23} fixed, for all datasets.

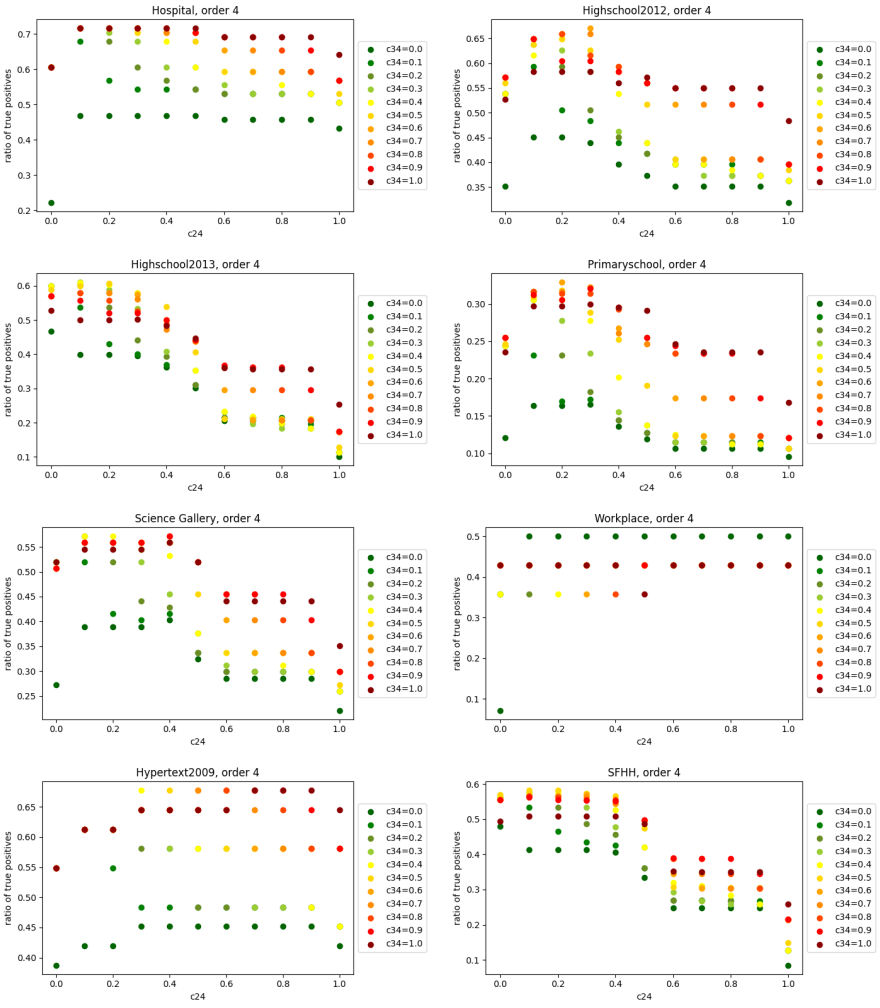


Figure A8: Ratio of true positives at order 4 as a function of c_{24}, with c_{34} fixed, for all datasets.

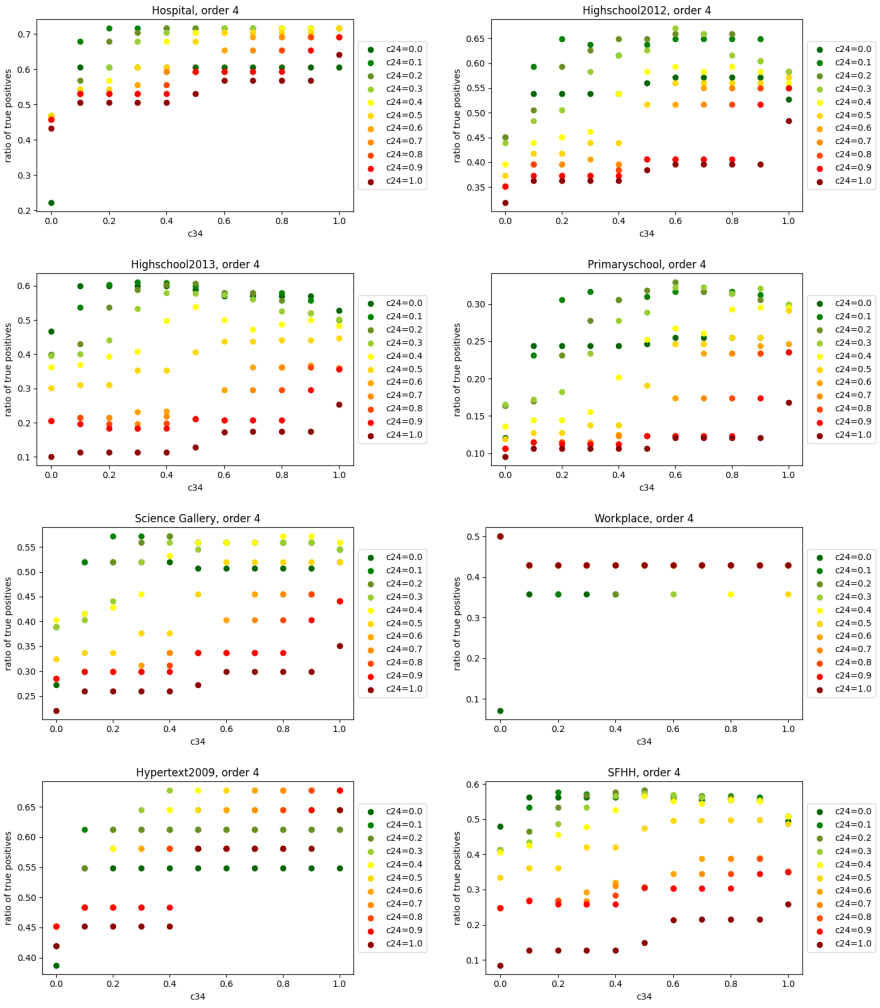


Figure A9: Ratio of true positives at order 4 as a function of c_{34}, with c_{24} fixed, for all datasets.