View Reviews

Paper ID
355

Paper Title
Continuous-Time Relationship Prediction in Dynamic Heterogeneous Information Networks

Track Name
Research

REVIEWER #1

REVIEW QUESTIONS

1. Overall Evaluation
Weak Reject

2. Originality
Medium

3. Importance
Medium

4. Summary of the contribution (in a few sentences)
The paper aims at solving the problem of predicting future continuous-time relationships in information networks that are dynamic and heterogeneous. The main goal is to predict when a link will emerge or appear between two nodes in the network. It differs from earlier works on link prediction that either concentrate on homogeneous networks or investigate either heterogeneity or dynamism (but not both) or predict whether a link will appear in the network (rather than when which is the focus of this work). The authors propose the use of a Non-Parametric Generalized Linear Model to solve the problem of continuous-time relationship prediction based on the extracted features. The benefit of such a non-parametric solution is that there is no need to know the exact distribution of the relationship building time. The authors perform evaluations with both synthetic and DBLP datasets to evaluate the benefit of the technique proposed.

5. List 3 or more strong points, labelled S1, S2, S3, etc.
S1. Nicely written paper.
S2. The problem addressed is both timely and interesting.
S3. The most interesting part of the paper is the evaluation with the DBLP dataset.

6. List 3 ore more weak points, labelled W1, W2, W3, etc.
W1. The experimental evaluation is weak.
There are no experiments that show how the proposed work improves the state of the art.
Also, I would have liked to see experiments with various and possibly more dynamic and heterogeneous types of networks (eg. Twitter, Facebook) so that the benefit of the proposed approach is validated in different settings and with different datasets.

7. Detailed evaluation, labelled D1, D2, D3 etc.
No experiments that show how the proposed work improves the state of the art.
The proposed approach is compared with distributions with different shapes.
I would have liked to see some comparison with one of the techniques that address heterogeneity and dynamicity such as in papers [16], [17] or [11 - 15].

I would have liked to see experiments with various and possibly more dynamic
and heterogeneous types of networks such as social networks so that the benefit
of the proposed approach is validated in different settings and with different
datasets.

Minor comments:

"..were determines.." -> ".. were determined.."

REVIEWER #2

REVIEW QUESTIONS

1. Overall Evaluation
Weak Accept

2. Originality
Medium

3. Importance
Medium

4. Summary of the contribution (in a few sentences)
The authors consider the novel problem setting of continuous-time relationship prediction in dynamic heterogeneous information networks. To this end, they extend existing works by introducing a model, which relies on metapath-based features capturing heterogeneity and considering temporal dynamics. Further, times at which future relations (links or more general metapath-relations) will form are predicted based on a non-parametric generalized linear model. The main contributions compared to the closest competitor approach (by Sun et al.) are the consideration of temporal dynamics and the use of a non-parametric model.

5. List 3 or more strong points, labelled S1, S2, S3, etc.
S1 A novel and relevant problem setting is explored. Previous approaches have either not considered general relationship prediction or are not dynamic and/or rely on parametric assumptions.
S2 Experiments are sufficient and show that the contributions lead to more effective prediction.
S3 The presentation of the paper is very good. It is well readable, figures and formulas are well selected and placed.

6. List 3 ore more weak points, labelled W1, W2, W3, etc.
W1 The paper closely follows the work by Sun et al. It is not too incremental, but also not especially original.
W2 The problem setting is interesting but rather specific, so the importance of the paper is rather medium than high.
W3 The dynamic features do not actually model temporal dynamics, they just focus on recent relations (see D3 below).

7. Detailed evaluation, labelled D1, D2, D3 etc.
D1 The definitions of V and E in II.B are not clear. The formal definition does not match the explaining sentence. Are all entities considered, which lifetimes intersect with the given time interval? If not, why is it restricted?
D2 The features in Definition 3 should be quite costly to compute. How are they computed? What is the computational complexity? For which node pairs are they computed (e.g. for DBLP, computation for all author-pairs should be infeasible)? Are they adapted over successive time windows or always computed from scratch?
D3 Definition 4 still computes an aggregation of the whole window, though the individual contributions are weighted with an aging factor. Thus, the features focus on more recent changes but do not model the temporal dynamics. For instance, they cannot model re-occurring/periodic patterns or other trends.
D4 The inference queries are relevant but could have been presented more concisely, especially because they are not considered anymore in the experiments. The space could have been used for more important content (see below).
D5 Experiments:
- Which interval sizes delta are used? Also which rate parameters are used? I would expect these parameters to have strong influence on performance in the dynamic setting. It would have been interesting to see experiments investigating their influence.
- More baselines could have been considered: For simple link prediction, the target relation could have been modeled as a link type between authors.

D6 The proposed related work by Sun et al. could have been presented in more detail, especially the differences to the proposed model on a technical level.
D7 Slight inconsistency in terminology: The Survival function is referred to as Reliability function in Table II.
D8 Typo in V.B Experiments Setup: ...were determines... -> ...were determined...

8. Required changes for a revision, if applicable. Labelled R1, R2, R3, etc. (Please mark the requests clearly.)
R1 Clarification of D1.
R2 Short elaboration on D2.
R3 Experiments: State the values used for parameters delta and mu. (see D5)

REVIEWER #3

REVIEW QUESTIONS

1. Overall Evaluation
Weak Reject

2. Originality
Medium

3. Importance
Medium

4. Summary of the contribution (in a few sentences)
This paper introduces a relatively new problem to predicate time for link building, and designs a non-parametric generalized linear model to handle it.

5. List 3 or more strong points, labelled S1, S2, S3, etc.
S1: This paper studies a relatively new problem to predict time for link building.
S2: A non-parametric generalized linear Model is designed to make prediction without a given probability distribution of the relationship building model.

6. List 3 ore more weak points, labelled W1, W2, W3, etc.
W1:This paper ignores the advances in machine learning fields, such as RNN, LSTM on the prediction on the sequence data.

W2: The feature selection in the dynamic graph is confusing.

W3: The experimental part should be strengthened.

7. Detailed evaluation, labelled D1, D2, D3 etc.
D1: As for the prediction on the sequence data, the techniques like RNN or LSTM are widely used, especially in the machine learning field. These methods requires less cost in the feature selection and achieve high performance. The authors should cite these papers and conduct a comparison.

D2:The description of dynamic network is confusing. In the definition of dynamic information network, each node has the birth and death date time. From the condition in the definition, the data object is the newly added and not removed in the interval. However, from the description, the data objects are newly added or removed in the interval. In addition, what is the role of objects whose birth date are before the interval?. The different meanings have impact on the features selection in Definition 3.

D3: Please explain more on the importance of distribution model. The paper claim that a conventional approach to modeling this distribution is to fix a parametric distribution for t (e.g. Exponential distribution) and then relate x to t using a Generalized Linear Model [9]. I think we can relate to t using a generalized linear model without explicitly inferring the distribution model. How about the solution?

D4: The weight assignment for different meta-paths is not discussed extensively in this paper. The paper may cite some paper on this issue, like HinDroid: An Intelligent Android Malware Detection System Based on Structured Heterogeneous Information Network in SIGKDD 2017.

D5: The experimental parts should be strengthened. More competitors should be introduced in the experiments. A case study is useful to illustrate the effectiveness of the method. What are the interesting results? Which kinds of meta paths are used? The real-life graph is not large. The authors extract two graphs from DBLP. I am not sure whether the methods can be used on the DBLP data directly.

D6: Minor

Page 5, In “A”ddition.
Page 5, Equation 1, missing = ?
Page 9: the rate of parameter of features were determines…
I do not believe that the work [9] to study the continuous time relationship prediction problem neglect the temporal dynamics of the network.