Synchronous Hyperedge Replacement Graph Grammars

Corey Pennycuff, Satyaki Sikdar, Catalina Vajiac, David Chiang, and Tim Weninger(✉)

University of Notre Dame, Notre Dame, IN 46556, USA
{cpennycu,ssikdar,cvajiac,dchiang,tweninge}@nd.edu

Abstract. Discovering the underlying structures present in large real world graphs is a fundamental scientific problem. Recent work at the intersection of formal language theory and graph theory has found that a Probabilistic Hyperedge Replacement Grammar (PHRG) can be extracted from a tree decomposition of any graph. However, because the extracted PHRG is directly dependent on the shape and contents of the *tree decomposition*, rather than from the dynamics of the graph, it is unlikely that informative graph-processes are actually being captured with the PHRG extraction algorithm. To address this problem, the current work adapts a related formalism called Probabilistic Synchronous HRG (PSHRG) that learns synchronous graph production rules from temporal graphs. We introduce the PSHRG model and describe a method to extract growth rules from the graph. We find that SHRG rules capture growth patterns found in temporal graphs and can be used to predict the future evolution of a temporal graph. We perform a brief evaluation on small synthetic networks that demonstrate the prediction accuracy of PSHRG versus baseline and state of the art models. Ultimately, we find that PSHRGs seem to be very good at modelling dynamics of a temporal graph; however, our prediction algorithm, which is based on string parsing and generation algorithms, does not scale to practically useful graph sizes.

Keywords: Graph generation · Hyperedge replacement
Temporal graphs

1 Introduction

The discovery and analysis of network patterns is central to the scientific enterprise. Thus, extracting the useful and interesting building blocks of a network is critical to the advancement of many scientific fields. Indeed the most pivotal moments in the development of a scientific field are centered on discoveries about the structure of some phenomena [14]. For example, chemists have found that many chemical interactions are the result of the underlying structural properties of interactions between elements [7]. Biologists have agreed that tree structures are useful when organizing the evolutionary history of life [8], sociologists find

L. Lambers and J. Weber (Eds.): ICGT 2018, LNCS 10887, pp. 20–36, 2018.
https://doi.org/10.1007/978-3-319-92991-0_2

that triadic closure underlies community development [11], and neuroscientists have found "small world" dynamics within the brain [4]. Unfortunately, current graph research deals with small pre-defined patterns [13] or frequently reoccurring patterns [15], even though interesting and useful information may be hidden in unknown and non-frequent patterns.

Pertinent to this task are algorithms that learn the LEGO-like building blocks of real world networks in order to gain insights into the mechanisms that underlie network growth and evolution. In pursuit of these goals, Aguinaga *et al.* recently expanded on the relationship between graph theory and formal language theory that allows for a *Hyperedge Replacement Grammar* (HRG) to be extracted from the *tree decomposition* of a graph [2]. In this perspective, the extracted HRG contains the precise building blocks of the graph as well as the instructions by which these building blocks ought to be pieced together [9,12]. In addition, this framework is able to extract patterns of the network's structure from small samples of the graph probabilistically via a Probabilistic HRG (PHRG) in order to generate networks that have properties that match those of the original graph [1].

In their typical use-case, context free grammars (CFGs) and their probabilistic counterpart PCFGs are used to represent and generate patterns of strings through rewriting rules. A natural language parser, for example, learns how sentences are recursively built from smaller phrases and individual words. In this case, it is important to note that the CFG production rules used by natural language parsers encode the way in which sentences are logically constructed, that is, the CFG contains descriptive information about how nouns and verbs work together to build coherent sentences. CFGs can therefore generate new sentences that are at least grammatically correct. This is not the case with HRGs.

On the contrary, the PHRG is completely dependent on the graph's tree decomposition. Unfortunately, there are many ways to perform a tree decomposition on a given graph, and even the optimal, *i.e.*, minimal-width, tree decomposition is not unique. As a result, the production rules in a standard HRG are unlikely to represent the temporal processes that generated the graph.

In the present work we address this problem by learning rules from a temporal graph, in which edges are added and removed at various timesteps, using a related formalism called Synchronous CFGs. SCFGs are a lot like regular CFGs, except that their production rules have two right hand sides, a *source* and a *target*. SCFGs are typically used to perform natural language translation by producing related sentences from a source language into a target language. We reproduce an example synchronous grammar from Chiang [5] for illustration purposes here:

$$
\begin{array}{rcccc}
\mathsf{S} & \rightarrow & \mathsf{NP}_1\ \mathsf{VP}_2 & : & \mathsf{NP}_1\ \mathsf{VP}_2 \\
\mathsf{VP} & \rightarrow & \mathsf{V}_1\ \mathsf{NP}_2 & : & \mathsf{NP}_2\ \mathsf{V}_1 \\
\mathsf{NP} & \rightarrow & \texttt{i} & : & \texttt{watashi ha} \\
\mathsf{NP} & \rightarrow & \texttt{the box} & : & \texttt{hako wo} \\
\mathsf{V} & \rightarrow & \texttt{open} & : & \texttt{akemasu}
\end{array}
$$

where the subscript numbers represent links that synchronize the nonterminals between the source and target RHSs. This grammar can then be used to generate sentences in both English and Japanese. Starting from the synchronous starting nonterminal S_1, we apply these rules to generate the following sentences synchronously (Table 1).

Table 1. Derivation of two sentences from a synchronous CFG.

Rule	English	Japanese
	S_1	S_1
1	NP_2 VP_3	NP_2 VP_3
2	NP_2 V_4 NP_5	NP_2 NP_5 V_4
3	`i` V_4 NP_5	`watashi ha` NP_5 V_4
4	`i` V_4 `the box`	`watashi ha hako wo` V_4
5	`i open the box`	`watashi ha hako wo akemasu`

From the synchronous grammar formalism we see that the production rules encode precise translations between the source and target language. Apart from word-to-word translations, an analysis of these rules would also show the differences in how sentences are constructed in each language. A machine translation system would translate a sentence by parsing the given sentence using the LHS and source-RHS rules. An application of the production rules from the resulting parse tree could regenerate the original source sentence if the source-RHS was applied, or it would generate a translation into the target language if the target-RHSs were applied instead.

Synchronous HRGs (SHRGs, pronounced "shergs") are to HRGs as synchronous CFGs are to CFGs. SHRGs have been proposed for performing natural language understanding and generation by mapping between strings and abstract meaning representations (AMRs), which are graphical representations of natural language meaning.

The present work expands upon work in SHRGs with the observation that the temporal dynamics of a graph, *i.e.*, the changes from one timestep to another, can be represented like a translation from a source language to a target language. *Towards this goal, this paper presents a SHRG extraction algorithm for connected, temporal hypergraphs.* We find that these SHRGs encode interesting information about the temporal dynamics of the graph, and we show that a parse of a graph can be used to predict its future growth.

2 Definitions

Before we describe the SHRG extraction method, some background definitions are needed.