Summarizing and Comparing Story Plans*

Adam Amos-Binks[1], David L. Roberts[2], and R. Michael Young[3]

1 **North Carolina State University, Raleigh, NC, USA**
`aaamosbi@ncsu.edu`
2 **North Carolina State University, Raleigh, NC, USA**
`dlrober4@ncsu.edu`
3 **North Carolina State University, Raleigh, NC, USA**
`rmyoung@ncsu.edu`

Abstract

Branching story games have gained popularity for creating unique playing experiences by adapting story content in response to user actions. Research in interactive narrative (IN) uses automated planning to generate story plans for a given story problem. However, a story planner can generate multiple story plan solutions, all of which equally-satisfy the story problem definition but contain different story content. These differences in story content are key to understanding the story branches in a story problem's solution space, however we lack narrative-theoretic metrics to compare story plans. We address this gap by first defining a story plan summarization model to capture the important story semantics from a story plan. Secondly, we define a story plan comparison metric that compares story plans based on the summarization model. Using the Glaive narrative planner and a simple story problem, we demonstrate the usefulness of using the summarization model and distance metric to characterize the different story branches in a story problem's solution space.

1998 ACM Subject Classification I.2.8 [Problem Solving, Control Methods, and Search] Plan execution, formation and generation

Keywords and phrases artifical intelligence, planning, narrative, comparison, story

Digital Object Identifier 10.4230/OASIcs.CMN.2016.9

1 Introduction

Branching story games, such as Mass Effect and The Walking Dead, have gained popularity for creating unique playing experiences by adapting story content in response to user actions. Research in interactive narrative (IN) uses *automated planning* to generate story plans as they offer an action-oriented and causally-related representation consistent with narrative [3, 19, 34]. For a given story problem, a story planner can generate multiple story plan solutions, all of which equally-satisfy the story problem definition but contain different story content. Characterizing the qualitatively different story plans in a story problem's solution-space would capture the story branches experienced by a user and enable plan-based INs to adapt in a story-theoretic manner to user actions.

A first step to characterizing the different story branches in a story plan problem is to develop the capability to compare two story plans based on their narrative-theoretic

* This material is based upon work supported in whole or in part with funding from the Laboratory for Analytic Sciences (LAS). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the LAS and/or any agency or entity of the United States Government.

7th Workshop on Computational Models of Narrative (CMN 2016).
Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No. 9; pp. 9:1–9:16
Open Access Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

properties. Conflict Partial Order Causal Link (CPOCL) plans represents story plans with additional narrative-theoretic structures, namely character subplans and conflict, derived from classical plan structures. While conforming to the formal structures of classical planning limits the narrative representation, it allows us to directly leverage the *automated planning* community of research for topics such as plan comparison. Thus, in this paper we limit the scope of narrative-theoretic plan comparisons to those with a CPOCL representation.

Historically, plan comparison research has focused on metrics to assess minimal length or optimal plans. More recently however, mixed-initiative planning systems have employed additional advanced distance metrics to capture the differences in plan syntax. Still, these distance metrics have been restricted to domain-independent properties and do not account for domain-specific semantics of story plans. We address this limitation by differentiating between the necessary *syntactical* structure required by a story plan to be considered a valid plan and the additional *semantic* structures derived by CPOCL. This difference allows us to define a story plan *summary*. By first summarizing a story plan's narrative-theoretic content and then making comparisons allows us to emphasize the narrative-theoretic representation in story plans to more accurately capture a users experience.

The first contribution we make is the definition of a computational model to summarize the story semantics of CPOCL story plans. The model captures key story semantics of the logical progression of plot and character believability, both factors associated with narrative comprehension. Our second contribution is the definition of a CPOCL story plan distance metric. This metric builds upon previous work on plan distance metrics to calculate differences in story semantics between plans. We bring these two models into operation together on a story domain-problem pair to demonstrate their sensitivity to changes in story plan semantics and contrast them to existing domain-independent plan distance metrics.

2 Previous Work

In this section, we review key areas of research that support developing a model of CPOCL story plan summarization and distance metric. First, we outline previous work in story planning to highlight the specific story plan structures that are applicable to plan summarization research. This is followed by assessing research on plan comparison for concepts that translate to story plan comparison. Finally, we discuss story comparison from a wide variety of disciplines to identify principles to operationalize in story plan comparison.

2.1 Story Plan Summarization

The origins of automated story generation can be traced to the Tale-Spin system [17]. Tale-Spin makes use of an inference engine to direct character action when an initial state and environment are provided. Automated planning was later used to capture author goals in the Universe system [15], which focused on causal coherence of actions in the stories it generated and ensured specific author outcomes in a goal state. Following the advances made by Universe, the Fabulist system [26] developed the Intent-driven Partial Order Causal-Link (IPOCL) planner which determines the intentions each character could have and motivates these intentions through story actions. This supports character believability which is linked to narrative comprehension. Conflicting character goals are viewed as an essential literary element of stories [1, 12]. Work to operationalize a notion of conflict by Ware *et al.* [33] resulted in the Conflict Partial Order Causal Link (CPOCL) algorithm, which uses non-executed steps to model foiled character goals. While other story generation mechanisms are less formal and explore less quantifiable aspects of narrative (e.g MEXICA [25]), we limit the scope of this paper to story plans with more formal representation as it allows the direct application of previous work in classical planning to plan summarization and comparison.

The work of Clement *et al.* [5] uses abstraction in a Hierarchical Task Network (HTN) planner to dramatically reduce the run time of planning algorithms. Myers *et al.* [20] investigated more appropriate ways to interact with a user with an approach that seeks an explanation of a plan from the knowledge stored in a meta-theory, rather than from the syntax of a plan. Another attempt to provide a human-centric interface to plan summaries was the research by Mellish *et al.* [18] that used Natural Language Generation (NLG) to elucidate the steps of a generated plan. Additional work by Myers [21] uses a temporal based domain-theory to summarize key temporal regularities and exceptions in support of domain-independent summarization techniques in temporal plans.

The use of Grice's maxim of quantity [11] motivated the work of Young [35] to find an equilibrium between the level of detail and abstraction communicated from plan structures. Plan summaries using the Local Brevity Algorithm (LBA) are computed by weighting the importance of individual plan steps based on the causal inferences of story events as identified by Trabasso and Sperry [31]. Specifically, Trabasso and Sperry [31] define a causal chain to contain an opening, a closing, and to continue the chain of events. The causal chains of six folk-tales were used to compute both a story event's membership on the causal chain and causal connectedness. The experimental results showed that these two factors account for a significant portion of the variance in agreement of story event recall among study participants and did not have a statistically significant interaction. This leads the authors to conclude that causal inferences are foundational to the process of story representation in memory and a source of how we privilege information. This finding will prove to be valuable research, as it will form the basis for our arguments in Section 3.2.

2.2 Story Plan Comparison

Like plan summarization, many plan distance metrics are domain independent. In classical planning, plan distance is well-served by assessing a candidate plan's syntactic structure to determine whether it is *minimal* or *redundant* with respect to an optimal plan. The application of planning algorithms to new domains where human decision-making is augmented, termed mixed-initiative planning, has increased interest in developing more robust plan distance metrics to capture differences in plan structure.

Research by Srivastava *et al.* [30] and subsequently by Nguyen *et al.* [23] into domain-independent plan distance metrics founded on Jaccard distance resulted in the action, causal-link and state-space distance metrics. The authors define action distance as

$$\delta_A(p, p') = 1 - \frac{|A(p) \cap A(p')|}{|A(p) \cup A(p')|}, \tag{1}$$

where A is the actions of a plan and p, p' are complete plans. Similarly, the causal link distance is

$$\delta_C(p, p') = 1 - \frac{|C(p) \cap C(p')|}{|C(p) \cup C(p')|}, \tag{2}$$

where C is the set of causal links between plan actions of a plan and p, p' are complete plans. These two metrics are of particular interest, as they capture the differences in actions and causal links, which enable plans to model story properties.

There have also been several other domain-independent distance metrics developed. Goldman and Kuter [10] designed *normalized compression distance* as a way to measure the conditional information from one plan to another. Plan landmarks, on the other hand, capture the essential conditions all solution plans must contain and are used by Bryce [4] to assess the differences between plans in the *landmark distance* metric. Lastly, Roberts *et al.* [28] use the differences in plan length to compute *parsimony*. Each of these metrics capture syntactic differences between plans; however, none can be immediately applied to story plan semantics.

To date, computational models of story comparison have tended to apply comparison methods from other disciplines. For example, representing story actions as a sequence of string characters enables the use of sequence-alignment tools commonly used in textual analysis (spell-checking) and bioinformatics (DNA sequence alignment). Porteous *et al.* [24] manipulate characters' social relationships to explore the qualitative differences in interactive narratives as reflected in Levenshtein distance [16]. Another sequence-alignment algorithm is evaluated by Fay *et al.* [8] who make use of the Needleman-Wunsch algorithm [22] to compare linear story sequences from Genesis and show its ability to greatly reduce the compute time for matching and comparison of stories. Finally, the use of an intelligent Drama Manager (DM) to negotiate the balance between authorial intent and player autonomy in interactive systems motivated the work of Jones and Isbell [13], who empirically evaluated story similarity metrics when gameplay is represented as a Targeted Trajectory Distribution-Markov Decision Process (TTD-MDP) [27]. While methods from other disciplines have shown to be helpful in computing the differences in stories, they do not make use of human-centered models of story comparison.

Human-centered research on story comparison has also been a recent focus, such as the work by Fisseni and Lowe [9] on non-structural dimensions of narratives for the purpose of story equivalence. Of particular relevance to the work presented in this paper, Kypridemou and Michael [14] validate the *common summary* of two stories as a model of their similarity. Through human subject validation, the authors confirm that the more appropriate a common summary of two stories, the more similar the two stories are judged to be. While they did not employ a computational model for generating summaries of the stories, their work offers a valuable human-centered principle upon which to define a computational model of a story summary for use in story comparisons.

2.3 Summary of Previous Work

Story generation models that adhere to classic planning principles offer a rich causal representation of events, which has lead to the capability to represent character intentions [26] and conflict [32] as story plan semantics. Unfortunately, plan comparison metrics have focused on syntactic properties of plans [4, 10, 23, 28, 30] not domain-specific semantics related to story plan semantics. Reasonably, computational models of story comparison have leveraged methods from other disciplines [8, 13, 24]; however, human-centered models of story comparison have shown that a common summary of two stories is important to similarity assessments [14]. A plan summarization model based on a story event's causal degree and membership on the causal chain [35] offers a principled plan-based representation to begin comparing CPOCL story plans.

3 Story Planning

As discussed in Section 2.3, a limitation of domain-independent comparison metrics is their focus on measuring syntactic differences between plans. This results in the difference between the story semantics of two plans not being represented in current plan comparison metrics.

We address this limitation by first defining the story plan semantics that support a story plan summarization model. This narrative-theoretic summarization model captures character intentions and causally significant actions from the syntax required to conform to planning formalisms. This is followed by defining a story plan distance metric which computes the differences between two story plan summaries.

3.1 Story Plan Semantics

Story planning algorithms generate story plans to solve a story planning problem, of which there can be many equally-satisfying solutions. The story semantics between these story plan solutions can vary and are not captured in current plan distance metrics. We provide precise definitions of story plans and their related parts to both leverage previous work in classical planning and to ground work from other disciplines in this context. In the tradition of being precise, the following definitions are formalized based on the established body of work in story planning.

A planning domain theory models the way in which the world can change through the application of actions to a world state. Story planners implement algorithms which instantiate and combine actions from a domain theory to reach a goal state from some initial state.

▶ **Definition 1** (Action)**.** An action is defined as the preconditions that must be satisfied before its execution and the effects that result. A precondition is a function-free positive literal in a state space and the conjunction of an action's preconditions must evaluate to true before it can execute. An action's effects are function-free positive literals whose conjunction is the result of the change in state space when an action is executed.

Together with an action's name and parameter list, the precondition and effects describe an *action schema*. An action schema can be instantiated into various forms, dependent on the literals to which the variables unify. Note that the use of the terms *action, step,* and *operator* are used interchangeably.

▶ **Definition 2** (Story planning problem)**.** A story planning problem Φ is a four-tuple $\langle \mathcal{I}, A, \mathcal{G}, \Lambda \rangle$ where $\mathcal{I}$ is a conjunction of function-free ground literals which are true in the initial state, A the set of symbols referring to character agents, $\mathcal{G}$ a conjunction of function-free ground literals which are true in the goal state, and Λ a set of action schemata. This definition is in the way of Riedl & Young [26].

An action's preconditions are satisfied through causal links to an earlier step's effects and, in turn, its effects can satisfy preconditions to later steps.

▶ **Definition 3** (Causal Links)**.** Causal links are denoted $s \overset{p,q}{\to} u$, where s, u are steps in S of story plan P, with an effect p and precondition q respectively. In this case, q is satisfied for u because s had p as an effect and there exists a literal r that unifies with p and q. The step s is a causal parent of u, while u is the causal child of s. The causal parents of s are causal ancestors of u and the causal children of u are the causal descendants of s. These transitive relations extend until an action with no parents or children is reached.

Steps which establish causal links for subsequent steps are at risk of some other step's effect undoing a causal parent's effect, thereby transforming the state of the world into one which the causal child action cannot execute.

▶ **Definition 4** (Causal link threat)**.** A causal link threat occurs when a causal link is established $s \overset{p,q}{\to} u$, and some other step w has the effect $\neg p$ and could be executed after s but before u. Executing w in this interval means the precondition q of u is no longer satisfied by s and u will not execute.

A causal link threat to $s \overset{p,q}{\rightarrow} u$ can be resolved by introducing an ordering constraint such that w is executed before s or after u. Resolving all causal link threats to this causal link guarantees that p remains true after s is executed until u is executed, preserving the causal link with u.

▶ **Definition 5** (Ordering constraint). An ordering constraint is of the form $w \prec s$ and is interpreted as w must be executed at some time before s.

Establishing causal links between steps requires maintaining which precondition and effect variable pairs must unify. Additionally, it may be desired that the free variables in action parameters are not equivalent. Binding constraints are used to capture both these conditions.

▶ **Definition 6** (Binding constraint). A binding constraint is a pair of variables (u, v) or negated pair $\neg(u, v)$ where the pair must unify or not be allowed to unify, respectively.

Characters are an important element of stories. In support of this, character intention frames were implemented in story plans to justify individual character actions in service of their goals.

▶ **Definition 7** (Intention Frame). An intention frame in a plan P is a tuple of five elements $\langle c, g, m, \sigma, T \rangle$ where c is a character, a goal that c intends to make true is represented by g. The motivating step m is in $S(P)$ with the effect $intends(c, g)$, the satisfying step σ is also in $S(P)$ and has g as an effect. The set of steps T is a subset of steps from $S(P)$ taken by c to achieve the goal effect g. T is called the character's subplan to achieve the goal effect g. All steps in T must occur after the motivating step m and before σ.

The POCL (Partial Order Causal Link) family of planners searches through a story plan space to produce story plans of the form defined below.

▶ **Definition 8** (Story plan). A story plan P is a tuple of five elements $\langle S, B, O, L, I \rangle$ where the set of steps is S (with executed steps denoted S_e and non-executed S_{ne}), the set of binding constraints on the free variables of S is defined as B, the partial ordering of the steps in S defined as O, L the set of causal links joining steps from S, and finally I, the set of intention frames which define character subplans in S. This definition is consistent with the definition of a CPOCL plan in Ware *et al.* [33].

The differentiation of executed and non-executed steps is made in support of representing character conflict. Non-executed steps are part of foiled sub-plans that a character had the intent of completing, but could not due to an unresolved causal link threat with another step.

▶ **Definition 9** (Story plan solution). A story plan P is a *solution* to the story plan problem Φ if its actions, which has $\mathcal{I}(\Phi)$ as effects of the first action and $\mathcal{G}(\Phi)$ as preconditions to the goal step, have no open preconditions and is *consistent*. A story plan is *consistent* if there are no cycles in the ordering constraints $O(P)$ and no causal link threats remain between two *executed* steps in $S_e(P)$.

While these definitions may appear verbose, they are necessary to formally characterize the representation of a story plan solution. This representation leads to a story planning problem to, in fact, have many story plan solutions and they can differ in both their syntax (*e.g.* actions) and in more human centric ways. The next sections focus on operationalizing the ability of humans to recall certain events as a model of summarization and comparison.

3.2 Story Plan Summarization

The story semantics detailed in Section 3.1 lay the principles for a computational model of a story plan summary. The importance of causal degree and the causal chain in the recall of story events [31] suggests a mechanism to summarize the logical progression of plot. Additionally, intention frame summaries play an important role in the comprehension of a story. These two properties will form the basis of a story plan summary, which is incrementally defined below.

We capture a story plan's causal links in the $n \times n$ causal matrix $\mathcal{S}_P$,

$$\mathcal{S}_P = \begin{bmatrix} 0 & s_{12} & s_{13} & \dots & s_{1n} \\ 0 & 0 & s_{23} & \dots & s_{2n} \\ 0 & s_{32} & \ddots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & s_{n-1n} \\ 0 & 0 & \dots & \dots & 0 \end{bmatrix}. \quad (3)$$

where $n = |S(P)|$. The s_{ij} entry contains the number of effects of step i used as preconditions by step j. Note the zeroes in the initial state's column (s_1) and of the goal state's row (s_n), which denote the lack of preconditions and effects, respectively. It is necessary to compute $\mathcal{S}_P$ for each story plan solution, and not for the action schemata, as $\mathcal{S}_P$ captures the number of times the effects of a step are used in a story solution plan, which can be equal to, greater than, or less than the effects in the action schema.

To support the identification of important events in a story plan, the causal degree of the step i, s_i, is computed directly from $\mathcal{S}_P$ by adding both the ith row sum,

$$deg_c^-(s_i) = \sum_{j=1}^{n} s_{ji}, \quad (4)$$

a step's used effects, and column sum,

$$deg_c^+(s_i) = \sum_{j=1}^{n} s_{ij}, \quad (5)$$

a step's satisfied preconditions.

▶ **Definition 10** (Causal degree)**.** The causal degree of a step s_i in a plan P is the sum of the step's preconditions $deg_c^-(s_i)$ and the sum of the effects used by each causal child $deg_c^+(s_i)$, defined in Equation6:

$$deg_c(s_i) = deg_c^-(s_i) + deg_c^+(s_i). \quad (6)$$

The criteria for a causal chain having a motivation, beginning, and end emerged from Trabasso and Sperry's work [31] and is directly applicable to story plans as character goals, the initial state, and the goal state, respectively.

▶ **Definition 11** (Causal chain)**.** A causal chain C of a story plan solution P, $C(P)$, is a subset of $S(P)$, which consists of all the steps that are causal ancestors of the goal step, plus those steps' causal descendants. The causal chain excludes both the initial and goal step, s_1 and s_n, respectively.

The causal chain excludes the initial and goal steps. While they are both elements of $S(P)$, they are never executed by the planner. Rather they are only states to delineate the start and end of a plan. Including these steps in the causal chain would only serve to complicate deriving story plan summaries.

▶ **Definition 12** (Important steps)**.** The important steps E of a story plan solution P is the set of *executed* steps in the causal chain $C(P)$ with the *highest* causal degree computed from the matrix $\mathcal{S}_P$.

While non-executed steps provide a construct to identify conflicts between characters, they are excluded from important step calculations as their effects are not realized in the story world state. Additionally, we use highest causal degree to mean an exact number, and it is expected that the number of important steps in story plan will be small.

The set of steps T of an intention frame $I = \langle c, g, m, \sigma, T \rangle$ (Definition 7) captures the means by which a character achieves a goal and it has been used to reflect character traits [2]; however, this could include steps without any story plan semantics, such as movement actions. While such actions are necessary for the progression of the story plan, we do not have a character-centric model to capture which steps are meaningful in story plans. In order to avoid these steps influencing our comparisons, we simply remove T to form an intention frame summary, until such a time that plan steps can be validated as a robust character trait model.

▶ **Definition 13** (Intention frame summary). An intention frame summary j of some intention frame in I from the story plan P is a four-tuple $\langle c, g, m, \sigma \rangle$ where c,g,m and σ are preserved from Definition 7. The set of intention frame summaries of P is denoted as $J(P)$, where each intention frame in $I(P)$ has a corresponding intention frame summary in $J(P)$.

We combine the causally important steps (Definition 12) and intention frames summaries (Definition 13) to define the *Important-Step Intention-Frame* story plan summary as

$$\psi^{ISIF}(P) \quad = \quad \langle E, J \rangle, \tag{7}$$

where P is a story plan solution, important steps E are computed from the steps $S(P)$ and intention frame summaries $J(P)$ derived from the intention frames $I(P)$.

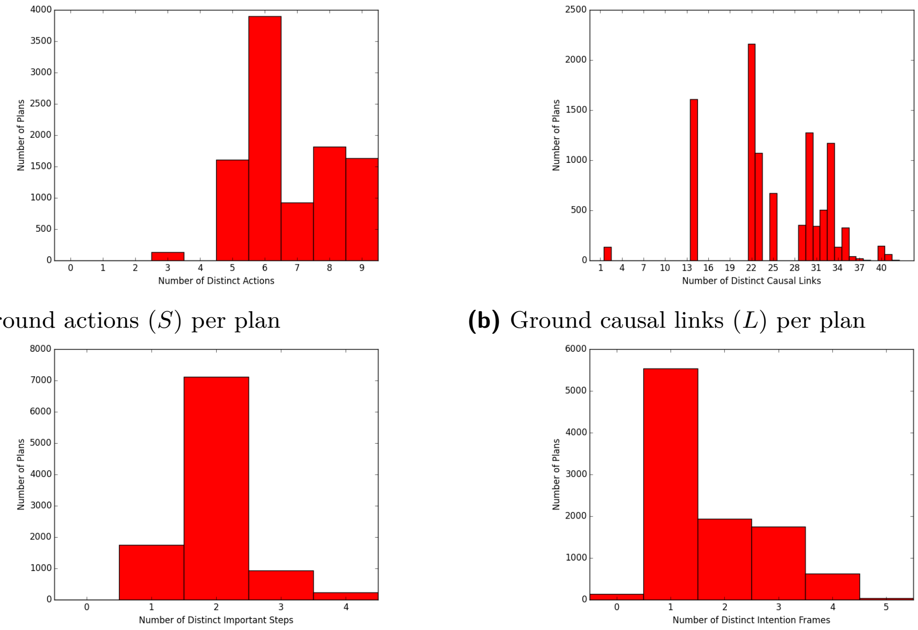
3.3 Story Plan Distance Metric

Previous plan comparison metrics compare, on a limited basis, the semantics between story plans as they apply uniform importance to all plan syntax. We address this limitation by using the formalization of story plan summaries in Section 3.2, which extracts the story plan semantics to make them available for direct comparison. We first define the notion of the distance between two story plan summaries formally.

▶ **Definition 14** (Story plan distance metric). A distance metric between two story plan summaries, ψ_1 and ψ_2, is a function Δ where $\Delta(\psi_1, \psi_2) \longrightarrow [0, 1]$. The value of zero represents perfect similarity between the two summaries, where the value of one denotes perfect dissimilarity.

In order for any Δ to be mathematically consistent as a distance metric, it needs to satisfy the identity, symmetry, and triangle inequality properties. As a result, proper subsets of plans must not be made by the metric itself, as it violates the identity property of a distance metric. As an example, a metric that only compared the penultimate steps of two plans would assess them as being identical when their penultimate step is equal, when in fact the rest of the steps of the plan could be all different. To maintain mathematical consistency, the distance metric defined in this section is based on *story plan summaries*.

Previous work in plan comparison has made use of the Jaccard distance metric to compare plans. While not theoretically motivated or explicitly denoted by the authors in [23, 30], Jaccard distance is often used as a simple distance metric for assessing non-normal data, with no underlying distribution and no linear relationships. A key feature of the Jaccard distance is the use of the *intersect* set operator in the numerator as it accounts for the



(a) Ground actions (S) per plan

(b) Ground causal links (L) per plan

(c) Ground important steps (E) per plan

(d) Unique intention frames (I) per plan

Figure 1 Solution space distributions of Π.

common elements between two story plan summaries, a factor for human judgment of story similarity [14].

We define the Jaccard distance metric between two story plan summaries as

$$\delta_{ISIF}(\psi_1^{ISIF}, \psi_2^{ISIF}) = 1 - \frac{1}{2}\left(\frac{\left|E(\psi_1^{ISIF}) \cap E(\psi_2^{ISIF})\right|}{\left|E(\psi_1^{ISIF}) \cup E(\psi_2^{ISIF})\right|} + \frac{\left|J(\psi_1^{ISIF}) \cap J(\psi_2^{ISIF})\right|}{\left|J(\psi_1^{ISIF}) \cup J(\psi_2^{ISIF})\right|}\right), \qquad (8)$$

where ψ_1^{ISIF} and ψ_2^{ISIF} are ISIF story plan summaries (Equation 7), E is the important events of a story plan summary, and J is the intention frame summaries of a story plan summary. We introduce a factor of $\frac{1}{2}$ over the Jaccard similarity of E and J to ensure the metric remains between 0 and 1. This also has the consequence of equally weighting both E and J. We can interpret the δ_{ISIF} distance metric as story plan summaries which have little in common in terms of intention frames and causally important steps will have a score close to 1, whereas similar summaries will have a δ_{ISIF} close to 0.

3.4 Summary of Definitions

Past research in plan comparison has classified distance metrics into domain-independent and domain-specific. While the distance metric described in Section 3.3 can be classified as a domain-specific metric, the use of narrative-theoretic constructs in the story plan summary affords the metric to generalize to all story domains and would classify as a *Narrative Metric* under the StoryEval framework [29]. In practice, story plan distance metrics are constrained to the representational capability of story planners. In Section 4, we turn our focus to determining whether the δ_{ISIF} distance metric can capture differences in story semantics in the solution space of a simple story planning problem.

4 Experimental Results

We desire to characterize the story branches in an IN that uses automated planning for story generation. This requires differentiating between story plans which equally-satisfy a single story planning problem. With this in mind, the evaluation uses a single solution plan-set generated from the Glaive narrative planner [32], which has been previously used for IN story generation, and does not consider comparing plans across story planning problems or story planning domains.

Our commitment to the CPOCL story plan representation introduces another consideration in evaluating the δ_{ISIF} story plan distance metric. We must consider what alternative methods exist to compare two CPOCL story plans. Previous story comparison efforts fall into two classes. The first are human judgments of story comparisons [9, 14], making them difficult to formalize to a CPOCL representation. A second class uses existing domain-independent measures that accept the author's representation as input, such as the use of string comparison [8, 24] and subgraph isomorphism [7]. In short, previous story comparison methods either require capturing the stories in CPOCL story plans or adapting CPOCL story plans to the method's required input. Rather than undertake the task of further formalization, we use CPOCL plans' adherence to automated planning formalisms to find appropriate plan-based domain-independent distance metrics to compare the δ_{ISIF} story plan distance metric. Conveniently, established plan-based distance metrics which capture the action-oriented and causally-linked representation overlaps between narrative and automated planning already exist, namely the action (δ_A) and causal-link (δ_C) distance metrics.

The goal in the preliminary evaluation in the following section is two-fold. Firstly, to investigate the ability of δ_{ISIF} story plan distance metric and existing distance metrics to capture the story branches in an IN story problem's solution plan-set. Secondly, demonstrate that when comparing a pair of plans the $\delta_{ISIF} \bullet$ story plan distance metric is more sensitive to small syntactical differences that result in significant story semantic differences than existing automated-planning distance metrics.

4.1 Story Domain and Problem

To demonstrate the differences between the different distance metrics, we use the *space* story domain, which was previously defined to evaluate the Glaive planner and highlight the CPOCL representation [32]. The space story domain was the simplest of the domains used in the evaluation and consists of two variable types, eleven predicates, and ten actions. This domain was used as it presents enough complexity to highlight the differences between distance measures, but not so much as to lead the analysis into extreme cases.

The story problem used is one in which an astronaut ("Zoe") must resolve conflicting intentions of exploration and self-preservation as a volatile planet becomes inhospitable. This story problem is referred to as the *exploration* problem. The initial state of the *exploration* problem consists of fourteen predicates and five character intentions, three belonging to Zoe and two to an alien. A key feature of this problem is the goal step consisting of a single predicate; $\neg$(*habitable surface*), which is only possible by one action in the domain; *begin-erupt*.

4.2 Solution Plan-Set Analysis

The Glaive heuristic search planner (HSP) [32] was able to generate a solution plan-set of $10,000$ CPOCL solution plans, Π, to the exploration problem. While each story plan is

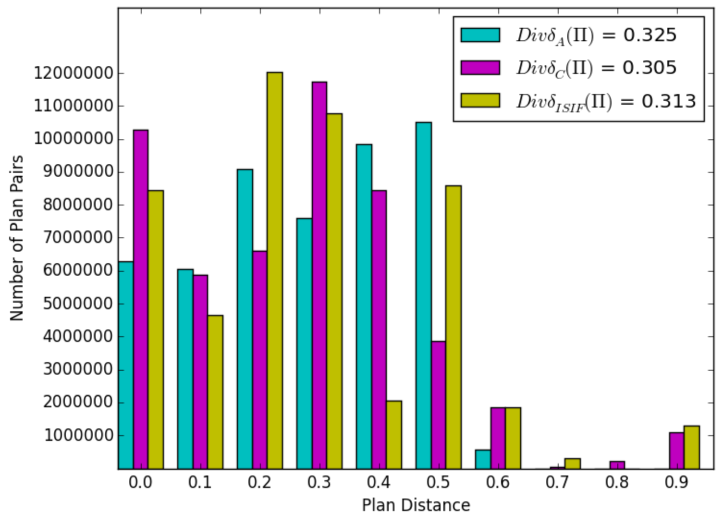


Figure 2 Solution diversity using different distance metrics.

equally-satisfying in their ability to solve the story planning problem, significant differences exist between their story semantics. Of the plan comparison metrics reviewed, δ_A (Equation 1) and δ_C (Equation 2) are the most relevant to compare with δ_{ISIF} as they capture first principle overlaps between planning and stories: namely, action-oriented nature and causally-related events. Additionally, the metrics do not require total orderings, a necessary criteria when comparing partial order plans.

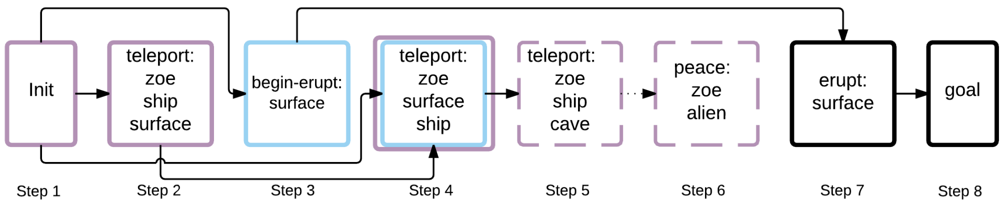
We can obtain some insight of the pertinent elements of Π to each distance metric in Figures 1a–1d. Specifically, both the distribution of steps (Ground actions, Figure 1a) and causal links (Figure 1b) show that a non-trivial number of elements are used in the calculations made by the δ_A and δ_C distance metrics. While the number of important steps (Figure 1c) and intention frames (Figure 1d) are smaller relative to ground actions and causal links, they are combined together in the δ_{ISIF} metric to achieve greater size. To sum, we can see from Figures 1a–1d that despite the simplicity of the space exploration problem, the distribution of the relevant properties in plans of Π support substantive syntactic and story semantic comparisons.

We use a standard calculation from the automated planning community to capture the differences in the solution space. The *plan-set diversity* [6] is a pairwise comparison made between every solution plan in Π,

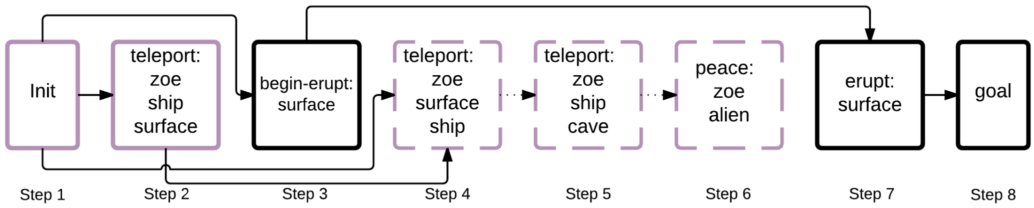
$$Div(\Pi) \quad = \quad \frac{\sum\limits_{\pi,\pi' \in \Pi} D(\pi,\pi')}{\frac{|\Pi| \times (|\Pi|-1)}{2}}, \tag{9}$$

where π and π' are plans in a plan-set, Π, and $D(\pi,\pi')$ is a distance metric. The distribution of plan-pair comparisons using the three distance metrics and their associated diversity measures are shown in Figure 2.

We can observe the agreement in plan-set diversity using the three metrics (0.325, 0.305, 0.313); yet, this does not capture the differences in distribution. Take for instance the lack of scores near 1.0 for the δ_A metric but the presence of such scores for δ_{ISIF}. This is an artifact of the *exploration* problem's goal state containing the $\neg$(*habitable surface*) predicate, which requires all solutions to have the erupting volcano action as it is the only action with this



(a) Plan π_1



(b) Plan π_2

Figure 3 Story plans with intention frames in color.

Table 1 Story plan properties.

π_i	$\|S(\pi_i)\|$	$\|L(\pi_i)\|$	$\|I(\pi_i)\|$	$\|E(\pi_i)\|$	$C(\pi_i)$	$E(\pi_1)$
π_1	8 (6 executed, 2 non-executed)	35	2	2	$\{s_2, s_3, s_4, s_7\}$	$\{s_2, s_4\}$
π_2	8 (5 executed, 3 non-executed)	35	1	1	$\{s_2, s_3, s_7\}$	$\{s_2\}$

predicate as an effect. The presence of scores at 1.0 when using the δ_{ISIF} metric indicates that there is more than a single narrative experience in the solution plan-set. While we can glean some insights from the plan-set diversity measure when using the δ_{ISIF} metric, plan-set diversity does not capture, for instance, the exact number of branches in a story problem's solution space. We leave the development and application of other plan-set measures that capture important IN characteristics for future work.

4.3 Solution Plan-Pair Analysis

The solution space distance metric results presented in Section 4.2 illustrate qualities of the solution space coupled to properties of the story domain and story problem; however, specific examples of the story plan distance measure are needed to demonstrate the effectiveness of the story summary distance metric. The following example uses two story plans, π_1 and π_2, from Π to illustrate when small syntactic differences have large semantic differences between story plans.

We observe the difference in structure of the two plans in Figure 3, where executed actions are solid rectangles, non-executed actions dashed rectangles, and the links between them represent the existence of one or more causal links. The plans have a similar syntactic structure, containing the same number of steps and causal links (8 and 35 respectively) with a single difference in s_4 (executed vs non); however, this single syntactic difference changes the story semantics in a significant manner by affecting the number of important events and intention frames.

We capture the common intention frame between the two plans in purple and the additional intention frame in π_1 with blue. The common intention frame in each plan is Zoe's goal to make peace with the alien, which is motivated in s_1, and her first action towards

$$\mathcal{S}_{\pi_1} \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ deg_c^- \end{array} \begin{array}{c} \begin{array}{ccccccccc} & & & & & & & & deg_c^+ \end{array} \\ \left[\begin{array}{ccccccccc} 0 & 6 & 1 & 5 & 5 & 6 & & & 23 \\ 0 & & & 1 & & & & & \mathbf{1} \\ 0 & & & & & & 1 & & 1 \\ 0 & & & & 1 & & & & \mathbf{1} \\ 0 & & & & & & & & x \\ 0 & & & & & & & & x \\ 0 & & & & & & & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{6} & 1 & \mathbf{6} & x & x & 1 & 1 & \end{array}\right] \end{array} \qquad \mathcal{S}_{\pi_2} \begin{array}{c} \begin{array}{ccccccccc} & & & & & & & & deg_c^+ \end{array} \\ \left[\begin{array}{ccccccccc} 0 & 6 & 1 & 5 & 5 & 6 & & & 23 \\ 0 & & & 1 & & & & & \mathbf{1} \\ 0 & & & & & & 1 & & 1 \\ 0 & & & & & & & & x \\ 0 & & & & & & & & x \\ 0 & & & & & & & & x \\ 0 & & & & & & & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{6} & 1 & x & x & x & 1 & 1 & \end{array}\right] \end{array}$$

Figure 4 Causal Matrices.

Table 2 Example story plan-pair properties.

π_i, π_k	$\delta_A(\pi_i, \pi_k)$	$\delta_C(\pi_i, \pi_k)$	$\delta_{ISIF}(\pi_i, \pi_k)$
π_1, π_2	0.11 (1 - $\frac{8}{9}$)	0.00 $(1 - \frac{35}{35})$	0.50 $(1 - \frac{1}{2}(\frac{1}{2} + \frac{1}{2}))$

completing this goal is s_2; however, the surface begins to erupt in s_3 making Zoe $\neg safe$. Zoe now has conflicting goals and must choose between continuing her existing goal while $\neg safe$ or adopting a new goal of ensuring her life is preserved by teleporting back to the ship and becoming $safe$. We can observe that s_4 is executed in π_1, which constitutes Zoe's first and satisfying step of her goal to be $safe$ instead of continuing her intention to make peace with the alien. This additional intention frame contrasts with π_2 where s_4 is a non-executed step, indicating Zoe does not adopt the goal of being $safe$, thus only a single intention frame of her search for the alien.

The change of s_4 from non-executed to executed step also affects the number of causally-important steps. Both plans have s_2 as an important step, since it is both on the causal chain and has the highest causal degree, seven as $deg_c^-(s_2) = 6$ and $deg_c^+(s_2) = 1$ (see Figure 4). Identically, we observe that the causal degree of s_4 is also seven, $deg_c^-(s_4) = 6$ and $deg_c^+(s_4) = 1$; however since the definition of a causal chain requires steps to be executed, s_4 is excluded from $E(\pi_2)$. This additional member of $E(\pi_1)$ demonstrates how important steps can vary from equally-satisfying solutions and have a semantic impact on story plan content.

The semantic properties, as well as the syntactic properties, which are relevant for computing the three distance metrics between π_1 and π_2, are shown in Table 1. We observe in Table 2 that the syntactic change in s_4 between the plans results in a difference of 0.11 when calculated using the δ_A metric. The same change results in the δ_C metric calculated as 0.00, as the causal structure is identical from π_1 to π_2. In contrast, the δ_{ISIF} calculation results in a much more significant 0.50 as it captures the semantic importance of the small syntactic change. This relatively large difference between the domain-specific (δ_{ISIF}) and domain-independent (δ_A, δ_C) metrics demonstrates the ability of a metric based on story plan semantics to more accurately capture subtle differences in story plan syntax.

4.4 Summary of Experimental Results

In this section, we described the *space* domain's *exploration* problem then evaluated three plan distance metrics on their ability to compare two equally-satisfying story plans. The δ_A metric provided a useful insight into the composition of the problem definition, namely all solution plans will contain the *begin-erupt* step. It further suggests that meaningful insights could be ascertained by describing the solution space using the δ_{ISIF} metric. Lastly, we used two equally-satisfying story plans with a single syntactic difference to illustrate the difference in story plan semantics that δ_{ISIF} captures. These results support semantic summaries as a promising approach to story plan comparison.

5 Limitations

We have presented and demonstrated formalized models of story plan summarization and comparison for a CPOCL story plan representation. It is worth emphasizing that while CPOCL represents validated forms of character intentions and conflict, its adherence to a strict automated planning formalism limits its capability to represent narrative. Story comparisons extend far beyond those stories which CPOCL and the models presented in the paper capture. As more advances are made in deriving story semantics from classic plan representations into story plans, summarization models and comparison metrics will be able to build on the contributions of this paper.

6 Conclusions

Branching stories have enjoyed success in recent games and quantifying the story branches in a narrative planner requires a story-plan comparison metric. In support of addressing this need, this paper draws together research from both computer science and narrative theory to synthesize two primary contributions to advance CPOCL story plan comparison.

The first is the definition of a computational model for a story plan summary supported by cognitive psychology and narratology. A story plan summary uses story plan steps' causal degrees, in addition to character intention frame summaries to capture story plan semantics while ignoring syntactic structure of less importance. The resulting story plan summary is a concise semantic representation of a story plan and provides a foundation upon which to compare other story plans.

A second contribution is the definition of the δ_{ISIF} story plan distance metric. The distance metric leverages previous work in plan comparison, namely the use of Jaccard similarity, to compare domain-specific properties. While the formulation of the distance metric weights the score equally over the important step and intention frame components, it avoids learning weighted parameters and violating the definition of a distance metric.

We demonstrated the above contributions using a CPOCL solution plan-set to the *exploration* problem. At both the solution plan-set level and in specific story plan-pair examples, we calculate the differences between two existing syntactic plan comparison metrics and the δ_{ISIF} story plan comparison metric. We exhibit that when using the δ_{ISIF} distance metric in plan-set comparisons it can confirm the existence of mutually-exclusive story plan summaries, while at the story plan-pair level it captures small syntactic differences that have larger semantic differences. These results are significant, not only do the story plan summarization model and distance metric enable more robust comparisons of story plans but we expect them to generalize to all story plan domains due to their cognitive psychology and narratology foundations. We expect to validate these models in a human subject evaluation in the near future.

References

1 H. Porter Abbott. *The Cambridge Introduction to Narrative*. Cambridge University Press, second edition, 2008. doi:10.1017/CBO9780511816932.

2 Julio César Bahamón and R. Michael Young. Toward a computational model for the automatic generation of character personality in interactive narrative. In *Intelligent Virtual Agents*, pages 520–522. Springer, 2012.

3 Mieke Bal. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, 1997.

4 Daniel Bryce. Landmark-based plan distance measures for diverse planning. In *ICAPS*, pages 56–64, 2014. URL: `http://www.aaai.org/ocs/index.php/ICAPS/ICAPS14/paper/view/7903/8011`.

5 Bradley J. Clement, Edmund H. Durfee, and Anthony C. Barrett. Abstract Reasoning for Planning and Coordination. *Journal of Artificial Intelligence Research*, 28:453–515, 2007. `doi:10.1613/jair.2158`.

6 Alexandra Coman and Héctor Muñoz avila. Generating Diverse Plans Using Quantitative and Qualitative Plan Distance Metrics. In *AAAI Conference on Artificial Intelligence*, pages 946–951, 2011.

7 David K. Elson. Detecting Story Analogies from Annotations of Time, Action and Agency. In *Proceedings of the Third Workshop on Computational Models of Narrative*, pages 91–99, 2012.

8 Matthew P. Fay. Story Comparison via Simultaneous Matching and Alignment. *Computational Models of Narrative Workshop*, pages 100–104, 2012.

9 Bernhard Fisseni and Benedikt Löwe. Which dimensions of narrative are relevant for human judgments of story equivalence? *Computational Models of Narrative Workshop*, pages 114–118, 2012. URL: `http://dare.uva.nl/document/362429`.

10 Robert P. Goldman and Ugur Kuter. Measuring Plan Diversity : Pathologies in Existing Approaches and A New Plan Distance Metric Normalized Compression Distance for Plan. In *AAAI Conference on Artificial Intelligence*, 2015.

11 H. Paul Grice, Peter Cole, and Jerry L. Morgan. Syntax and semantics. *Logic and conversation*, 3:41–58, 1975.

12 David Herman. *Narrative theory and the cognitive sciences*. Center for the Study of Language and Information, 2003.

13 Joshua Jones and Charles L. Isbell. Story Similarity Measures for Drama Management with TTD-MDP. In *International Conference on Autonomous Agents and Multi-agent Systems*, pages 77–84, 2014. URL: `http://dl.acm.org/citation.cfm?id=2615747`.

14 Elektra Kypridemou and Loizos Michael. Narrative Similarity as Common Summary. In *Workshop on Computational Models of Narrative*, pages 129–146, 2013.

15 Michael Lebowitz. Story-telling as planning and learning. *Poetics*, 14(6):483–502, 1985. `doi:10.1016/0304-422X(85)90015-4`.

16 Vladimir I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.

17 James R. Meehan. TALE-SPIN: An Interactive Program that Writes Stories. In *International Joint Conference on Artificial Intelligence*, volume 77, pages 91–98, 1977. URL: `http://www.ijcai.org/PastProceedings/IJCAI-77-VOL1/PDF/013.pdf`.

18 Chris Mellish and Roger Evans. Natural language generation from plans. *Computational Linguistics*, 15(4):233–249, 1989.

19 David S. Miall. Experiencing narrative worlds: On the psychological activities of reading. *Journal of Pragmatics*, 32(3):377–382, Feb 2000. `doi:10.1016/S0378-2166(99)00017-X`.

20 Karen L. Myers. Metatheoretic Plan Summarization and Comparison. *Artificial Intelligence*, pages 182–191, 2002.

21 Karen L. Myers. Temporal Summarization of Plans. In *International Conference on Automated Planning and Scheduling*, 2007.

22 Saul B. Needleman and Christian D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453, 1970. `doi:10.1016/0022-2836(70)90057-4`.

23 Tuan Anh Nguyen, Minh Do, Alfonso Emilio Gerevini, Ivan Serina, Biplav Srivastava, and Subbarao Kambhampati. Generating Diverse Plans to Handle Unknown and Partially

Known User Preferences. *Artificial Intelligence*, 190:1–31, 2012. doi:10.1016/j.artint.2012.05.005.

24 Julie Porteous, Fred Charles, and Marc Cavazza. NetworkING : Using Character Relationships for Interactive Narrative Generation. In *International Conference on Autonomous Agents and Multi-agent Systems*, pages 595–602, 2013.

25 Rafael Pérez ý Pérez and Mike Sharples. MEXICA: A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence*, 13(2):119–139, 2001. doi:10.1080/09528130118867.

26 Mark O. Riedl and R. Michael Young. Narrative Planning : Balancing Plot and Character. *Journal of Artificial Intelligence Research*, 39:217–267, 2010.

27 David L. Roberts, Andrew S. Cantino, and Charles L. Isbell. Player Autonomy versus Designer Intent: A Case Study of Interactive Tour Guides. *Proceedings of the Third Artificial Intelligence and Interactive Digital Entertainment Conference*, pages 95–97, 2007. URL: http://www.aaai.org/Papers/AIIDE/2007/AIIDE07-020.pdf.

28 Mark Roberts, Adele E. Howe, and Indrajit Ray. Evaluating Diversity In Classical Planning. In *ICAPS*, pages 253–261, 2014.

29 Jonathan P. Rowe, Scott W. McQuiggan, Jennifer L. Robison, Derrick R. Marcey, and James C. Lester. STORYEVAL: An Empirical Evaluation Framework for Narrative Generation. In *AAAI Spring Symposium: Intelligent Narrative Technologies*, pages 103–110, 2009.

30 Biplav Srivastava, T. A. Nguyen, and A. Gerevini. Domain Independent Approaches for Finding Diverse Plans. In *International Joint Conference on Artificial Intelligence*, pages 2016–2022, 2007. URL: http://www.aaai.org/Papers/IJCAI/2007/IJCAI07-325.pdf.

31 Tom Trabasso and Linda L. Sperry. Causal Relatedness and Importance of Story Events. *Journal of Memory and Language*, 24:595–611, 1985. doi:10.1016/0749-596X(85)90048-8.

32 Stephen G. Ware and R. Michael Young. Glaive: A State-Space Narrative Planner Supporting Intentionality and Conflict. In *International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2014. URL: http://stephengware.com/publications/ware2014glaive.pdf.

33 Stephen G. Ware, R. Michael Young, Brent Harrison, and David L. Roberts. A Computational Model of Plan-based Narrative Conflict at the Fabula Level. *IEEE Transactions on Computational Intelligence and AI in Games*, 6(3):271–288, 2014. doi:10.1109/TCIAIG.2013.2277051.

34 R. Michael Young. Notes on the Use of Plan Structures in the Creation of Interactive Plot. In *AAAI Fall Symposium on Narrative Intelligence*, pages 164–167, 1999. URL: http://www.aaai.org/Papers/Symposia/Fall/1999/FS-99-01/FS99-01-028.pdf.

35 R. Michael Young. Using Grice's Maxim of Quantity to Select the Content of Plan Descriptions. *Artificial Intelligence*, 115(2):215–256, 1999. doi:10.1016/S0004-3702(99)00082-X.