Accepted for COSIT Workshop on
Identifying Objects, Processes and Events in Spatio-Temporally Distributed Data
September 12th 2011, Belfast, Maine. http://nav.spatial.maine.edu/cosit/

Boundary Provenance Relations

John Stell

School of Computing, University of Leeds, U.K.
j.g.stell@leeds.ac.uk

Abstract. Spatially nested objects splitting and merging have been studied formally as relations between nodes of adjacency trees representing the start and end states of a process. This gives provenance for objects at the end – specifying those objects at the start which were involved in the formation of those at the end. This paper shows that using relations between edges as well as nodes of the trees gives a more informative model. The appropriate kind of relation between trees is identified, and sequences of atomic changes (merging, splitting, creation and destruction as well as no change) are shown to be sufficient to generate all possible relations of this particular kind between adjacency trees.

Keywords: spatio-temporal change, provenance, adjacency tree

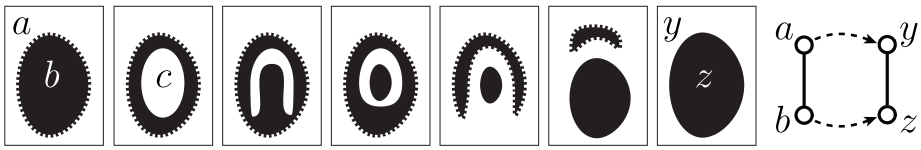
1 Introduction

In [1] provenance is described as 'information about the origin, ownership, influences upon, or other historical or contextual information about an object'. Objects can be as diverse as intermediate results in a numerical calculation, facts in a database, or, going beyond the context of [1], spatially extended entities such as crowds of people, wildfires, or bodies of water. Provenance of spatial objects is significant in the study of spatio-temporal processes. In studying coastal erosion and deposition we may need to determine how the sand dunes present at the end of a period were formed through the combination and separation of those dunes identified at the start of the period. Analogous questions can be asked about crowds of people, wildfires, etc. In querying the provenance of a specific object in one of these contexts an adequate answer must provide as a minimum the set of those objects present initially that contributed to its formation.

In terms of a formal model, the association of a set of initial objects to each final object is equivalent to providing a relation between the initial and final sets of objects. This has already been investigated in [2, 3] where the spatial arrangement of nested objects is modelled as an adjacency tree, and where events of merging, splitting, creation and destruction of objects compose to generate relations between the nodes of two trees. In the present paper the approach of [3] is extended by modelling provenance between spatial objects as a relation between trees which takes account of the edges of the tree as well as the nodes. The significance of this extension is that it allows more information about provenance to be recorded. The following two scenarios show that two relations between given trees can be the same on the nodes but differ in their action on edges.

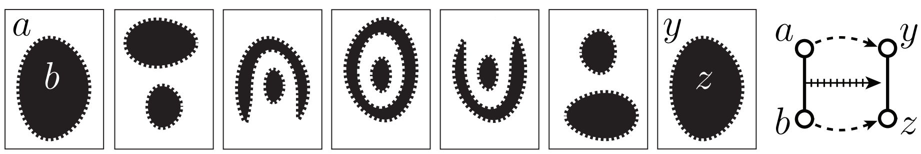
In both scenarios we start with two objects: a (white) and b (black). The boundary between these objects is shown toothed in the diagrams and this decoration is maintained when this boundary splits or merges.

Scenario 1



In the first scenario a new object, c, appears within b and splits b into two parts. In the process c merges with a to form the object labelled y in the rightmost snapshot, and one of the two parts of b is deleted leaving only the other part labelled z. The provenance of y is a and that of z is b, but the boundary between y and z is not derived from the boundary between a and b.

Scenario 2



In the second scenario, b splits into two parts one of which encircles the other, simultaneously capturing part of a. The two pieces of b merge to form z allowing the two parts of a also to merge and form y. As in the first case, the provenance of y is a and that of z is b, but additionally the boundary between y and z is derived from that between a and b.

An important distinction between the two scenarios is that in the first z comprises only part of b. The fact that the two edges are not related tells us that this must be the case. However, the converse of this does not hold. It is possible to construct a scenario in which the edges are related without z comprising all of b. The difference in the way the boundary behaves is apparent in the relations between trees given at the right of the diagrams above. In these diagrams related nodes in the tree are shown by dashed lines and related edges thus: .

2 Tree Relations

We need to establish some basic concepts, starting with graphs which we need to be undirected and without loops or multiple edges.

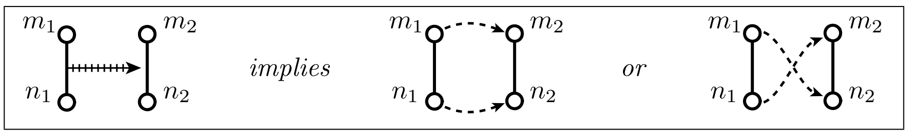
Definition 1 *A* ***graph****, G, is a pair (N, E) where N is a set and E is a set of subsets of N each having exactly two elements. The elements of N are called the* ***nodes*** *of G and those of E are called the* ***edges****.*

Definition 2 *A* ***tree****, T, is a graph (N, E) such that given any nodes $m, n \in N$ there is a unique sequence of distinct nodes $n_0, n_1, \ldots, n_k$ where $m = n_0$, $n = n_k$, and $\{n_{i-1}, n_i\} \in E$ for $i = 1, \ldots, k$. The number k is called the* ***distance*** *between m and n, and is denoted $d(m, n)$.*

Definition 3 *Let $T_1 = (N_1, E_1)$ and $T_2 = (N_2, E_2)$ be trees. A* ***tree relation*** *$R : T_1 \to T_2$ is a pair (R_N, R_E) where $R_N \subseteq N_1 \times N_2$ and $R_E \subseteq E_1 \times E_2$ are relations between the sets of nodes and edges respectively subject to the following* ***compatibility condition*** *for all edges $\{m_1, n_1\} \in E_1$ and $\{m_2, n_2\} \in E_2$.*

$$\{m_1, n_1\}\ R_E\ \{m_2, n_2\} \quad \textit{implies} \quad \begin{cases} (m_1\ R_N\ m_2 \textit{ and } n_1\ R_N\ n_2) \textit{ or} \\ (m_1\ R_N\ n_2 \textit{ and } n_1\ R_N\ m_2). \end{cases}$$

We can express this condition diagrammatically as follows.



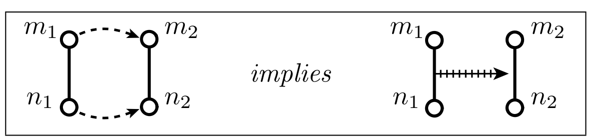
The **composition** of tree relations $R : T_1 \to T_2$ and $S : T_2 \to T_3$ is defined by composing the node and edge relations separately, $R\,;S = (R_N\,; S_N, R_E\,; S_E)$. The **converse** is given by $R^{-1} = ({R_N}^{-1}, {R_E}^{-1})$.

Definition 4 *A tree relation $R : T_1 \to T_2$ is* ***bipartite*** *if for all $m_1, n_1 \in N_1$ and for all $m_2, n_2 \in N_2$*

$$m_1\ R_N\ m_2 \textit{ and } n_1\ R_N\ n_2 \textit{ implies } d(m_1, n_1) \equiv d(m_2, n_2) \pmod 2.$$

Bipartite tree relations model relations between nested regions where there are just two types of region (foreground and background) and where any two related regions are both of the same type.

Definition 5 *A tree relation $R : T_1 \to T_2$ is* ***full*** *if for all edges $e_1 = \{m_1, n_1\} \in E_1$ and $e_2 = \{m_2, n_2\} \in E_2$, the conditions $m_1\ R_N\ m_2$ and $n_1\ R_N\ n_2$ imply $e_1\ R_E\ e_2$. Diagrammatically this appears as follows.*



For any tree T the **identity tree relation** provides an example of a full tree relation. This has its node and edge components as the identity relations on the nodes and edges of T respectively. More generally a tree relation R is an **isomorphism** if R_N is a bijective function on the nodes of the trees which induces a bijection of edges, and if also R is full.

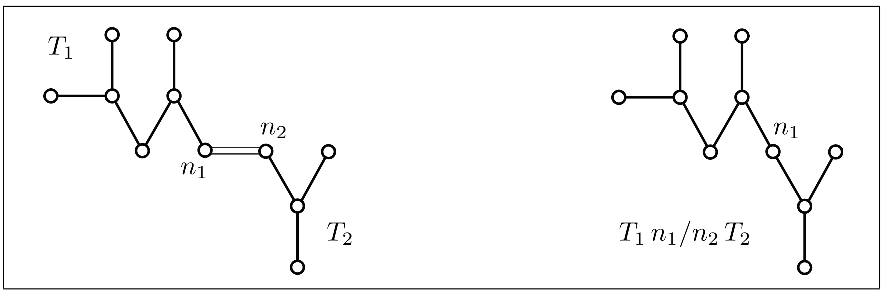
3 Gluing Tree Relations Together

In order to analyse tree relations in general we need to be able to describe how complex relations can be obtained by combining simpler components. We start with gluing trees together and then extend this to relations between trees.

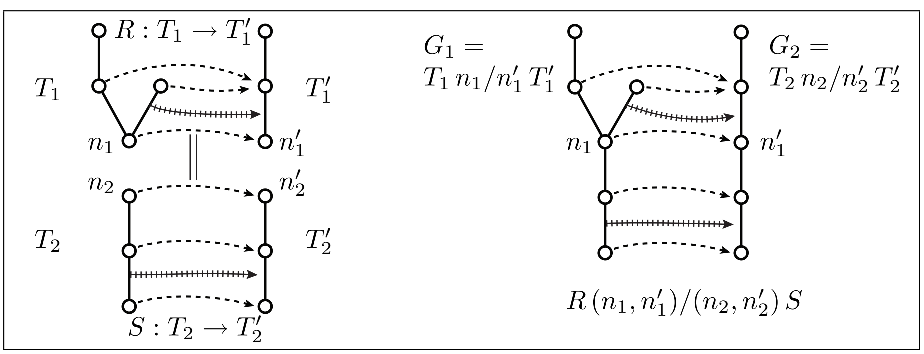
Suppose T_1 and T_2 are trees containing nodes n_1 and n_2 respectively and that $(N_1 - \{n_1\}) \cap (N_2 - \{n_2\}) = \varnothing$. We define the tree $T_1\, n_1/n_2\, T_2$ to have nodes N and edges E where $N = N_1 \cup N_2 - \{n_2\}$ and

$$E = E_1 \cup (E_2 \cap \mathcal{P}(N)) \cup \{\, \{n_1, m\} \mid \{n_2, m\} \in E_2\}.$$

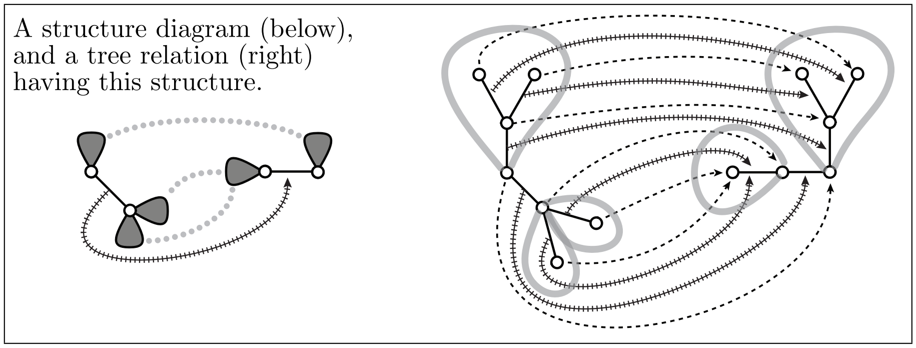
This tree is called the result of gluing T_1 to T_2 at n_1 and n_2, an example is provided below.



Suppose now we have tree relations $R : T_1 \to T_1'$ and $S : T_2 \to T_2'$ and that there are nodes $n_1 \in T_1$, $n_1' \in T_1'$, $n_2 \in T_2$, and $n_2' \in T_2'$ such that $n_1\ R\ n_1'$ and $n_2\ S\ n_2'$. Suppose also that the nodes of T_1 and T_1' are sufficiently disjoint that we may form $G_1 = T_1\, n_1/n_1'\, T_1'$, and similarly for T_2 and T_2' permitting the gluing $G_2 = T_2\, n_2/n_2'\, T_2'$. The tree relation $R\,(n_1, n_1')/(n_2, n_2')\,S : G_1 \to G_2$ is constructed as illustrated below; the formal definition is omitted here due to limitations of space.

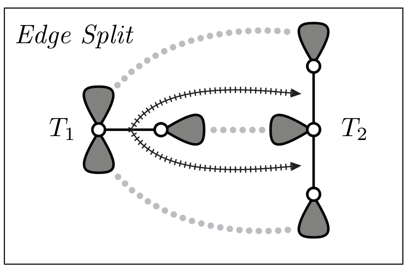


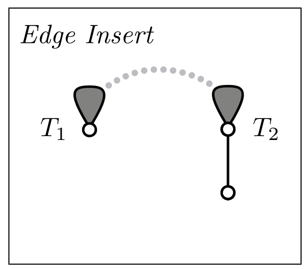
Tree relations can often be described as the result of gluing together a number of simpler relations. Often these simpler relations are isomorphisms and this makes it convenient to use an extension of the structure diagram notation introduced in [3] and illustrated in the next diagram. In this notation a subtree with a distinguished node is shown as an open circle for the node and a dark grey shaded loop for the rest of the subtree. An isomorphism between two subtrees is shown by a light grey dotted line. Any relationships between explicitly indicated edges are indicated by the style of line used in earlier relations.



4 Boundary Provenance Relations

Definition 6 *An* ***atomic change*** *is a tree relation of one of five forms. It is either an isomorphism or an edge split or an edge insert (see below) or the converse of one of these last two (an edge merge or an edge delete respectively). Note that these atomic changes relate both edges and nodes.*





A ***boundary provenance relation*** *is any tree relation that can be obtained by composing a sequence of atomic changes.*

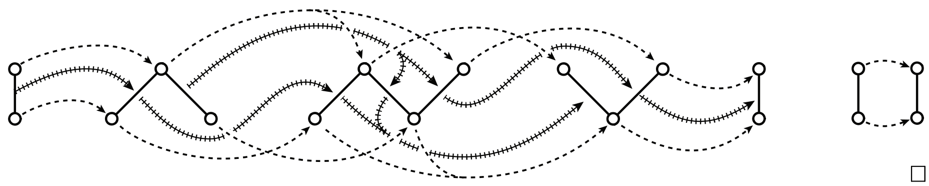
In [3] Stell and Worboys showed that five kinds of atomic change affecting nodes alone were sufficient to generate all possible relations between nodes that modelled a division into nested background and foreground regions. The following result generalizes this to take account of edges as well as nodes. It shows that any bipartite node relation can be extended to a boundary provenance relation by relating the edges in any way that satisfies the compatibility condition of Definition 3.

Theorem 1 *For any trees, T_1 and T_2, a tree relation $R : T_1 \to T_2$ is a boundary provenance relation if and only if it is bipartite.*

To prove this we first need two lemmas, the first of which expresses the relation of Scenario 1 in the introduction as a sequence of atomic changes.

Lemma 2 *Let T be the tree consisting of a single edge, and let $R : T \to T$ be the tree relation which is the identity relation on nodes and the empty relation on edges. Then R is a boundary provenance relation.*

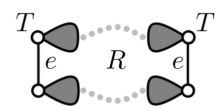
Proof. The relation can be written as a sequence of four atomic changes: an edge insert, an edge split, an edge merge, and an edge delete as shown below.



□

Lemma 3 *Let $T = (N, E)$ be a tree and e an edge of T. The tree relation R where R_N is the identity relation on N and where $f, g \in E$ are related by R_E if and only if $f = g$ and $f \neq e$, is a boundary provenance relation.*

Proof. The relation R, which has the structure shown in the adjacent diagram, can be obtained by gluing together three boundary provenance relations. Two are isomorphisms and the third is the relation from Lemma 2. □



The methods used in the proof of the main result in [3] apply only to node relations, but it is possible to extend them to tree relations which are full and thus to establish the following. Space precludes giving further details here.

Theorem 4 *Full bipartite tree relations are boundary provenance relations.* □

This allows us to establish Theorem 1 since an arbitrary bipartite tree relation can be written a composite of a full bipartite tree relation and a number of relations of the form appearing in Lemma 3.

5 Conclusions and Further Work

This paper has extended the theories of topological change for nested regions by relating boundaries as well as interiors of dynamic regions. A boundary has no physical counterpart in cases such as a crowd of people, but the two scenarios in the introduction show that the behaviour of the corresponding edges in the adjacency trees can lead to deductions having practical significance. Unrelated edges can imply that two related crowds do not contain the same people.

Further work will consider regions that have more complex spatial relationships than just nesting. This will require graphs that are more general than trees. Another issue is to investigate how a similar analysis can be developed for dynamic regions in various kinds of discrete space.

References

1. Cheney, J., Ahmed, A., Acar, U.A.: Provenance as dependency analysis. In: Arenas, M., Schwartzbach, M. (eds.) DBPL 2007. LNCS, vol. 4797, pp. 138–152 (2007)
2. Jiang, J., Worboys, M.: Event-based topology for dynamic planar areal objects. International Journal of Geographical Information Science 23(1), 33–60 (2009)
3. Stell, J., Worboys, M.: Relations between adjacency trees. Theoretical Computer Science (in press), (Preprint: `www.comp.leeds.ac.uk/jgs/AdjacencyTrees.pdf`)