

Article

An Efficient Graph-based Method for Long-term Land-use Change Statistics

Yipeng Zhang ^{1,2}, Yunbing Gao ^{1,2,3,*}, Bingbo Gao ^{1,2}, Yuchun Pan ^{1,2,4,5} and Mingyang Yan ^{1,2,6}

Received: 13 September 2015; Accepted: 10 December 2015; Published: 29 December 2015

Academic Editors: Yichun Xie and Xinyue Ye

¹ Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; zypcastle@hotmail.com (Y.Z.); gaobb@nercita.org.cn (B.G.); panyc@nercita.org.cn (Y.P.); ymygis@outlook.com (M.Y.)

² National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

³ College of Information and Electrical Engineering, China Agriculture University, Beijing 100083, China

⁴ Key Laboratory of Agri-informatics, Ministry of Agriculture, Beijing 100097, China

⁵ Beijing Engineering Research Center of Agricultural Internet of Things, Beijing 100097, China

⁶ College of Geoscience and Surveying Engineering, China University of Mining & Technology, Beijing 100083, China

* Correspondence: gaoyb@nercita.org.cn; Tel.: +86-10-5150-3410

Abstract: Statistical analysis of land-use change plays an important role in sustainable land management and has received increasing attention from scholars and administrative departments. However, the statistical process involving spatial overlay analysis remains difficult and needs improvement to deal with mass land-use data. In this paper, we introduce a spatio-temporal flow network model to reveal the hidden relational information among spatio-temporal entities. Based on graph theory, the constant condition of saturated multi-commodity flow is derived. A new method based on a network partition technique of spatio-temporal flow network are proposed to optimize the transition statistical process. The effectiveness and efficiency of the proposed method is verified through experiments using land-use data in Hunan from 2009 to 2014. In the comparison among three different land-use change statistical methods, the proposed method exhibits remarkable superiority in efficiency.

Keywords: land management; land-use change; spatio-temporal model; graph theory; network flow

1. Introduction

Land use change is a key component of the Earth changes being found in all of the continents and having impacts on the Earth resource and services used by humankind [1–5]. Land use changes are resulting in the disturbance of soil, vegetation, and water resources and services, and at the same time determining and modifying the fate of Earth life system, chemical, and hydrological cycles [6–8]. Certain analytic indices [9,10] (e.g., relative change rate, dynamic degree, location quotient, etc.) and dynamic models (e.g., Markov forecasting [11,12], spatial autocorrelation [13], etc.) have been applied widely to studies on land-use change patterns.

Type transition statistics are an important kind of spatio-temporal statistics that provide databases for the analytic indices and dynamic models. It concerns the measurable areal relationship characterizing changes between features and answers queries such as “how much farmland has changed to residential land over a period”. Transition statistical process (TSP), which involves complicated relational and topological operations, remains a challenge in land data statistical applications. Although TSP can be obtained through overlay analysis (Figure 1), the process requires substantial time and resources when handling massive practical land-use data [14]. Based on the

query optimizations for spatio-temporal database (STDB), the computation of TSP can be reduced to a certain extent, but still unable to satisfy the practical needs. The lack of an accurate and efficient TSP method for mass data is the bottleneck problem of land-use change analysis and data mining.

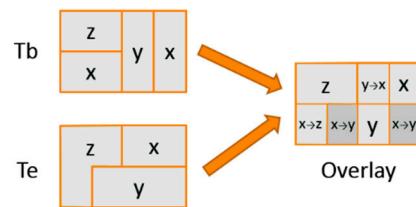


Figure 1. Transition Flow Calculation by overlay analysis.

Graph representation can be used to facilitate change analysis in vector spatial data through the discovery of long temporal patterns from short-term relations. In this paper, we firstly analyze land use change in graph approach and establish restrict mathematical relationship between graph and spatio-temporal change. Then, we propose and illustrate a new graph-based TSP method based on the graph analysis. Finally, we compare this new TSP method with the basic overlay analysis method and a query-optimized method using practical land-use data.

2. Related Works

2.1. STDB for Land Management

Spatio-temporal databases have been applied to land use data for effective storage and management of the historical data. It can be used to reconstruct the past states, trace the changes, and analyze the statistics or trends of land use state. The basic spatio-temporal models that are suitable for effective land-use data management include a base state with an amendment model, object-oriented model, and event-based model [15–17]. One typical spatio-temporal model that combines the characteristics of these basic model types is proposed by Teng *et al.* [18] (Figure 2). The model stores geographical entities as data object in a present database (Present DB) and a history database (History DB). Every object has semantic attributes, geometric shape (Shape), and lifespan (BeginTime, EndTime). When geographical entities change, the new objects are stored in the Present DB, and the old objects are stored in the History DB. The model also records the inheritance relationship between old objects and new objects in the Event Table, by which local spatio-temporal changes can be traced and the evolution of inheritances can be understood.

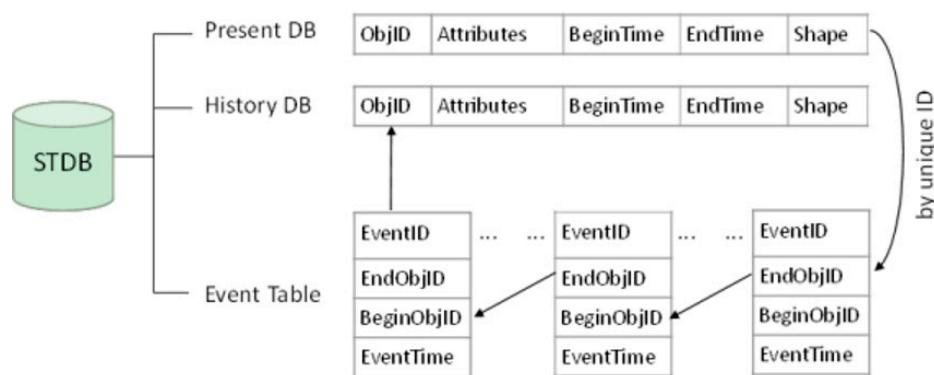


Figure 2. STDB based on a typical spatio-temporal model (refer to Teng, 2005 [18]).

Topological information plays an important role in land use databases. Point-set methods are generally used in vector GIS [19]. Topology is generally used in maintaining data quality. An

important function of topological information is storing a spatial partition since, without topological information, it is difficult to control whether a given set of regions forms a partition or not [20]. Spatio-temporal topology represents inheritance relations in land use changes which is involved in some queries. The spatio-temporal topology is used for land use database updating in China [21].

2.2. Statistical Process in STDB

STDB organized with Teng's model, or its analogues, stores spatio-temporal topology for the changes and enables various forms of query when spatio-temporal query languages [22] are integrated. The STDB can be used in GIS to retrieve entities in snapshots and optimize static statistical processes. However, GIS cannot facilitate query or analysis computations for entities that are beyond the representation capabilities of its data models [23], and the optimization of transition statistical processes is strictly limited unless the data model reveals further connections between simple topological relations.

Ruan *et al.* introduced a land-use temporal statistics model based on transaction mechanics in 2011 [24]. In the concept of "transaction", complex changes are divided into "change basic events", then a "change index table" is built to advance the statistical query process. However, Ruan's statistical method specified to land transition cannot be applied to long term changes where multiple complex changes occurred at the same place during a period.

3. Analysis of Land Use Change in Graph Theory Approach

Spatio-temporal graphs (STG) show how entities are related in space and time. This section focuses on analysis of the relationship between long-interval land use transition and spatio-temporal topology of short-term local events in a STG. We extend STG into a flow network where the change area of local events is mapped as edge capacity and the long-term type transition is modeled as a special kind of network flow that is mapped from spatio-temporal transitions. The properties of this network flow ensures that it can keep constant in reducible network, which is defined to be the basis of a fast TSP method.

3.1. Characterization of Spatio-Temporal Flow Network

Wilcox classified the spatio-temporal model into four types based on the continuity in space and time and discusses the methods of modeling space-discrete, time-continuous, spatio-temporal data in a graph-based approach [25]. Based on his research, Yin *et al.* presented a space-discrete time-discrete graph model for cadastral changes [26]. Spéry *et al.* first investigated the application of directed acyclic graph (DAG) to the distributed cadastral management [27]. Recently, Rodriguez *et al.* researched the formalization to extend spatio-temporal graph expressive power and the mapping from spatio-temporal graph to STDB [28].

Land-use change can be described by Yin's STG model, since the change is space-discrete and time-discrete (see Figure 3). Vertices represent spatio-temporal entities whose temporal relationship forms a partial order. A directed edge connects a pair of spatio-temporal entities that intersect in space and meet in time, and the intersection area of the two entities is recorded for the edge. An STG is acyclic in that time is invertible.

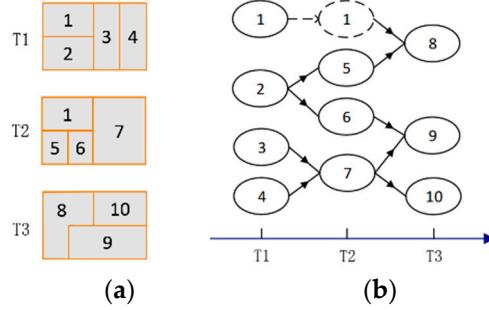


Figure 3. (a) Land-use snapshots at three different timestamps; (b) corresponding spatio-temporal graph (refer to Yin, 2003 [23]).

If we specify two timestamps, namely, T_b and T_e and a geographic region R , the 5-tuple (V, E, S, T, c) can be used to represent a spatio-temporal flow network (STFN). V and E , respectively, denote the vertex set and edge set in the STG that is associated with the change between T_1 and T_2 within R . Source set S and sink set T respectively denote the spatial temporal entities existing at T_b and T_e within R . We define capacity function c as the mapping from edge (u, v) to the intersection area between entities u and v . In Figure 4, the land-use change process described in Figure 3 is represented by STFN.

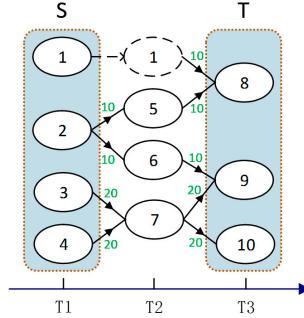


Figure 4. Spatio-temporal flow network (note: The capacities are labeled on each edge).

Let the function g gets the geometric shape of a spatial-temporal entity, and the function $Area$ calculates the area of a geometric shape. Then, c can be formulated below:

$$c(u, v) = Area(g(u) \cap g(v)), \quad (u, v) \in E \quad (1)$$

The birth and death of any specified entity is expected to be record in edge set E . Therefore, it can be deduced that:

$$g(w) = \bigcup_{u \in V, (u,w) \in E} g(u) \cap g(w), \quad \forall w \in V \quad (2)$$

$$g(w) = \bigcup_{v \in V, (w,v) \in E} g(u) \cap g(w), \quad \forall w \in V \quad (3)$$

If we define:

$$c^-(w) = \sum_{u \in V, (u,w) \in E} c(u, w), \quad \forall w \in V \quad (4)$$

$$c^+(w) = \sum_{v \in V, (w,v) \in E} c(w, v), \quad \forall w \in V \quad (5)$$

For every internal vertex (vertex that is not a source or sink) w , Equations (2) and (3) implies that:

$$c^-(w) = \text{Area}(g(w)) = c^+(w), \quad \forall w \in V - S - T \quad (6)$$

Equation (6) means that the capacity of inflow edges and outflow edges for any internal vertex is balanced.

3.2. Modelling Long-Term Transition as Multi-Commodity Flow

A multi-commodity flow problem is a kind of network flow problem where multiple commodities are routed between different sources and sinks [29]. In STFN, we use different commodities to indicate the distinction between the topological relations of different connected $S-T$ pairs (the term “A connect to B” is used for simplicity in a directed graph and means exactly that A can reach B or B can reach A). Let $f_{ij}(u, v)$ be the intersection area of entities s_i, t_j, u , and v , specifically:

$$f_{ij}(u, v) = \begin{cases} \text{Area}(g(s_i) \cap g(t_j) \cap g(u) \cap g(v)), & (u, v) \in E, \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

Since the geometric shapes at any given timestamp are mutually exclusive and collectively exhaustive, for any source-sink pair, we can deduce that:

$$\sum_{(u,w) \in E} f_{ij}(u, w) = \text{Area}(g(w) \cap g(s_i) \cap g(t_j)) = \sum_{(w,v) \in E} f_{ij}(w, v), \quad \forall w \in V - S - T \quad (8)$$

and:

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij}(u, v) = \text{Area}(g(u) \cap g(v)) = c(u, v), \quad \forall (u, v) \in E. \quad (9)$$

Let F_{ij} be the transition area from entity i to entity j , it follows that:

$$F_{ij} = \text{Area}(g(s_i) \cap g(t_j)) = \sum_{(u,v) \in E} f_{ij}(u, v) \quad (10)$$

Hence, determining multi-commodity flow values in STFN is equivalent to calculating transition areas.

3.3. Constant Multi-Commodity Flow Condition

Based on graph theory [30], we prove that the multi-commodity flow is constant under certain conditions.

Preliminary. A transportation network with multiple sources and sinks is a 5-tuple (V, E, S, T, c) where m sources $s_1, s_2, \dots, s_m \in S$ and n sinks $t_1, t_2, \dots, t_n \in T$ belong to the vertex set V , and $c : E \rightarrow \mathbb{R}_+$ is the capacity function for edges in the directed edge set E .

An $S-T$ path P is a directed path from one source to one sink. Denote the complete set of $S-T$ paths by \mathcal{P} the set of $S-T$ paths from s_i to t_j by \mathcal{P}_{ij} , and the set of $S-T$ paths containing edge e by \mathcal{P}_e . i.e. $\mathcal{P}_{ij} \{ P \in \mathcal{P} | s_i \in P, t_j \in P \}$, $\mathcal{P}_e \{ P \in \mathcal{P} | e \in P \}$.

A (path-form) multiflow $f \in \mathbb{R}_+^{\mathcal{P}}$ is an assignment for flows along $S-T$ paths. A feasible multiflow is nonnegative and subject to the capacity constraint:

$$\sum_{P \in \mathcal{P}_e} f_p \leq c(e), \quad \forall e \in E \quad (11)$$

We suppose that the commodities routed between different connected $S-T$ pairs are different, and use F to denote the multi-commodity flow with respect to the $S-T$ pairs,

$$F_{ij} \sum_{P \in \mathcal{P}_{ij}} f_p$$

is the flow value for $S-T$ pair (s_i, t_j) .

Unlike residual network for single-commodity flow, the residual network for a multiflow f does not contain a directed edge that is opposite to the edges in the original network. It contains edges with capacity $c_f(e) = c(e) - \sum_{P \in \mathcal{P}_e} f_p$ if $c_f(e) > 0$.

Let $R(v)$ be the set of vertices that can reach v or v can reach (include vertex v); i.e., $R(v) = \{u \in V | (\exists P) u \in P, v \in P\}$. Obviously, $u \in R(v) \Leftrightarrow v \in R(u)$. The network presented in this paper is assumed to have no directed paths connecting two different sources or two different sinks; that is, $|R(s) \cap S| = 1$ for each $s \in S$ and $|R(t) \cap T| = 1$ for each $t \in T$. We also assume that the network is $S-T$ connected, which implies $|R(v) \cap S| > 0$ and $|R(v) \cap T| > 0$ for each $v \in V$.

Definition 1. A vertex v is called *mixed vertex* if more than one source can reach v and v can reach more than one sink, i.e.:

$$|R(v) \cap S| > 1 \text{ and } |R(v) \cap T| > 1 \quad (12)$$

Definition 2. A *one-way cut* is a cut $[X, \bar{X}]$ separating S from T where no directed edges have a tail in \bar{X} and a head in X .

Definition 3. An *$S-T$ exclusive cut* $\delta(X)$ is a one-way cut that only contains edges which can be reached from exactly one source and can reach exactly one sink, i.e.:

$$|R(u) \cap S| = |R(v) \cap T| = 1, \quad \forall (u, v) \in \delta(X) \quad (13)$$

Lemma 1. A directed acyclic network without mixed vertices has an $S-T$ exclusive cut.

Proof. Let X_i be the set of s_i and all non-terminal vertices that only s_i among S can reach; i.e., $X_i = \{u | R(u) \cap S = \{s_i\}\}$, For every edge (u, v) in cut $[X_i, \bar{X}_i]$, there are only two possible circumstances about the its head v :

- $v \in T$, then $|R(v) \cap T| = 1$
- $v \in V - T$, then v must be reachable from other sources than s_i (otherwise it will be included in X_i), so it must connect only one sink (or it is a mixed vertex). $|R(v) \cap T| = 1$

Since X_i for each sources do not intersect, the number of reachable terminals for each vertex in the union set $X \cup_{i=1}^m X_i$ is not changed. $\forall (u, v) \in [X, \bar{X}], |R(u) \cap S| = |R(v) \cap T| = 1$. Obviously, $[X, \bar{X}]$ is a one-way cut, so $[X, \bar{X}]$ is an $S-T$ exclusive cut. Q.E.D.

Remark 1. The above proof also implies a method to find an $S-T$ exclusive cut, which is to be used in the optimization algorithm.

Definition 4. A multiflow f on a network is *saturated* if the total flow on each edge is maximized. i.e.:

$$\sum_{P \in \mathcal{P}_e} f_p = c(e), \quad \forall e \in E \quad (14)$$

Definition 5. A network is *capacity-balanced* if for all internal vertices, the capacity sum of inflow edges equals the capacity sum of outflow edges. Let $c^-(v) \sum_{u \in V} c(u, v), c^+(v) \sum_{u \in V} c(v, u)$. Hereby, the capacity-balanced constraint is represented as:

$$c^-(v) = c^+(v), \quad \forall v \in V - S - T \quad (15)$$

Remark 2. This section intends to determine on what circumstances the saturated multiflow on a directed acyclic capacity-balanced network (DACP network) has constant multi-commodity flow with respect to $S-T$ pairs. Although the result could be achieved by solving linear programming

directly [31], its variable size is $O(|E| \times |S| \times |T|)$. Only small instances of multiflow problem can be solved as linear programming in practice. Hence, we need to find a simpler approach by studying the properties of DACB network.

Fact 1. In a DACB network, the residual network for any given feasible flow is still a DACB network.

Fact 2. In a DACB network, if P is the only $S-T$ path that includes edge (u, v) , then (u, v) has the smallest capacity among edges along P .

Proposition 1. There always exists a saturated multiflow in a DACB network.

Proof. We construct a multiflow by sequentially adding a maximum feasible flow on each $S-T$ path: When the first k $S-T$ paths is maximized with flows, denote the residual network by N^k . Fact 1 implies that N^k is still a DACB network for $k \in [1, |\mathcal{P}|-1]$. In the construction process, if the n th $S-T$ path P_m is the last routable path for an edge e_n , e_n must be one of the edges with the smallest capacity in N^{m-1} because of Fact 2, and $c(e_n)$ is saturated when maximizing the flow along P_m . It ensures that each edge e_n will be saturated after all flows on routable paths P ($e_n \in P$) are added. Thus, the multiflow constructed in this way is saturated.

Lemma 2. In a DACB network, if an $S-T$ exclusive cut $\delta(X)$ exists, then F^* , the saturated multi-commodity flow with respect to $S-T$ pairs, is constant.

Proof. Since $\delta(X)$ is a one-way cut, $\pi_p \{\mathcal{P}_e | e \in \delta(X)\}$ is a partition of \mathcal{P} . Therefore:

$$F_{ij} = \sum_{P \in \mathcal{P}_{ij}} f_P = \sum_{\mathcal{P}_e \in \pi_p} \text{sum}(f, \mathcal{P}_e \cap \mathcal{P}_{ij}) \quad (16)$$

where $\text{sum}(f, Q) \sum_{P \in Q} f_Q$.

Note that for any e in the $S-T$ exclusive cut $\delta(X)$:

$$\mathcal{P}_e \cap \mathcal{P}_{ij} = \begin{cases} \mathcal{P}_e, & R(u) \cap S = \{s_i\} \text{ and } R(v) \cap T = \{t_j\}, \\ \emptyset, & \text{otherwise} \end{cases} \quad (17)$$

and for saturated multiflow f^* , $\sum_{P \in \mathcal{P}_e} f_P^* = c(e)$. Then, all $\text{sum}(f^*, \mathcal{P}_e \cap \mathcal{P}_{ij})$ in Equation (6) are constants, and F_{ij}^* is also constant regardless of f^* . Q.E.D.

Lemma 3. In a DACB network, if a mixed vertex exists, then F^* , the saturated multi-commodity flow with respect to $S-T$ pairs, is not constant.

Proof. The $S-T$ path matrix [32] is denoted by $M_p [p_{ij}]_{|E| \times |\mathcal{P}|}$, where:

$$p_{ij} = \begin{cases} 1, & \text{if the } i\text{th edge lies in the } j\text{th ST path,} \\ 0, & \text{otherwise} \end{cases}$$

Then we can formulate saturated multiflow problem as standard linear programming:

$$M_p \cdot f = c \quad (18)$$

$$f \geq 0 \quad (19)$$

Now, f is the column vector of f_P whose entries are in the same order with $S-T$ paths, and c is the column vector of $c(e)$ whose entries are in the same order with edges. We need to determine whether each $S-T$ pair commodity flow F_{ij} is constant for all feasible solutions in this linear programming.

Suppose that v_m is the mixed vertex. Then it is easy to find two $S-T$ paths, P_{ac} and P_{bd} , which both contain v_m and route between different sources s_a, s_b and different sinks t_c, t_d , respectively. We can set any feasible positive flow concurrent f_{ac} and f_{bd} for these two paths. According to Fact 1 and Proposition 1, the residual network for $(f_{ac} + f_{bd})$ have saturated multiflow f' and, hence, $f^* = f_{ac} + f_{bd} + f'$ is a particular solution for the linear equations (18). While P_{ac} and P_{bd} meet at v_m , we can find another $S-T$ path P_{ad} , which share the same path with P_{ac} between s_a and v_m and share

the same path with P_{bd} between v_m and t_d . We can find the $S-T$ path P_{bc} , similarly. If constructing a vector $f^0 = (f_1^0, \dots, f_k^0, \dots, f_{|\mathcal{P}|}^0)^T$, where the real number $\lambda \neq 0$ and:

$$f_k^0 = \begin{cases} -\lambda, & \text{the } k\text{th path is } P_{ac} \text{ or } P_{bd}, \\ \lambda, & \text{the } k\text{th path is } P_{ad} \text{ or } P_{bc}, \\ 0, & \text{elsewise} \end{cases}$$

then, $M_p f^0 = 0$, which suggests that f^0 is a nontrivial solution of (18)'s corresponding homogeneous equations. So $f = f^* + f^0$ is also a solution of (19). Given that the components for P_{ac} and P_{bd} in f^* are positive, there exist solutions also feasible for the non-negative constraint (19) when $\lambda > 0$. As f_{ac}, f_{bd}, f_{ad} , and f_{bc} contribute to different F_{ij} , F cannot remain constant.

Theorem 1. (Constant Multi-commodity Flow Condition)

In a DACB network, the following statements are equivalent.

- The saturated multi-commodity flow F is constant
- No mixed vertices exist
- An $S-T$ exclusive cut exists

Proof. This theorem is directly derived from Lemmas 1–3.

3.4. Reducible or Unreducible

As is shown in Section 3.1, STFN is directed, acyclic, and capacity balanced. Equations (9) and (10) shows that flow value for every $S-T$ pair in an STFN saturated-commodity flow is the intersection area of corresponding parcels, which is used in TSP (see Section 1, Figure 1). An STFN is considered reducible if the saturated multi-commodity flow with respect to $S-T$ pairs is uniquely determined so that TSP can be reduced into simple operations of summing numbers.

Theorem 1 implies whether a network is reducible or not depends on the existence of mixed vertices. Let R_s and R_t respectively be the set of sources and sinks connecting to mixed vertices. Clearly, a sub-network without any connection between R_s and R_t contains no mixed vertices; hence, such sub-network is reducible; an $S-T$ exclusive cut can be found in a reducible network and, thus, be used to calculate the flow value F_{ij} (see Lemma 2). In contrast, the commodity flows F_{ij} that routes between R_s and R_t cannot be uniquely determined by finding an $S-T$ exclusive cut (a simple corollary of Theorem 1). Finding a network partition with the maximum reducibility network is the key to simplify TSP.

4. Description of the Graph-Based TSP Method

Based on the analysis in Section 3, we introduce the graph-based method for TSP in this section. We assume that the land use data is already stored in an STDB and queries of land use type transitions in terms of any two timestamps is required. We firstly construct an STFN and use it to optimize the calculation of TSP. The network partitioning (*PartitionSTFN*) is the key algorithm of this method.

In the graph construction algorithm (*BuildSTFN*), STFN is created as a 5-tuple (V, E, S, T, c) using a typical STDB (introduced in Section 2.2) with the materialization of change events. We then retrieve source set S , sink set T , and related event list by spatio-temporal queries, extend the vertex set V by simply iterating the event list (edge set E), and connect the vertices and edges to form an STG. No spatio-temporal entity information other than their identities is read in this graph-construction process. Retrieving edges and attaching them on STG forms an STFN. Hash techniques may be used to accelerate this process.

Section 3.4 shows that the spatio-temporal network can be divided into two sub-networks, namely, the reducible one and the unreducible one. In the reducible sub-network, there exists an $S-T$ exclusive cut, so the transition flow can be calculated directly. In the unreducible sub-network,

the transition flow is still calculated by the overlay between spatio-temporal entities corresponding to the sources and sinks of the network. Given that a few mixed vertices can be found in a STFN of practical data, in the function *PartitionSTFN*, we divide the global STFN into maximized reducible and minimal irreducible networks in order to calculate the transition statistics with maximum efficiency.

The pseudo code of the partitioning algorithm is presented. We use a list of triple $(s, t, c) Fr$ to store the flows for $S-T$ pairs in the reducible network. The queue Q is used for breadth first search (BFS), the basic search strategy in this algorithm. This algorithm turns networks into a triple (Rs, Rt, Fr) . For each vertex v , we use $v.Rs$, $v.Rt$ to denote the reachable terminals, and use $v.color$ to represent the vertex set X .

```

PartitionSTFN(N)
/* Label T-Reachability */
Q ← Ø;
for vertex  $t \in N.T$ 
     $t.Rt \leftarrow t;$ 
    Enqueue  $t$  to  $Q$ ;
end for
Begin backward BFS on  $N$  using  $Q$ ;
while BFS is not complete do
    Set  $v$  to the next vertex in BFS;
    for vertex  $w$  in  $v$ 's parents
         $w.Rt \leftarrow v.Rt \cup w.Rt;$ 
    end for
end while
/* Label S-Reachability and get decomposition results*/
Set  $v.color$  to white for all vertices in  $N.V$ ;
Q ← Ø;
for each vertex  $s \in N.S$ 
     $s.Rs \leftarrow s;$ 
    Enqueue  $s$  to  $Q$ ;
end for
Begin forward BFS on  $N$  using  $Q$ ;
while BFS is not complete do
    Set  $v$  to the next vertex in BFS;
    if  $|v.Rs| > 1$  and  $|v.Rt| > 1$  then /* mixed vertex */
         $Sm \leftarrow Sm \cup v.Rs;$ 
         $Tm \leftarrow Tm \cup v.Rt;$ 
         $v.color \leftarrow black;$ 
    end if
    for vertex  $w$  in  $v$ 's children
        if  $v.color = black$  then
             $w.color \leftarrow black;$ 
        else if  $|v.Rs| = 1$  and  $|w.Rt| = 1$  then
            /* find a potential flow in S-T exclusive cut*/
            Add  $(s, t, N.c(v,w))$  to  $Fr$ ;
             $w.color \leftarrow black;$ 
        end if
         $w.Rs \leftarrow v.Rs \cup w.Rs;$ 
    end for
end while
if /* exclude flows from irreducible networks */
Delete  $(s, t, c)$  rows in  $Fr$  where  $s \in Sm$  or  $t \in St$ ;
return  $Rs, Rt, Fr$ ;

```

e.g., (Figure 5) v_1, v_2, v_3 and v_4 are the sources and v_8, v_9 and v_{10} are the sinks. We observe that v_7 simultaneously connects sources v_3, v_4 and sinks v_9, v_{10} , so it is a mixed vertex which makes the lower sub-network irreducible. In the upper sub-network, edge (v_1, v_8) only connects source v_1 and sink v_8 , edge (v_2, v_5) only connects source v_2 and sink v_8 , and edge (v_2, v_6) only connects source v_2 and sink v_9 ; therefore, $\{(v_1, v_8), (v_2, v_5), (v_2, v_6)\}$ is an S-T exclusive cut, and the flows can be calculated: $Fr = [(v_1, v_8, 10), (v_2, v_8, 10), (v_2, v_9, 10)]$.

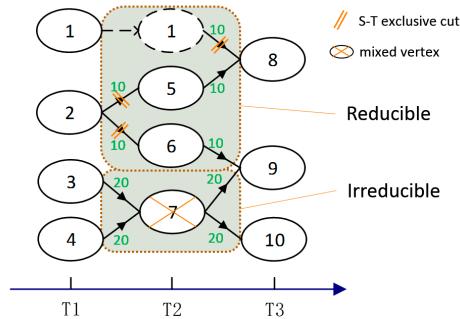


Figure 5. Network partitioning.

The complete procedures of proposed graph-based TSP is presented in Figure 6: we primarily construct a dynamic STFN N according to user's spatio-temporal query. Then, we partition the entire STFN and get the reducible flow list Fr , irreducible sources Sm , and irreducible sinks Tm . Therefore, transition flows on the reducible sub-network is achieved by simple grouping operations on Fr . At the same time, transition flows on the irreducible sub-network (between Sm and Tm) is calculated through the spatial overlay analysis. The final transition statistics result is achieved by summing up statistics from both sub-networks.

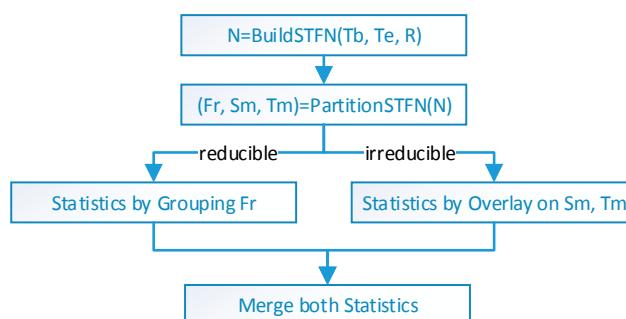


Figure 6. Graph-based TSP procedures.

5. Data Experiments

5.1. Sample Data

China has experienced noticeable land-use changes for years resulting from rapid economic development, accelerated urbanization, and the implementation of ecological protection strategy [33]. The contradiction between economic development, farmland protection, and ecologic sustainability is a problem that draws increasing attentions of the land management and planning departments. To effectively manage and analyze the land use status in China, the Second National Land Survey was launched (2007–2009). From then on, the map accuracy, data storage methods, and land classification system have been significantly improved. The land use databases have been continuous updated due to annual land change survey mechanism.

Our sample data (Figure 7) was actually measured from two towns during the Second National Land Survey and annual land change survey up to 2014. In this sample region with a land area of 15,373 hectares, the amount of land parcels ranges from 18,502 to 18,660, and the approximate average number of polygon's vertices is 79 (there are also line objects in snapshots, of an amount that ranges from 13,615 to 18,502, with average number of vertices 4). In 2009, agricultural land (farmland, garden, and forest) accounted for 78.56%, construction land accounted for 18.40%, and unused land accounted for 2.90% in this area. We observed that the land use changes of in the sample region are diversified: they were resulting from construction, reclamation, reforestation, and many other human activities; the transition from farmland to construction land is the highest (345 hectares, up to 2014), and changes between other land use types also occurred; as to the pattern of changing parcels, merging, splitting, and partial geometric changes are all discovered.



Figure 7. Experimental land-use data in Hunan. The first picture displays base state at 2009, and the following five pictures display incremental data indicating the location and time of change events (2009–2014).

5.2. Method for Evaluation

We consider three kinds of statistical method for comparison: basic TSP, the query-optimized TSP and the graph-based TSP. The basic TSP uses a spatial overlay to calculate transition areas and provides a lower bound of performance; the query-optimized TSP only retrieve changed parcels from STDB for the spatial overlay and reflects the performance of the present level; the graph-based TSP is the proposed method in this paper which only perform spatial overlay for the parcels with regard to the irreducible network.

The time complexity of TSP is mainly determined by the amount of polygons involved in the spatial overlay. Polygon intersection operation can achieve a time complexity of $O(p \times \log_2 p)$ [14,34], where p is the vertex number of each polygon. When we overlay two polygon layers containing n polygons each, then n^2 intersection operations have to be performed. So, the polygon overlay results in a time complexity of $O(n^2 \times p \log_2 p)$ and it determined the efficiency of basic TSP. In the query-optimized TSP, a spatio-temporal query is performed before the spatial overlay, but the query takes time no more than $O(m)$, where m is the amount of entities in the region. When calculating graph-based TSP with an STG of $|V|$ vertices and $|E|$ edges, the function *BuildSTFN* can achieve a time complexity $O(|E|)$, and the function *DecomposeSTFN* is based on BFS, which is known for $O(|V| + |E|)$ complexity. All three TSP methods have to perform spatial overlay, whose complexity has a quadratic relationship with the amount of polygons and a faster than linear relationship with polygon vertex number, which have a time expense far greater than all other procedures in any of the TSP methods when n and p are large enough. In addition, polygons processed by spatial overlay are necessary to read into memory and determines the space cost of the methods.

Notice that n varies with method. In basic TSP, all parcels in both timestamps T_b and T_e is processed. In the query-optimized TSP, only changed parcels between T_b and T_e are processed. In the graph-based TSP, the number of parcels process by spatial overlay operations is indicated by the sources and sinks in the irreducible network. Since spatial overlay dominate the performance of TSP, we can evaluate three methods by counting n .

5.3. Discussion of the Results

We tested TSP methods by calculating land use type transition matrix for the sample data. Table 1 shows the amount and ratio of the polygons that are processed by spatial overlay in TSP for the sample data.

Table 1. Number of polygons processed by spatial overlay in the experiment.

Query Condition	Basic TSP		Query-Optimized TSP		Graph-Based TSP	
	Amount	Ratio	Amount	Ratio	Amount	Ratio
2009–2010	37,162	1.0000	2105	0.0566	0	0.0000
2009–2011	37,254	1.0000	2530	0.0679	13	0.0003
2009–2012	37,311	1.0000	2613	0.0700	13	0.0003
2009–2013	37,320	1.0000	3723	0.0998	20	0.0005
2009–2014	38,938	1.0000	3914	0.1005	25	0.0006

Results in Table 1 imply that graph-based TSP is exceptionally efficient. It only performs a spatial overlay of a data size that is about 1/1558 of the size in the basic TSP and about 1/157 of the size in the query-optimized TSP for a five-year statistics. The results correspond to the actual performance of the program using ArcGIS Engine: the graph-based TSP completes calculations almost instantly, whereas the other two methods take several minutes to obtain the same results.

In general, land use change is sporadic in space and time [35] and mostly directional in terms of feature types [36]. These characteristics contribute to the high reducibility of STFN in this graph modeling. The efficiency of the method is verified by the sample data, which contains spatio-temporal changes diversified in reasons, transition types and geometric change forms. For the above reasons, the graph-based method is probable to retain high efficiency for land-use data elsewhere.

6. Conclusions

The spatio-temporal topological information in land-use databases can be utilized for the improvement of statistical processes. Land-use data organized with typical event-based spatio-temporal models can be described as a spatio-temporal graph, and further modeled as a transportation network with several interesting properties. Based on the proof that transition flows on such a transportation network can be determined under certain condition, a graph-based method is proposed to reduce the amount of polygons involved in spatial overlay pressure, which determines the efficiency of transition statistical process. The proposed method has excellent performance compared with other mentioned methods as is demonstrated in experiments with practical data from Changsha, and is expected to retain high efficiency for land use data elsewhere. The proposed method can be used to support statistical analysis or administration applications where efficient and accurate statistical processes are required. This approach can also be used for the change statistics on properties other than land use type, and can be possibly applied to transition calculations for STDB in land cover and cadastral fields, *etc.*

Acknowledgments: This research was funded by National Science and Technology Support Program of China (No. 2013BAJ05B01). The authors appreciate the experimental data provided by Hunan Information Center of Land and Resources. Additionally, we would like to thank Zhou Jinxin and Zhang Ran from Beijing Jiaotong University for helpful comments in mathematical modeling. The authors are very grateful to the anonymous referees for valuable remarks and comments, which significantly contributed to the quality of the paper.

Author Contributions: Yunbing Gao and Yuchun Pan laid the foundation of the research. Yipeng Zhang and Mingyangyan preprocessed the experimental data. Yipeng Zhang designed the method and wrote the paper manuscript. Yipeng Zhang, Yunbing Gao, Bingbo Gao and Yuchun Pan participated in the discussion of the key issues of the method. Yipeng Zhang, Bingbo Gao and Mingyang Yan revised the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ziadat, F.M.; Taimeh, A.Y. Effect of rainfall intensity, slope and land use and antecedent soil moisture on soil erosion in an arid environment. *Land Degrad. Dev.* **2013**, *24*, 582–590. [[CrossRef](#)]
- De Müelenrae, S.; Frankl, A.; Haile, M.; Poesen, J.; Deckers, J.; Munro, N.; Veraverbeke, S.; Nyssen, J. Historical landscape photographs for calibration of Landsat land use/cover in the Northern Ethiopian highlands. *Land Degrad. Dev.* **2014**, *4*, 319–335. [[CrossRef](#)]
- Muñoz-Rojas, M.; Jordán, A.; Zavala, L.M.; de la Rosa, D.; Abd-Elmabod, S.K.; Anaya-Romero, M. Impact of land use and land cover changes on organic carbon stocks in mediterranean soils (1956–2007). *Land Degrad. Dev.* **2015**, *26*, 168–179. [[CrossRef](#)]
- Zhao, G.; Mu, X.; Wen, Z.; Wang, F.; Gao, P. Soil erosion, conservation, and eco-environment changes in the loess plateau of China. *Land Degrad. Dev.* **2013**, *24*, 499–510. [[CrossRef](#)]
- Saha, D.; Kukal, S.S. Soil structural stability and water retention characteristics under different land uses of degraded lower himalayas of north-west India. *Land Degrad. Dev.* **2015**, *26*, 263–271. [[CrossRef](#)]
- Keesstra, S.D.; Geissen, V.; van Schaik, L.; Mosse, K.; Piiranen, S. Soil as a filter for groundwater quality. *Curr. Opin. Env. Sust.* **2012**, *4*, 507–516. [[CrossRef](#)]
- Gao, X.; Wu, P.; Zhao, X.; Wang, J.; Shi, Y. Effects of land use on soil moisture variations in a semi-arid catchment: Implications for land and agricultural water management. *Land Degrad. Dev.* **2014**, *25*, 163–172. [[CrossRef](#)]
- Berendse, F.; van Ruijven, J.; Jongejans, E.; Keesstra, S.D. Loss of plant species diversity reduces soil erosion resistance of embankments that are crucial for the safety of human societies in low-lying areas. *Ecosystems* **2015**, *18*, 881–888. [[CrossRef](#)]
- Wang, X.; Bao, Y. Study on the method of land use dynamics change research. *Prog. Geogr.* **1999**, *18*, 83–89.
- Zhu, H.; Li, X. Discussion on the index method of regional land use change. *Acta Geogr. Sinica* **2003**, *5*, 643–650.
- Rutherford, G.N.; Bebi, P.; Edwards, P.J.; Zimmermann, N.E. Assessing land-use statistics to model land cover change in a mountainous landscape in the European Alps. *Ecol. Model.* **2008**, *212*, 460–471. [[CrossRef](#)]
- Han, H.; Yang, C.; Song, J. Scenario simulation and the prediction of land use and land cover change in Beijing, China. *Sustainability* **2015**, *4*, 4260–4279. [[CrossRef](#)]
- Xie, H.L.; Liu, L.M.; Li, B.; Zhang, X.S. Spatial autocorrelation analysis of multi-scale land-use changes: A case study in Ongniud Banner, Inner Mongolia. *Acta Geo. Sin.* **2006**, *61*, 389–400.
- Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.; Stein, C. Computational geometry. In *Introduction to Algorithms*, 3rd ed.; The MIT Press: Cambridge, MA, USA, 2009; pp. 1022–1027.
- Liu, N.; Liu, R.; Zhu, G.; Xie, J. A spatial-temporal system for dynamic cadastral management. *J. Environ. Manage.* **2006**, *78*, 373–381. [[CrossRef](#)] [[PubMed](#)]
- Gao, Y.; Pan, Y.; Gao, B.; Zhang, X.; Gao, J.; Zhang, Y. Key Technologies for land use survey oriented spatio-temporal database construction. *Sci. Surv. Mapp.* **2015**, *40*, 49–54.
- Chen, X.; Wu, H.; Li, X.; Zhang, W. Study on event-based spatio-temporal data model for land use change. *J. Imag. Graph.* **2003**, *8*, 957–963.
- Teng, L.; Liu, R.; Liu, N. A study on spatio-temporal data model based on feature and event. *J. Rem. Sens.* **2005**, *9*, 634–639.
- Clementini, E.; Di Felice, P. A comparison of methods for representing topological relationships. *Inform. Sci. Appl.* **1995**, *3*, 149–178. [[CrossRef](#)]
- Tøssebro, E.; Nygård, M. Representing topological relationships for spatiotemporal objects. *GeoInformatica* **2011**, *15*, 633–661. [[CrossRef](#)]
- Tang, Y. The linkage mechanism and incremental extraction of land use updating. Ph.D. Thesis, Zhejiang University, Hangzhou, China, 2011.
- Schneider, M. Spatial and spatio-temporal data models and languages. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer US: Medford, MA, USA, 2009; pp. 2681–2685.
- Worboys, M.F.; Hearnshaw, H.M.; Maguire, D.J. Object-oriented data modelling for spatial databases. *Int. J. Geogr. Inform. Syst. Appl. Rem. Sens.* **1990**, *4*, 369–383. [[CrossRef](#)]
- Ruan, M.; Liu, R.; Liu, N.; Teng, L. Research on land utilization temporal statistic model based on event. *Appl. Resear. Comput.* **2005**, *7*, 31–33.

25. Wilcox, D.J.; Harwell, M.C.; Orth, R.J. Modeling Dynamic Polygon Objects in Space and Time: A New Graph-based Technique. *Cartogr. Geogr. Inform. Sci.* **2000**, *27*, 153–164. [[CrossRef](#)]
26. Yin, Z.; Li, L.; Ai, Z.X. A study of spatio-temporal data model based on graph theory. *Acta Geod. Cartogr. Sinica* **2003**, *32*, 168–172.
27. Spéry, L.; Claramunt, C.; Libourel, T. A spatio-temporal model for the manipulation of lineage metadata. *GeoInformatica* **2001**, *5*, 51–70. [[CrossRef](#)]
28. Del Mondo, G.; Rodriguez, M.; Claramunt, C. Modeling consistency of spatio-temporal graphs. *Data Knowl. Eng.* **2013**, *84*, 59–80. [[CrossRef](#)]
29. McBride, R. Advances in solving the multicommodity-flow problem. *Interfaces* **1998**, *28*, 32–41. [[CrossRef](#)]
30. West, D.B. *Introduction to Graph Theory*, 1st ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2000; p. 470.
31. Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.; Stein, C. Linear programming. In *Introduction to Algorithms*, 3rd ed.; The MIT Press: Cambridge, MA, USA, 2009; pp. 862–863.
32. Wing, O.; Kim, W.H. The path matrix and switching functions. *J. Frankl. Instit.* **1959**, *268*, 251–269. [[CrossRef](#)]
33. Liu, J.; Kuang, W.; Zhang, Z.; Xu, X.; Qin, Y.; Ning, J.; Chi, W. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *J. Geogr. Sci.* **2014**, *2*, 195–210. [[CrossRef](#)]
34. Žalik, B. Two efficient algorithms for determining intersection points between simple polygons. *Comput. Geosci.* **2000**, *26*, 137–151. [[CrossRef](#)]
35. Peuquet, D.J. Time in GIS and Geographical Databases. *Geogr. Inform. Syst.* **1999**, *1*, 91–102.
36. Wang, S.; Zhang, Z.; Zhou, Q.; Wang, C. Study on Spatial temporal features of land use/land cover change based on technologies of RS and GIS. *J. Rem. Sens.* **2002**, *6*, 223–228.



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).