

# Uso de Grafos de Proveniência para Análise Temporal de Uso do Solo em Centros Urbanos: uma Abordagem Prática\*

Raffael Paranhos<sup>1</sup>, Marcos Lage<sup>1</sup>, Daniel de Oliveira<sup>1</sup>

<sup>1</sup>Instituto de Computação - Universidade Federal Fluminense (IC/UFF)

raffaelp@id.uff.br, {mlage,danielcmo}@ic.uff.br,

**Resumo.** *Estudar a evolução do uso do solo é essencial para descobrir a origem de problemas de infraestrutura e identificar áreas com urbanização inadequada. Identificar dificuldades atuais e erros passados no uso do solo é crucial para definir políticas públicas eficazes. No entanto, analisar a dinâmica das modificações dos terrenos ao longo do tempo é desafiador, especialmente devido à complexidade espacial do problema. Este artigo descreve a abordagem LandEvol-PROV, que identifica transformações nos terrenos ao longo dos anos e cria um grafo de proveniência para análise espaço-temporal do uso do solo. A avaliação utilizando o conjunto de dados MapPLUTO confirmou a viabilidade da abordagem proposta.*

**Abstract.** *Studying the evolution of land use is crucial to uncover infrastructure problems and identify areas with inadequate urbanization. Identifying current difficulties and past errors in land use is vital for formulating effective public policies. However, analyzing the dynamics of land modifications over time is challenging, especially due to spatial complexity. This paper describes the LandEvol-PROV approach, which identifies land transformations over the years and creates a provenance graph for space-time analysis of land use. An evaluation using the MapPLUTO dataset confirmed the feasibility of the proposed approach.*

## 1. Introdução

O crescimento das cidades, especialmente no último século, exigiu que os governos estabelecessem leis para determinar como o solo pode ser ocupado [Munneke 2005]. Um exemplo é a definição dos gabaritos de construção e dos tipos de empreendimentos permitidos em diferentes áreas da cidade. O objetivo destas regras é controlar o tamanho das construções e evitar interferências negativas em seu entorno, como sombreamento de outras edificações. As leis e regulamentações sobre o uso do solo estabelecem o chamado *Zoneamento Urbano*. No entanto, as regras de zoneamento estão sujeitas a modificações ao longo do tempo, usualmente motivadas por fatores socioeconômicos. Portanto, é essencial estudar a evolução do uso do solo ao longo do tempo, tanto para compreender problemas que possam ter ocorrido no passado, quanto para estabelecer políticas públicas de ocupação do solo no futuro. Para realizar esse estudo, é necessário dispor de *datasets* que descrevam os terrenos da cidade, incluindo seus atributos e geometrias.

Nos últimos anos, impulsionadas principalmente por políticas de dados abertos [Yang and Wu 2021], muitas cidades têm disponibilizado *datasets* que descrevem o uso do solo ao longo do tempo, incluindo a geometria de cada terreno e vários metadados associados. Um exemplo é o MapPLUTO [NYCDP 2023] da cidade de Nova Iorque, que possui

---

\*O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Os autores gostariam ainda de agradecer ao CNPq (grant 311898/2021-1) e FAPERJ (grant E-26/202.806/2019) pelo apoio financeiro.

mais de 80 atributos para cada terreno e, somente nas regiões de *Manhattan* e do *Brooklyn*, detalha informações de 300.000 lotes a cada ano. Embora *datasets* como o MapPLUTO representem um grande avanço, sua análise ainda apresenta desafios, especialmente devido à natureza espacial envolvida. Por exemplo, ao longo dos anos, terrenos podem ser criados ou combinados, o que exige não apenas uma atualização dos metadados, mas também da geometria desses lotes.

Considerando que as transformações nos lotes ao longo do tempo podem ser representadas por meio de um grafo (em que uma transformação recebe um ou mais terrenos como entrada e gera um ou mais terrenos como saída), os dados de proveniência [Freire et al. 2008] surgem como uma opção para registrar o histórico de transformações dos terrenos ao longo dos anos. Isso proporciona uma visão abrangente da evolução do zoneamento em uma cidade, permitindo que os tomadores de decisão consultem e analisem como os terrenos se modificam a cada ano, por meio de consultas aos metadados e transformações geométricas dos lotes. Embora o uso de dados de proveniência não seja algo recente, até o momento foi pouco explorado no contexto de análise de uso do solo. Mesmo os trabalhos que já aplicam dados de proveniência nesse contexto [Bennett et al. 2011, Al-Ageili and Mouhoub 2022] focam em analisar o tipo de atividade exercida nos terrenos (*i.e.*, comercial, residencial) e não a evolução espacial dos terrenos.

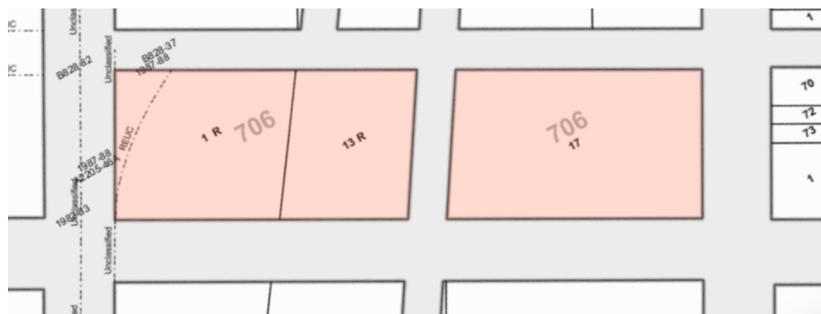
Este artigo apresenta a abordagem intitulada LandEvol-PROV, que formaliza operações de transformação nos lotes (*e.g.*, subdivisão e agregação) implicitamente descritas em *datasets* sobre uso dos solos. A partir das transformações identificadas, a abordagem gera uma base de dados de proveniência e exporta grafos de proveniência, que auxiliam os tomadores de decisão na análise da evolução temporal da geometria dos terrenos da cidade. Para avaliar a viabilidade da LandEvol-PROV, utilizamos o *dataset* MapPLUTO disponibilizado pela prefeitura de Nova Iorque. Os resultados obtidos demonstraram o potencial da abordagem desenvolvida. O artigo está organizado em quatro seções, além da Introdução. A Seção 2 apresenta o referencial teórico e os trabalhos relacionados. A Seção 3 detalha a abordagem proposta. A Seção 4 realiza um estudo de viabilidade e, por fim, a Seção 5 apresenta as conclusões do artigo.

## 2. Referencial Teórico e Trabalhos Relacionados

### 2.1. Conjuntos de Dados de Uso de Solo: PLUTO e MapPLUTO

Existem vários *datasets* que pretendem representar o uso do solo [Fritz et al. 2017, Krumm and Krumm 2019], mas esses *datasets* não focam exclusivamente no uso do solo em cidades. Ao contrário desses *datasets*, o PLUTO (*Property Land Use Tax lot Output*) [NYCDP 2023] descreve as características de todos os terrenos tributáveis da cidade de Nova Iorque ao longo dos anos. O MapPLUTO é uma versão do PLUTO que contém, além dos atributos fornecidos no PLUTO, a geometria dos lotes. Sendo assim, o MapPLUTO (fragmento visualizado na Figura 1) é um conjunto de dados espaço-temporal. Observamos que tanto geometrias quanto os metadados dos terrenos podem sofrer alterações ao longo do tempo.

Esse *dataset* é atualizado pelo menos uma vez ao ano e o arquivo relativo à atualização é disponibilizado em formato *shapefile*, que ocupa cerca de 850MB de espaço em disco. Cada terreno no MapPLUTO é identificado unicamente por um código BBL (concatenação dos códigos de distrito, quarteirão e terreno) e um mesmo terreno pode ser decomposto ou combinado com outros terrenos ao longo de diversos anos. Cada versão do



**Figura 1. Visualização do bloco 706 em Manhattan a partir dos dados do MapPLUTO. Os três terrenos desse bloco (1, 13 e 17), estão assinalados em vermelho claro**

*dataset* é composta por uma coleção de cinco arquivos, sendo um para cada região administrativa de Nova Iorque. Esses arquivos mantêm indexados os dados espaciais com os metadados não espaciais dos lotes. Além disso, eles podem ser convertidos para GeoJSON, formato mais amigável e aceito por múltiplas bibliotecas para análise de dados espaciais.

## 2.2. Trabalhos Relacionados

Apesar de dados de proveniência já serem explorados com sucesso em múltiplas áreas do conhecimento [de Oliveira et al. 2019], seu uso ainda é pouco explorado no contexto de análise de uso de solo. [Bennett et al. 2011] apresentam um arcabouço baseado em proveniência para compreender a dinâmica do uso do solo no estado de Montana (EUA). Entretanto, somente introduzem os dados de proveniência para compreender como foi gerado um modelo de previsão de uso de solo. Os autores registram a varredura de um grande número de parâmetros para investigação de um espaço de soluções grande. [Al-Ageili and Mouhoub 2022] propõem a aplicação de ontologias em um arcabouço que identifica o uso do solo ao longo dos anos. Entretanto, o trabalho foca apenas em analisar a mudança de propósito do uso do solo, sem considerar uma análise das transformações espaciais dos terrenos. Existem trabalhos que abordam o uso do MapPLUTO, mas não no contexto de análise da evolução do solo. [Aerts et al. 2013] o utilizam o MapPLUTO para estimar o dano financeiro causado por enchentes causadas por furacões. [Ma et al. 2020] propõem o uso de métodos de Aprendizado de Máquina para a construção de um modelo a partir do MapPLUTO para identificar os fatores que determinam o valor de um terreno.

## 3. Abordagem Proposta: LandEvol-PROV

Apesar de representar um avanço no que tange a disponibilização de dados de uso de solo, a estrutura de *datasets* como o MapPLUTO ainda limita a capacidade analítica do usuário. Como o MapPLUTO organiza seus dados por ano, uma análise temporal dos dados requer o desenvolvimento de *scripts* específicos. A ideia do LandEvol-PROV é identificar automaticamente um conjunto de transformações nos terrenos e estruturar essa informação por meio de um grafo de proveniência que pode ser consultado pelo usuário, sem a necessidade de desenvolver *scripts* próprios. A arquitetura da LandEvol-PROV é apresentada na Figura 2, e é composta por quatro componentes: (i) Importador, (ii) Extrator de Proveniência, (iii) Base de Proveniência e (iv) API.

O processo se inicia com o *Importador*, que obtém os arquivos do MapPLUTO, e os armazena em um sistema de arquivos local. Uma vez que todos os arquivos relativos aos anos de interesse já se encontram armazenados, o *Extrator de Proveniência* é invocado.

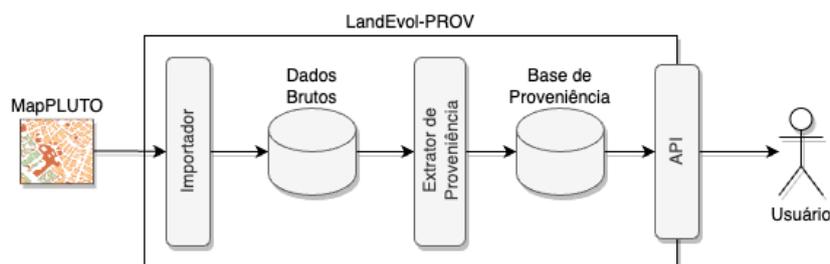


Figura 2. Arquitetura da LandEvo1-PROV.

Ele compara os arquivos ano a ano, e por meio da execução de uma intersecção espacial de um terreno com todos os terrenos de uma mesma região do ano seguinte, identifica as transformações realizadas entre um ano  $a_m$  e  $a_{m+1}$ . A intersecção espacial descreve quais terrenos estão ocupando o mesmo espaço entre os anos e permite caracterizar diferentes transformações, *e.g.*, um *split*, em que um território que antes era ocupado por um único terreno e passou a ser ocupado por dois. Nesta abordagem propomos quatro tipos: (i) *Merge*, onde dois ou mais terrenos se tornam um, (ii) *Split*, onde um terreno se torna dois ou mais, (iii) *Maintain*, quando um terreno não sofre transformação e (iv) *Rearrange*, quando a transformação é a composição de *Merges* e *Splits*. A Figura 3 apresenta mudanças nos terrenos de uma região ao longo do tempo. O terreno *E* na Figura 3(b) é um *Merge* entre os terrenos *A* e *B* da Figura 3(a), enquanto que os terrenos *F* e *G* na Figura 3(c) foram formados a partir de um *Split* do terreno *E* da Figura 3(b). Foram utilizadas as bibliotecas *PyShp* e *Shapely* para a manipulação de dados geométricos, e o cálculo da área de intersecção foi realizado pela função *intersection* do *Shapely*.

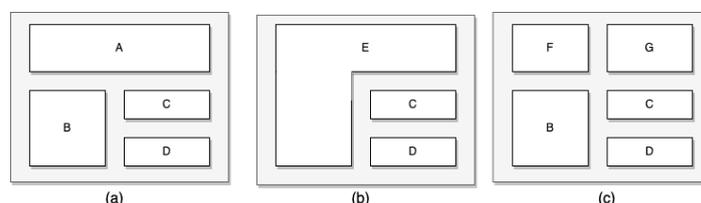


Figura 3. Transformações possíveis na abordagem LandEvo1-PROV.

Uma vez que as transformações tenham sido identificadas, o *Extractor de Proveniência* carrega os dados de proveniência em uma *Base de Dados de Proveniência*. Na sua versão atual, a base foi instanciada no Neo4J. No grafo gerado, cada terreno é representado como um nó para cada ano em que o mesmo aparece no *dataset*. Cada nó do grafo possui todos os metadados presentes no MapPLUTO correspondentes a seu lote. Sempre que uma transformação é identificada, uma aresta ligando os dois nós atingidos é incluída no grafo. A aresta possui como atributos a área de intersecção e qual a porcentagem de sobreposição da área dos terrenos em questão. Uma vez que a *Base de Proveniência* está populada, os usuários podem submeter consultas ao LandEvo1-PROV. As consultas são submetidas por meio de uma API REST. A API submete consultas parametrizáveis em Cypher (linguagem de consulta do Neo4J) por meio de quatro métodos: (i) *Get Edges*, que retorna todos os nós e arestas adjacentes que correspondem aos critérios de busca, (ii) *Get Splits*, que retorna todos os nós que participam de um *split* e arestas adjacentes, que correspondem aos critérios de busca, (iii) *Get Merges*, que retorna todos os nós que participam de um *merge* e arestas adjacentes, que correspondem aos critérios de busca, e (iv) *Get Rearranges*, que retorna todas os nós que participam de um *rearrange* e arestas adjacentes, que correspondem aos critérios de

busca. Todos os métodos permitem ao usuário definir o intervalo de pesquisa e os números dos blocos (área de interesse no mapa). Além disso, nos métodos *Get Edges*, *Get Splits* e *Get Merges*, também é possível filtrar pelos metadados importados do MapPLUTO.

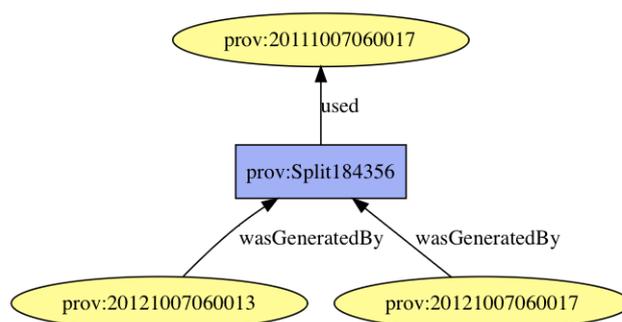
O resultado da busca é retornado no formato PROV-JSON, que é uma representação compacta e fiel da recomendação PROV do W3C [Groth and Moreau 2013] para representação de proveniência. Esta representação permite buscas eficientes. No arquivo PROV-JSON exportado, todos os terrenos são definidos como *Entidades* e as transformações *Merge*, *Split*, *Maintain* e *Rearrange* como *Atividades*. Toda transformação utiliza (relacionamento *used*) um ou mais terrenos, e todo terreno é gerado por uma transformação (relacionamento *wasGeneratedBy*). A proveniência final será composta então por uma série de Entidades interligadas por Atividades que descrevem ano a ano, qual a transformação sofrida por aquele terreno. Também foi implementada uma opção que permite ao usuário escolher quais tipos de transformação deseja exportar. Finalmente, tanto os terrenos quanto as transformações incluem atributos como área e ano. A LandEvol-PROV pode ser obtida em <https://github.com/UFFeScience/LandEvol-PROV>.

#### 4. Estudo de Viabilidade: *Hudson Yards*

Para avaliar a LandEvol-PROV, realizamos um estudo de viabilidade. Como região de análise, foi escolhida a área do *Hudson Yards* na parte oeste de *Manhattan*. Esta região possui um projeto de desenvolvimento imobiliário homônimo que tem como objetivo transformar a área por meio da construção de múltiplos arranha-céus, parques e espaços para arte e cultura. Por conta desse projeto, são esperadas muitas mudanças na região ao longo dos anos. A análise aqui apresentada foca nos terrenos do bloco 706 (Figura 1). No Bloco 706, há múltiplos *Splits* de terrenos a partir de 2012 (Figura 4), mas um *Split* em especial (de 1 terreno para 2) apresentou uma incongruência de cerca de 72% da área de intersecção, *i.e.*, o terreno original corresponde a apenas 28% da área dos terrenos novos. Neste *Split*, o terreno 17 perde 72% da sua área total e o terreno 13 é criado a partir do terreno 17, e utiliza 1,8% da área do terreno 17 (área tracejada em vermelho na Figura 4). Analisando fontes de informações externas [NYCDOF 2023], descobrimos a criação de uma rua na região, o que explica a diminuição da área total envolvida originalmente no terreno 17. Esses terrenos também fizeram parte de um *Rearrange* de 7 terrenos para 4 e compreendem uma série de mudanças que ocorreram no bloco 706 anteriores ao projeto de revitalização (os eventos ocorreram em 2011 e o projeto iniciou em 2012). É provável que estas operações representem as transações imobiliárias realizadas e a legalização da composição final dos terrenos do empreendimento, no ano anterior ao início das obras.

#### 5. Conclusão

Analisar o uso do solo nas cidades ao longo dos anos é uma tarefa prioritária para os governos, já que permite que tomadores de decisão possam elaborar políticas públicas eficientes. Apesar de existirem *datasets* de uso do solo já disponíveis em muitas cidades, a análise desses dados pode não ser trivial por conta da sua complexidade espacial. Em geral, os usuários precisam usar ferramentas não triviais ou desenvolver seus próprios *scripts* para desempenhar essa tarefa. Esse artigo apresenta a abordagem LandEvol-PROV que importa *datasets* de uso de solo e identifica as alterações geométricas realizadas nos terrenos ao longo dos anos. Essas transformações são estruturadas em uma base de dados de proveniência que pode ser consultada, facilitando assim a análise. A abordagem foi avaliada em um estudo de viabilidade com dados do MapPLUTO de Nova Iorque e se mostrou promissora. Trabalhos futuros incluem a exploração de novos *datasets* de outras cidades.



**Figura 4. Fragmento do grafo de proveniência da criação do terreno 13 a partir de um cisão do lote 17 no bloco 706 de *Manhattan* entre 2011 e 2012.**

## Referências

- [Aerts et al. 2013] Aerts, J. C., Lin, N., Botzen, W., Emanuel, K., and de Moel, H. (2013). Low-probability flood risk modeling for new york city. *Risk Analysis*, 33(5):772–788.
- [Al-Ageili and Mouhoub 2022] Al-Ageili, M. and Mouhoub, M. (2022). An ontology-based information extraction system for residential land-use suitability analysis. *Int. J. Softw. Eng. Knowl. Eng.*, 32(7):1019–1042.
- [Bennett et al. 2011] Bennett, D. A., Tang, W., and Wang, S. (2011). Toward an understanding of provenance in complex land use dynamics. *J. of Land Use Science*, 6(2-3):211–230.
- [de Oliveira et al. 2019] de Oliveira, D. et al. (2019). *Data-Intensive Workflow Management: For Clouds and Data-Intensive and Scalable Computing Environments*. M. & Claypool.
- [Freire et al. 2008] Freire, J., Koop, D., Santos, E., and Silva, C. T. (2008). Provenance for Computational Tasks: A Survey. *Computing in Science & Engineering*, 10(3):11–21.
- [Fritz et al. 2017] Fritz, S. et al. (2017). A global dataset of crowdsourced land cover and land use reference data. *Scientific Data*, 4(1):170075.
- [Groth and Moreau 2013] Groth, P. and Moreau, L. (2013). W3C PROV. Available at <https://www.w3.org/TR/prov-overview/>.
- [Krumm and Krumm 2019] Krumm, J. and Krumm, K. (2019). Land use inference from mobility traces. In *Proc. of the 3rd ACM SIGSPATIAL*, page 1–4. ACM.
- [Ma et al. 2020] Ma, J. et al. (2020). Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques. *Land use policy*, 94:104537.
- [Munneke 2005] Munneke, H. J. (2005). Dynamics of the urban zoning structure: An empirical investigation of zoning change. *Journal of Urban Economics*, 58(3):455–473.
- [NYCDP 2023] NYCDP (2023). Pluto. [www.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page](http://www.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page). Acesso: Junho 2023.
- [NYCDOF 2023] NYCDOF (2023). Department of finance digital tax map. <http://gis.nyc.gov/taxmap/>. Acessado em: Junho 2023.
- [Yang and Wu 2021] Yang, T. and Wu, Y. (2021). Looking for datasets to open: An exploration of government officials’ information behaviors in open data policy implementation. *Gov. Inf. Q.*, 38(2):101574.