Identifying Street Name Evolution in Semantic, Temporal, and Geographic Spaces

Jina Kim[a,1], Min Namgung[a,1], Johannes H. Uhl[b], Keith Burghardt[c], Yao-Yi Chiang[a], Stefan Leyk[b], Kristina Lerman[c]

[a] *University of Minnesota, Minneapolis, MN, USA*
[b] *University of Colorado Boulder, Boulder, CO, USA*
[c] *University of Southern California, Los Angeles, CA, USA*

Abstract

Streets are often named after landmarks, famous individuals, professions, and other concepts, which are often related to the region's cultural, historical, and political background (Bancilhon et al., 2021; Chen et al., 2021; Rusu, 2019). For example, Chen et al. (2021) classified 31 street names with semantic information and analyzed the distribution of categorized semantic types based on historical events to explain urban development. However, the existing approaches have dealt with a limited number of semantic types, which is difficult to generalize to wide-ranging geographic areas. Thus, we propose an automatic pipeline that can capture the semantic meanings of street names in the United States and analyze their evolution in geographic space and time.

To investigate how street names and their semantic meanings change over space and time, we use names, geographic locations, and age estimates of streets in 2,651 counties of the U.S. This dataset has been created based on Zillow's Transaction and Assessment Dataset (ZTRAX), an industry-assembled property and real-estate database, holding rich attributes for over 150,000,000 cadastral built-up properties in the U.S. Attributes include the construction year of buildings within a property, their address, geolocation, function, size, and other physical and ownership-related attributes. Based on the property addresses in ZTRAX, we group the property records by street name, separately for each place (i.e., city, town, or village) and derive several street-level statistics, such as the earliest construction year per street, as a rough age estimate of a given street. We assume high levels of coherence

[1]Equal contribution.

between the evolution of buildings and roads to infer street age based on nearby building age. This is a common assumption in historical road network modeling (Barrington-Leigh and Millard-Ball, 2015; Burghardt et al., 2022), and has been employed to estimate the age of built-up areas over 200 years (Leyk and Uhl, 2018). This method is robust to bias introduced by building replacements, as a single original building per street is sufficient to obtain a realistic age estimate. This way, we create a unique, spatio-temporal street dataset covering large parts of the U.S. Importantly, this dataset does not take into account street renaming activities and instead reflects the current naming. Figure 1 shows a subset of this dataset for a part of Boulder (Colorado).

We preprocess the street names by converting them into title cases and removing the street suffix abbreviations and numbers. We then use the preprocessed street names to find their corresponding entities in existing knowledge graphs, such as Wikidata (Vrandečić and Krötzsch, 2014). We use both partial and exact matches between the entity name and preprocessed street name to find the entity that could represent the semantic meaning of a street name. The matched cases are one-to-one or one-to-many, in which one street name is matched to either a single entity or to multiple entities. For example, in the one-to-one match case 'Daylight Drive', there can be one matched case (e.g., the light of day). In contrast, given a street name such as 'Adams Boulevard', there are multiple entities in the knowledge graph that contain the name 'Adams' (e.g., 'Arthur Adams', 'John Adams', and 'Adams County'). To retrieve the fixed-size representation of matched entities, we employ pretrained embeddings that capture the semantic meaning of entities and their relations in the knowledge graph, such as TransE (Bordes et al., 2013).

To select a single semantic meaning from the one-to-many matched cases, we are inspired by Tobler's First Law of Geography (Tobler, 1970) and assume that nearby street names are semantically similar (i.e., share similar naming concepts). This would seem to occur in both planned neighborhoods and organically grown roads. We first map the matched entities and their relations to other entities (e.g., 'John Adams' is an *instance of human*, and his *occupation* is a *politician*) into embeddings in a new representation space using graph representation learning algorithms (Bordes et al., 2013; Hamilton, 2020). This way, each street name receives a list of embedding candidates, each representing the semantic meaning of a matched entity in the knowledge graph. We then compare the embedding candidates of nearby streets. If a

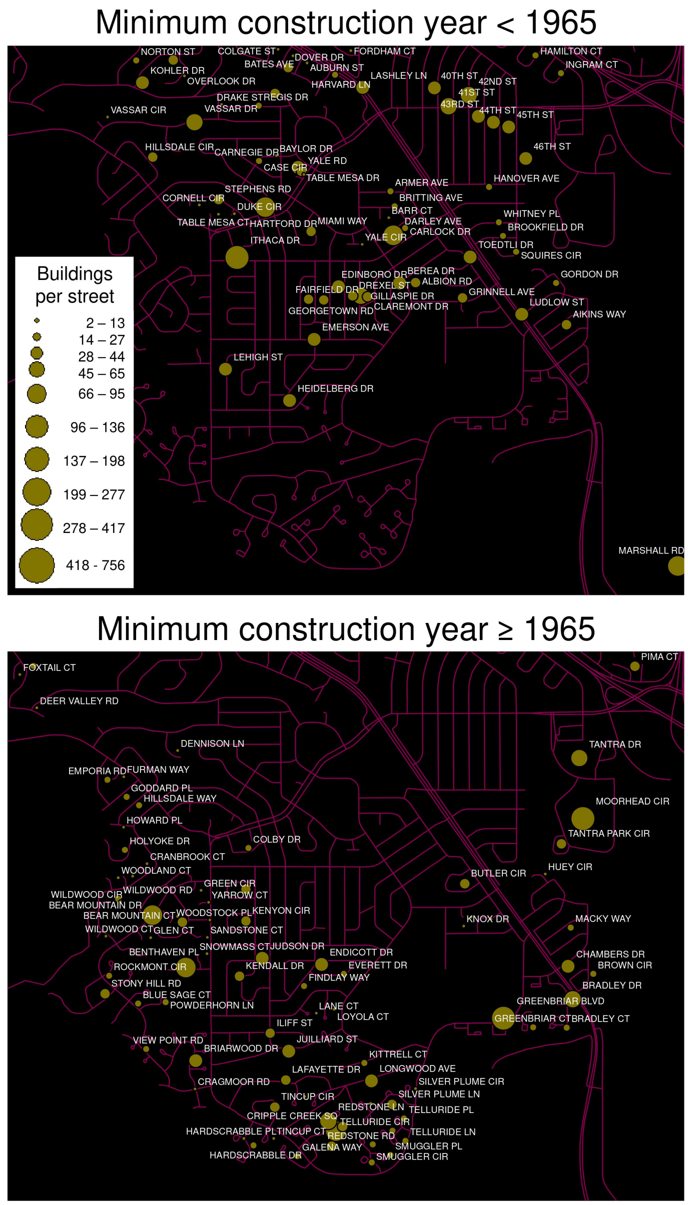

Figure 1: Spatial representation of ZTRAX-based street level aggregate statistics (i.e., earliest construction year and total number of buildings per street) shown for a part of Boulder, Colorado. Each street is represented by the centroid coordinates of all properties along the street (based on the street name retrieved from the property addresses), and the dot size represents the number of these properties per street. The top panel shows streets where the earliest construction year of the properties per street is <1965, and the bottom panel shows streets with a minimum construction year >1965. Basemap: 2019 National Transportation Dataset road network data.

group of similar embeddings (i.e., semantically similar) frequently appears on the candidate list of multiple streets in a local neighborhood, we use them as the final embedding to represent the semantic meaning of a street name in that neighborhood. For example, a street named 'Washington' near 'Adams' street in Los Angeles is more likely to be an '*instance of*' *human* whose '*occupation*' is *politician* than an '*instance of*' *capital*.

After identifying the semantic meanings of each street name, we explore patterns of semantically similar street names over time and space. We first group the streets by specific time intervals using the street age estimate and compare the spatially clustered semantic meanings of new street names with the clusterings of older street names. By identifying the semantic contexts of street names and their evolution over time, across space, and along the rural-urban continuum, the proposed approach has the potential to improve our understanding of the socio-historical processes surrounding infrastructure changes, (sub)-urbanization and land development processes in the recent past in the U.S. and elsewhere where comparable data are available.

References

Bancilhon, M., Constantinides, M., Bogucka, E.P., Aiello, L.M., Quercia, D., 2021. Streetonomics: Quantifying culture using street names. Plos one 16, e0252869.

Barrington-Leigh, C., Millard-Ball, A., 2015. A century of sprawl in the united states. Proceedings of the National Academy of Sciences 112, 8244–8249.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O., 2013. Translating embeddings for modeling multi-relational data. Advances in Neural Information Processing Systems, 26.

Burghardt, K., Uhl, J.H., Lerman, K., Leyk, S., 2022. Road network evolution in the urban and rural united states since 1900. Computers, Environment and Urban Systems 95, 101803.

Chen, Y., Zhang, F., Li, X., Zhang, C., Chen, N., Du, Z., Liu, R., Wang, B., 2021. Deep understanding of urban dynamics from imprint urban toponymic data using a spatial–temporal–semantic analysis approach. ISPRS International Journal of Geo-Information 10, 278.

Hamilton, W.L., 2020. Graph representation learning. Synthesis Lectures on Artifical Intelligence and Machine Learning 14, 1–159.

Leyk, S., Uhl, J.H., 2018. Hisdac-us, historical settlement data compilation for the conterminous united states over 200 years. Scientific data 5, 1–14.

Rusu, M.S., 2019. Shifting urban namescapes: Street name politics and toponymic change in a romanian (ised) city. Journal of Historical Geography 65, 48–58.

Tobler, W.R., 1970. A computer movie simulating urban growth in the detroit region. Economic geography 46, 234–240.

Vrandečić, D., Krötzsch, M., 2014. Wikidata: a free collaborative knowledgebase. Communications of the ACM 57, 78–85.