

Week Four – Assignment Centrality Measures

Authors

Tony Fraser and Mark Gonsalves

Datasource overview:

- This is the **CAL FIRE Damage Inspection Program (DINS)** database of structures damaged or destroyed by wildland fires in California **since 2013**, as documented by CAL FIRE and partnering agencies. Structures damaged before 2013 do not have a digital record. Fires in LRA (Local Responsibility Area) or FRA (Federal Responsibility Area) responsibility areas may or may not be included.
- Starting in 2018, the DINS program began collecting data on all structures (damaged and non-damaged). Before 2018, only damaged/destroyed structures were recorded.
- This database includes structures impacted by wildland fire that are **inside or within 100 meters** of the fire perimeter. Structure type, construction features, and defensible space attributes are determined as accurately as possible, even when the structure is completely destroyed. Some attributes may be missing if they could not be determined.
- Fire damage and poor access can limit inspections. While all inspections follow a systematic process, some impacted structures may not be identified, leading to a small margin of error.
- The database contains two address fields:
 - Field-determined address: Street number, street name, and street type, entered by the inspector based on on-site observations.
 - Parcel-based address: Address (parcel) and APN (parcel), added through a spatial join after data collection.

Datasource statistics:

- Rows: 130,717
- Columns: 46
- Update Frequency: Monthly or after a big fire
- Coverage: 2013 to
- Last Update: February 11, 2025, 4:07 PM (UTC-08:00)

Loading the data:

- We will load the data the same way we did for assignment 2.
- `import os`
- `import pandas as pd`
- `import numpy as np`
- `import networkx as nx`
- `import pickle`
- `from scipy.spatial import cKDTree`
- `from geopy.distance import geodesic`
- `from data620.helpers.glimpse import glimpse`
- `from data620.helpers.dins_utils import clean_column_names`
-
- `dins = pd.read_csv("https://tonyfraser-data.s3.us-east-1.amazonaws.com/calfire/raw/POSTFIRE_MASTER_DATA_SHARE_2064760709534146017.csv")`

Categorical Variables:

We have many categorical variables that can be used for the analysis depending on which comparisons we decide to utilize:

Some of the categorical variables are:

- Structure Type
- Structure Category
- Damage Severity
- CALFIRE Unit
- Roof Construction
- Vent Screen
- Exterior Siding

Our comprehensive analysis will include:

- Network construction using geographical proximity of structures
- Calculation and visualization of centrality measures
- Statistical analysis comparing centrality across structural categories
- Investigation of how building characteristics correlate with fire damage spread

Hypothetical Outcomes predicted from comparing degree centrality across categorical groups:

- By comparing degree centrality across the categorical groups, we can predict that structures with a higher degree of centrality in the “Destroyed” category were more likely in buildings that were residential as compared to commercial.