

GeoGraph

COVID-19 Public Health Hotspot Modeling

Company Introduction

SignalFrame maintains a geospatial temporal graph platform that relates human movements through space using detectable WiFi and Bluetooth signals emitted by devices like smartphones, wearables, electronics, cars, and network routers. The graph is built and refreshed by signal observations crowdsourced from approximately 30 million monthly enabled smartphones and tablets moving around the world.

Core applications are focused on surfacing physical-social networks, community-based analytics, and tracking network evolution over time. For example, SignalFrame reveals co-location between people (i.e., which devices share the same space at the same time), and the nature and weight of relationships between those individuals (or groups of individuals). Similarly, co-occupation of spaces (i.e., which locations tend to share visitors across time) highlights relationships between different places and surfaces locations of significance to a social network.

GeoGraph Overview

SignalFrame has developed a “geographical-graph” (GeoGraph) model for public health applications to assess risk levels based on historical and current information about the spread of the COVID-19 virus. GeoGraph links geographic locations based on sets of mobile devices that those locations share in common across time. Geographic linkages rely on underlying social networks (excluding people who are merely in “transit”), and are updated in real-time as behaviours change.

For example, individuals that regularly frequent a certain Community Center may live in a distinct and far-away Residential Neighborhood. As a result, the Community Center and Residential Neighborhood form part of a geo-network linked by the social network of visitors. Without GeoGraph, that linkage would not be obvious given the lack of physical proximity between the two locations.

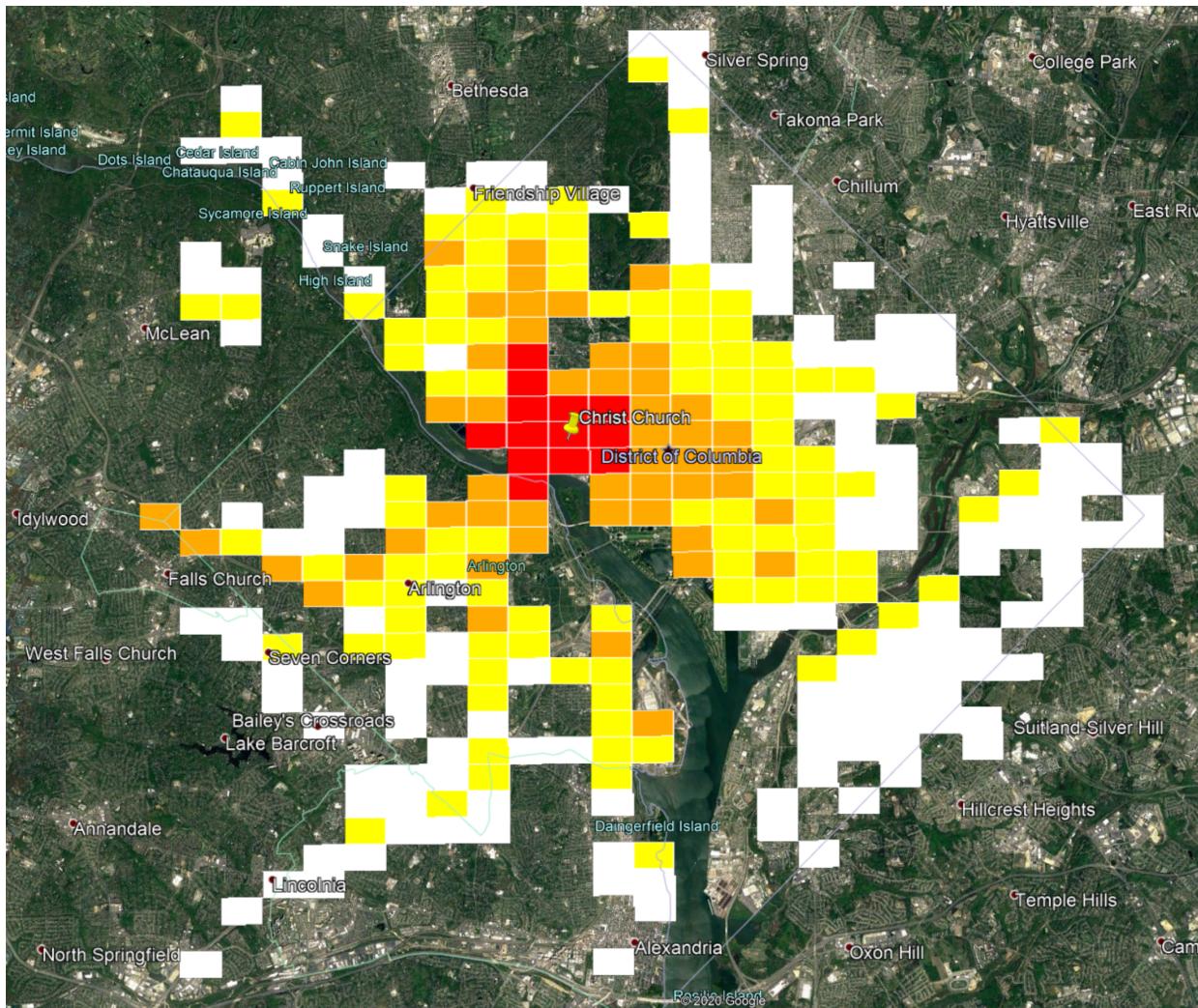
Unlike a simpler model that tracks individuals only on the basis of physical proximity at moments in time, SignalFrame’s approach helps predict COVID-19 spread based on GeoGraph’s linkages. Whereas the person-to-person tracking model suffers from data sparsity and location inaccuracies, GeoGraph shows where the virus may be “active” – even after an infected individual has left – and which other locations it is likely to spread to through human networks.

Applied to a COVID-19 outbreak, the analysis extends to using newly confirmed cases to predict how the virus may spread through more and more locations over time. Updated as new data becomes available, GeoGraph linkages also provide *real-time* monitoring of social-distancing failures and alert decision-makers with risk-scores for related locations.

Example: Christ Church, Washington, DC

On March 7, 2020, the leader of Christ Church at 3116 O Street NW in the Georgetown neighborhood of Washington became the District of Columbia's first confirmed COVID-19 case. The previous Sunday, March 1, the individual was present for three services, during one of which he distributed Communion. After his diagnosis, 500 parishioners who may have come into contact with him were urged to self-isolate, and in the days since at least five others involved with the church have tested positive for COVID-19.

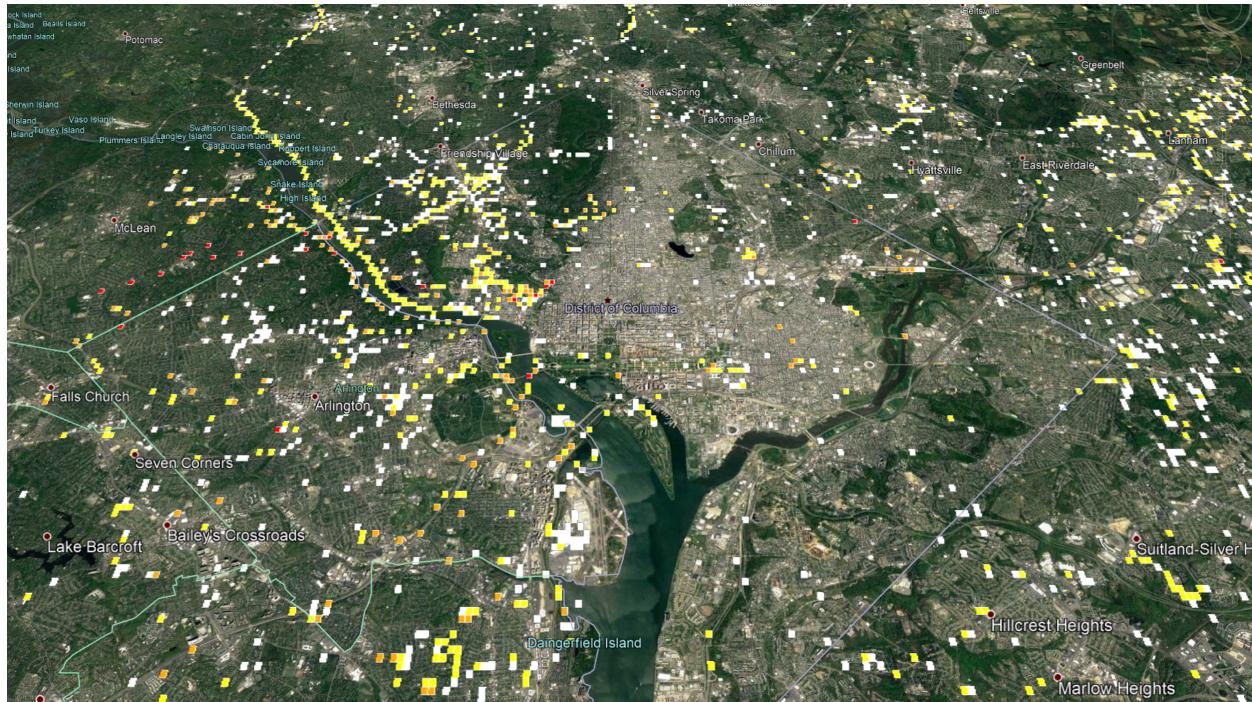
A common approach may map spread risk through contact-tracing using an incomplete dataset of mobile-device identifiers. For this real-time example, SignalFrame's GeoGraph instead uses the church as an epicenter to help inform where the response and allocation of resources should be focused.



The image above shows GeoGraph centered around the church (denoted by the pindrop). Each polygon has an area of 600m x 600m. Polygons are color-coded to model COVID-19 "hops" spreading outward from the church through prior-established GeoGraph physical-social

networks and across time. Immediate risk locations are in red, orange is one hop removed, followed by yellow, and finally white polygons for three hops removed. Each additional hop represents the potential spread of the virus within two to three days. This high-level image supports a common-sense assumption that the downtown Washington, DC-area would need to be prioritized for social distancing to curb the spread from the church.

However, GeoGraph modelling using finer-grained locations (100m x 100m polygons) produces an image that reveals far more accurate linkages between the church and other (at-risk) locations.



In this example, decision-makers receive real-time, actionable insight that their resources and social-distancing policies would be most effective to curb the COVID-19 spread from the church when directed at the McLean and Arlington suburbs in Virginia – the next potential hotspots as denoted by the number of red and orange polygons – and *not* downtown DC.

Significantly, policies and individuals change human behavior during a pandemic. Accordingly, GeoGraph predictions are updated as data is ingested, ensuring ongoing relevance to real-time decision-making, and providing a feedback-loop on response effectiveness, including social distancing.

GeoGraph's Edge

A number of public and private organizations are repurposing mobile-location data (i.e., location data collected from smartphones, commonly used for mobile advertising purposes) to conduct contact tracing between COVID-19-positive individuals and their contacts. Like contact-tracing from the church in Washington, DC, such efforts aim to mitigate the virus' spread by identifying

and isolating those individuals deemed to be at-risk as a result of their direct relationship(s), over a certain timeframe, with other individuals who are known to be infected. This approach is – by definition – reactive and cannot predict how the virus may spread without observed contact. Moreover, given the sparsity of location data (due notably to reporting inconsistencies, disconnected data sources, and privacy concerns), and a reliance on pin-point accuracy, mobile-location tracing can yield only limited and narrow results.

SignalFrame's GeoGraph, in contrast, makes predictions on viral spread based on historical social priors between locations, combined with real-time data streams. It allows decision-makers to make real-time assessments on response targeting and resource allocation.

GeoGraph can:

1. Provide a broad risk-assessment to all locations:
 - *not limited to single tracked individuals,*
 - *not limited by the frequency of location tracking (location scans),*
 - *not limited to “co-observation” of different individuals to deliver actionable information.*
2. Update risk profiles in real-time with new information:
 - each confirmed case-location (and relevant available movement trajectories) *updates risk profiles for all other geographies,*
 - highlights *new potential hotspot geographies.*

Finally, GeoGraph's linkages can be strengthened using different data sources merged on geographic keys, including:

- *additional data elements attached to geographies,*
- *data from multiple sources*, including government, telecommunications, and media organizations.

Learn More

To learn more and get started today, contact SignalFrame at contact@signalframe.com.

Technical Appendix: GeoGraph's System Model

GeoGraph is a set of streaming computation steps implemented on SignalFrame's SignalGraph platform:

1. Transformation of mobile-location data into polygon descriptions.
2. Clustering of polygon descriptions into geo identities.
3. Computing similarities between geo identities.
4. Community detection over the induced GeoGraph.

These steps form a streaming pipeline, where a step's output is fed as the input to its successor. The final step pushes data into Web-based mapping user interfaces and other desktop-based

KML/GeoJson tools, or stores it and makes it available to real-time continuous queries. Each step is detailed below.

Step 1: Computing Polygon Descriptions

SignalGraph continuously ingests data from mobile devices as (adtech identifier, timestamp, wifi signals, bluetooth signals, [lat, lon]) tuples. Geographic coordinates (i.e., lat/lon pairs) are optional and can be inferred using the underlying SignalGraph platform.

Each input tuple is transformed into a (polygon_identifier, timestamp, adtech_identifier) tuple. Polygon identifiers are selected from a finite and mutable set of geojson polygon descriptors. If the given tuple's (possibly inferred) geographic coordinates fall within a defined polygon, then a corresponding polygon tuple is generated. Note that one adtech tuple may generate multiple polygon tuples.

SignalFrame processes tens of millions adtech tuples per hour. The platform accommodates additional data sources, such as social media streams, and is not limited to adtech data.

Step 2: Computing Geo Identities

Polygon tuples are grouped together using polygon_identifiers and time windows. A sliding window technique uses 12 or 24-hour collection windows, with a 12-hour shift. All collected tuples are merged into an *observation* vector, whose dimensions are the collected attributes (such as adtech identifiers). This follows similar encoding schemes found in NLP/IR systems.

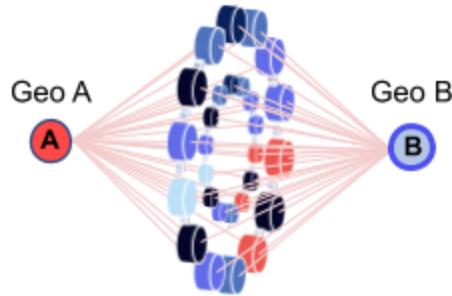


Figure 1: Geo identities are related to one another through the sum total of **all available attributes**.

A set of observation vectors, per polygon identifier, is collected according to (2 week, 1 week) sliding windows. Each set of vectors (per given time window) is clustered using SignalFrame's proprietary graph-embedding algorithm. In particular, the algorithm is designed to operate over infinite streams and vector dimensions (attributes). Common attributes are the basis for computing similarities between geo identities (see Figure 1 above).

Each timed cluster (i.e., embedding) is added to the cluster list of the respective polygon identifier. This set of clusters forms an identity. As these sets are monotonically increasing, stale clusters are regularly dropped. In the current implementation clusters older than 3 weeks

from the *current* timestamp (i.e., from now) are dropped. Each identity persists in a permanent graph database storage.

Step 3: Computing Similarities between Geo Identities

GeoGraph is induced over computed geo identities where edges are weighted using a similarity score between the identities (see Figure 2 below). Similarity scores are re-computed per polygon identifiers when the underlying cluster set (identity) is updated. Note that the computations are localized using a reverse index over clusters' dimensions.

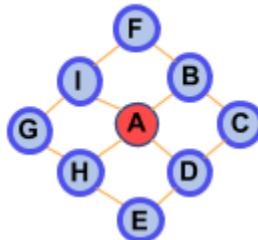


Figure 2: Induced GeoGraph from **similarities** between geo identities.

The similarity score reflects the underlying vectors' magnitudes, as well as the number of clusters. In short, similarity is higher if the underlying attributes are shared, and those attributes persist in an identity across time.

Step 4: Computing Communities of Geo Identities

A community of geo identities is defined as a set of geo identities that shares a large number of triangles between its members (a formal definition is available as part of detailed technical documentation). When edge changes are detected between identities, the communities' memberships are also updated accordingly.

Communities persist in a graph database (as part of the SignalGraph platform) and can be queried in real-time to surface up-to-date edges and neighbors.

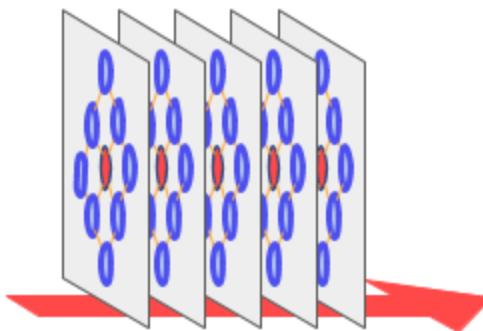


Figure 3: GeoGraph maintains **current** edges as **similarities** and identities change over time.