Enhancing spatial query results using semantics and multiplex networks

Manesh Pillai
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD 21250
USA
mpillai1@umbc.edu

George Karabatis
Department of Information Systems
University of Maryland, Baltimore County
Baltimore, MD 21250
USA
georgek@umbc.edu

ABSTRACT

Geographic information retrieval (GIR) research mainly focuses on finding relevant geographic artifacts along one layer, although answers to queries should preferably include artifacts across multiple layers. This paper addresses this limitation, by exploiting relationships that exist between geographical objects of the same or different types. Given the existence of these relationships we can infer additional indirect semantic relationships between geographical objects that otherwise would be ignored. This semantic information contains multiple dimensions. This research creates a mathematical model of semantically relevant geographic objects that belong to different layers using the construct of multiplex networks. The proposed model respects and extends the "layered" structure of geographic data, and enhances the results of queries made to a GIS. A prototype system has been implemented over an existing GIS relational database. Experiments with the prototype show significant improvements in the results of queries, measured using precision and recall.

CCS CONCEPTS

• **Information Systems** → Geographic Information Systems • **Theory of Computation** → Dynamic graph algorithms

KEYWORDS

Geographic Information Retrieval; Multiplex Networks; Semantic Link Networks

1 INTRODUCTION

While geographic information retrieval (GIR) models have been proposed as a viable answer to the "what is relevant to the location of interest" question, there are several problems with current GIR methods. First, GIR research remains focused on finding relevant geographic artifacts within a single layer (a database table with associated geographic objects) when such queries should include geographic artifacts across multiple layers or tables.

For example, a local government council discusses pending legislation affecting a neighborhood. During this discussion one council member remembers that there was previous legislation affecting this neighborhood from a few years ago but cannot remember the legislation number. In this scenario, a method to query related heterogeneous data (legislation, zoning, streets) would improve productivity in a way that most effectively provides that council member with relevant answers.

This search for relevant information is an example of geographic information retrieval. It showcases the desire to retrieve relevant legislation from the perspective of a user investigating a neighborhood and its history. Artifacts from other layers, such as nearby parks or zoning ordinances could also be relevant and desirable to be included in the query results.

Currently, the most common type of geographic search involves one layer and one metric, a generalized "similarity" measure, e.g. finding relevant legislation that affects the streets near the user's location. While distance is relevant, it is not the only metric that should be important for this type of query. Other relevant dimensions for legislation that could affect relevance could be the year the legislation was passed or the votes on that legislation.

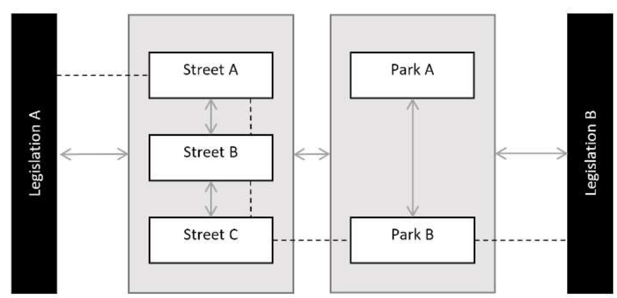

Figure 1: An example of how two pieces of legislation can be related through nearby relevant geographic artifacts.

In Figure 1 we show two pieces of legislation and two types of geographic objects: streets and parks. While legislation is not directly part of a GIS the text of a legislation can include keywords that reference a location in a GIS such as a street or a park. Relationships between types of objects (legislation, streets and parks) are shown as horizontal gray arrows. Relationships between objects of the same class (for example, two streets near each other) are represented by vertical arrows. Given the existence of these relationships we can infer a relationship between the two pieces of legislation by chaining these relationships, via a black dotted line. Legislation A references street A and legislation B references park B. Street A is close to street C and street C is adjacent to park B. By chaining these relationships, we infer a tangential relationship between Legislation A and Legislation B.

In addition to possibly incorporating multiple metrics describing the relationships between any two objects, GIR research generally assumes that the user only wishes to find relevant geographic artifacts in one table when there could be data in multiple tables and databases relevant to the query. For example, in addition to searching for relevant legislation, the council might be interested in construction projects, zoning ordinances or permits issued in that neighborhood.

The research question this work addresses is: How can we craft a GIR model to incorporate these new dimensions (multiple metrics and multiple tables)? Our proposed solution takes into consideration recent research into multiplex networks [1, 2] and semantic link networks (SLNs).

A **multiplex network** is a set of networks (or layers) containing objects connected across multiple dimensions. Objects could include neighborhoods, streets or parks and the connections between them could reflect a similarity (or relevance) between them. In many cases, a transitive relationship between objects may exist. For example, when compared, two pieces of legislation might not appear to be similar but a shared reference to a geographic location (such as a neighborhood) indirectly links them together. **Semantic link networks** (SLNs) exploit these transitive relationships. The contributions of this paper are:

- Creation of a mathematical model of geographic relevance based on the properties of multiplex networks and semantic link networks.
- Enhancement of GIS query results with semantically relevant artifacts that provide additional related information.
- Implementation of the model in a prototype system over a relational database.
- Evaluation of the implemented model through a series of experiments with public GIS datasets measured through precision and recall.

The paper is organized as follows: Section 2 provides an overview of related work, Section 3 outlines our approach, Section 4 describes our implementation with results and Section 5 provides our conclusions.

2 RELATED WORK

This research aims to use the mathematical principles behind multiplex networks and semantic link networks to create a system for geographic information retrieval. In this section, we first explore previous research into GIR and its applications and then we discuss the development of multiplex networks and semantic link networks.

2.1 Geographic Information Retrieval

Geographic Information Retrieval (GIR) can be defined as the relationship between a user's geographic needs and the spatio-temporal expression of geographic objects in the user's surrounding environment [3, 4]. A user's needs can be defined in many ways, such as the nearest object to a location or the boundary containing a location. In the GIS field these are known as topological constraints [5]. Examples of topological constraints include:

- *Boundary constraints* – What boundaries are relevant to the location of interest? For example, the school district containing a house location.
- *Nearest neighbor* – Finding the nearest (by distance) geographic object to the location of interest. For example, the park nearest to the user's location. This type of metric can be complicated by restrictions on how the distance is calculated. For example, if a park can only be accessed by a road network then we must travel along that network to determine the distance to that park. Also, if we want to prioritize geographic artifacts by distance, how does this translate into a relevancy score? [4]
- *Adjacency* – If the location of interest is a line or polygon then we might be interested in adjacent geometries. For example, if commercial zoning is adjacent to residential zoning.

Geographic Information Retrieval research focuses on two areas. The first is the acknowledgement that an increasing amount of computing happens on mobile devices such as phones and that given their limited screen real estate it is expedient to focus the output of mobile computing systems on geographically relevant results [6]. The second is an emphasis on identifying geographically relevant artifacts from text corpora [7].

Many GIR queries come in the form of a triplet, <what, relation, where>, where "what" and "where" are geographic objects with a relationship between the two [8]. There are two types of relationships we can define of this type: relationships between classes of objects and relationships between objects of the same class. In GIR we define relationships between classes of objects by establishing an ontology on those classes. For example, we can define the class "address point" to be contained within "school boundary." Within an individual layer, we can define geographic relationships between them. For example, two addresses can have a relationship defined by the relative distance between them.

There are several limitations with current GIR techniques which we hope to address: First, most techniques define web-based frameworks (like XML) to define geographic relationships and do not apply a rigorous mathematical framework. Second, while these types of XML-based querying systems could be used to imply transitive relationships, these transitive geographic relationships could use a more formal mathematical definition. We hope that we can use the research on both multiplex networks and semantic link networks to address these two issues. We discuss the background of both in the next sections.

2.2 Multiplex Networks

Network or graph theory has attempted to model natural (or real-world) systems as a network of interconnected nodes. This type of modeling has produced many theories as to the properties and organizational principles of these networks. One theory includes the premise that even randomly generated graphs have certain observable properties that mimic those in the natural world [9, 10].

The net result of this research was that networks themselves, if their purpose is to model real-world properties of nature, must become less homogeneous (uniform in structure) and more adaptable to the flexible and dynamic structure of real-world systems [11].

Graph theory generally describes objects that have only one relationship between them, but this approach becomes limited when there could be multiple relationships between the same nodes [12]. Specifically, this limitation is an over-simplification of the natural complexity of these systems which could lead to misleading results [13]. Because of this, we examine the properties of a multiplex network, or a network of networks where each network is interactive and interdependent with each other.

The applications of multiplex networks have been investigated in many research domains including infrastructure networks [14], social networks [15-17], transportation [18, 19], and biology [20].

A **multiplex network** consists of a series of individual networks (layers) and a set of relationships between those layers [21]. Each **layer** represents a distinct interaction between the nodes on the network [22]. A geospatial example could be two coordinates that share a similar political boundary as well as a similar elevation. Attempting to aggregate these similarities into one measure would be problematic because how does one create one singular measure to determine the similarity between the location of interest and a park or a fire station? In Figure 2 we have a multiplex network with three layers. If M represents the number of layers in a multiplex network, then in this example M = 3. Each layer can be thought of as an individual network (or graph) where a node represents an object (such as a street or a park) and an edge between the nodes represents a relationship between those nodes. In each layer, there are a total of five nodes (N = 5) and these five nodes exist in each of the layers.

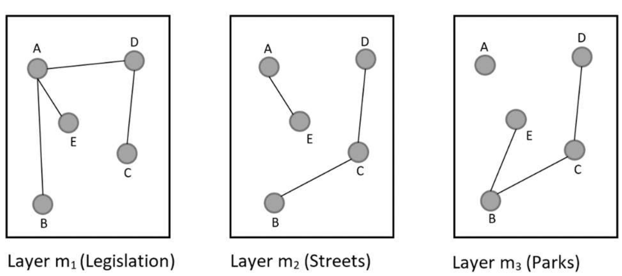

Figure 2: An example of a multiplex network. Each layer m_i can be represented as an individual graph.

In a multiplex network, given N nodes we can represent the complete set of nodes in our multiplex as:

$$\vec{N} = (n_1, \dots, n_N)$$

Given M layers we can represent the complete set of layers in our multiplex as:

$$\vec{M} = (m_1, \dots, m_M)$$

Representing three different layers $m_1 \dots m_3$ as isolated graphs presents problems when attempting to fully model the connectivity of each node. For example, if we want to examine the relationship between node A in layer m_3 we are stuck because that node has no connectivity in that layer (it has degree = 0). We need to properly model the interconnected nature of these layers. That is, if there is a relationship between the layers we need a place to represent that. Therefore, we look at the **inter-layer correlations** between the layers [18].

We treat changing layers as a step in our network with its own dedicated edge and a value (known as a penalty) for traversing from one layer to another [23]. We assign a penalty because switching layers represents an indirect relationship between two objects. For example, if two roads are not directly related to each other (they are distant) but they share identical zoning then the two roads are related but not as strongly as if the roads themselves were adjacent to each other.

To reduce the complexity of our multiplex construction we choose to only allow travel between layers at the same node. Therefore, we institute a rule where we only allow travel between nodes on the same layer or between layers at the same node [24].

2.3 Semantic Link Networks

We would like to examine the transitive relationships between geographic objects. In our motivating example, we examine the relationship between two pieces of legislation that would not normally be directly related. We inferred a relationship between them because the first piece of legislation described a street and the second piece of legislation described a park and the street and park were geographically adjacent to each other.

A **Semantic Link Network (SLN)** is defined as "a self-organized semantic data model for semantically organizing resources, which can be abstract concepts or specific entities such as texts, images, videos, and audios" [25]. SLNs have been used in network applications, knowledge management and to enhance search [26].

SLNs are mathematically defined as adjacency matrices where the values of the matrix cells represent the strength of the relationship between two objects. SLNs are augmented with mathematical reasoning rules to define new linkages between objects that might not have existed otherwise.

2.4 Multiplex Semantic Link Networks

We combine the mathematical and intuitive properties of multiplex networks and semantic link networks to create a multiplex semantic link network (mSLN). We have previously applied the principles of mSLNs in the field of cybersecurity [27].

Our purpose is to generate a multiplex semantic link network and use this to create linkages between geographic objects to aid in querying those objects.

3 RESEARCH APPROACH

Our purpose is to create a multiplex model of geographic relationships to aid in geographic information retrieval. As Figure 3 illustrates, we generate a model of geographic relevance and then apply this model to incoming queries to provide additional and quite relevant geographic artifacts.

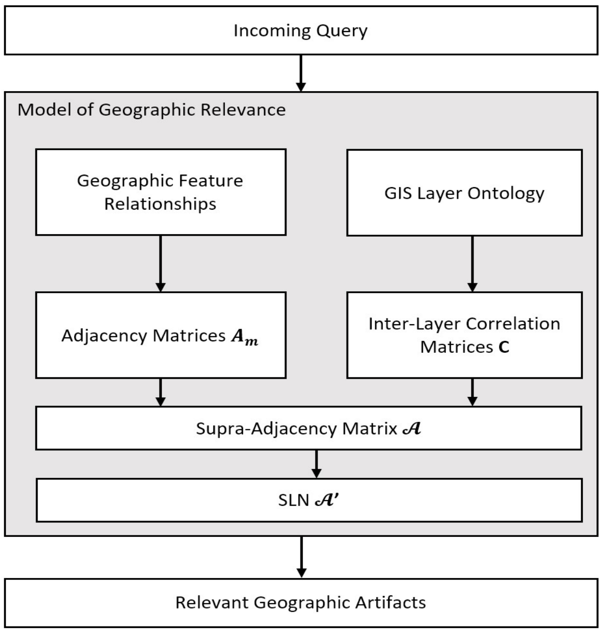

Figure 3: Workflow to create a geographic information retrieval model using mSLN.

3.1 Creation of a Static Model

Our multiplex model is a static model, in that the model does not change based on incoming queries. Given a set of layers and geographic objects in those layers we examine the relationships between those layers and between the objects in those layers. We assign probabilities that indicate relationships between those objects to generate a multiplex construction and then use the properties of SLNs to evaluate indirect (or transitive) relationships between those objects, resulting in the static model. These steps are detailed in this section.

GIS Layer Ontology. We first need to determine how different GIS layers interact with each other. How is a specific address related to a park, a fire station or a school district? How are school districts related to the area's designated zoning? To answer these questions, we first need to have an ontology to assist in forming the answers.

An **ontology** describes the relationships between different classes of objects. For example, there is a relationship between schools and streets in that schools are adjacent to streets. Geospatial ontologies have been explored in [28-31].

There are various methods to create geographic ontologies and we assume that the ontologies have already been developed by experts.

Inter-Layer Correlation Matrices. When a relationship is defined between two classes of objects we define that relationship between them as a probability value representing the strength of the relationship, $\omega \in [0,1]$.

For example, we might determine that the relationship between school districts and fire stations is weak and we might assign it a lower $\omega = 0.1$. However, the relationship between an agricultural preservation district and the zoning of an area might be high since areas which are agricultural preservation districts have a higher propensity to be in an area with rural zoning, so we might assign it a higher $\omega = 0.8$.

We use these ω values to populate the values in our inter-layer correlation matrices:

$$\boldsymbol{C}_{ab} = \begin{pmatrix} \omega_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \omega_N \end{pmatrix}$$

where $\omega_n \mid n \in N$ is the penalty of moving from layer a to layer b at node n. $\boldsymbol{C}_{ab}$ is a diagonal matrix because we are only interested in relationships between layers at the same node (or location).

Geographic Feature Relationships. Given an individual layer and a location of interest, we need to determine which features on that layer are the most relevant. For example, if we are have identified a point of interest and we have determined that streets are relevant to our query then we need a way to determine which streets are relevant to our point of interest. We create probability values that represent the relevance of each feature on the layer to the location of interest. For example, we might find streets closer to our point of interest more relevant than streets farther away and assign a higher probability value to those streets. The probability values form the layer adjacency matrices $\boldsymbol{A_m}$.

$$\mathrm{A_m} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

An element a_{ij} of $\boldsymbol{A_m}$ represents the relationship of geographic object i to object j as a decimal number $\in [0,1]$.

We use a similar set of relationships to those used when developing the ontology. That is, we can use intersections, adjacencies and distances to determine the relative significance (or relevance) of geographic objects within one class to others of the same class. So, if we are looking at parks, we can say that parks that are closer to us are more important than parks that are farther away.

We define our significance $a_{ij} \in [0,1]$ where i and j are individual geographic objects in our environment. A value of 0 represents no relationship between the two objects and a value of 1 represents the strongest possible relationship between any two geographic objects.

Supra-Adjacency Matrix. Now that we have the values needed to populate both our adjacency matrices $\boldsymbol{A_m}$ and our inter-layer correlation matrices $\boldsymbol{C}$ (to define the relationships of geographic objects of different classes), we need to combine them to form a supra-adjacency matrix. This matrix defines the relationships between a given geographic object and all objects in all layers, not just the layer containing the geographic object. We use this matrix to expand our search results to objects in all layers of interest.

We define our supra-adjacency matrix as follows:

$$\mathcal{A} = \begin{pmatrix} \boldsymbol{A_1} & \boldsymbol{C_{12}} & \dots & \boldsymbol{C_{1M}} \\ \boldsymbol{C_{21}} & \boldsymbol{C_{22}} & \cdots & \boldsymbol{C_{2M}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C_{M1}} & \boldsymbol{C_{M2}} & \dots & \boldsymbol{A_M} \end{pmatrix}$$

The supra-adjacency matrix $\mathcal{A}$ is itself an adjacency matrix, whose elements are matrices themselves. $\mathcal{A}$ is an M x M matrix where each element is an n x n matrix making $\mathcal{A}$ an (NxM) x (NxM) matrix. We next need to generate an SLN from the supra-adjacency matrix.

Initial SLN. Given our supra-adjacency matrix defining direct relationships (as probabilities) between geographic objects we need to discover indirect relationships between these objects. That is, if park A and park B are related and if park B is related to street C, we need to define the relationship between park A and street C. We do this by preparing a semantic link network (SLN) that represents the relationship of geographic objects in our environment. This process is described in the procedure outlined in [2].

A single row in our supra-adjacency matrix represents the strength of the relationships between a geographic object (representing one row in the matrix) and all other geographic objects in all layers. We need to convert the values in $\mathcal{A}$ into values that represent the propensity of connecting that one geographic object to every other geographic object in our system by normalizing them to produce a right stochastic matrix $\mathcal{A}'$. If $\mathcal{A}$ is a p x p matrix (where p = N x M):

$$\mathcal{A'}_{ij} = \frac{\mathcal{A}_{ij}}{\sum_{m=1}^{p} \mathcal{A}_{im}}$$

$\mathcal{A}'$ is our initial SLN and represents a Markov chain describing the probability that a given geographic object is relevant to the geographic object of interest. Given park A, this represents the probabilities that any other object in our system is relevant to park A. The initial SLN $\mathcal{A'}_{ij}$ represents the probability of moving from node i to node j. It is referred to as an initial SLN because it only reflects direct relationships between geographic objects.

Reasoning Rules. Once we have our initial SLN $\mathcal{A}'$ we need to apply reasoning rules to derive new transitive relationships between geographic objects that might not been apparent before this application. We describe a reasoning rule as follows:

$$n_i \xrightarrow{\alpha} n_j, n_j \xrightarrow{\omega} n_{j*}, n_{j*} \xrightarrow{\beta} n_{k*} \Longrightarrow n_i \xrightarrow{\gamma} n_{k*}$$
$$\mid \alpha, \beta, \omega, \gamma \text{ are weights on links and } \alpha \cdot \omega \cdot \beta = \gamma$$

This describes the relationship between two objects in two different layers. As Figure 4 illustrates, if Street A and Street B have a relationship value α, Park A and Park B have a relationship β, and the classes "streets" and "parks" have a relationship ω then we can imply a relationship γ between Street A and Park A. Since these relationships are represented by probability values between 0 and 1, we multiply these probabilities to generate a new probability between the street and the park.

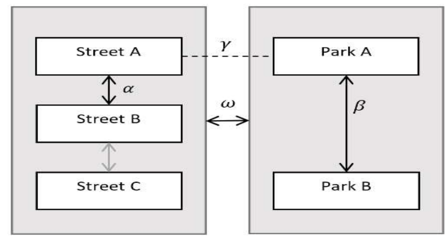

Figure 4: The Multiplex SLN Schema applied to our motivating example.

If this derived probability is higher than the original probability, then we replace the link's probability with the derived one. If we describe the relationship between two objects of two different classes then this probability would be initially zero. That is, there would be no direct relationship between a given street and a given park before reasoning rules are applied.

The result of these calculations is an adjacency matrix, called a multiplex SLN, that represents the optimized relationships (after taking potential indirect relationships into account) between all geographic objects across all layers. A **Multiplex Semantic Link Network (mSLN)** for a system can be defined as a graph representation of the semantic relationships between its nodes (representing objects) among several different layers.

Now that we have established relationships between layers and between objects in those layers to construct an mSLN, we use this model to enhance incoming queries.

3.2 Processing Incoming Queries

Once the static model is generated, we can start using it and submit queries that take advantage of these extra relationships that have been formed and saved in our model. When we perform a query, we want to discover related objects. For example, when we submit a query on a piece of legislation, we might want to find out what relevant streets and parks are related to that legislation. We might also want to identify other pieces of legislation that share a similar geography.

These types of queries take an incoming object (such as legislation), examine the text of that legislation to extract geographic keywords (references to streets, parks, zones, etc.) and uses the optimized mSLN to identify relevant objects. We process these queries as follows:

- The subject of our query already has a layer associated with it as well as an object ID within that layer. For example, "Deep Run Park" belongs to the "parks" layer and has a unique identifier within that layer.

- Every object in the mSLN identifies a row in it. There would be a row in the mSLN for "Deep Run Park."
- Given this row we identify relevant objects based on the highest probability of column values on that row. These probabilities have already been calculated when the model was generated.

Identifying relevant objects for a query becomes a simple linear lookup on the model. The number of results that are returned is determined by a threshold ϴ. For example, if ϴ = 0.6, we only return relevant geographic objects with a calculated probability higher than 0.6. This threshold is manually chosen to balance returning meaningful results from the query and to avoid returning too many irrelevant query results.

4 IMPLEMENTATION AND RESULTS

We use publicly accessible data from Howard County, Maryland for the analysis of our results. Following our motivating example we look at legislation passed by the county government to identify legislation that might be related based on the geographic locations they describe.

4.1 Datasets

We use the following datasets to construct our mSLN: A dataset with legislation related information, another one with zoning information and finally a third dataset containing street segments (see Table 1).

Table 1: Description of Datasets used to construct mSLN

Dataset Name	Dataset Size
Legislation	154 pieces of legislation
Zoning	531 zone polygons
Streets	12,139 local street segments

Following our motivating example, the legislation is one of our test datasets. This includes legislation (government laws such as council bills and resolutions) from January 2006 to May 2017 that includes a reference to a street in its description. The two records in Table 2 show two pieces of legislation that reference streets in their short description. They both reference a street closing. Given the 154 pieces of legislation in our original dataset 38 of them were tagged as sharing a common theme. These themes included closing streets, agricultural preservation and zoning issues.

Table 2: Sample rows from legislation dataset with GIS keywords.

Leg. Number	GIS Keyword	Short Description
CR35-2017	FOREST AVE	A RESOLUTION to close a portion of Forest Avenue…
CR1-2017	WINTER THICKET RD	A RESOLUTION to close all of Winter Thicket Road…

This data can be accessed at https://cc.howardcountymd.gov. The GIS datasets can be downloaded at https://data.howardcountymd.gov.

4.2 Submitting queries

For this implementation, we create a scenario where a user, looking at a piece of legislation, is interested in legislation geographically related to it. The query takes a single piece of legislation as input, looks at its related geographic properties (both the street associated with the legislation and zoning associated with the property associated with the legislation) and compares it to the geographic properties of other legislation, and generates related legislation as output.

For this scenario, our static geographic model contains zones and streets and we apply the following rules:

- **Defining Search Radii** – We vary the search radius of datasets to examine the quantity and quality of search results as well as the processing time involved. Given a geographic feature (like a street) we could compare it to every other street in the dataset but since we are only interested in local relationships it does not make as much sense to compare every street in the dataset. Therefore, we define fixed search radii to limit the scope of our comparisons.
- **Relationships Within datasets** - We assign a similarity score between any two geographic objects in the model as a linear relationship based on the distance between the two objects compared proportionally to the search radius used. If two streets are 5 miles away from each other and our search radius is 5 miles then the two streets are assigned a similarity score of 0 and if two parks are adjacent to each other then they are assigned a similarity score of 1. For both zones (which are polygon objects) and streets (which are line objects) we use the centroids of the objects to calculate similarity.
- **Relationships Between datasets** – In this simplified example we only have two geographic datasets: zones and streets. To simplify our model the inter-layer correlation between the two datasets ω is set to 1.

4.2 Results

We tested two scenarios: One comparing legislation using a single layer (streets) and another using two layers, the first representing streets and the second layer representing zoning information (forming a multiplex construction with two layers). With 38 legislations tagged with streets as GIS keywords (see Table 2) there are 1,444 possible relationships formed. Of those, 96 were manually tagged as having reasonably legitimate relationships. Those 96 relationships represent the ground truth of our scenarios.

As a baseline, we attempted to identify relevant legislation based on the streets identified in their text alone. We varied the search radius and the threshold ϴ limiting the scope of the results that would return for any given legislation.

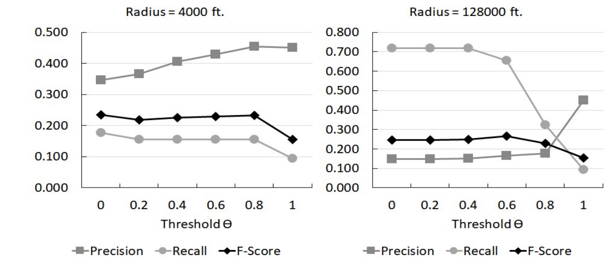

Figure 5: Results of single-layer (streets) with varying radii and threshold ϴ.

As Figure 5 shows, increasing the search radius or decreasing the threshold predictably increases recall at the expense of precision. However, increasing the search radius increases the processing time. An increase in radius from 4000 ft. to 128000 ft. increases the average number of streets to be compared 70-fold. While we could establish a relationship between any piece of legislation and another by comparing every street countywide this could become unwieldy at big data scale.

We then submitted the same query using two layers, streets and zoning, to identify relevant legislation that would not be identifiable otherwise. For example, two pieces of legislation might be related because they both refer to agricultural preservation and a comparison of street proximity alone might not reveal this relationship. However, since agricultural preservation only applies to properties that are agriculturally zoned we can infer this relationship.

Legislation is tagged with streets and streets are related to zoning in that a street segment is contained inside a zone. We choose a search radius of 4000 ft. on the zoning layer since for zoning, unlike streets, we are not as interested in distance.

As an example of the benefits of using two layers we focus on an individual piece of legislation. The legislation with number "CR15-2013" describes an effort by Howard County to purchase farmland for agricultural preservation. Using streets alone, the only two matches the query returns with a threshold ϴ > 0.8 are "CR67-2015" which authorizes the county to sell property it owns and "CR64-2009" which describes issuing municipal bonds. Neither legislation reference agricultural preservation. When we use two layers (adding zoning) with the same threshold the query returns 5 additional matches that also relate to agricultural preservation.

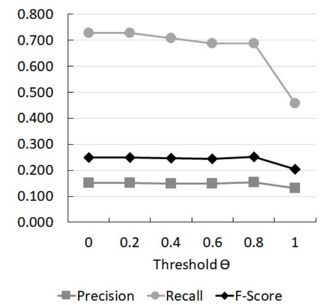

Figure 6: Two-layer (streets and zoning) results with varying threshold.

Adding two layers increases the total number of correct responses returned as reflected in a larger recall score shown in Figure 6. However, this comes at the expense of precision, like the single-layer case with an expanded search radius. One key difference is in the number of calculations. Calculating similarity for thousands of street segments is more computationally intensive than calculating similarity for hundreds of zoning polygons. The reduced amount of computation provides a boost to the performance of the system while at the same time it does not sacrifice any relevant information from the results.

For example, one of the first steps is to calculate the similarity between the geographic objects linked to the legislation. As shown Table 3, if we compare streets, even at a radius of 4000 ft. with our dataset we must make about 2.3 million comparisons. This increases to a maximum of about 147 million comparisons to compute the similarity between all 12,139 segments. Performing the comparison with zoning with a radius of 4000 ft. only requires about 12,000 comparisons. Finding the zone associated with a street in a standard geospatial database is a trivial calculation assuming both tables have proper spatial indexes, but calculating the similarity for both tables is a brute force calculation.

Table 3: Effect of radius on similarity calculations required.

Radius	Street Comparisons	Zoning Comparisons
4000 ft.	~2.3 million	12,119
16000 ft.	~24.6 million	61,545
64000 ft.	~132 million	257,193
128000 ft.	~147 million	281,693

5 CONCLUSION

This paper uses the concepts of multiplex networks and semantic link networks to create a mathematical model of geographic relevance. Based on our results this model succeeded in expanding the search results returned by a geographic query (increased recall) while reducing the amount of computation required to do so.

However, this increase in accuracy and computational performance was accomplished at the expense of precision. While the system returned more relevant results to the user it also returned more irrelevant results. A possible avenue for future research would be to refine the query results to isolate relevant results from irrelevant results, improving precision. Another area of future focus would be the computational performance. One of the reasons for utilizing zoning in our experiment was that it helped bypass computing similarity for over 11,000 street segments. However, there might be situations where this might not be possible, like big data environments. Therefore, future research might involve examining the query radius even further to more quickly generate relevant results.

References

[1] Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. and Porter, M. A. Multilayer Networks. *arXiv.org*, physics.soc-ph (Sep 27 2013).

[2] Aleroud, A. Contextual information fusion for the detection of cyber-attacks. (2014).

[3] Raper, J. Geographic relevance [information retrieval systems]. *Journal of Documentation*, 63, 6 (2007), 836-852.

[4] De Sabbata, S. and Reichenbacher, T. *A probabilistic model of geographic relevance*. ACM, New York, New York, USA, 2010.

[5] Zaila, Y. L. and Montesi, D. Geographic information extraction, disambiguation and ranking techniques. In *Proceedings of the Proceedings of the 9th Workshop on Geographic Information Retrieval - GIR '15* (Paris, France, 2015).

[6] Reichenbacher, T. and De Sabbata, S. Geographic relevance - different notions of geographies and relevancies. *SIGSPATIAL Special*, 3, 2 (2011), 67-70.

[7] Wallgrün, J. O., Klippel, A. and Karimzadeh, M. Towards contextualized models of spatial relations. In *Proceedings of the Proceedings of the 9th Workshop on Geographic ...* (2015).

[8] Mata, F. Geographic Information Retrieval by Topological, Geographical, and Conceptual Matching. *GeoS*, 4853, Chapter 7 (2007), 98-113.

[9] Barabasi, A. L. and Albert, R. Emergence of scaling in random networks. *Science*, 286, 5439 (Oct 15 1999), 509-512.

[10] Dorogovtsev, S. N., Goltsev, A. V. and Mendes, J. F. F. Critical phenomena in complex networks. *arXiv.org*, cond-mat.stat-mech, 4 (Apr 30 2007), 1275-1335.

[11] Kenett, D. Y., Gao, J., Huang, X., Shao, S., Vodenska, I., Buldyrev, S. V., Paul, G., Stanley, H. E. and Havlin, S. Network of Interdependent Networks: Overview of Theory and Applications. *Networks of Networks: The Last Frontier of Complexity*, Chapter 1 (2014), 3-.

[12] Havlin, S., Kenett, D. Y., Ben-Jacob, E., Bunde, A., Cohen, R., Hermann, H., Kantelhardt, J. W., Kertész, J., Kirkpatrick, S., Kurths, J., Portugali, J. and Solomon, S. Challenges in network science: Applications to infrastructures, climate, social systems and economics. *The European Physical Journal Special Topics*, 214 (Nov 2012), 273-293.

[13] De Domenico, M., Sole-Ribalta, A., Omodei, E., Gomez, S. and Arenas, A. Centrality in Interconnected Multilayer Networks. *arXiv.org*, physics.soc-ph (Nov 12 2013), 6868.

[14] Min, B., Do Yi, S., Lee, K.-M. and Goh, K. I. Network robustness of multiplex networks with interlayer degree correlations. *arXiv.org*, physics.soc-ph (Jul 04 2013).

[15] Sole-Ribalta, A., De Domenico, M., Gomez, S. and Arenas, A. Centrality rankings in multiplex networks. In *Proceedings of the WebSci '14: Proceedings of the 2014 ACM conference on Web science* (Jun, 2014).

[16] Estrada, E. and Higham, D. J. Network Properties Revealed through Matrix Functions. *SIAM review*, 52, 4 (2010), 696-714.

[17] Horvat, E.-A. and Zweig, K. A. One-mode Projection of Multiplex Bipartite Graphs. In *Proceedings of the Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on* (2012).

[18] Estrada, E. and Gómez-Gardeñes, J. Communicability reveals a transition to coordinated behavior in multiplex networks. *Physical Review E*, 89, 4 (Apr 30 2014), 042819.

[19] Kurant, M. and Thiran, P. Layered Complex Networks. *Physical Review Letters*, 96, 1 (Apr 2006), 138701.

[20] De Domenico, M., Nicosia, V., Arenas, A. and Latora, V. Layer aggregation and reducibility of multilayer interconnected networks. *CoRR abs/1409.5253*, physics.soc-ph (2014), 6864.

[21] Dorogovtsev, S. N., Mendes, J. F. F., Samukhin, A. N. and Zyuzin, A. Y. Organization of modular networks. *arXiv.org*, cond-mat.stat-mech, 5 (Mar 24 2008), 056106.

[22] De Domenico, M., Sole-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., Gomez, S. and Arenas, A. Mathematical Formulation of Multi-Layer Networks. *arXiv.org*, physics.soc-ph, 4 (Jul 18 2013), 041022.

[23] Min, B., Gwak, S. H., Lee, N. and Goh, K. I. Layer-switching cost and optimality in information spreading on multiplex networks. *Scientific reports*, 6 (2016), 21392.

[24] De Domenico, M., Sole, A., Gomez, S. and Arenas, A. Random Walks on Multiplex Networks. *arXiv.org*, physics.soc-ph, 23 (Jun 03 2013), 8351-8356.

[25] Zhuge, H. Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning. *Knowledge and Data Engineering, IEEE Transactions on*, 21, 6 (2009), 785-799.

[26] Pons, A. P. Object prefetching using semantic links. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 37, 1 (Jan 31 2006), 97-109.

[27] Pillai, M. and Karabatis, G. *Using Multiplex Networks to Model Cybersecurity Attack Profiles*. Springer International Publishing, City, 2016.

[28] Stock, K. and Cialone, C. *An Approach to the Management of Multiple Aligned Multilingual Ontologies for a Geospatial Earth Observation System*. Springer, Berlin, Heidelberg, City, 2011.

[29] Sen, S. *Two Types of Hierarchies in Geospatial Ontologies*. Springer, Berlin, Heidelberg, City, 2007.

[30] Third, A., Bennett, B. and Mallenby, D. *Architecture for a Grounded Ontology of Geographic Information*. Springer, Berlin, Heidelberg, City, 2007.

[31] Torres, M., Quintero, R., Moreno, M. and Fonseca, F. *Ontology-Driven Description of Spatial Data for Their Semantic Processing*. Springer, Berlin, Heidelberg, City, 2005.