A Framework for the Development of Digital Spatial Weights using Social Media Data: Methodological Foundation and Initial Analysis

Erica Erin E. Elazegui[1], Ariel C. Blanco[1], Bienvenido G. Carcellar III[1]

[1]Department of Geodetic Engineering, University of the Philippines Diliman, Quezon City 1101 Philippines
(eeelazegui, acblanco, bgcarcellar)@up.edu.ph

Keywords: Digital spatial weights, Social media data, Twitter data analysis, Spatial weight matrices, Geospatial analysis.

Abstract

The rapid growth of internet users has significantly impacted human activities, prompting a re-examination of spatial definitions in digital contexts. This study provides a framework for representing space digitally using social media data (Twitter, now X) and compares digital space representations with traditional physical space representations. Spatial weight matrices were developed to represent both physical and digital spaces in selected cities in Metro Manila and Quezon City (QC) barangays. Four digital spatial weight schemas were created using Twitter data to facilitate this comparison. Digital and physical spatial weights were then compared through matrix and weight maps. The results show that digital and physical spaces represent distinct spatial realities. Although exploratory in nature, the study highlights the temporal and spatial effects of using digital spatial weights compared to physical spatial weights. A key advantage of digital space representation is that it bypasses the disconnected spatial units issue present in most common physical representations, and temporal variations can be incorporated. This allows for more subtle relationships to be captured in spatial and temporal analysis. Since most phenomena occur in both physical and digital spaces, a combined representation of both may provide a more accurate model. Overall, the comparisons reveal the differences between digital and physical spaces in terms of matrix structures and weight maps. It is recommended to further explore spatial weights in the context of modeling implementations and the integration of both spatial representations.

1. Introduction

1.1 Background of the Study

Traditional spatial analysis relies on *physical space* concepts measured by Euclidean distance and transportation networks (Janc, 2015). However, the rapid expansion of Internet access has created new patterns of human activities online (Kemp, 2022). The Philippines exemplifies this transformation, becoming the "social media capital of the world" by 2021 (Lopez, 2021). This shift enabled humans to conduct various activities online, altering analyses of distance and proximity to consider digital human activity effects.

As Zook et al. (2011) observed, "Geography and technology shape each other in different ways," necessitating evolution in spatial analysis to include *digital space* representations. This digital space depicts space in the context of digital presence or online activities. This depiction of space will be helpful for urban planning and smart cities, public health management, general resource allocation and management, and sustainable mobility planning. While some studies have explored digital spatial relationships through social media and mobile data, there remains a significant gap in systematically comparing digital and physical space measures using spatial weight matrices.

In geospatial statistics, spatial relationships are commonly represented by spatial weights matrices (Getis and Aldstadt, 2004). These matrices, originally based on binary contiguity and distance-based measures, have evolved to capture potential interactions between spatial units. There are several variations in the computation of the spatial weights varying from perimeter contiguity, consideration of area, accessibility, based on social network theory, or combinations of all mentioned. The choice of spatial weight matrix should align with the specific spatial analysis being conducted. Given the complexity of selecting appropriate weight matrices, Anselin (1988) emphasized that researchers must clearly present their matrix specifications as constraints and explain how these specifications support their spatial analysis.

1.2 Research Objectives

This research developed a framework for creating digital spatial weights from social media data. Particularly, this study aims to: (1) acquire and analyze Twitter data, applying various constraints and determining the geospatial location of each data point, (2) develop digital spatial weight matrices for selected Metro Manila cities and Quezon City barangays using different periods of time, and (3) demonstrate the framework's application through comparative analysis with traditional physical spatial weights by comparing the weights matrices and maps.

The physical and digital spatial weights developed were used to produce weights matrix maps and spatial weights maps of administrative units for comparison. Three matrix constraints options were analyzed for both physical and digital weights: row-scaling, diagonal inclusion in scaling, and diagonal exclusion in final matrix. The choice between these configurations depends on the specific spatial analysis application and the desired interpretation of self-influence versus neighbor influence of the weights.

1.3 Scope and Limitations of the Study

This framework focuses on geotagged tweets from seven Metro Manila cities and 142 Quezon City barangays from January 2019 to June 2021, spanning pre-COVID and lockdown periods to capture temporal variability. Geotagged tweets are representative of population mobility patterns despite character restrictions and location inconsistencies (Jurdak et al., 2015). The Twitter API (Twitter, Inc, 2023) allows access only to public data with no private information included. API access was revoked in July 2023, limiting complete coverage to the cities shown in Figure 1.

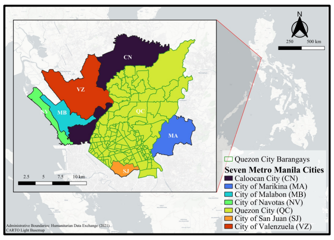

Figure 1. Seven cities of Metro Manila and 142 Quezon City barangays with complete data. Extracted from the Philippine Administrative Boundaries (United Nations, 2021).[1]

Several reports precaution that the social media user numbers in the Philippines do not show unique users as an individual may have multiple accounts. Nevertheless, social media data is an abundant source of information and extensive across areas (Kuchler et al., 2022), the difficulty in its usage lies in the pre-processing, access to data, and identification of restrictions.

No specific preprocessing was implemented to correct for known Twitter data biases such as demographic representativeness, bot activity, or multiple accounts per user. Our approach instead leveraged temporal averaging across study periods to minimize the influence of outliers and sporadic activity, and employed multiple analytical frameworks (tweet-based, user-based, and hybrid measures) to provide different perspectives on the same spatial phenomena.

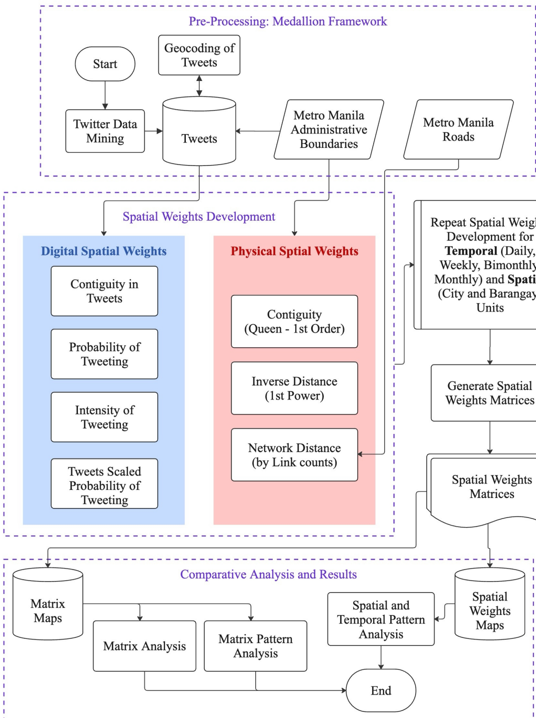

Figure 2. Three-phase approach: data pre-processing using medallion framework, parallel development of digital and physical spatial weights, and comparative analysis.[1]

An important limitation, in any research involving digital media, is the difficulty of recommending future technologies, methodologies, and the verification of results over time. The dynamic change of digital space over time suggests performing the research methodologies at various time periods (Janc, 2015). The current situation of a city in terms of general popularity, administration, reputation, language, and economy among many things limits the user-generated data for analysis.

Digital space components and research are exploratory in nature as there are no standards or definitions set in stone for space representations of the events in the Internet other than flows, user rates, and indices. Thus, the digital space representations were usual as patterned with the physical space concepts.

It should be noted that physical distance and mobility may continuously be defined by physical factors like the Euclidean distance and transportation networks; but multiple studies (Bailey et al., 2018) (Crampton, 2009) (Graham and Zook, 2011) (Janc, 2015) (Zook et al., 2011) and data show that an understanding of distance and proximity in terms of digital space provides a new aspect of analysis in geospatial information studies.

2. Methodology

We performed the data acquisition and pre-processing, weights generation, and analysis of results shown in Figure 2. For all stages of the study, we mainly used Python (van Rossum, 1995) programming, QGIS (QGIS Development Team, 2009), GeoDa (Anselin et al., 2010), and Microsoft Power BI (Microsoft, 2025).

2.1 Data Acquisition and Pre-Processing

The geotagged tweets were gathered using Twitter API via tweepy package (Roesslein, 2009) in Python 3.8.5. An Academic Research access of the Twitter API was acquired for the study. This limited the data gathering to 10,000,000 tweets per month with additional rate restrictions (Table 1).

API Feature	Rate Limits	
	General	Per Tweet
Total Tweets	10 million/month	-
Getting Tweets	300 requests/15 min	500 tweets/3 sec
Geocoding Tweets	75 requests/15 min	1 tweet/12 sec

Table 1. Twitter API rate limits for academic access upon conduct of this study.

Each geotagged tweet was geocoded from the ‘geo’ attribute containing: (1) ‘coords’ attribute with actual point coordinates, and (2) ‘place id’ attribute as Twitter ID for specific locations. The ’place id’ attribute is more common than ‘coords’, requiring additional geocoding. It imposed a time complication with geocoding the ‘place id’ as shown in Table 1.

Like other studies, the datasets used in this study have varying temporal characteristics, geocoding needs, and potential data quality issues. It was crucial for the later stages of the study that data pre-processing is done to have a consistent set of data. We followed the medallion framework (Databricks, 2023) for pre-processing, this framework has three stages: raw data (Bronze), cleaned data (Silver), and processing-ready data (Gold).
The medallion framework was selected for its systematic approach to handling large-scale, heterogeneous data that requires progressive quality improvement and validation; particularly suitable for exploratory research involving complex social media datasets. The Bronze stage preserved all original data without modification to maintain a historical archive and enable reprocessing if needed. The Silver stage performed comprehensive data cleaning operations including standardization of coordinate reference systems and data types,

[1] For enlarged figures, go to doi.org/10.6084/m9.figshare.30438350.v2

timestamp formatting, geocoding of place ID attributes, and spatial filtering. The Gold stage contains analysis-ready datasets optimized for spatial weight matrix generation, with tweets segmented into manageable files and consistent temporal and spatial indexing applied to all datasets. This staged approach enabled iterative quality control, computational efficiency for spatial operations, and methodological reproducibility.

The (Bronze) data acquired from Twitter was paginated into 44,554 CSV files with textual information. The latitude and longitude were geocoded from the 'geo' attribute of the tweet containing the coordinates or a place ID. The tweets were split into 21 GeoDataFrame which are converted to vector point files (GPKG) with approximately 1,000,000 points for better file management for processing.

After several trials and errors (Silver) of reading and writing data frames and GPKG files, it was found for better file management to produce split files with fewer rows than to produce one file with all data. This enables less RAM usage when processing and reading the data. There are 6,440,672 tweets within the bounds of the cities considering the study period. The centroid of QC was deleted in the barangay set which totalled 282,649 points.

The points split files (Gold) were used to intersect with the cities and barangay boundaries to get the set of points within the administrative units (barangay and city level) for mapping.

2.2 Exploratory Data Analysis

Data exploration involved summary statistics, visualizations, and temporal analysis to understand data quality and patterns. Mean and median users and tweets per month were calculated across the study period for both administrative levels to establish baseline patterns and correlations.

As there are multiple possibilities for summarizing the values, an interactive dashboard was made in Microsoft PowerBI to allow for analysis while viewing. The main advantage of creating a dashboard is to simultaneously show output and interactively edit the visuals by establishing the data models and visualizations. Although PowerBI allows for many visuals and possibilities of customizations, limitations on the scope and aggregations were imposed to simplify the output dashboards.

2.3 Physical Spatial Weights Development

The physical space weights chosen for the study are three of the most common and representative of physical space relationships in varying degrees: (1) contiguity, (2) distance, and (3) network.

Queen contiguity and inverse distance weights (Anselin, 1988) were created using GeoDa, generating text files with neighbor lists and weight values. Ermagun and Levinson (2018) suggested using the network weight matrix to consider the connectivity in terms of travel to properly model the real-world networks. ESRI (2023), however, suggests that network weights are only recommended when the measurements involve accessibility and movement in the network. Freeman (1978) offered a compromise by following the measure of network in terms of degree centrality, the number of links between locations: link being a set of road lines connecting the locations. The network links were generated via Python using GeoPandas (Jordahl et al., 2020), counting road connections by using roads and boundaries.

For the discussion in this paper, short forms for the physical spatial weights were used in the figures and tables. Specifically, the study focuses on the (1) Queen Contiguity (First Order) [assigned short form: *queen*], (2) Inverse Distance (First Power) [assigned short form: *invdst*], and (3) Network Links [assigned short form: *links*] to depict the physical space.

Since these representations are fixed in nature and would not change in a span of time, it was assumed that the weights for both the city and barangay level are fixed throughout the study.

2.4 Digital Spatial Weights Development

The related literature on digital space studies show that they are exploratory and suggestive in nature. There are no established standards or measures for digital space components like the Internet, social media, or other online presence.

Bailey et al. (2018) formalized the strong similarity of digital relationships and physical movement (Zook et al., 2011) (Janc, 2015) through the Social Connectivity Index (SCI) given by Equation 1. SCI measures the relative probability of social connectedness of Facebook (FB) users across two locations.

$$SCI_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i \times FB\_Users_j} \tag{1}$$

where $FB\_Connections_{i,j}$ is the total FB connections between friends in locations i and j, FB_Users_i is the number of users in location i, and FB_Users_j is the number of users in location j.

When COVID-19 struck, Kuchler et al. (2022) compared SCI and physical distance with the spread of COVID-19 and found that the social proximity changes in an area strongly affected the increase in COVID-19 cases in the area. SCI slightly improves prediction models, but the analysis highlights the value of having similarly accessible data.

The digital spatial weighting schemes presented in this study are adapted from the definitions of physical weights (Anselin, 1988) (Freeman, 1978) or digital media research (Bailey et al., 2018) (Zook et al., 2011) (Janc, 2015).

The equations are based on the locations of city or barangay matching per row, i, and column, j. The weights matrices were generated per period; the tweets and/or users are filtered based on the given range of dates or period which formed the matrix elements according to their corresponding equations. Each weighting scheme was stored in a separate folder with the naming based on the administrative boundary (city or brgy for city or barangay) and period (dly, wly, bmo, mon for daily, weekly, bi-monthly, and monthly) with start date (YYYY-MM-DD).

Short forms for each digital space weighting scheme suggested were also assigned for use in the processing, figures, and tables. Four digital space weighting schemes are suggested for the study: (1) Contiguity in Tweets [assigned short form: *contiguity*], (2) Probability of Tweeting [assigned short form: *probability*], (3) Intensity of Tweeting [assigned short form: *intensity*], and (4) Tweets-Scaled Probability of Tweeting [assigned short form: *probscale*]. These weighting schemes are based on the physical space weights with *contiguity* based on the *queen* with a combination of distance-based weighting like *invdst*; and the *probability* based on SCI. Both *contiguity* and *probability* being based on the number of users tweeting in the area at a given period while *intensity* and *probscale* consider the tweets made by the users in the area at a given period making them somehow improved *probability*.

Since the digital space weights mainly associates the users tweeting in locations i and j at a given period, the computations will produce matrices that have non-zero and non-one matrix main diagonal elements. The decision on the inclusion of the unit itself and row-standardization (scaling) were based on the intended definitions of the weighting scheme. To maintain the spatial weights for use in the comparisons, the consistency on the non-inclusion of the unit itself is done for all the matrices. On the other hand, the row-scaling decision was based on the magnitude and meaning of the rows of the matrices.

2.4.1 Contiguity in Tweets: This describes the existence of tweets in both locations i and j as shown in Equation 2 by getting the total number of users based on their tweeting habits in both locations.

$$Contiguity_{i,j} = \sum user\ [contiguity] \tag{2}$$

Using the Iverson Bracket to indicate a function in $[contiguity]$,

$$[contiguity] = \begin{cases} 1 & ,if\ user\ tweets\ in\ both\ i\ and\ j \\ 0 & ,otherwise \end{cases} \tag{3}$$

Like *queen* in physical space, this weight will show the presence of users in both spatial units' relationship in the digital space.

2.4.2 Probability of Tweeting: This suggested weighting scheme, given by Equation 4 is based on SCI (Equation 1) where they related friendship links via the probability of the friendship counts across administrative boundaries.

$$Probability\ of\ Tweeting_{i,j} = \frac{Users_{i,j}}{Users_i + Users_j} \tag{4}$$

where $Users_{i,j}$ is the total unique users tweeting in both locations i and j, $Users_i$ is the total users tweeting in location i, and $Users_j$ is the total users tweeting in location j.

2.4.3 Intensity of Tweeting: This represents the number of tweets the users make in locations i and j shown in Equation 5. This is a novel weighting scheme focusing on the number of tweets more than the users tweeting in the locations.

$$Intensity\ of\ Tweeting_{i,j} = \sum_{m=1,2,...}^{n} (user_m) \left(\frac{tweets_{m,i}}{tweets_{m,i,j}} \right) \tag{5}$$

where n = number of users tweeting in locations i and j

2.4.4 Tweets-Scaled Probability of Tweeting: Adding to the novelty of *intensity*, this weighting scheme (Equation 6) combines *intensity* and *probability* to both consider the users and tweets in the area at the given period. Using the short form in the equation,

$$probscale_{i,j} = \frac{Users_{i,j} \times Tweets\ Scale\ of\ Users_{i,j}}{Users_i + Users_j} \tag{6}$$

The tweets scale of users (Equation 7) mainly follows the *intensity* definition but considers the users in i and j.

$$Tweets\ Scale\ of\ Users_{i,j} = \sum_{m=1,2,...}^{n} (user_m) \left(\frac{tweets\_user_{m,i}}{tweets\_user_{m,i,j}} \right) \tag{7}$$

2.5 Spatial Weights Comparison

Berry (1990) suggested the use of matrix visualization via color mapping to depict the values in each row and column instead of analyzing the magnitudes of each matrix element, this is called matrix maps. Although the study was mainly done for large matrices, the visualizations presented a better understanding of the pattern of the matrix values across the matrix configuration.

Going into the geospatial unit, the weights are applied to the Metro Manila cities and municipality mapping to depict the spatial weights in terms of the city and municipality boundaries (Duncan et al., 2017). The matrix maps produced will cover a per administrative boundary – per time aggregation while the spatial weights maps will be per actual barangay or city, per time aggregation in the study period.

Aside from the visual comparisons of the matrix and weights maps, the constraints on the creation of the weights were also tallied for reference using the weights for spatial modeling.

3. Results and Discussion

3.1 Twitter Data Processing and Analysis

Quezon City demonstrated significantly higher digital activity, accounting for approximately 60% of total tweets, with consistent concentration in commercial and educational areas including Bagong Pag-asa, Central, Loyola Heights, and South Triangle barangays. Figure 3 shows that the spatial distribution revealed clear concentration of tweets in southern QC regions with higher population density and mixed-use development.

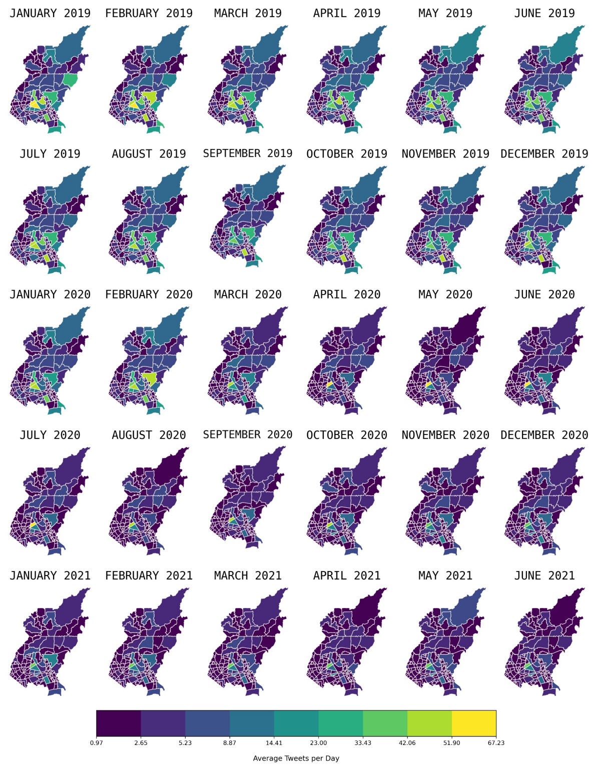

Figure 3. Monthly average tweets per day for QC barangays.[1]

The March 2020 lockdown produced distinct temporal-spatial effects validating the framework's sensitivity to societal changes. Most cities showed 30-50% increases in tweet volume during lockdown initiation, while unique user counts decreased by 15-25%, indicating intensified activity per user rather than expanded user base. This divergent behavior between tweet volume and user distribution demonstrates the framework's capability to capture changing digital spatial relationships during major events.

Successful geocoding exceeded 90% across all administrative levels. The temporal analysis revealed adequate data coverage for digital weight matrix generation, with spatial variations aligning with documented urban land use characteristics of the barangays and cities based on QC land use classifications (Quezon City Government, 2016) and general knowledge of the areas, as check. The 54% variation in tweet patterns versus 15% variation in user patterns across cities suggests different spatial processes governing digital activity intensity versus digital presence. At the barangay level, this differentiation becomes more pronounced as seen in the bivariate analysis done with tweets and user counts. Figure 4 shows barangays that consistently had concentrated activities correlating with specific land use types and institutional presence. This reveals fine-scale digital spatial processes that administrative boundaries alone cannot capture.

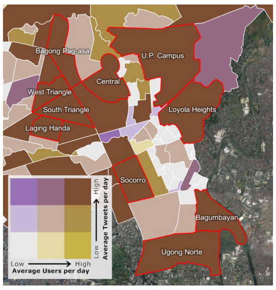

Figure 4. Part of the March 2020 bivariate map of tweets and users in QC barangays. Barangays with consistently high users and tweets across the study period are outlined in red. [1]

3.2 Physical Spatial Weights

There are three physical space weights implemented in the barangay and city levels for a total of six physical space weights for the study. The physical space weights were used based on the common constraints in using spatial weights for modeling.

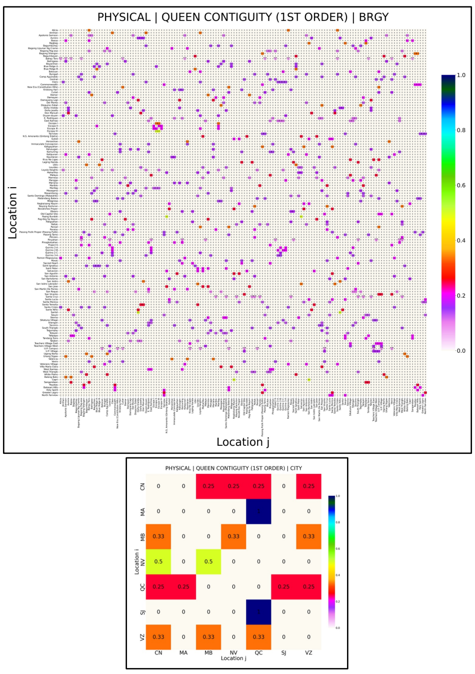

Figure 5. Queen contiguity (first order) spatial weights for QC barangays (top) and seven Metro Manila cities (bottom). [1]

Figure 5 shows the *queen* matrices. Being the most common spatial weight used thus far in spatial analysis studies, *queen* was utilized in its simplest first order form with the binary contiguity in the queen direction. Binary contiguity spatial weights are recommended to be row-normalized (row-scaling) to have the sum of the weights' row equal to unity. This is done to maintain the effect of the neighbors to the variable values and not exceed the actual value by huge amounts (if we use 1's and 0's only).

We used the *invdst* spatial weight with the distance in meters raised to the 1st power. Unlike *queen*, *invdst* is not immediately row-scaling unless necessary for the target spatial relationship being defined. Row-scaling distance-based weights in general will scale the weights but will not represent the distance decay function in most cases. For this study, *invdst* was used to depict the distance decay function, so row-scaling was avoided.

The last physical space weighting scheme used is a representation of the network weights, *links* (Freeman, 1978). Since there are links within the spatial units, the main diagonal of the weights matrix will contain link counts as values. A comparison of the inclusion or exclusion of the main diagonal for row-scaling was done to decide on which constraint will be set for this spatial weight. It was found that the row-normalized without the main diagonal have better representation closer to the COVID-19 cases scenario for the study period. Caution should be taken here as this should depend on the spatial relationship being represented.

Table 2 shows the summarized constraints for each physical spatial weight. This notes the row-standardization, matrix main diagonal consideration, and final matrix main diagonal of zero.

Spatial Weight	Order/ Method	Row-Scaling	Diagonal included in Scaling	Zero Diagonal in Final Weights
queen	1st Order	Yes	No	Yes
invdst	1st Power	No	-	Yes
links	Link count	Yes	No	Yes

Table 2. Physical Space Spatial Weights Constraints.

In GeoDa, an option is given to consider the main diagonal of the spatial weight. To keep to the meaning of physical space as defined here, the main diagonal for the three physical weights were set to 0 (a unit is not related to itself physically).

3.3 Digital Spatial Weights

Digital spatial weight generation produced 1,131 matrices per administrative boundary across four temporal aggregations: 912 daily, 129 weekly, 60 bi-monthly, and 30 monthly. Each of the four digital weighting schemes generated 2,262 total matrices, requiring 9-10 hours processing time per scheme due to iterative temporal filtering and user-tweet relationship calculations.

Digital weight matrices exhibit fundamentally different characteristics from physical spatial weights. Administrative scale effects are pronounced in digital weights. City-level matrices (Figure 6) show clearer, more stable patterns than barangay-level (Figure 7) due to spatial aggregation of the same underlying tweet dataset. Larger spatial units capture more user activities, producing higher values and more robust relationships.

Matrix sparsity varies significantly with temporal aggregation: daily periods produce predominantly sparse matrices with extensive zero values, particularly at barangay level, while bi-monthly and monthly aggregations capture more complete spatial relationships. This temporal dependency demonstrates that digital spatial relationships require accumulation periods to establish meaningful patterns, suggesting that sustained digital engagement over time may be necessary to reflect stable spatial connections. Daily aggregation may be useful for event-specific day patterns or emergency response studies while weekly and bi-monthly might be suitable for policy effects with moderate temporal sensitivity, and monthly or longer aggregation for models with long-term trends. The choice of temporal scale should align with both the research question's temporal requirements and the expected digital activity levels in the study area. Areas with lower digital activity may require longer aggregation periods to achieve meaningful spatial weight matrices, while high-activity areas can support shorter temporal scales for more temporally sensitive analysis.

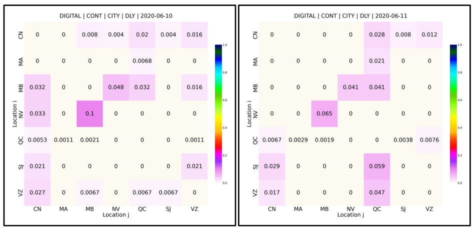

Figure 6. Sample digital spatial weights (*contiguity*) for city level with daily period in June 10 and June 11, 2020. [1]

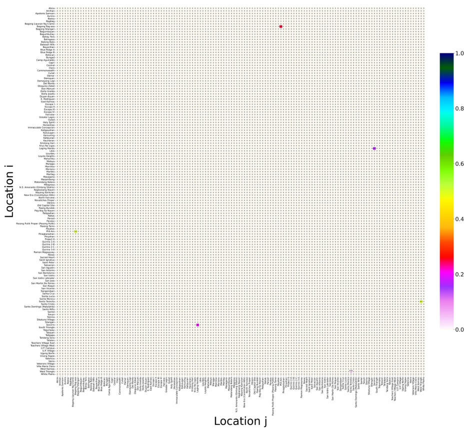

Figure 7. Sample digital spatial weights (*contiguity*) for barangay level with daily period in June 10, 2020.[1]

Each digital weighting scheme captures distinct aspects of spatial relationships. Contiguity in Tweets follows binary logic similar to physical queen contiguity but based on user presence rather than geographic adjacency. Results show this scheme effectively captures areas where users maintain dual presence, regardless of physical distance. Following the contiguity basis, binary contiguity, it was set to have row-standardization to scale the effect of neighboring spatial units per location.

Probability of Tweeting (based on SCI) measures relative likelihood of spatial connectedness through digital activity. This scheme produced the most stable temporal patterns, suggesting probability-based measures are less sensitive to short-term activity fluctuations. It is the only weighting scheme with no apparent dimension suggesting not to row-standardize anymore. Intensity of Tweeting emphasizes activity volume over user presence, revealing areas of concentrated digital engagement. High-intensity connections often correspond to institutional or commercial areas (e.g., UP Campus, Central business districts) where users generate substantial tweet volumes.

Tweets-Scaled Probability of Tweeting combines user presence and activity intensity, producing nuanced connectivity measures that account for both breadth (user presence) and depth (activity levels) of digital relationships.

With digital space weights, the question of row-standardization also poses the question of including the main diagonal values or excluding them from the standardization. Unlike physical weights where standardization primarily scales neighbor effects, digital weight standardization must account for varying user activity levels and temporal accumulation effects. Since the weighting schemes are based on Twitter users and tweets, the effect of a spatial unit to itself based on user activity should be accounted for when the focus is existence of users, strength of existence based on their tweets, and scaled probability of tweeting. Including diagonal values in standardization accounts for self-connectivity (users tweeting within their location) while maintaining spatial weight conventions by setting the main diagonal to zero in final weights. This allows digital weights to preserve the meaning of user presence within spatial units during dimension removal.

Table 3 shows the summary of constraints for the digital space spatial weights. These constraints were made to be simple according to the spatial weights' basic concepts and usage.

Spatial Weight	Order/ Method	Row-Scaling	Diagonal included in Scaling	Zero Diagonal in Final Weights
contiguity	-	Yes	Yes	Yes
probability	-	No	-	Yes
intensity	-	Yes	Yes	Yes
probscale	-	Yes	No	Yes

Table 3. Digital Space Spatial Weights Constraints.

3.4 Spatial Weights Comparison

Spatial weights maps were generated to show how the weighting scheme affected the spatial relationship between units. Much more than the number of matrix maps, the spatial weights maps took more time to generate – around 20 to 22 hours per weighting scheme. A total of 168,996 spatial weights maps were produced.

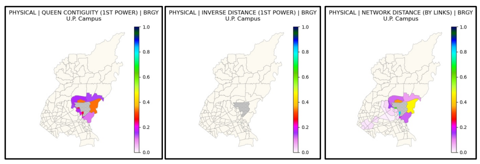

Figure 8. Physical space spatial weights maps of UP Campus (shaded gray): (left to right) *queen*, *invdst*, and *links*. [1]

Figure 8 and Figure 9 show the physical space and digital space weights maps of UP Campus to serve as a sample comparison of the two spaces. Physical space weights demonstrate clear distance-decay patterns with higher connectivity assigned to geographically proximate units, following established spatial analysis principles.

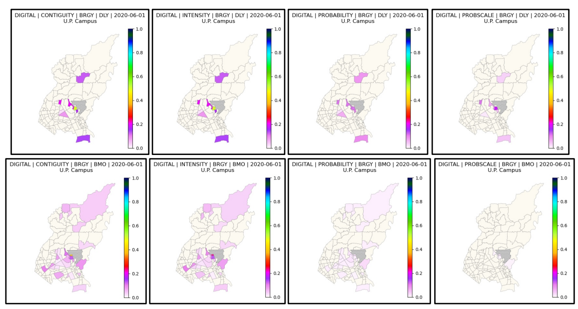

Figure 9. Digital spatial weights maps of UP Campus (shaded gray): (left to right) *contiguity*, *intensity*, *probability*, *probscale* (top: daily, bottom: bi-monthly) for June 1, 2020 range. [1]

Digital space weights (Figure 9) exhibit different connectivity patterns that transcend physical proximity, establishing relationships based on user activity rather than geographic adjacency. Areas like UP Campus show digital connections to physically distant locations, capturing mobility and interaction patterns not represented in traditional physical spatial weights.

Figure 10 illustrates how *queen* and *links* schemes show different connectivity patterns for UP Campus and its neighbor UP Village, with *links* providing more realistic representation of actual accessibility while the digital spatial weights for both barangays also show variations of connectivity.

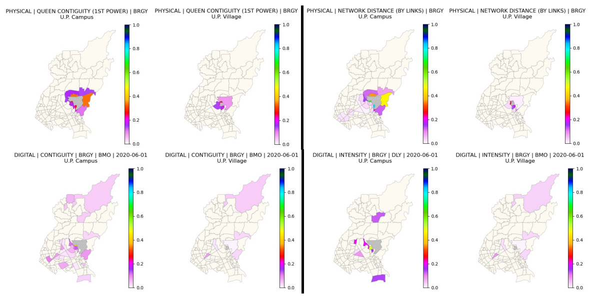

Figure 10. UP Campus (shaded gray) and UP Village (shaded gray) weights maps using different weighting scheme. [1]

City-level analysis reveals clearer patterns than barangay-level due to spatial aggregation effects. The same tweet dataset produces higher weight values at city level, demonstrating that administrative boundary definition significantly impacts digital weight magnitudes and connectivity patterns.

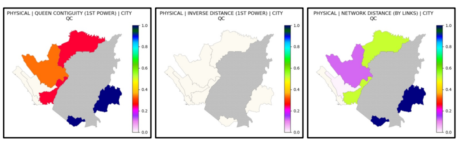

Figure 11. Physical space spatial weights maps of QC (shaded gray) in order from left to right: *queen*, *invdst*, and *links*. [1]

A key finding illustrated in Figure 11 and Figure 12 is that digital weights bypass the "island problem" inherent in physical spatial analysis. The comparison between QC's physical weights (Figure 11) and digital weights (Figure 12) reveals how physically disconnected areas can exhibit strong digital connectivity through user activity patterns. Areas like NV display digital connections to QC despite limited physical network connections, suggesting digital weights capture interactions that transcend immediate physical connectivity.

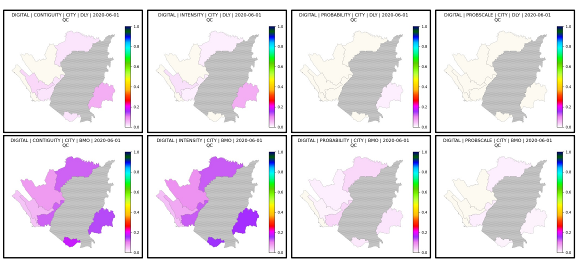

Figure 12. Digital space spatial weights maps of QC (shaded gray): (left to right) *contiguity*, *intensity*, *probability*, *probscale* (top: daily, bottom: bi-monthly) for the June 1, 2020 range. [1]

Figure 13 shows digital weights for QC across pre-pandemic, during-lockdown, and post-lockdown periods. Although conclusive patterns are difficult to establish, the digital space variation in time provides good insights of human behavior compared to proxy representations using physical space. Notably, barangays in areas not associated with universities or commercial establishments show stronger relationships to physically nearer neighbors for contiguity and intensity schemes, similar to city-level patterns. This suggests digital space can capture physical distance-based relationships but only partially, as other related units are physically distant, indicating no full adherence to distance decay functions.

Across all figures, the *invdst* produced notably low values (0-1 range) in both physical and digital comparisons, suggesting potential scaling corrections or kilometer-dimension adjustments may be necessary. This pattern is consistently visible in Figures Figure 8, Figure 11, and the physical weight comparisons, stressing parameter optimization importance in spatial weight development.

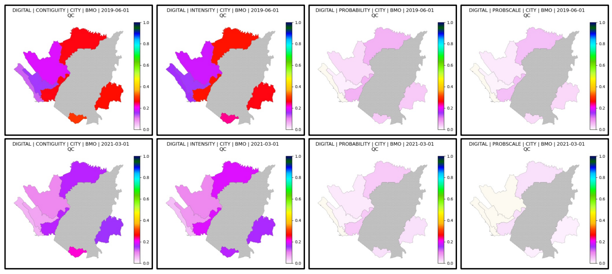

Figure 13. Digital spatial weights maps of QC (shaded gray) in bi-monthly: (left to right) *contiguity*, *intensity*, *probability*, *probscale*; Pre-pandemic on June 1, 2019 and post-lockdown on March 1, 2021. During lockdown period in Figure 12 [1]

The figures also reveal that physical and digital weights represent complementary rather than competing spatial realities. Physical weights effectively capture proximity-based relationships and infrastructure-dependent interactions, while digital weights reveal activity-based connectivity and behavioral spatial patterns. The *"nearness=relatedness"* assumption fundamental to physical spatial analysis does not govern digital space relationships, where connectivity depends on user activity and temporal dynamics rather than geographic proximity.

The distinct characteristics shown across Figure 8 to Figure 13 suggest that combined approaches may provide more comprehensive spatial models than either method alone. Most urban phenomena occur simultaneously in physical and digital spaces, making integrated spatial weight frameworks potentially more accurate for contemporary spatial analysis applications.

4. Conclusions

This study demonstrates that digital spatial weights are fundamentally different from physical spatial weights, successfully answering the core research question using Twitter data. The framework generated digital spatial weight using four schemes (*contiguity*, *probability*, *intensity*, *probscale*) across multiple temporal periods, enabling direct comparison with traditional physical spatial weights (*queen*, *invdst*, *links*).

Digital weight matrices show fundamentally different connectivity patterns based on user activity rather than geographic proximity, successfully addressing the "disconnected spatial units" problem in physical spatial analysis. Digital weights demonstrate temporal sensitivity that are not present in physical weights, capturing behavioral changes. Administrative scale effects revealed that larger spatial units produce more stable digital weight matrices due to activity aggregation.

Data collection was constrained to seven Metro Manila cities due to Twitter API restrictions. No bias correction was applied for demographic representativeness or bot activity. The exploratory nature of digital space research necessitates continued methodological development as platforms evolve. Future research should explore integration methodologies, validation through modeling applications, expansion to other platforms, and standardized evaluation metrics. Importantly, testing this framework across diverse social media platforms (Facebook, Instagram, LinkedIn, location-based services) would help establish the generalizability of digital spatial weight concepts and identify platform-specific characteristics that influence spatial relationship representation.

The distinct characteristics of digital versus physical spatial weights suggest they capture different dimensions of spatial relationships. While this study demonstrates their differences through descriptive analysis, future research should test whether combined physical and digital approaches provide improved explanatory power in spatial modeling applications. Particularly, through spatial panel data models that can assess the relative contribution of digital versus physical spatial relationships in explaining different phenomena. A hybrid weight matrix can be constructed by combining standardized digital and physical weight matrices based on their relative contributions identified. Alternatively, context-specific integration can be employed where digital weights may be applied for mobility-based spatial analyses while physical weights may be applied for static spatial relationships. This allows researchers to leverage the temporal sensitivity of digital weights for dynamic processes while maintaining traditional physical relationships for stable phenomena.

This framework establishes a foundation for incorporating digital human activity into spatial analysis, though empirical validation of integrated approaches requires future modeling studies.

References

Anselin, L. E., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Anselin, L., Syabri, I., Kho, Y., 2010. Geoda: An introduction to spatial data analysis.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., Wong, A., 2018. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives* 32(3), 259–280. doi.org/10.1257/jep.32.3.259

Berry, M.W., 1990. The Use of Matrix Visualization in Algorithmic Design. *Computing Systems in Engineering* 1(1), 63–73.

Crampton, J.W., 2009. Cartography: Maps 2.0. *Progress in Human Geography* 33(1), 91–100.

Databricks, 2023. Medallion Architecture.

Duncan, E. W., White, N. M., Mengersen, K., 2017. Spatial smoothing in Bayesian models: A comparison of weights matrix specifications and their impact on inference. *International Journal of Health Geographics* 16.

Ermagun, A., Levinson, D., 2018. An Introduction to the Network Weight Matrix. *Geographical Analysis* 50(1), 76-96. doi.org/10.1111/gean.12134

ESRI, 2023. Generate network spatial weights (spatial statistics). Retrieved from ArcGIS Pro.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. *Social Networks* 1, 215–239. doi.org/ 10.1016/0378-8733(78)90021-7

Getis, A., Aldstadt, J., 2004. Constructing the spatial weights matrix using a local statistic. *Geographical Analysis* 36(2), 90-104. doi.org/10.1111/j.1538-4632.2004.tb01127.x

Graham, M., Zook, M., 2011. Visualizing global cyberscapes: Mapping user-generated placemarks. *Journal of Urban Technology* 18, 115–132. doi.org/ 10.1080/10630732.2011.578412

Janc, K., 2015. Visibility and Connections among Cities in Digital Space. *Journal of Urban Technology* 22, 3-21. doi.org/ 10.1080/10630732.2015.1073899

Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L. J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, Leblanc, F., 2020. geopandas/geopandas: v0.8.1.

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D., 2015. Understanding human mobility from Twitter. *PLoS ONE* 10. doi.org/ 10.1371/journal.pone.0131469

Kemp, S., 2022. Digital 2022: Global overview report.

Kuchler, T., Russel, D., Stroebel, J., 2022. The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook. *Journal of Urban Economics* 127, doi.org/10.1016/j.jue.2020.103314.

Lopez, R. M., 2021. Speech of secretary ramon m. lopez, philippine digital convention 2021.

Microsoft, 2025. Power BI.

QGIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation.

Quezon City Government, 2016. Land Use and Zoning Map. City Planning and Development Department (CPDD).

Roesslein, J., 2009. Tweepy. Tweepy Python Package.

Twitter, Inc, 2023. Twitter Developer Portal platform.

United Nations, 2021. Philippines - subnational administrative boundaries. Retrieved from The Humanitarian Data Exchange.

van Rossum, G., 1995. Python tutorial. Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam.

Zook, M., Devriendt, L., Dodge, M., 2011. Cyberspatial proximity metrics: Reconceptualizing distance in the global Urban system. *Journal of Urban Technology* 18(1), 93–114. doi.org/10.1080/10630732.2011.578411