# Spatial Neighbourhood Matrix Computation – Inverse Distance Weighted versus Binary Contiguity

João NEGREIROS

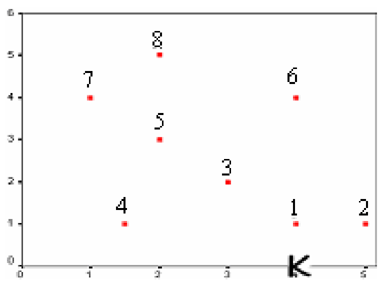

## Abstract

A particular GIS aspect is the existence of the spatial neighbourhood weighted matrix (**W**) needed for their assessment and commonly setup by the binary or inverse distance weight (IDW) contiguity. This article recommends that if the former approach fulfils **W** specification quite well, the distance decay contiguity of IDW, by itself, is blinded to its natural neighbours because the exponent of $1/d^{exp}$ already underlies the considered number of neighbours. In a more wise approach, IDW should be used mutually with the binary one.


## 1 Introduction

Quite often, spatial weight matrix (formal expression of spatial dependence between spatial units) quantification is commonly based on IDW, rook, lengths of shared borders, $n^{th}$ nearest neighbour distance, network links, centroids within d distance, and bishop and queen binaries approaches. Yet, if spatial objects are points, the vicinity boundaries become unclear, reflected in their **W** matrix contents and, therefore, in spatial autocorrelation and clustering results. The choice of geographical weights for spatial statistical modelling is not a clear cut issue.

It is clear that different weights lead to an assorted Moran I and Moran scatterplot results. As well, the **W** neighbourhood matrix is not influenced by the region value because **W** calculus does not include sample values. By choosing different neighbour limits for the same layout, observations can only exchange positions between quadrants I and IV or II and III of the Moran scatterplot in a vertical movement whose amplitude is, in average, 15 % of the overall mean. According to the same author, the contribution of $1/d^5$ and $1/d^4$ to each sample, regarding K point (see Fig. 1), holds a massive power for the closest two samples. If binary contiguity were considered, the same weight would be given to its natural neighbours (1, 2, 3 and 4, for instance) but leading to a more smooth weight situation among neighbours.




| | 1/d | $1/d^2$ | $1/d^3$ | $1/d^4$ | $1/d^5$ |
|---|---|---|---|---|---|
| 1) | 28.4% | 47.7% | 62.4% | 73.6% | 81.4% |
| 2) | 19.9% | 23.0% | 22.7% | 19.1% | 15.1% |
| 3) | 12.9% | 9.6% | 5.8% | 3.1% | 1.5% |
| 4) | 12.9% | 9.6% | 5.8% | 3.1% | 1.5% |
| 5) | 7.6% | 3.5% | 1.3% | 0.4% | 0.1% |
| 6) | 7.1% | 3.2% | 0.9% | 0.2% | 0.0% |
| 7) | 5.6% | 1.6% | 0.4% | 0.1% | 0.0% |
| 8) | 5.6% | 1.6% | 0.4% | 0.1% | 0.0% |

**Fig. 1:** An impact comparison of $1/d^n$ contiguities over **W** weight matrix

## 2 Case Study

The city of San Diego, California, is not a homogeneous region where reasonable small sections of the metropolitan area follow a distinct spatial autocorrelation tendency like major cities in this world (see Fig. 2). Because Rancho Santa Fe is the richest suburban area with house prices nearly three times higher than the next highest district, La Jolla, its value was removed because of the lack of robustness of spatial autocorrelation statistics (GETIS & ORD 1992). The total housing costs mean equals \$192.81 while variance, skewness and kurtosis are 5523, 0.83 and 0.20, respectively.

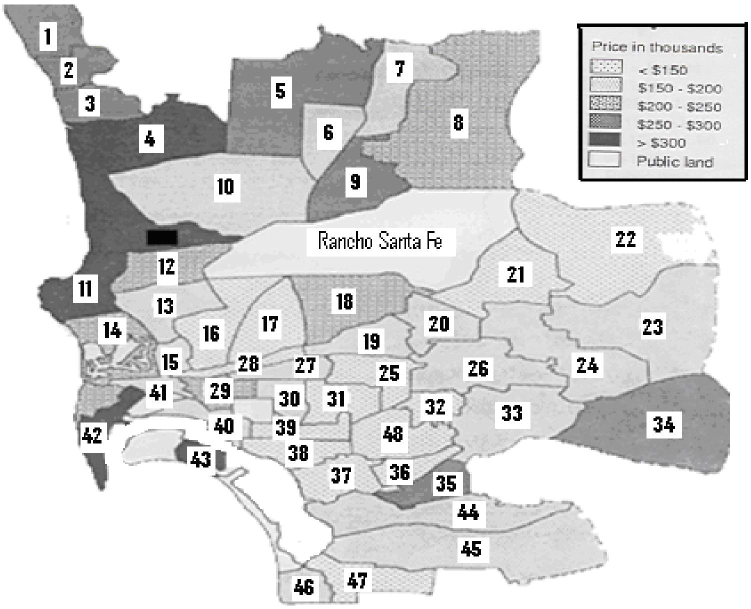



**Fig. 2:** The layout map of San Diego, CA, county

The Moran scatterplot (housing ward cost versus neighbourhood housing ward cost) were computed for five different situations: I (**W**=0 or 1); II (**W**=1/d); III (**W**=$1/d^2$); IV (**W**=$1/d^3$); V (**W**=$1/d^4$). Yet, a certain refinement of knowledge can be gained: (A) IDW (II) holds the capability of limiting points around the Moran scatterplot centre (44 wards), confirmed by the weak Moran I of +0.0742. (B) $1/d^2$ (III) and $1/d^3$ (IV) recognize high and low patterns better than IDW. This conclusion is reinforced by their Moran I (+0.1630 and +0.1809 for $1/d^2$ and $1/d^3$, respectively). (C) The binary contiguity puts more emphasis on the rich northern districts (Encinitas (1), Cardiff (2), Del Mar (4), Lake Hodges (5), Rancho Penasquitos (6), Rancho Bernardo (7), La Jolla (11), Beaches (14), Bay Park (15) and Kearny Mesa (16)) then other **W** matrices (noticed that those districts follows a long and thin design layout).

# 3 Discussion

When a local pattern respects a long but narrow arrangement such as the rich northern area of San Diego, binary contiguity fits best because it avoids the smoothing effect of inverse distance decay contiguity and, thus, it places those observations farther away from the Moran scatterplot centre. With binary contiguity, each neighbour's weight is inversely proportional to its total number of neighbours and, therefore, it becomes more force resistant to the gravity centre force than IDW. Cardiff (2), for instance, is the only Encinitas (1) neighbour receiving, thus, 100 % of the total weight. With the 1/d approach, Cardiff (2), Solana Beach (3) and Del Mar (4) receives 13 %, 7 % and 5 %, a more even situation. Concerning $1/d^4$ assessment, those three regions contribution reaches 85 %, 11 % and 1 %, respectively. As the distance difference among neighbours gets larger, more weight is given to the closest regions (see Tab. 1).

**Table 1:** The average weights for the first four distance decay contiguities versus binary

| Distance decay contiguity weights | | | | | Binary |
|---|---|---|---|---|---|
| **Nearest neighbour** | **$1/d^4$** | **$1/d^3$** | **$1/d^2$** | **1/d** | |
| 1st | 80 % | 65 % | 40 % | 32.5 % | 100 % divided by the total direct number neighbours |
| 2nd | 15 % | 17.5 % | 25 % | 27.5 % | |
| 3rd | 5 % | 7.5 % | 12.5 % | 20 % | |
| 4th | | 10 % | 10 % | 12.5 % | |
| 5th | | | 12.5 % | 5 % | |
| Others | | | | 2.5 % | |

If binary contiguity is applied (note that boundary length or others secondary vicinity rings are not considered here, a restrictive constraint), the eight neighbours of Centre of San Diego (48) receive 12.5 % weight each. With 1/d methodology, the neighbourhood weights range between 2.9 % and 17.4 %, leading to a smoother neighbourhood average. With $1/d^4$, this scope fluctuates between 0.8 % and 52 %. This trend is confirmed by the following domains of their neighbourhood average: [218.67, 168.87] of 1/d against [283.44, 114.58] of $1/d^4$.

The major concern of IDW specification is the smoothing effect by of concentrating their positions close to the Moran scatterplot centre and, thus, a smaller Moran I. On the other hand, as $1/d^n$ decreases ($n \rightarrow \infty$), the neighbourhood average is reduced to the closest sample unless the distance difference between the closest and the following ones are quite small. With $1/d^5$, for instance, the closest neighbour holds, on average, 90 % of all the weights. This gives the suggestion that the more the distance decay ($1/d^n$) decreases, the wider the neighbourhood range turn out to be and, thus, the more anti-gravitational the Moran scatterplot layout becomes. What, then, is the best distance decay contiguity? No rules can be setup.

From the statistical viewpoint, this contiguity process is correct. Yet, it does not make sense to reduce vicinity to the closest sample in spite of the fact that neighbourhood is a fuzzy concept. Another major concern is the dissimilarity found in the Moran I regarding binary contiguity (+0.27) against distance decay contiguity, a scale issue. With $1/d^6$, for example, this index equals -0.21. However and looking closely at its distribution map, the city of San

Diego pursues a distinct spatial autocorrelation pattern and, thus, a positive Moran I is expected like major spatial Mother Nature processes.

The key question regarding distance decay contiguity results of the expected number of neighbours considered by this methodology. With 1/d, the nearest five neighbours represent 97.5 %, in average. With $1/d^2$, the nearest four neighbours hold 87.5 % of the total weight. With $1/d^3$, the nearest three neighbours stand for 90 %. With $1/d^4$, the nearest two neighbours own 95 %. With $1/d^5$, the nearest one retains around 97 %. Therefore, inverse distance decay is not sensitive to the total number of natural neighbours that any region might hold. With 1/d, the nearest five neighbours are the ones that really counted for regardless of the true number of direct neighbours. If $1/d^4$ is considered, this number is reduced to two. Yet, this discrepancy of expected neighbours becomes quite obvious, if binary contiguity is setup. Consistent with these last results, the standard deviational ellipse consider twenty-one counties with only one or two natural neighbours, fifteen with three, four or five direct neighbours while the rest of the eleven counties hold between six and eight neighbours.

For ANSELIN (1992), the neighbourhood first order contiguity structure for West Virginia counties averages five links per county, although their composition varies quite considerably. This supports the San Diego argument: one ward with eight neighbours, thirteen counties encloses two/three connections while thirty four register between four and seven neighbours. There are no island constraints and anisotropy is not being taken into account. Distance decay contiguity does not respect the real direct neighbours that any location might hold. Furthermore, distances between centroids of less than one may create problems regarding IDW computation although distance matrix rescaling such that the smallest distance equals one can be made.

## 4 Conclusion

The fact is that distance decay contiguity by itself is not a good process for **W** matrix specification because it is blind and insensitive to its obvious natural neighbours, unlike the binary approach. Contrary to major on-going practice, IDW must be used in combination with binary approach to recognize direct neighbours. Then, their standard weights should be closely proportional to their central-neighbours distances. Returning to the eight neighbours of San Diego Centre (48) ward, for instance, their **W** weights would become 17 % (Spring Valley), 8.3 % (Paradise Hills), 12.6 % (National City),12.6 % (Logan Heights),12.6 % (West San Diego), 11.4 % (East San Diego), 16 % (La Mesa) and 9.5 % (Lemon Grove) with this mix approach leading, therefore, to a more prudent situation.

## References

ANSELIN, L. (1992), SpaceStat Tutorial – A Workbook for using SpaceStat in the Analysis of Spatial Data. University of Illinois, Urbana-Champaign.

GETIS, A. & ORD, J. (1992), The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis, 24, pp. 189-206.