

CUADERNO DE TRABAJO

Diseño y aplicación metodológica para la definición de límites de ciudades, mediante la Teoría de la Percolación

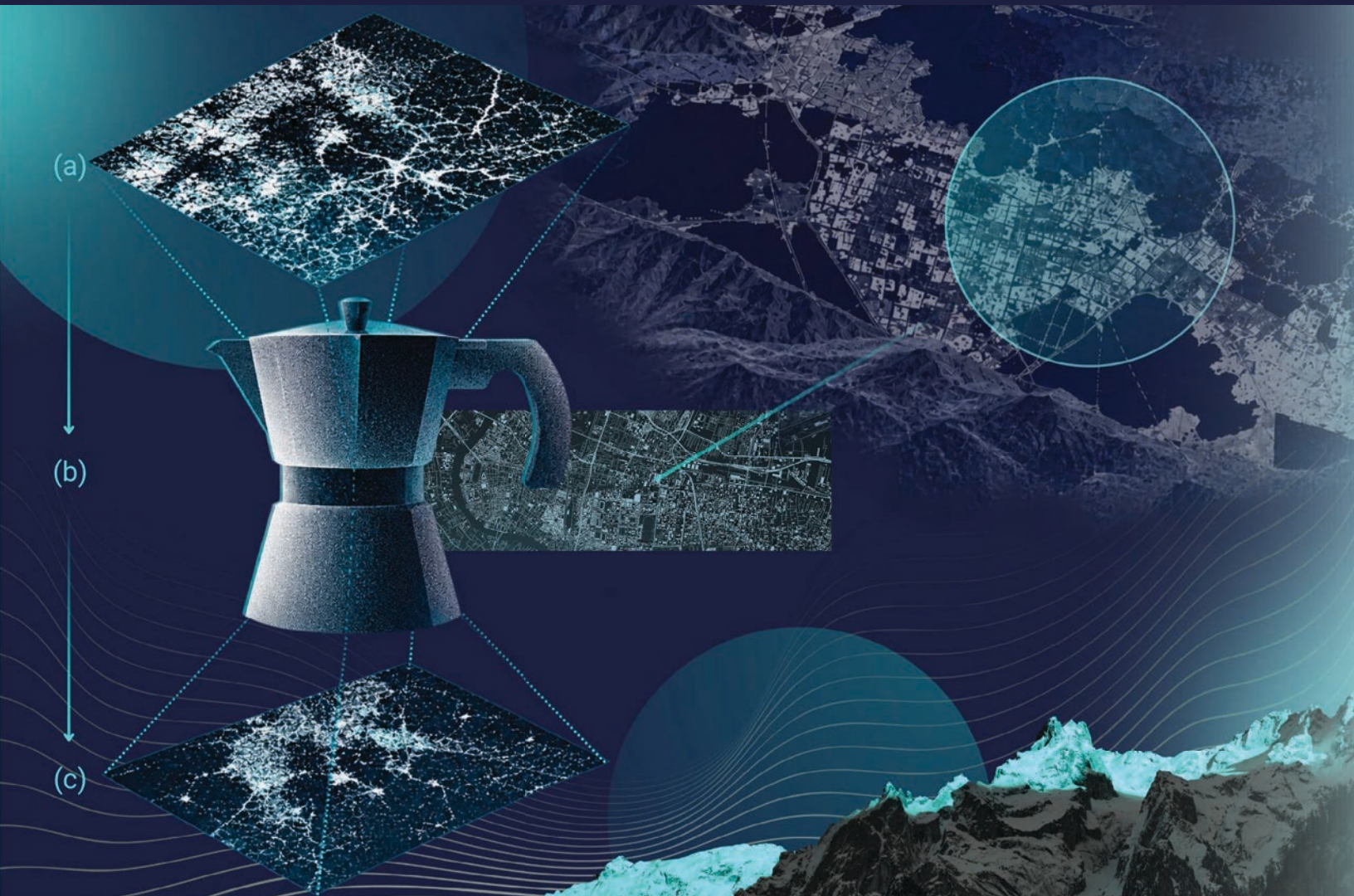
Jorge Alberto Montejano Escamilla

Camilo Alberto Caudillo Cos

Rodrigo Tapia McClung

Felipe Gerardo Ávila Jiménez

EDITORIAL
CentroGeo



Ciencia y Tecnología
Secretaría de Ciencia, Humanidades, Tecnología e Innovación



CentroGeo
Centro de Investigación en
Ciencias de Información Geoespacial, A.C.

Diseño y aplicación metodológica para la definición de límites de ciudades, mediante la **Teoría de la Percolación**

Jorge Alberto Montejano Escamilla
Camilo Alberto Caudillo Cos
Rodrigo Tapia McClung
Felipe Gerardo Ávila Jiménez

Primera edición, 2026
Derechos patrimoniales en trámite.
Publicación no comercializable y de distribución gratuita.
Prohibida su venta.

D.R. © 2026.
Centro de Investigación en Ciencias de Información Geoespacial, A.C.
Contoy, número 137, Colonia Lomas de Padierna, Código Postal
14240, Alcaldía Tlalpan, Ciudad de México, México.
www.centrogeo.org.mx

Diseño y producción: Unidad de Diseño y Comunicación Visual del
CentroGeo

Diseño de interiores: Unidad de Diseño y Comunicación Visual del
CentroGeo

Diseño de portada: Unidad de Diseño y Comunicación Visual del
CentroGeo

Se permite la reproducción total o parcial de esta obra, en cualquier
forma o medio, con propósitos educativos y sin fines de lucro.

Impreso y hecho en México/Printed and made in Mexico

Coordinación General de la Publicación

Comité Editorial del Centro de Investigación en Ciencias de
Información Geoespacial, A.C.

Responsables del Proceso de Publicación

Subcomité Editorial integrado por:

Mauricio Pablo Cervantes Salas
Aleajndra Aurelia Lopez Caloca
Daniela Alejandra Moctezuma Ochoa
Angelina Espejel Trujillo
Jorge Omar López Martínez

Directorio

Pablo López Ramírez

Director General del CentroGeo

Jorge Alberto Montejano Escamilla

Investigador Titular C del CentroGeo

Mauricio Pablo Cervantes Salas

Presidente del Comité Editorial del CentroGeo

Landis Córdova de la Cruz

Secretaria General del CentroGeo

Contenido

Tabla de contenido

Contenido	2
Introducción	4
I. Contexto	6
II. Bases conceptuales	8
La dimensión Fractal y la Entropía	9
Dimensión fractal	9
Entropía	11
Información y entropía	11
Teoría de la percolación	13
Historia	14
Cálculo del parámetro crítico	15
Universalidad	15
Etapas	16
Subcrítico y supercrítico	16
Criticidad	17
Diferentes modelos	17
Percolación para identificación de áreas urbanas	18
Ciudades Naturales	21
Metodología para delinear ciudades	28
El Grado de Urbanización nivel 1	28
Paso 1: Clasificación de cuadrícula	29
Paso 2: Clasificar las unidades locales	29
El Grado de Urbanización nivel 2	30
Paso 1: Clasificación de cuadrícula	30
Paso 2: Clasificación de la Unidad Local	31
Apuntes para la definición de ciudad	32
Conceptualizaciones y definiciones	32
Características esenciales	34
Hacia una definición sintética de ciudad	36
Formas de medir las ciudades	36
África	36
Norte América	37
América del Sur	38
Asia	38

Europa	40
Oceanía.....	41
De los conceptos y métodos utilizados para la delimitación de las ciudades	41
Evaluación de algoritmos de agrupación.....	42
Descripción	42
Argumentos.....	42
Detalles.....	43
Valor	44
Clustering basado en densidad DBSCAN y variantes	44
Potenciales aplicaciones	45
Métodos de clustering.....	45
Entidades mínimas por clúster	46
Distancia de búsqueda (DBSCAN y OPTICS).....	47
Clustering en espacio y tiempo (DBSCAN y OPTICS).....	49
Sensibilidad de clúster (OPTICS)	52
Comparación de los métodos.....	52
Resultados	56
III. Enfoque metodológico (inicial y ajustes).....	57
IV. Minería de datos y procesamiento de las geobases de datos	59
V. Aplicación del método y resultados	61
Comparación entre el SUN y resultados de DBSCAN.....	67
Comparación ciudad por ciudad.....	69
Comparación con el índice de GINI (pureza).....	71
Índice de Jaccard.....	72
Obtención de los rangos	72
Rangos y ley de Zipf de cada conjunto	76
VI. Conclusiones	78
Bibliografía	79

Introducción

Este Documento de Trabajo es un prueba piloto del proyecto editorial del Centro de Investigación en Ciencias de Información Geoespacial (CentroGeo), y surge del Proceso Prioritario 4 (PP4), denominado: “Lanzar la serie “Documentos de Trabajo y Reportes Técnicos” para difundir resultados preliminares y metodologías de proyectos de investigación e incidencia”.

El objetivo planteado para el desarrollo de esta prueba piloto, fue explorar el proceso de institucionalización y estandarización editorial para la publicación de la serie *Documentos de Trabajo y Reportes Técnicos* de CentroGeo, y a través de ello, recopilar la experiencia que nos permita avanzar en la calidad de los contenidos que se presentan en este tipo de publicaciones, así como en definir la identidad visual de la serie, teniendo en cuenta los procesos a seguir desde su recepción y evaluación inicial, hasta su publicación, difusión y monitoreo de impacto. El propósito final, es lograr que cada Documento de trabajo se convierta en un producto académico que permita la comunicación interna de los avances de investigación y cuente con potencial de incidencia en comunidades científicas, educativas y de política pública, fortaleciendo así la proyección externa del proyecto editorial de CentroGeo.

En el marco de esta prueba piloto, el Comité Editorial de CentroGeo conformó un subcomité que, a partir de un pequeño grupo de tres documentos propuestos por dos investigadores, decidió desarrollar este piloto con el trabajo que aquí se presenta. El documento recibido, originalmente titulado “Diseño metodológico para la definición de límites de ciudades, mediante la Teoría de la Percolación”, representa el informe final de la propuesta de investigación presentada y aprobada en el marco de la Convocatoria de Ciencia Básica y/o Ciencia de Frontera Modalidad: Paradigmas y Controversias de la Ciencia 2022 emitida por la de la Comisión Nacional de Ciencia y Tecnología (CONACYT).¹

Este Documento de trabajo, presenta la propuesta de investigación que dicho proyecto persiguió, es decir, el desarrollo de un método parsimonioso, que apoye la determinación de los límites de las ciudades mexicanas y compare los resultados con la delimitación de las 401 ciudades consideradas en el Sistema Urbano Nacional del año 2018 (SUN). El método incluye una serie de procesos y criterios que con apoyo de las geotecnologías permite automatizarlo y disminuir la carga subjetiva. La teoría de percolación –derivada de la física–, y la teoría de la Entropía Máxima (Emax) sirvieron de base para el desarrollo metodológico, cuyos criterios cuantitativos fueron de gran utilidad. En términos metodológicos, se revisaron materiales científicos que abordan temas relacionados con las bases teóricas señaladas anteriormente.

El trabajo de Cao *et.al.* (2020), denominado: Cuantificación de áreas urbanas con datos de múltiples fuentes basados en la teoría de la percolación. El concepto de “Ciudades Naturales” es desarrollado por Bin Jiang y Yufan Miao, (2014). Concepto retomado por Lingbo Liu *et.al.* (2019), y expuesto en su trabajo “Delimitación de la Ciudad Natural con Puntos de Interés en función del Área de Servicio y Método de Máxima Entropía”. Trabajos que fueron de gran utilidad; además de literatura relacionada con el concepto de ciudad, con las teorías de la Percolación, de la Entropía y sobre los algoritmos: City Clustering Algorithm (CCA) y el

¹ Mediante el acuerdo Pp F003 5/VIII-E/2022 de la 8ª Sesión Extraordinaria 2022 del Comité Técnico y de Administración (CTA) del Programa F003 “Programas nacionales estratégicos de ciencia, tecnología y vinculación con los sectores social, público y privado”, dicho Comité autoriza el proyecto:

Density-based spatial clustering of applications with noise (DBSCAN), utilizados en algunas de las investigaciones antes referidas.

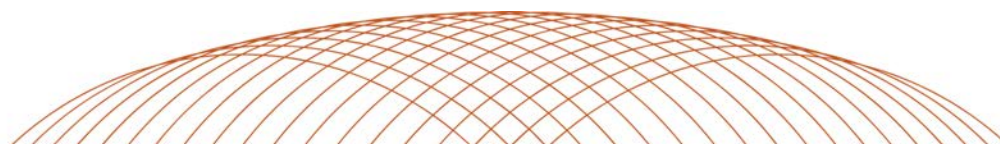
Los resultados obtenidos de los algoritmos y sobre todo del DBSCAN, sirvieron de base para la obtención de los criterios, mediante la percolación de puntos para determinar los umbrales de la clasificación (urbana-rural); para cuatro bases de datos (capas geográficas de puntos) de cobertura nacional. La obtención de la Entropía Máxima (E-Máx) para cada conjunto de puntos geográficos, permitió ajustar los umbrales, filtrar los cúmulos en función de la E-Máx y poder validarlos con el Sistema Urbano Nacional del año 2018 (SUN). La validación consideró los puntos filtrados de los conjuntos procesados, para el cálculo del índice de Jaccard, la coincidencia espacial y la Ley de Zipf.

Los resultados obtenidos permitieron: 1. Identificar la E-Máx, para cada uno de los conjuntos de puntos y tener un criterio para filtrar, 2. La similitud en términos de agrupación y correspondencia espacial de puntos, entre cada una de las cuatro bases geográficas y el SUN fue muy alta; y 3. Se determinó con base en la comparación de jerarquía de las 401 ciudades del SUN, una alta correspondencia con la jerarquía de los cúmulos de puntos percolados a su E-Máx.

Palabras clave: Percolación, Límites Urbanos, DBSCAN, Máxima Entropía, Ley de Zipf.

Finalmente, el trabajo que se presenta sirvió de base para la redacción de dos artículos ya publicados:

- Montejano Escamilla, J. A., Caudillo Cos, C. A., & Ávila Jiménez, F. G. (2023). Estimación de los usos del suelo urbano a nivel nacional. *Realidad, Datos y Espacio. Revista Internacional de Estadística y Geografía*, 14(1).
- Caudillo-Cos, C. A., Montejano-Escamilla, J. A., Tapia-McClung, R., Ávila-Jiménez, F. G., & Barrera-Alarcon, I. G. (2024). Defining urban boundaries through DBSCAN and Shannon's entropy: The case of the Mexican National Urban System. *Cities*, 149, 104969.



I. Contexto

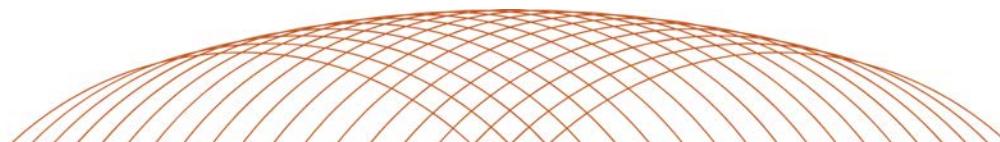
La determinación de lo que es urbano y lo que no lo es no es una tarea trivial. Diferentes grupos de investigación en México y en el mundo han buscado diversas formas de, por ejemplo, delimitar Áreas Metropolitanas con base en la interdependencia funcional de los asentamientos humanos y mediante la aplicación de costosas encuestas origen-destino, sin alcanzar un consenso.

En 2015, colegas del Center for Advanced Spatial Analysis (CASA), University College London encontraron una relación muy prometedora entre la dimensión fractal obtenida a través de la teoría de la percolación y cómo el sistema urbano británico está organizado jerárquicamente (Arcaute et al., 2016). En su trabajo, utilizaron la teoría de la percolación sobre la red de calles de Gran Bretaña, estableciendo jerarquías en diferentes umbrales de percolación. Varias regiones se formaron con cada distancia de percolación (es decir, varios pequeños "parches" surgieron cuando el proceso de percolación se calculó a una distancia muy baja, por ejemplo, 10 m, y se formó un solo grupo cuando el proceso de percolación se calculó a la distancia máxima de percolación). Pero lo más relevante fue que al calcular la dimensión fractal de los cúmulos extraídos, "observaron que la dimensión fractal presenta un máximo sobre todos los diferentes umbrales de distancia [y que] los cúmulos obtenidos en este máximo se corresponden en gran medida con la definición morfológica de las ciudades dada por imágenes de satélite" (p. 1). En otras palabras, encontraron un umbral en el que los conglomerados emergentes pueden identificarse como "ciudades" al analizar su dimensión fractal.

La percolación es un modelo de probabilidad cuyas raíces se pueden encontrar en problemas relacionados con la física. Stauffer y Aharony (1994) proporcionan una explicación simple de la teoría de la percolación. Supongamos que en una gran malla reticular, determinados cuadrados se llenan de puntos en sus centros mientras que otros se dejan vacíos. Al hacerlo, podemos observar puntos más cercanos a otros, formando así grupos. Estos grupos se forman cuando los puntos tienen al menos un cuadrado vecino que tiene un lado en común. Estos patrones observados —en número y propiedades— se encuentran entre los temas de estudio de la teoría de la percolación. La teoría de la percolación trata principalmente de la probabilidad de que un cuadrado esté ocupado por un punto. Como afirman Stauffer & Aharony: "La teoría de la percolación se ocupa de los cúmulos que se forman, en otras palabras, de los grupos de sitios vecinos ocupados" (1994, 3).

Por otro lado, las ciudades son sistemas complejos y, por tanto, su geometría también lo es (Zarza, 1996). Dado que su dinámica está relacionada con la no linealidad (ver Bettencourt et al., 2007), se podría decir que su forma es el resultado de procesos caóticos que, al mismo tiempo, pueden asociarse a una geometría fractal (p. 49). Dado que las ciudades son sistemas complejos, y su geometría puede medirse con métricas fractales (aptas para geometrías no euclidianas), se acude a la métrica de la dimensión fractal del sistema urbano para caracterizar su grado de complejidad.

Autores anteriores y posteriores han estado utilizando la teoría de la percolación para: a) comprender si los patrones de crecimiento urbano están correlacionados con la percolación (i.e., Makse et al., 1998); y b) encontrar límites urbano-rurales que, al final, definan los límites de la ciudad (i.e., Cao et al., 2020). Particularmente, Cao y colegas percolaron diez sistemas urbanos en todo el mundo para obtener un umbral de área urbana óptimo utilizando tres fuentes de datos: población, red de carreteras y luces nocturnas. Luego, validaron sus resultados



comparando las áreas urbanas derivadas de la percolación de esas fuentes de datos al valor máximo de la entropía de Shannon de cada conjunto de datos con las imágenes de Landsat, obteniendo un muy buen resultado con los datos de referencia. En esta investigación, lo relevante es que calcularon la entropía de Shannon para cada sistema de conglomerados, encontrando que la entropía alcanza un máximo alrededor del punto crítico, el cual coincide con el punto crítico de la dimensión fractal.

En CentroGeo (Tapia-McClung et al., s.f.) hemos estado investigando si el mismo fenómeno sucede en el Sistema Urbano Mexicano, encontrando que se cumple una máxima similitud de lo que nosotros llamamos “urbano” con la red nacional de calles percolada en el punto máximo de su dimensión fractal y de su entropía. Lo anterior, sugiere que existe una relación entre lo que es urbano y ambas medidas del sistema urbano nacional (entropía y dimensión fractal) que representan un grado de complejidad determinado, lo que a su vez puede derivar en un método simple y menos subjetivo para definir lo que es urbano y lo que no, simplemente encontrando la máxima dimensión fractal del sistema y la máxima entropía de Shannon.

II. Bases conceptuales

Se presentan los temas de interés para esta investigación, iniciando con los fundamentos de la Teoría de Fractales, la Entropía Teoría y de la Percolación; seguido de los trabajos identificados en la literatura nacional e internacional, relacionados con la delimitación metodológica de ciudades. Además, se consideró necesario analizar el concepto de ciudad en el contexto internacional para identificar la concepción prevaleciente en México. También se revisaron los algoritmos utilizados para la percolación y clusterización.

La dimensión Fractal y la Entropía

Dimensión fractal ²

La dimensión fractal se basa en la **dimensión Hausdorff $H(X)$** de un objeto fractal **X** y mide el número **N** de conjuntos de longitud **L** que hacen falta para cubrir **X** utilizando objetos de esa longitud **L**.

Si comenzamos con un segmento de longitud 1, y lo dividimos en segmentos de longitud **L** obtendremos $N(L)$ partes, de manera que

$$N(L) \cdot L^1 = 1$$

Cualquiera que sea L :

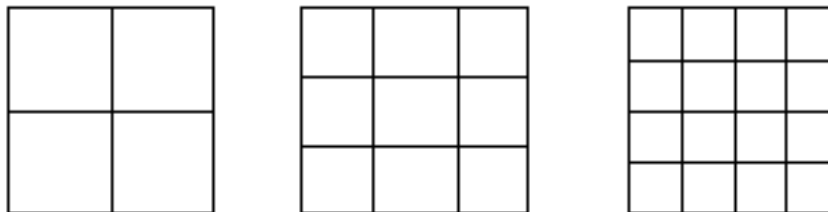


Esta relación está clara en el caso de un segmento rectilíneo. El exponente al que va elevado L nos da la dimensión, en este caso $d = 1$.

Si el objeto inicial es un cuadrado de superficie 1, y lo comparamos con unidades cuadradas, cuyo lado tenga de longitud L , el número de unidades que es necesario para recubrirlo $N(L)$, cumple

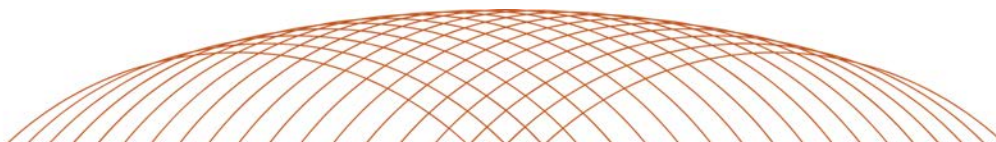
$$N(L) \cdot L^2 = 1$$

Cualquiera que sea L . En este caso $d = 2$



En general,

² <https://marzomates.webs.ull.es/fractales-la-geometria-del-caos/#Dimension>



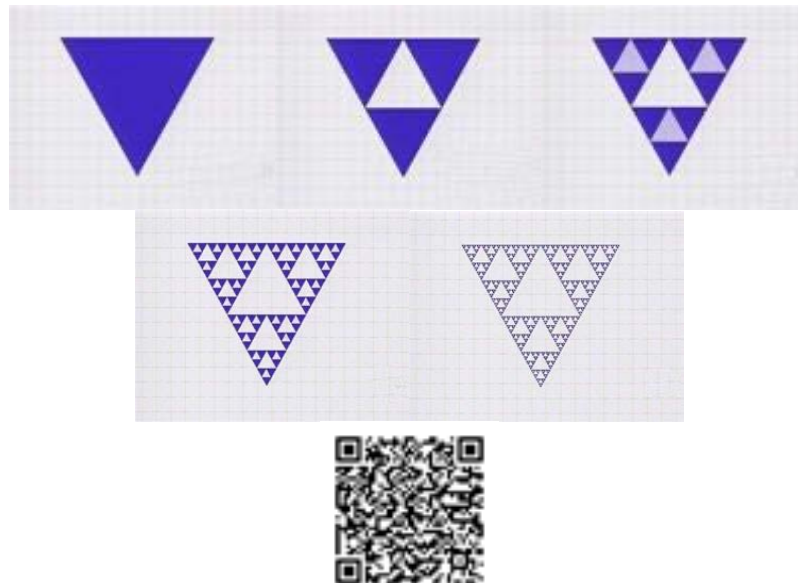
La dimensión fractal de un objeto geométrico es d si

$$N(L) \cdot L^d = 1$$

Es decir,

$$d = \frac{\log[N(L)]}{\log\left(\frac{1}{L}\right)}$$

Veamos la dimensión del triángulo de Sierpinski.



Triángulo de Sierpinski

Observa la segunda figura: el lado de cada triángulo es $L = 1/2$; para cubrir la figura necesitamos 3 triángulos, es decir $N(L) = 3$

$$d = \frac{\log(3)}{\log(2)} = 1.584 \dots$$

Entropía ³

En física y química a:

- Entropía termodinámica, una magnitud que mide la parte de la energía que no puede utilizarse para producir un trabajo; es el grado de desorden que poseen las moléculas que integran un cuerpo, o también el grado de irreversibilidad alcanzada después de un proceso que implique transformación de energía.
- Entropía de formación, la diferencia de entropía o desorden en el proceso de formación de sustancias a partir de sus elementos constituyentes.
- Entropía de Kolmogórov objeto o dimensión estudiado en la física y matemáticas a partir de las homotecias (un tipo de transformación geométrica).

En astrofísica y cosmología: a la entropía de los agujeros negros.

En lingüística y semiótica: es el nivel o grado de información discursiva frecuentemente ponderado por la cantidad de lexemas. Así se considera que un discurso con muchos neologismos es más entrópico que uno con pocos neologismos (notar que el mayor grado de neologismos puede aportar más información pero también —si es exagerado— caos en la información).

En teoría de la información a:

- Entropía en la información, el grado de incertidumbre que existe sobre un conjunto de datos.
- Entropía de Shannon.
- Entropía de Alfred Rényi; (d) Entropía condicional.

En ecología la entropía es una medida asociada a la biodiversidad.

En matemáticas a:

- Entropía topológica, la correspondiente a la cantidad real asociada a todo sistema topológicamente dinámico.
- Entropía métrica, la correspondiente a la cantidad real asociada a todo sistema dinámico medible.

Información y entropía

Los conceptos de información y entropía están ampliamente relacionados entre sí, aunque se tardó años en el desarrollo de la mecánica estadística y la teoría de la información para hacer esto aparente. Nos centraremos en la entropía desde la teoría de la información. Esta entropía se llama frecuentemente entropía de Shannon, en honor a Claude E. Shannon.

El concepto básico de entropía en teoría de la información tiene mucho que ver con la incertidumbre que existe en cualquier experimento o señal aleatoria. Es también la cantidad de "ruido" o "desorden" que contiene o libera un sistema. De esta forma, podremos hablar de la cantidad de información que lleva una señal.

Como ejemplo, consideremos algún texto escrito en español codificado como una cadena de letras, espacios y signos de puntuación (nuestra señal será una cadena de caracteres). Ya que, estadísticamente, algunos caracteres no son muy comunes (por ejemplo, 'y'), mientras

3

https://guiasjuridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAAAAAAAAEAMtMSbF1jTAAASNjU3MDtBLUOuLM_DxbIwMDS0NDA1OQQGZapUt-ckhlQaptWmJOcSoAHWB9zUAAAA=WKE

otros sí lo son (como la 'a'), la cadena de caracteres no será tan "aleatoria" (o equiprobable) como podría llegar a ser. Obviamente, no podemos predecir con exactitud cuál será el siguiente carácter en la cadena, y eso la haría aparentemente aleatoria. Pero es la entropía la encargada de medir precisamente esa aleatoriedad, y fue presentada por Shannon en su artículo de 1948 "A Mathematical Theory of Communication".

Shannon ofrece una definición de entropía que satisface las siguientes afirmaciones:

La medida de información debe ser proporcional (continua). Es decir, el cambio pequeño en una de las probabilidades de aparición de uno de los elementos de la señal debe cambiar poco la entropía.

Si todos los elementos de la señal son equiprobables a la hora de aparecer, entonces la entropía será máxima.

Ejemplos de máxima entropía pueden ser: Suponiendo que estamos a la espera de un texto, por ejemplo un cable con un mensaje. En dicho cable solo se reciben las letras en minúscula de la (a hasta la z), entonces si el mensaje que nos llega es "qalmbphijcdgketrsvfxyzwño" el cual posee una longitud de 27 caracteres, se puede decir que este mensaje llega a nosotros con la máxima entropía (o desorden posible) ya que es poco probable que se pueda pronosticar la entrada de caracteres ya que estos no se repiten y además no están ordenados en una forma predecible.

Formalmente, la entropía nos indica el límite teórico para la compresión de datos. También es una medida de la información contenida en el mensaje. Su cálculo se realiza mediante la siguiente fórmula:

Una variable aleatoria χ puede tomar distintos valores posibles en distintas repeticiones (realizaciones del experimento). Como algunos valores de χ son más probables que otros, existe una distribución de probabilidad de los valores de χ , la cual depende del experimento elegido.

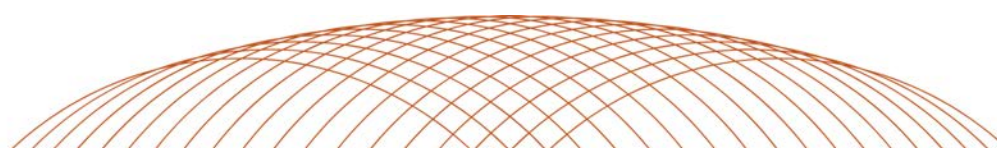
La entropía asociada a la variable χ es un número que depende directamente de la distribución de probabilidad de χ e indica lo plana que es esa distribución. Una distribución es plana (tiene alta entropía) cuando todos los valores de χ tienen probabilidades similares, mientras que es poco plana cuando algunos valores de χ son mucho más probables que otros. En una distribución de probabilidad plana (con alta entropía) es difícil poder predecir cuál es el próximo valor de χ que va a presentarse, ya que todos los valores de χ son igualmente probables.

Si a cada posible valor de χ que pueda ocurrir se le asigna una cierta combinación de dígitos binarios 0 o 1 para diferenciarlo de los demás, la cantidad promedio de dígitos binarios que hay que asignarle a los distintos valores de χ es (aproximadamente) igual a la entropía de la distribución de probabilidad de χ . Los valores 0 o 1 usados suelen llamarse bits. Además, la metodología comúnmente usada para asignar combinaciones de valores 0 o 1 a los distintos valores posibles de χ se conocen con el nombre de codificación Huffman.

Así, la entropía definida por Shannon, referida a la teoría de la información, hace referencia a la cantidad media de información que contiene una variable aleatoria o, en particular, una fuente de transmisión binaria. La información que aporta un determinado valor χ_i de una variable aleatoria discreta χ se define como:

$$I(\chi_i) = \log_2 (1/p(\chi_i))$$

Cuya unidad es el bit utilizándose el logaritmo en base 2 (por ejemplo, cuando se emplea el logaritmo neperiano se habla de nats). La entropía o información media de la variable



aleatoria discreta, χ , se determina como la información media del conjunto de valores discretos que puede adoptar (también medida en bits):

$$H(\chi) = \sum p(\chi_i) \times \log_2 (1/p(\chi_i))$$

Además de su definición y estudio, Shannon demostró analíticamente que la entropía es el límite máximo al que se puede comprimir una fuente sin ninguna pérdida de información.

Finalmente, una mención especial a la entropía de Kolmogórov, que se define como principio que mide la pérdida de información a lo largo de la evolución del sistema. También es definida como la suma de exponentes de Liapunov. Esta entropía tiene una gran importancia en su aplicación a sistemas de los que no se dispone más que de series temporales de valores de una determinada variable. Su fórmula matemática viene definida por subconjuntos, de tal manera que:

$$H(C) = \log(N)$$

El tamaño es de $2n$ subconjuntos. Si lo definimos en sistema binario, este tiene una extensión de N bits.

Una fórmula alternativa de la entropía de Kolmogórov es:

$$H(C) = \log(L/2e)$$

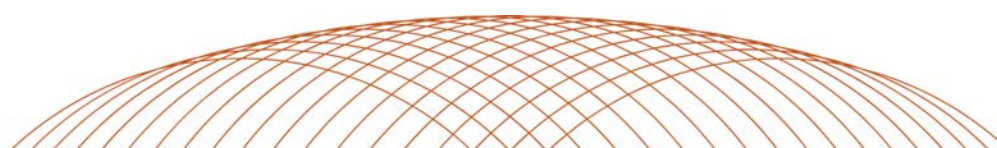
Ideada también por el matemático Kolmogórov, después de que Claude E. Shannon fuera tachado de demasiado orientado a la tecnología.

Teoría de la percolación ⁴

En física estadística y matemáticas, la teoría de la percolación describe el comportamiento de una red cuando se agregan nodos o enlaces. Este es un tipo geométrico de transición de fase, ya que en una fracción crítica de la adición, la red de pequeños grupos desconectados, se fusionan en un grupo conectado significativamente más grande, denominado grupo de expansión. Las aplicaciones de la teoría de la percolación a la ciencia de los materiales y en muchas otras disciplinas se analizan aquí y en los artículos Teoría de redes y percolación.

Una pregunta representativa (y la fuente del nombre) es la siguiente. Suponga que se vierte un poco de líquido sobre un material poroso. ¿Podrá el líquido pasar de un agujero a otro y llegar al fondo? Esta cuestión física se modela matemáticamente como una red tridimensional de $n \times n \times n$ vértices, generalmente llamados "sitios", en los que el borde o los "enlaces" entre cada dos vecinos pueden estar abiertos (permitiendo el paso del líquido) con probabilidad p , o cerradas con probabilidad $1 - p$, y se supone que son independientes. Por lo tanto, para un p dado, ¿Cuál es la probabilidad de que exista un camino abierto (es decir, un camino, cada uno de cuyos enlaces es un enlace "abierto") de arriba hacia abajo? El comportamiento para n grande es de interés primordial. Este problema, llamado ahora percolación de enlaces, fue introducido en la literatura matemática por Broadbent & Hammersley (1957), y ha sido estudiado intensamente por matemáticos y físicos desde entonces.

⁴ <https://academia-lab.com/enciclopedia/teoria-de-la-percolacion/>



En un modelo matemático ligeramente diferente para obtener un gráfico aleatorio, un sitio está "ocupado" con probabilidad p o "vacío" (en cuyo caso se eliminan sus bordes) con probabilidad $1 - p$; el problema correspondiente se llama percolación del sitio. La pregunta es la misma: para un p dado, ¿Cuál es la probabilidad de que exista un camino entre arriba y abajo? De manera similar, uno puede preguntar, dado un gráfico conectado, en qué fracción $1 - p$ de fallas el gráfico se volverá desconectado (sin un componente grande).

Se pueden hacer las mismas preguntas para cualquier dimensión de red. Como es bastante típico, en realidad es más fácil examinar redes infinitas que solo redes grandes. En este caso la pregunta correspondiente es: ¿Existe un cúmulo abierto infinito? Es decir, ¿Hay un camino de puntos conectados de longitud infinita "a través" de la red? Según la ley cero-uno de Kolmogorov, para cualquier p dada, la probabilidad de que exista un grupo infinito es cero o uno.

Dado que esta probabilidad es una función creciente de p (prueba a través del argumento de acoplamiento), debe haber un p crítico (denotado por p_c) por debajo del cual la probabilidad es siempre 0 y por encima del cual la probabilidad es siempre 1. En la práctica, esta criticidad es muy fácil de observar. Incluso para n tan pequeño como 100, la probabilidad de un camino abierto desde la parte superior hasta la inferior aumenta bruscamente desde muy cerca de cero hasta muy cerca de uno en un intervalo corto de valores de p .

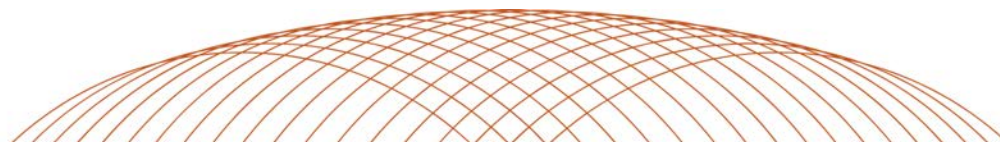
Historia

La teoría de Flory-Stockmayer fue la primera teoría que investigó los procesos de percolación. La historia del modelo de percolación tal como lo conocemos tiene sus raíces en la industria del carbón. Desde la revolución industrial, la importancia económica de esta fuente de energía impulsó numerosos estudios científicos para comprender su composición y optimizar su uso. Durante los años 30' y 40', el análisis cualitativo por química orgánica dejó cada vez más espacio a estudios más cuantitativos.

En este contexto, se creó en 1938 la British Coal Utilization Research Association (BCURA). Se trata de una asociación de investigación financiada por los propietarios de las minas de carbón. En 1942, Rosalind Franklin, recién licenciada en química por la Universidad de Cambridge, se incorporó al BCURA. Inició investigaciones sobre la densidad y porosidad del carbón. Durante la Segunda Guerra Mundial, el carbón fue un importante recurso estratégico. Se utilizó como fuente de energía, pero también fue el componente principal de las máscaras antigás.

El carbón es un medio poroso. Para medir su densidad 'real', se sumergía en un líquido o un gas cuyas moléculas son lo suficientemente pequeñas como para llenar sus poros microscópicos. Al tratar de medir la densidad del carbón utilizando varios gases (helio, metanol, hexano, benceno) y al encontrar diferentes valores según el gas utilizado, Rosalind Franklin demostró que los poros del carbón están hechos de microestructuras de varias longitudes que actúan como un tamiz microscópico para discriminar los gases. También descubrió que el tamaño de estas estructuras depende de la temperatura de carbonatación durante la producción del carbón. Con estas investigaciones obtuvo el doctorado y dejó el BCURA en 1946.

A mediados de los años cincuenta, Simon Broadbent trabajaba en el BCURA como estadístico. Entre otros intereses, estudió el uso del carbón en máscaras de gas. Una pregunta es entender cómo un fluido puede difundirse en los poros del carbón, modelado como un laberinto aleatorio de túneles abiertos o cerrados. En 1954, durante un simposio sobre métodos



Monte Carlo, le hace preguntas a John Hammersley sobre el uso de métodos numéricos para analizar este modelo.

Broadbent y Hammersley introdujeron en su artículo de 1957 un modelo matemático para modelar este fenómeno, que es la percolación.

Cálculo del parámetro crítico

Para la mayoría de los gráficos de celosía infinita, P_c no se puede calcular con exactitud, aunque en algunos casos p_c tiene un valor exacto. Por ejemplo:

- Para la red cuadrada $\mathbb{Z}$ en dos dimensiones, $P_c = 1/2$ para la percolación de bonos, un hecho que fue una pregunta abierta durante más de 20 años y finalmente fue resuelto por Harry Kesten a principios de la década de 1980, ver Kesten (1982). Para la percolación del sitio, el valor de P_c no se conoce a partir de la derivación analítica, sino solo a través de simulaciones de redes grandes.
- Un caso límite para celosías de grandes dimensiones viene dado por la celosía de Bethe, cuyo umbral está en $P_c = 1/z - 1$ para un número de coordinación z . En otras palabras: para el árbol regular de grado z , P_c es igual a $1/(z-1)$.
- Para una red aleatoria tipo árbol sin correlación grado-grado, se puede demostrar que dicha red puede tener un componente gigante, y el umbral de percolación (probabilidad de transmisión) está dado por

$$p_c = \frac{1}{g'_1(1)},$$

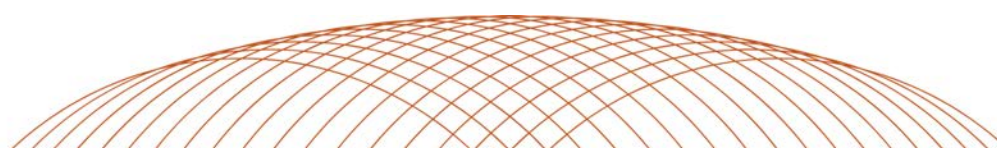
- donde $g_1(z)$ es la función generadora correspondiente a la distribución de exceso de grado. Entonces, para redes Erdős-Rényi aleatorias de grado medio $\langle k \rangle$, $p_c = 1/\langle k \rangle$.
- En redes con agrupamiento bajo $0 < C \ll 1$, el punto crítico se escala de $(1 - C)^{-1}$ tal manera que:

$$p_c = \frac{1}{1 - C} \frac{1}{g'_1(1)}.$$

Esto indica que para una distribución de grado dada, el agrupamiento conduce a un umbral de filtración mayor, principalmente porque para un número fijo de enlaces, la estructura de agrupamiento refuerza el núcleo de la red con el precio de diluir las conexiones globales. Para redes con alta agrupación, la agrupación fuerte podría inducir la estructura de núcleo y periferia, en la que el núcleo y la periferia podrían filtrarse en diferentes puntos críticos, y el tratamiento aproximado anterior no es aplicable.

Universalidad

El principio de universalidad establece que el valor numérico de p_c está determinado por la estructura local del gráfico, mientras que el comportamiento cerca del umbral crítico, p_c , se caracteriza por exponentes críticos universales. Por ejemplo, la distribución del tamaño de los conglomerados en la criticidad decae como una ley de potencia con el mismo exponente para



todos los retículos 2d. Esta universalidad significa que para una dimensión dada, los diversos exponentes críticos, la dimensión fractal de los grupos en p es independiente del tipo de red y del tipo de percolación (p. ej., enlace o sitio). Sin embargo, recientemente se ha realizado una percolación sobre una red estocástica plana ponderada (WPSL) y se ha encontrado que aunque la dimensión de la WPSL coincide con la dimensión del espacio donde está incrustada, su clase de universalidad es diferente a la de todas las redes planas conocidas.

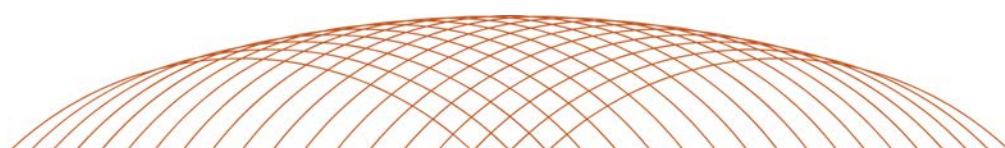
Etapas

Subcrítico y supercrítico

El hecho principal en la fase subcrítica es el "decaimiento exponencial". Es decir, cuando $p < p_c$, la probabilidad de que un punto específico (por ejemplo, el origen) esté contenido en un grupo abierto (es decir, un conjunto máximo conectado de bordes "abiertos" del gráfico) de tamaño r decae a cero exponencialmente en r — Menshikov (1986) y de forma independiente Aizenman y Barsky (1987) demostraron esto para la percolación en tres y más dimensiones. En dos dimensiones, formaba parte de la demostración de Kesten de que $P_c = 1/2$.

El gráfico dual de la red cuadrada $\mathbb{Z}$ también es la red cuadrada. De ello se deduce que, en dos dimensiones, la fase supercrítica es dual a un proceso de percolación subcrítico. Esto proporciona esencialmente información completa sobre el modelo supercrítico con $d = 2$. El resultado principal para la fase supercrítica en tres y más dimensiones es que, para N suficientemente grande, existe un cúmulo abierto infinito en la losa bidimensional $\mathbb{Z} \times [0, N]$. Esto fue probado por Grimmett & Marstrand (1990).

En dos dimensiones con $p < 1/2$, existe con probabilidad uno un cúmulo cerrado infinito único (un cúmulo cerrado es un conjunto máximo conectado de bordes "cerrados" del gráfico). Así, la fase subcrítica puede describirse como islas abiertas finitas en un océano cerrado infinito. Cuando $p > 1/2$ ocurre justo lo contrario, con islas cerradas finitas en un océano abierto infinito. El cuadro es más complicado cuando $d \geq 3$ ya que $P_c < 1/2$, y hay coexistencia de infinitos cúmulos abiertos y cerrados para p entre P_c y $1 - P_c$.



Criticidad

La percolación tiene una singularidad en el punto crítico $p = P_c$ y muchas propiedades se comportan como una ley de potencia con $p - P_c$, cerca de P_c . La teoría del escalamiento predice la existencia de exponentes críticos, en función del número de dimensiones, que determinan la clase de la singularidad. Cuando $d = 2$, estas predicciones están respaldadas por argumentos de la teoría del campo conforme y la evolución de Schramm-Loewner, e incluyen valores numéricos predichos para los exponentes. La mayoría de estas predicciones son conjeturas excepto cuando el número d de dimensiones satisface $d = 2$ o $d \geq 6$. Incluyen:

- No hay clústeres infinitos (abiertos o cerrados)
 - La probabilidad de que haya un camino abierto desde algún punto fijo (digamos el origen) a una distancia de r decrece polinomialmente, es decir, es del orden de $r^{-\alpha_d}$ para algún α_d no depende de la red particular elegida, ni de otros parámetros locales. Depende solo de la dimensión d (este es un ejemplo del principio de universalidad). α_d disminuye desde $d = 2$ hasta $d = 6$ y luego permanece fijo.

$$\alpha_2 = -5/48$$

$$\alpha_6 = -1.$$

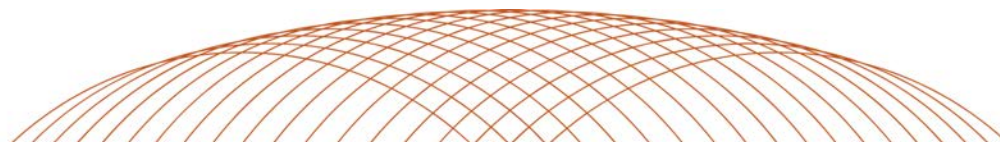
- La forma de un grupo grande en dos dimensiones es conformemente invariante.

Véase Grimmett (1999). En 11 o más dimensiones, estos hechos se prueban en gran medida utilizando una técnica conocida como expansión de encaje. Se cree que una versión de la expansión de encaje debería ser válida para 7 o más dimensiones, quizás con implicaciones también para el caso umbral de 6 dimensiones. La conexión de la percolación con la expansión del encaje se encuentra en Hara & Slade (1990).

En dos dimensiones, el primer hecho ("sin percolación en la fase crítica") se prueba para muchas redes, usando dualidad. Se ha logrado un progreso sustancial en la percolación bidimensional a través de la conjetura de Oded Schramm de que el límite de escala de un gran cúmulo puede describirse en términos de una evolución de Schramm-Loewner. Esta conjetura fue probada por Smirnov (2001) en el caso especial de percolación del sitio en la red triangular.

Diferentes modelos

- La percolación dirigida que modela el efecto de las fuerzas gravitatorias que actúan sobre el líquido también se introdujo en Broadbent & Hammersley (1957) y tiene conexiones con el proceso de contacto.
- El primer modelo estudiado fue la percolación de Bernoulli. En este modelo todos los enlaces son independientes. Los físicos llaman a este modelo percolación de enlaces.
- Luego se introdujo una generalización como el modelo de conglomerados aleatorios de Fortuin-Kasteleyn, que tiene muchas conexiones con el modelo de Ising y otros modelos de Potts.
- La percolación de Bernoulli (enlace) en gráficos completos es un ejemplo de un gráfico aleatorio. La probabilidad crítica es $p = 1/N$, donde N es el número de vértices (sitios) del gráfico.
- La percolación Bootstrap elimina las celdas activas de los clústeres cuando tienen muy pocos vecinos activos y analiza la conectividad de las celdas restantes.



- Percolación del primer paso.
- Percolación de invasión.

Percolación para identificación de áreas urbanas

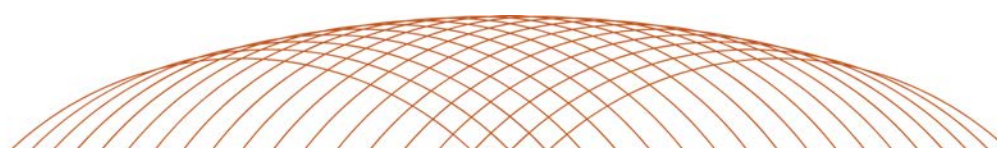
Para abordar el tema de la teoría de la percolación, se revisó el trabajo de Wenpu Cao, Lei Dong, Lun Wu y Yu Liu (2020). Denominado “Cuantificación de áreas urbanas con datos de múltiples fuentes basados en la teoría de la percolación”. Los autores argumentan que la delimitación de las áreas urbanas es crucial para abordar los problemas urbanos asociados, a los problemas ambientales y de sostenibilidad. En este sentido, señalan que se ha trabajado el tema con los datos de teledetección y especialmente las imágenes de luz nocturna. Las cuales se han utilizado ampliamente para delimitar áreas urbanizadas en todo el mundo. Además de algunos datos urbanos emergentes, como son: la información geográfica voluntaria (p. ej., OpenStreetMap); y los datos de detección social (p. ej., teléfonos móviles y redes sociales). Estos datos, también han mostrado un gran potencial para revelar los límites y las dinámicas urbanas. Sin embargo, los autores consideran que los métodos para cuantificar las áreas urbanas, al utilizar datos de múltiples fuentes cuestionan su consistencia y solidez.

Derivado de lo anterior, Cao et.al. (2020) proponen un método basado en la percolación, para extraer áreas urbanas utilizando datos urbanos de múltiples fuentes. Para ello, encontraron el umbral urbano/no urbano óptimo, que resultó de considerar la naturaleza crítica de los sistemas urbanos, con el apoyo de la teoría de la percolación. Además, aplicaron su método a tres conjuntos de datos de código abierto (población, carreteras y luz nocturna) para un total de 28 países. Con ello, demostraron que el método propuesto, logra capturar características urbanas similares en términos de áreas urbanas. Posteriormente, los conglomerados resultantes los contrastaron con los principios de la ley de Zipf, encontrando que se cumple en la mayoría de los países. Para evaluar la precisión de las áreas urbanas definidas por su método las validaron utilizando como referencia los datos de Landsat en 10 ciudades⁵.

La definición de las áreas urbanas es una preocupación de estos investigadores, los cuales consideran que las áreas urbanas unidades espaciales básicas para la planificación, los estudios urbanos y formulación de políticas públicas. Su importancia también radica en el surgimiento de un gran número de regiones de rápida urbanización en todo el mundo, aunada a la complejidad del sistema urbano y a la difusa transición urbano-rural. Además de señalar que muchas de la políticas se implementan en función de divisiones administrativas que responden a criterios históricos políticos y geográficos, situación que hace difícil analizar la dinámica socioeconómica de las ciudades. Situaciones que hacen casi imposible realizar comparaciones entre diferentes países y periodos. Ante esta realidad, señalan de manera oportuna que la construcción de datos de movilidad y de tipo censal para el seguimiento de las ciudades, resultan costosos y de mucho tiempo; además de la carencia de criterios estandarizados para definir lo urbano.

Parte de su aportación, radica en la utilidad de datos urbanos emergentes generados por personas considerados sensores. Al producto de esta actividad lo identifican como información

⁵ Cao y compañía eligieron China como principal país de estudio para demostrar la eficacia de su método. Lo aplicaron a 28 países para comprobar su universalidad y validaron los resultados en 10 países utilizando Landsat. Eligieron los países más grandes de cada continente, con un área no menor a 100,000 km², incluidos tanto los países desarrollados como los países en desarrollo.

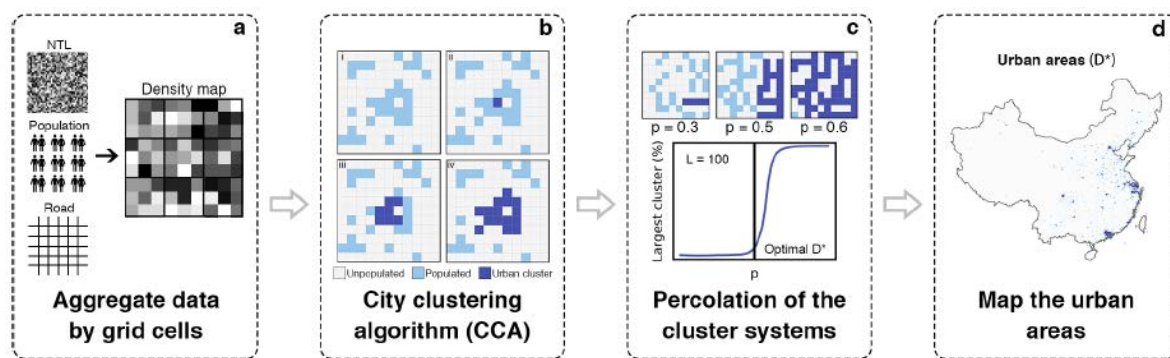


geográfica voluntaria (VGI, por ejemplo, Open-StreetMap) y los datos de detección social (por ejemplo, teléfonos móviles y redes sociales), los cuales también han mostrado un gran potencial en el acercamiento a la identificación de los límites socioeconómicos de las ciudades.

No solo reconocen las bondades de este tipo de información; también consideran algunos métodos nuevos para delimitar áreas urbanas y específicamente hacen referencia al City Clustering Algorithm (CCA). Este método desde el punto de vista de estos autores, es sencillo y eficiente para definir a las ciudades, a partir del reconocimiento de las áreas pobladas máximamente conectadas. No obstante, la transición entre lo urbano y lo rural sigue siendo difuso, más aún cuando se utilizan modelos univariados, que resultan insuficientes para reflejar la complejidad de las ciudades. Aunque señalan que dentro de esta complejidad se asoman patrones universales de auto organización, que pueden ser explicados por la Ley de Zipf, leyes de escala y/o de fractales.

Para desarrollar su trabajo, utilizaron tres conjuntos de datos: 1. luz nocturna (datos de detección remota) disponibles a través de <https://www.ngdc.noaa.gov/eog/viirs>; 2. Población (datos de detección social) <https://www.worldpop.org>; y 3. Redes de carreteras (datos VGI) disponibles a través de <https://download.geofabrik.de>.

Su algoritmo de agrupamiento de ciudades basado en la percolación (PCCA) incluye tres pasos. Primero, agregaron los datos urbanos a escala fina por $0.5' \times 0.5'$ celdas de cuadrícula para unificar diferentes conjuntos de datos. En segundo lugar, aplicaron el CCA para fusionar celdas de cuadrícula en grupos urbanos bajo cada umbral potencial. En tercer lugar, realizaron el análisis de percolación en los grupos detectados para encontrar el umbral óptimo y luego mapear las áreas urbanas en el umbral óptimo.



En el esquema anterior que expresa la esencia del método de agrupamiento de ciudades basado en percolación, se presentan de manera gráfica los siguientes pasos:

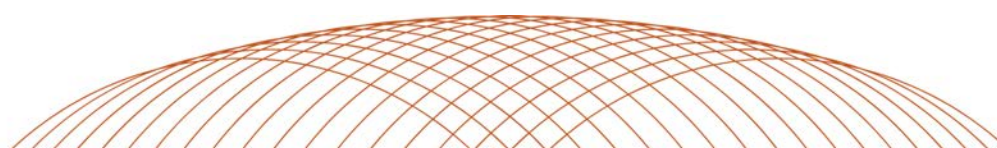
- a) Los datos urbanos de múltiples fuentes se agragan por celdas de cuadrícula.
- b) Usamos el CCA para fusionar unidades urbanas en un sistema de grupos bajo cada umbral potencial. CCA: i) Las celdas con un valor superior al umbral potencial se marcan como unidades urbanas (azul claro). ii) Se selecciona una celda urbana sin procesar para formar un nuevo grupo urbano. iii) Este grupo urbano agrega recursivamente las celdas urbanas más cercanas hasta que se hayan procesado todos los vecinos más cercanos. Usamos los ocho vecinos más cercanos en nuestra investigación.

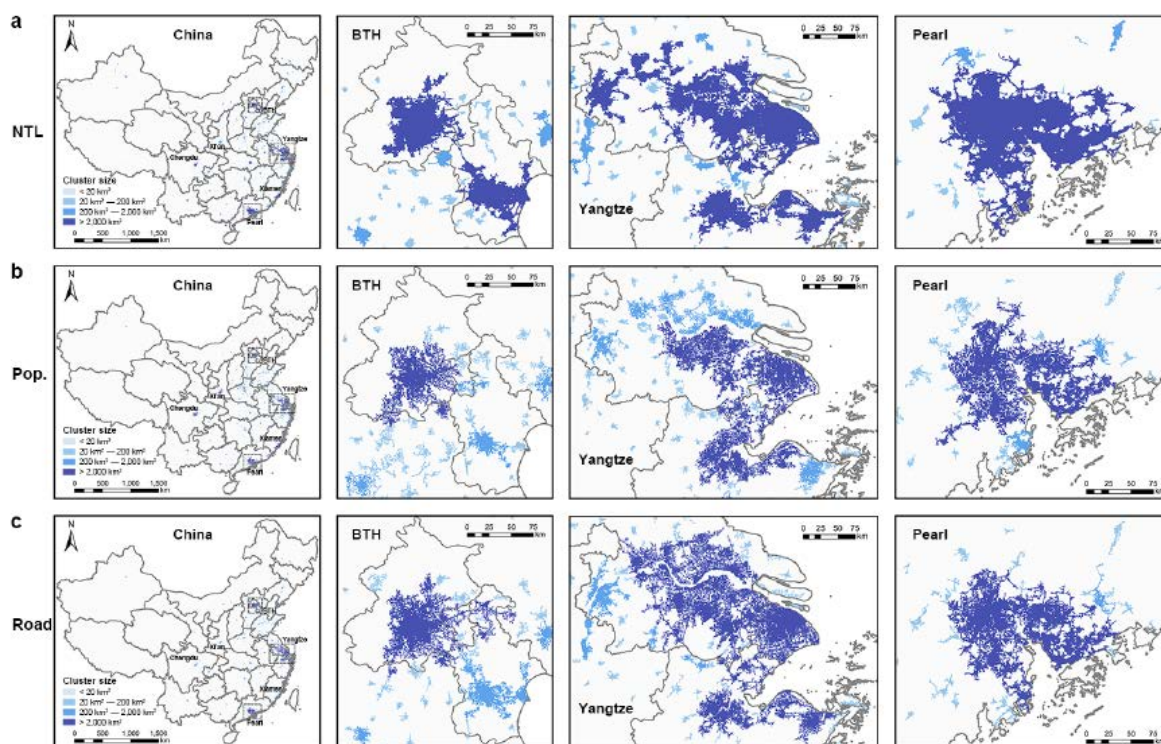
- iv) Se forma un grupo urbano (azul oscuro). El proceso continua hasta que todas las células urbanas pertenecen a un clúster urbano.
- c) Modelo de percolación del sitio bidimensional. En cuanto a un $L \times L$ en la malla, cada sitio puede ser ocupado con probabilidad p , y los sitios ocupados adyacentes forman un grupo (azul claro). Como p aumenta, el cúmulo más grande (azul oscuro) permanece estable al principio, pero rápidamente se convierte en uno gigante en el punto crítico. Consideramos cada umbral potencial como probabilidad p en percolación y encontrar el umbral óptimo D^* en este punto crítico.
- d) Áreas urbanas con umbral óptimo D^* .

Como resultado principal obtienen las áreas urbanas extraídas por PCCA, aplicando los siguientes pasos que se presentan de manera resumida: Primero aplicaron el PCCA a los conjuntos de datos de China. Siguiendo la sección de Métodos, se obtuvieron los mapas de densidad de población, intersecciones de carreteras y luz nocturna de China. Luego, se extrajeron los sistemas de clústeres bajo todos los umbrales potenciales y aplicaron el análisis de percolación a los sistemas de clústeres para encontrar el umbral óptimo. Con base en el tamaño mínimo de una ciudad en China, que equivale a 20 km², establecieron el criterio para filtrar aquellas áreas fragmentadas.

Encontraron el punto crítico, el cual indica una transición de fase continua. Este fenómeno refleja las características del sistema urbano dentro de la ciudad, son mucho más fuertes que las conexiones entre ciudades, las conexiones entre ciudades débiles se rompen a medida que aumentaron el umbral. También calcularon la Entropía de Shannon para cada conjunto de datos. Encontraron que las entropías en el punto crítico están cerca de los grupos de más de 20 km², que son 5.88 (población), 5.32 (carretera) y 5.34 (luz nocturna), lo que indica distribuciones de tamaño similares de áreas urbanas extraídas de diferentes fuentes de datos.

Mapearon las áreas urbanas obtenidas y evidenciaron que estas áreas rompen las restricciones geográficas de los límites administrativos y tienen conexiones altamente integradas. Sus resultados confirman altas integraciones.





Por último, generaron un análisis con la Ley de Zipf, una ley importante para la distribución del tamaño del sistema urbano, se cumple para su definición de áreas urbanas. La ley de Zipf refleja la naturaleza autoorganizada de los sistemas urbanos y la encontraron en la mayoría de los países analizados.

Fusionaron los tres resultados de las áreas urbanas, dado que se derivan de conjuntos de datos de múltiples fuentes, así generaron un mapa de área urbana fusionado para mejorar los resultados.

No lograron encontrar un umbral óptimo para el conjunto de países analizados, lo que si encontraron es que para cada país las entropías en el punto crítico en tres conjuntos de datos son similares, lo que indica distribuciones de tamaño similares de las áreas urbanas delineadas.

Ciudades Naturales

El concepto de “Ciudades Naturales” acuñado por Bin Jiang y Yufan Miao, publicaron en el 2014 su trabajo “La evolución de las ciudades naturales desde la perspectiva de las redes sociales basadas en la ubicación”. Su investigación examina el antiguo medio social basado en la ubicación *Brightkite*⁶, en un periodo de tres años. El concepto de ciudades naturales se refiere a eventos geográficos agrupados espacialmente, como los parches aglomerados agregados desde las ubicaciones de los usuarios individuales de las redes sociales. Aplicaron la regla de

⁶ Brightkite era un sitio web de redes sociales basado en la ubicación. Los usuarios pudieron "registrarse" en lugares mediante el uso de mensajes de texto o una de las aplicaciones móviles y pudieron ver quién está cerca y quién ha estado allí antes. El servicio fue creado en 2007 por Brady Becker, Martin May y Alan Seideman, quienes anteriormente fundaron el servicio de notificación por SMS Loopnote. En abril de 2009, Brightkite fue adquirida por la red social móvil Limbo.

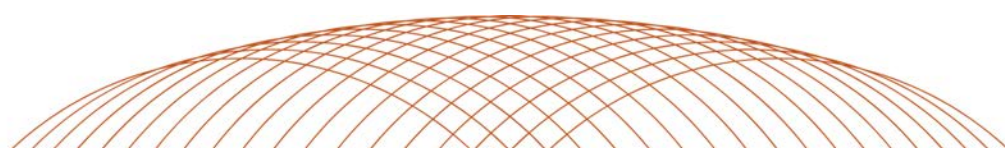
división cabeza/cola para derivar ciudades naturales, basándonos en el hecho de que hay muchas más ubicaciones de baja densidad que ubicaciones de alta densidad en la superficie terrestre. Específicamente, generaron una red irregular triangulada, compuesta por ubicaciones de usuarios únicas e individuales, y posteriormente categorizan pequeños triángulos (por debajo del promedio) como ciudades naturales para los Estados Unidos mensualmente.

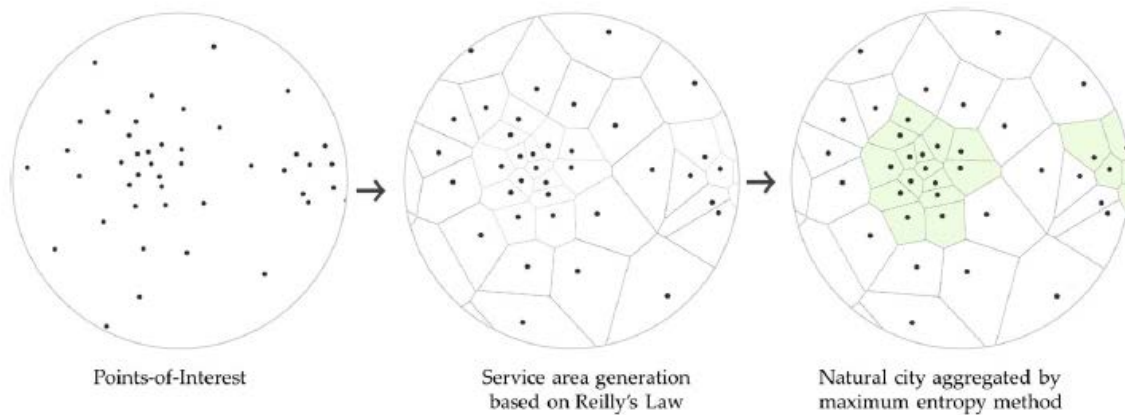
El concepto de ciudades naturales proporciona un medio poderoso para desarrollar nuevos conocimientos sobre la evolución de las ciudades reales, porque prácticamente no hay datos disponibles para rastrear la historia de las ciudades, a lo largo de toda su vida y en escalas espaciales y temporales muy finas. Por lo tanto, consideran que el método de las ciudades naturales, puede actuar como un buen *proxy* de las ciudades reales, en el sentido de comprender las interacciones subyacentes a nivel global, en lugar de predecir las ciudades a nivel individual. Aparte de los datos producidos y los métodos aportados, colocan desde su perspectiva, nuevos conocimientos sobre la estructura y la dinámica de las ciudades naturales. Por ejemplo, la idea de que las ciudades naturales evolucionan de manera no lineal en las dimensiones espacial y temporal, a nivel global, más que de predecir ciudades, a nivel individual.

Otra variante para la delimitación de ciudades naturales, se encuentra en el trabajo “Delimitación de la Ciudad Natural con Puntos de Interés en función del Área de Servicio y Método de Máxima Entropía”, de Lingbo Liu, Binxin Xia, Hao Wu, Jie Zhao, Zhenghong Peng y Yang Yu. Trabajo publicado en el año 2019. En este, se borda y desarrolla el concepto de la Ciudad Natural, considerado por ellos como esencial para comprender la escala física urbana e identificar la expansión urbana en los estudios urbanos. Su método refleja los límites urbanos funcionales de la ciudad definidos por las actividades humanas en lugar de los límites administrativos.

Sustentan su trabajo a partir del reconocimiento de que la mayoría de los estudios tienden a utilizar datos del entorno físico, como redes de calles y datos de sensores remotos para delimitar la ciudad natural; sin embargo, es posible que dichos datos no coincidan con la distribución real de la densidad de actividad humana en las nuevas ciudades o incluso en las ciudades fantasma en China. Es por ello, que sugieren agregar el límite natural de la ciudad, a partir de los polígonos del área de servicio de los puntos de interés, con base en la Ley de Gravitación Minorista de *Reilly* y el método de máxima entropía (MaxEnt), ya que la mayoría de los puntos de interés brindan servicios a las comunidades circundantes, lo que refleja la vitalidad de forma ascendente.

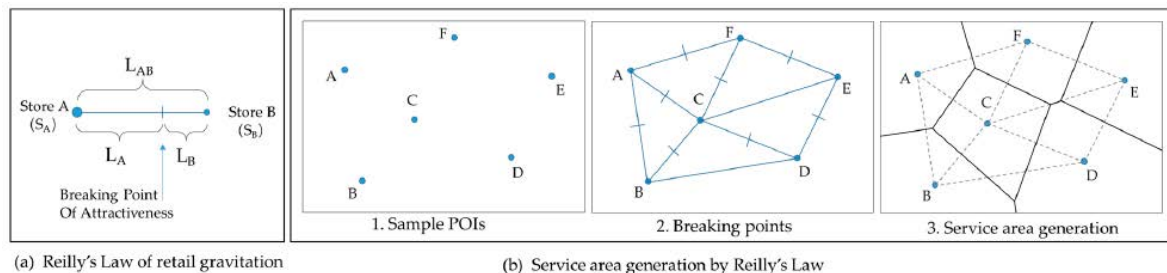
Este artículo propone un método SA-MaxEnt que integra la Ley de Gravitación Minorista de *Reilly* y el método de máxima entropía (MaxEnt) para abordar estos dos temas proponen el siguiente esquema de trabajo.





De acuerdo a Ley de *Reilly*, los clientes tienden a visitar los centros comerciales cercanos, cuyo atractivo está determinado por su tamaño y la distancia a otros centros comerciales, obedeciendo a la ley física de la gravedad, lo que indica que el área de servicio está estrechamente relacionada con la población circundante. Así, el área de servicio que representa el atractivo podría ser la mejor unidad geográfica para formar el límite natural de la ciudad. Además, el tamaño de los PDI podría ignorarse cuando la cantidad de ciertos tipos de PDI es grande suficiente, lo que convierte el cálculo del área de servicio en una generación de un diagrama de Voronoi de puntos de interés. Luego se aplica el método de máxima entropía para calcular el valor umbral óptimo para seleccionar polígonos de Voronoi para agregar la ciudad natural.

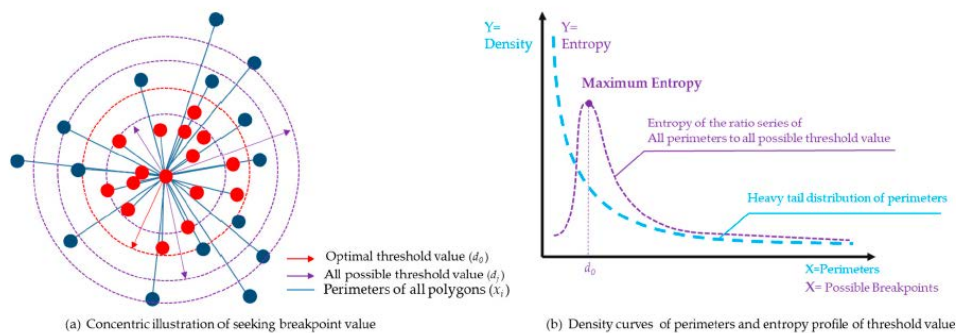
El problema central de su discusión está en el ámbito de la delimitación de los límites naturales de la ciudad. Para ello, utilizan puntos de interés basados en servicios, como tiendas minoristas, farmacias, bancos y restaurantes, que brindan servicios localmente a los residentes de los alrededores, cuya densidad varía en función de la población cercana; para aplicar la ley de *Reilly*. Toman en cuenta los puntos de interés en grandes cantidades para ignorar la influencia del tamaño, tratando los puntos de interés como puntos de tamaño similar. Con esto, generan el área de servicio que representan mediante el método clásico del polígono de Thiessen (o diagrama de Voronoi), como se muestra en el siguiente esquema.



Los límites naturales de la ciudad se agregan a partir de los polígonos mediante un método de extracción apropiado.

Para la agregación utilizan el método de entropía máxima, la cual es una forma de medir la incertidumbre en el sistema complejo. La entropía en termodinámica fue definida por primera vez por Clausius, describiendo la energía incapaz de realizar trabajo durante el procedimiento de transferencia de energía⁷. Boltzmann usó la entropía como una medida del desorden en la estadística termodinámica, que fue introducida en la teoría de la información y luego definida como incertidumbre por Shannon⁸. Jaynes desarrolló el principio de máxima entropía a partir de la expresión de Shannon⁹. Wilson introdujo el principio de máxima entropía en el modelado urbano y regional¹⁰. El popular método del peso de la entropía también puede tomarse como una aplicación del método de la máxima entropía¹¹.

Como el método de máxima entropía es adecuado para resolver problemas de incertidumbre, lo utilizan para buscar el valor umbral óptimo para extraer polígonos de área de servicio. Mientras tanto, los perímetros de los polígonos del área de servicio se eligieron como el factor para buscar el valor del umbral, evitando contener polígonos extremadamente estrechos con áreas pequeñas mientras se usaba el área como factor. Según el principio de máxima entropía para problemas de incertidumbre, la distribución de probabilidad que mejor representa el estado actual del conocimiento es la de mayor entropía. Asumen que puede existir un valor umbral óptimo d_0 en el conjunto de datos de longitud aleatoria $D(d_j)$ para todos los perímetros del conjunto de datos $X(X_i)$, donde la función de proporción $Fd(X)$ de d_j y X_i alcanzará el valor máximo en d_0 (figura a). Por otro lado, la ley de potencia que muestra los perímetros del área de servicio en una forma de distribución de cola pesada puede transformarse en una distribución normal gaussiana aplicando el método de máxima entropía (figura b). Tal transformación refleja el orden interno, una característica autoorganizada del sistema de complejidad urbana¹².



⁷ Liu, L.; Peng, Z.; Wu, H.; Jiao, H.; Yu, Y.; Zhao, J. Identificación rápida de expansión urbana basada en agrupamiento de KMeans con densidad de población y entropía espacial local. *Sostenibilidad* 2018, 10, 2683.

⁸ Cabral, P.; Augusto, G.; Tewolde, M.; Araya, Y. Entropía en Sistemas Urbanos. *entropía* 2013, 15, 5223–5236.

⁹ Leopoldo, LBEI concepto de entropía en la evolución del paisaje; papel profesional; Imprenta del Gobierno de EE.UU.: Washington, DC, EE. UU., 1962.

¹⁰ Wilson, AG Entropía en Modelado Urbano y Regional. *economía Geogr.* 2011, 48, 446–447.

¹¹ Sahoo, MM; Patra, KC; Swain, JB; Khatua, KK Evaluación de la calidad del agua con la aplicación de la regla de Bayes y el método del peso de la entropía. *EUR. J. Medio Ambiente. civ. Ing.* 2016, 21, 730–752.

Delgado, A.; Romero, I. Análisis de conflictos ambientales utilizando un método integrado de agrupamiento gris y peso de entropía. *Reinar. Modelo. suave* 2016, 77, 108–121.

Li, X.; Wang, K.; Liu, L.; Jing, X.; Yang, H.; Gao, C. Aplicación del Método Peso de Entropía y TOPSIS en la Evaluación de Seguridad de Minas de Carbón. *Procedia Ing.* 2011, 26, 2085–2091.

¹² Li, X.; Wang, K.; Liu, L.; Jing, X.; Yang, H.; Gao, C. Aplicación del Método Peso de Entropía y TOPSIS en la Evaluación de Seguridad de Minas de Carbón. *Procedia Ing.* 2011, 26, 2085–2091.

Porque cualquier perímetro X_i que se desvía del valor umbral d_j , no importa si es más pequeño o más grande que el umbral, impondrá un efecto más significativo en la proporción de perímetros en un valor de umbral, por lo tanto, la función de proporción original $F_d(X) = X_i/d_j$ fue revisado como:

$$F_d(X) = m_i | \ln X_i - \ln d_j |$$

Para cada valor de umbral específico d_j , las secuencias de razón correspondientes PAG_{y0} de $F_d(X)$ por cada perímetro X_i es:

$$PAG_{y0} = \sum \frac{F_d(X_i)}{F_d(X_i)}$$

Según la definición de entropía, la entropía total de Shannon H_j de $F_d(X)$ es: j

$$H_j = \sum PAG_{y0} \ln P_{y0}$$

Por cada valor d_j en datos de longitud D , tiene un valor de entropía correspondiente H_j en, que alcanzaría su valor máximo en el óptimo $\sqrt{y0}$ valor $d0$ en términos del principio de máxima entropía. Este artículo también compara $\sqrt{y0}$ el valor de $3/2 d0$, sobre el cual se argumenta que el valor umbral $d0$ debe multiplicarse con $3/2$ según la definición de distribución gaussiana y la regla 3-sigma.

La ciudad de Wuhan, ubicada en el centro de China, es su área de estudio de caso. Tomaron tres tipos de datos de puntos de interés recopilados como información geográfica voluntaria en Wuhan durante el 2016 de *Baidu Map API*, que proporciona datos abiertos de información geográfica voluntaria (PDI) en China. Dichos datos incluían restaurantes (norte=14,233), tiendas de conveniencia (norte=9819) y bancos (norte=4487). A modo de comparación, las estaciones base de telefonía móvil (MP) (norte=33,758) también fueron utilizados. Estos datos fueron proporcionados por un operador de telecomunicaciones asociado cuya participación de mercado era de aproximadamente el 60 % y se verificó que son proporcionalmente representativos de toda la distribución de la población en Wuhan.

Durante el proceso para la generación del área de servicio, observaron que hay muchos más polígonos pequeños que polígonos grandes. Y que el área de servicio de las estaciones base de telefonía móvil presenta una característica más pareja que la de los PDI, donde la primera es planificada por el gobierno y la segunda construida por la mano invisible del mercado.

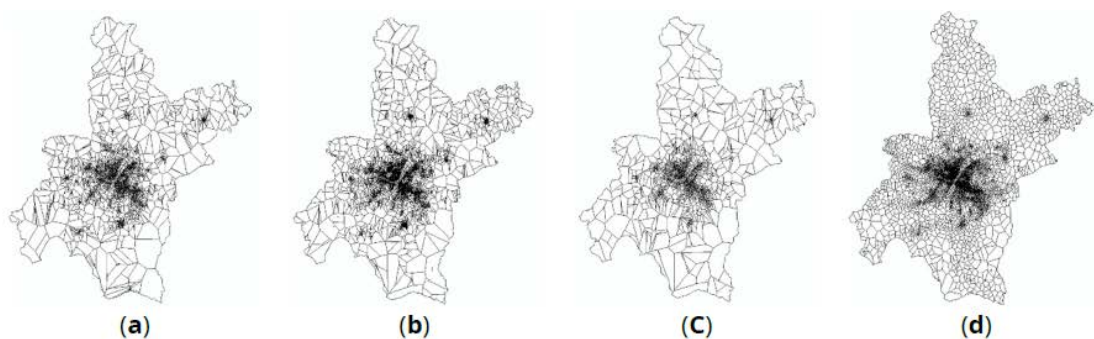


Figura 5. Diagramas de expansión urbana basados en la distribución de la población en cuadrícula. (a) Restaurantes; (b) Historias; (c) Bancos; (d) Estaciones base MP.

Se realizó un análisis estadístico de las áreas de servicio de diferentes puntos de interés.

	Restaurant	Stores	Bank	MP Base Station
Amount	13207	9678	4289	10212
Service Area	(km ²)			
Min area	0.000008	0.000180	0.000133	0.000012
Max area	185.819	130.447822	180.126999	42.908001
Mean	0.650851	0.888088	2.002382	0.842049
Standard Error	4.5672	14.107599	9.164582	2.361717
Perimeter	(km)			
Min Perimeter	0.017092	0.061354	0.057230	0.016712
Max Perimeter	57.826401	64.695572	59.245899	27.1527
Mean	1.520791	2.074611	3.13577	2.358477
Standard Error	3.530842	3.84682	5.797607	3.044902

Como el método de ruptura de cabeza/cola ha indicado que las características fractales dentro de un sistema complejo urbano constituyen una distribución Zipf en rango-tamaño del entorno físico humano, una transformación logarítmica de rango-tamaño mostraría un ajuste lineal. Tal suposición fue consistente con los cuatro conjuntos de datos, donde las tiendas dan el R2 más alto, luego los bancos y restaurantes, y las estaciones base de teléfonos móviles (MP) tuvieron las más bajas.

Para definir el umbral óptimo que define el límite de la ciudad natural, realizaron graficar la ecuación de la entropía máxima para cada conjunto de datos. Obteniendo los siguientes resultados:

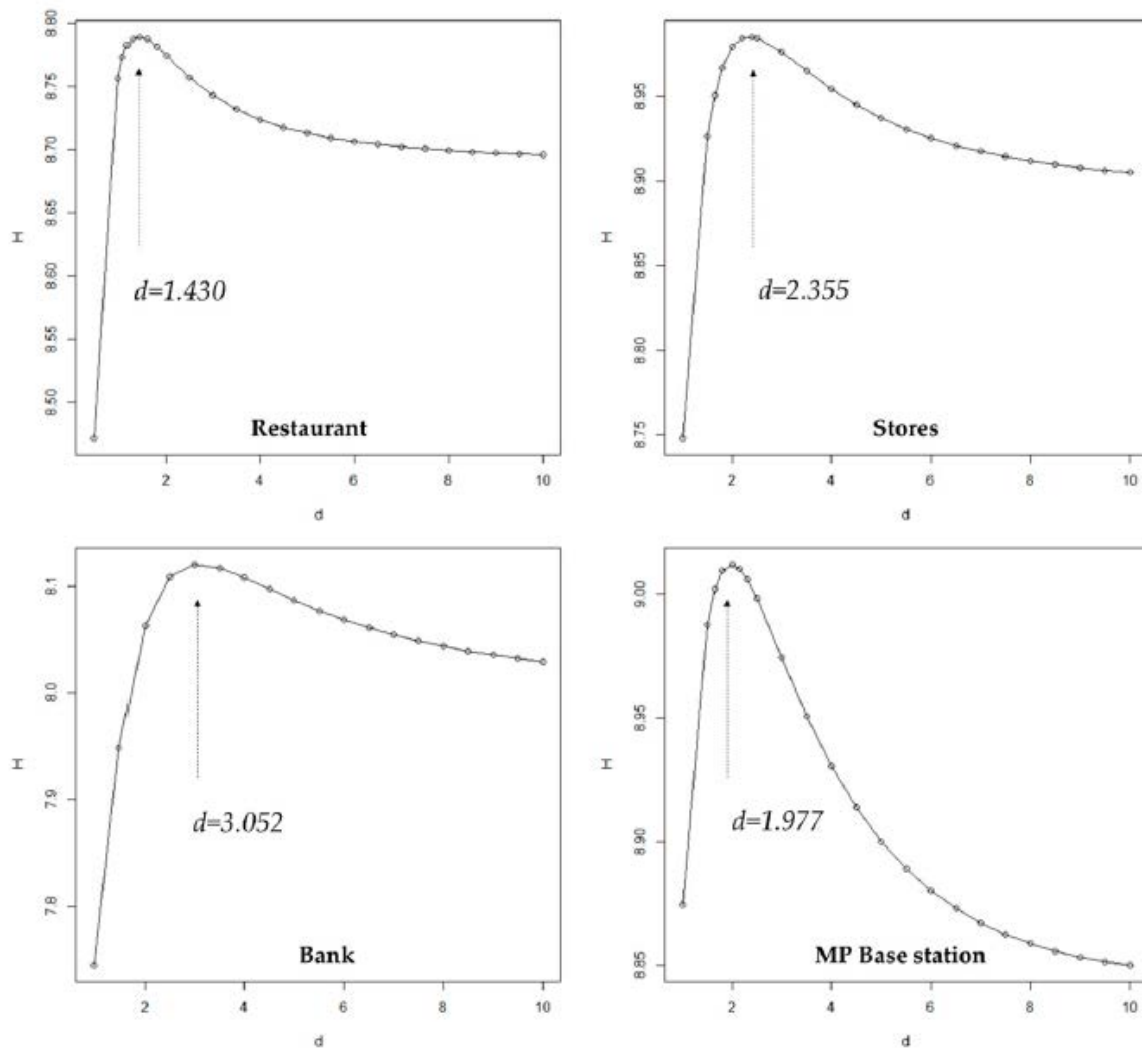


Figura 7. Valor de umbral de entropía máxima.

Tabla 2. Comparación de valor umbral.

	Media (M)	MaxEnt (H)	Relación (H/METRO)	$\frac{3}{\sqrt{2}}H(3\sigma)$
Restaurantes	1.520791	1.430	0.940300	3.0334876
Historias	2.074611	2.355	1.135152	4.9957086
Bancos	3.13577	3.052	0.973285	6.4742686
Estaciones base MP	2.358477	1.977	0.838252	4.1938496

Realizaron una validación a partir de la agregación del límite natural de la ciudad creado mediante cruces de carreteras por Beijing City Lab (BCL). Con base en ello, determinaron que las tiendas ubicadas más uniformemente en la zona urbana generan la mayor cantidad de área natural de la ciudad y que comparada con el área de validación fue muy parecida.

Table 3. Area of Natural city boundaries delimited by the different threshold values.

	Mean(M)	MaxEnt(H)	$3/\sqrt{2}H$	TIN-Head/Tail Breaks
Restaurants	204.62 km ²	186.24 km ²	427.18 km ²	240.47 km ²
Stores	337.76 km ²	388.77 km ²	815.77 km ²	134.44 km ²
Banks	282.21 km ²	271.33 km ²	617.9 km ²	199.10 km ²
MP Base Stations	536.48 km ²	424.07 km ²	1017.832 km ²	846.17 km ²
Road Junctions (BCL)	807.34 km ²			

Los puntos de interés de las tiendas mostraron el mejor desempeño en la delimitación del límite natural de la ciudad, reflejaron que la distribución de las tiendas está más acorde con la densidad de las actividades humanas que los restaurantes y los bancos. En otras palabras, la distribución espacial de las tiendas minoristas está más relacionada con la densidad de población circundante.

Metodología para delinear ciudades ¹³

Recientemente se ha publicado una propuesta metodológica para delinear ciudades, áreas urbanas y rurales, para comparaciones estadísticas internacionales (2020). Esta pretende armonizar los criterios para definir los límites de las ciudades y poder realizar comparaciones internacionales. Para ello, propone el “Grado de Urbanización, clasifica todo el territorio de un país en tres clases: 1) ciudades, 2) pueblos y zonas semidensas y 3) zonas rurales. Tiene dos extensiones. La primera extensión (grado de urbanización nivel 1) identifica: ciudades, pueblos, áreas suburbanas o periurbanas, aldeas, áreas rurales dispersas y áreas mayoritariamente deshabitadas. La segunda extensión (grado de urbanización nivel 2) agrega una zona de desplazamiento alrededor de cada ciudad para crear un área urbana funcional o área metropolitana” (pp. 2).

El Grado de Urbanización nivel 1

El Grado de Urbanización clasifica las unidades locales como 1) “ciudades” o “áreas densamente pobladas”, 2) “pueblos y áreas semi-densas” o “áreas de densidad intermedia” y 3) “áreas rurales” o “áreas escasamente pobladas” con base en sobre la densidad de población, el tamaño de la población y la contigüidad utilizando celdas de cuadrícula de 1 km². Cada unidad local pertenece exclusivamente a una de estas tres clases. Las unidades locales pueden ser unidades administrativas, como los municipios, o unidades estadísticas, como las áreas de empadronamiento del censo.

La base para el Grado de Urbanización es una cuadrícula de población de 1 km². Cada celda de la cuadrícula tiene la misma forma y tamaño, por lo que se evitan las distorsiones causadas por el uso de unidades que varían en forma y tamaño. Esta es una ventaja considerable

13 Comisión Europea – Eurostat y DG de Política Regional y Urbana–OIT, FAO, OCDE, ONU-Hábitat, Banco Mundial. 2020. Fifty-first sesión. Items for discussion and decision: demographic statistics. Disponible en: https://ec.europa.eu/eurostat/cros/content/recommendation-method-delineate-cities-urban-and-rural-areas-international-statistical-comparisons_en

en comparación con los métodos basados en el tamaño y la densidad de la población de las unidades administrativas locales u otras unidades estadísticas.

Paso 1: Clasificación de cuadrícula

El Grado de Urbanización identifica tres tipos de celdas de cuadrícula:

- Un centro urbano (o un clúster de alta densidad) consta de celdas de cuadrícula contiguas con una densidad de al menos 1,500 habitantes por km². Un centro urbano tiene una población de al menos 50,000 habitantes. Los espacios en este grupo se rellenan y los bordes se suavizan. Si es necesario, se pueden agregar celdas con un 50% de acumulación¹⁴.
- Un clúster urbano (o clústeres de densidad moderada) consta de celdas de cuadrícula contiguas con una densidad de al menos 300 habitantes por km² y tiene una población de al menos 5,000 habitantes en el clúster. (Los centros urbanos son subconjuntos de los grupos urbanos correspondientes).
- Celdas de cuadrícula rural (en su mayoría celdas de baja densidad) son celdas que no pertenecen a un conglomerado urbano. La mayoría de estos tendrán una densidad por debajo de los 300 habitantes por km². Algunas celdas rurales tendrán una mayor densidad, pero no forman parte de un conglomerado con un tamaño de población lo suficientemente grande como para ser clasificado como un conglomerado urbano.

Paso 2: Clasificar las unidades locales

Una vez que todas las celdas de la cuadrícula se han clasificado como centros urbanos, conglomerados urbanos y celdas de la cuadrícula rural, el siguiente paso consiste en superponer estos resultados a las unidades locales.

1. Ciudades (o áreas densamente pobladas): unidades locales que tienen al menos el 50% de su población en centros urbanos.
2. Pueblos y áreas semi-densas (o áreas de densidad intermedia): unidades locales que tienen menos del 50% de su población en centros urbanos y menos del 50% de su población en celdas de cuadrícula rural.
3. Zonas rurales (o zonas escasamente pobladas): unidades locales que tienen al menos el 50% de su población en celdas de cuadrícula rural. Áreas urbanas consisten en ciudades más pueblos y áreas semi-densas. Sin embargo, como este método fue desarrollado para capturar el continuo rural urbano, recomienda reportar indicadores para las tres clases en lugar de solo para la dicotomía urbano-rural. Esto es importante porque los pueblos y las áreas semidensas pueden diferir significativamente tanto de las ciudades como de las áreas rurales. Las políticas que se aplican uniformemente a estas tres clases pueden no ser adecuadas y podrían beneficiarse si se adaptan a cada uno de estos tipos de áreas. Una consecuencia importante de este nuevo método es un llamado a más investigación sobre políticas para esas áreas de densidad intermedia, que podrían tener en cuenta las complementariedades e interdependencias entre los tres tipos de territorios.

¹⁴ En unos pocos países con un desarrollo urbano de densidad relativamente baja y una fuerte separación de las funciones de uso del suelo, el Grado de Urbanización genera múltiples centros urbanos para una sola ciudad. La creación de centros urbanos que utilicen celdas con una densidad de al menos 1.500 habitantes y celdas que estén construidas al menos en un 50 % resuelve este problema. Estas celdas altamente urbanizadas suelen contener parques de oficinas, centros comerciales, fábricas e infraestructura de transporte.

Las comparaciones con una selección de definiciones nacionales han mostrado un alto nivel de acuerdo para las dos clases extremas: las ciudades generalmente se clasifican como urbanas y las áreas rurales como rurales. Sin embargo, las ciudades y las áreas semidensas a veces se clasifican como urbanas ya veces como rurales según las definiciones nacionales. Al ubicar estas áreas en una categoría intermedia, este método intenta acomodar estos diferentes puntos de vista y enfatiza que estas áreas de densidad intermedia están a medio camino entre una ciudad y un área rural.

Las áreas semidensas en países de ingresos bajos y medianos a menudo se describen como áreas periurbanas. En los países de altos ingresos, generalmente se los describe como suburbios. En ambos casos, estas áreas tienen una densidad moderada y se encuentran en la transición entre una zona rural y una ciudad o pueblo.

Tenga en cuenta que la calidad de la cuadrícula de población puede tener un impacto significativo en las proporciones de población en las tres clases del grado de urbanización. Este es particularmente el caso de la diferencia entre áreas rurales y ciudades y áreas semidensas, ya que esto a menudo requiere una distinción fina dentro de las unidades administrativas. Una red que concentre demasiado la población sobrestimaré la población en pueblos y áreas semidensas. Una cuadrícula que disperse demasiado a la población sobreestimaré la población en las zonas rurales.

El Grado de Urbanización nivel 2

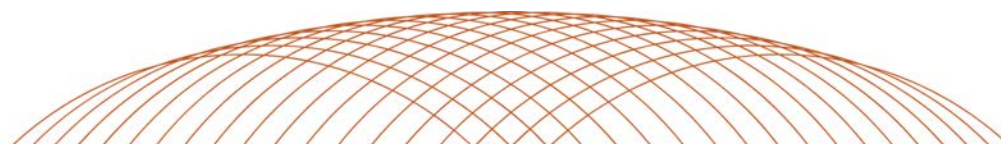
El Grado de Urbanización nivel 2 se implementa con el mismo enfoque de dos pasos. Las primeras celdas de la cuadrícula se clasifican según la **densidad de población**, el **tamaño de la población** y la **contigüidad**. Posteriormente, las unidades locales se clasifican según el tipo de cuadrícula en la que viven sus habitantes. El Grado de Urbanización nivel 2 es una subclasificación del nivel 1. Fue creado para identificar asentamientos medianos y pequeños, es decir, pueblos y aldeas.

Paso 1: Clasificación de cuadrícula

Un núcleo urbano se identifica de la misma forma que en el Grado de Urbanización nivel 1:

- Un centro urbano consiste en celdas de cuadrícula contiguas con una densidad de al menos 1,500 habitantes por km². Un centro urbano tiene una población de al menos 50,000 habitantes. Los espacios en este grupo se rellenan y los bordes se suavizan. Si es necesario, se pueden agregar celdas que tengan un 50% de acumulación.
- Las células del clúster urbano que no forman parte de un centro urbano se puede subdividir en tres tipos.
- Un denso grupo urbano consiste en celdas contiguas con una densidad de al menos 1,500 habitantes por km², con una población de al menos 5,000 y menos de 50,000 en el clúster.
- Un conglomerado urbano semidenso consistir en celdas de cuadrícula contiguas con una densidad de al menos 300 habitantes por km² y tiene una población de al menos 5,000 habitantes (es decir, un grupo urbano) y este grupo no es contiguo ni está dentro de los 2 km de un grupo urbano denso o un centro urbano¹⁵.

¹⁵ Medido como fuera de un búfer de tres celdas de cuadrícula de 1 km² alrededor de densas agrupaciones urbanas y centros urbanos.



- Células suburbanas o periurbanas son las restantes células de conglomerados urbanos, es decir, aquellas que no forman parte de un conglomerado denso o semidenso. Estas celdas son parte de un grupo urbano que es contiguo o dentro de los 2 km de un grupo urbano denso o un centro urbano.
- Las celdas de cuadrícula rural se pueden clasificar en tres tipos.
- Un grupo rural consiste en celdas contiguas con una densidad de al menos 300 habitantes por km² y una población entre 500 y 5,000 en el clúster.
- Celdas de red rural de baja densidad son celdas de cuadrícula rural con una densidad de al menos 50 y no forman parte de un clúster rural.
- Celdas de red rural de muy baja densidad son celdas con una densidad de menos de 50 habitantes por km².

Paso 2: Clasificación de la Unidad Local

Las unidades locales se clasifican como ciudades de la misma manera que en el Grado de Urbanización nivel 1:

- Una ciudad consiste de una o más unidades locales que tengan al menos el 50% de su población en un centro urbano.
Las unidades locales clasificadas como “pueblos y áreas semi-densas” se pueden dividir en tres clases:
- Pueblos densos tienen una mayor proporción de su población en conglomerados urbanos densos que en conglomerados urbanos semidensos (es decir, es denso) y una proporción mayor en conglomerados urbanos densos más semidensos que en células suburbanas o periurbanas (es decir, es una ciudad).
- Pueblos Semidensos tienen una mayor proporción de población en conglomerados urbanos semidensos que en conglomerados urbanos densos (es decir, es semidenso) y una mayor proporción en conglomerados urbanos densos más semidensos que en células suburbanas o periurbanas (es decir, es una ciudad).
- Áreas suburbanas o periurbanas tienen una mayor proporción de población en las células suburbanas o periurbanas que en los grupos urbanos densos y semidensos.
- Las ciudades densas y semidensas se pueden combinar en ciudades. Esto reduce el número de clases y puede ser útil especialmente si la proporción de población en las ciudades semidensas es baja.
Las unidades locales clasificadas como “áreas rurales” se pueden dividir en tres clases:
- Pueblos tienen la mayor parte de su población de celdas de cuadrícula rural que vive en un grupo rural.
- Zonas rurales dispersas tienen la mayor parte de su población de celdas de cuadrícula rural viviendo en celdas de cuadrícula rural de baja densidad.
- En su mayoría zonas deshabitadas tienen la mayor parte de su población de celdas de cuadrícula rural viviendo en celdas de cuadrícula rural de muy baja densidad.

Apuntes para la definición de ciudad ¹⁶

Para los fines de este trabajo y del proyecto de percolación, se consideran los conceptos urbano y ciudad como términos similares, que ambos son útiles para fines prácticos en el diseño del modelo de percolación para identificar el suelo urbano en el contexto nacional.

Como parte de la revisión documental para precisar sobre la definición de ciudad o bien del término urbano, se revisó el escrito de Horacio Capel (1975) quien señala que la distinción de lo rural y lo urbano, cada vez se va diluyendo, debido a los efectos y desarrollo de los medios de comunicación, así como la evolución de transporte, la transmisión de mensajes e información; además de la homogeneización de muchas pautas de comportamiento, formas de vida y actitudes con relación al nivel de vida y la acción generalizada de los medios de comunicación de masas. Estos factores según Capel han contribuido en los países industrializados a borrar muchas de las antiguas diferencias entre ciudad y campo, haciendo confusa y problemática esta distinción. No obstante, señala que en países periféricos aún sigue siendo un tema relevante en términos de la investigación.

En su texto Capel presenta muchas de las definiciones teóricas de ciudad y la determinación del límite inferior de lo urbano; además de señalar la existencia de un continuo rural-urbano. EL mismo autor reconoce dos vertientes: 1. La definición teórica de lo urbano en dicotomía de lo rural, que trata de identificar los rasgos esenciales de la ciudad; 2. La definición técnica o concreta utilizada en cada país o región, con fines estadísticos de lo urbano y buscando fijar límites a partir de los cuales se puede hablar de ciudad.

Relacionado con la primera vertiente, se identifican dos perspectivas teóricas:

1. Las definiciones que se basan en una o dos características esenciales de lo urbano, y
2. Aquellas que son consideradas como eclécticas, que intentan definir la ciudad a partir de su complejidad. A continuación se presenta el compilado de definiciones de autores representativos en este tema y presentadas por Capel:

Conceptualizaciones y definiciones

El sociólogo francés R. Maunier (1910) al definir la ciudad como "una sociedad compleja, cuya base geográfica es particularmente restringida con relación a su volumen y cuyo elemento territorial es relativamente débil en cantidad con relación al de sus elementos humanos" (pp. 266).

H. Dörries (1952) cuando sostiene que una ciudad se reconoce "por su forma más o menos ordenada, cerrada, agrupada alrededor del núcleo fácil de distinguir y con un aspecto muy variado, acompañada de los elementos más diversos"¹⁷ (pp. 182).

Ratzel (1897) en su *Anthropogeographie* al considerar a la ciudad como "una reunión duradera de hombres y de viviendas humanas que cubre una gran superficie y se encuentra en la encrucijada de grandes vías comerciales"¹⁸ (pp. 15).

16 Como insumo para el proyecto "Definición de límites de ciudades mediante la teoría de la percolación", se resume el trabajo de Capel, Horacio. (1975). La definición de lo urbano. En *Estudios geográficos*. Año 1975, Vol. 36, Número 138-139. Dedicado a: Homenaje a D. Manuel de Terán (I). pp. 265-302. Disponible: <https://dialnet.unirioja.es/ejemplar/203439>

17 DÖRRIES, H.: *Der gegenwärtige standt der Stadtgeographie*, 1930. Cit. por SORRE, M.: *Les fondaments de la Géographie Humaine, L'habitat*, vol. III, París, A. Colin, 1952, pág. 182.

18 RATZEL, F: *Anthropogeographie*, 1891. Cit. por CHABOT, G.: *Les villes*, París, A. Colin, 1948, pág. 15. Trad. cast. Barcelona, Labor, S. A., 1972.

Para Richthofen (1963), una ciudad es "un agrupamiento cuyos medios de existencia normales consisten en la concentración de formas de trabajo que no están consagradas a la agricultura, sino particularmente al comercio y a la industria"¹⁹ (pp. 30).

El norteamericano Auroousseau (1961), "rurales son aquellos sectores de población que se extienden en la región y se dedican a la producción de los artículos primarios que rinde la tierra; los sectores urbanos, en cambio, incluyen a las grandes masas concentradas que no se interesan, al menos en forma inmediata, por la obtención de materias primas, alimenticias, textiles o de confort en general, sino que están vinculadas a los transportes, a las industrias, al comercio, a la instrucción de la población, a la administración del Estado o simplemente a vivir en la ciudad"²⁰ (pp. 42).

R. E. Dickinson (1951), que caracteriza a las ciudades de la Europa occidental y de América del Norte como "núcleos de poblamiento compacto, dedicados principalmente a ocupaciones no agrícolas".²¹

W. Sombart (1932), consideró a la ciudad "como un establecimiento de hombres que para su mantenimiento han de recurrir al producto de un trabajo agrícola exterior"²² (pp. 449).

J. M. Casas Torres (1957), "la ciudad, grande o chica, se abastece desde fuera, no se basta para su aprovisionamiento, y existe en función de una región más amplia a la que organiza, a la que sirve, para la que es el nexo de unión con el resto del mundo"²³ (pp. 262).

P. Deffontaines y J. Brunhes (1926), "hay ciudad cuando la mayor parte de los habitantes pasan la mayor parte del tiempo en el interior de la aglomeración"²⁴

P. George (1968), plantea el problema de las pequeñas ciudades, y propone considerar como tales a las agrupaciones densas de viviendas en el interior de las cuales todos los desplazamientos funcionales se realizarían a pie; esto lleva a individualizar núcleos cuya población oscila entre menos de 10,000 y más de 50,000 habitantes, según las regiones.²⁵

Georg Simmel, (1966 y 1968) la economía del dinero estimula en el hombre la tendencia a la abstracción y favorece el desarrollo de las facultades intelectuales, provocando una despersonalización de las relaciones humanas; "economía monetaria y vida del intelecto están íntimamente ligadas".²⁶

¹⁹ RICHTHOFEN, F.: Vorlesungen über allgemeine Siedlungs- und Verkehrsgeographie, 1908. Cit. por BEAUJEU GARNIER, J. y CHABOT, G.: *Traité de Géographie urbaine*, París, A. Colin, 1963. Trad. cast. Barcelona, Vicens Vives, 1970, pág. 30.

²⁰ AUROUSSEAU, M.: "The distribution of population: a constructive problem", *The Geographical Review*, New York, vol. XI, núm. 4, octubre 1921, págs. 563-592. Cit. por DICKINSON, R. E.: *Ciudad, región y regionalismo*, trad. cast. por Angel Abascal, Barcelona, Editorial Omega, 1961, pág. 42.

²¹ DICKINSON, R. E.: *The Western European city. A Geographical Interpretation*, Londres, 1951. (El capítulo 15 ha sido reproducido en MAYER, H. M., y KOHN, C. F.: *Readings in Urban Geography*, op. cit. en nota 47).

²² SOMBART, W.: *L'apogée du capitalisme*. Trad. franc., París, Payot, 1932, vol. 1, tercera parte, "La ville", págs. 449 y sigs.

²³ CASAS TORRES, J. M.: "Ciudades, urbanismo y geografía", *Estudios Geográficos*, Madrid, vol. 18, 1957, núms. 67-68, pág. 262.

²⁴ BRUNHES, J., y DEFFONTAINES, P.: "Géographie humaine de la France", vols I y II de la *Histoire de la nation française*, dirigida por G. Hanotaux, París, Plon, 1926.

²⁵ GEORGE, P.: *Scheme d'Etude internationale des petites villes*, en U. G. I.: *21th International Geographical Congress. Abstracts of Papers*, Calcuta, 1968, num. 584.

²⁶ La importancia de la obra de Simmel ha sido destacada por Don MARTINDALE: "Prefactory remarks: The theory of the city" (introducción a la trad. inglesa de la obra de Max WEBER: *The city*, Nueva York, 1958; edic. de 1966, págs. 9-62); y, siguiendo a este autor, por CLAVAL, P.: "La théorie des villes", *Revue Géographique de l'Est*, Besançon, 1968, 1-2, págs. 3-56.

Para Manuel Castells (1974) "la organización social y el sistema cultural dependen de algo más que el número y la heterogeneidad de los individuos; hay que tener también en cuenta la estructura tecno-social, fundamento organizativo de la sociedad".²⁷

M. de Terán (1964), considera que "este elemento de la heterogeneidad social ha de ser retenido como especialmente válido como factor explicativo de la diferenciación social del espacio urbano y del paisaje en que adquiere expresión material".²⁸

Richard L. Meier, (1972) se plantea el problema de qué es lo que determina las diferencias de densidad de productividad del trabajo y de valor del suelo entre una comunidad rural y una metrópoli, y concluye que es fundamentalmente la capacidad de intercambio, en su sentido más amplio, desde la información a las mercancías, que existe en la segunda.²⁹

Para Remy (1966), la ciudad es un elemento esencial del sistema económico precisamente por su condición de lugar de intercambio, de elección y de innovación.³⁰

R. Abler, J. S. Adams y P. Gould, (1972) afirman que "una ciudad es una organización espacial de personas y actividades especializadas diseñadas para maximizar los intercambios; a nivel local, la ciudad es el mejor medio de interrelacionar actividades sociales y económicas para máximo beneficio de todas ellas; a nivel regional, aparecen sistemas de ciudades para organizar intercambios entre lugares distantes y para facilitar a las áreas circundantes de carácter no urbano los bienes y servicios que necesitan".³¹

Toschi, (1966) la ciudad "es un agregado complejo y orgánico de edificios y viviendas, ejercitando una función de centro coordinador para una región más o menos vasta, en el cual la población, las construcciones y los espacios libres se desarrollan diferenciados por las funciones y por la forma, coordinados unitariamente en función del grupo social localizado, y en desarrollo hasta constituir un típico organismo social".³²

Características esenciales

Al respecto Capel señala que "los rasgos que con más frecuencia se han considerado para caracterizar el hecho urbano han sido, fundamentalmente, el tamaño y la densidad, el aspecto del núcleo, la actividad no agrícola y el modo de vida, así como ciertas características sociales, tales como la heterogeneidad, la "cultura urbana" y el grado de interacción social" (pp. 266).

En tanto, para la definición concreta el **tamaño medido a partir del número de habitantes**, es una de las características fundamentales; no obstante, a nivel teórico se utilizan además de esta, otras características. Con mayor frecuencia se utiliza la **densidad de habitantes o de edificación**. Su construcción resulta relativamente sencilla de calcular; pero la dificultad se presenta cuando se requiere fijar un parámetro que permita definir lo que es urbano y lo que es rural. Para hacer frente a esta situación, un gran número de autores no dudan en asignar

²⁷ CASTELLS, M.: Ibid., pág. 55. El carácter ideológico de la confusión entre cultura urbana y forma espacial ha sido tratado también por CASTELLS, M.: La question urbaine, op. cit. en nota 66, cap. II.

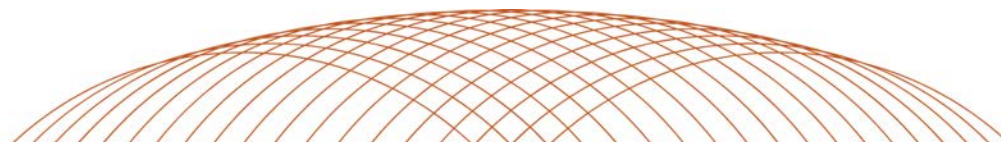
²⁸ TERAN, Manuel de: "Geografía humana y sociología. Geografía social", Estudios Geográficos, Madrid, vol. 25, núm. 97, nov. 1964, pág. 464.

²⁹ MEIER, Richard L.: A communication theory of urban growth, The Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, The M. I. T. Press, 1962. Existe trad. francesa por P. U. F., París, 1972.

³⁰ REMY, Jean: La ville, phénomène économique, Bruselas, Les Editions Vie Ouvrière, 1966, 298 págs., sobre todo las págs. 91-98, 124, 125, 226 y 254.

³¹ ABLER, Ronald; ADAMS, John S., y GOULD, Peter: Spatial organization. The Geographer's view of the World, Londres y Englewoods Cliffs, Prentice Hall International Inc., 1972, pág. 354.

³² TOSCHI, M.: La città, Turin, U. T. E. T., 1966, pág. 42



caracteres urbanos. La solución podría estar quizás en la utilización de unos gradientes de densidades.

La segunda característica fundamental es la **forma**, entendida como el aspecto de la aglomeración; aunque vale la pena mencionar que a casi cinco décadas, no se contaba con las herramientas, datos y métodos que hoy día se utilizan para medir la forma de las ciudades.

Tercera, la función económica, referida concretamente a las **actividades no agrícolas**. Característica que es compartida por numerosos autores al momento de definir a la ciudad. Una forma diferente de presentar esta definición funcional es la que insiste en el hecho de que los ciudadanos no producen directamente sus alimentos. Sin embargo, existe una dependencia del exterior que va unida necesariamente a lazos de dominación, intercambio con un espacio más o menos amplio, al de la ciudad. Además, el pasar la mayor parte del tiempo dentro de la aglomeración como algunas concepciones lo proponen, plantea el problema de las ciudades dormitorio afectadas por movimientos laborales diarios.

Para G. Simmel, las características principales que diferencian a las ciudades del campo son: "en una gran ciudad la vida es más intelectual"; y "una intensificación de la vida nerviosa". La ciudad es un lugar cosmopolita, con alto nivel de individualización originado por la división del trabajo. Posteriormente Max Weber (1958) señala que los intercambios comerciales, la función política y militar; instituciones y organización social, son características esenciales de las ciudades.³³

Con mayor claridad Wirth señala las siguientes características esenciales: "aislamiento social; secularización; segmentación de los roles o papeles desempeñados; normas poco definidas; relaciones sociales caracterizadas por la superficialidad, el anonimato y el carácter transitorio y utilitario; especialización funcional y división del trabajo; espíritu de competencia, frente a la solidaridad de las sociedades rurales; gran movilidad; economía de mercado, predominio de las relaciones secundarias e impersonales sobre las primarias, que serían características de las sociedades rurales; debilitación de las estructuras familiares y desaparición de las relaciones con parientes lejanos; en relación con ello, paso de la comunidad a la asociación; dimisión del individuo respecto a las asociaciones; control de la política por asociaciones de masas" (Capel, 1975 pp. 5).

De manera sintética, el mismo Wirth señala tres características: 1. tamaño y crecimiento de las aglomeraciones urbanas; 2. La densidad; y 3. Diversificación y heterogeneidad.

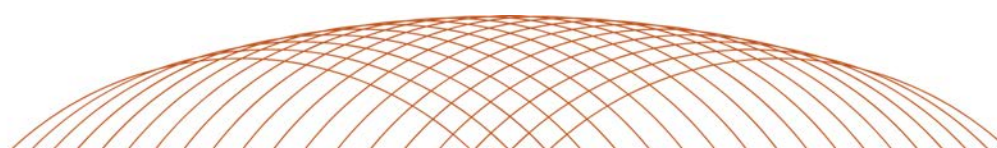
En la geografía moderna, varios autores han fundamentado sus trabajos los planteamientos de Wirth, haciendo énfasis en la heterogeneidad, la cual ha sido considerada como elemento distintivo de lo urbano; sin embargo para la sociología urbana, lo que cuenta es la intensidad de las interrelaciones y no solo que tan heterogénea es la ciudad.

Remy (1974), analizó la incidencia de las innovaciones tecnológicas sobre las formas de vida social, ha puesto de relieve el hecho de que las innovaciones en la transmisión de mensajes y en el desplazamiento de personas han roto el monopolio que hasta ahora poseía la densidad física del hábitat para crear una densidad de comunicación social.³⁴

Desde el punto de vista sociológico, tal como lo expresa Capel (1975) cabe plantearse el problema si tiene sentido seguir hablando de la ciudad y de lo urbano como contrapuesto a

³³ WEBER, M.: The city, Trad. inglesa, The Free Press, Glencoe, Illinois, 1958. Prefacio de Don Martindale. Véase también THIRY, J. Théories sur le phénomène urbain, op. cit. en nota 3, págs. 83 y sigs.

³⁴ REMY, Jean: Utilización del espacio, innovación tecnológica y estructura social, trad. cast en BERINGUIER, ch. y otros: Urbanismo y práctica política, Barcelona, Editorial Los Libros de la Frontera, 1974.



lo rural en las regiones de elevado desarrollo tecnológico. La pregunta ha sido formulada y, además, contestada negativamente por numerosos sociólogos y antropólogos: en el coloquio sobre "Sociología y urbanismo", celebrado en Royaumont, Francia, en 1968, fueron numerosos los participantes que a partir de estas premisas pusieron en duda la validez actual de la contraposición entre lo rural y lo urbano en los países altamente industrializados. Queda, sin embargo, incluso desde la perspectiva sociológica, el problema de los países subdesarrollados, donde esta contraposición, nos parece, continúa teniendo validez, aunque sólo sea por el hecho de que las condiciones técnicas en ellos imperantes hacen que sigan siendo precisamente las ciudades los lugares donde se consigue el grado más elevado de interacción social (pp. 8).

Hacia una definición sintética de ciudad

En general, todas estas definiciones de ciudad aceptan los caracteres tamaño, densidad y actividad no agrícola, sumando diversos criterios, según los casos. En la interpretación de algunas definiciones, Capel señala que los servicios altamente especializados y de funciones de organización, son criterios complementarios y a su vez definitorios de las ciudades.

Formas de medir las ciudades

De acuerdo con Capel (1975) existen diversos métodos y criterios para definir de manera concreta el hecho urbano, que son particulares de cada país. Estos se pueden agrupar en criterios cualitativos que consideran: estatuto jurídico, la función administrativa, el aspecto del núcleo, la existencia de determinados servicios; y los cuantificables: número de habitantes, la densidad o la existencia de una determinada cifra de trabajadores no agrícolas.

Para fines de este estudio, los que merecen mayor atención son los criterios de tipo cuantitativo. Entre estos se encuentra el volumen de población, como diferenciador entre lo urbano y lo rural; no obstante, Capel (1975) señala que los criterios son extraordinariamente diversos; al igual que la densidad tanto poblacional como edificada. La siguiente clasificación es tomada de ONU-Habitat³⁵, en la que presenta las definiciones de urbano que cada país ha definido de acuerdo al *Demographic Yearbook* de Naciones Unidas.

África

Botswana: Aglomeración de 5,000 habitantes o más donde el 75% de la actividad económica no es agrícola.

Burundi: Comuna de Bujumbura.

Comoras: Centros administrativos de prefecturas y localidades de 5,000 o más habitantes.

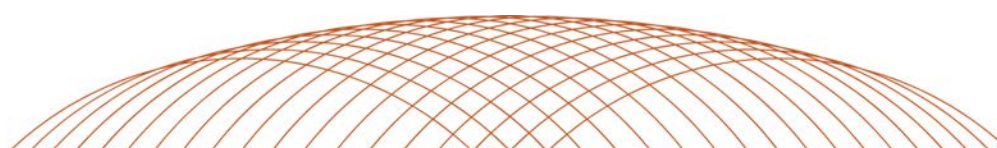
Egipto: Gobernaturas de El Cairo, Alejandría, Port Said, Ismailia, Suez, demarcaciones fronterizas y capitales de otras provincias, así como capitales de distrito (Markaz).

Guinea Ecuatorial: Centros de distrito y localidades con 300 viviendas y / o 1,500 habitantes o más.

Etiopía: Localidades de 2,000 habitantes o más.

Liberia: Localidades de 2,000 habitantes o más.

³⁵ ONU-Habitat, disponible en: <https://onuhabitat.org.mx/index.php/las-diferentes-definiciones-de-urbano-en-el-mundo#:~:text=Dos%20tipos%20de%20%C3%A1reas%20urbanas,por%20milla%20cuadrada%20o%20m%C3%A1s.>



Malawi: Todos los municipios y áreas de planificación urbana.
Mauricio: Ciudades con límites legales proclamados.
Níger: Capital, capitales de los departamentos y distritos
Senegal: Aglomeraciones de 10,000 o más habitantes.
Sudáfrica: Lugares con alguna forma de autoridad local.
Sudán: Localidades de importancia administrativa y / o comercial o con una población de 5,000 o más habitantes.
Swazilandia: Las localidades proclamadas como urbanas.
Túnez: Población que vive en comunas.
República Unida de Tanzania: 16 municipios proclamados.
Zambia: Localidades de 5,000 o más habitantes, la mayoría de los cuales dependen de actividades no agrícolas.

Norte América

Canadá: Lugares de 1,000 habitantes o más, con una densidad de población de 400 o más por kilómetro cuadrado.

Costa Rica: Centros administrativos de los cantones.

Cuba: Población que vive en un núcleo de 2,000 habitantes o más.

República Dominicana: Centros administrativos de municipios y distritos municipales, algunos de los cuales incluyen zonas suburbanas de carácter rural.

El Salvador: Centros administrativos de los municipios.

Groenlandia: Localidades de 200 o más habitantes.

Guatemala: Departamento de la Municipalidad de Guatemala y centros oficialmente reconocidos de otros departamentos y municipios.

Haití: Centros administrativos de comunas.

Honduras: Localidades de 2,000 habitantes o más, que tienen características esencialmente urbanas.

México: Localidades de 2,500 o más habitantes.³⁶

Nicaragua: Centros administrativos de municipios y localidades de 1,000 habitantes o más con calles y luz eléctrica.

Panamá: Localidades de 1,500 o más habitantes que tienen características esencialmente urbanas. A partir de 1970, las localidades de 1 500 o más habitantes con características urbanas tales como calles, sistemas de suministro de agua, sistemas de alcantarillado y luz eléctrica.

Puerto Rico: Aglomeraciones de 2,500 habitantes o más, generalmente con densidades de población de 1,000 personas por milla cuadrada o más. Dos tipos de áreas urbanas: áreas urbanizadas de 50,000 o más habitantes y grupos urbanos de al menos 2,500 y menos de 50,000 habitantes.

Estados Unidos: Aglomeraciones de 2,500 habitantes o más, que en general tienen densidades de población de 1,000 personas por milla cuadrada o más. Dos tipos de áreas

³⁶ Para el caso mexicano, el Instituto Nacional de Estadística y Geografía (INEGI), el concepto de urbano se refiere al espacio físico construido con diversas edificaciones (vivienda, fábricas, edificios, bodegas) e infraestructura de servicios (drenaje, tuberías de agua, tendidos eléctricos); habitan poblaciones mayores de 2,500 personas. Lo rural son localidades donde habitan poblaciones menores de 2,500 personas, como bosques, praderas y áreas agrícolas.
<https://cuentame.inegi.org.mx/glosario/r.aspx?tema=G>

urbanas: áreas urbanizadas de 50,000 o más habitantes y grupos urbanos de al menos 2,500 y menos de 50,000 habitantes.

Islas Vírgenes de los Estados Unidos: Aglomeraciones de 2,500 habitantes o más, que generalmente tienen densidades de población de 1,000 personas por milla cuadrada o más. Dos tipos de áreas urbanas: áreas urbanizadas de 50,000 o más habitantes y grupos urbanos de al menos 2,500 y menos de 50,000 habitantes (A partir del Censo 2000, no se identifican áreas urbanizadas en las Islas Vírgenes de EE. UU.).

América del Sur

Argentina: Centros poblados con 2,000 habitantes o más.

Bolivia: Localidades de 2 000 habitantes o más.

Brasil: Zonas urbanas y suburbanas de centros administrativos de municipios y distritos.

Chile: Centros poblados que tienen características urbanas definidas, como ciertos servicios públicos y municipales.

Ecuador: Capitales de provincias y cantones.

Islas Malvinas: Falkland Islands): Ciudad de Stanley.

Paraguay: Ciudades, pueblos y centros administrativos de departamentos y distritos.

Perú: Centros poblados con 100 o más viviendas.

Suriname: Ciudad de Paramaribo.

Uruguay: Ciudades.

Venezuela, República Bolivariana: Centros con una población de 1,000 habitantes o más.

Asia

Armenia: Ciudades y localidades de tipo urbano, oficialmente designadas como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

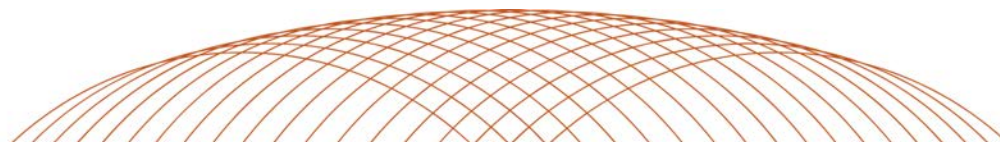
Azerbaiyán: Ciudades y localidades de tipo urbano, oficialmente designadas como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Bahrein: Comunas o pueblos de 2 500 habitantes o más.

Camboya: Ciudades.

China: Las ciudades solo se refieren a las ciudades designadas por el Consejo de Estado. En el caso de ciudades con establecimiento de distrito, la ciudad propiamente dicha se refiere a toda el área administrativa del distrito si su densidad de población es de 1,500 personas por kilómetro o más; o la sede del gobierno del distrito y otras áreas de calles bajo la administración del distrito si la densidad de población es menor a 1,500 personas por kilómetro. En el caso de ciudades sin establecimiento de distrito, la ciudad propiamente dicha se refiere a la sede del gobierno de la ciudad y otras áreas de calles bajo la administración de la ciudad. Para el distrito de la ciudad con una densidad de población inferior a 1,500 personas por kilómetro y la ciudad sin establecimiento de distrito, si la construcción urbana del distrito o la sede del gobierno de la ciudad se ha extendido a alguna parte de las ciudades vecinas designadas o municipio (s), la ciudad propiamente dicha incluye toda el área administrativa de la (s) ciudad (s) o municipio (s).

Chipre: Las áreas urbanas son aquellas definidas por los planes de las ciudades locales.



Georgia: Ciudades y localidades de tipo urbano, oficialmente designadas como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

India: Ciudades (lugares con corporación municipal, comité de área municipal, comité de ciudad, comité de área notificado o junta de acantonamiento); también, todos los lugares con 5 000 o más habitantes, una densidad de no menos de 1,000 personas por milla cuadrada o 400 por kilómetro cuadrado, características urbanas pronunciadas y al menos tres cuartas partes de la población masculina adulta empleada en actividades distintas de la agricultura.

Indonesia: Lugares con características urbanas.

Irán: República Islámica de): cada distrito con un municipio.

Israel: Todos los asentamientos de más de 2,000 habitantes, excepto aquellos en los que al menos un tercio de los hogares, que participan en la fuerza de trabajo civil, se ganan la vida con la agricultura.

Japón: Ciudad (shi) que tiene 50,000 habitantes o más con el 60% o más de las viviendas ubicadas en las principales zonas urbanas y el 60% o más de la población (incluidas sus dependientes) dedicadas a la fabricación, el comercio u otros tipo de negocio urbano. Alternativamente, un shi que tenga las instalaciones y condiciones urbanas definidas por el orden prefectural se considera urbano.

Kazajistán: Ciudades y localidades de tipo urbano, designadas oficialmente como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

República de Corea: Población que vive en ciudades, independientemente del tamaño de la población.

Kirguistán: Ciudades y localidades de tipo urbano, oficialmente designadas como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Malasia: Áreas registradas con una población de 10,000 y más.

Maldivas: Malé, la capital.

Mongolia: Centros de capitales y distritos.

Pakistán: Lugares con corporación municipal, comité de ciudad o acantonamiento.

Sri Lanka: El sector urbano se compone de todas las áreas municipales y municipales.

República Árabe Siria: Ciudades, centros Mohafaza y centros Mantika, y comunidades con 20 000 habitantes o más.

Tayikistán: Ciudades y localidades de tipo urbano, oficialmente designadas como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Tailandia: Áreas municipales.

Turquía: Población de lugares de asentamiento, 20,001 y más.

Turkmenistán: Ciudades y localidades de tipo urbano, designadas oficialmente como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Uzbekistán: Ciudades y localidades de tipo urbano, oficialmente designadas como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Vietnam: Las áreas urbanas incluyen los distritos urbanos de las ciudades, los barrios urbanos y las ciudades. Todas las demás unidades administrativas locales (commues) pertenecen a áreas rurales.

Europa

Albania: Ciudades y otros centros industriales de más de 400 habitantes.

Austria: Comunas de más de 5,000 habitantes.

Bielorrusia: Ciudades y localidades de tipo urbano, designadas oficialmente como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Bulgaria: Ciudades, es decir, localidades legalmente establecidas como urbanas.

República Checa: Localidades con 2,000 habitantes o más.

Estonia: Ciudades y localidades de tipo urbano, designadas oficialmente como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Finlandia: Comunas urbanas. 1970: Localidades.

Francia: Comunas que contienen una aglomeración de más de 2,000 habitantes que viven en casas contiguas o con no más de 200 metros entre casas, también comunas de las cuales la mayor parte de la población es parte de una aglomeración multicomunal de esta naturaleza.

Grecia: Población de municipios y comunas en la que el mayor centro de población tiene 10,000 o más habitantes. Incluyendo también la población de las 18 aglomeraciones urbanas, como se definieron en el censo de 1991, a saber: Gran Atenas, Salónica, Patra, Iraklio, Volos, La Canea, Irannina, Chalkida, Agrinio, Kalamata, Katerini, Kerkyra, Salamina, Quíos, Egio, Rethymno, Ermoupolis y Sparti.

Hungría: Budapest y todas las ciudades legalmente designadas.

Islandia: Localidades de 200 o más habitantes.

Irlanda: Ciudades y pueblos, incluidos los suburbios de 1,500 o más habitantes.

Letonia: Ciudades y localidades de tipo urbano, designadas oficialmente como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Lituania: La población urbana se refiere a las personas que viven en ciudades y pueblos, es decir, las áreas de población con viviendas permanentes estrechamente construidas y con una población residente de más de 3,000 de los cuales 2/3 de los empleados trabajan en la industria, infraestructura social y negocios. En algunas ciudades, la población puede ser inferior a 3,000 ya que estas áreas ya tenían los estados de "ciudad" antes de que se aplicara la ley (julio de 1994).

Países Bajos: Urbano: Municipios con una población de 2,000 y más habitantes. Semiurbano: Municipios con una población inferior a 2,000 habitantes pero con no más del 20% de su población masculina económicamente activa dedicada a la agricultura, y municipios residenciales específicos de viajeros.

Noruega: Localidades de 200 habitantes o más.

Polonia: Ciudades y asentamientos de tipo urbano, p. asentamientos de trabajadores, asentamientos de pescadores, centros de salud.

Portugal: Aglomeración de 10,000 habitantes o más.

República de Moldova: Ciudades y localidades de tipo urbano, oficialmente designadas como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Rumanía: Ciudades, municipios y otras ciudades.

Federación de Rusia: Ciudades y localidades de tipo urbano, designadas oficialmente como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Eslovaquia: 138 ciudades con 5,000 habitantes o más.

España: Localidades de 2,000 habitantes o más.

Suiza: Comunas de 10,000 habitantes o más, incluidos los suburbios.

Ucrania: Ciudades y localidades de tipo urbano, oficialmente designadas como tales, generalmente de acuerdo con los criterios de número de habitantes y predominio de la agricultura, o número de trabajadores no agrícolas y sus familias.

Oceanía

Samoa Americana: Aglomeraciones de 2,500 habitantes o más, que generalmente tienen densidades de población de 1,000 personas por milla cuadrada o más. Dos tipos de áreas urbanas: áreas urbanizadas de 50,000 o más habitantes y grupos urbanos de al menos 2 500 y menos de 50,000 habitantes. (A partir del Censo 2000, no se identifican áreas urbanizadas en Samoa Americana).

Guam: Aglomeraciones de 2,500 habitantes o más, que generalmente tienen densidades de población de 1,000 personas por milla cuadrada o más, denominadas "agrupaciones urbanas".

Nueva Caledonia: Nouméa y las comunas de Païta, Nouvel Dumbéa y Mont-Dore.

Nueva Zelanda: Todas las ciudades, además de distritos municipales, pueblos y ciudades rurales con una población de 1,000 o más.

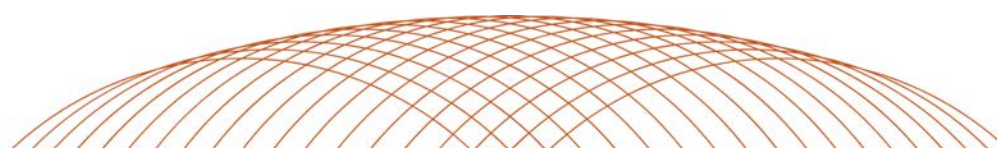
Islas Marianas del Norte: Aglomeraciones de 2,500 habitantes o más, que generalmente tienen densidades de población de 1,000 personas por milla cuadrada o más. Dos tipos de áreas urbanas: áreas urbanizadas de 50,000 o más habitantes y grupos urbanos de al menos 2 500 y menos de 50 000 habitantes.

Vanuatu: Centro de Luganville y Vila urbana.

De los conceptos y métodos utilizados para la delimitación de las ciudades

Las reglas a aplicar para la realización de esta delimitación pueden ser absolutamente contextuales, ya que se trata de resolver una problema muy concreto (por lo general las situaciones complejas se resuelven mediante mecanismos [*ad hoc*]), pero en ningún caso deberían ser ambiguas. En este sentido, Naciones Unidas, al presentar sus estadísticas sobre población urbana, define ésta como aquella que reside en áreas urbanas, y éstas a su vez se definen según los criterios de cada país (UNSTAT, 2003) que dieran, y mucho, tanto en sus aspectos cuantitativos como cualitativos (Capel, 1975; UNSTAT, 2010).

En este caso resulta imprescindible partir de una claridad conceptual: ¿Cuáles son los atributos de lo urbano? ¿Son todos ellos imprescindibles para definir un espacio como urbano? ¿Cómo calificar los espacios que cuentan sólo con algunos de estos atributos? En definitiva, ¿Qué nos dice esta posible calificación/clasificación sobre las características concretas de cada espacio?



Lewis Mumford (1961) se encuentra muy temprano con una aparente paradoja: al tiempo que considera ciudad y civilización como dos realidades inseparables. Para mantener su hipótesis debe plantear una forma peculiar de urbanización. Este ejemplo nos ilustra sobre las dificultades que nos encontramos igualmente al intentar conceptualizar la realidad urbana contemporánea. Nuestras ciudades tienen, cada una, su propia historia y han cambiado en mayor o menor medida también en función de su exposición a las tendencias globales.

Nos encontramos con pequeñas ciudades de provincia que siguen ejerciendo el papel de cabeza administrativa y comercial de comarcas netamente rurales, otras que han crecido hasta tamaños descomunales, mutando a nuevas formas de urbanización, y otras que sin crecer tanto, se han visto envueltas o absorbidas por el crecimiento de ciudades vecinas. Ciudades que se han subido a cada nueva ola de innovación técnica, otras que se han hiper-especializado, otras que parecen haberse quedado estancadas bajo una lenta e inexorable decadencia. Y todo ello dejando huella en términos de patrimonio urbano o arquitectónico, tradiciones culturales, estructuras políticas y sociales, que han llegado al presente en mayor o menor medida, conviviendo con las transformaciones más recientes.

Lo urbano se define por un conjunto de atributos diversos (como vimos arriba) que incluyen aspectos materiales, sociales, políticos y económicos; automáticamente lo rural se define por la ausencia de estas propiedades. Sin embargo, al ser un conjunto heterogéneo de atributos, resulta posible, cuando no habitual, que determinados espacios presenten sólo algunos de ellos, y que además lo hagan con intensidades variables.

Podemos definir el grado de urbanidad de un espacio en función del número de atributos presentes, o jerarquizar dichos atributos y decidir que algunos de ellos son imprescindibles para definir lo urbano y obtener una delimitación precisa.

Evaluación de algoritmos de agrupación

Algoritmo de agrupamiento de ciudades (CCA) ³⁷

Descripción

El método inicia seleccionando una celda poblada arbitraria que se quema (alcanzados). Luego, los vecinos poblados también son quemados. El algoritmo sigue haciendo crecer el clúster quemando iterativamente a los vecinos de las celdas quemadas, hasta que no haya más celdas vecinas pobladas. A continuación, se selecciona otra celda poblada no ha sido quemada y se repite el procedimiento hasta que todas las celdas pobladas se asignan a un grupo.

El City Clustering Algorithm (CCA) se basa en el algoritmo de quemado y se introdujo por primera vez en el contexto de las ciudades. Entre otras cosas, también se utilizó para un estudio global de percolación urbana.

Argumentos

data	Datos a agrupar. Puede ser un ráster, una matriz o un marco de datos. Ver detalles.
-------------	---

³⁷ Ver fuente: <https://rdrr.io/cran/osc/man/cca.html> documentación de OSC compilada el 19 de diciembre de 2019 a las 5:06 p. m. Autores: Steffen Kriewald, Till Fluschnik, Dominik Reusser, Diego Rybski.

s	El radio/tamaño de la carcasa del procedimiento de grabación (es decir, qué tan tolerante es el algoritmo a los pequeños espacios). La unidad es 'número de celdas' si los datos son una matriz numérica y 'metros' si los datos son un ráster o un marco de datos.
x	La posición inicial en la dirección <i>x</i> . Solo se usa si los datos son una matriz numérica.
y	La posición inicial en la dirección <i>y</i> . Solo se usa si los datos son una matriz numérica.
mode	El algoritmo para una matriz no georeferenciada viene en tres versiones que afectan qué celdas cercanas se incluyen en el clúster considerado: (modo=1) vecinos más cercanos (modo=2) celdas dentro de un caparazón (es decir, cuadrados de cierto tamaño) (modo=3) celdas dentro de un radio. Mientras que (modo=1) es equivalente a (modo=3) con $r=1$ y (modo=2) con $r=1$ es equivalente a (modo=3) con $r=2$. Solo se usa si los datos son una matriz numérica.
count.cells	Establezca esta opción en VERDADERO, si desea saber el número de celdas que pertenecen a cada grupo. Solo se usa si los datos son una matriz numérica.
count.max	Esto define el número máximo de clústeres. Se establece de forma predeterminada en $ncol*3$. Solo se usa si los datos son una matriz numérica.
res.y	La resolución del conjunto de datos expresada como la distancia entre dos centros de celda en grados (sistema de coordenadas geográficas). Solo es necesario si <i>data</i> es un <i>data.frame</i> y <i>unit</i> ="m".
res.x	Resolución <i>X</i> . Solo es necesario si <i>data</i> es un <i>data.frame</i> y <i>unit</i> ="m".
cell.class	Solo se requiere si los datos son un ráster. Especifique qué clase de celda (p. ej., tipo de uso del suelo) se agrupará. Puede ser un número entero o un vector y se puede combinar con la opción de comparación.
unit	Si <i>unidad</i> = "m" (metro) el algoritmo de conglomerado se realizará para una distancia de conglomerado en 'metro'. De lo contrario, la agrupación se realiza en grados. Si desea hacer una agrupación en píxeles, elija la resolución como distancia de agrupación.
compare	Si <i>compare</i> = "g", se elegirán las celdas mayores que la clase de celda dada. Si <i>compare</i> = "s" celdas más pequeñas, entonces se elegirá la celda. clase.

Detalles

CCA se implementa en dos versiones, según el formato de los datos. Para matrices numéricas, se llama una versión basada en matrices. Para datos basados en ráster y marcos de datos, se utiliza una versión basada en listas, que es más rápida para matrices dispersas y grandes distancias de clúster. (Ver viñeta, sección: Comparación - matriz vs lista).

Para matriz: La matriz es una matriz numérica simple. Un valor igual a 0 o menor se trata como una celda sin importancia, un valor superior a 0 se trata como una celda de interés. Se identifican grupos de células conectadas.

Para ráster: Se llamará a una subfunción para extraer las coordenadas de un tipo de celda dado (`cell.class`). Además, es posible establecer un umbral que determine qué celdas se pueden quemar usando `compare = "g"` (p. ej., población mínima para considerar una celda como poblada) Pasos siguientes; ver marco de datos.

Para `data.frame`: Un marco de datos con dos columnas que especifican las coordenadas de longitud (primera columna) y latitud (segunda columna). El algoritmo identifica todos los puntos con una distancia entre sí menor que la distancia del grupo `s`. Si `unit="m"` la distancia ortodrómica, de lo contrario se utilizará la distancia euclidiana.

Valor

Para matriz: Una matriz que define para cada celda a qué grupo pertenece.

Para raster / `data.frame`:

Lista con dos entradas: 1. marco de datos con las coordenadas de longitud y latitud de las celdas y el `cluster_id`. y 2. un vector que da el tamaño del grupo en unidades de la celda primitiva. El primer número es el tamaño del clúster con `cluster_id` 1, el segundo el tamaño del clúster con `cluster_id` 2 y así sucesivamente.

Clustering basado en densidad DBSCAN y variantes

38

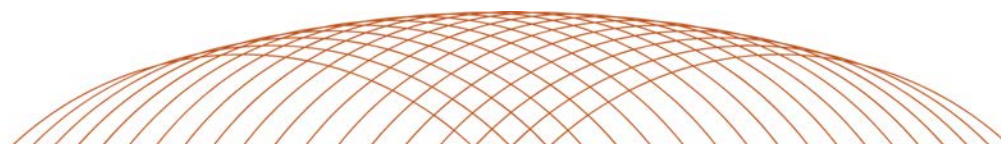
La herramienta *Clustering* basado en densidad se basa en la detección de en qué áreas existen concentraciones de puntos y dónde están separados por áreas vacías o con escasos puntos. Los puntos que no forman parte de un clúster se etiquetan como ruido. Opcionalmente, el tiempo de los puntos se puede usar para buscar grupos de puntos que se agrupan en clústeres de espacio y tiempo.

Esta herramienta utiliza algoritmos de *clustering* de aprendizaje de máquina no supervisados que detectan automáticamente patrones basándose puramente en la ubicación espacial y en la distancia a un número de vecinos especificado. Estos algoritmos se consideran no supervisados porque no requieren ninguna formación sobre en qué consiste un clúster.

Sugerencia:

El clustering, el agrupamiento y la clasificación son algunas de las técnicas más utilizadas en aprendizaje de máquina. El Clustering multivariante y la herramienta Clustering multivariante restringido espacialmente también utilizan métodos no supervisados de aprendizaje de máquina para determinar los clústeres naturales presentes en sus datos. Estos métodos de clasificación se consideran no supervisados porque no requieren un conjunto de entidades previamente clasificadas para guiar o entrenar, ni para encontrar clústeres en sus datos.

³⁸ ArcGIS Pro 2.8|Otras versiones| Archivo de ayuda. [Clustering basado en densidad](#)



Potenciales aplicaciones

Estas son algunas formas en que se puede aplicar esta herramienta:

- Las redes de suministro urbano de aguas son un importante activo subterráneo oculto. El clustering de las roturas y reventones de tuberías puede indicar problemas en ciernes. Mediante la herramienta Clustering basado en densidad, los ingenieros pueden determinar dónde se encuentran esos clústeres y tomar medidas preventivas en las zonas de alto riesgo de sus redes de suministro de aguas.
- Suponga que dispone de datos de posición de todos los tiros con enceste y sin enceste de los jugadores de la NBA. La herramienta Clustering basado en densidad puede mostrarle los diferentes patrones de posiciones de tiro con enceste y sin enceste de cada jugador. A continuación, esta información se puede usar como base para desarrollar estrategias de juego.
- Suponga que está estudiando una enfermedad concreta propagada por plagas y tiene un *dataset* de puntos que representa a los hogares de su área de estudio, algunas de ellas infestadas y otras no. Mediante la herramienta Clustering basado en densidad, puede determinar los mayores clústeres de hogares infestados para ayudarlo a delimitar un área en la que iniciar el tratamiento y la erradicación de las plagas.
- La geolocalización de tuits después de una catástrofe natural o un ataque terrorista puede organizarse en clústeres para informar sobre las necesidades de rescate y evacuación en función del tamaño y la ubicación de los clústeres identificados.

Métodos de clustering

El parámetro **Métodos de clustering** de la herramienta Clustering basado en densidad ofrece tres opciones con las que se pueden encontrar clústeres en sus datos de puntos:

- **Distancia definida (DBSCAN):** Utiliza una distancia especificada para separar los clústeres densos del ruido más disperso. El algoritmo DBSCAN es el método de clustering más rápido, pero solo es apropiado si se puede utilizar una **Distancia de búsqueda** muy clara, y funciona bien con todos los clústeres potenciales. Para ello, se requiere que todos los clústeres significativos presenten densidades similares. Este método también permite utilizar los parámetros **Campo de tiempo** e **Intervalo de tiempo de búsqueda** para buscar clústeres de puntos en espacio y tiempo.
- **Autoajuste (HDBSCAN):** Utiliza toda una variedad de distancias para separar clústeres de densidades variables del ruido más disperso. El algoritmo HDBSCAN es el método de clustering que más se basa en los datos y, por tanto, requiere la menor participación del usuario.
- **Escala múltiple (OPTICS):** Utiliza la distancia entre entidades vecinas para crear un diagrama de alcanzabilidad que, a continuación, se utiliza para separar los clústeres de densidades diferentes respecto del ruido. El algoritmo OPTICS ofrece la mayor flexibilidad en el afinamiento de los clústeres detectados, a pesar de que requiere una gran potencia de cómputo, en especial si la **Distancia de búsqueda** es grande. Este método también permite utilizar los parámetros **Campo de tiempo** e **Intervalo de tiempo de búsqueda** para buscar clústeres de puntos en espacio y tiempo.

Esta herramienta toma **Entidades de puntos de entrada**, una ruta para las **Entidades de salida** y un valor que representa el número mínimo de entidades requerido para considerar

que existe un clúster. En función del **Método de clustering** seleccionado, puede ser necesario especificar parámetros adicionales, como se describe a continuación.

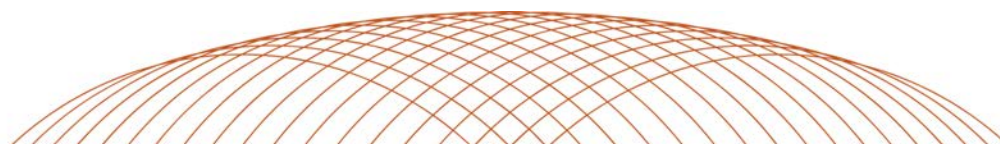
Entidades mínimas por clúster

Este parámetro determina el número mínimo de entidades requerido para considerar que un agrupamiento de puntos es un clúster. Por ejemplo, si tiene varios clústeres diferentes con un tamaño de entre 10 y 100 puntos y selecciona un valor de 20 en **Entidades mínimas por clúster**, todos los clústeres que tengan menos de 20 puntos se considerarán como ruido (porque no forman un agrupamiento lo suficientemente grande como para ser considerado un clúster), o bien se fusionan con los clústeres cercanos para satisfacer el número mínimo de entidades requerido. Por contra, al seleccionar un valor de **Entidades mínimas por clúster** menor del que consideraría su clúster significativo más pequeño, podría ocurrir que los clústeres significativos quedaran divididos en clústeres más pequeños. En otras palabras, cuanto más pequeño es el valor de **Entidades mínimas por clúster**, más clústeres se detectan. Cuanto mayor es el valor de **Entidades mínimas por clúster**, menos clústeres se detectan.

Sugerencia:

El valor ideal de **Entidades mínimas por clúster** depende de qué está intentando capturar y la pregunta que se hace en su análisis. Debe ser el agrupamiento de menor tamaño que desee considerar como un clúster significativo. Un aumento del valor de **Entidades mínimas por clúster** podría dar lugar a la fusión de algunos de los clústeres más pequeños.

El parámetro **Entidades mínimas por clúster** también es importante para el cálculo de la distancia de núcleo, que es una medida utilizada por los tres métodos a la hora de encontrar clústeres. Conceptualmente, la distancia de núcleo de cada punto es una medida de la distancia que es necesario recorrer desde cada punto hasta el número mínimo de entidades. Por tanto, si se elige un valor elevado en **Entidades mínimas por clúster**, la distancia de núcleo correspondiente será mayor. Por tanto, si se elige un valor pequeño en **Entidades mínimas por clúster**, la distancia de núcleo correspondiente será menor. La distancia de núcleo está relacionada con el parámetro **Distancia de búsqueda**, que se utiliza tanto en el método **Distancia definida (DBSCAN)** como en el método **Escala múltiple (OPTICS)**.



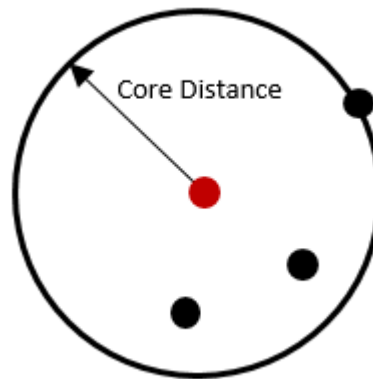
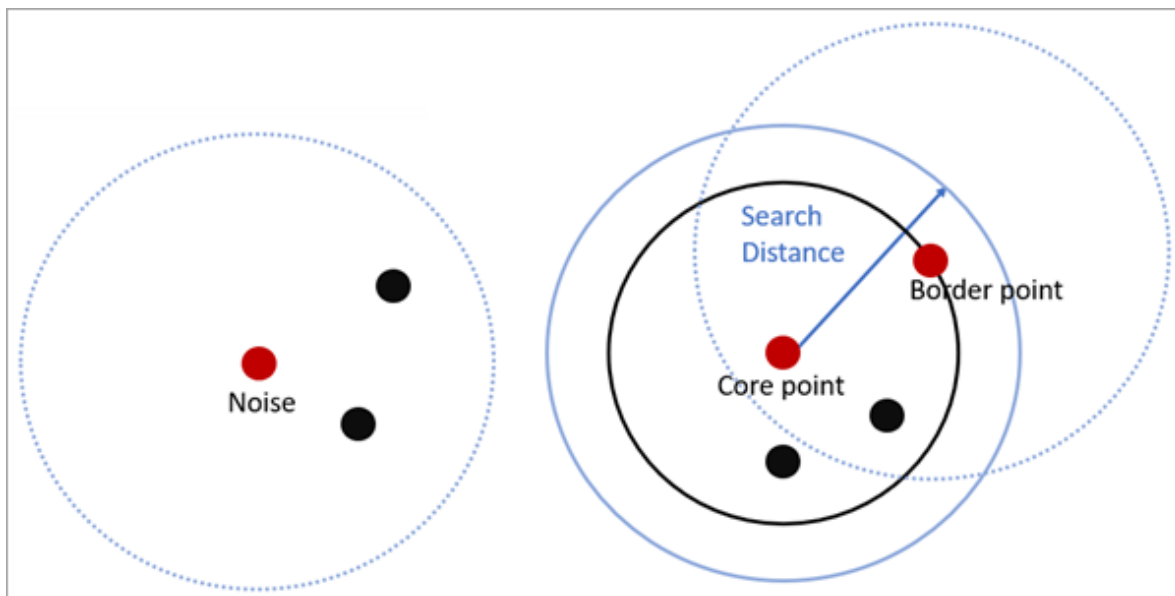


Ilustración de la distancia de núcleo, medida como la distancia desde una entidad determinada que es necesario recorrer para crear un clúster con un mínimo de 4 entidades, incluida la propia entidad.

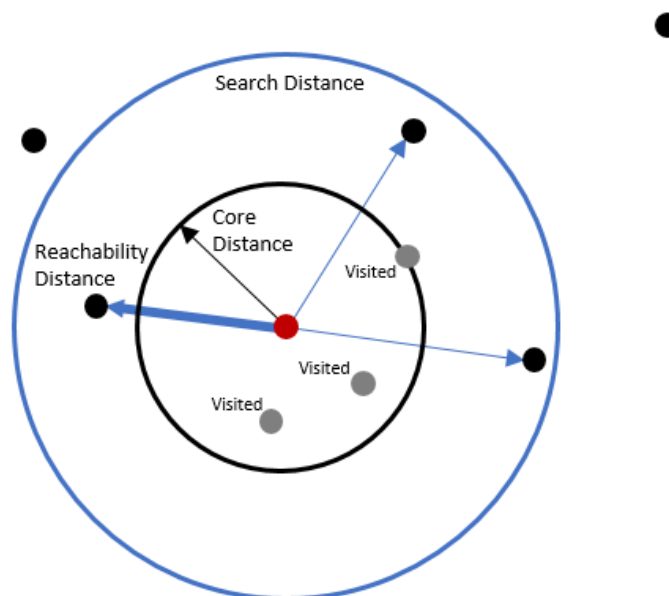
Distancia de búsqueda (DBSCAN y OPTICS)

En **Distancia definida (DBSCAN)**, si el número de **Entidades mínimas por clúster** se encuentra dentro de la **Distancia de búsqueda** a partir de un punto en particular, ese punto se marcará como punto de núcleo y se incluirá en un clúster, junto con todos los puntos dentro de la distancia de núcleo. Un punto de borde es un punto que se encuentra dentro de la distancia de búsqueda de un punto de núcleo, pero no tiene el número mínimo de entidades dentro de la distancia de búsqueda. Cada clúster resultante se compone de puntos de núcleo y puntos de borde, donde los puntos de núcleo tienden a estar en el centro del clúster y los puntos de borde se encuentran en el exterior. Si un punto no tiene el número mínimo de entidades dentro de la distancia de búsqueda y no está dentro de la distancia de búsqueda de otro punto de núcleo (es decir, no es ni un punto de núcleo ni un punto de borde), se marca como punto de ruido y no se incluye en ningún clúster.



Se muestra un clustering de cuatro entidades mínimas por clúster. El punto de núcleo tiene cuatro vecinos dentro de la distancia de búsqueda (incluido él mismo). El punto de borde solo tiene tres entidades dentro de la distancia de búsqueda, pero es vecino de un punto de núcleo, de modo que se incluye en el clúster. El punto de ruido no tiene cuatro entidades dentro de la distancia de búsqueda y no es vecino de un punto de núcleo, por lo que no se incluye en un clúster.

En el caso de **Escala múltiple (OPTICS)**, el valor de **Distancia de búsqueda** se trata como la distancia máxima que se comparará con la distancia de núcleo. **Escala múltiple (OPTICS)** utiliza un concepto de distancia mínima alcanzable que es la distancia desde un punto hasta su vecino más cercano que aún no se haya visitado en la búsqueda. (Nota: OPTICS es un algoritmo ordenado que comienza con la entidad que presenta el Id. más bajo y va desde este punto al siguiente para crear un diagrama. El orden de los puntos es fundamental para el resultado). **Escala múltiple (OPTICS)** busca todas las distancias vecinas dentro de la **Distancia de búsqueda** especificada, comparando cada una de estas distancias con la distancia de núcleo. Si cualquier distancia es menor que la distancia de núcleo, se asigna a la entidad esa distancia de núcleo como su distancia de alcanzabilidad. Si todas las distancias son mayores que la distancia de núcleo, la más pequeña de estas distancias se asigna como distancia de alcanzabilidad. Si no hay más puntos dentro de la distancia de búsqueda, el proceso se reinicia en un nuevo punto que no se ha visitado anteriormente. En cada iteración, las distancias de alcanzabilidad se recalculan y se ordenan. La menor de las distancias se utiliza para la distancia de alcanzabilidad final de cada punto. A continuación, estas distancias de alcanzabilidad se utilizan para crear el diagrama de alcanzabilidad, que es un diagrama ordenado de las distancias de alcanzabilidad utilizadas para detectar los clústeres.



Una vez visitadas todas las entidades presentes dentro de la distancia de núcleo, la Distancia de alcanzabilidad asignada a una entidad es la distancia más pequeña entre la entidad seleccionada y todas las demás entidades presentes dentro del umbral de **Distancia de búsqueda**.

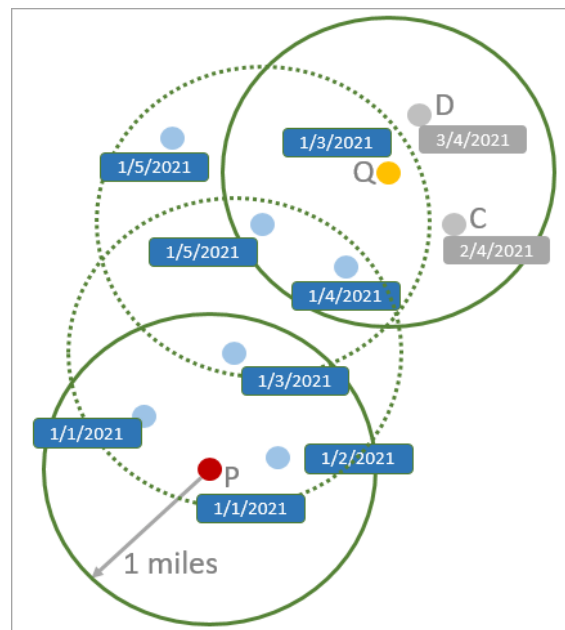
Tanto para **Distancia definida (DBSCAN)** como para **Escala múltiple (OPTICS)**, si no se especifica ninguna distancia, la **Distancia de búsqueda** predeterminada es la distancia de núcleo más alta encontrada en el dataset, excluidas aquellas distancias de núcleo que se encuentren dentro del 1 % superior (en otras palabras, se excluyen las distancias de núcleo más extremas).

Clustering en espacio y tiempo (DBSCAN y OPTICS)

En dos de los métodos de clustering, se puede proporcionar el tiempo de cada punto en el parámetro **Campo de tiempo**. Si se proporciona, la herramienta buscará clústeres de puntos cercanos entre sí en espacio y tiempo. El parámetro **Intervalo de tiempo de búsqueda** es análogo a la distancia de búsqueda y se utiliza para determinar los puntos de núcleo, los puntos de borde y los puntos de ruido de los clústeres de espacio-tiempo.

Para **Distancia definida (DBSCAN)**, al buscar miembros de clústeres, las **Entidades mínimas por clúster** se deben encontrar dentro de los valores de **Distancia de búsqueda** e **Intervalo de tiempo de búsqueda** para que sean un punto de núcleo de un clúster de espacio-tiempo.

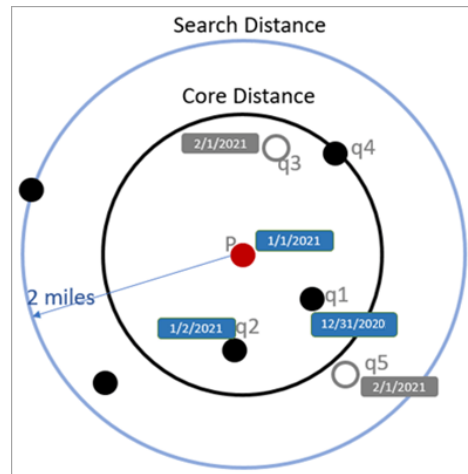
En la siguiente imagen, la distancia de búsqueda es de 1 milla, el intervalo de tiempo de búsqueda es de 3 días y el número mínimo de entidades es 4. El punto P es un punto de núcleo porque hay cuatro puntos dentro de la distancia y el intervalo de tiempo de búsqueda (incluido él mismo). El punto Q es un punto de borde porque solo hay tres puntos dentro de la distancia y el intervalo de tiempo de búsqueda, pero el punto se encuentra dentro de la distancia y el intervalo de tiempo de búsqueda de un punto central del clúster (el punto con marca de tiempo 1/5/2021). Los puntos C y D son puntos de ruido porque no son puntos de núcleo ni puntos de borde del clúster, y todos los demás puntos (incluidos P y Q) forman un clúster único.



Para **Escala múltiple (OPTICS)**, todos los puntos que queden fuera del rango de **Intervalo de tiempo de búsqueda** de un punto se excluirán cuando el punto calcule su distancia

de núcleo, busque todas las distancias vecinas dentro del valor de **Distancia de búsqueda** especificado y calcule la distancia de alcanzabilidad.

En la siguiente imagen, la distancia de búsqueda es de 2 millas, el intervalo de tiempo de búsqueda es de 3 días y el número mínimo de entidades es 4. La distancia de núcleo del punto P se calcula omitiendo el punto q3 porque está fuera del intervalo de tiempo de búsqueda. En este caso, la distancia de núcleo de P es igual a la distancia de alcanzabilidad de P.



La siguiente imagen muestra el cálculo de la distancia de alcanzabilidad cuando ya se han visitado algunos puntos vecinos. Los puntos q1, q2 y q4 se omitieron porque se visitaron anteriormente y puntos q3 y q5 se omitieron porque están fuera de la ventana de tiempo de búsqueda.

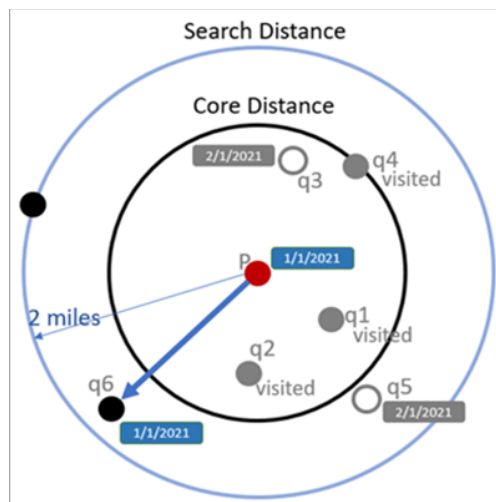
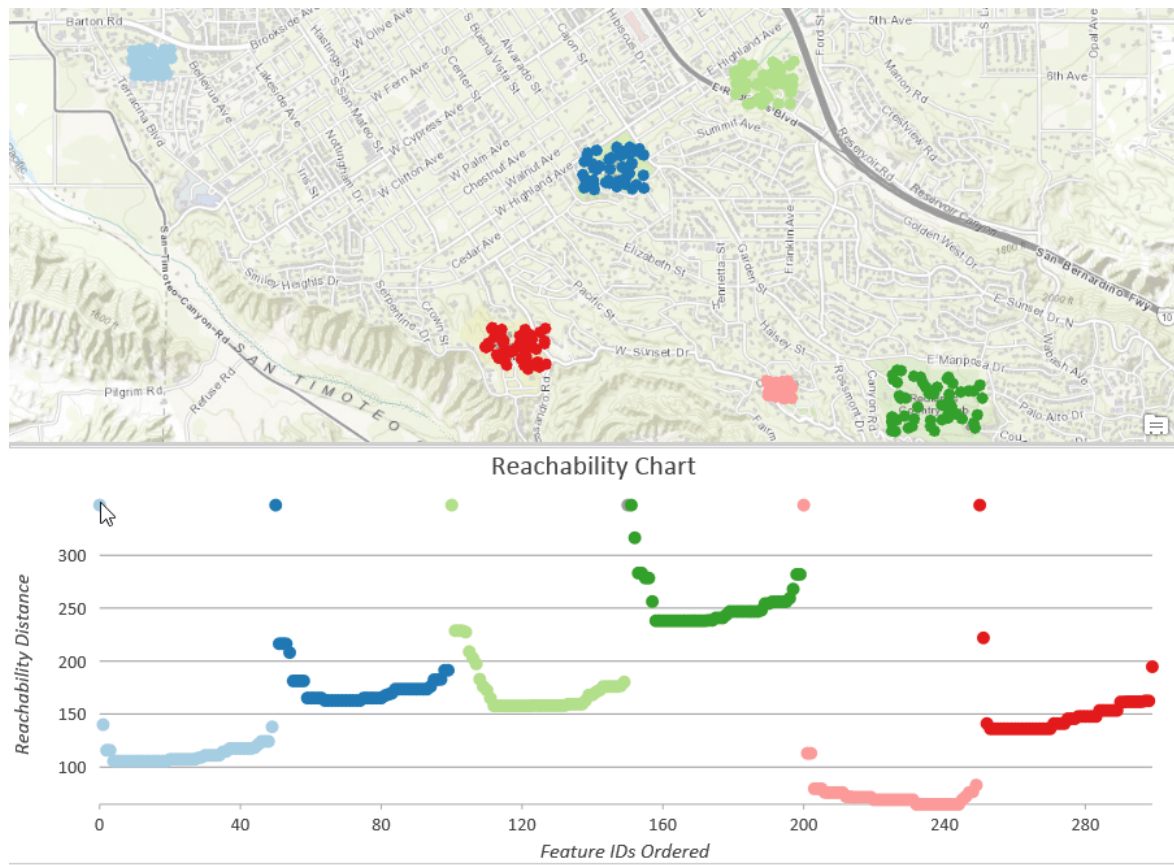
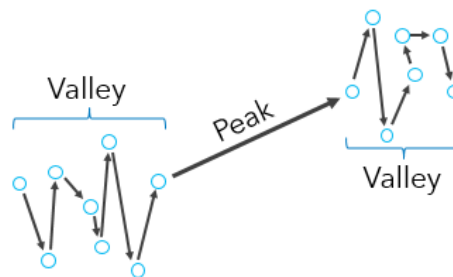


Diagrama de alcanzabilidad (OPTICS)

Una vez calculadas todas las distancias de alcanzabilidad de todo el dataset, se construye un diagrama de alcanzabilidad que ordena las distancias y revela la estructura de clustering de los puntos.



La presencia de valles en el diagrama de alcanzabilidad implica que es necesario recorrer una distancia corta de un punto al siguiente. Por tanto, los valles representan clústeres diferenciados dentro del patrón de puntos. Cuanto más denso es un clúster, menores serán las distancias de alcanzabilidad, y más bajo es el valle en el diagrama (el clúster de color rosa, por ejemplo, es el más denso del ejemplo anterior). Los clústeres menos densos presentan distancias de alcanzabilidad mayores y mayores valles en el diagrama (el clúster en verde oscuro, por ejemplo, es el menos denso del ejemplo anterior). Los picos representan las distancias que es necesario recorrer para viajar de un clúster a otro, o de un clúster a ruido y de nuevo a un clúster, en función de la configuración del conjunto de puntos.

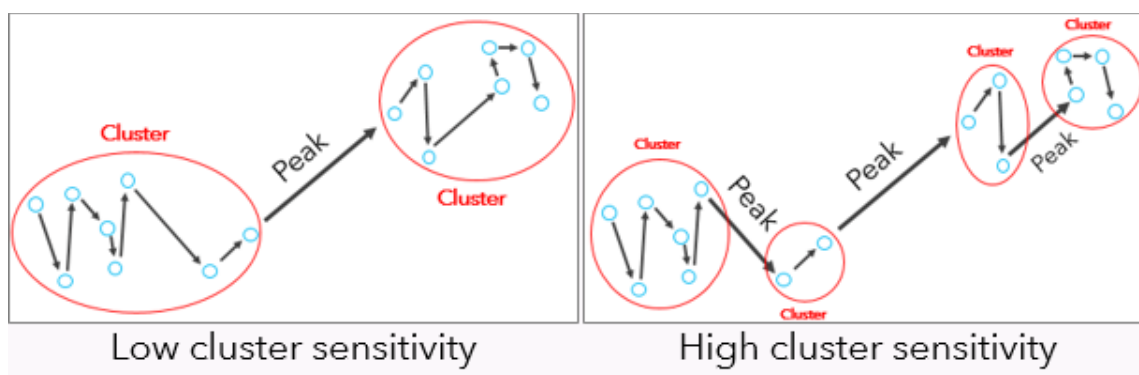


Las distancias de alcanzabilidad forman picos y valles en el diagrama de alcanzabilidad. La medición de distancias más largas entre puntos da lugar a resultados de pico en el diagrama de alcanzabilidad.

El diagrama de alcanzabilidad puede presentar mesetas si la **Distancia de búsqueda** es menor que la distancia de núcleo más grande. Un aspecto clave de utilizar el método de clustering OPTICS es determinar cómo detectar los clústeres en el diagrama de alcanzabilidad, lo cual se realiza mediante el parámetro **Sensibilidad de clúster**.

Sensibilidad de clúster (OPTICS)

El parámetro **Sensibilidad de clúster** determina cómo la forma (tanto la pendiente como la altura) de los picos dentro del diagrama de alcanzabilidad se utilizará para separar los clústeres. Un valor de **Sensibilidad de clúster** muy elevado (cercano a 100) tratará incluso el pico más pequeño como una separación entre clústeres, lo que resulta en un mayor número de clústeres. Un valor de **Sensibilidad de clúster** muy bajo (cercano a 0) tratará únicamente los picos más inclinados y elevados como una separación entre clústeres, lo que resulta en un menor número de clústeres.

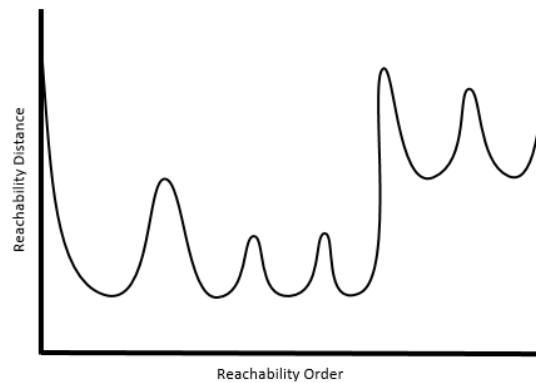


Baja Sensibilidad de clúster frente a alta Sensibilidad de clúster

El valor predeterminado de **Sensibilidad de clúster** se calcula como el umbral en el cual agregar más clústeres no agrega información adicional; se basa en la Divergencia de Kullback-Leibler entre el diagrama de alcanzabilidad original y el diagrama de alcanzabilidad suavizado que se obtiene tras el clustering.

Comparación de los métodos

Aunque solo **Escala múltiple (OPTICS)** utiliza el diagrama de alcanzabilidad para detectar clústeres, el diagrama puede usarse para explicar conceptualmente las diferencias entre estos métodos. A modo de ilustración, el diagrama de alcanzabilidad que aparece a continuación se utiliza para explicar las diferencias entre los 3 métodos. El diagrama revela clústeres con distintas densidades y distancias de separación. Exploraremos el resultado de usar cada uno de los métodos de clustering con estos datos ilustrativos.



Un diagrama de alcanzabilidad conceptual que contiene clústeres de puntos con densidades diferentes y separados por distintas distancias.

Con **Distancia definida (DBSCAN)**, imagine que dibuja una línea a través del diagrama de alcanzabilidad en la **Distancia de búsqueda** especificada. Las áreas situadas por debajo de la **Distancia de búsqueda** representan clústeres, mientras que los picos situados por encima de la **Distancia de búsqueda** representan puntos de ruido. **Distancia definida (DBSCAN)** es el método de clustering más rápido, pero solo es apropiado si se puede utilizar una **Distancia de búsqueda** muy clara como valor de corte y esta **Distancia de búsqueda** funciona bien con todos los clústeres. Para ello, se requiere que todos los clústeres significativos presenten densidades similares.

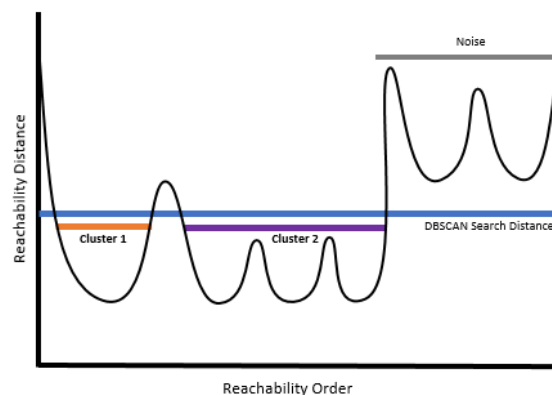


Ilustración de **Distancia de búsqueda** en el algoritmo DBSCAN

Con **Autoajuste (HDBSCAN)**, las distancias de alcanzabilidad son como niveles anidados de clústeres. Cada nivel de clustering resultaría en la detección de un conjunto de clústeres diferente. **Autoajuste (HDBSCAN)** elige qué nivel de clústeres de cada serie de clústeres anidados creará óptimamente los clústeres más estables, incorporando tantos miembros como sea posible sin incorporar ruido. Para más información sobre el algoritmo, consulte la excelente documentación de los creadores de HDBSCAN. **Autoajuste (HDBSCAN)** es el método de clustering que más se basa en los datos y, por tanto, requiere la menor participación del usuario.

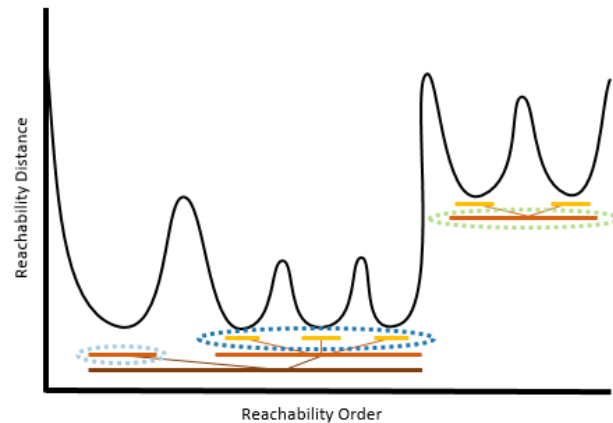


Ilustración de los niveles jerárquicos utilizados por el algoritmo HDBSCAN para encontrar los clústeres óptimos y maximizar la estabilidad

En **Escala múltiple (OPTICS)**, el trabajo de detección de clústeres no se basa en una distancia en particular, sino en los picos y valles dentro del diagrama. Digamos que cada pico tiene un nivel que puede ser Pequeño, Mediano o Grande.

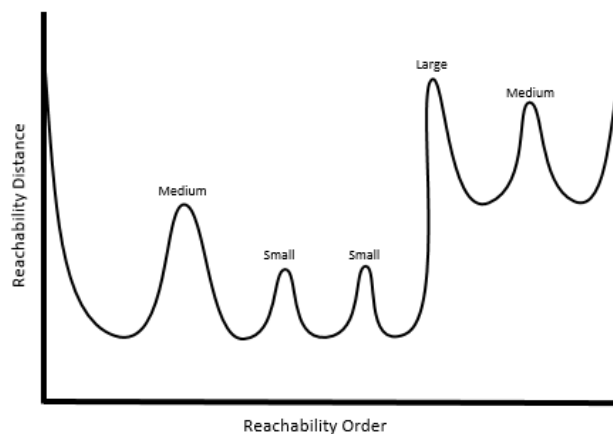
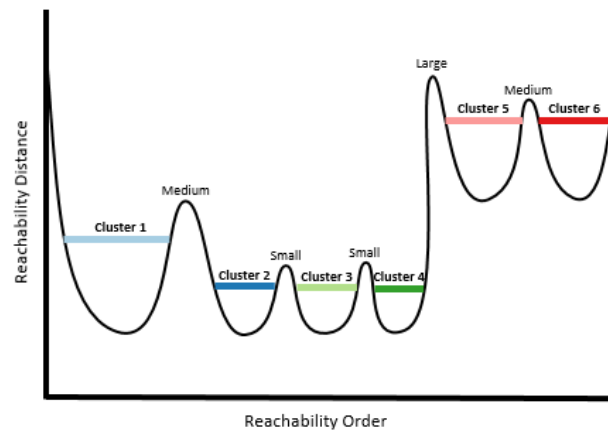


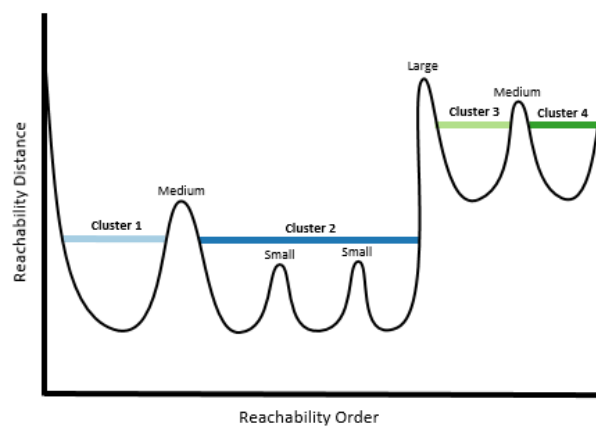
Ilustración de la intensidad de los picos en el diagrama de alcanzabilidad

En esencia, cuando se elige un valor de **Sensibilidad de clúster** muy elevado, todos los picos, ya sean pequeños o grandes, actuarán como separaciones entre clústeres (lo que resulta en un mayor número de clústeres).



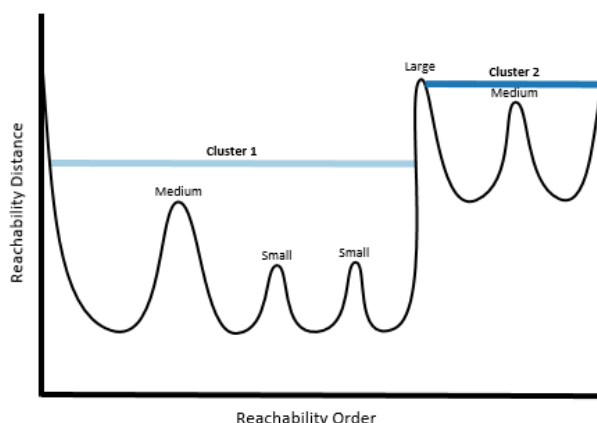
Una ilustración del impacto de un valor elevado en **Sensibilidad de clúster** empleado por OPTICS, con los clústeres correspondientes

Al seleccionar un valor moderado en **Sensibilidad de clúster**, se utilizarán los picos medianos y grandes, pero no los picos pequeños.



Una ilustración del impacto de un valor moderado en Sensibilidad de clúster empleado por OPTICS, con los clústeres correspondientes

Si se selecciona un valor muy bajo en **Sensibilidad de clúster**, solo se utilizarán los picos grandes, lo que resulta en la detección del mínimo número posible de clústeres.



Una ilustración del impacto de un valor bajo en el parámetro Sensibilidad de clúster empleado por OPTICS, con los clústeres correspondientes

Escala múltiple (OPTICS) ofrece la mayor flexibilidad en el afinamiento de los clústeres detectados, aunque también es el más lento de los tres métodos de clustering.

Resultados

Esta herramienta produce una clase de entidad de salida con un nuevo campo de tipo entero **CLUSTER_ID**, que le muestra de qué clúster es miembro cada entidad. La representación en pantalla predeterminada se basa en el campo **COLOR_ID**. Si hay varios clústeres, se asigna un color a cada uno. Los colores se asignarán y repetirán de forma que cada clúster sea visualmente diferente de sus clústeres vecinos.

Si el **Método de clustering** elegido fue **Autoajuste (HDBSCAN)**, la clase de entidad de salida también contendrá estos campos: **PROB**, que es la probabilidad a la que pertenece la entidad en su grupo asignado; **OUTLIER**, que designa a la entidad quizá un valor atípico dentro de su propio clúster (cuanto mayor sea el valor, la entidad tendrá mayor probabilidad de ser un valor atípico) y **EXEMPLAR**, que denota las entidades más prototípicas o más representativas de cada clúster.

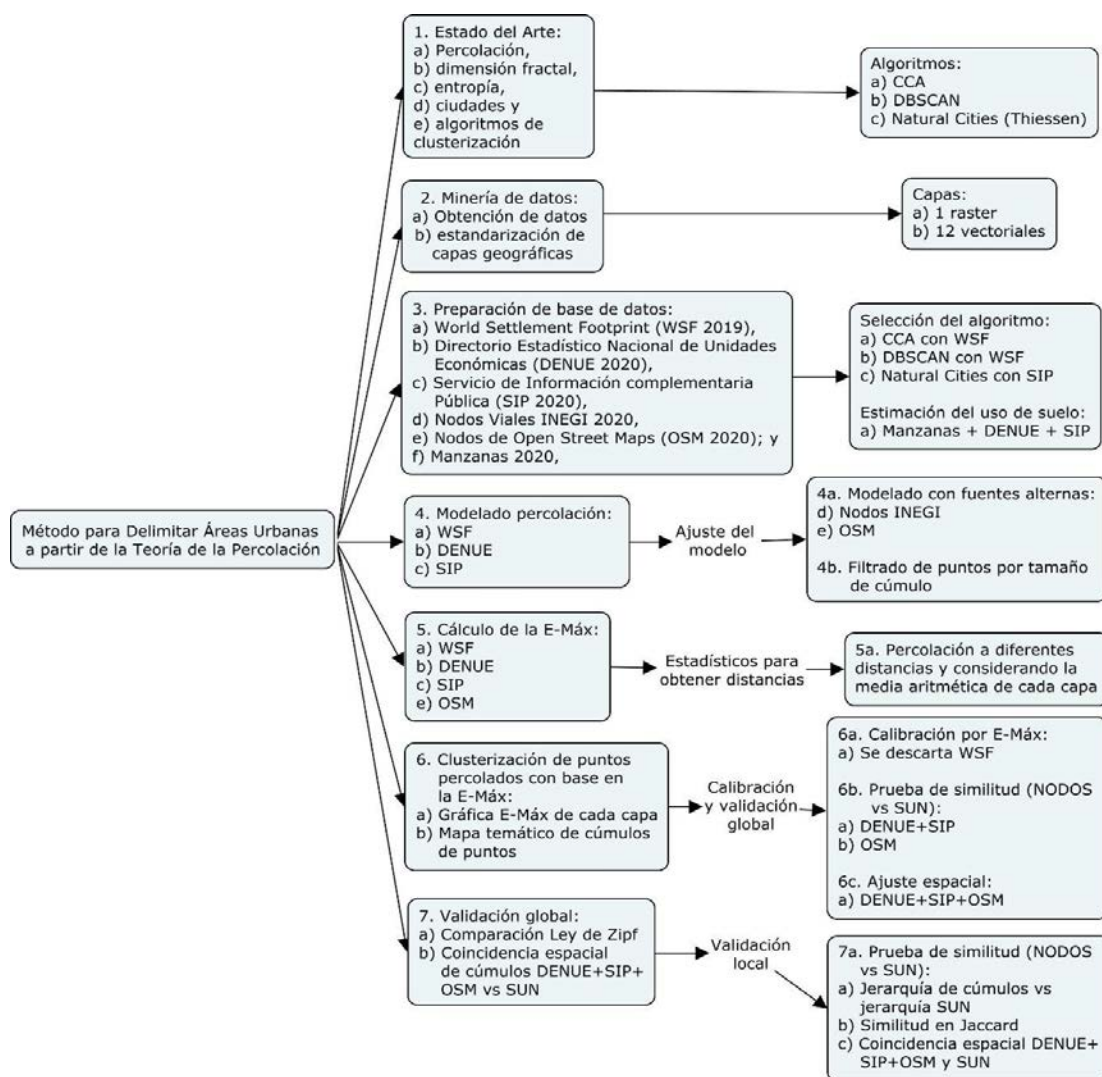
Esta herramienta también crea mensajes y gráficos para ayudarle a comprender las características de los clústeres identificados. Puede acceder a los mensajes desplazándose sobre la barra de progreso, haciendo clic en el botón emergente y ampliando la sección **Mensajes** en el panel **Geoprocesamiento**. También puede acceder a los mensajes de la anterior ejecución de la herramienta Clustering basado en densidad a través del historial de geoprocesamiento. Los gráficos están disponibles desde el panel **Contenido**.

Además del diagrama de alcanzabilidad que se crea cuando se elige **Escala múltiple (OPTICS)**, todos los métodos de clustering crean también un gráfico de barras que presenta todos los Id. de clúster únicos. Este diagrama se puede utilizar para seleccionar fácilmente todas las entidades pertenecientes a un clúster específico, así como para explorar el tamaño de cada uno de los clústeres.

III. Enfoque metodológico (inicial y ajustes)

Se parte del supuesto de que la similitud encontrada entre los valores máximos de la Entropía de Shannon al momento de percolar a diferentes distancias variables (en este caso de puntos) que representan un sistema de asentamientos humanos, puede ser de gran utilidad para definir límites urbanos. Con base en lo anterior, se espera que el método capte una mayor complejidad de relaciones intraurbanas y menor complejidad de relaciones inter-urbanas, permitiendo ello la definición más precisa de los límites urbano-rural –dada la asunción de una mayor complejidad en zonas urbanas–.

Para comprobar la hipótesis de partida, se presenta el siguiente esquema que simplifica las actividades desarrolladas de la investigación; como la aplicación de procesos geoestadísticos para la aplicación del método.



Fuente: Elaboración propia.

A continuación se describen las etapas presentadas en el esquema conceptual:

1. Revisión y conjuntar el Estado del Arte en materia de percolación, dimensión fractal, entropía, complejidad en ciudades y diferentes métodos de obtención de cúmulos. Se logró conjuntar y revisar los artículos relacionados con los conceptos de la Teoría de la Percolación (17), para la Entropía (9), algoritmos DBSCAN y CCA (2) y para el concepto de Ciudad (12).
 2. Minería de datos para la obtención de pre-procesamientos de bases de datos (fuentes oficiales e internas). Se recopilaron y procesaron las siguientes bases de datos geográficas del INEGI y de la Agencia Espacial Europea (geobase WSF). Archivo WSF_2015_250_epsg_3857.tif recortada para el territorio nacional. 12 capas geográficas del Marco Geoestadístico Nacional del INEGI 2020. Dos capas geográficas procesadas con el Sistema Urbano Nacional 2018 y con la Estrategia Nacional de Ordenamiento del Territorio ENOT, 2020.
 3. Preparación de base de datos: Clasificación de las actividades del DENUÉ-INEGI, en categorías diferentes (Industria, Comercio y Servicios); geolocalización a nivel manzana para todo el país, el uso del suelo y vivienda e imagen raster a escala adecuada para el modelo de percolación. Las siguientes bases geográficas se estandarizaron y preprocesaron para su incorporación al modelo: Capa geográfica del Directorio Estadístico Nacional de Unidades Económicas DENUÉ 2020. Archivo WSF_2015_250_epsg_3857.tif recortada para el territorio nacional y resampleado a 30x30 mts. Capa geográfica a nivel de manzana del Uso Estimado del Suelo Urbano 2020 a nivel nacional. Este producto se obtuvo del procesamiento y reclasificación de las categorías obtenidas del DENUÉ, SCINCE (capa SIP) 2020 y el Marco Geoestadístico Nacional 2020.
 4. Aplicación y selección del algoritmo para la percolación con base en los resultados obtenidos del modelado con dos métodos y a diferentes distancias.
 5. Se calculó la E-Máx para cada una de las cuatro geobases de puntos de todo el país. Además se obtuvieron los estadísticos descriptivos para la distribución de puntos de cada base geográfica.
 6. Selección del método de percolación y modelado utilizando los valores de entropía y elaboración de mapas comparativos tipo raster que representan los valores máximos de la entropía mediante un índice de similitud.
 7. Validación de resultados contrastados con el SUN y los índices de correspondencia espacial, similitud de Jaccard y la jerarquía de cúmulos con la Ley de Zipf.
- La aplicación del modelo conceptual-procedimental se presenta en el apartado V de este documento.

IV. Minería de datos y procesamiento de las geobases de datos

Tanto el proceso de minería de datos; como el procesamiento de las capas obtenidas, se pueden dividir en tres momentos:

- a) En un primer momento, como inicio de los trabajos, se realizó la recopilación de bases de datos de dos principales fuentes: datos geográficos del INEGI y de la Agencia Espacial Europea (geobase que será referida como WSF 2019). Para la primera fuente, se obtuvieron un total de 12 capas geográficas del Marco Geoestadístico Nacional del INEGI para el año censal 2020. De las cuales, se seleccionó 1. El Directorio Estadístico Nacional de Unidades Económicas DENUÉ 2020, 2. Capa geográfica a nivel de manzana 2020 de cobertura nacional y, 3. Capa geográfica de puntos Servicios con Información complementaria de tipo Puntual, derivada del Marco Geoestadístico, INEGI 2020. Esta última capa contiene la localización puntual de los equipamientos públicos, áreas verdes e infraestructura dotada por el Estado. La primera capa contiene la localización puntual de las Unidades Económicas (UE), clasificadas en función del Sistema de Clasificación Industrial de América del Norte (SCIAN-2018). La segunda fuente del orden anterior, que se generó fue la estimación del uso del suelo urbano al año 2020. Esta capa se consideró como una posible variable para afinar el modelo; la cual representa a nivel de polígonos de manzanas el uso habitacional y sus densidades; así como la mezcla de usos a partir del cruce con los puntos del DENUÉ 2020. Esta *geobase* se calculó a nivel nacional, considerando la posibilidad de adherir una variable de aprovechamiento del suelo para afinar el modelo³⁹.

El proceso de preparación de las bases de datos siguió el siguiente orden:

1. Descarga de la base de datos geográfica tipo ráster: World Settlement Footprint (WSF) 2019. Se tuvo acceso al Repositorio para la descarga de una capa binaria de resolución de 10 m por píxel; la cual describe el alcance de los asentamientos humanos a nivel mundial derivado por medio de imágenes multitemporales Sentinel-1 y Sentinel-2 de 2019. El conjunto de datos está organizado en 5,138 archivos GeoTIFF (proyección EPSG4326), cada uno de los cuales hace referencia a una porción de tamaño de 2x2 grados (~222x222 km en el suelo) con un búfer adicional de 0,1 grados para evitar cualquier discontinuidad entre teselas vecinas. Cada mosaico se identifica por las coordenadas de la esquina inferior izquierda especificadas en el nombre del archivo [por ejemplo, el mosaico WSF2019_v1_12_18.tif cubre el área entre (12E;18N) y (14E;20N)]. Los asentamientos están asociados al valor 255; todos los demás píxeles están asociados con el valor 0. El WSF2019 tiene licencia bajo CC-BY-4.0. Actualmente se está finalizando una publicación completa con todos los detalles técnicos y cifras de precisión. Por el momento, consulte: Marconcini, M., Metz-Marconcini, A., Esch, T., Gorelick, N. (2021). Comprender las tendencias actuales en la urbanización global: la suite World Settlement Footprint. *GI_Forum*, 1, 33-38. https://doi.org/10.1553/giscience2021_01_s33. Para obtener más información, póngase en contacto con: mattia.marconcini@dlr.de

³⁹ Ver artículo publicado en: <https://rde.inegi.org.mx/index.php/2023/01/03/estimacion-de-usos-del-suelo-con-base-en-datos-abiertos-para-la-planeacion-urbana-en-mexico/>

2. Se descargaron todos los mosaicos tipo TIFF
3. Los mosaicos se reproyectaron a EPSG3857
4. Se reclasificaron para que estuvieran a 1 BIT
5. Se resampearon para obtener pixeles de 30X30m
6. Se realizó la unión de todos los mosaicos
7. Se extrajeron a partir de un Clip, los mosaicos correspondientes a México.
8. Posteriormente, se generaron los centroides de cada pixel, dando como resultado un total de 41'120,102 puntos; los cuales corresponden solo a las áreas con asentamientos humanos.

La preparación de la capa de puntos del DENU 2020 que contiene la localización puntual de las 5'326,696 Unidades Económicas del país, consideró:

1. Descarga masiva de las 32 capas geográficas a nivel nacional
2. Unión de las 32 capas geográficas, para cubrir el territorio nacional
3. Reproyección de la capa unida (de Cónica Conforme Lambert a EPSG3857)
4. Creación de tres columnas, a) concatenación de la clave censal a nivel de manzana; b) columna con el valor numérico del Personal Ocupado (se tomó el valor absoluto del límite superior de la clase de referencia); y c) columna con el sector de actividad económica, correspondiente a los dos primeros dígitos de la clave de actividad de cada Unidad Económica (punto).

Para el procesamiento de los polígonos de manzanas se realizó lo siguiente:

1. Relación entre la tabla geográfica, con la tabla de datos estadísticos
2. Selección de las variables de la población total y del total de viviendas
3. Clasificación de las localidades urbanas y rurales, a partir de la población (se aplica el criterio de manzanas pertenecientes a localidades menores a 2,500 habitantes, como rurales y las mayores como urbanas)
4. Se obtuvieron 1'596,975 puntos correspondientes al total de manzanas urbanas y rurales del país.

La capa geográfica de puntos Servicios con Información complementaria de tipo Puntual, referida en este trabajo como SIP, se procesó de la siguiente manera:

1. Descarga de las 32 capas estatales
2. Unión a nivel nacional
3. Reproyección geográfica de Cónica Conforme Lambert a EPSG3857
4. Depuración de los elementos clasificados como “templos” con el fin de considerar solamente equipamientos e infraestructuras públicas
5. Se realizó una revisión topológica de los 441,794 puntos
- b) El segundo momento inicia posterior a la obtención de los primeros resultados de la percolación de los puntos del WFS 2019. Se detectaron algunos vacíos de datos, por ello se volvió a repetir el procesamiento hasta obtener una cobertura total nacional (ver detalles en el apartado V). Para tener más posibilidades el equipo optó por descargar una capa de la Red Nacional de Caminos del año 2020, del INEGI (INEGI 2020); la cual se procesó como sigue:
 1. La red nacional de caminos, se reproyectó geográficamente
 2. Se obtuvieron solamente los nodos de la red (15'872,494 puntos).

Además de la capa anterior, se obtuvo la red de calles de Open Street Maps del año 2020 (OSM 2020), representada por 3'440,394 puntos a nivel nacional. Se procesó de manera similar que la red de caminos del INEGI.

- b) En el tercer momento, se realizó la generación del uso estimado del suelo en manzanas urbanas de todo el país; aplicando un método de reclasificación de la unión de las capas de manzanas, con el DENUÉ y el SIP (para ver detalles metodológicos ver la referencia del apartado a) de esta sección).

Al final se trabajó con cuatro capas geográficas preprocesadas y preparadas para el desarrollo del método: 1. Puntos urbanos de la **WSF**; 2. Capa de puntos de las Unidades Económicas del **DENUÉ** (capa que representa la inversión del capital privado); 3. Capa **SIP** de los equipamientos e infraestructuras promovidas por el sector público; y 4. Nodos de **OSM**, que representan la dimensión de la infraestructura vial de los asentamientos humanos.

Para el proceso de validación se procesaron y se estandarizaron las siguientes capas:

1. 1,809 polígonos de municipios que conforman las 401 ciudades del Sistema Urbano Nacional de CONAPO 2018
2. Se clasificaron las 50,308 localidades del país en urbanas y rurales con base en el criterio de población mencionado al inicio de este apartado.

V. Aplicación del método y resultados

Una vez que se contó con las bases de datos debidamente homologadas y estandarizadas. Se inicia el proceso de conceptualización del modelo a operar. Se parte de la idea desarrollar un método sencillo que permita comparar los resultados obtenidos del proceso de percolación de los puntos vectoriales de la capa (ráster) WSF (recordar que fue resampleada a 30x30m), en un primer momento se pensó hacer la comparación con el mismo raster que clasifica lo urbano y lo no urbano. El método debería permitir la obtención de una métrica libre de variables funcionales o rasgos distintivos de áreas urbanas; es decir, solo trabajar con los centroides de los píxeles de la capa antes mencionada.

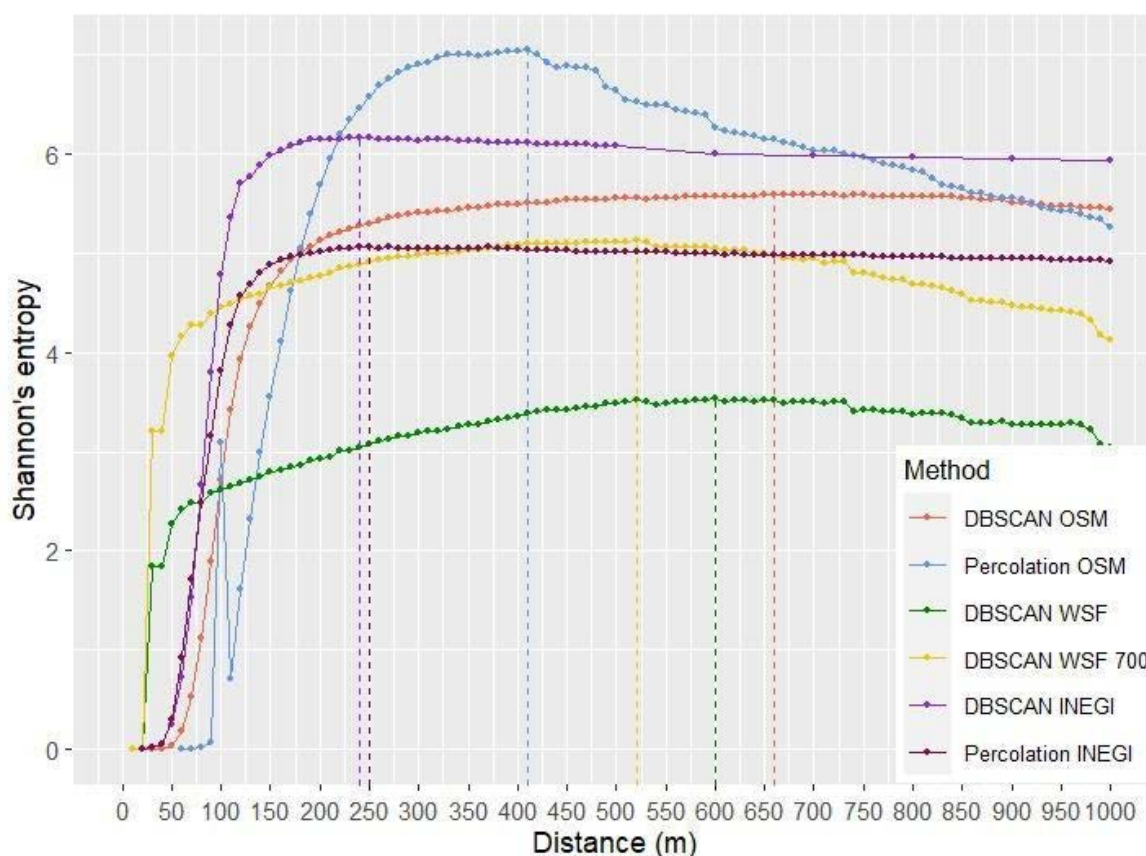
Se discutió sobre los procesos para definir criterios de comparabilidad entre los resultados obtenidos de la percolación, con algunas otras bases de datos. Surgió la posibilidad de comparación con fuentes voluntarias como Open Street Maps (OSM) o con la red carretera del INEGI.

Para ello, se habló sobre la escala a utilizar como un criterio espacial que cubriera a los centroides de manzanas; no obstante si se ampliaba la distancia no resolvía el problema para generar una malla que permitiera la definición de una envolvente que definiera lo urbano.

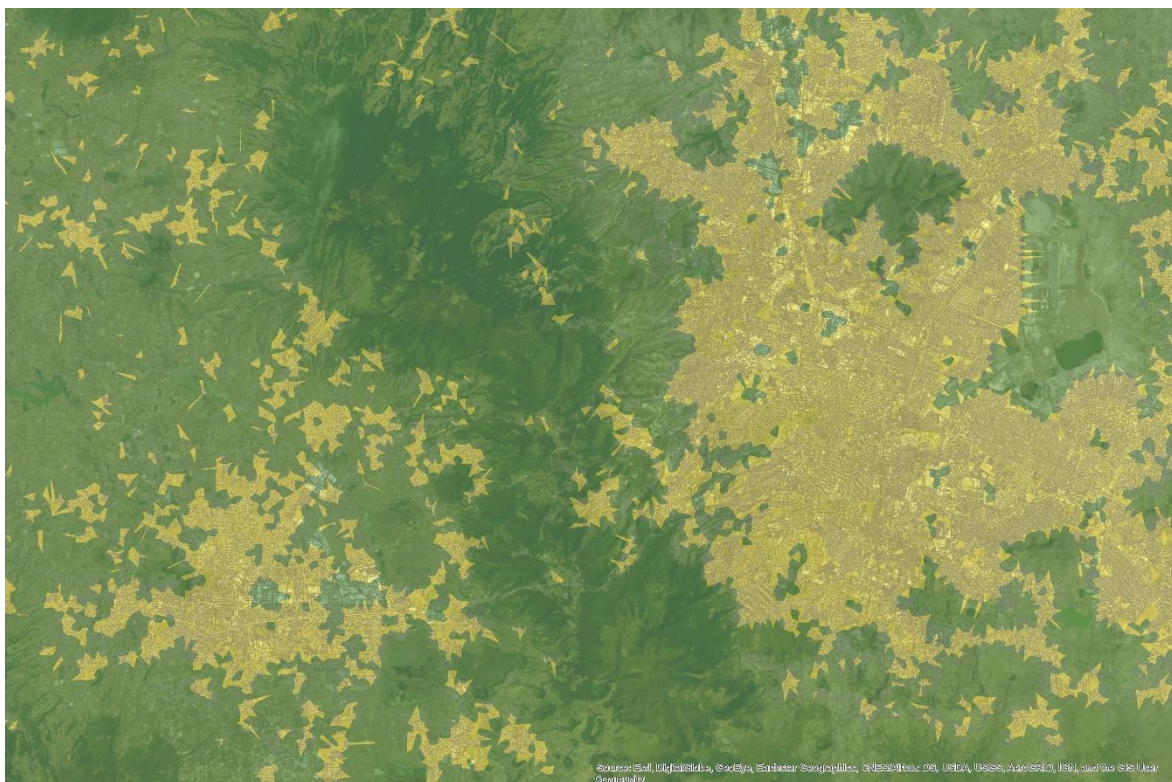
Posteriormente se consideró como una posibilidad de utilizar la asignación de clave de pertenencia a los puntos percolados y los que resultaron de la aplicación del DBSCAN, para fines comparativos con las ciudades consideradas en el SUN2018. Se planteó la posibilidad de que habría que considerar un paso posterior al desarrollo de los procesos mencionados, este consiste en realizar pruebas que permitieran medir la congruencia entre los resultados con la realidad, es decir con lo que se considera urbano (aquí había dudas con respecto a que información se utilizaría como referencia). Se profundizó sobre la conformación de clústers que se obtienen de las primeras corridas con DBSCAN, cuyas distancias iniciales se determinaron a 250, distancia obtenida del proceso de percolación (se utilizó el lenguaje de programación de R). Se analizaron los resultados para algunas ciudades como la de Toluca, en la que se encontraron que la distancia máxima de clusterización era menor a los 250 metros.

Vale la pena mencionar, que la base de puntos WFS se decidió desecharla del método debido a que si se continuaba utilizando, llegaríamos a confirmar solamente que la base permitía definir lo urbano y lo no urbano; además estaríamos sesgando el estudio. La opción fue utilizar la red de nodos de calles del INEGI y la de nodos de Open Street Maps, de éstas se hablará más adelante.

La discusión continuó hasta llegar a la decisión de considerar las bases de datos del SIP, la cual se procesó con el DBSCAN (recordemos que este permite generar clústers o cúmulos de puntos) para realizar corridas a distintas distancias: 700, 1,000 y 2,000 metros (recordemos que DBSCAN considera como necesaria la definición de la distancia máxima y número de puntos mínimos). Se obtuvieron los siguientes resultados:



Vale la pena mencionar que de la capa SIP se omitieron los puntos que hacen referencia a los templos, debido a que se acordó que la presencia de estos elementos, se localizan en prácticamente todas las localidades sin distinción de urbanas o rurales. Para este ejercicio se consideró operar el concepto de ciudades naturales, haciendo polígonos de *Thiessen* para completar y comprobar los resultados obtenidos con esa base y con ese método.



Una vez revisados los resultados del proceso anterior, se tomó la decisión de unificar en una sola base de datos los puntos del SIP (sin templos) con los del DENUE, para volver a correr la percolación. Los criterios para el DBSCAN inicialmente se utilizaron 750m y con 5 puntos mínimos. Se realizaron pruebas a distancias 600 y 750m en el DBSCAN, equivalentes a pruebas de sensibilidad, aunque en este momento la preocupación era encontrar la E-máx y el filtrado de puntos de ruido. Esto llevó al equipo a analizar las gráficas de para ver si eran consistentes las distancias máximas derivadas del proceso de percolación de las bases de puntos.

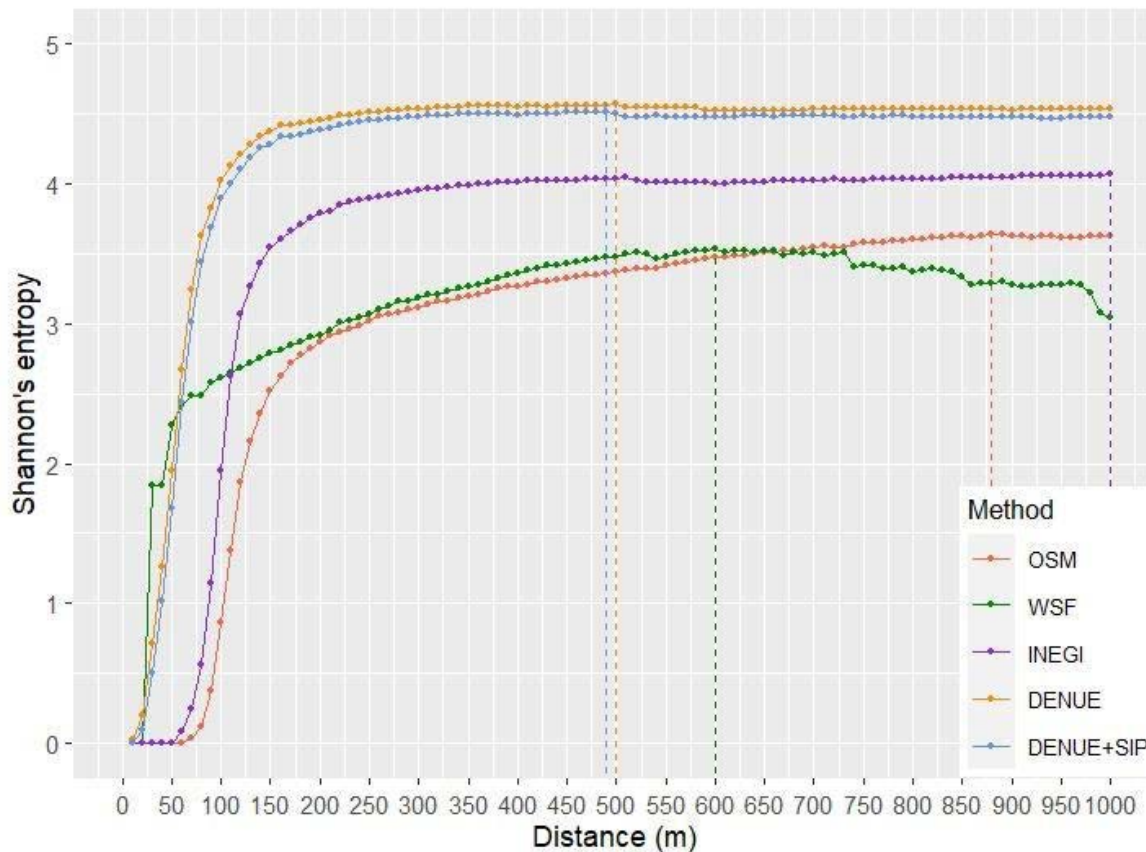
Se abrió la discusión sobre el proceso de filtrado de los datos identificados como ruido, es decir, aquellos puntos que no formaron parte de algún cúmulo; pero se presentó nuevamente otra interrogante: ¿Cómo establecer un criterio de selección del ruido? Esto llevó al grupo a tomar la decisión de generar estadísticas descriptivas de cuatro conjuntos de datos geográficos: Nodos de calles (INEGI), DENUE, SIP y OSM.

ESTADÍSTICOS DESCRIPTIVOS	RED*		EQUIPAMIENTOS*		UNIÓN	WSF
	INEGI*	SM	ENUE	DENUE+SIP	DENUE+SIP+OSM	POP_2020<2,499
SUM	5,060,686	89,091	61,049	168,857	1,014,152	9,359,030
MAX	2 ,659	45	90	575	1,002	9,736
MIN	2 1			1	1	0
MEDIA	12	4,690	8	6	24	202
DESV EST	1 23	8	4	23	40	262

ESTADÍSTICOS DESCRIPTIVOS	RED*		EQUIPAMIENTOS*		UNIÓN	WSF
	INEGI*	SM	ENUE	DENUE+SIP	DENUE+SIP+OSM	POP_2020<2,499
AREA MAX (m2)	396,586	66,947		212,320		876,240
√AREA MAX (m)	630	17		461		936
MAX = Número de puntos máximos promedio en localidades < 2,500 habitantes						
* Puntos actualizados con la Red Vial Nacional 2020						

El proceso utilizado para obtener los estadísticos consistió en asignar las claves de las localidades consideradas por el INEGI como rurales a nivel nacional, a los puntos de las cuatro capas arriba referidas. Posteriormente, obtener el número de puntos promedio que pertenecían a los polígonos rurales (localidades menores a 2,500 habitantes). Con estas métricas se obtuvieron dos criterios fundamentales para esta investigación: 1. El número de puntos promedio para las localidades rurales, que sirvió para determinar el criterio de puntos mínimos; 2. Las distancias promediadas entre todos los puntos pertenecientes a este tipo de localidad; con ello se obtuvo el criterio de la distancia máxima.

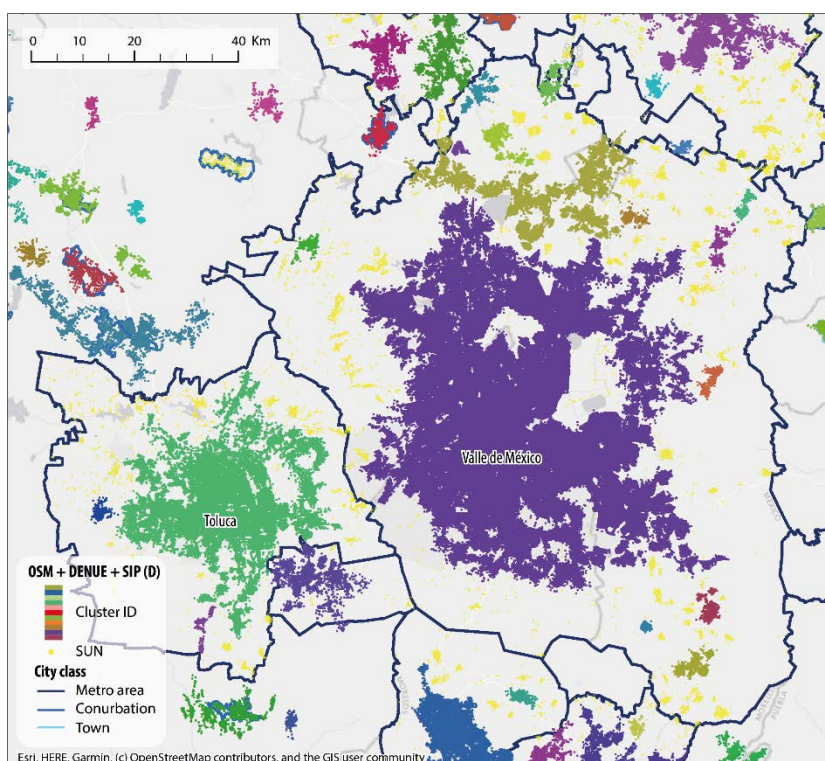
Los parámetros se utilizaron para desarrollar los procesos de la percolación y aplicar el método DBSCAN, hasta encontrar la entropía máxima (E-máx), que diera lugar a definir los umbrales de los cúmulos resultantes. Las percolaciones arrojaron los siguientes resultados:



Una vez que se analizaron los gráficos que mostraban la E-máx, y dado que la curva de la geobase de Nodos de calles del INEGI nunca bajó⁴⁰, es decir, no se encontró a diferentes distancias del percolado la E-máx se decidió desecharla del modelo. Posteriormente, se continuaron haciendo pruebas de sensibilidad en la percolación de puntos a diferentes distancias, considerando el valor máximo de puntos contenidos en localidades menores a 2,500 habitantes, para encontrar: 1. La distancia máxima y 2. La E-máx.

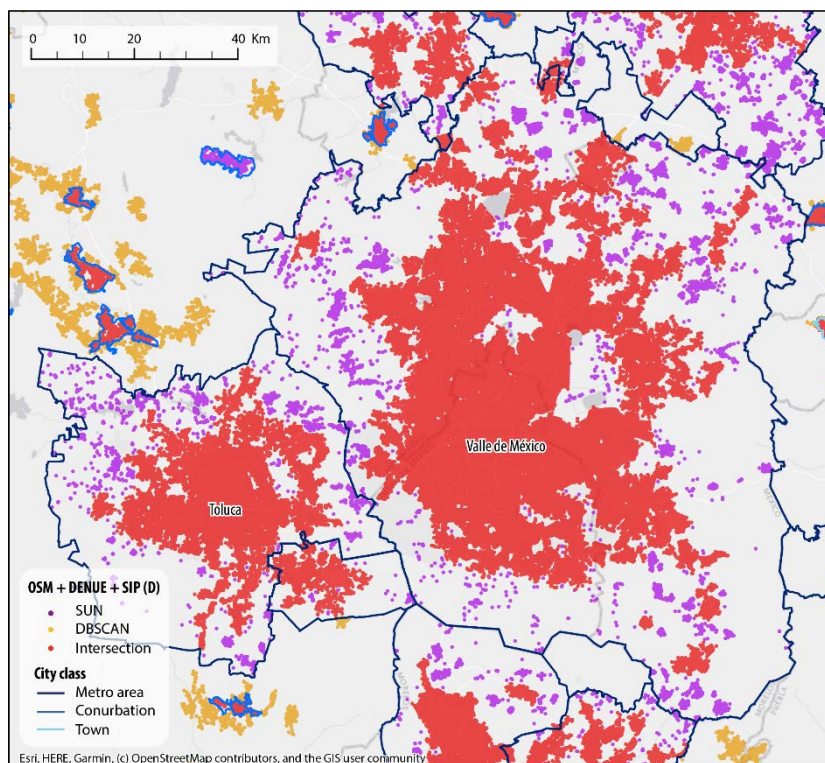
Obtenidos estos parámetros se inició la discusión sobre la pertinencia de realizar no solo la comparación de clústers con el SUN; sino también con la Ley de Zipf, con el objetivo de comprobar si la conformación de cúmulos y principalmente sus tamaños se relacionaban con esta ley. Como hallazgo relevante, se pudo considerar que en otros trabajos la E-máx no se relaciona directamente con el tamaño de las ciudades a partir de la aplicación de la Ley de Zipf. En este caso encontramos que existe una relación entre la E-máx para las ciudades que conformaron cúmulos y de acuerdo con su tamaño, existe una relación con la Ley de Zipf (resultados que más adelante se presentarán).

Recordar que se comprendió que la E-máx es considerada como la mayor homogeneidad con respecto a la distribución de las frecuencias de las clases (en este caso los clústers) y mayor heterogeneidad entre las cualidades de los clústers (en este caso en la identidad de clústers). Por ejemplo: si tuviéramos en una ciudad un solo tipo de uso del suelo la entropía sería cero, lo que supondría que sería homogénea la distribución.



⁴⁰ Se realizaron varias corridas percolando los puntos para encontrar la E-máx y se llegó a una distancia de mil metros y en la curva no se encontró la E-máx; motivo por el cual se decidió desechar la geobase de puntos del nodos de calles del INEGI.

Los resultados obtenidos de los análisis de sensibilidad para el conjunto general de las bases de puntos, se seleccionaron tres conjuntos con los cuales se trabajó en la representación cartográfica de dos cúmulos: urbanos y no urbanos (ya con el ruido filtrado dentro del límite del SUN).



Una de las conclusiones a la que nos llevó la distribución de los puntos reflejados en el mapa anterior es: **1.** Hay correspondencia entre los puntos que el modelo considera como urbano y no urbano, **2.** Dentro de las ciudades del SUN (recordar que son grupos de municipios) existen grupos de puntos que el modelo no considera como urbanos, y **3.** En algunos centros de población aislados del SUN, el modelo agrega más puntos que clasificó como urbanos; es decir, el modelo determina que el centro de población con base en los criterios del modelo debe ser considerados como urbanos. No obstante, en algunos casos los resultados del DBSCAN, no logró clasificar como urbanos algunos centros de población debido a que no alcanzaron el número mínimo de puntos para conformar un cúmulo. Cabe mencionar que en este caso se eliminaron los puntos que no formaron parte de un cúmulo y que no estuvieran dentro de los límites del SUN. Para este caso se utilizó la distancia en la que se encontró la E-máx (770m), obtenida de la percolación de la base unida de DENUE-SIP.

Una vez que se tuvieron definidos los tres *dataset* de puntos, con sus distancias en las que se ubica la E-máx para cada una, se abrió la posibilidad de realizar un índice comparativo que permitiera realizar la **similaridad global** (lo que es urbano y lo que no es urbano) de los

clústers obtenidos, con la conformación de las ciudades del SUN⁴¹. El índice seleccionado fue Jaccard⁴².

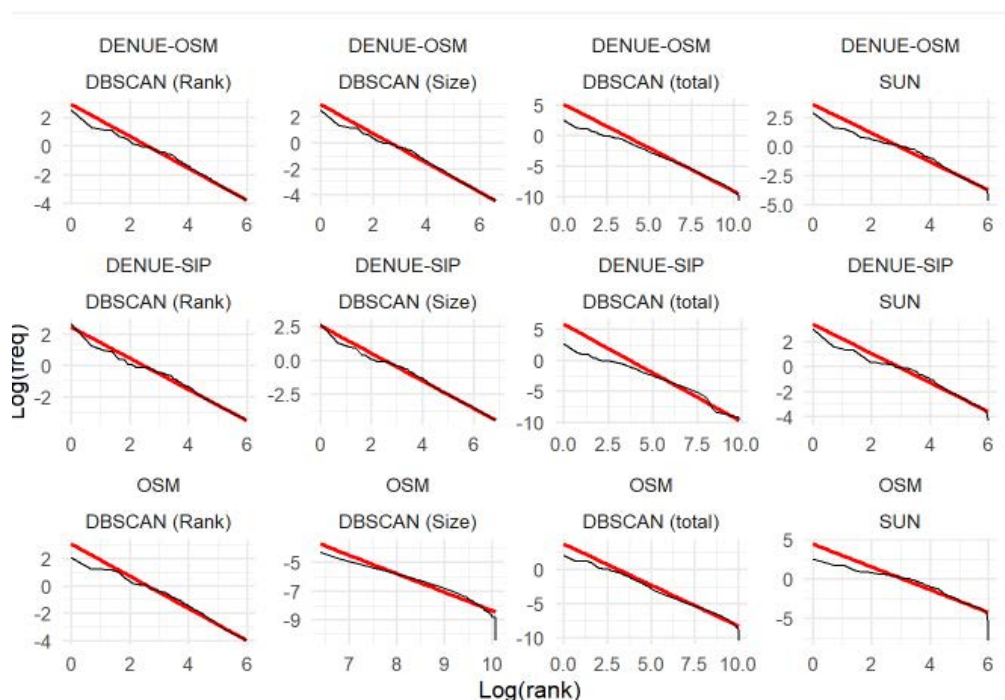
Comparación entre el SUN y resultados de DBSCAN

En un reporte previo nos ocupamos de la comparación global mediante el índice de Jaccard para identificar la similitud entre lo que se considera urbano de acuerdo con el SUN y lo que resulta como la distancia en la que se ubica la máxima entropía después de percolar los datos con DBSCAN y filtrar según el umbral de las localidades rurales. El reporte mencionado previamente incluye la comparación local, de los elementos que forman parte del SUN, cuantificamos la intersección con los puntos percolados dentro del SUN.

Ahora nos ocupamos de las comparaciones locales

Escala local por cada ciudad del SUN. Esto se logró de dos formas: a) por cada SUN, qué proporción de la i-ésima ciudad se interseca con los resultados de DBSCAN con el umbral determinado por el máximo de puntos en localidades rurales. b) por cada SUN, qué pureza hay en los cúmulos de DBSCAN con mismo umbral, esta es una forma de explorar no sólo la coincidencia, sino la homogeneidad de los puntos que se clasifican como cúmulos por DBSCAN.

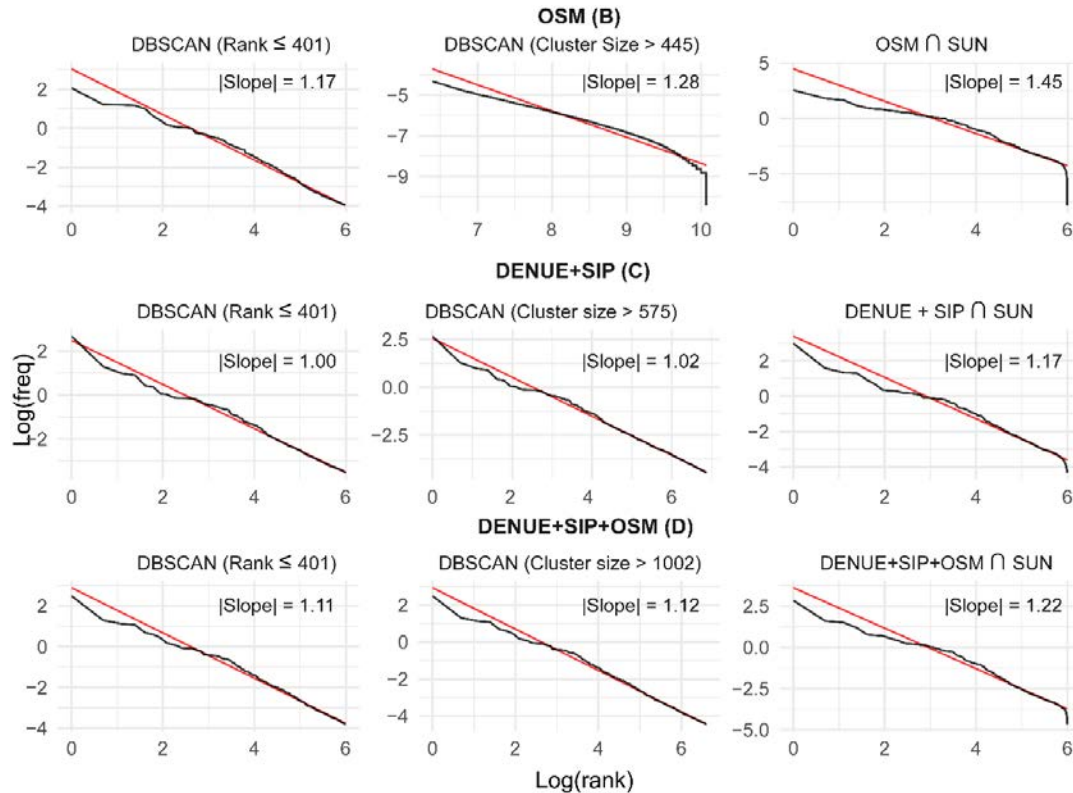
De acuerdo con el rango de las ciudades del SUN (cuántos nodos/píxeles/unidades económicas hay en las n ciudades del SUN) vs rango de cúmulos filtrados por DBSCAN (de los mismos n cúmulos más grandes). El primero gráfico resultante es el siguiente:



⁴¹ Se utilizó el conjunto de datos WSF puntos, asignándoles el identificador de la ciudad a la que pertenece de acuerdo con la clasificación del SUN; base que sirvió para hacer la comparación y validez de los resultados de nuestro método.

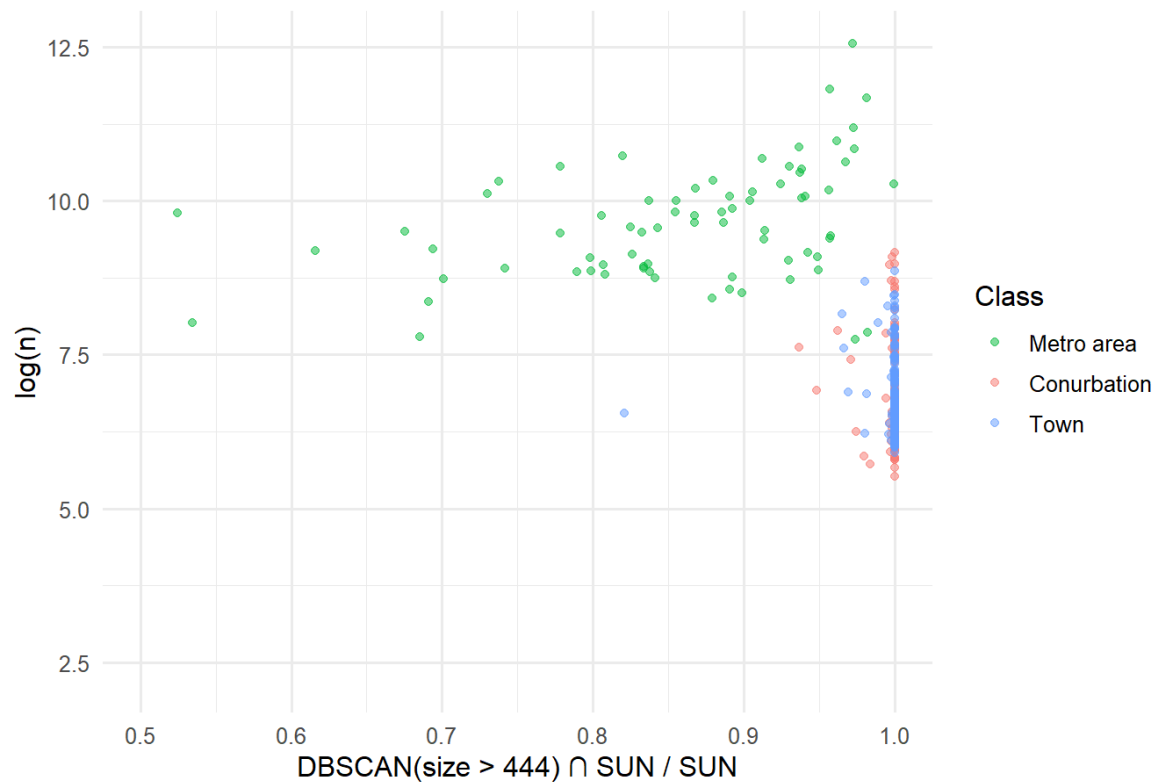
⁴² Se utilizó para 'Globalmente' hacer la comparación entre cada una de las bases de datos filtradas contra todos los nodos que componen el SUN (todos los nodos dentro del SUN son urbanos; los que quedan fuera del SUN se consideran rurales).

En el gráfico se puede presentar en un panel con todos los diagramas. Evaluar si vale la pena utilizar los cúmulos con sólo el criterio de tamaño, son muy parecidos a los cúmulos sin filtrar.



Se puede resaltar que la definición del SUN incluye criterios múltiples que en ningún resultado se ajusta a la ley de Zipf. El DENUE-SIP con 401 cúmulos de DBSCAN es el conjunto que más se acerca a la ley de Zipf. La definición del SUN incluye criterios múltiples que en ningún resultado se ajusta a la ley de Zipf, al cien por ciento, no obstante el ajuste del DENUE-SIP es el que menor desviación a la línea teórica tiene con relación al tamaño de las ciudades. En los gráficos de tamaño, se incluyeron la base completa de cúmulos para todo el país. El DENUE-SIP se utilizaron los 401 cúmulos de DBSCAN es el conjunto que más se acerca a la ley de Zipf.

Cabe señalar que se hicieron varias gráficas que mostraron el comportamiento de las bases de puntos para encontrar la **similaridad local** con el SUN. Lo local se entiende como las comparaciones para cada cúmulo de puntos por cada ciudad del SUN. Las siguientes gráficas reflejan las comparaciones utilizando los conjuntos de puntos, que fueron analizados para decidir qué *dataset* era el de mayor explicación del modelo. Por esta razón, se realizaron los gráficos para las bases de puntos por separado y con la agregada por OSM+DENUE+SIP.

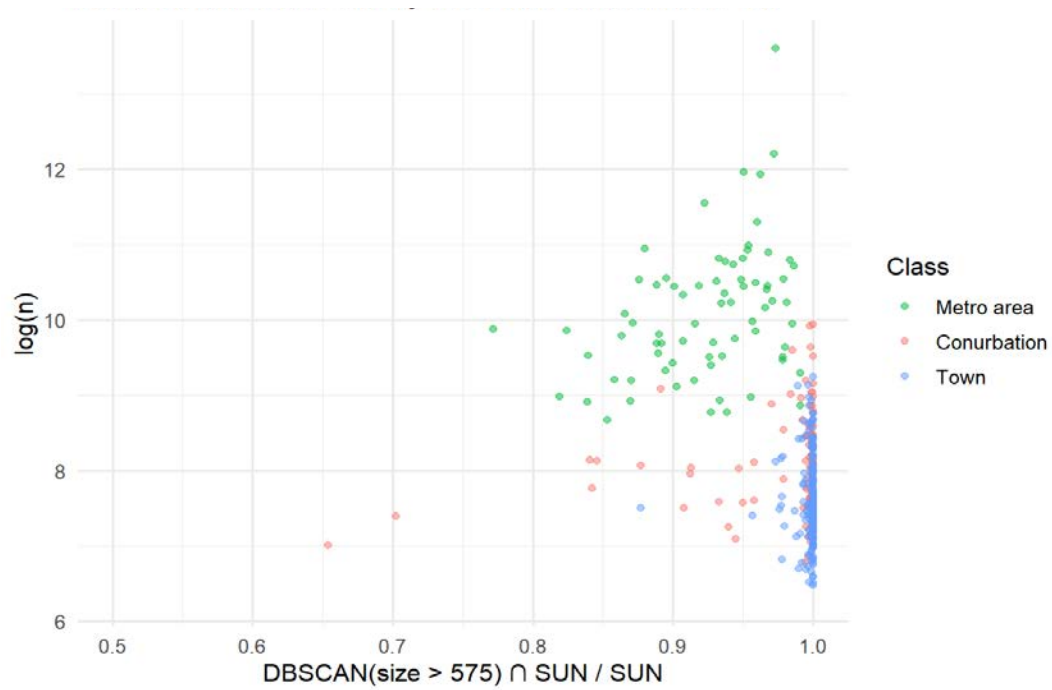


Coincidencia entre SUN y DBSCAN con OSM

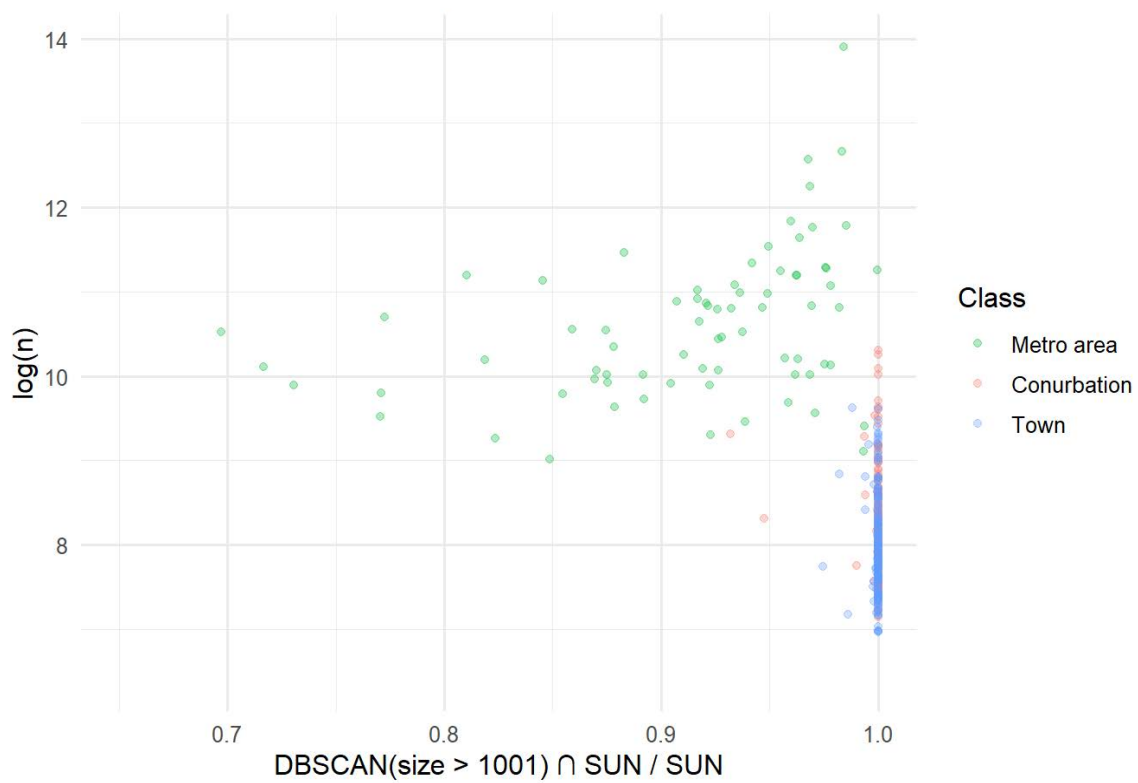
De lo más general a lo particular, el promedio de coincidencia o cociente de la intersección dividida por el total es 0.8523234, que parece bastante bueno. En términos de error, se debe resaltar que 18 conurbaciones y 30 poblados del SUN no tienen correspondencia con los cúmulos de OSM mayores a 445 nodos.

Comparación ciudad por ciudad

Una forma simple de comparar es ver qué porción del conjunto del SUN representa la intersección con DBSCAN para cada ciudad. Esto no incorpora el rango de los cúmulos. Esto supondría la utilidad del método como una forma de clasificación más general sin tomar en cuenta la jerarquía.

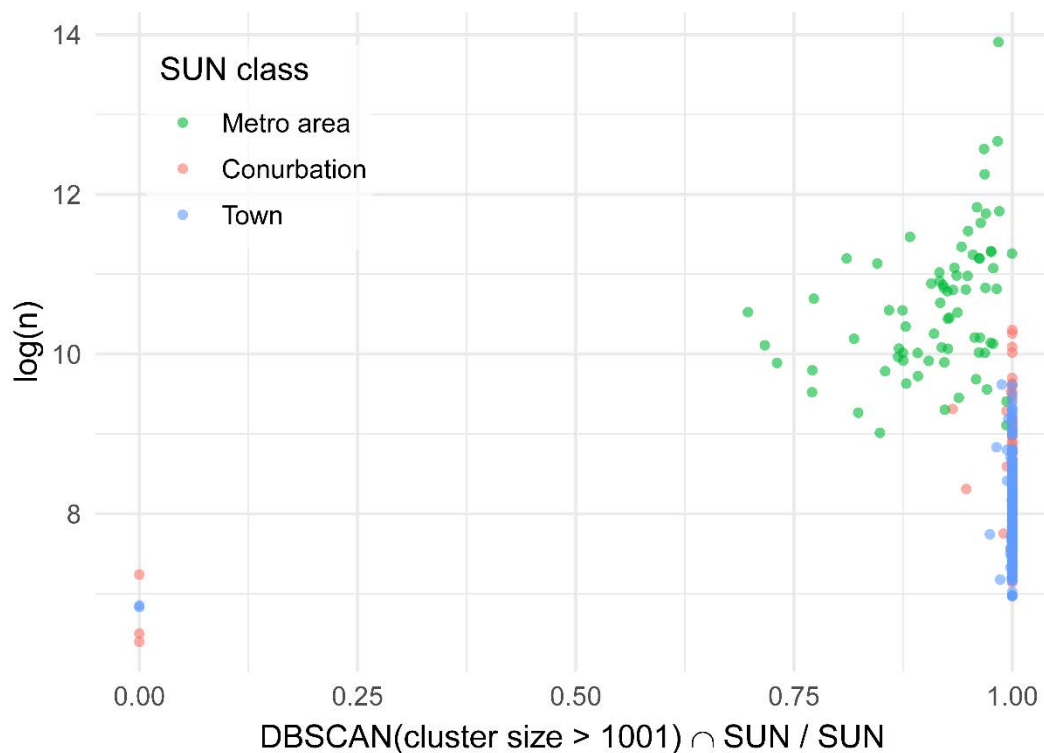


Coincidencia entre SUN y DBSCAN con DENUE-SIP



Coincidencia entre SUN y DBSCAN con DENUE-OSM

La siguiente gráfica resultó ser la ganadora en términos de cumplir con el supuesto de comparabilidad.



Coincidencia entre SUN y DBSCAN con OSM+DENUE+SIP

Comparación con el índice de GINI (pureza)

Se utilizó el índice de Gini para determinar las coincidencias interpretando que los gráficos muestran valores de 0 a 1, cuando tienden a 1 la similaridad es mayor entre los puntos clusterizados con los puntos del SUN. A la derecha se muestran los puntos de las ciudades que coinciden en más del 75% con los identificados con el SUN, y a la izquierda se muestran 5 ciudades que coinciden en menor proporción con el SUN.

El índice de Gini se utiliza como criterio de pureza para algunos algoritmos de clasificación basados en árboles, en nuestro caso en lugar de utilizar la curva de Lorenz completa, omitiremos únicamente la proporción que aportan los cúmulos fuera de DBSCAN (que sólo se encuentran en el SUN). Sin embargo, en principio sí tomaremos en cuenta el total de la curva. Nuestro índice será la proporción de nodos y la proporción de nodos en los cúmulos.

De esta forma únicamente los casos en los que un sólo cúmulo de DBSCAN sea exactamente una ciudad del SUN el índice de Gini tendrá un valor de 1. En primer lugar necesitamos todos los cúmulos que se encuentren en la intersección con el SUN, utilizaremos un nuevo ID_CLUST como identificador de los cúmulos de DBSCAN que cumplen con la condición de que tienen un tamaño según el umbral de lo urbano para cada conjunto de datos.

Vamos a construir una función para graficar en un diagrama de dispersión el tamaño contra el índice de Gini y la jerarquía como color.



La primera opción, ver la pureza o concentración por clave, nos indica que algunas conurbaciones y poblados son los que mejor se clasifican (y también peor, 47 no tienen correspondencia). Las zonas metropolitanas tienen resultados más heterogéneos.

También podemos comparar qué pasa si tomamos únicamente los cúmulos del mismo rango que las ciudades del SUN.

Índice de Jaccard

Vamos a codificar como 1 todos los nodos que pertenecen a alguna ciudad del SUN y como 2 a todos los nodos que sean de un cúmulo > 445 nodos, de este modo la variable 'jac' es la suma de las dos variables previas y los valores 1 serán sólo SUN, 2 sólo DBSCAN y 3 es la intersección. Estos son los insumos para calcular el índice de Jaccard.

Obtención de los rangos

Necesitamos los cúmulos organizados de acuerdo con la jerarquía, a diferencia de los estudios clásicos que utilizan población, nosotros utilizamos el tamaño de las ciudades medido a través de los nodos, unidades económicas o elementos urbanos incluidos en cada conjunto de datos.

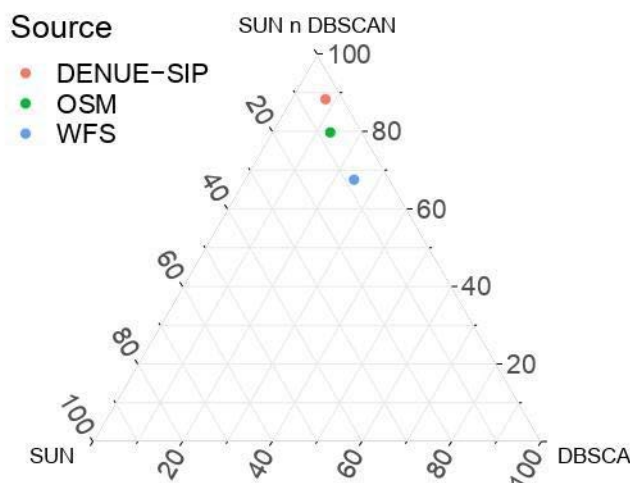


Diagrama ternario de los conjuntos

El índice de Jaccard en el caso de dos clases es fácil de representar por medio de los porcentajes que cada componente acumula, en el diagrama de abajo, la intersección (en violeta) expresada en porcentaje es precisamente el índice de Jaccard, es 0.797. Es una clasificación mejor que conservando únicamente los cúmulos de más de 50 nodos. Los nodos exclusivamente en el SUN son 0.072 y en DBSCAN 0.131.

Resultado de intersecciones

CVE_SUN	SUN Name	Rank SUN	Rank DBSCAN	Intersection
M09.01	Valle de México	1	1	0.932
M14.01	Guadalajara	2	2	0.97
M19.01	Monterrey	3	3	0.907
M21.01	Puebla-Tlaxcala	4	4	0.943
M15.02	Toluca	5	7	0.944
M11.03	León	6	6	0.983
M22.01	Querétaro	7	5	0.912
M02.03	Tijuana	8	8	0.891
M31.01	Mérida	9	9	0.904
M05.01	La Laguna	1	13	0.812
M16.02	Morelia	1	11	0.942
M24.02	San Luis Potosí	1	12	0.976

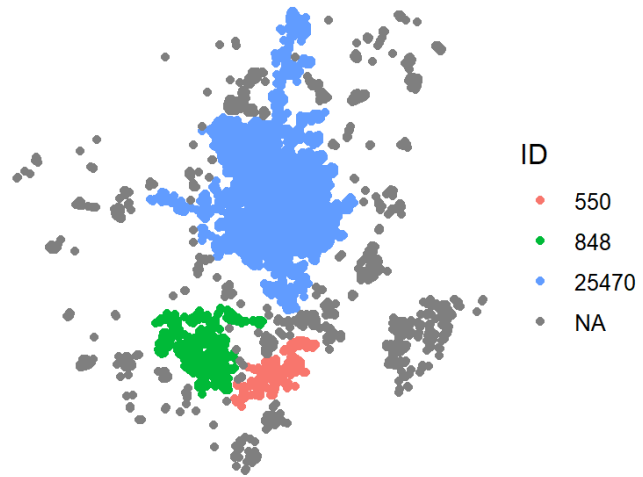
M08.04	Juárez	1 3	14	0.976
M20.01	Oaxaca	1 4	10	1
M01.01	Aguascalientes	1 5	15	0.955

La representación de los cúmulos de ciudad por ciudad, se muestran a continuación:

Jaccard

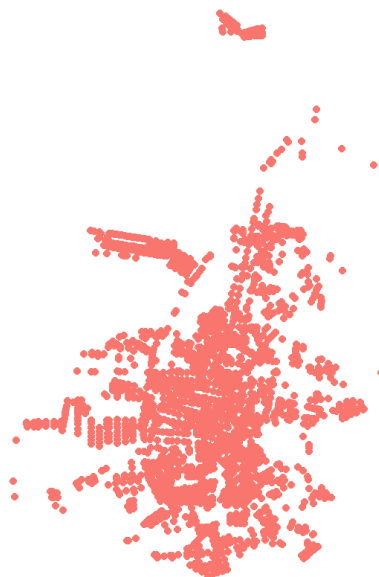


CLUSTER_ID (size > 444)

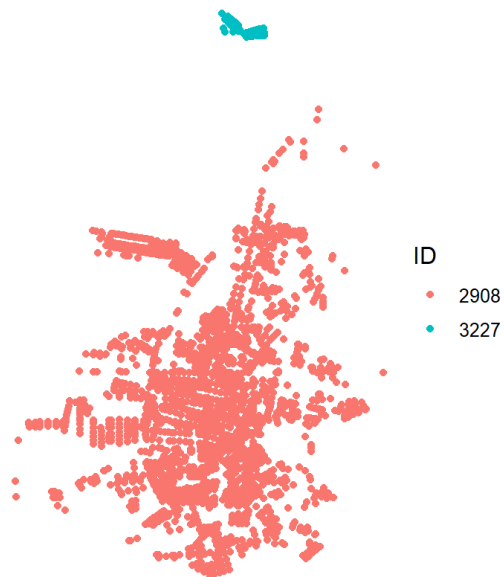


ZM de Aguascalientes OSM data

Jaccard



CLUSTER_ID (size > 444)



Villahermosa OSM data

Jaccard

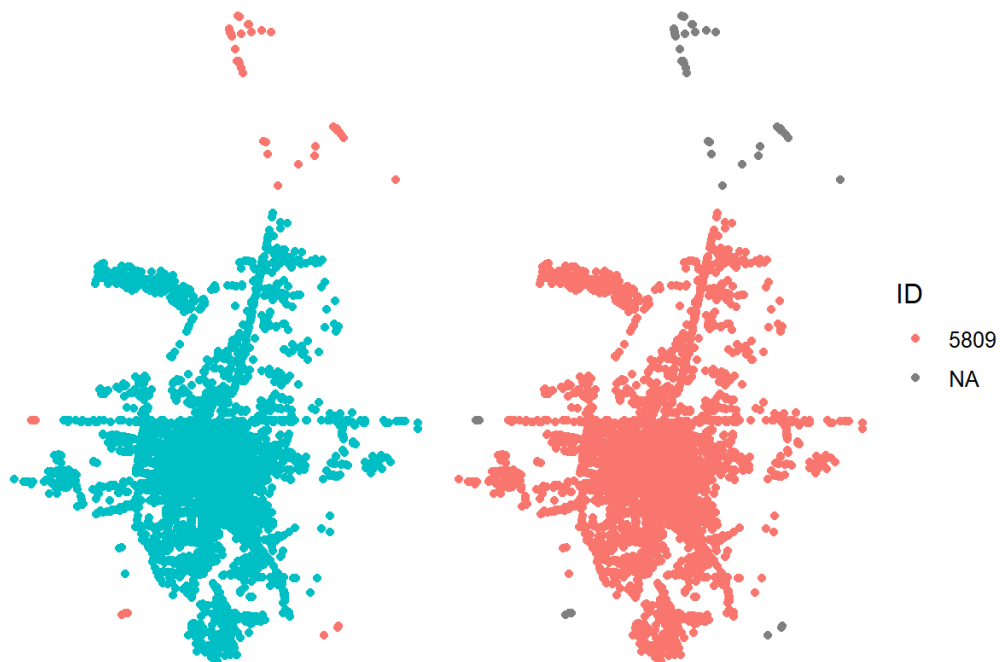
CLUSTER_ID (size > 575)



ZM de Aguascalientes DENU-SIP data

Jaccard

CLUSTER_ID (size > 575)



Villahermosa DENU-SIP data

Jaccard

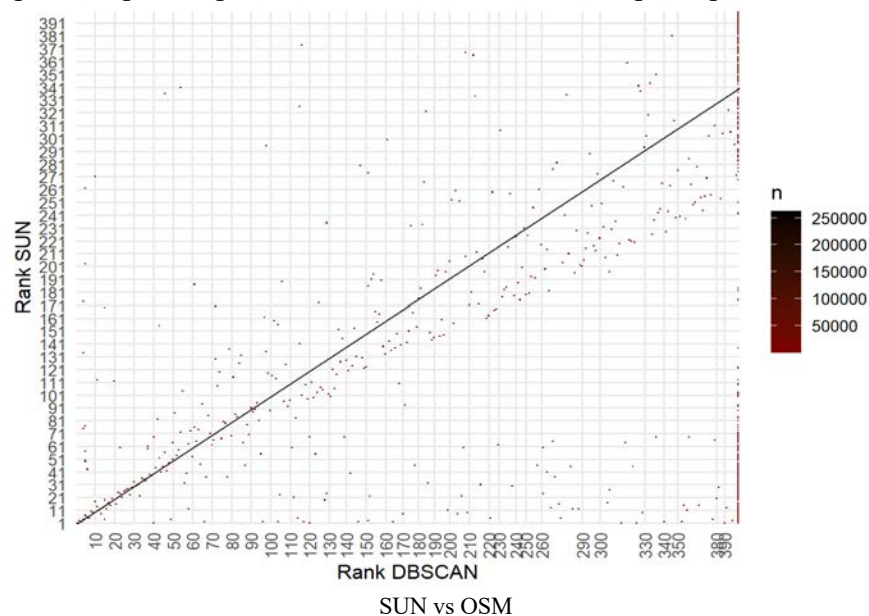
CLUSTER_ID (size > 1002)



ZM de Aguascalientes DENUE-OSM data

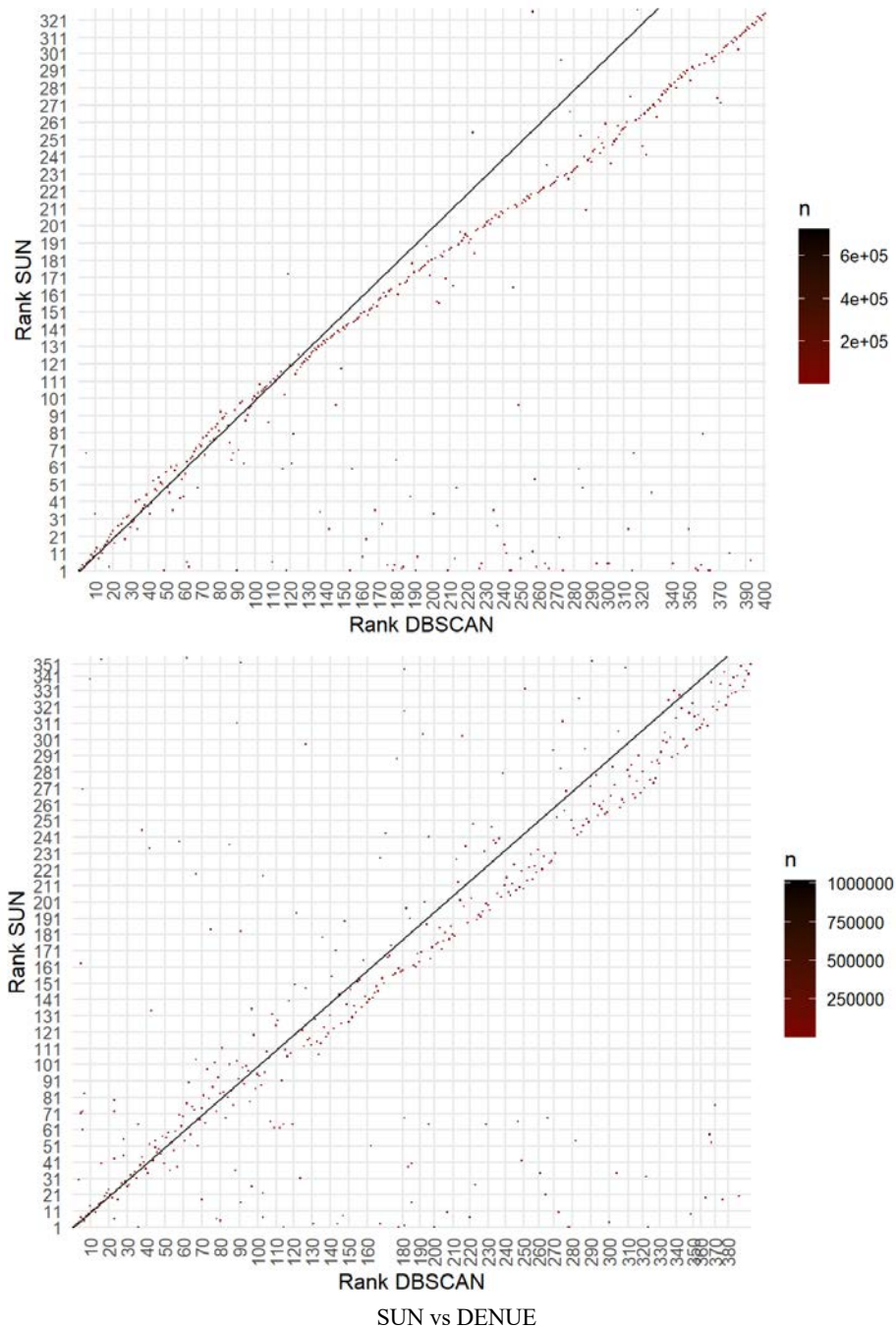
Rangos y ley de Zipf de cada conjunto

Con base en el rango, también podemos probar si la relación rango-frecuencia obedece a la ley de Zipf, que no es otra cosa que la pendiente entre el modelo de regresión lineal $\log(\text{Frecuencia}) \sim \log(\text{Rango})$, por lo general el ajuste no es muy bueno en los extremos. Modelos de regresión: La pendiente es el parámetro importante para identificar si se cumple la ley de Zipf. Las pendientes son relativamente cercanas a -1, para el SUN es 1.45, DBSCAN completo 1.21 y los 400 cúmulos más grandes es 1.16. La única que es casi perfecta es la del DENUE-SIP para los puntos percolados con los 401 cúmulos principales.



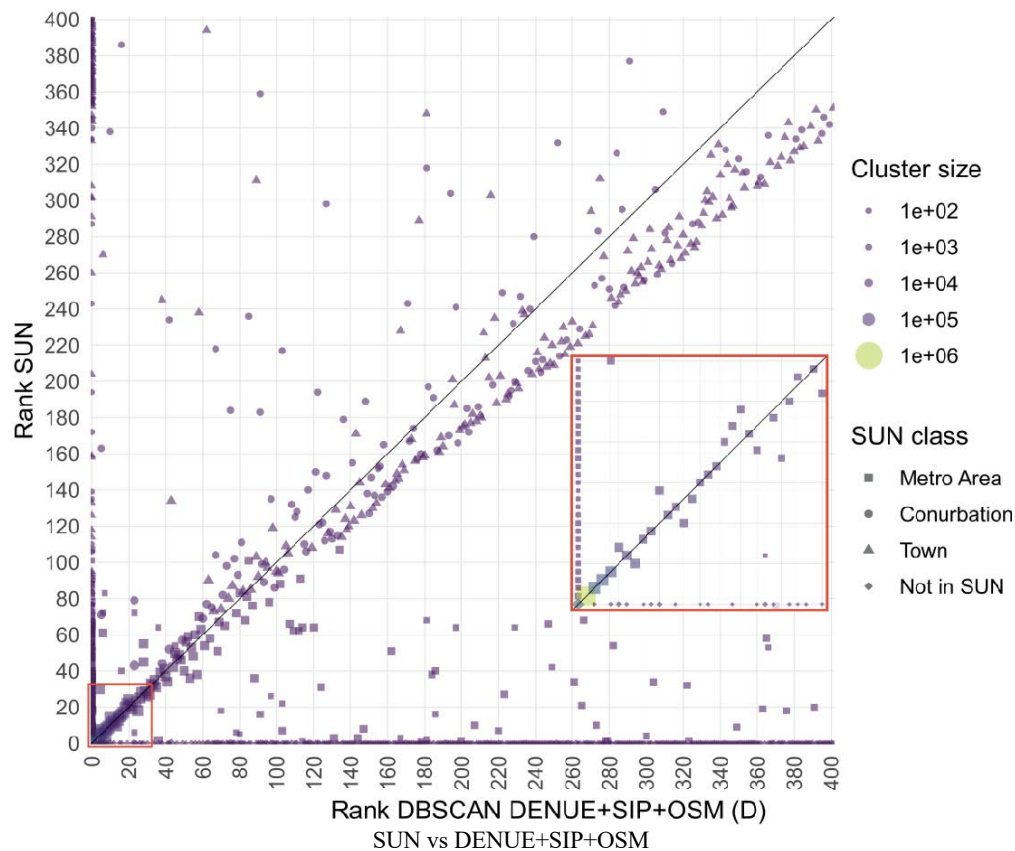
SUN vs OSM

Se ve claramente cómo la matriz se sale rápidamente de la diagonal, lo que indica una desviación del rango en ambos grupos.



A la izquierda del gráfico, se muestran los clústers de mayor tamaño de puntos que corresponden a las zonas metropolitanas del SUN y en sentido contrario, se localizan los agrupamientos de menor tamaño, es decir, muestra mayores coincidencias por tamaño en los

agrupamientos con mayor número de puntos, los cuales en su mayoría corresponden con ciudades más grandes.



Nuestro Rank obedece a la ley de Zipf mientras que el SUN no tanto, los puntos por debajo de la línea corresponden a ciudades mayores en comparación con la clasificación del SUN.

VI. Conclusiones

Se confirma el supuesto inicial de esta investigación y para la construcción conceptual del modelo. Se obtuvo un método que logró captar la mayor complejidad de relaciones intraurbanas y menor complejidad en las inter-urbanas. Esto permitió la definición –desde nuestro punto de vista– más precisa de los límites urbano-rurales.

Se obtiene un método que opera de manera sencilla y clara con variables que representan: la configuración social del espacio urbano (DENU); de la inversión pública en equipamientos e infraestructuras (SIP); e información voluntaria (OSM) que da cuenta de la configuración territorial de la infraestructura viaria, que refleja en gran medida la estructura territorial de las ciudades.

Los resultados obtenidos además de ser asequibles, con ellos se pueden diseñar políticas públicas, dirigidas a los clústeres específicos detectados. Abre la posibilidad de profundizar en los estudios urbanos estudiando las diferencias en la dinámica urbana entre los clústeres

detectados. Disminuye la subjetividad debido a que se encontraron umbrales para la delimitación de lo urbano, a partir de la relación entre la máxima entropía y la ley de Zipf.

Las 'ciudades' pronosticadas están alineadas con otras investigaciones que se han ocupado de encontrar ciudades naturales.

A pesar de esto, varias tareas quedan pendientes y justifican en un futuro profundizar en: 1. Ajustar el umbral de urbano-rural, así como analizar el número de puntos de interés, servicio o nodos mínimos de una red vial, que precise el concepto de ciudad; 2. Trabajar en la delimitación formal que transite de un cúmulo de puntos a geometrías mínimas de límite; 3. Trabajar con un método difuso, en lugar de uno fijo para afinar el modelo; y 4. Debería probarse en otros países para comprobar si los resultados se mantienen.

Bibliografía

- Aizenman, M., & Barsky, D. J. (1987). Sharpness of the phase transition in percolation models. *Communications in Mathematical Physics*, 108(3), 489–526. DOI: 10.1007/BF01212322
- Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P., & Batty, M. (2016). Cities and regions in Britain through hierarchical percolation. *Royal Society open science*, 3(4), 150691.
- Aurousseau, M. (1921). The distribution of population: A constructive problem. *Geographical Review*, 11(4), 563–592. DOI: 10.2307/208239
- Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17), 7301–7306. DOI: 10.1073/pnas.0610172104
- Broadbent, S. R., & Hammersley, J. M. (1957). Percolation processes: I. Crystals and mazes. *Proceedings of the Cambridge Philosophical Society*, 53(3), 629–641. DOI: 10.1017/S0305004100032680
- Cao, W., Dong, L., Wu, L., & Liu, Y. (2020). Quantifying urban areas with multi-source data based on percolation theory. *Remote Sensing*, 12(1), Article 137. DOI: 10.3390/rs12010137
- Castells, M. (1974). *La cuestión urbana*. Siglo XXI Editores.
- Franklin, R. (1942). *Density and porosity of coals and carbonaceous materials*. British Coal Utilization Research Association (BCURA).
- George, P. (1968). *Schème d'étude internationale des petites villes*. En *21st International Geographical Congress: Abstracts of Papers* (Abstract No. 584). U.G.I.
- Grimmett, G. R. (1999). *Percolation* (2.^a ed.). Springer. DOI: 10.1007/978-3-662-03981-6

- Grimmett, G. R., & Marstrand, J. M. (1990). The supercritical phase of percolation is well behaved. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 431(1883), 439–459. DOI: 10.1098/rspa.1990.0143
- Jiang, B., & Miao, Y. (2014). The evolution of natural cities from the perspective of location-based social networks. *The Professional Geographer*, 67(2), 295–306. DOI: 10.1080/00330124.2014.968886
- Kesten, H. (1982). *Percolation theory for mathematicians*. Birkhäuser. DOI: 10.1007/978-1-4899-2730-9
- Liu, L., Xia, B., Wu, H., Zhao, J., Peng, Z., & Yu, Y. (2019). Delineating natural city boundaries with points of interest based on service area and maximum entropy method. *Remote Sensing*, 11(13), Article 1544. DOI: 10.3390/rs11131544
- Makse, H. A., Andrade, J. S., Batty, M., Havlin, S., & Stanley, H. E. (1998). Modeling urban growth patterns with percolation theory. *Physical Review E*, 58(6), 7054–7062. DOI: 10.1103/PhysRevE.58.7054
- Marconcini, M., Metz-Marconcini, A., Esch, T., & Gorelick, N. (2021). Understanding current trends in global urbanisation – The World Settlement Footprint Suite. *GI_Forum*, 1, 33–38. DOI: 10.1553/giscience2021_01_s33
- Menshikov, M. V. (1986). Coincidence of critical points in percolation problems. *Soviet Mathematics - Doklady*, 33, 856–859.
- Mumford, L. (1961). *The city in history: Its origins, its transformations, and its prospects*. Harcourt, Brace & World.
- Ratzel, F. (1897). *Anthropogeographie*. J. Engelhorn.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x
- Simmel, G. (1968). La metrópolis y la vida mental. En *Sociología y psicología de la vida urbana* (pp. 247–263). Península.
- Smirnov, S. (2001). Critical percolation in the plane: Conformal invariance, Cardy's formula, crossing probabilities. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 333(3), 239–244. DOI: 10.1016/S0764-4442(01)01991-7
- United Nations Statistics Division. (2003). *Demographic yearbook 2001*. United Nations Publications.
- Weber, M. (1958). *The city* (D. Martindale & G. Neuwirth, Trans.). The Free Press.
- Wirth, L. (1938). Urbanism as a way of life. *American Journal of Sociology*, 44(1), 1–24. DOI: 10.1086/217913
- Zarza, A. (1996). *Una interpretación Fractal de la forma de la Ciudad*. Cuadernos de Investigación Urbanística. Instituto Juan de Herrera

Para obtener más información sobre DBSCAN:

Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, 60(1), 208-221.
<https://doi.org/10.1016/j.datak.2006.01.013>

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. En E. Simoudis, J. Han y U. Fayyad (Eds.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226-231). AAAI Press.

Para obtener más información sobre HDBSCAN:

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. En *Lecture Notes in Computer Science: Vol. 7819. Advances in Knowledge Discovery and Data Mining* (pp. 160-172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

Para obtener más información sobre OPTICS:

Agrawal, K. P., Garg, S., Sharma, S., & Patel, P. (2016). Development and validation of OPTICS based spatio-temporal clustering technique. *Information Sciences*, 369, 388-401. <https://doi.org/10.1016/j.ins.2016.06.048>

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. *ACM SIGMOD Record*, 28(2), 49-60. <https://doi.org/10.1145/304181.304187>

Stauffer, D., & Aharony, A. (2014). *Introduction to percolation theory* (2nd ed.). Taylor & Francis.

Rozenfeld, H. D., Rybski, D., Andrade, J. S., Batty, M., Stanley, H. E., & Makse, H. A. (2008). Laws of population growth. *Proceedings of the National Academy of Sciences*, 105(48), 18702-18707. <https://doi.org/10.1073/pnas.0807434105>

Fluschnik, T., Kriewald, S., García Cantú Ros, A., Zhou, B., Reusser, D. E., Kropp, J. P., & Rybski, D. (2016). The size distribution, scaling properties, and spatial organization of urban clusters: A global and regional percolation perspective. *ISPRS International Journal of Geo-Information*, 5(7), Article 110.
<https://doi.org/10.3390/ijgi5070110>

Jorge Alberto Montejano Escamilla<https://orcid.org/0000-0002-6914-1229>

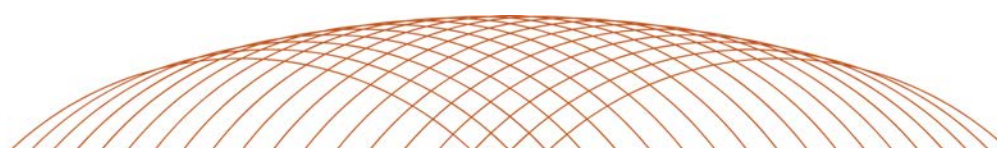
Dr. en Urbanismo por la Universitat Politècnica de Catalunya (Tesis Excelente Cum-Laude) y Arquitecto por la Universidad Iberoamericana. Actualmente es Profesor-Investigador y Coordinador de Proyectos en el Centro de Investigación en Ciencias de Información Geoespacial (CentroGeo). Es miembro del Sistema Nacional de Investigadores S.N.I. nivel II; miembro de número de la Asociación Mexicana de Urbanistas (AMU), y miembro de la Red de Estudios de la Forma Urbana (REFU). Sus líneas de investigación son metropolización territorial; crecimiento urbano; movilidad; espacio público; lógicas y modelos de localización, vivienda social, y Sistemas de Información Geográfica aplicados a la planeación y a los estudios urbanos. Entre sus publicaciones destacan MONTEJANO, J., COS, C. (Coords.) (2017). Densidad, Diversidad y Policentrismo: ¿Planeando Ciudades más Sustentables?. México: CentroGeo-CONACYT.; MONTEJANO, J., COS, C. Y CÁRDENAS, J. (2015). Contesting Mexico City's alleged polycentric condition through a centrality-mixed land-use composite index. Urban Studies, 0042098015588685. Como actividades profesionales en el campo de la planeación, diseño urbano y arquitectura, destacan la Actualización del Plan Regional de Desarrollo Urbano del Valle de Toluca (Dr. Alfonso Iracheta Cenecorta), CENTRO EURE, 2011; Diseño Urbano e implementación de programas de mejoramiento urbano en 4 comunidades del Estado de Campeche.

Camilo Alberto Caudillo Cos<https://orcid.org/0000-0002-2450-5295>

El Dr. Camilo Alberto Caudillo Cos es Profesor-Investigador del Centro de Investigación en Ciencias de Información Geoespacial (CentroGeo). Es Doctor en Ciencias de Información Geoespacial por CentroGeo, además de contar con una Maestría en Estudios de Población por la Facultad Latinoamericana de Ciencias Sociales (FLACSO-México) y una Maestría en Geomática por el propio CentroGeo. Es miembro del Sistema Nacional de Investigadores S.N.I. nivel I. Su formación académica integra enfoques de la geografía, la demografía, el análisis espacial y los estudios urbanos. Su trayectoria de investigación se ha enfocado en los campos de la geografía urbana, los estudios urbanos, la movilidad urbana, la segregación residencial, la geodemografía, la econometría espacial, el análisis de redes y el mapeo del crimen, especializándose en la aplicación de métodos cuantitativos y técnicas avanzadas de análisis espacial para el estudio de fenómenos territoriales y urbanos

Rodrigo Tapia McClung<https://orcid.org/0000-0002-7784-5867>

Investigador del Centro de Investigación en Ciencias de Información Geoespacial (CentroGeo), donde actualmente me desempeño como Coordinador Académico de la Maestría en Ciencias de Información Geoespacial. Soy físico por la Facultad de Ciencias de la UNAM, maestro en Estudios Ambientales por la Wilfrid Laurier University (Canadá) y doctor en Ciencias de Información Geoespacial por CentroGeo. Es miembro del Sistema Nacional de Investigadoras



e Investigadores, Nivel I. Su investigación se centra en la analítica geovisual, la ciencia de datos geoespaciales y el análisis espacio-temporal, con el objetivo de desarrollar formas de visualizar y explorar grandes volúmenes de datos espaciales y espacio-temporales que comuniquen información de manera dinámica, intuitiva y oportuna para apoyar la comprensión y la toma de decisiones. Su trabajo integra sistemas de información geográfica, percepción remota, visualización interactiva y plataformas geoespaciales para el estudio de fenómenos sociales, urbanos y ambientales como el crecimiento urbano, la delimitación de ciudades, la isla de calor urbano, la incidencia delictiva, la movilidad, la vulnerabilidad social y los sistemas alimentarios.

Sus investigaciones han dado lugar a publicaciones en revistas internacionales y al desarrollo de software y bases de datos especializados en ciencias de información geoespacial.

Felipe Gerardo Ávila Jiménez

<https://orcid.org/0000-0001-5590-5577>

Es licenciado en Diseño de los Asentamientos Humanos por la Universidad Autónoma Metropolitana. Tiene estudios en Planeación y Políticas Metropolitanas y es maestro en Gestión de la Información, por la Universidad de la Habana, Cuba. Se desempeña como Profesor Titular de tiempo parcial en la UAM-X, impartiendo el apoyo de Métodos Cuantitativos en la licenciatura de Planeación Territorial y actualmente se encuentra cursando el Doctorado en Proyectos, por la Fundación Universitaria Iberoamericana. Su formación académica y profesional le ha permitido especializarse en análisis espacial con aplicación al ordenamiento territorial y a la planeación urbana y ambiental. Ha colaborado con el Programa Universitario de Estudios Metropolitanos de la UAM. Su participación ha abordado aspectos de planeación urbana y ordenamiento del territorio; así como, análisis estadístico y espacial con aplicaciones tecnológicas y sistemas de información geográfica. Desde el año 2015 a la fecha colabora con CentroGeo en diversos estudios como investigador asociado.

CUADERNO DE TRABAJO

Diseño y aplicación metodológica para la definición de límites de ciudades, mediante la Teoría de la Percolación

En este cuaderno de trabajo se describe una propuesta metodológica para identificar los límites de las ciudades mexicanas mediante el uso de la teoría de la percolación, la entropía de Shannon y algoritmos espaciales como DBSCAN y CCA. A partir del procesamiento de datos geoespaciales provenientes de fuentes como INEGI, DENUe y la Agencia Espacial Europea, el estudio busca superar la subjetividad y los costos de los métodos tradicionales al identificar umbrales críticos de complejidad urbana. Los resultados de este modelo fueron validados comparándolos con el Sistema Urbano Nacional (SUN 2018) y evaluando su correspondencia espacial mediante la Ley de Zipf y el índice de Jaccard.

