Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/4/8/eaar4644/DC1

Supplementary Materials for

Toward cities without slums: Topology and the spatial evolution of neighborhoods

Christa Brelsford*, Taylor Martin, Joe Hand, Luís M. A. Bettencourt

*Corresponding author. Email: brelsfordcm@ornl.gov

Published 29 August 2018, *Sci. Adv.* **4**, eaar4644 (2018)
DOI: 10.1126/sciadv.aar4644

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/4/8/eaar4644/DC1)

Movie S1 (.mp4 format). Urban topological invariance.

Supplementary Text

Section SA. Topology of access systems

Recently, growing availability of data from digital geo-referenced urban maps and analysis tools, especially from complex network theory, have led to a growing number of studies to characterize the spatial layout of cities.

One area of intense activity has been the statistical analysis of street patterns, once these are represented as complex networks (graphs). Much like our construction of the network graph *Y*, below, city streets are often abstracted from the actual access surface to form a *primal graph* where nodes represent intersections and edges represent centerlines e.g. (*21*). Alternatively, the access system has also been represented as a *dual graph*, where nodes represent streets and edges represent street intersections e.g. (*17*, *22*). This process of dual graph construction is analogous to the first-stage weak duals we used to prove the city block theorems of section SE, but clearly used in a different way and at a different scale (block vs. city). The authors of reference (*21*) argue for certain advantages of the primal approach.

Once these graphs have been constructed based on urban and regional data, many standard complex network properties have been measured, including statistical distributions of street lengths, several measures of centrality, and the distribution of block areas.

In most cases, except in intensively planned layouts these statistical distributions are very broad (well described by power-laws or lognormals) pointing to the large variability of street patterns and block areas and, in some cases, even to their apparent scale independence.

Although combinations of these measures have been used to attempt to classify urban shape classes the large variability makes the *geometric classification* of cities somewhat arbitrary, depending on the choice of index quantities and cutoffs between categories. This has been known more qualitatively through many attempts to distinguish city geometries (*29*) based on street patterns.

Hence, we believe that accepting that urban geometry is variable while the topology of cities, as derived here, is universal provides us with the most fertile starting point to analyze the geography of built space in cities. In particular, this allows us to emphasize the primary character of cities as built spaces that co-evolve with their socioeconomic life.

As introduced in the main text, we can systematically account for all urban built space by dividing it into two categories: 1) the *Access System*, including all roads, and paths; and 2) *Places*, which include buildings and private and public spaces. In this section, we discuss the *topology of the access system*. The relationship between these two spaces and a proof of their topological equivalence is given in section SF, while the general topology of places within city blocks is characterized in section SB, with proofs of theorems given in section SC.

The urban access system is the total physical volume of paths and roads in any given city. A schematic example is shown in Fig. S2A, and a small part of Las Vegas' urban access system is shown in Fig. S3.

The access system of every city is always a compact, path-connected, orientable, 2-dimensional surface. *Path-connected* means that any point on this surface can be reached from any other point by traveling on the surface. *Orientable* means that there is a global unambiguous definition of up from down. This means that all real access systems in cities are never like familiar non-orientable surfaces, such as a Möbius strip. *Compact* is a somewhat more abstract concept that refers to the topological finiteness of a surface. In practice, a surface is compact if and only if any triangulation uses a finite number of triangles. This means in practice that we can imagine dividing the total access surface of any city into a number of adjacent triangles. We can do this using a smaller or larger number of triangles, and it is clear that one would need only a finite (though possibly very large) number. The 2-dimensional (2D) character of the access system is obvious in that we can only move on road and path surfaces: we cannot use the underside of these pathways.

Moreover, the urban access system is a 2D surface with *boundaries*. Unlike familiar compact closed surfaces, such as the surface of a sphere or of a torus, the urban access system has both internal and external *boundaries*: On the outside, there is a boundary that traces city limits. On the inside, there is the boundary between accesses and each city block, which consists of *place*s. A schematic representation is shown in Fig. S2.

Given these five characteristics of the urban access system (compact, connected, orientable, 2D, bounded) we can now characterize its topology using well-known mathematical results.

The only less familiar aspect of the access surface relative to classical examples in topology is the existence of boundaries. The topology of surfaces with boundaries is analyzed with reference to a corresponding surface without boundaries, effectively by plugging each hole with a disk. Then, the resulting surface without boundaries can be analyzed using classical results: the two surfaces are related by keeping track of the number of boundary components. For a city with *b* blocks it is easy to count boundary components: there are *b internal boundaries* and *one external boundary* so that the total number of boundary components, *B*, is *B*=*b*+1.

Analogously to the entire city we define a city *subsection* as a set of contiguous city blocks and surrounding access system, including an external boundary that defines the physical limits of the subsection. When a subsection includes all blocks it is equivalent to the city, and shares its entire access system.

We now formalize our procedure, by applying well-known results in topology to any abstract urban access system. Our goal in this section is the derivation of a topological invariant that characterizes any given city and establishes the topological equivalence between the access systems of different cities and/or city subsections.

In mathematics, two dimensional manifolds are classified by orientability and number of "handles," called genus. A sphere is thus the only orientable, genus zero, two-dimensional space. As an urban access system is necessarily an orientable surface with no genus, it follows that:

Theorem: Topological Classes of Urban Access Systems
The access system of any city with *b* blocks is topologically equivalent to a sphere with *B*=*b*+1 disks removed.

Corollary:
For any value of *b*, two cities with *b* blocks have access systems that are topologically equivalent, since both cities' access systems are topologically equivalent to a sphere with *b*+1 disks removed**.**

Corollary:
Any subsection of one city with *b* blocks has an access system that is topologically equivalent to that of another subsection of another city with *b*' blocks, if and only if *b*=*b*'.

Here, we define a city subsection as a set of *b* blocks and an external boundary that defines that subsection in such a way that it is isomorphic to a circle. Then, the theorem above guarantees that the subsection is topologically equivalent to a sphere with *b*+1 disks removed. The two subsections will only be equivalent to each other if and only if they are both equivalent to a sphere with the same number of disks removed. Hence, they will be topologically equivalent if and only if *b*=*b*'**.** See also below.

Corollary:
The access system of an entire city with *b* blocks is topologically equivalent to a subsection of any other city with the same number of blocks.

Since the access system of a subsection is the same type of surface as an entire city, they are topologically equivalent spaces if they have the same number of blocks**.**

Finally, we wish to stress the universal character of the topology of any city access system by characterizing it in terms of a topological invariant. Recall that a topological invariant is defined as a quantity α if, whenever two objects *X* and *Y* are topologically equivalent, then α(X) = α(Y).

We can characterize urban access system surfaces in terms of the familiar Euler characteristic, χ. The Euler characteristic, χ, can be applied to objects (complexes) of any dimension.

For a graph (or 1-complex, consisting only of vertices and edges, where we do not take closed cycles as 2D surfaces), χ is defined as $\chi = v - e$, where *v* is the number of vertices, and *e* the number of edges. For any graph that is a tree, T, a well-known result is that $\chi(T) = 1$.

For surfaces (2-complexes), including planar graphs with faces, χ is defined as $\chi = v - e + f$. It can be shown that for *any* planar graph, Y, $\chi(Y) = 2$, a well-known result we will return to later.

Finally, we would like to compute the Euler characteristic for any urban access system. The final ingredient we need to consider is how to compute χ for a surface with boundaries.

To do this, consider a general surface *S* with *B* boundaries. Then define an associated surface *S**, where *S** is *S* with all *B* boundaries patched by disks, each sewn up along each of the boundaries of *S*.

We then conclude that the Euler characteristic for the two surfaces are related by

$$\chi(S^*) = \chi(S)+B$$

This is because, whatever the Euler characteristic for *S* is, it must be that of *S** minus that of *B* disjoint disks ($\chi(\text{Disk}) = 1$). Thus, we can compute the Euler characteristic of any surface with boundaries from that of the corresponding *S** surface without boundaries and then subtract the number of boundaries.

This allows us to conclude that the Euler characteristic of an urban access system with *b* blocks (entire city or subsection) is

$$\chi(\text{Access System with } b \text{ blocks}) = \ 1 - b$$

where we used the fact that *S** is a sphere and thus $\chi(S^*) = 2$.

To conclude our results, we employ a well-known result in topology. Oriented, compact connected surfaces with boundary are topologically equivalent if and only if they share the same number of boundary components and have the same Euler characteristic.

Corollary:
Urban Access Systems are topologically equivalent if and only if they have the same number of blocks.

These results allow us to conclude very generally that the access system of any subsection of any city is equivalent to another as long as they have the same number of blocks. Each one can be deformed into the other continuously, so that commensurate sets of blocks in Paris, New York City, Las Vegas or Harare, *etc*, are actually topologically equivalent even though they can have radically different geometries.

Finally, we show how the most obvious graph representation of the access system retains the topological signature of its surface, which can be read off from its "graph topology".

Corollary:
A 1-complex graph representation of an urban access system with *b* blocks, where edges correspond to road and path centerlines and nodes correspond to their intersections, called the *urban access network, Y*, has the same Euler characteristic as the urban access system, $\chi(Y) = \ 1 - b$.

Proof: We showed above that $\chi = \ 1 - b$ for an urban access system with *b* blocks. Now we show that $\chi = \ 1 - b$ for the corresponding urban access network, *Y*.

A 2-complex planar graph always has $\chi = v - e + f\ = 2$, while the corresponding 1-complex graph has $\chi = v - e$. A 2-complex graph representation, Y_2, of the urban access system that is constructed in the same manner as *Y*, will have the same number of faces as the urban access system has boundaries. There is one face for each block boundary, and an exterior face that corresponds to the access systems exterior boundary: $f = b+1$.

Then

$$\chi(Y_2) = v - e + f = \ 2$$
$$\rightarrow v - e = 2 - f = 2 - (b+1) = 1 - b = \ \chi(Y)$$

Thus, for any city, the Euler characteristic of the network graph, *Y*, coincides with that of the access system surface. For this reason, in other parts of this manuscript, we sometimes refer to the topology of the access system surface and the access network interchangeably.

Section SB. Topology of city blocks

We now consider a different, and in some ways more challenging, topological problem: the internal organization of city blocks. We will use the general term *parcel* to denote the decomposition of the city block land area into separate units: these are buildings, or more generally, separate land holdings and include public places that are not accesses.

The problem of analyzing the topology of city blocks is decomposed into two steps. The first step deals with the relationship of parcels to the access network and is treated in section SF. If a parcel is adjacent to any section of the extant access network then we simply call it *accessible*. A parcel that is not adjacent to the access network is *internal* to the block: its access in practice is mediated through other parcels.

If all parcels in a block are *accessible* we call the block *universally accessible.* As we show in section SF the *topology of these blocks is equivalent to that of the access network*. All blocks are universally accessible in developed cities. For this reason, we propose that all cities, as they develop, are eventually made up of universally accessible blocks and thus that the topology of cities is set by that of their access networks.

The second step deals with non-universally accessible blocks. These blocks are characteristic of dense informal neighborhoods where there was no overall formal planning procedure during construction. This situation is characteristic of many modern day slums, Figs. 1-2, S5-8; sections of ancient cities also often fall in this category, see Figs. S9. Many now developed cities once had neighborhoods with many non-universally accessible blocks.

We prove in section SC a set of theorems that characterize the topology of blocks in terms of their *accessible/non-accessible* character. The gist of these theorems is the identification of non-universal blocks, via the graph theoretical analysis of the spatial relationships between parcels. The first theorem gives a necessary and sufficient condition to identify non-universally accessible blocks, even in enormously complicated neighborhoods with many layers of internal parcels (high block complexity), see Figs. 1-2, S8. It also provides the underlying mathematics for the algorithmic topological optimization problem described in the main text and in section SD.

The next two theorems establish the minimal number of parcels that need to be crossed to render a neighborhood universally accessible. The second theorem makes use of graph theoretic concepts by mapping the number of crossings to the stage of a weak dual graph characterizing the neighborhood and thus the number of internal graph cycles that need to be opened to make the weak dual graph a tree.

Finally, the third theorem establishes the topology (and geometry) of the minimal length access networks that render the neighborhood universally accessible. In the same spirit of results in efficient transportation networks we prove –– in different ways from Rinaldo et al. (*26*) or West et al. (*27*)–, that such sections of access networks are *tree graphs*. Thus, minimal-length new path and road segments additions are necessarily culs-de-sac. These can be seen in our model solutions, Figs. 3A-B, S5, S7-8, and in many examples from real neighborhoods, see e.g. (*34*) Figs. 1C, S9.

These results, together with the topology of access networks, establish not only the universal topology of cities but also the mathematics of the spatial transformations necessary for poor and unplanned neighborhoods to develop gradually.

Because of the mathematical nature of these results, algorithms for optimal re-blocking and efficient access can then be readily created and applied to real-world situations.

Section SC. City block topological theorems

We complete our analysis of the topology of cities by dealing with city blocks that are *not universally accessible*. We use graph theory to prove a set of theorems that show how even blocks with extremely complicated parcel structures can be quantitatively analyzed and rendered connected with minimal disturbance.

Definition: A block S is called *universally accessible* if every parcel within S borders a road. Otherwise, S is not universally accessible.

Definition (minimal set of accesses)**:** Interior parcels can be connected to the urban access system by converting edges in the S_0 graph (see Fig. 1, S5) from parcel boundaries to roads. The *minimal set* of additional roads necessary to connect a given parcel to the road system is the set of edges with the shortest total length such that at least one node contained in the set of edges to be converted is part of an existing road, and at least one node is part of the face in S_0 that surrounds the parcel. The minimal set of accesses is the unique solution (section SC) of the strict topological optimization problem discussed in section SD.

Theorem One:
A block S is universally accessible if and only if its stage-two graph, S_2, is a tree.

Theorem Two:
If a parcel is represented with a node in the S_k graph, at least $\frac{k-1}{2}$ parcel boundaries must be crossed in order to reach it from the nearest section of the access system.

Theorem Three:
There will be no loops in the minimal set of additional roads necessary to connect all interior parcels to a road. Thus, newly constructed roads in the minimal set of accesses form a tree or set of trees (culs-de-sac).

Proofs:

We start by laying down some additional definitions and the procedure of (weak dual) graph construction. Consider a neighborhood, divided into parcels of land separated into connected components of land (blocks) by roads. Here, we will assume that no single parcel of land completely encloses another, no parcel of land touches a road at only points, and no two parcels share multiple non-contiguous borders.

Then, given a block S, we assign a *stage zero graph* S_0 where nodes and edges are created to represent the parcel's geometric boundaries (see also Fig. 1A, S5). This graph is a planar graph, meaning that edges only intersect at nodes. We think of the graph S_0 as sitting on the surface S.

Definition: In a graph G, a *cycle* is a collection of m vertices and m edges arranged so that each vertex has exactly two edges incident to it, where $m \geq 3$.

As usual, the *degree* of a vertex is the number of edges incident to it.

Definition: A *face* of a planar graph is a maximal region in the plane that contains no edge or vertex of the graph.

Note that every planar graph has one unbounded, exterior face. Here, we will disregard the exterior face so that each face in S_0 corresponds to a parcel in S. Now, given a block S and a stage zero graph S_0 for S, we can define a *stage one graph* S_1 in the following way:

Definition (weak dual graphs): For each bounded face of S_0, we assign a vertex in S_1. Two vertices of S_1 have an edge between them if and only if the faces of S_0 they represent share a common border of at least one edge in S_0. Then, S_1 is the *weak dual graph* of S_0. For a block S, we may then assign a *stage k graph*, S_k, defined recursively by repeating the process used to construct S_1 from S_0 on the stage k-1 graph S_{k-1}. See Fig. 1 for an example of this construction.

Definition: A vertex v of a graph G is called an *interior vertex* if there exists a cycle surrounding v so that deleting this cycle from G results in either:

1) Two connected components, one of which contains vertex v and all of its incident edges,

or

2) Just the vertex v and its incident edges. (See Fig. 1 for an example.)

Thus, a parcel n in block S does not border a road if and only if n is surrounded on all sides by other parcels in S. This is true if and only if the vertex v of S_1 that corresponds to parcel n is an interior vertex of S_1.

Definition: A graph G is called a *tree* if G contains no cycles.

Proof of Theorem: *A block S is universally accessible if and only if its stage two graph S_2 is a tree or set of trees***.**

First, we show that if S is universally accessible then S_2 is a tree. We do this by showing that if S_2 is not a tree then S is not universally accessible.

Suppose that for a given block S, S_2 is not a tree. This means that there exists an interior face of S_2 whose boundary is a cycle σ consisting of m vertices $\{x_1, x_2, \ldots x_m\}$ of S_2 and m edges. Each vertex x_i in σ represents a face f_i of S_1, where face f_i shares a common edge with face f_{i-1} (mod m) and face f_{i+1} (mod m). Furthermore, each of these shared edges is incident to a vertex v of S_1 that represents the interior face of S_2.

Thus, the cycle σ in S_2 corresponds to a subgraph of S_1 consisting of the m faces, $\{f_1, f_2, \ldots, f_m\}$ arranged in a circle around the vertex v. This means that vertex v is an interior vertex of S_1, so it corresponds to a parcel of the block S that does not border a road. This shows that block S is not universally accessible.

Now, we will prove that if a block S is not universally accessible, its stage two graph, S_2, is not a tree. We assume that there exists a parcel n of a block S that does not border a road. Thus, there is a vertex v_n of S_1 corresponding to parcel n that is an interior vertex of S_1.

Consider the subgraph V_1 of S_1 consisting of a minimal cycle surrounding vertex v_n, vertex v_n itself and all edges incident to vertex v_n. Now, we consider the subgraph V_2 of S_2 that represents V_1. V_2 will contain one vertex for each face of V_1 connected by one edge representing each edge incident to vertex v_n.

We conclude that the subgraph V_2 of S_2 is a cycle with m vertices, where m is the degree of vertex v_n in S_1. This says that the stage two graph of S contains a cycle, and is therefore not a tree. This concludes the proof of Theorem 1**.**

Note that if a stage k graph S_k is a tree, then the stage (k+1) graph S_{k+1} vanishes, as there are no interior faces in S_k, so there are no vertices in S_{k+1}.

Definition (block complexity)**:** From this, we may generalize a *complexity* on the block S as the smallest positive integer k such that S_k is a tree. Every block will be characterized by a positive, discrete value of this complexity. The complexity of universally accessible blocks is $k \leq 2$. Non-universally accessible blocks will have $k > 2$.

The complexity of a block S or a parcel represented in graph S_k is also useful for determining how many parcel boundaries must be crossed to reach an access from an interior parcel. This gives a topological measure of the difficulty of accessing the urban spaces from a given parcel and of achieving universal connectivity for a given block.

Theorem Two: If a parcel is represented with a node in the S_k graph, at least $\frac{k-1}{2}$ other parcel boundaries must be crossed in order to reach an access (road).

Proof of Theorem Two:

For any parcel n of block S, the minimum number of parcel boundaries that must be crossed in order to reach a road is represented by the minimum number of edges necessary to form a path from v_n, the vertex representing n in the S_1 graph, to an exterior vertex of S_1.

Observe that, in the algorithm for creating the S_k graph of a block S, parcels of S are represented by faces of S_k when k is even and nodes of S_k when k is odd. Furthermore, for even k, if a face of S_k touches an exterior vertex, that face is represented by an exterior vertex in S_{k+1}. Finally, observe that, for odd k, parcels represented by an exterior vertex of S_k are not represented at all in S_{k+1}.

Therefore, suppose a parcel n requires a path of length l to connect vertex v_n to an exterior vertex in S_1. It is clear that in S_3, the path from v_n to an exterior vertex will have length $l-1$, and so on. The vertex v_n will thus be an exterior vertex of the graph S_{1+2l}. Therefore, we see that, if vertex v_n appears in graph S_k, then $k \leq 1 + 2l$, which says that $\frac{k-1}{2} \leq l$. This concludes the proof of Theorem 2.

Theorem Three:

There will be no loops in the minimal set of additional roads necessary to connect all interior parcels to a road. Thus, newly constructed roads in the minimal set of accesses form a tree or set of trees (cul-de-sacs).

Proof of Theorem Three:
We may consider the access network of a given block as a subgraph of the stage zero graph S_0. In order to connect all parcels to a road, we consider parcel boundaries, which are represented by interior edges in S_0. We may then choose a set of such edges of S_0 to represent additional segments of road needed to ensure that the block is universally accessible. There will be several choices for this set of additional roads; we choose the one that has the fewest total geometric length of edges (*minimal* set of accesses).

Suppose that there exists a block for which the minimal set of additional roads is not a tree or set of trees. Let M denote the subgraph of S_0 consisting of edges belonging to the minimal set of roads along with the nodes incident to these edges. We are assuming that there is at least one cycle in M. Every face of S_0 representing an interior parcel must share at least one node with M in order for every parcel to be accessible via existing or new paths. However, all connected planar graphs have a spanning tree, which is a subgraph containing all nodes of the graph but no cycles. Then, we let M' be the subgraph of M consisting of spanning trees for each component of M. Thus, every face of S_0 representing an interior parcel will share a node with M', making every parcel accessible via existing or new roads, but M' has strictly fewer edges than M, as it is a subgraph containing no cycles. This contradicts the choice of M as minimal. Therefore, the set of newly constructed roads must form a tree or set of trees. This concludes the proof of Theorem 3.

The result of Theorem 3 explains the proliferation of culs-de-sac under strict topological optimization, a curious fact widely observed in the development of street patterns in dense old cities (*34*) and in modern informal settlements. However, this pattern – the "cul-de-sac effect" - is often actually desirable as it may promote privacy, safety and community cohesion (*30*) and is often also observed in *designed* planned suburban communities, e.g. in Las Vegas, NV (Fig. S3).

Section SD. Topological optimization: Minimal neighborhood reblocking

Finding the reblocking solution described above in theorem three for any given neighborhood requires general tools that transform a geometric spatial configuration of places into an algebraic problem. In this section, we define a problem of *topological optimization* by asking how to provide each parcel with access using the smallest amount of infrastructure (path length).

We define this *topological optimization* problem in two different ways. The first (*strict* optimization) is easier to understand and formulate but is too rigid for practical use. The second (*statistical* optimization) is more flexible and can be the basis for practical neighborhood upgrading tools discussed in the main text.

The *strict optimization* approach seeks to find the absolute smallest amount of new paths that need to be built to provide each interior parcel with access. In practice, there is always a unique solution to this problem.

Given a geo-referenced parcel map, the strict optimization problem is solved algorithmically by identifying the shortest path from each interior parcel (along parcel boundaries) to the existing road and measuring its geometric length. While optimal piecewise – parcel-by-parcel this identifies the smallest path length – this strategy may not be globally optimal when there are many internal parcels that are adjacent. This is because there may be solutions with shorter paths when two or more parcels are considered together rather than individually.

This problem of strict optimization can be pursued further, but typically leads to a search space that increases combinatorially (exponentially or faster) as pairs, trios, *etc*, of parcels are considered together. Though feasible for small blocks, this becomes slow or intractable for larger ones. Above all, building the strictly shortest set of accesses is likely not to be viable because of other local considerations, which the algorithm cannot *a priori* incorporate. These have come up in our discussions with resident associations and city agencies, especially in Mumbai, India. As such, we next emphasize the generation of an ensemble of possible access networks that can be discussed and edited by policy makers and residents, thus defining a *statistical optimization* problem.

The *statistical optimization* problem seeks to identify a set of path configurations that includes the solution of the strict optimization problem but also optimizes for other statistically small amounts of infrastructure introduced into the block. Because there may be many configurations of paths that are not strictly the shortest but that may have other advantages, this leads to a more flexible and realistic set of possible solutions.

In this way, we create a sample of solutions from a statistical ensemble of spatial configurations that render the block universally accessible. This optimization problem makes use of methods of statistical physics by defining a statistical distribution of neighborhood path configurations and sampling it using well-known Monte Carlo techniques of importance sampling (*28*).

To solve this second problem, a statistical ensemble of possible solutions is built in the following way: i) we define a probability distribution over paths, $p(l)$ where l is a path, which makes each path that solves the topological optimization problem more or less likely as a function of its characteristics (cost); ii) we choose a number of paths that connect each interior parcel with the outside and, thus, render the block universally accessible.

In practice, we make path cost a function of path length. Thus, shorter paths that require less infrastructure are more likely. Other properties that contribute to relative path cost – such as how straight it is - can also be included in defining this probability, if desirable.

We have generated an algorithm that implements this strategy: For each parcel we identify (strict optimization) the shortest path that makes it accessible. We then further construct n_l additional short paths with the same property. This set of n_l +1 short paths is identified for all interior parcels. In all practical examples developed here we chose n_l +1 =10. The choice of ten paths per interior parcel is arbitrary: a greater number can be chosen at larger computational cost and vice-versa.

Out of the set of sets of short paths for all interior parcels, a single path is selected with a probability, $p(l)$. We experimented with different functional forms of $p(l)$, which penalize to a greater extent longer paths. In Figs. 2-3, S5, S7-8, we show results for $p(l)$ inversely proportional to path length to the nth power, $p = \frac{1/l_i^n}{\sum l_i^n}$. This is equivalent to $p(l) \sim e^{-n \ln l}$, which thus penalizes long paths rather strictly, as a polynomial function of the *logarithm of path length*. In this context, the power of the logarithm (n=8 in Figs. 2-3, S5, S7-8) plays the role of an (dimensionless) inverse temperature in statistical mechanics (*28*).

Once selected, the path is converted from an interior edge to an access and the set of interior parcels is updated to reflect the new access layout. This path selection process is repeated until no interior parcels remain, and the S_2 graph for the block becomes a tree (section SC). We have built a demonstration website (openreblock.org) where each step in this path selection process can be viewed for two examples, and a beta version of an interactive website (beta.openreblock.org) where a user can upload the geometry for any block, and see the results generated by the algorithm.

Fig. 1D shows a single block from Epworth (Harare, Zimbabwe, block complexity k=3) with four interior (not accessible by road) parcels highlighted in red. Existing roads encircle the block. Fig. S5 shows an instance of the iterative road construction process for the block shown in Fig. 1. Fig. S6 shows the actual neighborhood geometry before and after a community driven re-blocking process. Fig. 2 shows a similar problem for a much larger block in Khayelitsha (Cape Town, South Africa), with block complexity k=9 and 433 parcels (and population of about 1,400 people), of which only 175 have direct road access; Fig. 2C shows a solution to the statistical optimization problem described here.

Section SE. Geometric optimization: Travel costs versus road construction

We have explained in section SD how different access configurations can be generated statistically to solve the *topological optimization problem* of granting access to each parcel in a city block.

Here, we formalize an additional *geometric* optimization problem that among possible solutions to the topological optimization problem can choose path configurations that reduce travel distance. Note that this second optimization problem deals with improving performance along a gradation of travel costs, whereas topological optimization is necessary and discrete: a city block is either universally accessible or it is not.

An important branch of literature considers the fundamental tradeoff between connectivity and construction costs to create generative algorithms for street networks. The literature of attempting to use algorithms based on these ideas to generate urban and regional layouts goes back at least several decades but recently these algorithms have become more sophisticated in terms of the optimization strategies invoked.

As we have discussed above, real city street networks lie on a continuum of morphologies between the two extremes of tree-like graph structures and fully connected (spatial, planar) networks. These two network extremes constitute the minimum and maximum number of links between a given number of nodes in a connected graph, respectively. Along this continuum one can define measures of network efficiency, "meshed-ness" coefficients and the balance between the information required to navigate the space and the distance of travel as possible targets of optimization for any street network design.

These measures have been used to consider how streets may be organized in the absence of a central planner, to consider possible principles about how city street networks arise or how roads should be laid out to facilitate optimal traffic flow, often by adapting algorithms from efficient transport networks in other complex systems.

The fundamental ingredient missing from current approaches is the fuller consideration of *places* as the end-points of these networks and, consequently, of emphasis on the *function* (rather than form) of urban road and street systems *in providing access between any two places*. Thus, such ideas do not yet generalize to a full theory of (built space) in cities. This is because the topological properties of street networks do not yet include full consideration of the relationship of streets to the places they are used to access or to consistent ways to topologically characterize each city and compare cities or parts of cities to each other. All these points have been well known to urbanists and we hope that a future convergence of more realistic optimization algorithms with topological and functional considerations will improve our current ability to tackle these issues in ways that emphasize the fundamental socioeconomic character of cities. We believe that the results of this paper constitute an important step in this direction and show the way to further developments.

To consider the issue of travel distance we introduce the following definitions. Consider *P*, the set of all accesses within a block, including all roads and paths. As a measure for evaluating the accessibility of an access system, we define the *n* by *n* travel cost matrix $\mathcal{T}$, where *n* is the number of parcels, with entries T_{ij} given by the distances between parcels *i* and *j* along the shortest path of *P*. This matrix has only non-negative entries and is traceless and symmetric. We define the average travel distance $\bar{\mathcal{T}} = \frac{\sum_{i,j} T_{ij}}{n^2}$, to be the average over all entries of $\mathcal{T}$. Note that the travel cost matrix is a non-trivial function of the block parcel configuration and its set of paths *P*, and in general needs to be evaluated computationally for each block configuration.

We now formulate the problem of finding the optimal block access network, *P**, that minimizes average travel costs $\bar{\mathcal{T}}$ under certain constraints. We take as a starting point the access network achieved as the solution to the topological optimization problem. From this, we are guaranteed a finite upper bound for $\bar{\mathcal{T}}$. In this section, we describe an algorithm for constructing new paths so as to have the greatest impact on travel cost reduction. These new paths will necessarily bisect the block. Note that the minimum possible value of $\bar{\mathcal{T}}$ would be obtained by connecting parcels all-to-all.

For small neighborhoods, the solution to this geometric optimization problem can be predicted visually. In general, however, this problem is known to be NP-complete. Thus, we describe an algorithm to successively construct new paths in the access system in a near optimal manner.

For a given parcel *p* in the neighborhood, the *geometric travel distance sum* G_p for *p* is defined to be the sum over all *n* parcels of the length of the line connecting the parcel centroid of *p* to each other parcel centroid in the neighborhood. The *on-network travel distance sum* N_p for *p* is defined to be the sum over all *n* parcels of the travel distance between *p* and each other parcel in the block, measured as the geometrically shortest path between the two parcels along network edges that are paths in the access system, not parcel boundaries. Note that this assumes that travel can only occur over roads and paths, a situation typical of dense urban neighborhoods. We then identify the parcel *p** for which the *travel distance ratio*, $\frac{G_p}{N_p}$, is minimal. This parcel *p** is least connected to the access network in the sense that the travel time from *p** to the rest of the neighborhood via roads and paths is comparatively much longer than the geometric distance from parcel centroid to parcel centroid. Once we have selected *p**, we identify the node *v* on the access network with the minimum travel distance ratio to *p** and we construct the shortest path between *p** and *v*. Note that this process guarantees that we create a through-fare, which bisects the block. We then construct a new travel cost matrix $\mathcal{T}$ reflecting the new path and compute the now reduced average travel cost $\bar{\mathcal{T}}$.

We have formalized an algorithm for making near optimal choices for reducing $\bar{\mathcal{T}}$, with minimal construction of new path infrastructure. New paths could be iteratively constructed until a given budget has been exhausted, minimizing $\bar{\mathcal{T}}$ given a budget constraint. Alternatively, new paths could be constructed until a desired $\bar{\mathcal{T}}$ has been achieved, thus meeting the objective of a specified reduction in travel costs with minimal new infrastructure. Finally, construction could continue until proposed new paths fail to meet a threshold of the ratio of new infrastructure cost to the benefits of reduced travel distances.

The on-network travel distance between any two parcels in a block is shown in Figs. 3, S7-8 as matrices in which parcels are ordered according to a hierarchical clustering procedure. This groups together parcels that are connected by short travel distances, which appear as darker blue matrix blocks. Parcels that are distant appear as orange and red entries in the same matrix. The latter are the primary target distances to be reduced by the geometric optimization described here.

Figs. 1D, 3A, S5-7 show the neighborhood layout and the matrix of parcel-to-parcel travel distance, in meters, for several possible path configurations for Epworth. Fig. S7A shows the solution to the topological optimization problem; panels B – D demonstrate the subsequent geometric optimization algorithm including three additional block bisections. The centroids of the two most distant parcels in this neighborhood are 205m apart, significantly smaller than the maximum on-road travel distance of 401m in the minimally connected case. The average travel distance decreases from 146m in the minimally connected case to 112m after a bisecting path of length 49.6m is introduced. The second bisection, of length 30m, reduces $\bar{\mathcal{T}}$ to 101m. The third bisecting path has length 72.7m and reduces $\bar{\mathcal{T}}$ to 90 m. For this neighborhood, the all-to-all minimal value of $\bar{\mathcal{T}}$ is 62m.

Figs. 2, 3B-C, and S8 show the neighborhood layout and the matrix of parcel-to-parcel travel distance, in meters, for several possible path configurations for Khayelitsha. Figs. 3C and S8 shows the travel cost matrix for this neighborhood after several through-fare paths have been introduced based on that same topological solution. The initial $\bar{\mathcal{T}}$ value (the result of the topological optimization described in section SD) is 214m, which is reduced to 186m after one bisection. The second bisection lowers $\bar{\mathcal{T}}$ to 161m, the fourth to 145m, and the sixth to 133m. These six bisections introduce a total of 118m of new paths.

Section SF. The topology of places is equivalent to the topology of the access system

We now discuss the topological relationships *between* places and the access system (roads and buildings). We describe in detail a method for building a continuous invertible map from parcels onto their surrounding paths and roads. We do this by building a graph representation of a city's accesses, a graph representation of the relationship between parcels and roads, and a mapping between the two that proves the they are homotopy equivalent. Homotopy equivalence is a specific type of topological equivalence.

In section SA, we analyzed the properties of the two-dimensional access system surface and described its topology. At the end of that section, we collapsed the two-dimensional access system down to the centerlines for the roads and paths that make up the access system, creating a graph representation that we call the *access network*. The access network, Y, is a graph where edges represent roads, paths, or other public rights of way, and nodes represent intersections between them. This is a common way to represent transportation "complex networks" in cities. In section SA we showed that the 1-complex access network, Y, corresponding to a given city's access system, has the same Euler characteristic as the access surface and is thus topologically equivalent to it.

In sections SB-C we demonstrated a general method for describing the topology of places, regardless of whether all structures are universally accessible or not. In sections SD-E, we demonstrated a method to transform the topology of a place in an optimal manner such that it is universally accessible, a necessary condition for the space of places to be topologically equivalent to the city's access system.

In this section we show that the space of places and their spatial relationships is topologically equivalent to the city's access system as long as the places are universally accessible. This is accomplished by showing that the space of parcels can be continuously retracted (i.e. "shrunk") into the access space. We start with a graph definition of the parcel space.

Definition (bridge graph)**:** We define a 1-complex *bridge graph*, X, representing the relationship of parcels to the access network, Y. Like Y, the bridge graph includes edges to represent roads, pathways and other public rights of way, and nodes to represent intersections between them. We define x as any element (edge or node) in X. In addition to containing all elements of Y, X also contains nodes to represent the centroid of each parcel in the city. This creates a set of initially disconnected nodes N, where each node n_i in N represents a specific parcel in the city. We then add a single edge, e_i, that connects node n_i to the edge or node of $x \in Y$ that the parcel most naturally accesses. If e_i connects n_i to an existing node n_i' in $x \in Y$ no further changes are needed. If e_i connects n_i to an edge e_Y in $x \in Y$, then e_Y is broken into e_{Y1} and e_{Y2} through a process called edge refinement and a node n_i' is added to represent the intersection through which this parcel enters the access network. Then e_{Y1} is an edge that connects nodes n_i' and n_Y.

In this way, X covers the space of Y with edges and nodes representing pathways and intersections between those pathways. X also includes nodes that represent intersections between public pathways and accesses to individual parcels. In this way, the nodes n_i representing each parcel should be thought of as representing the *intersections* of public and private spaces – for example one's front door.

With this definition it should be clear that *Y* is a subspace of *X*, although some nodes in *X* are contained in the edges of *Y* and some edges of *X* overlap only parts of edges of *Y*, see Fig. S4.

We construct a *homotopy equivalence* - a continuous deformation - of the topological space *X* into *Y*. The function we define below is a particular type of continuous deformation called a *strong deformation retraction*, defined as follows:

Definition: A continuous map *H(x,t)* is a *strong deformation retraction* of a space *X* onto a subspace *Y* if, for $x \epsilon X, y \epsilon Y,$ and $t = [0,1]$, three conditions hold:

1) $H(x, 0) = x,$
2) $H(x, 1) \epsilon Y,$ and
3) $H(y, t) = y.$

In layman's terms, a strong deformation retraction is a sequence of functions over a unit of time where points in the space *X* at "time" t=0 are pushed along into *Y* at "time" t=1. Meanwhile, points that start off in *Y* we stay fixed in *Y* for all values of time.

With these definitions, we can state the central result for this section:

Theorem: The space of *Places* and the *Access Network* of any city are homotopy equivalent for universally accessible blocks.

Proof: If there exists a strong deformation retraction from the bridge graph *X* to the access network *Y*, the two spaces are homotopy equivalent.

We prove this theorem by explicitly constructing a strong deformation retraction from *X* to *Y*. First, we define the map $r(x): X \rightarrow X$ by

$$r(x) = \begin{cases} x\,;\ x \in Y \\ n_i'\,;\ x \notin Y \end{cases}$$

Noting that {*r(x)*} is composed of nodes *n* and edges *e* and *Y* is composed of nodes Y_n and edges Y_e, we can define a second map *f(x)*: $X \rightarrow Y$

$$f(x) = \begin{cases} n \in Y_n\ \ ;\ r(x) \in Y_n \\ e \in Y_e\,; r(x) \in Y_e \\ n_Y\,;\ r(x) = n \notin Y_n \end{cases}$$

Then, the map *f* is a retraction from the bridge graph *X* to the *access network, Y*. The map *r* retracts all nodes representing parcels down to their intersection with the access network, and the map *f* retracts those private/public intersections along to the fully public intersections of roads and pathways in the access network, see Fig. S4 for an illustration. Fig. S4B clearly shows that *r(x)* retracts *X* to *Y*, but there is not complete node to node and edge to edge correspondence. The additional map *f(x)* shows that *r(x)* can be further refined to create perfect edge to edge and node to node equivalence between *Y* and a retraction of *X*.

We now define the map, *H(x,t)*

$$H(x, t) = (1 - t)x + t\, f(x)$$

and show that *H(x,t)* satisfies all three conditions for a strong deformation retraction:

Condition 1

$$H(x, 0) = (1 - 0) * x + 0 * f(x) = x$$

Condition 2, $H(x, 1) \epsilon Y$:

Based on the definition of *r(x)* and *f(x)*, all $x \notin Y$ are retracted to n_i'. If $n_i' \notin Y_n$, it is retracted to n_Y in the second step *f(r(x))*, where $n_Y \epsilon Y$. Thus $f(x) \epsilon Y \; \forall \; x \epsilon X$, and so

$$H(x, 1) = f(x) \epsilon Y$$

Condition 3, $H(y, t) = y$ for $y \epsilon Y$:

It is then clear from their definitions that *r(y)* = *y* and *f(y)* = *y*, and so $H(y, t) = y$ for any *t*.

Note also that we could have played the transformation in *H* "backwards in time" as *t* goes from 1 to 0, thus reconstituting *X* from *Y*. In this way, we have built a continuous map from the bridge graph *X* to the access network *Y*, defining a 1:1 correspondence between any parcel connected to the city's access network and some part of the access network. This shows that the topology of connected parcels in each universally accessible city block is homotopy equivalent to its access network. This concludes the proof.

This then allows us to conclude that the space of *places* and their interactions and the *access systems* in a city are topologically equivalent to each other and define the overall topology of any city.

The access network, *Y*, is likely to have trees entering blocks (culs-de-sac) and also trees that represent infrastructure connections to more distant cities (interstate highways, intercity railroads, *etc*). These trees do not influence the topology of *Y*, *X*, or the relationship between *X* and *Y*. The key factor that *r*(x) and *f*(*x*) rely upon to distinguish what is retracted or not is whether *x* or *n* are element of *Y*, so any trees in *Y* do not influence our ability to map *X* onto *Y*.

Note also that we could have shown even more explicitly that the graphs *X* and *Y* are themselves strong deformation retracts of the full urban space and the access system surface, respectively. The fact that these various surfaces eventually retract to *Y* shows their topological equivalence.

Together with the results of the previous sections we can now say that *any two cities, as sets of places and access systems, are topologically equivalent if they have the same number of blocks.*

This topological equivalence is demonstrated visually in Supplementary Video V1, which shows how a group of blocks from Midtown Manhattan in New York City can be deformed into part of the Summerlin neighborhood in Las Vegas, Nevada, and finally into part of Dharavi, a slum in Mumbai India.

This result does not apply to blocks where not every parcel is connected to the access system. If there exists a parcel represented by n_i that is not directly connected to the access system, no edge e_i will be created. It is then clear that *r(x)* cannot retract n_i along e_i into the space of *Y*. Then *H(x,t)* is not a deformation retraction from *X* to *Y* if *X* contains parcels that are not connected to the access system. Thus, two cities or neighborhoods are not in general topologically equivalent if their places are not universally accessible even if they share the same number of blocks.

The convergence of all these facts about cities and the generative ability of their formal topological characteristics is, in our opinion, an exciting prospect to understand their shape while simultaneously emphasizing the open-ended character of their built spaces and their functional socioeconomic roles. This formal understanding is needed to transform simple and static optimization processes into dynamic development dynamics for cities, thus hopefully breathing new (mathematical) life into the processes of spatial evolution in cities and urban systems.

Supplementary Figures

Fig. S1. Phule Nagar (Mumbai, India) path width. Measured path width for the Phule Nagar Settlement in Mumbai, India. Note that each place has some level of access for pedestrian movement, but that many paths are extremely narrow and preclude infrastructure provision, emergency access, or wheeled vehicle travel.

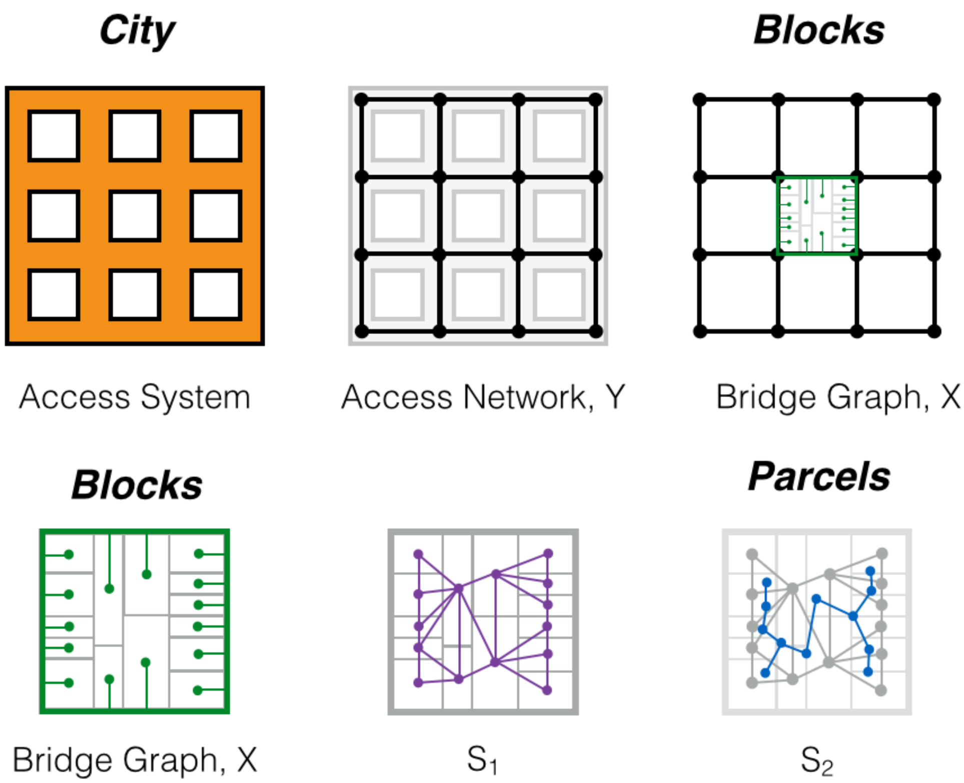


Fig. S2. Topological constructs for the systematic analysis of urban topology. From left to right in the top row these panels show how the built space of an entire city can be systematically analyzed in terms of the topology of its access system (orange), its access network *Y* (black), and the relationship from places (parcels) within blocks to their accesses represented as a bridge graph *X* (green). On the bottom row, the very local relationships between each parcel and its neighbors, represented by a hierarchy of S_k weak dual graphs for each city block (purple and blue).

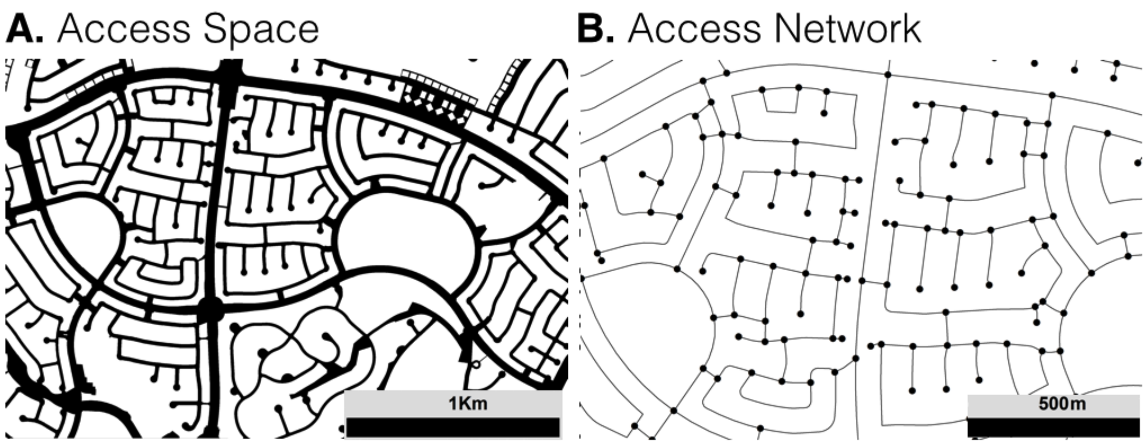


Fig. S3. Las Vegas (NV, USA) access system and access network. A shows the access system for a specific neighborhood. **B** shows the corresponding access network, *Y*. Data is provided by the Clark County Tax Assessor's Office. Note the prevalence of designed culs-de-sac, typical of many planned suburban communities. Note also the large special scale, and distances over the network in this neighborhood.

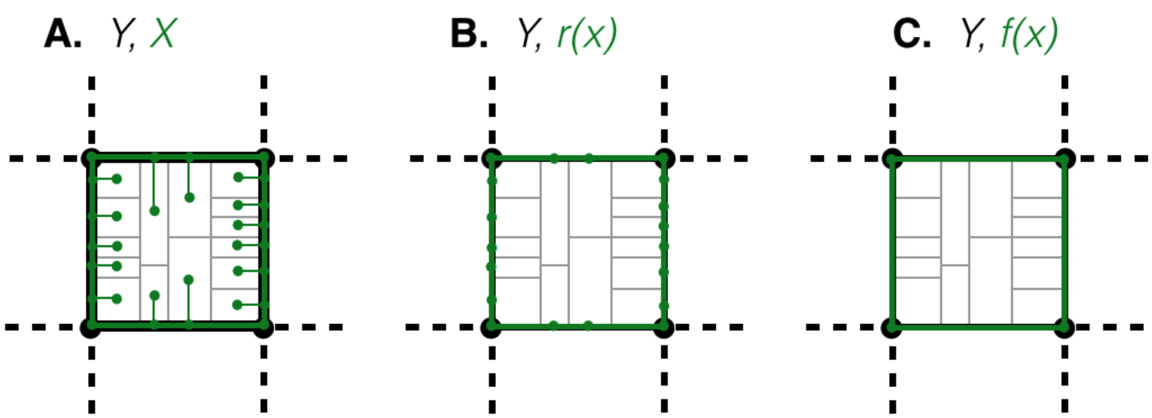


Fig. S4. Schematic bridge graph retraction. The two-step retraction from the bridge graph *X* (green), to the access network *Y* (black). **A** shows a subset of the original *X* and *Y* graphs for a single block. Dashed lines denote the continuation of *X* and *Y*. It is clear from **B** that *r(x)* is sufficient to map all nodes and edges of *X* into the space of *Y*, although there is not a node to node and edge to edge equivalence. **C** shows an additional map *f(x)*, which relies on the map *r(x)* to achieve full edge to edge and node to node equivalence between *f(x)* and *Y*, demonstrating explicitly the existence of a continuous deformation retraction from *X* to *Y*.

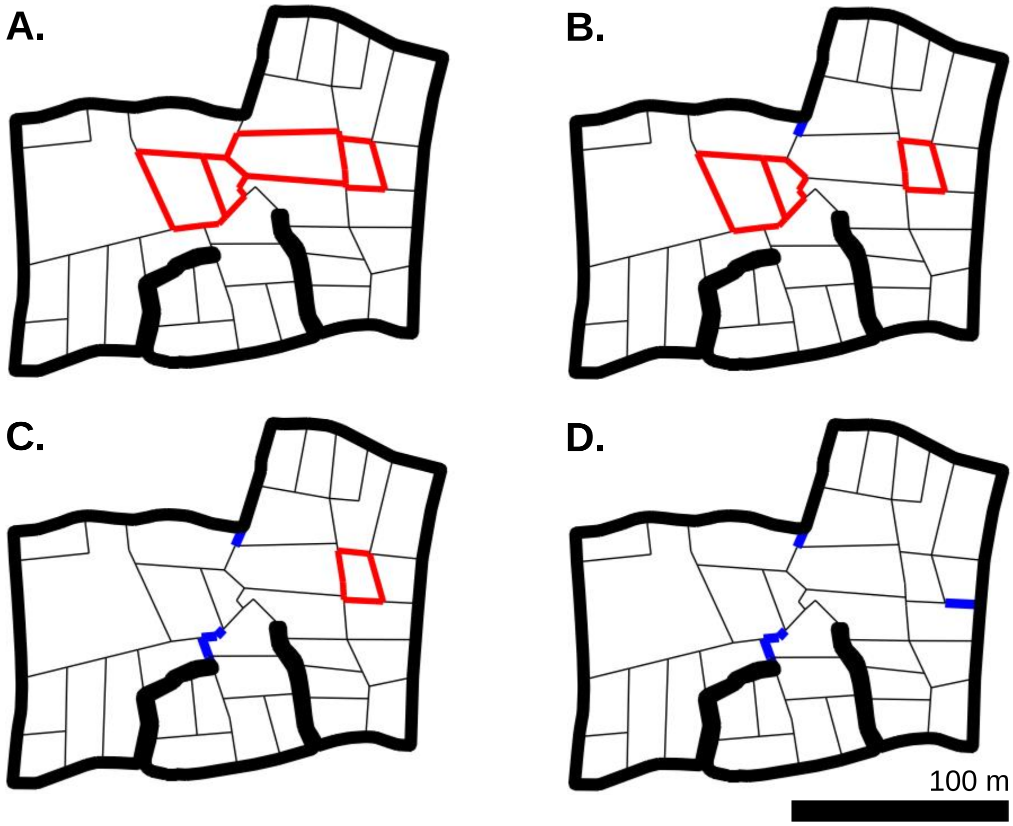


Fig. S5. Epworth (Harare, Zimbabwe) minimal reblocking. A. The example block shown in Figs. 1D & 3A, surrounded by the original access system (bold black) with four interior parcels (red). **B** One new path (blue) connects a single interior parcel to the access system. **C** An additional path (blue) connects two of the remaining interior parcels to the access system. **D** A third path (blue) connects the final interior parcel to the access system, making the block universally accessible.

Fig. S6. Epworth (Harare, Zimbabwe) before and after reblocking. A Many blocks in Epworth in their original layout before reblocking. **B** The planned outcome of the re-blocking procedure proposed by the resident community. Parcels highlighted (red) do not have direct road access. This community-driven re-blocking process has created access to the vast majority of internal parcels.

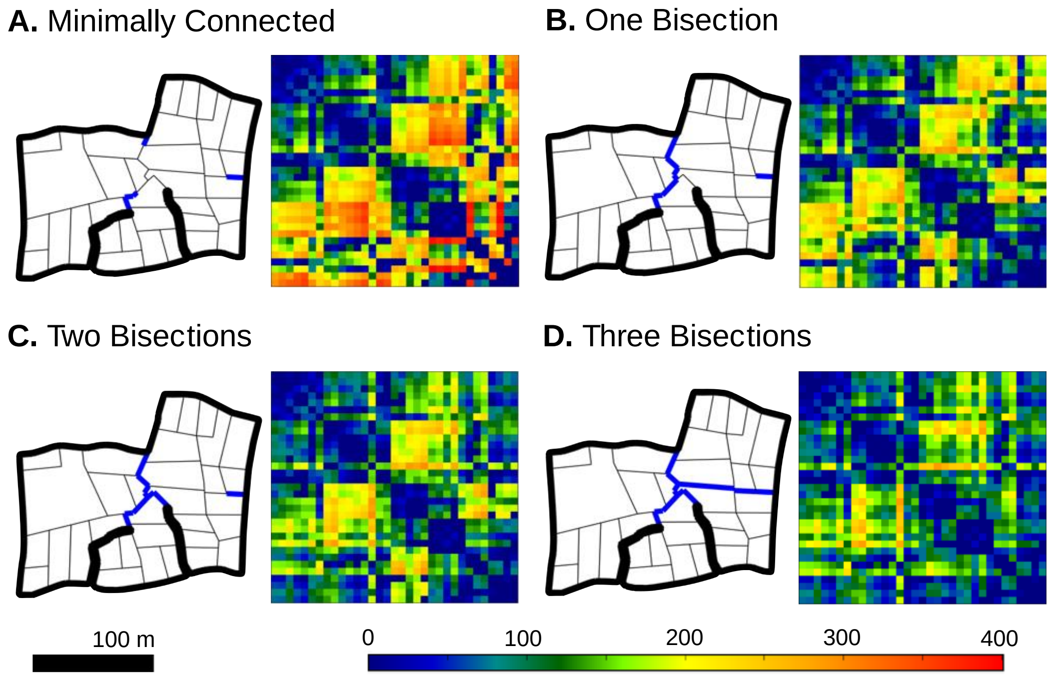


Fig. S7. Epworth (Harare, Zimbabwe): Geometric optimization and travel cost matrices. A. The minimal re-blocking strategy for the Epworth block of Fig. 1D (block complexity k=3). The right panel matrices show travel distance (in meters) between any two parcels in the block after the introduction of the three culs-de-sac (blue lines) that render the block universally accessible. Though all parcels are now accessible by roads, some remain distant to many others (red in the travel cost matrix). Such distances can be reduced through the introduction of more roads (see geometric optimization algorithm, section SG), leading to a non-minimal re-blocking strategy. **B.** shows the effect of connecting two of the cul-de-sacs through block bisection. This leads to a substantial reduction of the average distance between parcels (35 m) and of the maximal distance between parcels, from 401 m to 205 m. Unlike the change in block topology that results from providing access to all parcels, which is discrete, the process of geometric optimization of travel distances is gradual. Thus, this second process of introducing more roads and reducing inter-parcel distances has no clear stopping point and could be continued until all boundaries become roads. **C.** and **D.** show how inter-parcel distances can be further reduced by additional road construction, especially due to extending cul-de-sacs to further bisect the resulting blocks.

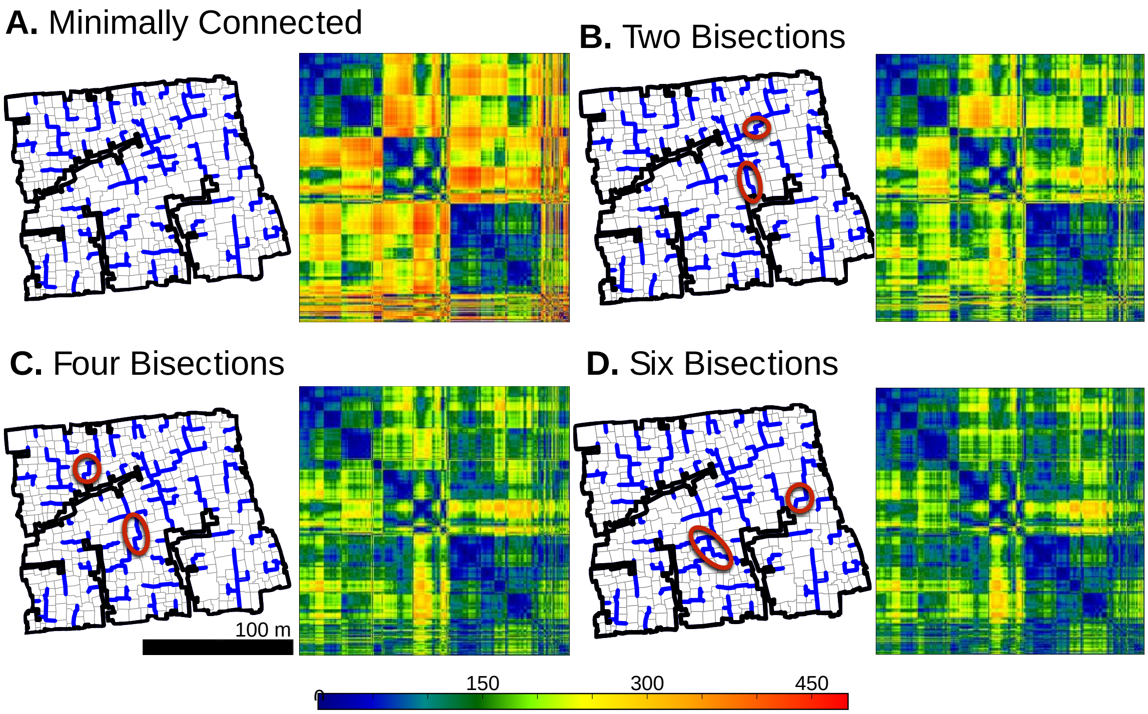


Fig. S8. Khayelitsha (Cape Town, South Africa): Reblocking and travel cost matrices. Mathematical methods become essential to upgrade large and dense slums, as in this example in Khayelitsha (block complexity *k*=9). There are 433 structures identified in this map, of which 258 do not have direct road access, for a population estimate of about 1,400 people. As in Fig. S7, **A.** shows the road geometry and layout for the minimally re-blocked case and the associated parcel to parcel travel cost matrix. **B.** The parcel-to-parcel travel cost matrix when two new road segments are added (circled in red). This leads to a significant reduction in average travel cost (53 m) in response to the addition of 50 meters of new road construction. **C.** and **D.** show how inter-parcel travel distances can be further reduced by further road construction.

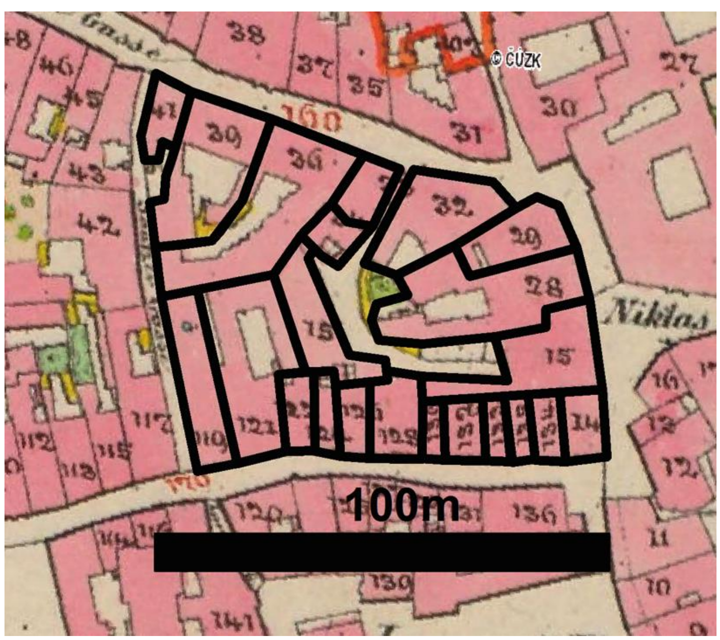
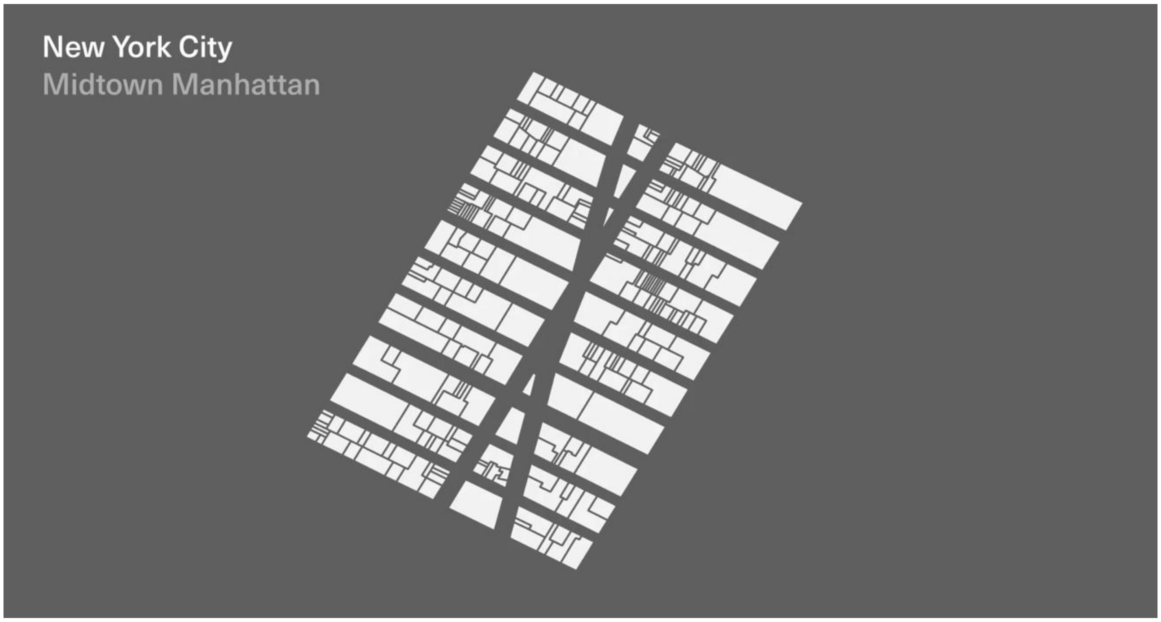


Fig. S9. Prague cadastral map and block parcel layout. The first Stable Cadastral Map of Prague, surveyed between 1817 and 1840, and completed in 1842. Open spaces and courtyards are included with the parcel they most likely border.



Movie S1. Urban topological invariance. This image shows a single frame from the Supplementary Video V1. The video demonstrates the topological transformation between 27 New York City blocks (shown in this image) into a set of blocks from the Summerlin in Las Vegas, NV, and then a set of blocks from Dharavi, a slum in Mumbai, India. The animation shows that sections of any city with the same number of blocks can be seamlessly transformed into each other by deforming each block relative to others. This transformation is similar to other familiar topological equivalences, such as how a mug can be transformed smoothly into a donut in 3D. For cities, the number of blocks is the fundamental element that must be preserved between any two topologically equivalent city street plans. The video was developed in collaboration with Nicholas de Monchaux (College of Environmental Design, UC Berkeley).