Growing Urban Bicycle Networks

Michael Szell*[a,b], Sayat Mimar[c], Tyler Perlman[c], Gourab Ghoshal[c], and Roberta Sinatra[a,b]

[a]*NEtwoRks, Data, and Society (NERDS), IT University of Copenhagen, 2300 Copenhagen, Denmark*
[b]*Complexity Science Hub Vienna, 1080 Vienna, Austria*
[c]*Department of Physics and Astronomy, University of Rochester, Rochester, NY 14627, USA*

Abstract

Cycling is a promising solution to unsustainable car-centric urban transport systems. However, prevailing bicycle network development follows a slow and piecewise process, without taking into account the structural complexity of transportation networks. Here we explore systematically the topological limitations of urban bicycle network development. For 62 cities we study different variations of growing a synthetic bicycle network between an arbitrary set of points routed on the urban street network. We find initially decreasing returns on investment until a critical threshold, posing fundamental consequences to sustainable urban planning: Cities must invest into bicycle networks with the right growth strategy, and persistently, to surpass a critical mass. We also find pronounced overlaps of synthetically grown networks in cities with well-developed existing bicycle networks, showing that our model reflects reality. Growing networks from scratch makes our approach a generally applicable starting point for sustainable urban bicycle network planning with minimal data requirements.

Introduction

Cities worldwide are scrambling for sustainable solutions to their inefficient, car-centric transport systems [1,2]. One promising, time-tested candidate is cycling. It is an efficient mode of sustainable urban transport that can account for the majority of intra-urban trips which are primarily short and medium-distance [3]. Cost-benefit analysis that accounts for health, pollution, and climate, reveals that in the EU alone cycling brings a yearly benefit worth €24 billion while automobility costs society €500 billion [4]. These insights provide further impetus for coordinated efforts to extend cycling infrastructure as one solution to the urban transport crisis and to effectively fight climate change [5,6].

*Corresponding author. Email: misz@itu.dk

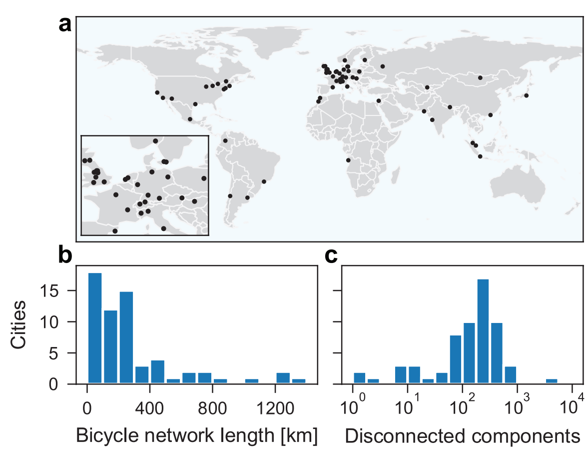

Figure 1: **The state of existing bicycle networks. a** We extract street networks from 62 cities covering different regions and cultures; many are considered modern and well developed. **b** The distribution of city-wide lengths of bicycle tracks indicates negligible existing cycling infrastructure that is also **c** split into hundreds of disconnected components. See more details in Supplementary Table 1.

In practice, however, bicycle infrastructure development struggles with a particularly pervasive political inertia due to "car culture" [7,8]: For example, Copenhagen took 100 years of political struggles to develop a functioning grid of protected on-street bicycle networks [9] that continues to be split into 300 disconnected components today [10]. Accordingly, the most developed, influential bicycle network planning guidelines, such as the Dutch CROW manual [11], acknowledge that building up bicycle networks happens typically through decades-long, piecewise refinements. Unfortunately, given the advanced stage of the planetary climate crisis [12], humanity has run out of time – there is now no other choice left than making urban transport sustainable as fast as possible [1,7,13]. While there has been historical inertia, the ongoing COVID-19 pandemic has prompted several cities to engage in successful accelerated network development, proving that such efforts are indeed possible [14,15]. A systematic exploration of city-wide, comprehensive development strategies is therefore urgently needed.

Although the prevailing, piecewise application of qualitative policy guidelines in existing bicycle network planning [11,16] might have a good track record in e.g. Dutch cities and Copenhagen [9], this process lacks rigorous scrutiny: Are the resulting networks truly optimal? Can such policies be replicated in other cities?

And are there fundamental topological limitations for developing a bicycle network? Indeed, an evidence-based, scientific theory of bicycle network development is missing.

There is a growing academic literature on analyzing bicycle networks of specific cities, for instance Montreal [17], Seattle [18], or recent data-driven approaches for Bogota [19] or London [20]. While such studies are invaluable in terms of local enhancements and data consolidation for a particular place, here we instead focus on a global analysis, in particular on the *fundamental topological limitations* of bicycle network development that are relevant for all urban environments, independent of the availability of traffic flow data [21]. This approach follows the idea of a *Science of Cities* [22] where we study the topological properties of bicycle networks that are *independent of place* using computational, quantitative methods of *Urban Data Science* [23].

The vast majority of cities on the planet has negligible infrastructure for safe cycling [6]. Indeed, urban transport infrastructure development worldwide has been heavily skewed towards automobiles since the 20th century, today featuring well connected networks of streets for motorized vehicles [10]. Rather than uprooting the existing infrastructure and replacing it with an entirely new one—an economically infeasible strategy—we investigate how to retrofit existing streets into bicycle networks. Sacrificing specificity for generalizability, our formulation contains as a starting point two ingredients: the existing street network of a city, and an arbitrary set of points of interests. With these minimal ingredients we explore different growth strategies that sequentially convert streets that were designed for only cars to streets that are safe for cycling [11, 24]. Inspired by the CROW manual [11], the objective of all explored strategies is to create *cohesive* networks, i.e. well connected networks that cover a large fraction of the city area (see Methods).

Across the realistic strategies we report a growth phase that initially leads to a diminishing trend of quality indicators, until a critical fraction of streets are converted, akin to a percolation transition observed in critical phenomena and also present in the growth of other forms of transportation infrastructure as well as patterns of traffic [19, 21, 25–28]. In other words, initial investments into building cycling-friendly infrastructure leads to diminishing returns on quality and efficiency until the emergence of a well-connected giant component. Once this threshold is reached, the quality improves dramatically, to an extent which depends on the specific growth strategy. We provide empirical evidence that the majority of cities effectively lie below the threshold which might be hindering further growth, implying fundamental consequences to sustainable urban planning policy: To be successful, cities must invest into bicycle networks with the right growth strategy, and *persistently*, to surpass a critical mass.

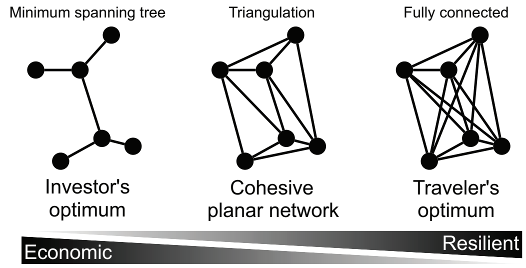

Figure 2: **Optimal connected network solutions.** Adapted from Ref. [29]. (Left) The investor's optimal strategy for a connected network is to invest as little as possible, minimizing total link length [10]. Its solution is a minimum spanning tree, maximally economic but minimally resilient with low directness, inadequate for travelers. (Right) The traveler's optimum connects all node pairs creating all direct routes. This solution is minimally economic, maximally resilient and direct, inadequate for investors. It also has crossing links and is therefore not a planar network. (Center) A both economic and resilient, as well as cohesive planar network solution in-between is the triangulation. In particular the minimum weight triangulation, approximated by the greedy triangulation, minimizes investment.

Results

The starting point for our analysis is the street networks of 62 cities, see Fig. 1(a) and Supplementary Table 1. Here, links represent streets and nodes are street intersections. Being embedded in a metric space, these constitute planar graphs [30]. We downloaded and processed these networks from OpenStreetMap using OSMnx [31] (see Methods).

Although many of the covered cities are from well developed regions, we observe that they have negligible bicycle infrastructure, Fig. 1(b). Additionally these are split into hundreds of disconnected components, Fig. 1(c), which has previously prompted analysis of strategies to merge them [10]. Although such strategies make sense in cities with already well established bicycle infrastructure, they are less useful in most other cities. Further, they produce minimum spanning tree-like solutions that are economically attractive but lack resilience and cohesion (Fig. 2), and they potentially reinforce socioeconomic inequalities by connecting only already developed areas while ignoring under-developed ones [21], prompting us here to grow new networks from scratch instead.

Growing bicycle networks from scratch

Our process of growing synthetic bicycle networks consists of four steps, Fig. 3, starting with the street network and the points of interest. For details see Methods; for an intuitive, interactive exploration see `http://growbike.net`.

Step 1) Points of interest (POIs). An arbitrary set of POIs is snapped to the intersections of the street network. We investigate two versions of POIs: i) arranged on a grid, and ii) rail stations.

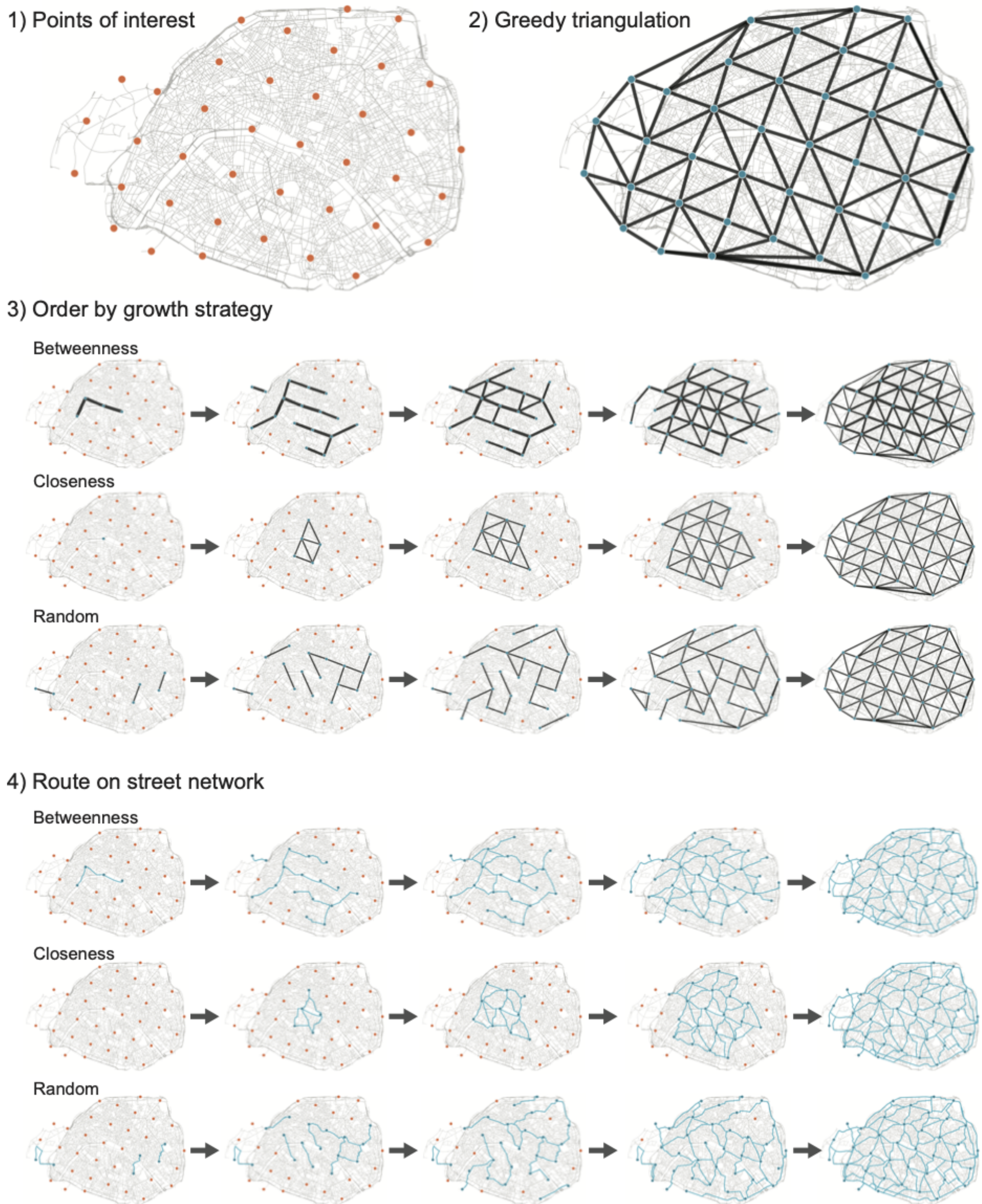

Figure 3: **Growing bicycle networks.** Explorable interactively at: `http://growbike.net`. Illustrated here for Paris. Step 1) Points of interest: A set of points of interests (POI, orange dots) is snapped to the intersections of the street network. Shown are grid POIs, alternatively we investigated rail stations. Step 2) Greedy triangulation: The POIs are ordered by route distance and connected stepwise without link crossings. Reached POIs are colored blue. Step 3) Order by growth strategy: One of three growth strategies (betweenness, closeness, random) is used to order the triangulation links from the strategy's 0-quantile (empty graph) to its 1-quantile (full triangulation), resulting in 40 growth stages. Shown are the five quantiles $q = 0.025, 0.125, 0.25, 0.5, 1$. Step 4) Route on street network: The links in the growth stages are routed on the street network. These synthetic bicycle networks are then analyzed for all 62 cities.

Step 2) Greedy triangulation. All pairs of nodes are ordered by route distance and connected stepwise as the crow flies. A link is added only if it does not cross an existing link. This greedy triangulation is an easily computable proxy for the NP-hard minimum weight triangulation [32]. It creates an approximatively shortest and locally dense planar network [33], and a connected, cohesive, and resilient network solution minimizing investment, therefore satifsying both traveler and investor demands, Fig. 2.

Step 3) Order by growth strategy. Each of three growth strategies is used to order the greedy triangulation links from the strategy's 0-quantile (empty graph) to its 1-quantile (full triangulation), resulting in a sequence of growth stages. To study this growth process in a high enough resolution we split the growth quantiles into 40 parts $q = 0.025, 0.05, \ldots, 0.975, 1$. The three strategies are:

1. **Betweenness** – orders by the number of shortest paths that go through a link. It can be interpreted as the simplest proxy for traffic flow (assuming uniform traffic demand between all pairs of nodes). Thus, growing by betweenness is an approach that aims to prioritize flow.
2. **Closeness** – starts with the "most central" node, i.e. the node that is closest to all other nodes. From this seed, the network is built up by connecting the most central adjacent nodes. This approach is the most local approach possible and leads to a linear expansion of a dense as possible network from the topological city center.
3. **Random** – adds links randomly and is used as a baseline. This strategy is not just a theoretical null model but well resembling how cities build their bicycle networks in practice, as we discuss later.

Step 4) Route on street network. The abstract links in the 40 stages are made concrete: They are routed on the street network. These synthetic bicycle networks are then analyzed for all 62 cities.

Different growth strategies optimize different quality metrics

We measure several network metrics to assess the quality of the synthetically grown networks and to compare them with existing bicycle networks. These metrics are: length L, length of the largest connected component (LCC) L_{LCC}, coverage C, POI coverage C_{POI}, directness of the LCC D, number of connected components Γ, global efficiency E_{glob}, local efficiency E_{loc}. We define coverage as the area of all grown structures endowed with a 500 m-buffer, see light blue areas in Fig 6(b) for an illustration. Directness measures the ratio of euclidian distance versus shortest distance on the network only for the LCC, while global efficiency

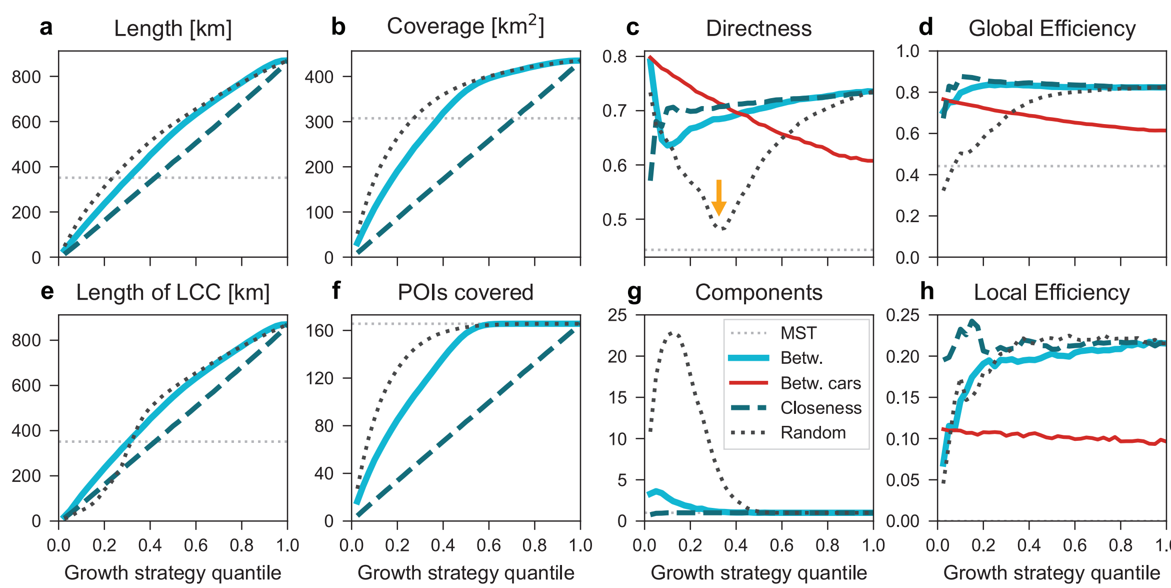

Figure 4: **Different growth strategies optimize different network quality metrics.** The three thick curves show the changes of network metrics with growth following three strategies (betweenness, closeness, or random) averaged over all 62 cities for grid POIs. By construction all curves arrive at the same endpoint, but they develop distinctly before that. For rail POIs and individual cities see Supplementary Figures 4-9. Red curves show the car network's simultaneous decrease of quality metrics if a five times decrease of speed limits is assumed for cars along all affected streets. Grey dotted lines show metrics for the minimum spanning tree (MST) that connects all POIs with minimal investment. Growth of **a** length, **b** coverage, **c** directness, **d** global efficiency, **e** length of LCC, **f** POIs covered, **g** connected components, **h** local efficiency. The yellow arrow highlights the substantial dip in directness until the critical threshold which is more pronounced for random growth than for betweenness growth.

provides a similar but network-wide measure that accounts for disconnected components [34]. See Methods for more details.

We first investigate how these quality metrics change throughout the growth process averaged over all cities, Fig. 4. The three thick curves show the change of the metrics with growth following the three strategies (betweenness, closeness, or random) for grid POIs. Similar results hold for rail stations, see Supplementary Figures 4-6. By construction all curves reach exactly the same point at the 1-quantile (full triangulation), but their development differs substantially before that. The minimum spanning tree (MST) solution is depicted as a baseline (grey dotted lines). This is the most economic connected solution that reaches all POIs, see Fig. 2; therefore any connected solution that reaches all POIs must be at least as long as the MST.

From Fig. 4(a) we observe that length grows linearly for closeness and slightly faster for the other strategies because closeness prioritizes close links which typically have similar length, while betweenness and random growth selects distant links earlier. Random growth adds single links scattered randomly across the city and therefore has the fastest growth of coverage, Fig. 4(b), followed by the betweenness-based strategy, while closeness leads to a linear growth. Directness, Fig. 4(c), displays a large dip for random growth, from $D \approx 0.73$ down to $D \approx 0.48$ at the 0.345-quantile, and a smaller dip from $D \approx 0.79$ to $D \approx 0.63$ for betweenness growth at the 0.1-quantile. Directness starts lower for closeness growth, around $D \approx 0.58$ but quickly overtakes the other strategies at the 0.05-quantile. Global efficiency, Fig. 4(d), starts at a high level, around $E_{\text{glob}} \approx 0.7$, and grows slightly until $E_{\text{glob}} \approx 0.82$ for both betweenness and closeness. Random growth starts instead much lower, around $E_{\text{glob}} \approx 0.33$. The length of LCC, Fig. 4(e), is almost identical as the growth of length for betweenness and closeness because the LCC makes up most of the network here. However, the LCC in random growth has a sigmoid growth pattern as it takes longer for the components to connect, Fig. 4(g). Coverage of POIs, Fig. 4(f), is similar to coverage but more pronounced for random and betweenness growth. On average all POIs are covered before the 0.6-quantile. Finally, local efficiency, Fig. 4(h), is steady around $E_{\text{loc}} \approx 0.22$ for closeness, but grows fast for both betweenness and random growth from around $E_{\text{loc}} \approx 0.05$.

To summarize, the different growth strategies optimize different quality metrics and come with different tradeoffs: 1) Use betweenness growth for fast coverage, intermediate connectedness and directness, and low local efficiency. 2) Use closeness growth for optimal connectedness and local efficiency but slow coverage. 3) Use random growth for fastest coverage but low directness, connectedness, and efficiency.

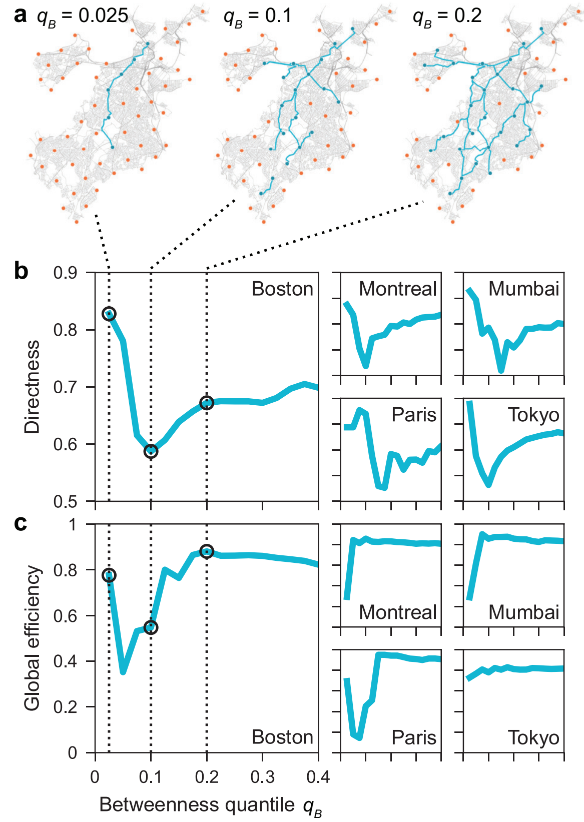

Figure 5: **Network consolidation: Bicycle network growth has a dip of decreasing directness.** **a** Three early stages of betweenness growth in Boston. **b** Directness initially decreases due to the initial largest connected component (LCC) being short (see the example $q_B = 0.025$ for Boston). As the LCC grows in a tree-like fashion, directness decreases to a minimum ($q_B = 0.1$), then consolidates into a single connected component with cycles ($q_B = 0.2$). The process is similar for most other cities as shown here for Montreal, Mumbai, Paris, Tokyo, and also holds for random growth, see Supplementary Figures 4,6,7,9. **c** We find mixed results for global efficiency: Mumbai and Montreal display a single jump, Tokyo is flat, while Boston and Paris shown an initial dip before increasing.

Network consolidation and non-monotonic gains in quality

The dips observed in directness, see yellow arrow in Fig. 4(c) for random growth, are akin to a phase transition in a percolation process from a disconnected set of components to a sudden emergence of a giant connected component, as known for e.g. random Erdős-Rényi networks [25]. Similar transitions have been observed in generalized network growth [35,36], including in various random spatial networks [30] and sidewalk networks [28], and similar flavors of bicycle network growth [19]. Figure 5(a) and (b) illustrates this consolidation process for individual cities: Links are added one by one, growing the largest connected component until a critical threshold at the curve's minimum (at $q_B = 0.1$ for Boston), at which the largest connected component consolidates the majority of the network and starts forming cycles that in turn increase directness. Because connectedness increases around the critical threshold, the evolution of connected components is inverse to the evolution of directness, Fig. 4(g). While the global efficiency averaged over all cities shows an initial increase followed by saturation, see Fig. 4(d), we find mixed trends at the level of individual cities: Mumbai and Montreal track the average trend, Tokyo has a flat global efficiency, while Boston and Paris show a dip before the critical threshold is reached with rapid gains thereafter (Fig. 5(c)).

This network consolidation has important implications for policy and planning. The point at which the transition happens represents substantial investments into building the network. Stopping investments and growth *before* this point leads to a net loss in investment as measured by infrastructure quality. Indeed, pushing past this threshold leads to substantive gains.

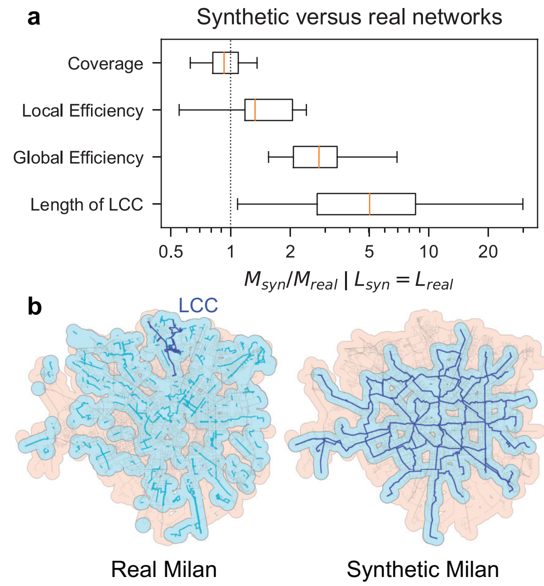

Figure 6: **Synthetic bicycle networks perform several times better than existing ones. a** We plot the distributions (over cities) of the ratios $M_{\text{syn}}/M_{\text{real}}$ between network metrics of synthetic and existing topologies fixed at same length ($L_{\text{syn}} = L_{\text{real}}$), for betweenness growth and grid POIs (for all other growths see Supplementary Figure 2). Synthetic networks have on average 5 times larger LCCs, 3 times the global efficiency, and higher local efficiency. Existing networks only tend to have better coverage because they are more scattered. **b** Illustration of high coverage (light blue area) due to extreme scattering and low length of LCC (dark blue sub-network) for Milan's existing bicycle network, versus its synthetic version at same length (185 km at $q_B = 0.425$). The LCC for synthetic Milan is the whole network.

The effect of bicycle network growth on the street network

While it is beneficial for a city to grow its bicycle network, it is important to ask how this growth affects the network of streets used by cars. The magnitude of this effect depends not only on the network topology, but also on the concrete bicycle infrastructure being implemented: shared spaces, unprotected cycle lanes, protected cycle tracks, bicycle streets, their width, and so on. To consider these factors, leading bicycle planning manuals consider a plethora of local variables [11, 37], for example road category, speed limit, volume of the motorized traffic, or car parking facilities. Therefore, to be conservative in our estimations, here we consider the strongest possible effect of new bicycle infrastructure on streets apart from complete replacement. We assume that along the affected road sections the effective speed limit for cars will be reduced by a factor of 5, i.e. to walking speed. In practice this assumption would be consistent with, for example, all infrastructure being built as a child-friendly "fietsstraat" or living street, i.e. as a shared traffic space where cyclists and pedestrians have priority and cars are tolerated to pass through in walking speed [11].

Given this strong constraint, we find the metric that is most affected is directness. Note that for calculating directness, we modify the affected road section lengths by a factor of 5, so this accounts for not just spatial, but spatio-temporal directness [11]. It decreases approximately linearly with the bicycle network's growth, from $D \approx 0.8$ to $D \approx 0.6$, as the network is grown using the betweenness strategy (red curve in Fig. 4(c)). In other words, in the absence of any bicycle infrastructure, car-routes deviate by around 25% from the Euclidean distance between any origin-destination, whereas once the full bicycle infrastructure is established, this increases to around 66%. At around the 0.4 betweenness quantile, the directness of the bicycle network exceeds that of the car network. The global efficiency decreases from around $E_{\text{glob}} \approx 0.75$ to $E_{\text{glob}} \approx 0.6$, (red curve in Fig. 4(d)), while the local efficiency decreases negligibly from $E_{\text{loc}} \approx 0.11$ to $E_{\text{loc}} \approx 0.10$ (red curve in Fig. 4(h)). Growing the bicycle network has no effect on the length and coverage of the automobile network, given that cars can still access all points on the street network, albeit in a longer time than they would without the bicycle infrastructure. We find almost identical behavior for the closeness and random growth strategies (Supplementary Figure 3).

One of the effects of modifying the street infrastructure is the redistribution of load on the street intersections, measured by the betweenness centrality. It has been shown that while the global distribution of the betweenness centrality remains unchanged due to change in density of streets, the spatial distribution and clustering of the high betweenness nodes tend to change, thus redistributing areas of higher traffic [38]. Two measures to quantify this effect are the spatial clustering and the anisotropy of the high betweenness nodes (see Methods). We find a slight increase (around 5%) in spatial clustering and anisotropy for nodes in the 90th percentile of betweenness values but the effect is marginal (Supplementary Figure 3). Thus bicycle networks do not substantially affect the congestion properties of existing street networks.

Comparing synthetic with existing network metrics

Although the growth processes described here are somewhat artificial, given the lack of accounting for practical limitations of bicycle network design—street width, incline, or political feasibility for instance—it is nevertheless prudent to compare the synthetic network with existing bicycle networks to gauge their general correspondence. To have a fair comparison in terms of length (which is a proxy for cost), we first select all cities that have a protected bicycle network with shorter length L_{real} than the fully grown synthetic network L_{syn} (42 out of 62 cities), and for each of them we fix the growth quantile where the synthetic length is equal to the real length, $L_{\mathrm{syn}} = L_{\mathrm{real}}$. Given this set of bicycle network pairs – real versus synthetic at same length – we then measure the ratio $M_{\mathrm{syn}}/M_{\mathrm{real}}$ between the synthetic quality metric M_{syn} and the quality metric of the existing infrastructure M_{real}. The results for the metrics of coverage, local efficiency, global efficiency, and length of LCC are reported in Fig. 6 (a).

We find that synthetic networks have on average 5 times larger LCCs, 3 times the global efficiency, and higher local efficiency. Existing networks only tend to have better coverage because they are more scattered, as illustrated in Fig. 6 (b) for Milan which has 230 disconnected components. Milan's scattered network provides an important lesson: Mere measures of total length or coverage are misleading when it comes to an efficient and safe infrastructure if the network is not well connected. Instead, if city planners were to develop and implement bicycle networks holistically, considering a city-wide rather than piece-wise local approach, much higher quality infrastructure could be derived to the benefit of the residents.

For completeness we also compare our results to the closeness and random growth approaches, Supplementary Figure 1(c) and (e). For closeness we find almost the same result as for betweenness, only with notably worse coverage which is to be expected given how closeness grows the covered area as slowly as possible. For the random growth approach we find the same coverage as existing infrastructure, and around 2 times the global efficiency and length of LCC. At first blush, this implies that even a naive random growth strategy can perform better than existing ones. However, this could be due to a number of reasons: For instance, in the random growth process described here, segments are added over at least 1.7 km in each step, whereas in real cities, segments are added in a more scattered fashion and in varying lengths. Further, cities can have non-negligible off-street bicycle tracks, for example through parks, a feature not considered in our analysis.

Comparing synthetic with existing network overlaps

To gain a better understanding into how the synthetically grown parts compare to existing infrastructure, and thus the extent to which the growth models approximate reality, we measure the percent overlap of synthetic infrastructure with existing bicycle infrastructure. Figure 7(a) and (b) report for rail station POIs the average overlaps for protected cycle tracks and for bikeable infrastructure respectively, where bikeable infrastructure is defined as the union of protected tracks and streets with speed limits ≤ 30 km/h (see Methods). Results are qualitatively similar for grid POIs, see Supplementary Figures 7–9.

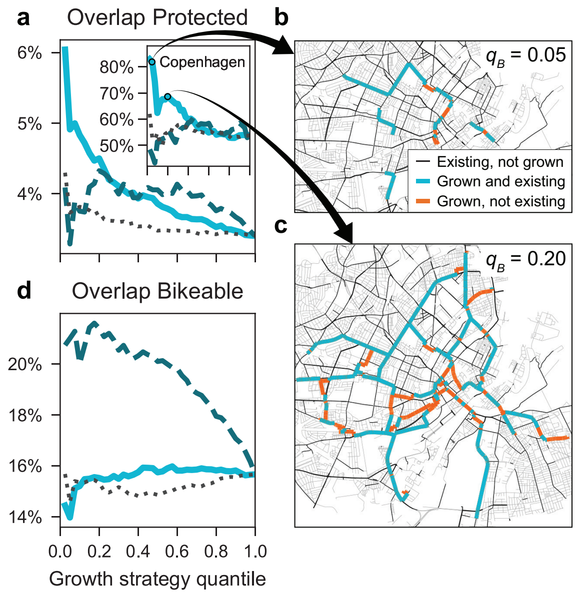

Figure 7: **First stages of synthetic growth recreate existing networks.** Shown are results for rail station POIs averaged over all cities. Same legend as Fig. 4. **a** Growth by betweenness starts with high, then decreasing overlap with existing protected bicycle infrastructure. Inset: The effect is especially strong in cities with well developed on-street bicycle networks such as Copenhagen. Here the growth algorithm starts with over 80% overlap. **b** Map of this high overlap in Copenhagen at the quantiles $q_B = 0.05$ and **c** $q_B = 0.20$. **d** The overlap with bikeable infrastructure has a notable effect only for growth by closeness due to traffic-calmed city centers: With increasing distance from the city center, overlap falls.

For protected infrastructure, Fig. 7(a), we find that growth by betweenness (solid line) starts with around 6% overlap on average, then decreases fast reaching 3.5%. We find a similar behavior for random growth (dotted line) but with a smaller effect. We find no clear effect for growth by closeness (dashed line). These observations suggest that cities take into account flow (betweenness) when building their cycling infrastructure, and that rail stations play some role – otherwise there would be no effect in the random growth. The betweenness overlap effect is especially strong in Copenhagen, Fig. 7(a) inset, which has a well-developed, cohesive on-street bicycle network. Figure 7(b) shows the synthetic growth stage at the second step, $q_B = 0.05$. Remarkably, at this step over 80% of the links suggested by the synthetic network model do already exist in reality. Even at $q_B = 0.20$ there is almost 70% overlap, Fig. 7(c). These values are far higher than expected by chance: Given the length of Copenhagen's on-street cycle tracks we would expect only at most 24% overlap in a random link placement.

The overlap for bikeable infrastructure looks different, Fig. 7(d). Here, there is no clear effect for betweenness and random growth, but a very clear effect for closeness, which starts with high values (above 20% on average), falling slowly to 16%. This observation is consistent with cities preferentially installing low speed limit areas in their city centers.

Discussion

We grew synthetic bicycle networks in 62 cities following three different growth strategies all aiming to generate a cohesive network, i.e. a network that is well connected and covers a large fraction of the city area. Studying the resulting networks we found a consistent critical threshold affecting directness in all cities and global efficiency in some of them, for the two most realistic strategies of growth by betweenness and random growth. This sudden network consolidation has a fundamental policy implication: To be successful, cities must invest into bicycle networks *persistently*, to surpass short-term deficiencies until a critical mass of bicycle infrastructure has been built up. Further, from a topological perspective, cities should avoid traditional "random growth-like" strategies that follow local, stepwise refinement. Such strategies substantially shift the critical threshold, thereby hold up the development of a functional cycling infrastructure which could fuel adversarial objections to bicycle network expansions along the lines of "We already built many bike tracks but nobody is using them, so why build more?" As we have shown, *it is not a network's length that matters but how you grow it.*

By comparing metrics and overlaps of synthetic with existing networks, we gained an insight into the realism of our models. Our observations suggest that growth processes of existing protected bicycle networks contain a strong random ingredient and a detectable consideration for flow (betweenness) and rail stations. The random ingredient can be explained by the traditionally slow-paced urban planning processes arising from political inertia [8, 16]. Unfortunately, the random strategy is also the slowest in terms of network consolidation: It needs at least three times the investments than the betweenness strategy to reach the critical threshold. The rail effect can be explained by transit-oriented development efforts, where bicycle facilities are planned close to transit lines [20, 39]. The remarkably high overlap with Copenhagen's well developed network suggests that our models could also be adapted to identify "missing links" in existing bicycle networks.

Only by being global and minimalistic, deliberately ignoring second order effects, does our approach uncover fundamental topological limitations of bicycle network growth independent of place. At the same time our results must be treated as *statistical solutions.* By no means do they suggest concrete recommendations for new bicycle facilities, as a vast array of local idiosyncrasies (second order effects) would need to be accounted for [11, 37], including: road category, speed limit, volume of motorized traffic, or aspects of comfort [40]. Despite the importance of these aspects, a transport network's geometry is its most fundamental limitation [41] which is the reason we explored it first. Although our approach here is not yet aiming to provide concrete urban design solutions, it could be useful for planning purposes for easily generating an initial vision of a cohesive bicycle network – to be refined subsequently. By publishing all our code as open source we facilitate such future refinements. Our minimal requirements on data are a deliberate limitation we impose for our framework to be applicable to data-scarce environments and thus to a large part of the planet [42]: no lane widths, inclines, traffic flows, etc. are needed to optimize network topology.

Starting from rail station POIs seems a reasonable approach, however care has to be taken to not amplify existing biases that are well-documented in the transport planning profession [43–45]: For example, planning bicycle infrastructure only along metro stations that were built following elitist or racist biases would reinforce them, neglecting under-served regions and their inhabitants even further. The strength of our POI approach lies in its arbitrariness that can bypass such issues: Grid POIs implement equal coverage and could be a starting point, to be refined carefully with e.g. population density or traffic demand models [21, 46]. The biggest limitation of our approach is the sole focus on retrofitting street networks for safe cycling. This approach has some issues because it only considers on-street but no off-street bicycle infrastructure. We discuss the technical details of this limitation, mostly relevant for concrete bicycle network planning in low urban density, in Supplementary Note 1, concluding that future research on bicycle network growth should consider off-street solutions wherever possible.

Finally we discuss the effect of growing bicycle networks on limiting street networks for car traffic. It is unclear whether to interpret a change from directness $D \approx 0.8$ to $D \approx 0.6$ as substantial or inconsequential. We deem it more important to discuss whether a small or a large change is *desired.* There are arguments for both sides: From the perspective of car-dependent transport planning the change should be small to not disrupt the existing system too abruptly [7, 47]. From the perspective of sustainability, human-centric urban planning, and climate research, the change should be large to boost efficient bicycle transport, livable cities, and to fight climate change effectively. Indeed, the CROW manual states that directness should be higher for cyclists than for cars [11]. On top of that it could be argued that our Copenhagen-style model of building a relatively sparse sub-network for cyclists is not going far enough. For example, it could be inverted into a Barcelona-style model where dense patches of living streets – Superblocks – are built within a sparse sub-network of automobile arterials [48]. In any case, resistance to such ideas needs to be anticipated [5,47], and reclaiming inefficiently used urban space for humans requires vigorous policy making following leading examples such as the Netherlands [8,49]. The necessity of doing so is backed by ample evidence from sustainability science, and if implemented persistently it will facilitate the transition to car-free cities to counteract climate change, providing extraordinary benefits to public health and urban livability on the side [2,13,50,51].

Methods

Data acquisition and processing

Infrastructure networks. We downloaded existing street and bicycle networks for 62 cities from OpenStreetMap (OSM) on 2021-02-26 using OSMnx [31]. For each city, three networks were downloaded: Street network, protected bicycle network, bikeable network. Each node is an intersection, each link is a connection between two intersections. A protected bicycle network is the union of all OSM data structures that encode protected bicycle infrastructure, both on-street and off-street. Following the cycling safety literature, we consider only protected bicycle networks in our main analysis because safe cycling in general conditions is only ensured through physical separation from vehicular traffic [6,11,24,37,49]. However, we also consider for additional analysis the "bikeable" network, which is the union of a protected bicycle network and all streets with speed limits $\leq$ 30 km/h or $\leq$ 20 m/h (including living streets). In special conditions such street segments can be considered safe for cycling, but not in general [11,24]. OSM data has been generally found to be of high quality and completeness [52,53], but multicity studies using bicycle infrastructure data such as ours could potentially suffer from some labeling inconsistencies, especially for less common types of bicycle infrastructure [54].

Points of interest (POI). Rail station POIs consist of all railway and metro stations. A few of the considered cities do not have rail stations. For creating grid POIs, we created grid points at a distance of 1707 m, ensuring a tolerable average distance of 167 m (2 minutes walking) over the whole city to the triangulated network in the worst case, see Supplementary Note 2. We then rotated this grid to align it with the city's most common street bearing [55], and snapped the grid points to the closest street network intersections within a 500 m tolerance. The rotation is mostly important for US cities that have a grid-like street network, e.g. Manhattan, for creating straight routes.

Greedy triangulation

The greedy triangulation orders all pairs of nodes by route distance and connects them stepwise as the crow flies. A link is added only if it does not cross an existing link. This triangulation is a $O(N \log N)$ computable proxy for the NP-hard minimum weight triangulation [32] with an approximation ratio of $\Theta(\sqrt{N})$ [56]. The greedy triangulation is fast and solvable for any set of nodes. Computing a quadrangular grid, as suggested by the CROW manual [11], or a quadrangulation, is in general not possible for arbitrary sets of nodes and also computationally less feasible [57].

Network metrics

Cohesion. The CROW manual [11] describes qualitatively what it means for a network to be *cohesive*: a "combination of grid size and interconnection". It states that this is the most elementary requirement for a bicycle network but without a rigorous definition. We interpret this concept as having both high connectedness (few disconnected components) and coverage, see below. A cohesive network should also be resilient, see below, which excludes pathological cases like the minimum spanning tree.

Coverage. We measure spatial coverage of the network as the union of the ε-neighborhoods of all network elements, i.e. a buffer of ε m around all links and nodes. Here we set $\varepsilon = 500$ m together with the grid POI distance, as this implies a theoretical coverage of 100% of the city area for a grid triangulation and an average distance to the network of 167 m, see Supplementary Note 2. In general, a cohesive bicycle network should cover the majority of the city area.

POI coverage. This metric refers to the number of POIs that have been covered by network elements (by the coverage defined above).

Resilience. By resilience we mean a general level of fault-tolerance [58]: A resilient bicycle network should provide an acceptable level of service in the face of faults and challenges to normal operation, for example interruptions due to road works. The removal of a small fraction of links should not have a substantial impact on network metrics.

Directness. The ratio of euclidean distances $d_E(i,j)$ and shortest path distances $d_G(i,j)$ on the network's largest connected component (LCC):

$$D = \frac{\sum_{i \neq j \in LCC} d_E(i,j)}{\sum_{i \neq j \in LCC} d_G(i,j)} \tag{1}$$

Directness is naturally defined only for the LCC because shortest path distances between nodes in disconnected components are by definition infinite. A similar, but network-wide, measure is given by the global efficiency, see below.

Local and Global Efficiency. A network's global efficiency is defined as [34]:

$$E_{\text{glob}} = \frac{\sum_{i \neq j} \frac{1}{d_G(i,j)}}{\sum_{i \neq j} \frac{1}{d_E(i,j)}} \tag{2}$$

A network's local efficiency is defined as the average of global efficiencies $E_{\text{glob}}(i)$ over each node i and its neighbors,

$$E_{\text{loc}} = \frac{1}{N} \sum_{i=1}^{N} E_{\text{glob}}(i) \tag{3}$$

Spatial clustering and Anisotropy. We first specify a threshold θ and identify the N_θ nodes with high betweenness above the θ-th percentile. Then, we compute their spread about their center of mass

$$x_{cm} = \frac{1}{N_\theta} \sum_{i=1}^{N_\theta} x_i$$

where x_i specifies their coordinates, normalizing for comparison across networks of different sizes via

$$C_\theta = \frac{1}{N_\theta \langle X \rangle} \sum_{i=1}^{N_\theta} \|x_i - x_{cm}\| \,, \tag{4}$$

where

$$\langle X \rangle = \frac{1}{N} \sum_{i=1}^{N} \|x_i - x_{cm}\|$$

is the average distance of all nodes in the network to the center of mass of the high betweenness cluster.

Transition between the topological and spatial regimes is quantified by the increasingly isotropic layout of the high betweenness nodes with increasing edge-density. The anisotropy factor is defined by the ratio

$$A_\theta = \frac{\lambda_1}{\lambda_2} \tag{5}$$

where $\lambda_1 \leq \lambda_2$ are the positive eigenvalues of the covariance matrix of the spatial position of the nodes with betweenness above the threshold θ [38].

For the largest 15 cities we calculated these values only at the 0, 0.5, and 1 quantiles of the growth strategies due to computational feasibility. Therefore, Supplementary Figure 3 reports average values over the 47 smallest cities.

Growth strategies

Betweenness centrality. This is a path-based measure that computes the fraction of paths passing through a given node i [59],

$$C_B(i) = \frac{1}{N} \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{6}$$

where σ_{st} is the number of shortest paths going from nodes s to t and $\sigma_{st}(i)$ is the number of these paths that go through i.

Closeness centrality. Measures the total length of the shortest paths from a node i to all other nodes in the network [60],

$$C_C(i) = \frac{N-1}{\sum_{i \neq j} d(i,j)} \tag{7}$$

Code availability

All code used in the research is open-sourced, available at: `https://github.com/mszell/bikenwgrowth`

Data availability

All data used and generated in the research are publicly available at zenodo [61]. Interactively growing networks, plots and video visualizations for all 62 cities can be explored and downloaded at the accompanying visualization platform: `http://growbike.net`

Acknowledgments

We thank Cecilia Laura Kolding Andersen and Morten Lynghede for developing and implementing the visualization platform. We are grateful to Anders Hartman, Anastassia Vybornova, and Laura Alessandretti for helpful discussions. We thank the ITU High-Performance Computing cluster for computing resources and support.

Author Contributions

M.S. designed the study with input from R.S. and G.G. M.S. wrote the manuscript with input from all authors. M.S. acquired and pre-processed the data with input from S.M. and T.P., and performed the simulations. M.S. directed the project, R.S. and G.G. helped supervise the project. M.S. measured the results, aided by S.M. T.P. performed the analysis on grid size and network coverage. All authors discussed the results.

References

[1] Banister, D. *Unsustainable transport: city transport in the new century* (Routledge, 2005).

[2] Nieuwenhuijsen, M. J. & Khreis, H. Car free cities: Pathway to healthy urban living. *Environment international* **94**, 251–262 (2016).

[3] Alessandretti, L., Aslak, U. & Lehmann, S. The scales of human mobility. *Nature* **587**, 402–407 (2020).

[4] Gössling, S., Choi, A., Dekker, K. & Metzler, D. The social cost of automobility, cycling and walking in the European Union. *Ecological Economics* **158**, 65–74 (2019).

[5] Gössling, S. Why cities need to take road space from cars-and how this could be done. *Journal of Urban Design* 1–6 (2020).

[6] Szell, M. Crowdsourced quantification and visualization of urban mobility space inequality. *Urban Planning* **3**, 1–20 (2018).

[7] Mattioli, G., Roberts, C., Steinberger, J. K. & Brown, A. The political economy of car dependence: A systems of provision approach. *Energy Research & Social Science* **66**, 101486 (2020).

[8] Feddes, F., de Lange, M. & te Brömmelstroet, M. Hard work in paradise. The contested making of Amsterdam as a cycling city. In *The Politics of Cycling Infrastructure: Spaces and (In) Equality*, 133 (Policy Press, 2020).

[9] Carstensen, T. A., Olafsson, A. S., Bech, N. M., Poulsen, T. S. & Zhao, C. The spatio-temporal development of Copenhagen's bicycle infrastructure 1912–2013. *Geografisk Tidsskrift-Danish Journal of Geography* **115**, 142–156 (2015).

[10] Natera Orozco, L. G., Battiston, F., Iñiguez, G. & Szell, M. Data-driven strategies for optimal bicycle network growth. *Royal Society Open Science* **7**, 201130 (2020).

[11] CROW. *Design manual for bicycle traffic* (2016).

[12] Ripple, W. *et al.* World scientists' warning of a climate emergency. *BioScience* (2019).

[13] Caiazzo, F., Ashok, A., Waitz, I. A., Yim, S. H. & Barrett, S. R. Air pollution and early deaths in the United States. Part I: Quantifying the impact of major sectors in 2005. *Atmospheric Environment* **79**, 198–208 (2013).

[14] Lovelace, R., Morgan, M., Talbot, J. & Lucas-Smith, M. Methods to prioritise pop-up active transport infrastructure (2020).

[15] Kraus, S. & Koch, N. Provisional COVID-19 infrastructure induces large, rapid increases in cycling. *Proceedings of the National Academy of Sciences* **118** (2021).

[16] Zhao, C., Carstensen, T. A., Nielsen, T. A. S. & Olafsson, A. S. Bicycle-friendly infrastructure planning in Beijing and Copenhagen-between adapting design solutions and learning local planning cultures. *Journal of transport geography* **68**, 149–159 (2018).

[17] Boisjoly, G., Lachapelle, U. & El-Geneidy, A. Bicycle network performance: Assessing the directness of bicycle facilities through connectivity measures, a Montreal, Canada case study. *International journal of sustainable transportation* **14**, 620–634 (2020).

[18] Lowry, M. & Loh, T. H. Quantifying bicycle network connectivity. *Preventive medicine* **95**, S134–S140 (2017).

[19] Olmos, L. E. *et al.* A data science framework for planning the growth of bicycle infrastructures. *Transportation research part C: emerging technologies* **115**, 102640 (2020).

[20] Palominos, N. & Smith, D. A. Identifying and characterising active travel corridors for London in response to COVID-19 using shortest path and streetspace analysis (2020).

[21] Mahfouz, H., Arcaute, E. & Lovelace, R. A road segment prioritization approach for cycling infrastructure. *arXiv preprint arXiv:2105.03712* (2021).

[22] Batty, M. *The New Science of Cities* (MIT Press, 2013).

[23] Resch, B. & Szell, M. Human-centric data science for urban studies. *ISPRS International Journal of Geo-Information* **8** (2019).

[24] Teschke, K. *et al.* Route infrastructure and the risk of injuries to bicyclists: a case-crossover study. *American journal of public health* **102**, 2336–2343 (2012).

[25] Erdős, P. & Rényi, A. On random graphs. *Publicationes Mathematicate* **6**, 290–297 (1959).

[26] Zeng, G. *et al.* Switch between critical percolation modes in city traffic dynamics. *Proceedings of the National Academy of Sciences* **116**, 23 (2019).

[27] Gross, B., Vakhin, D., Buldyrev, S. & Havlin, S. Two transitions in spatial modular networks. *New Journal of Physics* **22**, 053002 (2020).

[28] Rhoads, D., Solé-Ribalta, A., González, M. C. & Borge-Holthoefer, J. Planning for sustainable open streets in pandemic cities. *arXiv preprint arXiv:2009.12548* (2020).

[29] van Nes, R. *Design of multimodal transport networks*. Ph.D. thesis, Civil Engineering. Delft Technical University, Delft (2002).

[30] Barthélemy, M. Spatial networks. *Physics Reports* **499**, 1–101 (2011).

[31] Boeing, G. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* **65**, 126–139 (2017).

[32] Mulzer, W. & Rote, G. Minimum-weight triangulation is np-hard. *Journal of the ACM (JACM)* **55**, 1–29 (2008).

[33] Cardillo, A., Scellato, S., Latora, V. & Porta, S. Structural properties of planar graphs of urban street patterns. *Physical Review E* **73**, 1–7 (2006). 0510162.

[34] Latora, V. & Marchiori, M. Efficient behavior of small-world networks. *Physical Review Letters* **87**, 198701 (2001).

[35] Achlioptas, D., D'Souza, R. M. & Spencer, J. Explosive percolation in random networks. *Science* **323**, 1453–1455 (2009).

[36] Bollobás, B. & Thomason, A. G. Threshold functions. *Combinatorica* **7**, 35–38 (1987).

[37] NACTO. *Urban bikeway design guide* (Island Press, 2014).

[38] Kirkley, A., Barbosa, H., Barthelemy, M. & Ghoshal, G. From the betweenness centrality in street networks to structural invariants in random planar graphs. *Nature Communications* **9**, 2501 (2018).

[39] Ibraeva, A., de Almeida Correia, G. H., Silva, C. & Antunes, A. P. Transit-oriented development: A review of research achievements and challenges. *Transportation Research Part A: Policy and Practice* **132**, 110–130 (2020).

[40] Quercia, D., Schifanella, R. & Aiello, L. M. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media*, 116–125 (2014).

[41] Walker, J. To predict with confidence, plan for freedom. *Journal of Public Transportation* **21**, 12 (2018).

[42] Barrington-Leigh, C. & Millard-Ball, A. The world's user-generated road map is more than 80% complete. *PloS one* **12**, e0180698 (2017).

[43] Bullard, R. D., Johnson, G. S. & Torres, A. O. *Highway robbery: Transportation racism & new routes to equity* (South End Press, 2004).

[44] Hoffmann, M. L. *Bike lanes are white lanes: Bicycle advocacy and urban planning* (U of Nebraska Press, 2016).

[45] Pereira, R. H., Schwanen, T. & Banister, D. Distributive justice and equity in transportation. *Transport reviews* **37**, 170–191 (2017).

[46] Jafino, B. A. An equity-based transport network criticality analysis. *Transportation Research Part A: Policy and Practice* **144**, 204–221 (2021).

[47] Lamb, W. F. *et al.* Discourses of climate delay. *Global Sustainability* **3** (2020).

[48] Nieuwenhuijsen, M. J. Urban and transport planning pathways to carbon neutral, liveable and healthy cities; a review of the current evidence. *Environment international* 105661 (2020).

[49] Schepers, P., Twisk, D., Fishman, E., Fyhri, A. & Jensen, A. The Dutch road to a high level of cycling safety. *Safety science* **92**, 264–273 (2017).

[50] Klanjčić, M., Gauvin, L., Tizzoni, M. & Szell, M. Identifying urban features for vulnerable road user safety in Europe. *SocArXiv:89cyu* (2021).

[51] Prieto Curiel, R., González Ramírez, H., Quiñones Domínguez, M. & Orjuela Mendoza, J. P. A paradox of traffic and extra cars in a city as a collective behaviour. *Royal Society Open Science* **8** (2021).

[52] Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning B: Planning and Design* **37**, 682–703 (2010).

[53] Barrington-Leigh, C. & Millard-Ball, A. The world's user-generated road map is more than 80% complete. *PloS one* **12**, e0180698 (2017).

[54] Ferster, C., Fischer, J., Manaugh, K., Nelson, T. & Winters, M. Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *International Journal of Sustainable Transportation* 1–10 (2019).

[55] Boeing, G. Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science* **4**, 1–19 (2019).

[56] Levcopoulos, C. & Krznaric, D. Quasi-greedy triangulations approximating the minimum weight triangulation. *J. Algorithms* **27**, 303–338 (1998).

[57] Toussaint, G. Quadrangulations of planar sets. In *Workshop on Algorithms and Data Structures*, 218–227 (Springer, 1995).

[58] Zhang, X., Miller-Hooks, E. & Denny, K. Assessing the role of network topology in transportation network resilience. *Journal of Transport Geography* **46**, 35–45 (2015).

[59] Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35 (1977).

[60] Freeman, L. C. Centrality in social networks conceptual clarification. *Social Networks* **1**, 215–239 (1978).

[61] Szell, M. Urban bicycle networks, existing and synthetically grown (2021). URL `https://zenodo.org/record/5083049`.