Generic patterns in the evolution of urban water networks: Evidence from a large Asian city

Elisabeth Krueger, Christopher Klinkhamer, Christian Urich, Xianyuan Zhan, and P. Suresh C. Rao

Phys. Rev. E **95**, 032312 — Published 9 March 2017

DOI: 10.1103/PhysRevE.95.032312

Generic patterns in the evolution of urban water networks: Evidence from a large Asian city

Elisabeth Krueger[1,2*], Christopher Klinkhamer[1], Christian Urich[3], Xianyuan Zhan[1], and P. Suresh C. Rao[1,4]

[1] *Lyles School of Civil Engineering, 550 Stadium Mall Drive, Purdue University, West Lafayette, IN 47907, USA*

[2] *Helmholtz Center for Environmental Research - UFZ, Permoserstr. 15, 04318 Leipzig, Germany*

[3] *Department of Civil Engineering, 23 College Walk, Monash University Victoria 3800, Australia*

[4] *Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA*

*Corresponding Author (elisabethkrueger@purdue.edu)

Received: November 08, 2016, revised: January 23, 2017 & February 2, 2017

ABSTRACT

We examine high-resolution urban infrastructure data using every pipe for the water distribution network (WDN) and sanitary sewer network (SSN) in a large Asian city (≈4 million residents), to explore the structure, as well as the spatial and temporal evolution of these infrastructure networks. Network data were spatially disaggregated into multiple subnets for functional zones to examine intra-city topological differences for WDN and SSN, while time-stamped SSN data were examined to understand network evolution over several decades as the city expanded. Graphs were generated using a dual mapping technique (Hierarchical Intersection Continuity Negotiation - HICN), which emphasizes the functional attributes of these networks. Network graphs for WDN and SSN are characterized by several network topological metrics, and a double Pareto (power-law) model approximates the node-degree distributions of both water infrastructure networks (WDN and SSN), across spatial and hierarchical scales relevant to urban settings, and throughout their temporal evolution over several decades. These results indicate that generic mechanisms govern the networks' evolution, similar to those of scale-free networks found in nature. Deviations from the general topological patterns are indicative of: (1) incomplete establishment of network hierarchies and functional network evolution, (2) capacity for growth (expansion) or densification (e.g., in-fill), and (3) likely network vulnerabilities. We discuss the implications of our findings for the (re-)design of urban infrastructure networks to enhance their resilience to external and internal threats.

Keywords: complex networks, functional dual mapping, double power-law, HICN, water distribution system, sanitary sewers

PACS numbers: 89.75.Fb, 89.75.Da, 89.65.Lm, 89.20.Kk

I. INTRODUCTION

Urban infrastructure networks are designed and planned for each city and as new urban districts are added to suit the city's geography, to meet the demands of the growing urban population for critical services (energy, water, transportation, communication, etc.), and to comply with engineering design constraints based on local regulations. As cities around the world are growing at accelerating pace, it is of considerable interest to investigate how the structure and functions of urban infrastructure networks evolve over time and space. Specifically, what are the topological differences between urban infrastructure networks for water distribution and drainage? How does the network topology change over time as the city grows? How are the impacts of urban design changes and geographical constraints manifested in the spatial organization and the link between network structure and functions? These and related questions motivate our study, which examines high-resolution water infrastructure data for a rapidly growing, large city in Asia confronted with significant water security challenges.

Power-law relationships have been found for the geometries of cities [1–5], as well as for socio-economic metrics of urban areas, such as GDP, income, crime, innovation, etc. [7–10], and other functional attributes, such as traffic [6,11]. Many authors argue that, in comparison to socio-economic, biological or communication networks, urban infrastructure networks, such as roads, tend to show sparse structures with the absence of scale-free topologies [2,6,12]. A limited number of studies have analyzed the structure and function of below-ground urban infrastructure networks, and, to our knowledge, few have analyzed large networks, because such data are not as freely available as above-ground infrastructure [13]. For example, Yazdani and Jeffrey [14] analyzed the geometry of water distribution networks (WDN) of four small cities using a complex networks approach (primal mapping, see below). They found these networks, similar to the roads analyzed by other authors, to be sparse with an absence of degree-based hubs, with node-degrees ranging from 2 to 4 (average=2).

In complex networks analyses of infrastructure networks, using so-called *primal mapping*, nodes are usually conceived as intersections, and the segments crossing at these intersections as links. In contrast to this, *dual mapping* approaches rely on additional information of these infrastructure networks, such as hierarchies, to determine the nodes (pipes) and links (intersections) of a network, embedding it in so-called "information space". By recovering the inherent hierarchy of the network and removing the constraints of primal mapping, dual mapping allows for the hierarchical properties of the networks to emerge, and thus produces more useful information about the functional aspects of the network [15–17]. Kalapala et al. [18] found national road networks analyzed as dual maps for the US, Denmark and England to be scale-invariant. Masucci et al. [19] introduced a refined dual mapping approach, Hierarchical Intersection Continuity Negotiation (HICN), which is based on hierarchies, and was used to analyze the evolution of London's road network [19]. Their analysis showed that, while the entire street networks resulted in a robust lognormal distribution, the node-degree distributions for only the major roads resulted in a truncated double power-law (or double Pareto) distribution, and the road networks analyzed in [15] conform with these patterns.

Here, we investigate the temporal evolution of the sanitary sewer network (SSN) topology over several decades, as well as network topologies across space and functional hierarchies for both, WDN and SSN in a large Asian city. To our best knowledge, this is the first study to explore temporal and hierarchical evolution of urban water infrastructure networks. We find that earlier results found for the topology of *mature* road and sewer networks in a mid-size U.S. city (around 1 million residents; flat topography; temperate climatological setting) [20], and for major road networks in several countries [15,18,19] also apply to the sanitary sewer and water distribution networks of this large Asian city set in a very different geographical setting (arid climate, significant topography). We add several insights on the evolution of water infrastructure networks, on differences and similarities in the topologies of the two types of water infrastructure networks, as well as on the interpretation of deviations from the generic patterns found for urban infrastructure networks.

Our analysis is based on dual mapping of water infrastructure networks, where the pipe diameter, which determines the flow capacity (e.g., designed maximum discharge) of these pipes, is used to assign hierarchies. This mapping based on a functional attribute of the analyzed water networks results in generic patterns across spatial and temporal scales, as the networks grow along with population size and city area. Our analyses show that various topological metrics are determined primarily by network size. The NDD for both types of water networks can be approximated by a Pareto power-law [Eq. 1(a); large, mature networks], or double Pareto power-law distribution [Eq. 1(b); small, immature networks], described by a function in the form:

$$p(k) = ak^{-\gamma} \qquad (1a)$$

$$p(k) = ak^{-\gamma_{trunk}}\, bk^{-\gamma_{tail}} \qquad (1b),$$

for $2<k$, where the exponent, γ, of the trunk for both WDN and SSN converge above a threshold of network size, measured as dual-mapped nodes $N>10^2$. While the generality of power-law scaling of SSN is in agreement with earlier work [15,16,19,20] and extended to WDN in this study, we reveal here, for the first time, that variations in the tail part of the NDD indicate differences in the structure of the networks, their stage of evolution and potential functional vulnerabilities.

These insights about the evolution of water infrastructure networks are highly relevant in terms of: 1) reducing the extent of individual engineering planning necessary for constructing new or extending existing urban water pipe networks. Information about the city size allows the prediction of the topological features of the water infrastructure network (distribution of pipe hierarchies, i.e., diameters, and number of intersections) necessary to efficiently supply its population, because below-ground pipe networks unavoidably result in generic topological features; and 2) offering a simple and inexpensive approach to examine potential vulnerabilities of the networks, based on deviations from the expected topological features. Thus, these findings can have important implications for infrastructure network maintenance, retro-fitting, and (re-) design.

In the following, we first describe (Section II) the data analysis methods we deployed, and discuss data constraints that translate to limitations. In Section III we present topological features of dual-mapped networks for: 1) variously sized subsets of water networks clipped from the whole network, 2) the temporal evolution of the SSN as the city grows over a 47-year period (1969-2015), and 3) pipe networks of different hierarchies incrementally adding smaller diameter pipes to the main water conveyors. We set a particular focus on the double Pareto power-law functions characterizing the node-degree distributions of the networks. We close (Section IV) with a discussion of the practical implications of our analyses in terms of water infrastructure design, spatiotemporal evolution, vulnerabilities, and network resilience.

II. METHODS

A. Dual mapping

Converting a spatial map into a network graph allows topological analysis of the network, by simplifying spatial structures into network relations. In *primal mapping*, each network (e.g., pipe or road) segment is mapped as an edge (e), and the intersections of these segments are mapped as nodes (n), as done for water distribution networks in [21]. Conversely, in *dual mapping*, pipes are generally conceived as nodes, and intersections as edges. We applied the HICN *dual mapping* approach proposed by Masucci et al. [19]. The authors based the HICN method on a hybrid of two dual mapping techniques: the more widely used intersection continuity negotiation (ICN) [15,17,22,23] and the street name approach (SN), which are both described in Porta et al. [15]. ICN uses the geometrical properties of the planar map to derive the nodes of the graph, by merging aligned (straight) road segments across intersections. SN uses the "information space", and merges contiguous road segments into one node, if they have the same street name. Masucci et al. [19] combined these methods, by merging contiguous pipe segments according to the ICN method with a $\pi/2$ threshold (merging road segments that are connected with the convex angle > 90°) for different classes of roads as proposed in the SN method. Instead of using street names as classes of roads, the authors used road hierarchies as classes (motorways, class A roads, class B roads, minor roads), thus introducing the hierarchical element into dual mapping.

We applied the HICN method here to create a dual graph from a spatial map by merging multiple contiguous edge segments and re-defining them as one node if the convex angle between segments is >90°, and the hierarchy (in this case: pipe diameter) is unchanged (given that it does not cross a pipe of larger diameter). This approach recognizes the continuity of a pipe over a multitude of intersections, and organizes the network into functional units based on flow capacity (i.e., pipe diameter, and thus maximum designed flow). In a first step, we extracted the lists of nodes and edges of the primal graph. The edge list contains identifiers, source and target nodes, as well as the pipe diameter for later classification of their hierarchy. In the second step, contiguous pipe segments of the same hierarchy are merged to form a single node, starting from a randomly selected pipe (edge) in the network, and growing it in both directions until the angular threshold is reached, or the pipe hierarchy changes. This procedure is repeated until all pipes (primal edges) in the network are converted into dual nodes. In the final step, dual edges are created where two dual nodes share an intersection.

The benefit of this dual mapping approach over primal mapping is that primal mapping would partition functional pipe units into several edges connected by multiple nodes (intersections), and consequently restrict the topological analysis [e.g., in primal mapping, any node has a maximum of ≈ 4 edges, while a functional pipe unit of high order in the hierarchy (e.g., a main supply pipe) can connect to dozens of lower order pipes (e.g., supply districts or households)]. Figure 1 illustrates the primal mapping versus the dual mapping approach applied here.

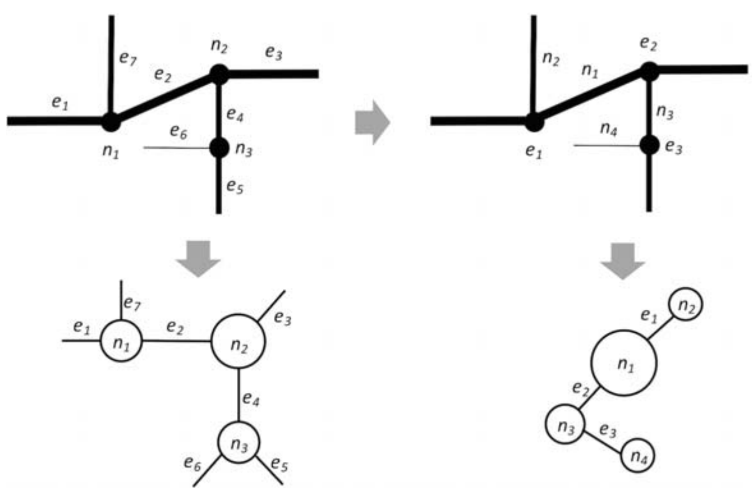

Fig. 1: Schematic of primal versus dual mapping approach applied in this study (left: spatial maps, right: network graphs). Top: primal mapping counts each pipe segment between intersections as edges (e), and the intersections as nodes (n), resulting in e=7, and n=3 in this schematic example. Bottom: Dual mapping creates a node from several pipe segments, which form a functional unit based on unchanged pipe diameter (flow capacity). Intersections connecting different functional pipe units form edges, resulting in e=3 and n=4.

B. Topological analysis of spatial and temporal evolution

Subnets of the water distribution and sewer pipe systems were created in four different ways, and were used to create dual mapped graphs, and to analyze spatial and topological urban metrics:

1) Subnets of WDN and SSN clipped from the entire pipe networks based on water Distribution Zones (DZs). DZs represent functional units for water distribution, which are each equipped with one (or several) water reservoir(s) from where water is supplied to the customers. We used the same DZ boundaries to clip sanitary sewer subnets from the whole network, and analyzed the largest connected component found therein.

2) Functional sewer units were extracted by creating a hierarchical network based on Strahler numbers, as is used for river networks [24–26], and then incrementally removing higher-order sewer pipes from the whole network, which created functional sewer subnets. Strahler numbers (first developed in hydrology by Horton and Strahler [24,25]) are used to assign hierarchies to branches of a mathematical tree, and were first employed to sewers in the generation of virtual drainage networks [27]. In the Strahler Ordering method, the smallest branches (in hydrology: headwater streams) are given a Strahler Order i=1 ("first-order" stream), two converging first-order streams create a second-order stream, and so on. Two converging streams of the same order (i) create a stream of order (i+1), but if a lower-order stream merges with a higher-order stream the number of the higher-order stream remains unchanged. The largest stream in the network has the highest Strahler number.

3) Sewer networks modeled for 10 time steps reproduced the functional sewer network evolution from 1969-2015 (Fig. 2 shows 6 of the 10 time steps). We used time-stamped SSN data in the form of construction year of sewer pipes, and adjustment of replaced pipes to the original installation date. This was done by determining the outlet of the sewer system at the treatment plant and creating sewer-sheds with the help of the bifurcating tree Strahler Ordering method described above, and then adjusting downstream pipe segments' age to the oldest upstream pipe.

4) Networks of different hierarchies were analyzed for different pipe diameters, starting with the largest pipe diameter, and incrementally changing the diameter thresholds to "grow" the network from the skeleton up to the entire network. Figure 3 illustrates the entire WDN and SSN with different pipe hierarchies.

The largest connected component (functional sub-units) for each water sub-network was analyzed, and treated as undirected networks for the analysis of the topological features. We applied this approach to the analysis of the spatial subnets clipped according to functional water DZs, as well as to different pipe diameters.

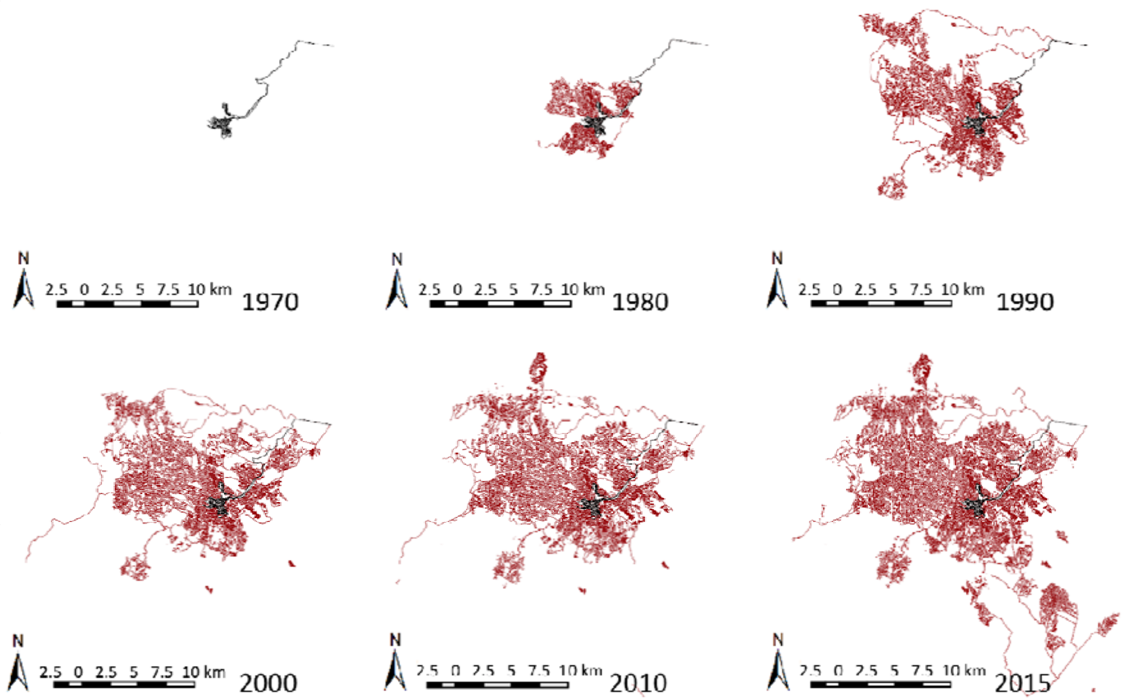

Fig. 2. Temporal evolution of sewer network 1970-2015. 1970 network is highlighted in all time steps. Topological analysis was performed for 10 time steps with results being consistent with those of the functional subnets; data for 6 time steps are shown.

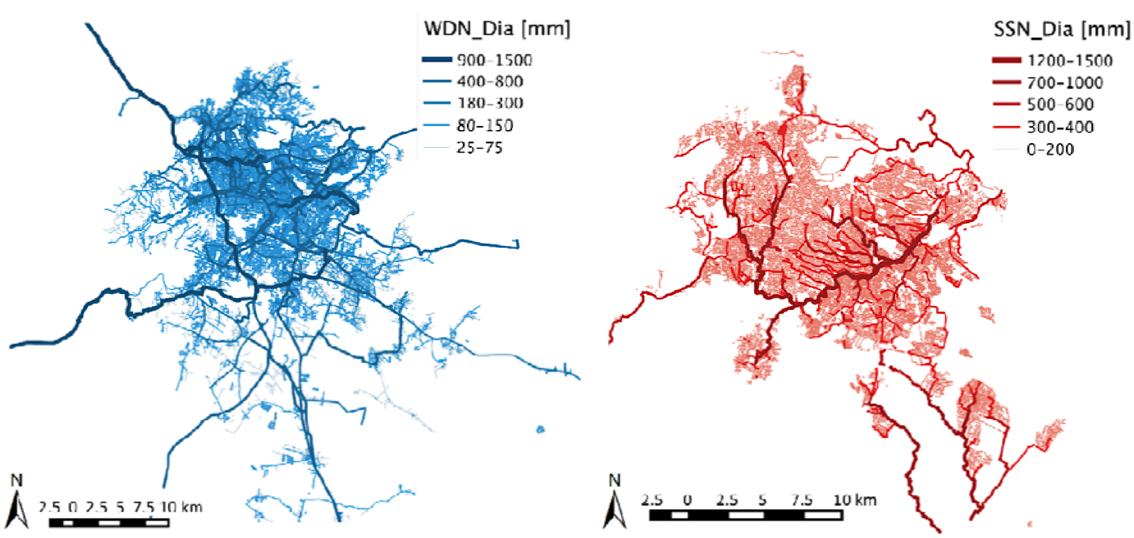

Fig. 3. Pipe hierarchies (diameters) of the entire WDN (left) and SSN (right). Network topologies were analyzed separately for the highest pipe hierarchy, and for networks with pipes added for an incrementally shrinking pipe diameter threshold. Networks shown here are the entire networks "grown" from the “backbone” (largest diameter pipes).

C. Data and analysis limitations

In the extraction of subnets for our analysis, we removed disconnected pipes from the network, and present the results of the network topological analyses by network size, represented by the number of dual mapped nodes. This eliminates a potential bias introduced by the reduction of SSN subnets by the disconnected pipes.

Finite size effects of real-world systems and data limitations challenge the statistically robust estimation of power-law (PL) parameters [28]. Patterns found at small scales can only repeat themselves at larger scales across a limited range of scales, and are subject to subtle changes as scales are changed, thus being limited in resembling the theoretical concept of "scale-free" networks [3]. We recognize these challenges and estimate PL *pdf*s *[p(k), probability distribution functions]* with frontal truncation to account for minimum node-degree and network resolution, and distal truncation to acknowledge finite-size effect. We fit double power-law functions [29] following the guidelines proposed by Clauset et al. [30], and refined by Corral and Deluca [28], using maximum-likelihood estimation and testing for goodness-of-fit for PL to our data. Minimum node-degree for frontal truncation is expected to be 2, representing a single pipe segment, connected at both ends. The generating mechanism (bounded preferential attachment), which produces power-law behavior, would adequately describe the evolution of urban water networks, which lends confidence to our chosen method. This physical generating mechanism has been explored by Carletti et al. [31]. See Appendix A for more information on 1) the methods of network extraction, 2) the algorithm applied for generating dual maps, and 3) for fitting of power-law distribution functions.

III. RESULTS AND DISCUSSION

Investigated total urban area focused on the city's water DZs, and was approximately 623 km^2 in area, with 8,725 km of water distribution pipes. Analyzed sanitary sewer lines (≈5,133 km) served around 80% of the total population in 2015. Data analyzed comprised all water supply and sanitary sewer lines from the source to the street connections (without house connections), and from the street connections to the wastewater treatment plant for the city area, respectively. Subnet creation according to water DZs resulted in subnets ranging in areas from about 1 to 110 km^2, with estimated populations from 56 to 300,272, respectively. Converted into dual mapped graphs, these networks contained between 11 and 4,029 dual nodes for WDN (82 to 33,588 nodes in primal mapping) and between 9 and 8,117 dual nodes for SSN (239 to 83,291 primal nodes). All topological network analyses were performed based on the dual graphs of the water pipe networks.

A. Topological Metrics of Water Networks

Network density is the fraction of links in a graph over the maximum possible number of links, indicating how well connected the nodes are within the network. While for primal mapped planar networks there is an upper boundary for the number of edges a node can have [12] $M \leq 3N\text{-}6$, in dual mapping, network density is defined as:

$$q = [2M/N(N-1)] \quad (2)$$

where N = number of nodes, M = number of edges. Network density values for the analyzed graphs fall within a single PL distribution with an exponent of 0.96, which strongly emphasizes the self-similarity of and homogeneity among the analyzed networks [Fig. 4(a)].

Average node-degree of the analyzed networks fell within a range between 1.8 and 2.5 for all subnets of sizes 10 to 8,117 nodes. With growing network size, the average node-degree increased to around 2.5 for WDN with significant scatter, with a mean of 2.2, while for SSN the average node-degree did not rise significantly above 2.0, and had a mean of 2.0 for all networks [Fig. 4(b)]. This is an interesting result, as an average node-degree of ≈2 (with little variance) is expected for branching trees in primal mapping. In dual mapping, even though the average node-degree remains between 2 and 3, we find a much larger variance than in primal mapping, with few nodes having as many as ≥50 links. This indicates the importance of looking at the shape of the distributions, not only at the mean topological metrics, which we expand on in Section III.B.

The clustering coefficient is the ratio of the number of edges between the neighbors of a node n, and the maximum number of edges that could possibly exist between the neighbors of n. It hence measures the number of triangles in a network. The clustering coefficient of a node is calculated as:

$$CC_n = [2e_n/(k_n(k_n-1))] \quad (3)$$

where k_n is the number of neighbors of n and e_n is the number of connected pairs between all neighbors of n. We calculated the average clustering coefficients for all nodes in each network, which is an indicator of modularity in the network [32]. The low clustering coefficients (<0.1 for all networks >60 nodes, and in 96% of all cases) show that the analyzed networks do not have small-world characteristics and modular organization is weak [33].

Compared with average node-degree the clustering coefficient increases with average node-degree, which may be an indicator of the network forming clusters, in this case in the form of subnets (for SSN), and increasing modularity of WDN as these networks grow. However, clustering was found to be higher in WDN than in SSN, which indicates a more modular structure of the WDN networks compared to the more tree-like structures expected for SSN [Fig.4(c)].

Network centralization indicates whether the network structure is decentralized (network centralization = 0) or star-like (centralization = 1). It is calculated as:

$$C = [N/N-2(max(k)/(N-1)-q)] = [max(k)/N - q] \quad (4)$$

where q=density [34]. Network centralization of the analyzed networks decreases with increasing network size [Fig. 4(d)]. Characteristic path length is the average shortest path connecting any two nodes in a network. This, too, increases with size for all WDN and SSN subnets [Fig. 4(e)].

Clustering and centralization metrics are in accordance with our knowledge of the city's water distribution system, which is organized into gravity-driven distribution zones, as well as the gravity-driven sanitary sewer system resulting in tree-like structures. The hilly terrain of the city makes these gravity-driven systems break up into relatively small natural watershed boundaries, following the undulating shape of the landscape, and hence create a collection of sub-watersheds and sub-sewer-sheds connected toward the inlets and outlets.

The network heterogeneity metric used here is the coefficient of variation of the node-degree distribution, and is defined as the coefficient of variation (CV) [34]. This metric reflects the tendency of a network to contain hubs. Network heterogeneity was between 0.5 and 1.5 for most analyzed networks [Fig. 4(f)], while a significant occurrence of hubs was found for networks with higher heterogeneity values (>1.5), which is in line with expected heavy-tailed distributions for CV>>1.

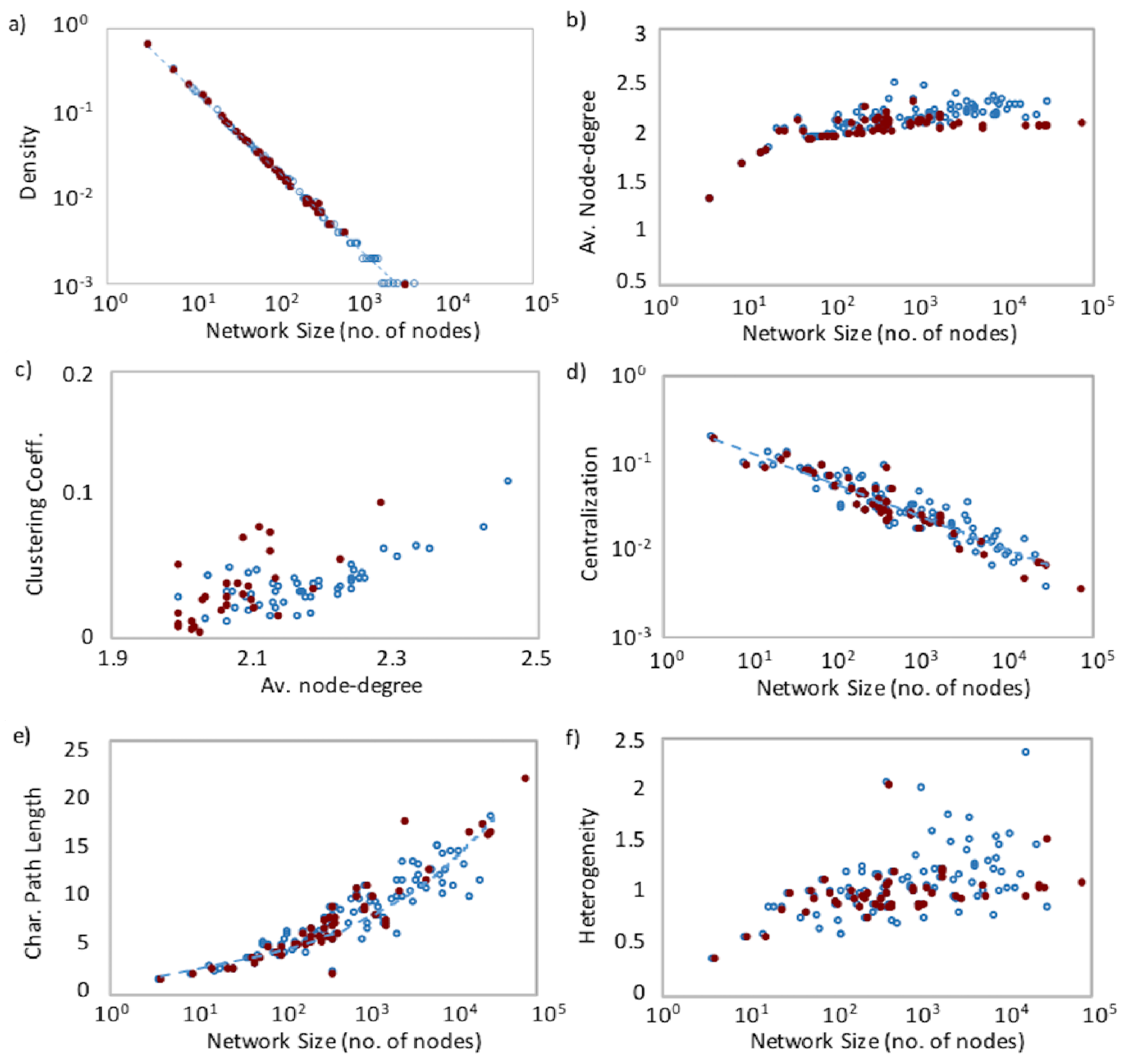

Fig. 4. Topological network metrics for all (110) analyzed subnets, including the entire networks for various time steps, and hierarchical subnetworks. WDN (blue circles) and SSN (red dots): (a) Network density follows a PL distribution ($y = 1.76x^{-0.96}$, $R^2=0.995$); (b) Average number of neighbors (av. node-degree) increases for small network sizes, and converges to ≈ 2 for SSN, and ≈ 2.3 for WDN); c) Clustering coefficient versus average node-degree; d) centralization follows a power law ($y = 2.03x^{-0.62}$, $R^2=0.85$); e) characteristic path length increases in the form: $y = 1.38x^{0.31}$, $R^2=0.83$); f) network heterogeneity.

B. Node-degree distributions

Besides these network topological metrics, we analyzed the node-degree distributions, *p(k)*, for each subnet, and find that dual-mapped infrastructure networks for both water distribution and sewer networks of various sizes, hierarchies and ages follow a truncated (double) power-law distribution. While for small networks (< 120 nodes) fitting a model to the empirical NDD *[p(k)]* delivered unreliable estimates, for larger networks we fitted double power-law functions, as well as exponentially truncated power-law functions to the data. We determined the breaking points between trunk and tail of the double Pareto power-law distributions by using the method introduced by [30], and fitted a truncated power-law function to the "trunk" segment ($k \geq 2$)

and the "tail" segment ($k > k_{break}$) of the distributions, respectively (see Fig. 5(a)). A threshold size of 120 nodes was set, as γ_{trunk} values below that strongly fluctuated, which is a result of fitting to too few data points for a network that is not fully developed. Two outliers (DZs 6 and 25) of the WDN subnets [see Fig. 5 (c)], with network sizes n = 511, and n = 735 dual nodes, respectively, also resulted in unsatisfactory results, when trying to fit a function to the NDD. The mean PL exponent for analyzed networks above 120 dual-mapped nodes was γ_{trunk} = 2.53 ±0.25 for WDN, and γ_{trunk} = 2.41 ±0.30 for SSN. For the tail of the distributions, we found γ_{tail} = 1.35 ±0.40 for WDN, and γ_{tail} = 1.45 ±0.55 for SSN [*p*-values from the KS-statistic for power-law fits ranged from 0.15-0.99 with a mean of 0.70].

Four major findings can be derived from these results:

1) The trunks of the distributions for large and mature networks converge at γ_{trunk} = 2.45 ±0.27 for network size ≥200 dual-mapped nodes [Fig. 5(c)], which emphasizes the generic patterns of these networks in spite of their geometric differences. This value is in the same range reported for sewer networks by Klinkhamer et al. [20].

2) The tails exhibit noise, and tails are reduced as the networks grow and mature. The noise can be explained by an imperfect process of preferential attachment that is limited at the local scale, as elaborated by Carletti et al. [31], because in real-world cases, information about the entire network is incomplete or spatial restrictions do not allow perfect preferential attachment. Carletti et al. [31] found that this partial information leads to an exponential tail, as opposed to a power-law tail, but that the power-law behavior is preserved over a finite small range of node-degrees. The partial information model of network growth [31] translates to constraints for link formation, in our case, spatial or design constraints for the attachment of water pipes. We fitted both, double Pareto, as well as exponentially truncated power-laws, and found that the former resulted in better fits. Based on our findings and according to the model presented by Carletti et al. [31], evolution of the water infrastructure networks analyzed here leads to convergence of the *pdfs* from the trunk towards the tail, as pipes are added to the network, and the tail part of the distribution is reduced, hence reducing the noise in the overall distribution. This is reflected by the increasing breaking points between the trunk and tail distributions [Fig. 5(d)]. High-degree, low-probability pipes form the backbone of the system. As the networks mature and more districts/households are connected to the networks by preferential attachment, the *pdfs* of NDD become more evidently (single and truncated) power-law [Fig. 5(a), Fig. 6(d)].

3) Both types of networks, WDN and SSN, produced surprisingly similar results [see Fig. 5(b)], in spite of their differences in pipe layouts. This could be explained by the organization of the city's water distribution system into multiple water DZs, each equipped with one or more water reservoirs from where the water is distributed to customers by gravity. As such, this WDN functions more as tree-like structures with "reversed" flows (DZs with a single source to multiple destinations), as compared to SSN (multiple sources to single (few) destination). Thus,

loops seem to a limited functional role in this WDN. In other WDN where pressure distribution and flow directions vary with demand (load) variations, loops play a more important role.

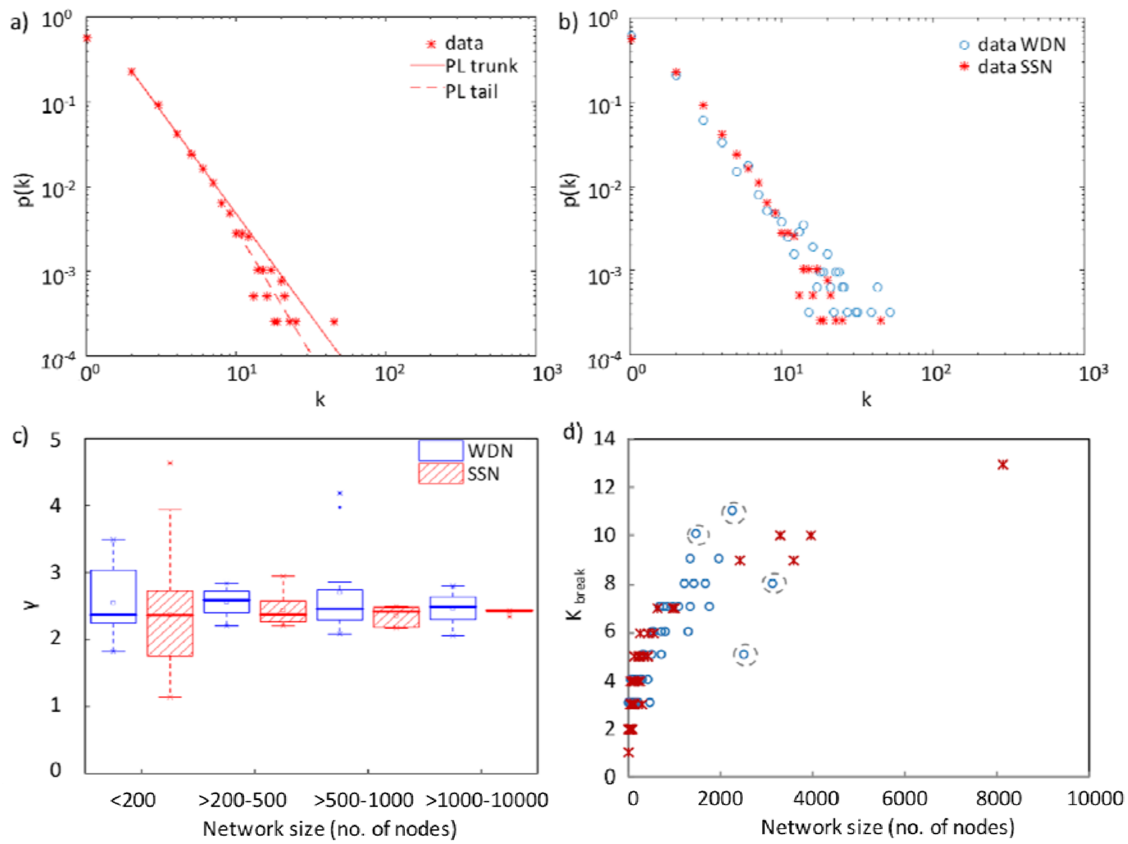

Fig. 5. Characteristics of node-degree distributions: (a) NDD of SSN in 2005 (3,938 dual-mapped nodes, 52,675 primal nodes) follows a (double) PL function with breaking point k_{break}= *10*, $p(k \geq 2)$ = 1.22$k^{-2.41}$ for the trunk, and $p(k > k_{break})$ = 2.74$k^{-2.95}$ for the tail; (b) Comparison of NDD of similarly sized subnets of WDN (DZ10, circles) and SSN (2005 network, asterisks) highlights the similarity of topologies among SSN and WDN; (c) Box plot showing heteroscedasticity of PL exponents for the trunks of WDN (hollow) and SSN (dashed) across the full range of subnet sizes; [mean (small squares), median (thick line), interquartile range (box), [25-75th-percentile ±(1.5* Interquartile Range)] (whiskers), outliers (diamonds)]. Υ_{trunk} converges at γ_{trunk} = 2.45 ±0.27 for network size ≥200 dual-mapped nodes (mean of γ_{trunk} = 2.53 for WDN, and γ_{trunk} = 2.41 for SSN), except for two WDN outliers (see text); (d) Breaking points between the two power-laws of NDD for WDN (blue circles) and SSN (red asterisks). Outliers are discussed in the text and shown in Fig. 6-8.

4) However, WDN had larger divergence between the scaling parameters of the trunk and the tail, than SSN, indicating differences in the hierarchical topologies between WDN and SSN (Υ_{trunk} = 2.53 ±0.25, Υ_{tail} = 1.35 ±0.40 for WDN, and Υ_{trunk} = 2.41 ±0.30, Υ_{tail} = 1.45 ±0.55 for SSN). The relatively flatter tail of WDN could be attributed to 1) redundancy of critical distribution lines, increasing the probability of high node-degree pipes as compared to SSN, and 2) potential for network growth (relative "overdesign" of supply pipes to allow for network growth, i.e., potential for adding lower-degree and terminal nodes, corresponding to street and house connections serving a limited number of customers). Table 1 summarizes the values discussed above for different network size groups.

Table 1: Summary of the results characterizing the NDD of WDN and SSN subnets and SSN temporal evolution. Displayed values are mean values for the respective size group.

No. of nodes	$<k_{break}>$		$<k_{max}>$		$<\gamma_{trunk}>$		$<\gamma_{tail}>$	
	WDN	SSN	WDN	SSN	WDN	SSN	WDN	SSN
>120-200	5.2	3.2	16.0	13.4	2.69	2.43	1.80	1.54
>200-500	5.3	4.8	30.0	18.8	2.64	2.44	1.70	1.91
>500-1000	7.6	6.8	31.8	20.3	2.43	2.33	2.05	2.25
>1000	10.4	10.2	41.4	36.8	2.46	2.41	2.12	2.53

We further explored this by examining the change in the breaking point (k_{break}) between the trunk and the tail segments of the node-degree distribution, and the consequential convergence of the trunk and tail for a given network. We chose two WDN subnets with $>10^3$ dual-mapped nodes with low k_{break}, which fall outside the trend, and two WDN with $k_{break} \geq 10$, highlighted in Fig. 5 (d), (dashed circles; DZs 11, 10, and 01, 32, respectively). As can be seen from Fig. 5, k_{break} increases and finally disappears, the slopes of the trunk and the tail of the distributions converge, and the hierarchies of the networks become more established [Fig. 6 (a-d)]. For the outliers falling well below the k_{break} trend, we can observe much flatter tails [DZs 11, 10; Fig. 6 (a, b)] compared to other subnets [DZs 01, 32; Fig. 6 (c, d)]. Discussions with the city's water utility indicate that these deviations might in fact be an indicator of network evolution in terms of providing network growth potential. The selected subnets with significantly lower k_{break} values and flatter PL tails were stated to contain capacity for network growth or expansion.

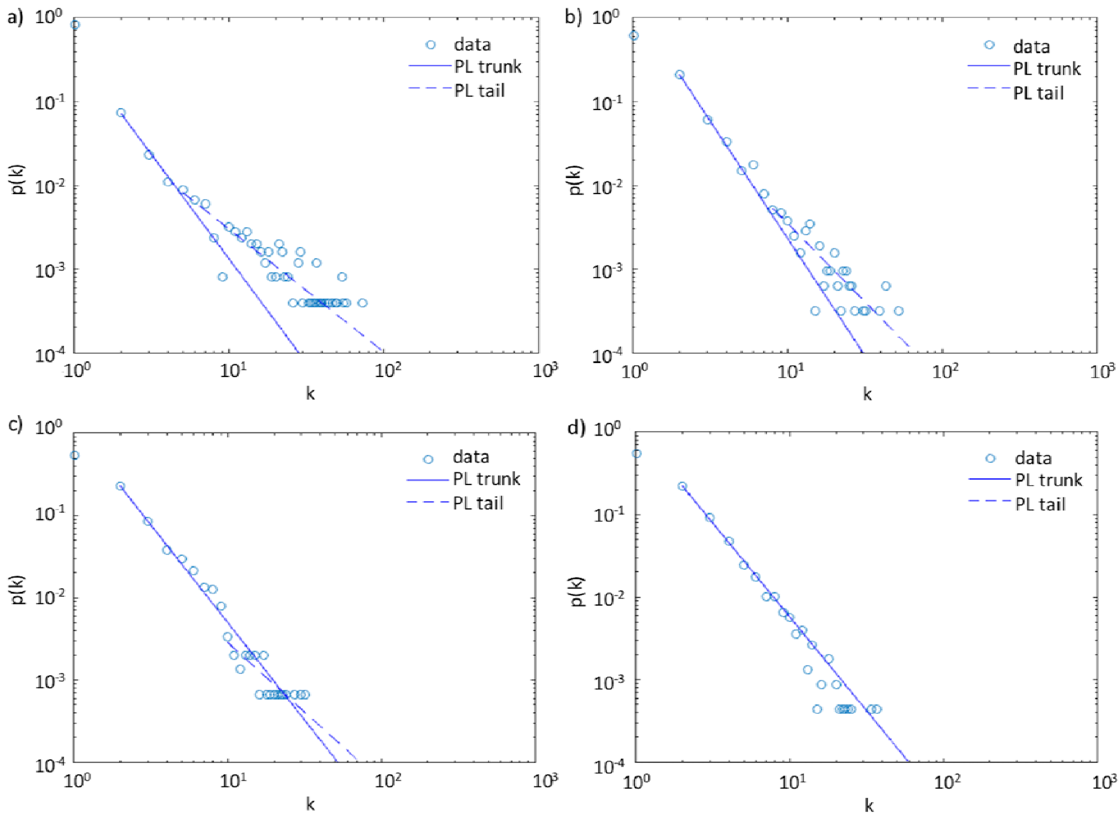

Fig. 6. WDN subnets along a gradient of breaking points between the power-laws of trunk and tail (k_{break} outliers from Fig. 6d): (a) DZ11: k_{break}= 5, $p(k)_{trunk}$= 0.40$k^{-2.49}$ and $p(k)_{tail}$= 0.09$k^{-1.47}$, n=2,425 (dual-mapped nodes); (b) DZ10: k_{break}= 8, $p(k)_{trunk}$= 1.46$k^{-2.80}$ and $p(k)_{tail}$= 0.26$k^{-1.88}$, n=3,179; (c) DZ01: k_{break}= 10, $p(k)_{trunk}$= 1.17$k^{-2.37}$ and $p(k)_{tail}$= 0.138$k^{-1.686}$, n=1,497; (d) DZ32: In this subnet power-law distributions of trunk and tail converge as the breaking point between the two power-laws increases (k_{break}= *20*), and *p(k)* = 1.07$k^{-2.27}$, (n=2,271).

The subnets shown in Fig. 6 (a) and (d) are shown in Fig. 7 (a) and (b) as network graphs, and as spatial maps in Fig. 8 (a) and (b), respectively, to allow for visual inspection of the differences in network structures. The small k_{break} value and flat, scattered tails found for DZ11 in Fig. 6 (a) indicate a significant hub-spoke structure [Fig. 7 (a)], while larger k_{break} values or distributions with converged trunk and tail found for DZ32 in Fig. 6 (d) show more regular network patterns indicative of mature networks [Fig. 7 (d)]. The network heterogeneity (h) also indicates the hub-spoke structure with DZ11 (Fig. 6a) having large network heterogeneity (h = 2.34). The existence of hubs for a given network size would emphasize the tail of a power-law distribution, as relatively more nodes with a higher number of links could be found in such a network, shifting these nodes towards the tail end of the distribution. The spatial maps do not seem to reveal these structural features (Fig. 8).

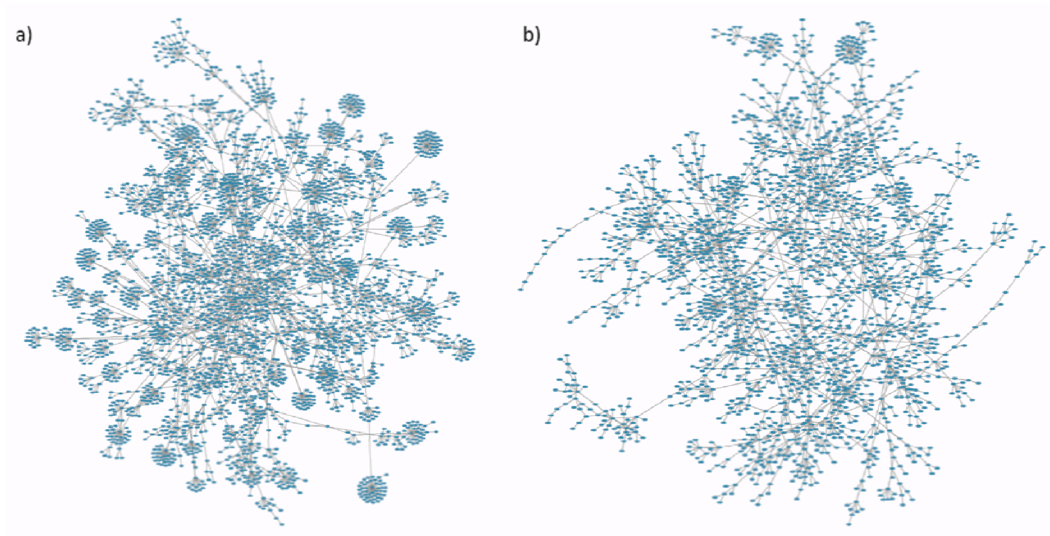

Fig. 7. WDN graphs of selected sub-networks: a) DZ11: tendency to contain high node-degree hubs, heterogeneity=2.34; b) DZ32: h=1.58.

Fig. 8. Spatial maps of the selected water distribution sub-networks (a: DZ11, b: DZ32).

The results presented above add another element to the power-law relationships found for the geometries of cities [1–5], as well as for socio-economic metrics of urban areas [7–10], and other functional attributes, such as traffic [6,11]. Adding to the topological investigations of the urban water networks, we also analyzed the patterns of the urban space occupied by these structures, the temporal evolution of population in comparison to SSN growth, and the economies of scale of the infrastructure networks. Interested readers can find the results of these analyses in Appendix B.

IV. IMPLICATIONS

Our analysis of functionally sampled subnets, temporal evolution of SSN over almost five decades, as well as hierarchical subnets from large to small diameter pipes produced highly consistent results, showing the dominant dependence of several topological metrics on network size, and convergence of γ_{trunk} values for WDN and SSN for N >200 nodes. We find the topological metrics of SSN to be stable over time, based on the temporal evolution of SSN over a 50-year period.

We identified a dominance of hub-spoke structures for deviations of k_{break} towards smaller values, as well as large heterogeneity values. We examined whether any topological changes could be observed for the evolution of the "skeleton" of our networks, which we assessed by stripping the networks from small-diameter pipes, then incrementally adding smaller diameter pipes and analyzing the resulting networks at each step. Again, in line with aforementioned results, the networks resulting from this procedure perfectly fitted into the general patterns found in our analysis, and indications of network "maturation" over time were not evident in SSN.

We conclude that the functional (dual mapped) topology of planned urban infrastructure networks starts out similar to that of river networks draining natural landscapes, where the "backbone" of the system is laid down early in its evolution, showing power-law characteristics from the beginning [35,36]. Of course, river networks evolve under natural forcing and over geologic time scales (making the temporal analysis of their evolution a challenge), orders of magnitude longer compared to urban infrastructure networks that are designed, built and maintained to provide specific urban services. Even when spatial maps of infrastructure networks appear to be random or grid-like [5], we observe that power-law functional traits characterize these networks.

The generality of our findings in terms of topological metrics for the two types of water infrastructure networks was surprising to us. We had expected to find 1) network topological indicators to change with evolution over time; and 2) different types of networks to have stronger differences in network topology, due to the differences in their functions and design. Instead, differences in network layout and design, particularly for WDN were evident in deviations from the respective k_{break} values, as well as network heterogeneity. Given the overall consistency of the results, it is these differences that bear the most interesting information for interpreting network structures. Discussions with the city's water utility indicate that these deviations might in fact be an indicator of network evolution in terms of providing network growth potential. The selected subnets with significantly lower k_{break} values and flatter PL tails were stated to contain capacity for network growth or expansion. According to the water utility, it is in these subnets that most of the network failures have occurred, hence bearing the highest vulnerabilities. These findings provide further support for the relevance of our findings for an efficient planning of new water pipe networks, or existing networks to be retro-fitted, as well as for the assessment of potential vulnerabilities of the networks based on deviations from the expected topological features.

The Asian city we examined here has a geographic setting with large elevation differences within the city set in a hilly terrain, and desert-like conditions and water scarcity force the water utility to run a rationed water supply schedule. In contrast, the U.S. city analyzed in [20] has a flat topography set in a temperate region, and continuous water supply. In spite of these differences in topography, climate, and water management, all of the analyzed infrastructure networks show similar patterns of Pareto power-law node-degree distributions both above ground (roads) and below ground (sewers, water distribution networks).

These findings point to generic mechanisms shaping urban infrastructure networks above *and* below ground. Further analyses of water infrastructure data are warranted to establish consistency among diverse cities in terms of size, age, water management, and geographic settings. Such evidence can contribute to establishing new concepts for resilient urban design and retro-fitting of degrading infrastructure networks subject to dynamic demands, as well as for targeted intervention into these structures, in order to maintain the resilience and reliability of critical urban services.

ACKNOWLEDGEMENTS

The authors acknowledge the local national authorities for their support, and the local water utility, who generously provided the infrastructure data used in this study.

Research reported here was initiated during the Synthesis Workshop "Dynamics of Structure and Functions of Complex Networks" hosted by Korea University, Seoul, South Korea. We would like to thank the organizers, mentors, and participants of this workshop for fruitful discussions and exchange of ideas, in particularly the host of the workshop, Prof. Kyungrock Paik.

The authors thank Ms. Soohyun Yang and Mr. Leonardo Bertassello for their support with coding.

Financial support for the lead author (EK) was furnished by Helmholtz Center for Environmental Research - UFZ, Germany and from a Lynn Fellowship awarded by ESE-IGP at Purdue University (Ecological Sciences and Engineering Interdisciplinary Graduate Program). Financial support to the last author (PSCR) was furnished by the Lee A. Rieth Endowment in the Lyles School of Civil Engineering, Purdue University.

This work was also supported by NSF Award Number 1441188 (Collaborative Research - RIPS Type 2: Resilience Simulation for Water, Power and Road Networks).

Constructive comments provided by an anonymous reviewer have helped improve the manuscript and are greatly appreciated.

The data used here are subject to security constraints and cannot be made available publicly. However, the authors are committed to act as the liaison to the data provider and work with those, who wish to work with the data.

APPENDIX A

Network extraction: The subnets analyzed here and extracted from ESRI shape files contained varying numbers of components and fractions of disconnected pipes (largest connected components ranging from 99 to 70 % of total nodes for WDN, and down to 30 % for sewer networks), which is partly due to imprecise mapping. Water DZ outlines were used to extract sewer subnets from the whole network. This sampling of sewer subnets resulted in a higher number of disconnected pipes and components, and hence reduction of subnet sizes. We considered extending the disconnected lines using a GIS extension, snapping or integration tool, but gap sizes were large and could have resulted in pipe links that are not in place in reality. We analyzed larger functional SSN for the temporal evolution of SSN, for pipe hierarchies, and with functional subnets using the Strahler Ordering method, which allowed us to compare a wide range of network sizes for both WDN and SSN. In addition, we presented the results of the network topological analyses by network size, represented by the number of dual mapped nodes. This eliminates a potential bias introduced by the reduction of SSN subnets by the disconnected pipes.

Dual mapping: Caution should be used, as the dual mapping approach used here can introduce some artefactual bias: our procedure chooses a random pipe segment and grows it in both directions to merge the segments into a dual node. While pipes selected early are more likely to have a higher degree, a pipe selected later will have fewer segments left for it to grow, and thus result in lower degree. Therefore, the process may result in some artificial hierarchy. However, because we are using pipe diameter as hierarchical classes, this effect should be minimal.

Topological analysis: Fitting power-laws to dual-mapped (HICN) node-degree distributions [*p(k); probability distribution functions, pdfs*] for urban infrastructure network data faces constraints related to data availability and also limitations of network data range: (1) urban agglomerations are usually ≤10^3 km^2, causing a "finite-size effect"; (2) even at the highest resolution available, total number of primal nodes are ≈10^4; and (3) dual-mapped maximum node-degree is in the order of ≤10^2. Thus, network data available do not cover multiple orders of magnitude to test for "pure" power-law *pdf*s. Given these constraints, statistically robust estimation of PL parameters is difficult [28]. These challenges become more apparent in our analyses when water network data for different sized subnets are analyzed for comparison, or when network growth over time is examined.

Our analysis recognizes these challenges and estimates PL *pdf*s with frontal truncation to account for minimum node-degree and network resolution, and distal truncation to acknowledge finite-size effect. We fit double power-law functions [29] following the guidelines proposed by [30] and refined by [28], using maximum-likelihood estimation and testing for goodness-of-fit for PL to our data. Minimum node-degree for frontal truncation is expected to be 2, representing a single pipe segment, connected at both ends. We chose this frontal truncation, because we are analyzing networks *without* house connections, and thus terminal nodes with $k=1$ occurring in the networks analyzed here have a lower probability than the house connections (or even higher resolution data, i.e. water pipes within each house) would have. However, PL functions also produced statistically robust results when fitted across all k, but caused a slight change in the exponent. Choice of truncation therefore needs to balance 1) choice of truncation for a more accurate fitting of slope to account for missing data, and 2) recognition of the fact that a frontally truncated power-law ignores a large portion of the data. Consistence in the method is critical for a comparison of the data.

We lend confidence to the suitability of fitting power-law functions to our data, as the generating mechanism (bounded preferential attachment), which produces power-law behavior, would adequately describe the evolution of urban water networks. This physical generating mechanism has been explored by Carletti et al. [31].

APPENDIX B

We also investigated the patterns of the space occupied by the infrastructure networks analyzed in the main part of this paper, the temporal evolution of population in comparison to SSN growth, and the economies of scale of the infrastructure networks.

The sizes of the districts, which are comprised within the water distribution zones, and the population within these districts can both be approximated by power-law probability distributions. The length of water pipes required to service each customer within the city also approximately follows a power-law (Fig. A-1). The latter is consistent with Maurer et al. [37], who found power-law economies of scale (sewer pipe length versus population) in a study of combined sewer systems for a Swiss case study.

The temporal evolution of sewer networks in our case study demonstrates the growth of the city, which experienced several waves of population increases due to migration. Population growth over five decades is exponential, as migration adds to natural (logistic) growth. The evolution of the sewer network follows these waves, with a more step-wise function for the growth of the SSN following major investment cycles (Fig. A-2).

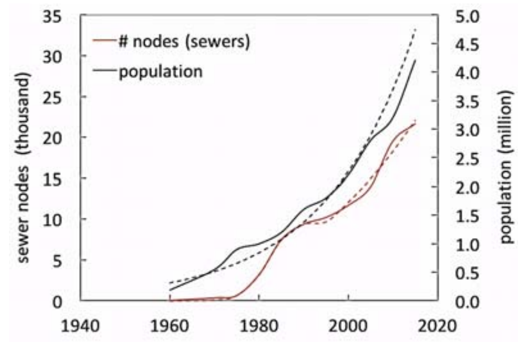

Fig. A-2. Growth of population and SSN in our case study city occurs in waves, with a distinct step-wise growth function for SSN. Dashed lines are fitted models (exponential growth model for population, super -positioned logistic growth model for SSN).

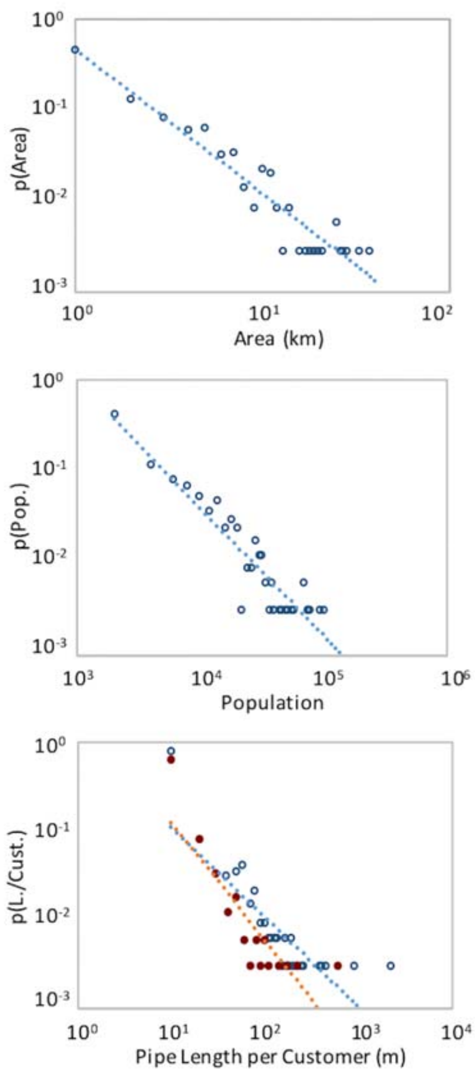

Fig. A-1. Geometric characteristics (*pdfs*) of the water districts (sub-zones of DZs): area: $y = 0.47x^{-1.58}$, $R^2 = 0.90$; population density: $y = 14x^{-1.38}$, $R^2 = 0.87$; pipe length per customer: SSN (closed circles): $y = 2.63x^{-1.32}$, $R^2 = 0.73$; WDN (open circles): $y = 1.11x^{-1.00}$, $R^2 = 0.68$).

References

[1] M. Barthélemy and A. Flammini, Networks Spat. Econ. **9**, 401 (2009).

[2] M. Barthelemy, in *Traffic Granul. Flow ’13*, edited by M. Chraibi, M. Boltes, A. Schadschneider, and A. Seyfried (Springer International Publishing Switzerland, 2015), pp. 317–337.

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

[4] S. Salat and L. Bourdic, TEMA - J. L. Use, Mobil. Environ. **5**, 55 (2012).

[5] J. Buhl, J. Gautrais, N. Reeves, and R. V Sol, Eur. Phys. J. B **49**, 513 (2006).

[6] S. Laemmer, B. Gehlsen, and D. Helbing, Phys. A Stat. Mech. Its Appl. **363**, 89 (2006).
[7] L. M. A. Bettencourt and J. Lobo, J. R. Soc. Interface **13**, (2016).
[8] J. Lobo, L. M. A. Bettencourt, D. Strumsky, and G. B. West, PLoS One **8**, (2013).
[9] L. M. A. Bettencourt, J. Lobo, and G. B. West, Eur. Phys. J. B **63**, 285 (2008).
[10] A. Gomez-Lievano, H. Youn, and L. M. A. Bettencourt, PLoS One **7**, (2012).
[11] X. Zhan, S. Ukkusuri, and P. S. C. Rao, Submitted 1 (2016).
[12] M. Barthélemy, Phys. Rep. **499**, 1 (2011).
[13] F. Blumensaat, M. Wolfram, and P. Krebs, Environ. Earth Sci. **65**, 1427 (2012).
[14] A. Yazdani and P. Jeffrey, Chaos **21**, 1 (2011).
[15] S. Porta, P. Crucitti, and V. Latora, Physica A **369**, 853 (2006).
[16] A. Perna, P. Kuntz, and S. Douady, Phys. Rev. E **83 066106**, (2011).
[17] B. Jiang, Physica A **384**, 647 (2007).
[18] V. Kalapala, V. Sanwalani, A. Clauset, and C. Moore, Phys. Rev. Lett. **73 026130**, (2006).
[19] A. P. Masucci, K. Stanilov, and M. Batty, Phys. Rev. E **89 012805**, (2014).
[20] C. Klinkhamer, E. Krueger, X. Zhan, F. Blumensaat, S. Ukkusuri, and P. S. C. Rao, Rev. (n.d.).
[21] A. Yazdani and P. Jeffrey, Chaos **16111**, (2011).
[22] M. Hu, R. Jiang, R. Wang, and Q. Wu, Phys. Lett. A **373**, 2007 (2011).
[23] J. Lin and Y. Ban, Transp. Rev. **33**, 658 (2013).
[24] A. N. Strahler, *Quantitative Geomorphology of Erosional Landscapes* (Algiers, 1954).
[25] A. N. Strahler, Trans. AGU **38**, 913 (1957).
[26] I. Rodriguez-Iturbe and A. Rinaldo, *Fractal River Basins - Chance and Self-Organization*, 2001st ed. (Cambridge University Press, 1997).
[27] R. Sitzenfrei, Stochastic Generation of Urban Water Systems for Case Study Analysis, Leopold Franzens University Innsbruck, 2012.
[28] Á. Corral and A. Deluca, Acta Geophys. **61**, 1351 (2013).
[29] G. Csányi and B. Szendröi, Phys. Rev. E **69 036131**, (2004).
[30] A. Clauset, C. R. Shalizi, and M. E. J. Newman, SIAM Rev. **51**, 661 (2009).
[31] T. Carletti, F. Gargiulo, and R. Lambiotte, Eur. Phys. J. B **88**, 18 (2014).
[32] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science (80-.). **297**, 1551 (2002).
[33] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[34] J. Dong and S. Horvath, BMC Syst. Biol. **20**, (2007).
[35] E. Strano, V. Nicosia, V. Latora, S. Porta, and M. Barthélemy, Sci. Rep. **2**, 1 (2012).
[36] K. Paik and P. Kumar, J. Stat. Phys. **142**, 862 (2011).
[37] M. Maurer, M. Wolfram, and H. Anja, Water Sci. Technol. **62**, 36 (2010).