



Topological clustering as a tool for planning water quality monitoring in water distribution networks

Kirstein, Jonas Kjeld; Albrechtsen, Hans-Jørgen; Rygaard, Martin

Published in:
Water Science and Technology: Water Supply

Link to article, DOI:
[10.2166/ws.2015.056](https://doi.org/10.2166/ws.2015.056)

Publication date:
2015

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Kirstein, J. K., Albrechtsen, H-J., & Rygaard, M. (2015). Topological clustering as a tool for planning water quality monitoring in water distribution networks. *Water Science and Technology: Water Supply*, 15(5), 1011-1018. <https://doi.org/10.2166/ws.2015.056>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Topological clustering as a tool for planning water quality monitoring in water distribution networks

Jonas Kjeld Kirstein¹, Hans-Jørgen Albrechtsen and Martin Rygaard

Technical University of Denmark, Department of Environmental Engineering,
Miljoevej 113, 2800 Kgs. Lyngby, Denmark,
Email: s103358@student.dtu.dk

Key words

Clustering; Graph theory; Monitoring; Sampling; Water distribution; Water supply.

Abstract

Topological clustering was explored as a tool for water supply utilities in preparation of monitoring and contamination contingency plans. A complex water distribution network model of Copenhagen, Denmark, was simplified by topological clustering into recognizable water movement patterns to: (1) identify steady clusters for a part of the network where an actual contamination has occurred; (2) analyze this event by the use of mesh diagrams; and (3) analyze the use of mesh diagrams as a decision support tool for planning water quality monitoring. Initially, the network model was divided into strongly and weakly connected clusters for selected time periods and mesh diagrams were used for analysing cluster connections in the Nørrebro district. Here, areas of particular interest for water quality monitoring were identified by including user-information about consumption rates and consumers particular sensitive towards water quality deterioration. The analysis revealed sampling locations within steady clusters, which increased samples' comparability over time. Furthermore, the method provided a simplified overview of water movement in complex distribution networks, and could assist identification of potential contamination and affected consumers in contamination cases. Although still in development, the method shows potential for assisting utilities during planning of monitoring programs and as decision support tool during emergency contingency situations.

Introduction

Documenting safe and adequate drinking water quality at the consumers' taps requires frequent and regular monitoring of the water quality in water distribution networks (WDNs) (European Parliament and Council, 1998). However, principles for where to collect the samples are not specified. Since it is practically impossible to check the water quality at every node in the pipe system, it is important to optimize the monitoring strategy to achieve the best possible coverage with as low number of samples as possible.

The increasing size and complexity of urban WDNs also increases the difficulties to assess water movement patterns in the system. Therefore, network segmentation has become a main field of research (e.g. Giustolisi & Ridolfi 2014; Perelman et al. 2015). The design of district metered areas, where in- and outgoing quantities of water are metered, has been introduced to improve leakage control (e.g. Morrison et al. 2007) or security during contamination events (Grayman et al. 2009). Usually, district metered areas are in the size of 1,000 to 5,000 connections (Grayman et al. 2009) and in case of large networks it can be difficult to analyse patterns in water

movement. Therefore, Deuerlein (2008) and Perelman & Ostfeld (2011) developed network simplification tools based on graph theory, which can provide deeper knowledge of the water movement inside complex WDNs. Moreover, network decomposition of large WDNs can be used for contaminant source identification in combination with application of source identification algorithms (Deuerlein et al. 2014). Such backtracking methodologies can be developed to find the most likely set of contaminating nodes based on reverse hydraulics and water quality sensor alarms (e.g. Salomons & Ostfeld 2010). Even though the execution time of such algorithms may be low, the approach is still challenged when an emergency response is needed in WDNs with random sampling locations and only few sensor results available.

In case of an emergency with an unknown contamination source, models and algorithms have to be modified and the utilities may waste valuable time. Thus, tools are needed to support immediate response plans for utilities. These can be prepared beforehand and ready to use, independent of where the contamination has been detected.

WDNs can be simplified through a topological analysis by dividing the WDN into “strongly connected clusters” (SCCs) and “weakly connected clusters” (WCCs) (Figure 1) (Perelman & Ostfeld 2011). A SCC is formed for all nodes u and v , with a directed path (sequence of distinct nodes) from u to v and a directed path from v to u . A WCC has only one directed path, either from u to v or from v to u . With this approach, a ranked connectivity matrix can be established to model a contaminant intrusion. The number of clusters generated in larger WDNs may be further minimized by merging smaller clusters, e.g. by setting upper and lower bounds for classification of WCCs as single WCCs or whether they should be merged with adjacent clusters (Perelman & Ostfeld 2012).

Originally, SCC boundary nodes or starting points, such as service reservoirs or tanks, served as root nodes for the identification of WCCs. This approach risked ambiguity in terms of whether one or two WCCs were identified when a SCC had two boundary nodes, but where only one node was directly connected downstream to the other node. Thus, the identification of one or two WCCs depended on the starting point of the algorithm. To avoid this ambiguity in cluster formations we previously further developed the method by merging adjacent WCCs to generate unique clusters (Kirstein et al. 2014) and consequently, no WCCs will be merged with SCCs. The length of the analysed time period strongly influences the number of generated clusters, but in time periods with more or less constant flow directions, patterns of clusters were recurring, and steady clusters were identified. With inspiration from the field of computational science, cluster connections can be analyzed and visualized with mesh diagrams (also known as cluster topology charts (Perelman and Ostfeld, 2011) or mesh topology (Solomon & Kim, 2011)).

To further investigate the strength of topological clustering our aims were to: 1) Identify steady clusters for a small defined part (Nørrebro district) of the Copenhagen WDN, where a contamination has occurred, 2) analyse the usefulness of the approach in an actual contamination event by the use of mesh diagrams, and 3) use mesh diagrams as support for establishing monitoring plans. Additionally it was investigated if the clustering method could be modified to

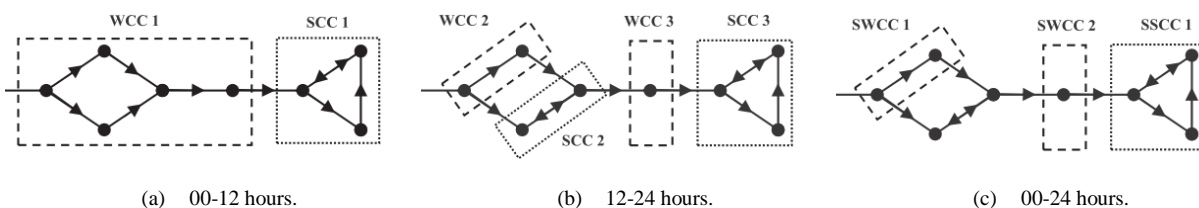


Figure 1. Clusters identified for two selected time periods between (a) 00 and 12 hours; (b) 12 and 24 hours. Clusters are steady when they exist in all time periods (c). WCC = weakly connected cluster; SCC = strongly connected cluster; SWCC = steady WCC; SSCC = steady SCC.

include a rank of importance by visualising the size of water flows through a district and add attention to sensitive consumers such as care institutions.

Method

Topological clustering analysis based on fundamental notions of graph theory (Perelman & Ostfeld 2011, 2012) was previously modified into a stepwise approach (Kirstein et al., 2014):

1. *Selection of time periods:* The analysis was conducted for selected time periods based on typical consumption patterns, in casu two periods: 00-07 hours and 07-22 hours. Due to anomalies in the flow patterns between 22-24, this period was left out to be subject for future studies.
2. *Identifying cluster formations:* For each selected time period an adjacency matrix was established and SCCs were identified with the depth first search (Tarjan, 1972). By deleting all strong connections in the matrix, only weak connections were left. Adding this modified adjacency matrix to its transposed, the depth first search algorithm was capable of identifying all adjacent WCCs (disguised as SCCs) (Figure 1).
3. *Steady cluster analysis:* Clusters of selected time periods were compared to reveal intersections to identify steady clusters which remained in all time periods (Figure 1).
4. *Mesh diagram visualization:* Connections between clusters of selected time periods were stored in an adjacency matrix. This matrix was used to visualize the connections between the clusters (denoted as mesh diagram). Arrows visualised the existence of a connection and flow direction of the water between the clusters. The size of the illustrated clusters was weighted based on e.g. the number of nodes or the water consumption.

The investigated network model for Copenhagen, Denmark (Greater Copenhagen Utility 2012) (Figure 2) was originally build in MIKE URBAN (DHI 2014) and modelled in EPANET (US EPA 2014; Kirstein et al. 2014). A close-up of the Nørrebro district revealed 1053 nodes and 1248 links and in this configuration 33 connections to the surrounding districts. Selected nodes

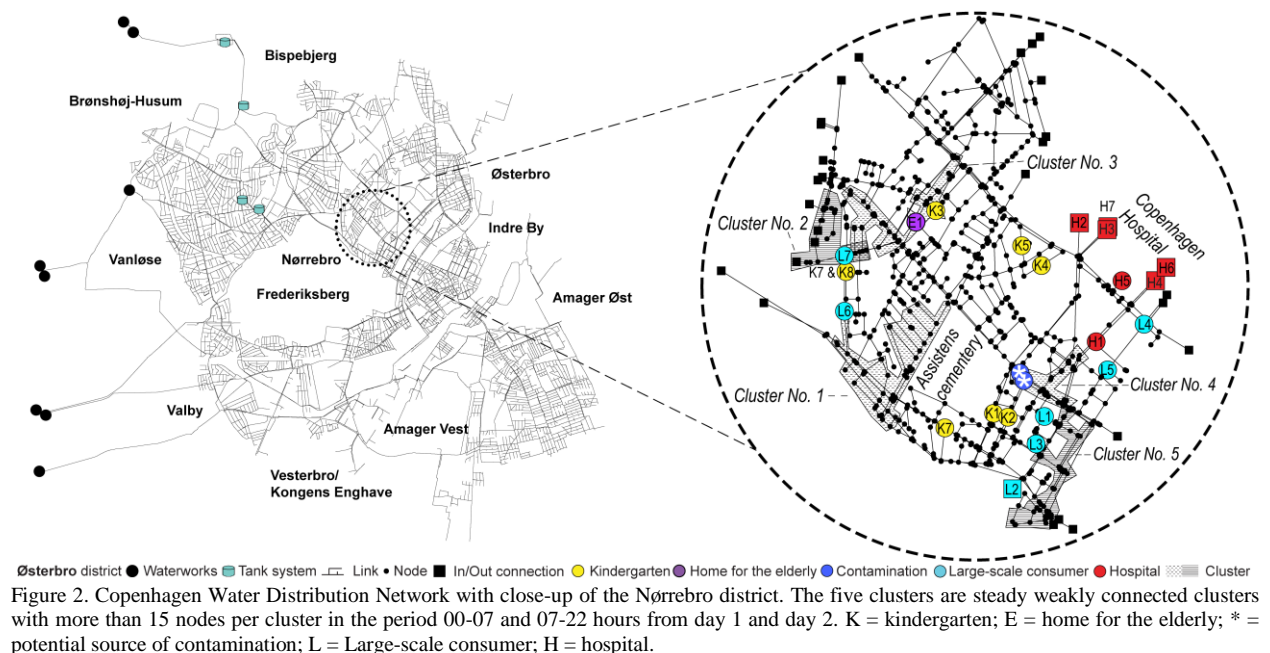


Table 1. Cluster formations in the Nørrebro district of Copenhagen's Water Distribution Network on day 1 and 2. The values given in parenthesis are representing the parameters listed in parenthesis in the capture of the column. SCC = strongly connected cluster; WCC = weakly connected cluster. Density = average nodes per cluster.

Time period (day)	SCCs (nodes)	SCC density	WCCs (nodes)	WCC density
00-07 (1)	72 (467)	6.5	91 (586)	6.4
07-22 (1)	67 (162)	2.4	27 (891)	33
00-07 (2)	77 (462)	6	87 (591)	6.8
07-22 (2)	59 (145)	2.5	29 (908)	31.3

with a particular sensibility for water quality deterioration such as hospitals (H1-7), kindergartens/nurseries (K1-8) and one elderly home (E1) were identified by Google Maps (2014) and the potential source of a contamination event in 2011 is marked (Figure 2). In 2011, rainwater infiltrated into the WDN and about 40,000 people were affected for five days by a boil-water advisory (Greater Copenhagen Utility 2011). Large-scale consumer nodes were considered of particular interest (L1-7) and were defined as nodes with a base demand > 1 L/s (Figure 2).

The relative importance of a cluster was defined by its water demand and visualised in a mesh diagram where the cluster size was relative to its total water consumption. In this paper, a person equivalent (pe) water consumption of 148 L/d was assumed for the WDN model of Copenhagen (City Of Copenhagen 2012). The EPANET simulation was based on a 24-hour demand pattern repeated over two days, representing typical water usage in Copenhagen (Kirstein et al. 2014).

Results & Discussion

Two time periods 00-07 and 07-22 hours were selected for the cluster formations on two days of simulation for the sub-district of Nørrebro (Kirstein et al. 2014) (Table 1). The number of identified clusters changed less than 12% from day 1 to day 2, so we focused on the results from day 1. Within the first time period (00-07 hours), the main tank system (located north of the Bispebjerg and Brønshøj-Husum districts, Figure 2) was filled and the consumption in the system was low. Therefore, the water frequently moved back and forth, and the density of nodes per SCC was higher in this period. Between 07-22 hours, more than 85% of all nodes were part of a WCC, reflecting dominating water transport with branched flow from the waterworks to the consumers.

Steady clusters

The two different cluster results of Nørrebro (Table 1) indicated that the movement of water changes with the time of day and thus the interpretation of a sample origin and distribution will depend on the sampling time. A steady cluster analysis for the four time periods of day 1 and 2 showed that hydraulic conditions varied greatly in the district between the time periods, since less than 50% of all nodes in Nørrebro were assigned to a steady cluster (Table 2). Thus, a sample collected at e.g. 05:00 hours probably represents a different upstream source and downstream distribution than a sample taken at 16:00 hours.

Table 2. Steady cluster formations in the Nørrebro district between 00-07 and 07-22 hours on day 1 and day 2. The values given in parentheses are representing the parameters listed in parenthesis in the capture of the column.

Steady SCCs (nodes)	Steady SCC density	Steady WCCs (nodes)	Steady WCC density	Steady single nodes (not part of steady cluster)
20 (44)	2.2	58 (439)	7.6	570

Sampling in a steady SCC, regardless of location, increased the likelihood of representing the same water quality for the entire cluster because the water frequently was exchanged between its

nodes. Thus, intruding contaminants would probably spread to all nodes in the cluster. In contrast to steady SCCs, several steady WCCs were large with an average 7.6 nodes per cluster (Table 2).

Among the nodes with a particular sensibility towards water quality deterioration, only L6 was located within a steady cluster with more than 15 nodes (cluster No. 1, Figure 2). Depending on the flow direction at the particular node, sampling at L6 either represented the downstream (flow direction points from this node into the steady cluster) or upstream connected nodes (flow direction points from steady cluster to this node) within the steady cluster over all four time periods (Figure 2). Here, our clustering method helped to delineate areas where the hydraulic conditions remained constant and where samples were more likely to be time-independent and thus better for assessing water quality deteriorations.

Mesh diagrams as a contamination contingency tool

Mesh diagrams were analysed as an emergency response tool for: 1) delineating the area affected by contamination, and 2) assisting in tracking the source of the contamination. In 2011, the WDN was contaminated and it was suspected to originate from a construction site in Nørrebro (Figure 2). In the mesh diagram of the first time period (00-07 hours, Figure 3a), the cluster size was based on the number of nodes per cluster, and the nodes that were considered as the potential source of the contamination were distributed within four different clusters. The contamination was first detected in cluster No. 2 (Figure 3a). Had the mesh diagram been available then, a worst case scenario could have been mapped by following all outgoing connections from cluster No. 2, showing the extent to which the contamination could be distributed. Assuming that the contamination event started at 00 hours and that the sample was taken before 07 hours, several clusters (e.g. cluster No. 3), could have been excluded as the potential origin due to the lack of a direct connection to cluster No. 2. Although in this situation the water in cluster No. 2 could originate from nine different clusters, cluster No. 1 was the only cluster which could have delivered the contamination to the sampling location within the selected time period. Following the flow from cluster No. 1 or No. 2, the visualization of vulnerable nodes emphasizes where an immediate response would be necessary, e.g. in the downstream connected hospital's nodes. Thus, quick emergency response actions could be enforced based on early detection of possible contaminated clusters with focus on demand size or sensitive consumers such as hospitals or kindergartens.

In case a contamination has been detected in samples from the SCC containing L4&5 (Figure 3a) the clustering method could predict the probable spreading of the contamination to the downstream connected clusters including hospital nodes H4-H7. Based on the cluster analysis, water utility companies could respond to this contamination by taking samples in the five upstream connected clusters to the SCC containing L4&L5 and if no further contamination was detected, the utility could delineate the affected area. First, the contamination would likely reach the mentioned hospital nodes. Second, assuming the mesh diagram included an entire network, it would be advisable to stop the connection between the WCC containing node H2&3 and the downstream connected SCC, since this is the only connection where the contamination could spread to further locations in the network. Also, analysing the spreading of the contamination by topological clustering may be useful, if the contamination was detected at several locations. Conducting such a task by conventional methods, such as backward and forward tracing of flow, as applicable in some hydraulic models (DHI 2014), would grow markedly in size with an increasing network and an increasing distance between the source and contaminated nodes.

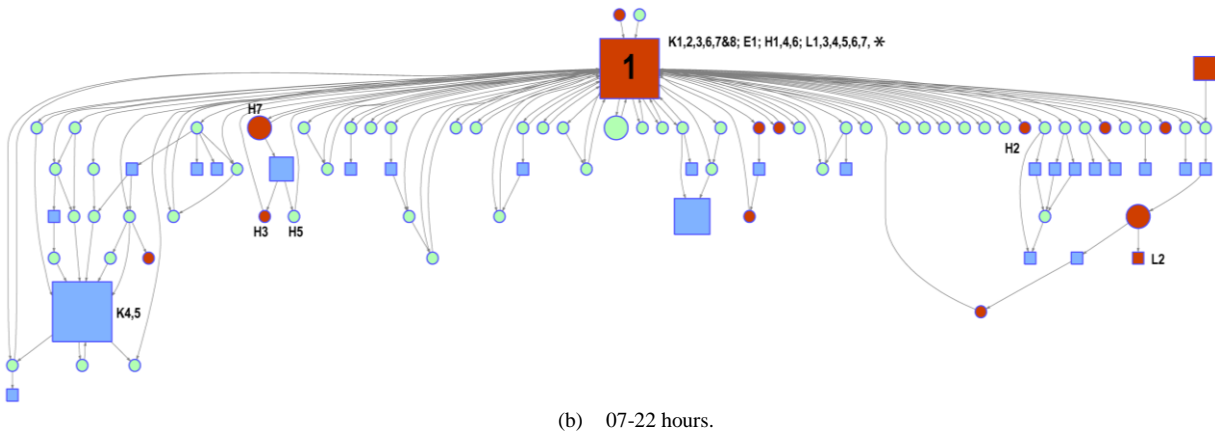
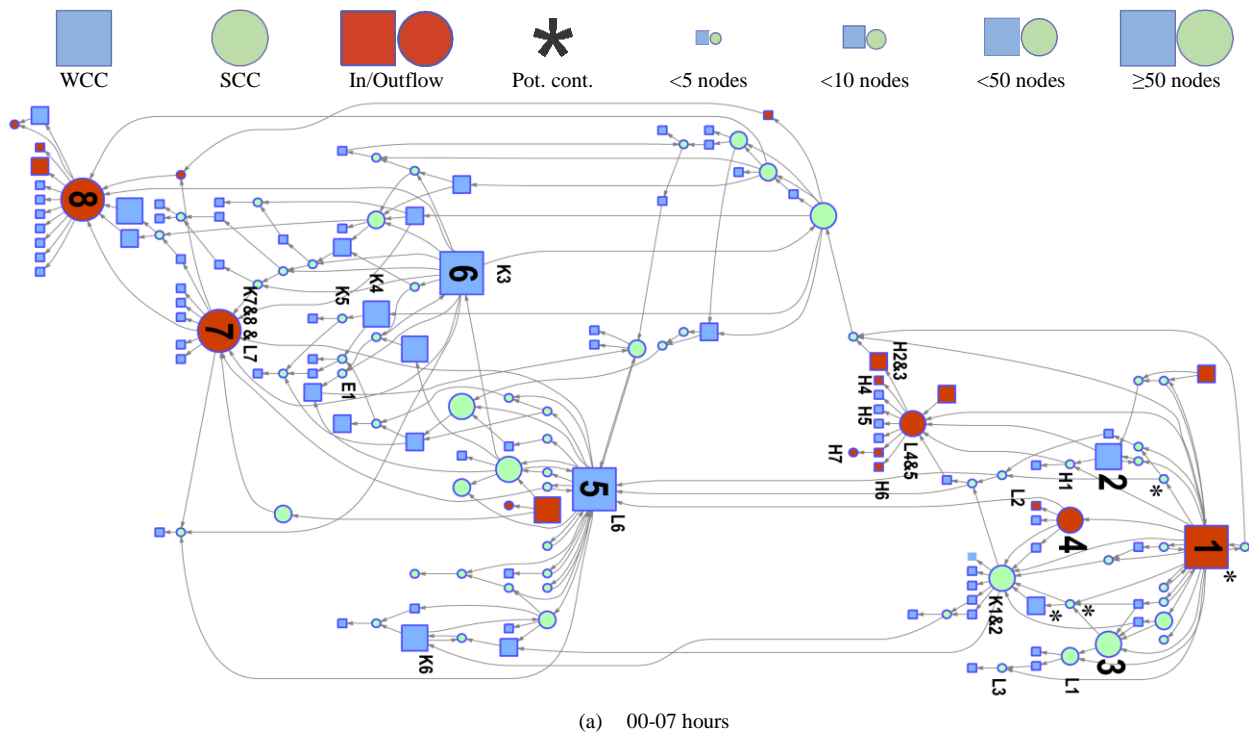


Figure 3. Mesh diagram for day 1, Nørrebro, time period: 00-07 hours (a) and 07-22 hours (b). Different cluster sizes illustrate the number of nodes per cluster. Clusters marked with red have an in- and/or outflow connection from the district. SCC = strongly connected cluster; WCC = weakly connected cluster. L = large-scale consumer; K = kindergarten; E = home for the elderly; H = hospital. A “*” illustrates the location of a potential contamination (Pot. cont.).

Compared to the first time period, it was not as easy to apply the 2011 contamination event on the second time period. At 07-22 hours, the fixed sampling location was located in cluster No. 1 (Figure 3b). Due to the impractical size of 751 nodes (71% of all nodes in Nørrebro), it was difficult to assess the general transport of the contamination. In this case, the WCC was oversimplified, although at the same time the high number of small SCCs did not reduce the complexity substantially.

Only two of the nodes suspected to be the contamination source (located in cluster No. 1 in Figure 3a and Figure 3b), were part of a larger steady WCC (named cluster No. 4, Figure 2).

Thus, the hydraulic conditions in most of the remaining nodes being suspected to be the origin of the contamination changed over the time periods, allowing the contamination easily to be spread to several areas. However, assuming that it is known that the contamination originated from the two nodes in Cluster No. 4 (Figure 2) and that this cluster has no further downstream connections, could have helped to narrow down the area affected by the contamination immediately. It could be assumed that the contamination would not spread to further clusters, since the flow directions were steady in the cluster.

Mesh diagrams as a tool for monitoring plans

Constructing another mesh diagram (Figure 4) for the same time period (00-07 hours) as previous (Figure 3a), but where the cluster size was based on the water consumption within the analysed period, some of the clusters, such as cluster No. 4, changed drastically in size due to the high water consumption of L2 located within this cluster. This emphasizes the importance of monitoring the water quality at this location because a high water consumption – and subsequently many consumers - are affected by this cluster. Additionally, cluster No. 1 and No. 3 had a very high consumption of more than 5000 pe, whereas cluster No. 6 and No. 8 decreased in size, indicating a low water consumption per node in these clusters.

However, during the second time period (07-22 hours), it was difficult to assess the general flow of the water through the district and thus to pinpoint a strategic sampling location. In this time period the mesh diagram weighted by consumption was similar to the mesh diagram weighted by the number of nodes, because almost all locations with a particular sensibility towards water quality deterioration, such as hospital or kindergarten nodes were present in the largest cluster (No. 1, Figure 3b). Identification of strategic sampling locations would benefit from further analysis of the nodes within this cluster.

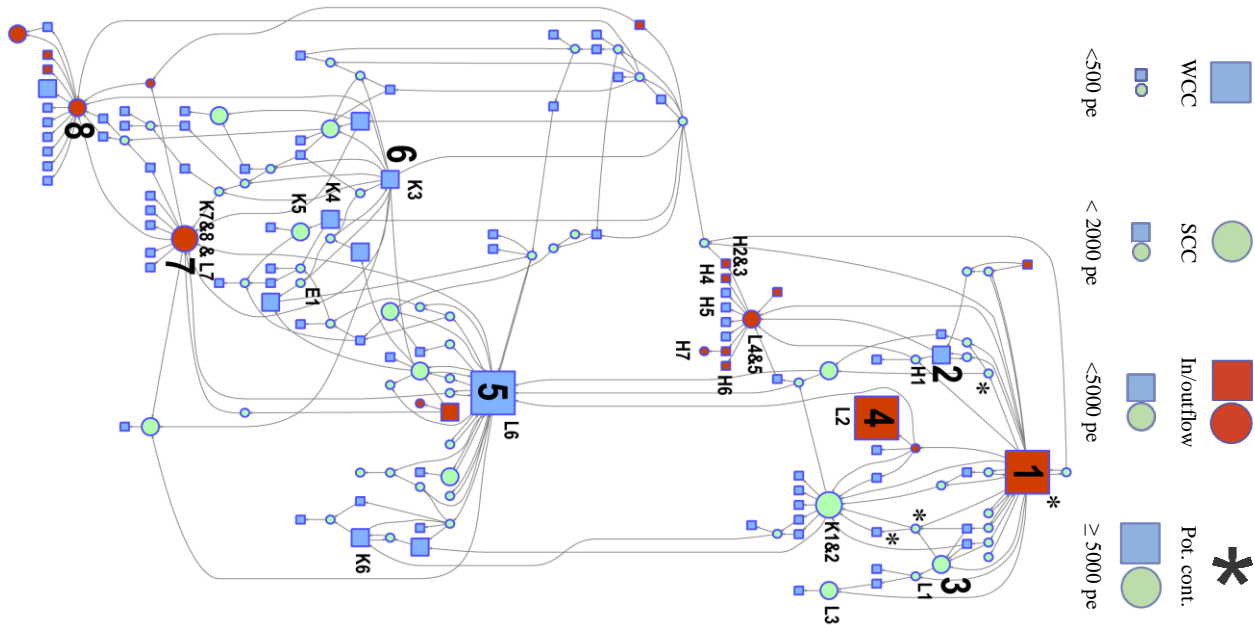


Figure 4. Mesh diagram for day 1, Nørrebro, time period: 00-07 hours. Different cluster sizes illustrate the water consumption within the selected period in person equivalents (pe). The total consumption is approx. 116,000 pe. Clusters marked with red have an in and / or outflow connection from the district. SCC = strongly connected cluster; WCC = weakly connected cluster. L = large-scale consumer; K = kindergarten; E = home for the elderly; H = hospital. .

Implications for the utility and further work

Our application of topological clustering to Copenhagen's WDN is considered a work in progress. The results show potential for future decision support both during planning and during emergency response situations. Some practical issues remain before it can be fully uptaken into the utility organization. We have identified the following future development potentials: 1) A better understanding of flow anomalies in the hydraulic model of Copenhagen's WDN. 2) Further analysis of temporal changes in cluster formation, so the analysis can be expanded to visualize changes in flow patterns with time, e.g. from hour to hour. 3) Incorporating the clustering method into software packages with a user friendly graphical interface.

Conclusion

- Topological clustering within selected time periods was demonstrated to be applicable for monitoring and emergency contingency planning of water distribution networks by generating mesh diagrams. Identification of steady clusters revealed areas of constant hydraulic conditions and areas where changes in the water quality were more easily spread to all nodes
- Mesh diagrams based on topological clustering can provide a simplified overview of the potential origin and spread of water in complex distribution networks
- Availability of mesh diagrams during a contamination in Copenhagen 2011 could have identified affected consumers and helped to pinpoint the source of contamination
- Considering water demand and consumers with particular sensibility for water quality deterioration in the mesh diagram could add a further dimension in the identification of clusters of special importance in monitoring and contingency situations
- Mesh diagrams can assist in identification of strategic sampling locations

Acknowledgment

We thank DHI and Greater Utility Copenhagen for sponsoring software and data used in the modelling. The project was partly funded by the Danish Agency for Science, Technology and Innovation (Project: Water in Urban Areas).

References

- City Of Copenhagen, 2012. Towards sustainable water consumption. URL <http://subsite.kk.dk/sitecore/content/Subsites/CityOfCopenhagen/SubsiteFrontpage/LivingInCopenhagen/ClimateAndEnvironment/CopenhagensGreenAccounts/Water.aspx> (accessed 3.6.15).
- Deuerlein, J., Piller, O., Montalvo, I., 2014. Improved Real-time Monitoring and Control of Water Supply Networks by Use of Graph Decomposition. *Procedia Eng.* 89, 1276–1281. doi:10.1016/j.proeng.2014.11.436
- Deuerlein, J.W., 2008. Decomposition Model of a General Water Supply Network Graph. *J. Hydraul. Eng.* doi:10.1061/(ASCE)0733-9429(2008)134:6(822)

- DHI, 2014. MIKE URBAN, GIS-based urban modeling system for water distribution systems and wastewater collection systems. URL <http://releasenotes.dhigroup.com/2014/MIKEURBANrelinf.htm> (accessed 3.6.15).
- European Parliament and Council, 1998. Directive 98/83/EC of 3 November 1998 on the quality of water for human consumption. Off. J. Eur. Communities L330, 32–54.
- Giustolisi, O., Ridolfi, L., 2014. New Modularity-Based Approach to Segmentation of Water Distribution Networks. *J. Hydraul. Eng.* 140, 04014049. doi:10.1061/(ASCE)HY.1943-7900.0000916
- Google Maps, 2014. Google Maps - Nørrebro. URL <https://www.google.com/maps/d/viewer?mid=zUuchysXD8k.kHgVHkTGPeqs&ie=UTF&msa=0> (accessed 3.6.15).
- Grayman, W.M., Murray, R., Savic, D.A., 2009. Effects of Redesign of Water Systems for Security and Water Quality Factors, in: *World Environmental and Water Resources Congress 2009*. American Society of Civil Engineers, Reston, VA, pp. 1–11.
- Greater Copenhagen Utility, 2011. Beredskab i praksis... Vandforurening i København, august 2011 (emergency response to drinking water contamination in Copenhagen, august 2011). URL <http://www.danva.dk/Admin/Public/Download.aspx?file=Files%2FFiler%2FMoeder+og+kurser%2FForsyningstr%C3%A6f%2FBeredskabssituation+fra+august+2011+i+KE+pr.+30.01.12.pdf> (accessed 3.6.15).
- Greater Copenhagen Utility, 2012. HOFOR, EPANET model of Copenhagen, version 2009, modified 2012.
- Kirstein, J.K., Albrechtsen, H.-J., Rygaard, M., 2014. Simplification of Water Distribution Network Simulation by Topological Clustering – Investigation of its Potential Use in Copenhagen’s Water Supply Monitoring and Contamination Contingency Plans. *Procedia Eng.* 89, 1184–1191. doi:10.1016/j.proeng.2014.11.248
- Morrison, J., Tooms, S., Rogers, D., 2007. District Metered Areas - Guidance Notes.
- Perelman, L., Ostfeld, A., 2011. Topological clustering for water distribution systems analysis. *Environ. Model. Softw.* 26, 969–972. doi:10.1016/j.envsoft.2011.01.006
- Perelman, L., Ostfeld, A., 2012. Water-Distribution Systems Simplifications through Clustering. *J. Water Resour. Plan. Manag.* 138, 218–229. doi:10.1061/(ASCE)WR.1943-5452.0000173
- Perelman, L.S., Allen, M., Preis, A., Iqbal, M., Whittle, A.J., 2015. Automated sub-zoning of water distribution systems. *Environ. Model. Softw.* 65, 1–14. doi:10.1016/j.envsoft.2014.11.025

Salomons, E., Ostfeld, A., 2010. Identification of Possible Contamination Sources using Reverse Hydraulic Simulation. *Water Distrib. Syst. Anal.* 2010 447–453. doi:10.1061/41203(425)42

Solomon, M.G., Kim, D., 2011. *Fundamentals of Communications and Networking*. Jones & Bartlett Learning.

Tarjan, R., 1972. Depth-first search and linear graph algorithms. *SIAM J. Comput.* 1, 146 – 160.

US EPA, 2014. EPANET, water distribution system modeling software. URL <http://www.epa.gov/nrmrl/wswrd/dw/epanet.html> (accessed 3.6.15).