Corona, Francesco; Mulas, Michela; Mikola, Anna; Kuokkanen, Anna; Heinonen, Mari; Vahala, Riku

Network representation and analysis of a large-scale wastewater treatment plant

Published in:
12th IFAC Symposium on Dynamics and Control of Process Systems

DOI:
10.1016/j.ifacol.2019.06.089

Published: 01/01/2019

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Corona, F., Mulas, M., Mikola, A., Kuokkanen, A., Heinonen, M., & Vahala, R. (2019). Network representation and analysis of a large-scale wastewater treatment plant. In *12th IFAC Symposium on Dynamics and Control of Process Systems* (pp. 364-369). (IFAC-PapersOnLine; Vol. 52, No. 1). Elsevier.
https://doi.org/10.1016/j.ifacol.2019.06.089

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 52-1 (2019) 364–369

Network representation and analysis of a large-scale wastewater treatment plant

Francesco Corona * **Michela Mulas** ** **Anna Mikola** ***
Anna Kuokkanen **** **Mari Heinonen** **** **Riku Vahala** ***

* *Department of Computer Science, Federal University of Ceará, Fortaleza (CE), Brazil, (e-mail: francesco.corona@ufc.br).*
** *Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza (CE), Brazil, (e-mail: michela.mulas@ufc.br).*
*** *Department of Built Enviroment, Aalto University, Espoo, Finland, (e-mail: {anna.mikola;riku.vahala}@aalto.fi)*
**** *Helsinki Region Environmental Services Authority, Helsinki, Finland, (e-mail: {anna.kuokkanen;mari.heinonen}@hsy.fi)*

Abstract: We mapped a large-scale wastewater treatment plant onto a complex network and we investigated how the structural properties of the graph evolve in time as the facility is operated. The Viikinmäki plant is mapped onto a dependence network in which the nodes are online process measurements and interconnectivity between nodes encodes pairwise correlations between the corresponding time series, as estimated over moving-windows. In this initial study, the construction of a graph of Viikinmäki is presented and results are discussed with the goal of understanding its usability as model for process interactions and encoder of latent structures.

Keywords: Process monitoring, Large-scale systems, Wastewater treatment, Complex networks.

1. INTRODUCTION

Recent years have witnessed a remarkable research effort in network science, with advancements fostered by studies of complex systems, such as those found in social, information and biological sciences (Barabási, 2012). Concepts from algebraic graph theory, probability theory and statistical mechanics are at the core of the theory of complex networks (Estrada, 2016; Crane, 2018). Complex systems are mapped onto graphs and their functioning studied from the internal structure of the resulting networks. For the task, a host of techniques has been developed (Dorogovtsev, 2010; Newman, 2018). Goals range from explorative to system analytical (Barrat et al., 2008; Kolaczyk, 2009; Lambiotte and Masuda, 2016). In explorative analyses, interest is in identifying statistical features, such as degree and path distributions, node and edge centralities, that characterise the system from the topological structure and communication paths of the graph. In network system analysis, interest is in the effect of a static graph connectivity on dynamical phenomena and in the dynamics of temporal graphs in which connectivity varies over time.

Although their use is not widely spread, complex networks and graphs have also been considered in process systems engineering. The earliest applications have been proposed by Gilles (1998) and Mangold et al. (2002) in the context of plant design and operation. More recently, the compartmental nature of process plants has been discussed by Preisig (2009) as a way of addressing online balancing of mass and energy flows. The flowsheets of chemical plants have been studied by Maurya et al. (2004) to define a graph-theoretical approach to fault-diagnosis. Andrade Jr. et al. (2006) also studied the properties of plant flowsheets and observed the existence of a small-world structure. Recently, Koeln and Alleyne (2018) propose a hierarchical model predictive control for network power flow system.

The literature seems to focus on the graph nature of process plants as it is induced by the exchange of mass and/or energy between compartments. In this work, on the other hand, we are interested on the more intangible connection that exists between sensor measurements, irrespective of the physical trajectory of process flows. We consider a large-scale wastewater treatment plant (WWTP) and we map it on a graph, in such a way that nodes correspond to on-line sensors and the edges between them represent their mutual dependence. The scope is to study the behaviour of such a representation over time, under a range of operating conditions, and to evaluate its applicability as a model for encoding the interactions between process variables.

The work is structured as follows. In Section 2, the Viikinmäki wastewater treatment plant is overviewed in terms of main process units and existing automation setting. In Section 3, introductory notions on graph theory are reviewed; emphasis is on structural properties that highlight the relevance of the nodes. Section 4 discusses how the Viikinmäki plant has been mapped onto a network and discusses some results that emerged from its analysis.

2. THE VIIKINMÄKI WWTP

The Viikinmäki WWTP (1.1M *Population Equivalent*) is the largest municipal wastewater treatment plant in the Nordic countries. The plant is built in a bedrock and treats an average influent of 280K m^3d^{-1} (with peaks of 800K m^3d^{-1}), 93% is domestic and 7% is industrial wastewater.

10.1016/j.ifacol.2019.06.089

The wastewater treatment line consists of bar screening, grit removal, pre-aeration, primary sedimentation, activated sludge process (ASP, 8+1 lines in DN-configuration), secondary sedimentation with (8+1)×2 tanks, and tertiary biological treatment (10 denitrifying post-filtration lines, FLT). Sludge is treated in 4 mesophilic digesters and dewatering systems. Biogas from sludge digestion is used for electricity and heat generation. Since 2004 (introduction of FLT lines) and since 2014 (introduction of 9th ASP line), a total nitrogen removal of ≈ 90% and a biochemical oxygen demand removal of ≈ 95% of yearly averages is achieved.

Activated sludge process The activated sludge process consists of nine plug-flow basins, one line is sketched in Figure 1. Each line begins with a mixing zone where pre-settled wastewater, return sludge from secondary sedimentation and internal recycle sludge are fed. Each basin is split in six sequential zones, the anoxic ones being in the beginning. The number of non-aerated zones (anoxic volume) depends on the aeration mode, which is adjustable.

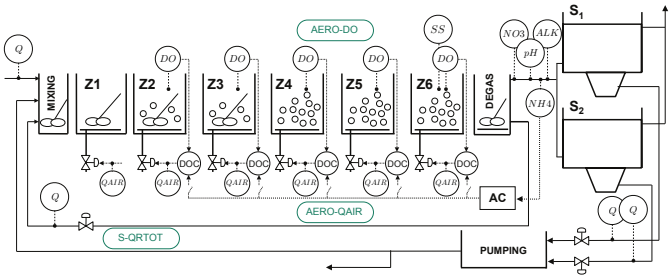

Fig. 1. Viikinmäki: Activated sludge process, one line.

The ASP lines are monitored with sensors providing online measurements for the most important process variables (Table 1). Influent from the primary sedimentation to the ASP lines is characterised in terms of flow-rate (I-Q). Flow-rates of sludge recirculation from the secondary sedimentation tanks (S1-QR and S2-QR) and internal recirculation (QA) are also measured. Dissolved oxygen (DO) concentration is measured in zones Z2 to Z6 (Z2/Z6-DO), while mixed liquor suspended solids are analysed only in zone Z6 (Z6-SS). The effluent is monitored in terms of ammonia (D-NH_4), nitrate (D-NO_3), alkalinity (D-ALK) and pH (D-pH). Lime is dosed upstream to control alkalinity within the optimal range for nitrification.

Table 1. Activated sludge process lines (Li, $i = 1, \ldots, 9$), measured and calculated data.

Name	Description	Units
Li-I-Q	Influent flow-rate	m^3/s
Li-Q-INT	Internal recycle flow-rate	m^3/s
Li-S1/S2-QR	Return sludge from S1 and S2	dm^3/s
Li-S-QRTOT	Total recycle flow-rate	m^3/s
Li-Z2/Z6-DO	Dissolved oxygen (Z2 to Z6)	mg/L
Li-Z1/Z6-QAIR	Air flow-rate (Z1 to Z6)	Nm^3/s
Li-Z6-SS	Suspended solids (Z6)	g/L
Li-D-NH4	Ammonia	mg/L
Li-D-NO3	Nitrate	mg/L
Li-D-ALK	Alkalinity	mmol/L
Li-D-pH	pH	–
Li-AERO-DO	Average DO (Z2 to Z6)	mg L^{-1}
Li-AERO-QAIR	Total QAIR (Z2 to Z6)	Nm^3/s

In each of the ASP lines, dissolved oxygen in all the zones (Z2-DO to Z6-DO) can be feedback-controlled by the corresponding air flow-rates. Aerobic zones Z6 to Z4 are always aerated, with a DO target of 3.5 mgL^{-1}. Aeration in zones Z3 to Z2 is progressively switched on only when the ammonia content D-NH4 exceeds a treatment threshold (4 mgL^{-1}). Zone Z1 is never aerated. When aerated, the DO targets in zones Z2 to Z3 are usually set to 3.5 mgL^{-1}. The number of aerated zones is used to meet the removal efficiency by adjusting the anoxic volume. As common practice for the external recycle, the flow-rates of return activated sludge from settlers (S1/S2-QR) are set to be proportional to the influent (I-Q). As for the internal recycle, a logic based on recirculated sludge and influent flow-rate and number of aerated zones, is used.

Denitrifying post-filtration The denitrifying postfiltration unit receives wastewater from the (8+1)×2 secondary sedimentation tanks. Nitrate removal is achieved by ten Biostyr filters (Figure 2 and Table 2) arranged in parallel.

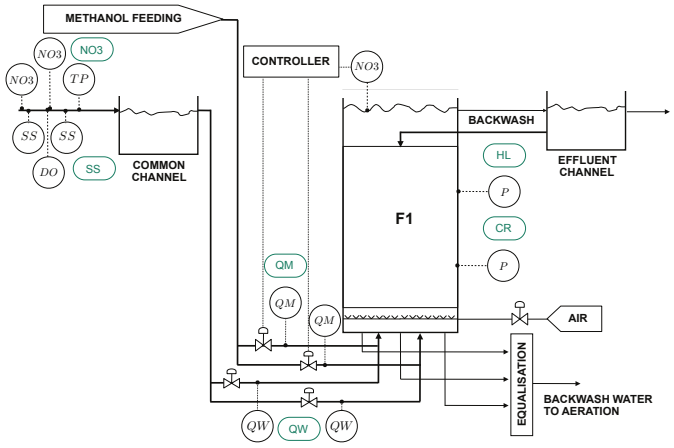

Fig. 2. Viikinmäki: Denitrifying post-filtration, one line.

Influent wastewater from sedimentation is distributed to the filter cells and, before each cell, the flow is split in two streams (QW1 and QW2), where methanol in water (10% dilution) is added (QM1 and QM2). Methanol provides a biodegradable organic substrate as energy source for the denitrifying bacteria. Methanol flow-rate into each filter is manipulated to control the nitrate concentration in the cell (NO3); the control is also based on nitrate and dissolved oxygen loads into the filters (I-NO3 and I-DO). Inside the filter, wastewater flows upwards through a floating support covered in biomass. To control clogging, the cells

Table 2. Denitrifying post-filtration lines (Fi, $i = 1, \ldots, 10$), measured and calculated data.

Name	Description	Units
F-I-NO3 (1,2)	Influent Nitrate (sensor 1,2)	mg/L
F-I-SS (1,2)	Influent Suspended solids (sensor 1,2)	mg/L
F-I-DO	Influent Dissolved oxygen	mg/L
F-I-TP	Influent Total phosphorus	mg/L
Fi-QW(1,2)	Wastewater flow-rate (line 1,2)	m^3/s
Fi-QM(1,2)	Methanol flow-rate (line 1,2)	m^3/h
Fi-P(1,2)	Pressure (bottom, top)	kPa
Fi-NO3	Nitrate	mg/L
Fi-CR (HL)	Clogging rate (Head-loss)	% (m)
F-I-NO3	Average influent NO3	mg/L
F-I-SS	Average influent SS	mg/L
Fi-QW-TOT	Total QW	m^3/s
Fi-QM-TOT	Total QM	m^3/h

are backwashed one at a time with effluent wastewater and a counter-current airflow. After filtration, wastewater from the filters is collected in the effluent channel, for discharge.

Influent wastewater to the denitrifying post-filtration unit is monitored, before division to each of the cells, in terms of dissolved oxygen (F-I-DO), suspended solids (F-I-SS(1,2)), nitrate (F-I-NO3(1,2)) and total phosphorus (F-I-TP).

3. BRIEF ON NETWORK CONCEPTS

Networks are understood as diagrammatic representations of a system whose components are interconnected. Informally, networks consist of nodes that represent the components of the system and pairs of nodes are joined by links, which represent interactions between components. Networks admit a mathematical representation as graphs. We provide an overview of graph theory notions and of some structural properties of simple graphs (Diestel, 2010).

3.1 General and algebraic notions

Consider a finite set $V = \{v_1, v_2, \dots, v_n\}$ of elements and let $V \otimes V$ be the set of all ordered pairs $[v_i, v_j]$ of the elements of V. A relation on set V is any subset $E \subseteq V \otimes V = \{e_1, e_2, \dots, e_m\}$. If $[v_i, v_j] \in E$ implies $[v_j, v_i] \in E$, then E is symmetric. If $[v, v] \in E$ for all $v \in V$, then E is reflexive. The relation is anti-reflexive if $[v_i, v_j] \in E$ implies $v_i \neq v_j$. A simple graph is defined as the pair $\mathcal{G}(V, E)$; V is the set of nodes and E is a symmetric and anti-reflexive relation on V. The elements of E are the edges or links.

The number $n = |V|$ of nodes corresponds to the number of entities in the system, the number $m = |E|$ of links represents the number of relations between them. Symmetry of E implies that simple graphs are undirected and interactions are bi-directional. In weighted graphs, edges are assigned a strength; E is replaced by $W = \{w_1, w_2, \dots, w_m\}$.

Two nodes u and v are adjacent if a link $e = \{u, v\}$ joins them; Nodes u and v are incident with link e and link e is incident with nodes u and v. Interactions between nodes occur even when they are not directly linked. A *walk* is a sequence of not necessarily distinct edges, $(u_1, v_1), (u_2, v_2), \cdots, (u_p, v_p)$, such that $v_i = u_{i+1}$, $i = 1, 2, \dots, (p-1)$. If $v_p = u_1$, the walk is closed. A *trial* is a walk whose edges (but not necessarily all the nodes) are distinct, a *path* is a trail whose nodes (and all the edges) are distinct. The walk/trial/path length is p and it can be shown that the *shortest walk* between distinct nodes is also the *shortest path*. A network is connected if there is a path between any two nodes. If a simple graph is not connected then it can be divided in disjoint connected components.

Adjacency relations Let $\mathcal{G}(V, E)$ be a simple network with $V = \{1, 2, \dots, n\}$. For all pairs $u, v \in [1, n]$, we define

$$a_{uv} = 1, \text{ if } (u, v) \in E; \quad (0, \text{ otherwise}).$$

The $|V| \times |V|$ matrix $\mathbf{A} = (a_{uv})$ is called the (*node-*) *adjacency matrix* of $\mathcal{G}$. Since E is symmetric also $\mathbf{A}$ is symmetric (undirected links) and since E is anti-reflexive, diagonal entries of $\mathbf{A}$ are zeros (no self-links). In an undirected graph, the u-th row/column of $\mathbf{A}$ has k_u entries, the number of nearest neighbours of node u. As any square matrix can be understood as a graph, if we let $a_{uv} = w_{uv}$ be the weight of an edge, then $\mathbf{A}$ induces a *weighted graph*.

The spectrum of the adjacency matrix is related to the structure of the network. The spectrum $\sigma(\mathbf{A})$ is the set of the m distinct eigenvalues of $\mathbf{A}$, with their multiplicities,

$$\sigma(\mathbf{A}) = \begin{Bmatrix} \lambda_1(\mathbf{A}) & \lambda_2(\mathbf{A}) & \cdots & \lambda_m(\mathbf{A}) \\ \nu[\lambda_1(\mathbf{A})] & \nu[\lambda_2(\mathbf{A})] & \cdots & \nu[\lambda_m(\mathbf{A})] \end{Bmatrix}.$$

The eigenvalues $\lambda(\mathbf{A})$ are the zeros of the characteristic polynomial $|\lambda\mathbf{I} - \mathbf{A}|$ and satisfy $\mathbf{A}\boldsymbol{\varphi} = \lambda(\mathbf{A})\boldsymbol{\varphi}$, where each of the non-zero vectors $\boldsymbol{\varphi}$ is an eigenvector. As $\mathbf{A}$ is non-negative, its eigenvalues are all real and the largest one $\lambda_1(\mathbf{A})$ is non-negative. $\lambda_1(\mathbf{A})$ has multiplicity one and a positive eigenvector $\boldsymbol{\varphi}_1(\mathbf{A}) = [\varphi_1^{(1)}, \varphi_1^{(2)}, \dots, \varphi_1^{(n)}]^T$, the Perron-Frobenius principal eigenvector. The spectrum of $\mathbf{A}$ makes partitions of the nodes that are determined by the sign pattern of the eigenvectors. As $\{\boldsymbol{\varphi}_1^{(i)}\}_{i=1}^n$ are positive, they are interpreted as all nodes are grouped in a single cluster. Eigenvector $\boldsymbol{\varphi}_2$ exhibits a bi-partition as some nodes associate to positive and other to negative entries.

Node-degree and node-degree distribution The *node degree* k_u of node u in $\mathcal{G}(V, E)$ is the number of edges (nearest neighbour nodes) that are incident on that node. If we denote by $\mathbf{1}$ a $|V| \times 1$ vector of ones, we get a column-vector of node degrees, $\mathbf{k} = [k_1, k_2, \dots, k_n]^T = (\mathbf{1}^T\mathbf{A})^T = \mathbf{A}\mathbf{1}$. For weighted networks, the weighted degree (node strength) k_u equals the sum of the weights of links incident with u.

Information about node degrees is used to analyse the structure of the network. The analysis is based on the *node-degree/strength distribution*. Let $p(k) = n(k)/n$ with $n(k)$ the number of nodes having degree k in a network of size n be the probability that a randomly uniformly selected node has degree k. A unimodal probability function can be summarised by $\langle k \rangle = (1/n)\mathbf{1}^T\mathbf{k}$, the *expected node-degree*.

3.2 Metric and structural notions

We can enumerate walks in $\mathcal{G}(V, E)$ by using powers $\mathbf{A}^p$ of its adjacency matrix; Entry $(\mathbf{A}^p)_{u,v}$ counts walks of length p between nodes u and v. $\mathbf{A}^p$ can be expressed from $\sigma(\mathbf{A})$ as $(\mathbf{A}^p)_{uv} = \sum_j \varphi_j^{(u)} \varphi_j^{(v)} \lambda_j^p(\mathbf{A})$. The shortest walk/path between u and v is obtained from powers of $\mathbf{A}$ as the smallest value of p for which $(\mathbf{A}^p)_{uv}$ is non-zero. The smallest through-p-link separation of nodes u and v is regarded to as network distance $d(u, v)$. Pairwise distances $d(u, v)$ can be arranged in a square symmetric matrix $\mathbf{D}$. The maximum entry for a row/column in $\mathbf{D}$ is the *node eccentricity*, it is used to denotes the graph *diameter*.

The notion of *node centrality* is used to determine which nodes are important, according to some criterion. There exists a number of criteria, from a node's ability to communicate with other nodes, to its closeness to many other nodes or its relevance in enabling communication between network parts. When formalised, these properties lead to a host of centrality measures (Freeman, 1979).

Degree centrality The *degree centrality* of a node corresponds to its degree and it measures its ability to directly communicate with the other nodes in the network. It accounts for short-range effects in the network. A node is more influential than another one if its degree is larger.

As $k_u = (\mathbf{A}\mathbf{1})_u = (\mathbf{A})_{uu}^2$ counts the number of walks of length 1 from node u, or the number of walks of length

2 starting and ending at node u. Let the probability of going from one node to another be $(\mathbf{P})_{uv} = (\mathbf{A})_{uv}/k_u$ and let $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_n]^T$ be the stationary vector of the Markov chain with $|V| \times |V|$ transition probability matrix $\mathbf{P}$, we get that the steady-state probability function of the stochastic process on the graph depends on centrality k_u

$$(\boldsymbol{\pi}\mathbf{P})_v = \sum_{v \in V} \pi_v (\mathbf{P})_{uv} = \frac{\sum_{u \in V} k_u (\mathbf{P})_{uv}}{\sum_{u \in V} k_u} = \frac{k_v}{\sum_{u \in V} k_v} = \pi_v.$$

Degree centrality thus models the flow of information through u in an infinitely long random walk. As the sum of weighted traces of the powers of $\mathbf{A}$ (the number of closed walks) converges to the trace of its matrix exponential, we define a measure of well-connectedness, the Estrada index

$$EE(\mathcal{G}) = \mathtt{tr}\big(\mathbf{I}+\mathbf{A}+\mathbf{A}^2/2!+\cdots\big) = \mathtt{tr}(e^{\mathbf{A}}) = \sum\nolimits_{i=1}^{m} e^{\lambda_i(\mathbf{A})}.$$

Closeness and betweenness centrality The *closeness centrality* of a node defines how relatively close that node is from the rest of the nodes in the network. The closeness of a node u is defined to quantify its vicinity to nodes v it is exchanging information with. Closeness is measured in terms of shortest path distances, as inverse of the distance sum of shortest paths $d(u, v)$ from u to all other nodes v,

$$c_u = (n-1)/\sum\nolimits_{v \in V} d(u, v).$$

The *betweenness centrality* specifies how important a node is in communication between other pairs of nodes in the network. The betweenness of node z accounts for the proportion of information that passes through that node in communications between other pairs (u, v) of nodes. The basic definition of betweenness assumes that information flow between two nodes occurs over shortest paths,

$$b_z = \sum_{u \in V} \sum_{v \in V} \frac{f(u, z, v)}{f(u, v)} \quad (u \neq v \neq z),$$

$f(u, v)$ is the number of shortest paths linking u and v and $f(u, z, v)$ is the number of such shortest paths through z.

Katz and eigenvector centrality We expressed node degree as number of walks of length one from that node. This idea is extended to longer walks, so that neighbouring nodes have more influence than distant ones. To combine walks of all lengths, we use a factor α that assigns more weight to shorter walks, as in Katz centrality (Katz, 1953)

$$\begin{aligned} K_u &= \big[(\mathbf{I} + \alpha^{-1}\mathbf{A}^1 + \alpha^{-2}\mathbf{A}^2 + \cdots + \alpha^{-p}\mathbf{A}^p + \cdots)\mathbf{1}\big]_u \\ &= \Big[\sum_{p=0}^{\infty} (\alpha^{-p}\mathbf{A}^p)\mathbf{1}\Big]_u \rightsquigarrow \Big[\Big(\mathbf{I} - \frac{1}{\alpha}\mathbf{A}\Big)^{-1}\mathbf{1}\Big]_u, \alpha \neq \lambda_1(\mathbf{A}) \end{aligned}$$

It is shown by Bonacich (1987) that entries of the principal eigenvector $\boldsymbol{\varphi}_1(\mathbf{A})$ are node centralities in the Katz's sense

$$\Big(\sum_{p=1}^{\infty} \alpha^{1-p}\mathbf{A}^p\Big)\mathbf{1} = \alpha\Big[\sum_j \frac{\lambda_j(\mathbf{A})}{\alpha - \lambda_j(\mathbf{A})}\boldsymbol{\varphi}_j^T\boldsymbol{\varphi}_j\Big]\mathbf{1} \underset{\alpha \uparrow \lambda_1(\mathbf{A})}{=} \boldsymbol{\varphi}_1.$$

Graph Laplacian An important operator defined on a network is the *discrete Laplacian*. Consider an arbitrary orientation of the links in $\mathcal{G}(V, E)$, suppose that for link $[u, v]$ we set u as positive and v as negative end. We can represent $\mathcal{G}$ as the $|V| \times |E|$ oriented incidence matrix $\nabla(\mathcal{G})$,

$$\nabla_{ij} = \begin{cases} +1, & v_i \text{ is the positive end of } e_j \\ -1, & v_i \text{ is the negative end of } e_j \end{cases}; \ (0, \text{ otherwise}).$$

Let $f : V \to \mathcal{R}$ be an arbitrary function which is assigned a value at each node and e an oriented graph link, we understand $\nabla f : E \to \mathcal{R}$ as discrete gradient operator. The operator $\nabla\nabla f : \mathcal{R}^{|V|} \to \mathcal{R}^{|V|}$ is the negative discrete Laplacian. By using the degree vector and adjacency matrix, we can write the Laplacian as $\mathbf{L} = \mathtt{diag}(\mathbf{k}) - \mathbf{A}$. The spectrum $\sigma(\mathbf{L})$ of the graph Laplacian matrix is the set of m distinct eigenvalues of $\mathbf{L}$ with their multiplicities

$$\sigma(\mathbf{L}) = \begin{Bmatrix} \lambda_1(\mathbf{L}) & \lambda_2(\mathbf{L}) & \cdots & \lambda_m(\mathbf{L}) \\ \nu[\lambda_1(\mathbf{L})] & \nu[\lambda_2(\mathbf{L})] & \cdots & \nu[\lambda_m(\mathbf{L})] \end{Bmatrix}.$$

$\mathbf{L}$ is positive semidefinite, with bounded eigenvalues and $\nu[\lambda_1(\mathbf{L})]$ equals the number of connected components.

Let us interpret the graph as a system of $|V|$ equal masses connected by $|E|$ equal springs. The Laplacian of the system is given by its kinetic energy minus the potential function from which force components are defined; As such, it corresponds to the graph Laplacian. The eigenvalues $\lambda(\mathbf{L})$ correspond to (squared) natural frequencies at which the network of harmonic oscillators vibrates in the absence of an external force. The smallest eigenvalue $\lambda_1(\mathbf{L}) = 0$ does not contribute to the vibrational energy, as all nodes move coherently in the same direction. The other eigenvectors represent node displacements due to harmonic oscillation.

4. THE VIIKINMÄKI COMPLEX NETWORK

In this section, we discuss the construction of the complex network of the Viikinmäki wastewater treatment plant from a collection of measurements over a period of about 7 years (Jan 1, 2012-May 7, 2017). We present the main structural features of the Viikinmäki network and their time evolution, hourly, on an illustrative selection of process episodes, on a shorter period (July 11-Aug 11, 2016).

4.1 Process data and network definition

The Viikinmäki network is a sequence of graphs $\mathcal{G}_t(V, E)$ built from online plant measurements collected as hourly averages. The set of nodes consists of $|V| = 314$ process variables; 22×9 variables from the ASP (Table 1), 11×10 from the FLT section and influent concentrations to FLT lines (Table 2), and air concentrations of N_2O and CO_2 in the ventilation channel. The edge set is fixed ($|V| = 3228$): Variables belonging to the same ASP or FLT lines are linked together (for instance, all L_i variables connect to each other) whereas, across lines, only consistent variables are linked together (for instance, L_i-Z6-SS links with L_j-Z6-SS, for all i and j). Edge strengths vary with time, as weights W_t quantify connectivity in terms of correlation between endpoint variables, as repeatedly estimated over a 7-day moving-window that shifts forward at every hour.

The network $\mathcal{G}_t(V, E)$, as observed on July 11, 2016 at 12PM, is depicted in the upper panel of Figure 3, in which a graph layout that places each node on a circle is used. Node relevance is encoded by using for node u a size that is proportional to its degree centrality k_u, whereas the strength of the edges w_{uv} is represented by dying the links on the basis of the current correlation between variables (u, v) at the endpoints (from green to yellow, as correlation increases). In the ASP lines, the nodes with largest degree centralities (that is, nodes that correlate more with neighbouring nodes) are found in the influent

(input and recycle flow-rates) and effluent (concentrations) variables, thus leaving variables associated to control loops as peripheral (DO concentrations and air flow-rates in the reactor zones). As for the FLT lines, the variables that are less central are those associated to the clogging conditions.

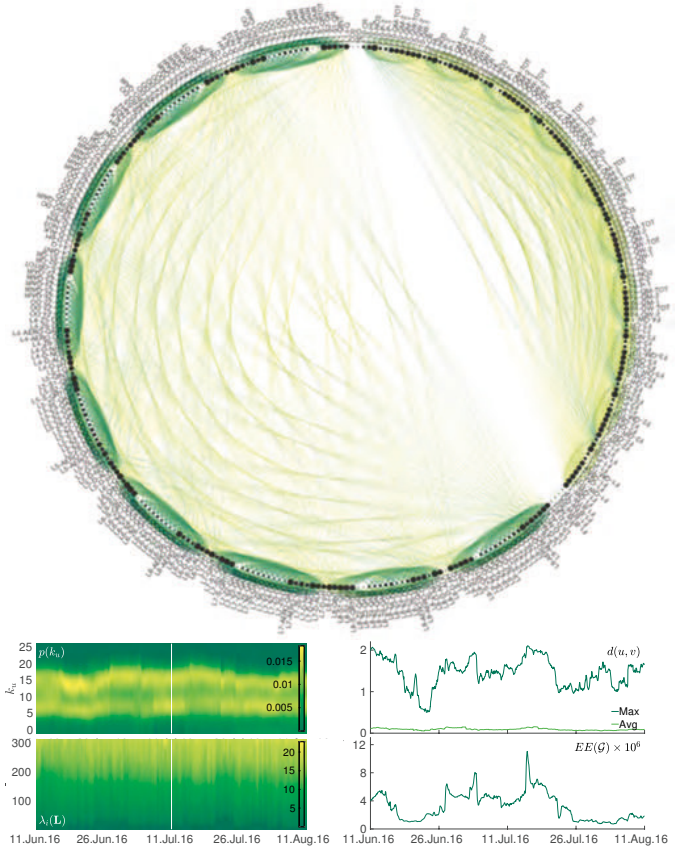

Fig. 3. Viikinmäki. The network of the WWTP (July 11, 2016 at 12PM, upper panel). The middle panels depict: left) the degree probability distribution, and; right) the maximum and average shortest path distances between pairs of nodes, over the given period. The bottom panels depict the Laplacian eigenvalues (left) and the Estrada index (right), during the period.

Figure 3 also displays the degree probability density distribution $p(k)$ (mid-left) and the maximum and average shortest path distances $d(u, v)$ (mid-right), over the selected period. The degree distribution shows a consistent bimodality, as a result of the strategy chosen for linking process variables between process sections. Bimodality becomes less evident as the network diameter, the maximum shortest path distance, gets larger. A similar behaviour can be observed also in terms of the smallest eigenvalues of the discrete Laplacian and the Estrada index (bottom panels).

The degree distribution of the graph observed on July 11 at 12PM (bottom panel of Figure 4) shows how the first mode ($k_u \approx 8$) of the degree distribution relates to ASP nodes, whereas the second one ($k_u \approx 15$) associates to FLT. The upper panels in Figure 4 show the portions of graph corresponding to the variables in the ASP and FLT sections (as of July 11, 2016 at 12PM). The layouts depict the spontaneous equilibrium configuration of a system of attractive forces between joined pairs of nodes and

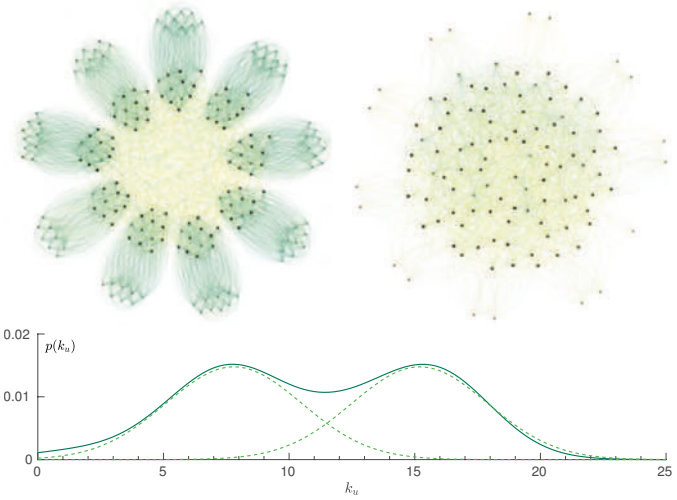

Fig. 4. Viikinmäki, on July 11, 2016 at 12PM. The partial networks for the ASP (upper left) and FLT (upper right) sections. The bottom panel depicts the bimodal node degree distribution (solid line), and the two Gaussian mixture components for ASP and FLT lines.

repulsive forces between all pairs of nodes. Node size shows degree centrality and edges are dyed based to correlation.

4.2 Structural analysis

To discuss the behaviour of the Viikinmäki network, we analyse the node centralities of all variables at both the network-scale (across ASP and FLT lines) and at the subscale of the individual lines, on two illustrative cases.

Figure 5 shows the line (or cumulative) degree- and closeness-centrality of all the ASP lines, over time. As line centralities are computed by summing up the centralities of all the variables belonging to that line, it is expected that lines subjected to similar operating and instrumental conditions be characterised by similar centrality values, along time. The depicted centralities, though characterised by different time series, show a coordinated behaviour of all ASP lines, over most of the period. It is, however, possible to note that, according to degree centrality, two drifts from coordinated operations can be detected. The detachments associate to line L9 ($\approx$ June 20) and line L2 ($\approx$ July 24).

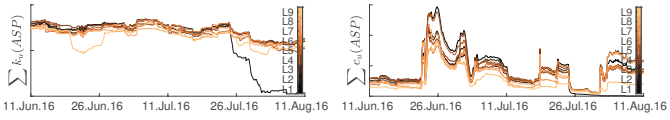

Fig. 5. Viikinmäki, ASP lines. Cumulative degree- (k_u, left) and closeness- centrality (c_u, right) of each line.

Although it is easy to associate the anomaly in L2 to the fact that the line had been shutdown for yearly maintenance, the identification of what induced the detachment of line L2 requires the inspection of the time series of the ASP lines and the centrality of individual variables.

From Figure 6, it is possible to see how, though ASP lines appear to be similarly operated, the degree centrality of L9-D-NO3/-NH4 and L9-D-ALK dropped significantly after June 20. As for the centrality of the other variables, only L9-AERO-QAIR shows an associated, but relatively

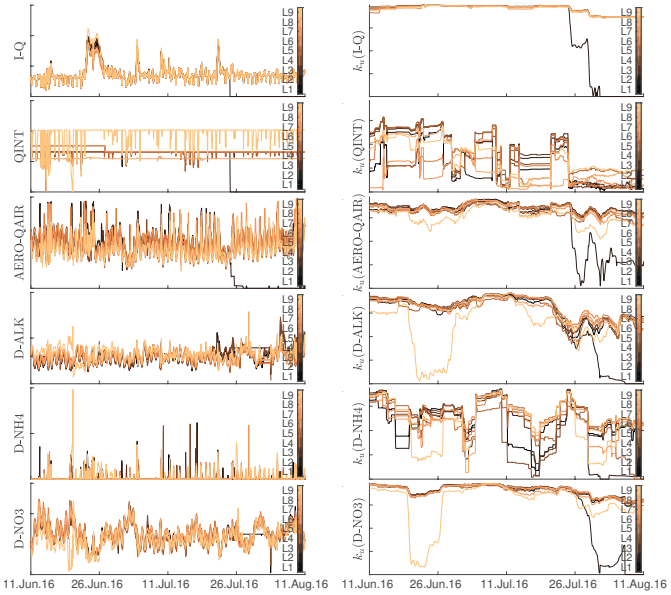

Fig. 6. Viikinmäki, ASP lines. Timeseries (left column) and degree centrality (right column) of selected variables.

smaller, change. A close analysis of the L9-variables shows a rapid variation in alkalinity around June 20. The variation, probably due to an adjustment of lime dosing, coincides with a short break of the internal sludge pump. The episode induced a further drop in alkalinity and a sharp peak in the effluent NH4. The NH4 peak can be associated to the low alkalinity. Hence, the low NO3 concentrations do not result from an efficient denitrification, but rather from a poor nitrification. The poor nitrification caused a persistent change in the diurnal pattern of NO3 that is not easy to visualise from the time series, but can be detected from the centrality. Although not reported, a similar behaviour was found in other centrality measures.

Being downstream along the plant, FLT lines are subjected to homogenised influent conditions and thus, usually, they are similarly operated. Figure 7 shows this coordinated operation of the filter lines in terms of total influent flow-rate (QM-1 + QM-2) and total methanol (QW-1 + QW-2).

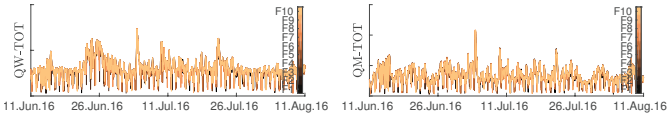

Fig. 7. Viikinmäki, FLT. Timeseries of selected variables.

This similarity is also reflected in the cumulative degree centralities (not shown). It is interesting to analyse how the node centrality of the process variables within an individual filter varies in response to changing conditions. As an illustrative example, consider the increase in influent flow-rate to the plant (Figure 6, influent flow-rate I-Q

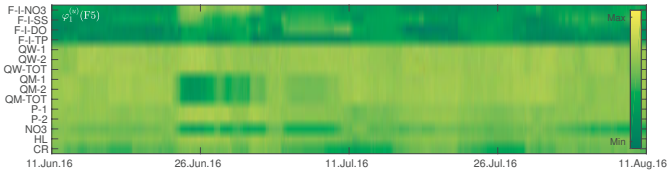

Fig. 8. Viikinmäki: FLT F5. All eigenvector centralities.

to ASP lines) on June 20, 2016. The variation induced a change in nitrogen removal efficiency in the ASP, as detected by the cumulative closeness centrality (Figure 5) and a subsequent effect in methanol dosage in the FLT line(s). In Figure 8, this is clearly visible in terms of the eigenvector centrality of the process variables in filter F5.

5. CONCLUSIONS

A large-scale wastewater treatment plant has been mapped onto a complex network and the structural characteristic of the resulting temporal graph have been studied, as the plant is subjected to routine operations. This qualitative study highlights how variations in structural graph properties, mainly diameter and node centralities, correspond to changes in operating and instrumental conditions in the plant. We speculate that such properties of the graph relate to hidden dynamical states of the plant and deserve to be further studied with system analytical techniques.

REFERENCES

J.S. Andrade Jr., D.M. Bezerra, J. Ribeiro Filho, and A.A. Moreira. The complex topology of chemical plants. *Physica A*, 360(2):637–643, 2006.

A.-L. Barabási. The network takeover. *Nature Physics*, 8: 14–16, 2012.

A. Barrat, M. Barthelemy, and A Vespignani. *Dynamical processes on complex networks*. Oxford University, 2008.

P. Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.

H. Crane. *Probabilistic Foundations of Statistical Network Analysis*. CRC, 2018.

R. Diestel. *Graph Theory*. Springer, 2010.

S.N. Dorogovtsev. *Lectures on Complex Networks*. Oxford University, 2010.

E. Estrada. *The Structure of Complex Networks: Theory and Applications*. Oxford University, 2016.

L.C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3):215–239, 1979.

E.D. Gilles. A network theory for chemical processes. *Chemical Engineering Technology*, 21(2):121–132, 1998.

L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

J.P. Koeln and A.G. Alleyne. Robust hierarchical model predictive control of graph-based power flow systems. *Automatica*, 96:127–133, 2018.

E.D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer, 2009.

R. Lambiotte and N. Masuda. *A Guide to Temporal Networks*. WSPC, 2016.

M. Mangold, S. Motz, and E.D. Gilles. A network theory for the structured modelling of chemical processes. *Chemical Engineering Science*, 57(19):4099–4116, 2002.

M.R. Maurya, R. Rengaswamy, and V. Venkatasubramanian. Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets. *Engineering Applications of Artificial Intelligence*, 17(5):501–518, 2004.

M. Newman. *Networks*. Oxford University, 2018.

H.A. Preisig. A graph-theory-based approach to the analysis of large-scale plants. *Computers and Chemical Engineering*, 33(3):598–604, 2009.