TECHNICAL Report:
Simulation of Fence Graph Causal Digital Twin and Counterfactual Experiments

Dr. PG Madhavan
Seattle, WA
September 2021

Theory, Algorithm & explanation at –
"Multichannel IoT Causal (MIC) digital twin: Counterfactual experiments on Fence Graphs"

https://pgmad.medium.com/multichannel-iot-causal-mic-digital-twin-counterfactual-experiments-on-fence-graphs-df884a9f5f35

CAUSALITY & ROOT-CAUSE ANALYSIS

"*Knowing the real root causes of events is critical to resolving problems rather than continuously dealing with the symptoms*".

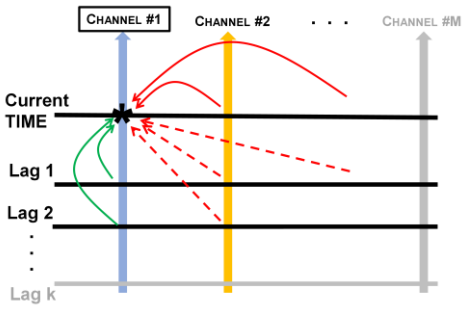
Condition monitoring is not new . . . from time immemorial, people have used their senses to assess the condition of their tools and machines. Once a condition of concern is detected, the next step is root-cause analysis; experts will get together, propose some hypotheses for what went wrong and conduct a series of simple experiments. Of course, this is an off-line cumbersome activity; also, the accuracy of root cause identification is limited by the time available and cleverness of the individual experts involved.

In my recent article, "Multichannel IoT Causal (MIC) digital twin: Counterfactual experiments on Fence Graphs", (August 2021), I discussed the power of counterfactual experiments for eliciting causes and their effects. Important solutions provided in this paper are the proper way to (1) Process multichannel data, (2) Estimate inter-channel (and "self-channel") cause-effects resulting in a (3) Multichannel IoT Causal (MIC) digital twin.

The reason to adopt MIC digital twin approach to move on beyond Condition Monitoring to root-cause analysis is that quantitative Cause-Effect metrics extracted from multiple channels of data alone will enable *automated root-cause analysis and optimization of processes that results in boosting productivity and efficiency.*

Any significant IoT deployment employs multichannel data collection . . . *Because end points that are monitored have some relationships*! Almost all current IoT applications today process single channels, one at a time; that may be sufficient for condition monitoring purposes . . .

Here is a picture of multichannel data. Each channel corresponds to an end-point being monitored. The RED arrows show that inter-channel relationships are NOT ACCOUNTED for in single channel processing; this has consequences!



- Single channel results can be erroneous
- Ignores CAUSE-EFFECT relationships among end-points

MIC digital twin is one way to avoid these missteps and automate root-cause analysis (using rigorous statistical estimation theory).

An overall solution is "FENCE Causal Graph" which is fully developed in my article linked above.

We need the FULL PICTURE to take IoT beyond Condition Monitoring to Causal Estimation if we are to deliver operations and productivity improvement prescriptions to business owners.

SVAR ESTIMATION

Multichannel IoT data (time series) is modeled as a Structural Vector Auto Regressive (SVAR) model.

Vector time series, $\mathbf{y_t}$, has the following model:

$$\mathbf{y_t} = \mathbf{S^0 y_t} + \sum_{\mathbf{d=1}}^{\mathbf{D}} \mathbf{S^d}\, \mathbf{y_{t-d}} + \mathbf{e_t} \qquad \ldots (1)$$

The first term is the "structural" or "instantaneous" causal model (SCM) since lagged data are not involved – this is the model of a "snap-shot" in time across channels. The second term accounts for the effect of lagged time series (lagged up to D lags) on "within" and "across" channels.

Given a multichannel vector time series, the unknown parameters, $\mathbf{S^0}$and $\mathbf{S^d}$ (d= 1 . . . D), are estimated in four steps as follows.

<u>Step 1:</u> VAR Estimation (Kalman Filter)

$$\boldsymbol{y}_{\mathrm{t}} = \sum_{\mathbf{d=1}}^{\mathbf{D}} \mathbf{M^d}\, \mathbf{y_{t-d}} + \mathbf{n}_{\mathrm{t}} \rightarrow \textit{Estimate } \widehat{\boldsymbol{M}}^{1..D}$$

<u>Step 2:</u> Get Residuals

$$\widehat{\mathbf{n}}_{\mathrm{t}} = \boldsymbol{y}_{\mathrm{t}} - \sum_{\mathbf{d=1}}^{\mathbf{D}} \widehat{\mathbf{M}}^{\mathbf{d}}\, \mathbf{y_{t-d}} \rightarrow \textit{Calculate } \widehat{\boldsymbol{n}}_t$$

<u>Step 3:</u> SCM Estimation (using ICA)

$$\widehat{\mathbf{n}}_{\mathrm{t}} = \mathbf{S}^0 \widehat{\mathbf{n}}_{\mathrm{t}} + \mathbf{e}_{\mathrm{t}} \rightarrow \text{Estimate } \widehat{\mathbf{S}}^0$$

<u>Step 4:</u> Get pure VAR

For d = 1. . .D,

$$\widehat{\mathbf{S}}^{\mathbf{d}} = \left(\mathbf{I} - \widehat{\mathbf{S}}^{\mathbf{0}}\right) \widehat{\mathbf{M}}^{\mathbf{d}} \rightarrow \text{Calculate } \widehat{\mathbf{S}}^{1..\mathrm{D}}$$

As an example, the derivation of $\widehat{\mathbf{S}}^{\mathbf{1}}$ in the last step is shown below:

<u>SVAR equation: *d=1 case*</u>

$$\left(\mathbf{I} - \mathbf{S^0}\right) \mathbf{y_t} = \mathbf{S^1}\, \mathbf{y_{t-1}} + \mathbf{e_t} \quad \textbf{Or} \quad \mathbf{y_t} = \left(\mathbf{I} - \mathbf{S^0}\right)^{-1} \mathbf{S^1}\, \mathbf{y_{t-1}} + \mathbf{e_t}$$

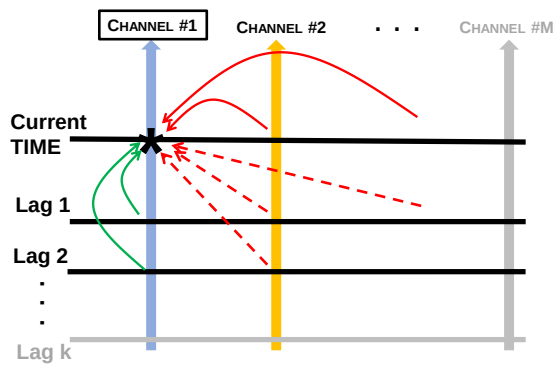
$$\textit{Using estimates}, \quad \mathbf{y_t} = \widehat{\mathbf{M}}^{\mathbf{1}}\, \mathbf{y_{t-1}} + \widehat{\mathbf{n}}_{\mathrm{t}}$$

Comparing, we get $\widehat{\mathbf{M}}^{\mathbf{1}} = \left(\mathbf{I} - \widehat{\mathbf{S}}^{\mathbf{0}}\right)^{-1} \widehat{\mathbf{S}}^{\mathbf{1}}$

Or $\widehat{\mathbf{S}}^{\mathbf{1}} = \left(\mathbf{I} - \widehat{\mathbf{S}}^{\mathbf{0}}\right) \widehat{\mathbf{M}}^{\mathbf{1}}$

FENCE CAUSAL GRAPH

In the multichannel picture on the right, the SOLID-RED arrows are "inter-node" **structural** (INS) causal factors and DOTTED-RED are "inter-node" **lagged** (INL) causal factors. GREEN arrows are the traditional AR(p) coefficients.



In the state-of-the-art IoT signal processing, only the GREEN arrows are considered and estimated, and all the RED arrows are ignored!

Referring to the last section,

- $\hat{\mathbf{S}}^{\mathbf{0}}$ matrix elements are the estimates of the INS causal factors (SOLID-RED arrows)
- $\hat{\mathbf{S}}^{\mathbf{d}}$ matrix elements are the estimates of the lagged – both self and inter-node – causal factors (DOTTED-RED & GREEN arrows)

We represent the INS and lagged causal factors in a "FENCE" graph. We use the example of a real data test to explain the graph.

Test in a real-life setting was done using the popular NASA Prognostics Data Repository's Bearing Dataset. The data is from a run-to-failure test setup of bearings installed on a shaft.

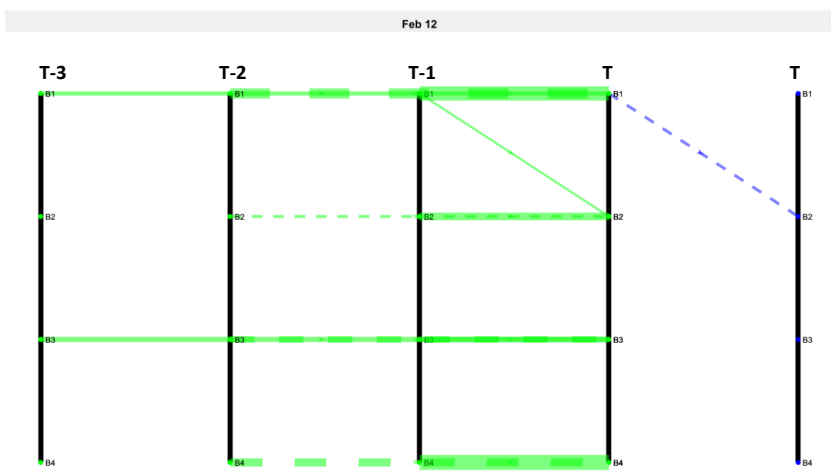


Figure 1. Fence Graph for Day-1 of testing – pristine condition

The five "fence posts" represent the time index, current time on the extreme right marked as "T", replicated again to be able to map "inter-node" **structural** (INS) causal factors. The rest are increasing lags from T-1 to T-3. In general, if there are M channels and p_L is the largest model order of AR(p) models among all of the single channel time series, there will be M horizontal rungs in each panel and (p_L + 2) vertical fence posts.

The four bearings, B1 to B4, are labeled vertically on each fence post with B1 at the top. BLUE lines indicate INS causal factors and GREEN indicate "inter-node" **lagged** (INL) causal factors. Solid line is for

positive causal factors and dotted line for negative (Cause will reduce the Effect). Line width is proportional to the causal factor magnitude.

You can see that there is a dotted BLUE line on the right-most panel; this means that at the current instant, B1 has a negative (instantaneous) effect on B2. In the (T1-T) panel, we see 4 solid GREEN horizontal links – these are the first coefficients of AR(3) model of each of the 4 time series corresponding to B1 to B4 bearing vibration data. What is shown in figure 1 is for Day-1 of the test when all the mechanical components are in pristine condition – hence there are minimal causal links that are not horizontal (horizontal links will always be present in the lagged panels unless the data is white noise).

On "D-day", B1 bearing started showing clear signs of failure. Fence Graph for a time block on that day is shown in figure 2. Bearing, B1, is the node on the very top.

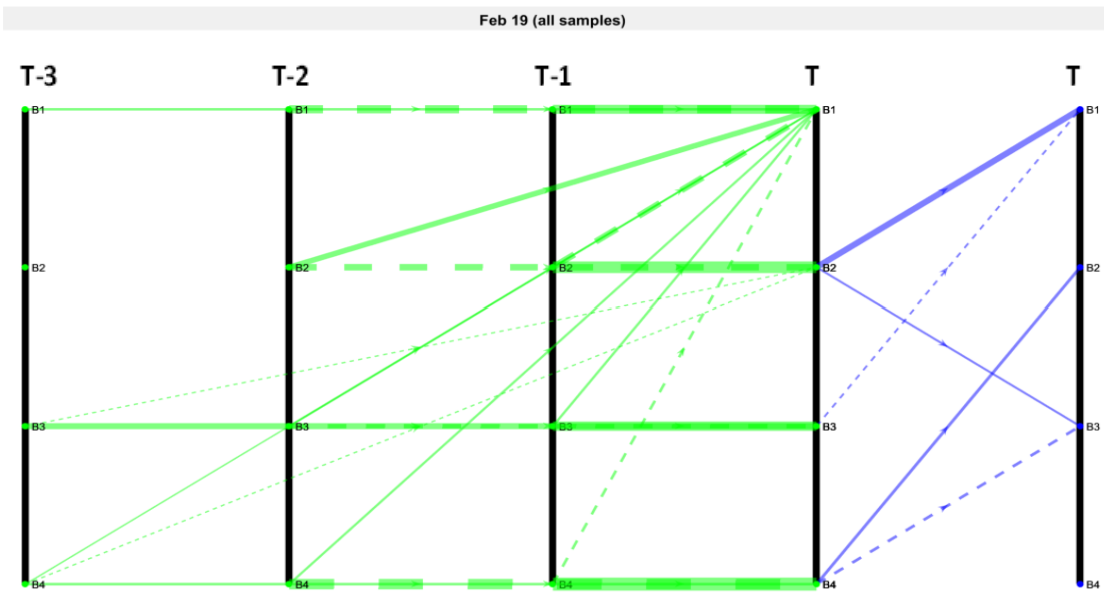
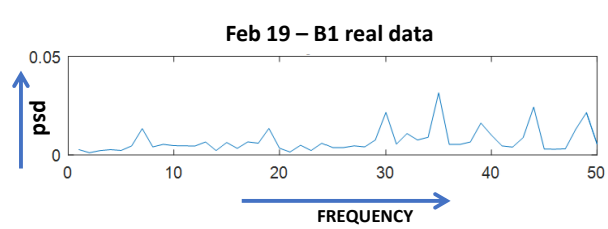


Figure 2. B1 failure on D-day (B1 – top node)

It is amply clear that B1 was attracting a lot of attention on D-day! One positive INS from B2 and INL from virtually every bearing at all lags were affecting B1 – no wonder it failed . . .

It is very revealing (to machine dynamics experts) to study the power spectral density of B1's vibration time series.



A machine vibration expert may tell you that the spectrum peak on Day-1 (corresponding to figure 1) is the "running frequency" driven by the A.C. induction motor of the test rig; the power at higher frequencies on D-day (corresponding to figure 2) is called "bearing frequencies" in the figure on the right that indicate roller element and inner/outer race defects; running frequency also exhibits harmonics at this time.

SIMULATION BASICS

Equation (1) the previous section can be re-written as –

$$y_t = (I - S^0)^{-1} \sum_{d=1}^{D} S^d y_{t-d} + (I - S^0)^{-1} e_t \qquad \ldots (2)$$

At this point, we have estimates, $\hat{\mathbf{S}}^{\mathbf{0}}$ and $\hat{\mathbf{S}}^{\mathbf{d}}$, the link strengths (causal factors) of the Fence Graph in hand. Using the PAST samples, $\boldsymbol{y_{t-d}}$, we generate current sample estimates, $\mathbf{y}_t^{est}$.

$$\mathbf{y}_t^{est} = \left(\mathbf{I} - \hat{\mathbf{S}}^0\right)^{-1} \sum_{d=1}^{D} \hat{\mathbf{S}}^d \mathbf{y}_{t-d} \qquad \ldots (3)$$

Once we have $\mathbf{y}_t^{est}$ in hand, we can perform very important tests to assure the quality of our estimation of Causal Structure and Causal Factors.

In equations (1) or (2), SVAR model assumes that $\mathbf{e_t}$ are NON-Gaussian and Independent – these assumptions are required for estimation <u>Step 3:</u> SCM Estimation (using ICA).

Residual Test:

Residual, $\hat{\mathbf{r}}_t = \mathbf{y_t} - \widehat{\boldsymbol{y}_t^{est}}$. Comparing to equation (2), we can estimate –

$$\hat{\mathbf{e}}_t = \left(\mathbf{I} - \hat{\mathbf{S}}^{\mathbf{0}}\right) \hat{\mathbf{r}}_t$$

If $\hat{\mathbf{e}}_t$ is non-Gaussian and independent, we are assured that our Fence Graph is statistically valid.

EXAMPLE:

Consider Fence Graph in figure 2 on D-day when Bearing 1 failed. Using the estimates, $\hat{\mathbf{S}}^{\mathbf{0}}$ and $\hat{\mathbf{S}}^{\mathbf{d}}$ in equation (3), we obtained $\mathbf{y}_t^{est}$. The results are shown in figure 3.

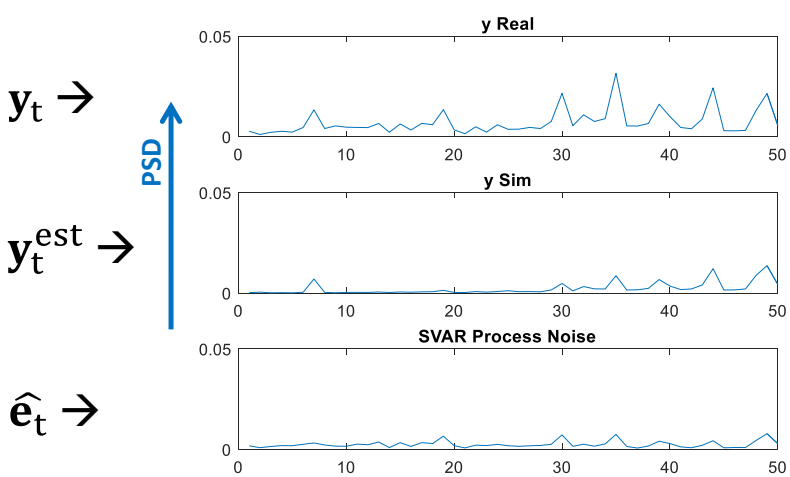
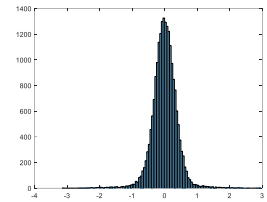


Figure 3. Power Spectral Density (PSD) of Bearing 1 real data, simulated data and model error

The features of the PSD of Bearing 1 real data was discussed elsewhere. Simulated Bearing 1 data in the middle panel shows similar spectral form (with a lower power). The estimated $\widehat{\mathbf{e}}_t$ has a nearly flat spectrum indicating uncorrelated $\widehat{\mathbf{e}}_t$.

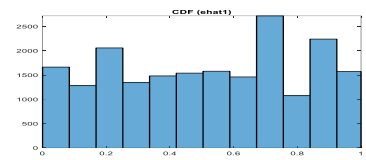
The histogram of $\widehat{\mathbf{e}}_t$ is as shown on the right.



Estimated Kurtosis = 5.7327

$\Rightarrow$ $\widehat{\mathbf{e}}_t$ is NON-Gaussian (also called "super" Gaussian)

It is known that if $\widehat{\mathbf{e}}_t$ is mapped using its CDF resulting in a Uniform distribution (implying $\widehat{\mathbf{e}}_t$ has maximum entropy), $\widehat{\mathbf{e}}_t$ is Independent.



From the histogram on the right, we see that **CDF** ($\widehat{e}_t$ of B1) → approx. UNIFORM.

$\Rightarrow$ $\widehat{\mathbf{e}}_t$ is INDEPENDENT

COUNTERFACTUAL EXPERIMENT USING FENCE CAUSAL GRAPH

We will consider an example to illustrate Counterfactual experiments on Fence Graph.

There two approaches:

1. A machine vibration analyst who is an expert may be able to suggest various changes in link strengths (causal Factors) to see how Bearing 1 failure can be prevented.
2. The interaction of various causal factors can be automated in a major optimization exercise since it is likely that causal factor changes to one link will interact with other links – TUNING the causal factors simultaneously in a large optimization run will become a compelling need as the graphs get bigger.

CASE 1: VARY LINK MAGNITUDES

Multiple Fence Graph links were removed as shown in figure 3. The expectation from this counterfactual experiment – "counterfactual" because links=0 was not a fact – is that B1 vibration data will not exhibit the high-frequency peaks (which are known to be indicators of imminent failure).

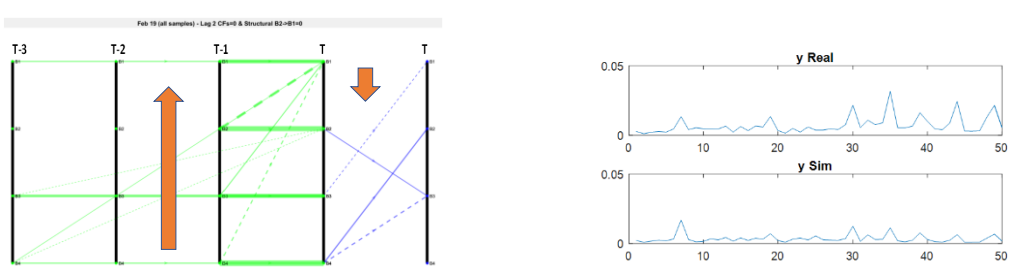


Figure 3. Links from Lag 2 to B1 & Structural link B2->B1 = 0 (indicated by RED arrows)

Power spectral density of the real B1 data and simulated B1 data are shown on the right. While high-frequency peaks have diminished, they have not disappeared. But a vibration pattern with this spectrum will look mostly within safe limits.

It is not clear that our expectation of this counterfactual experiment was fully met. For example, there may be other key links whose weight changes may affect B1 data more dramatically. Also, a large-scale optimization experiment may reveal a new pattern of links that minimizes B1 vibration which can become the basis of optimized operation.

CASE 2: OPTIMIZE LINKS TO ACHIEVE DESIRED OUTCOME BEHAVIOR

This is a very significant case. In addition to exploratory and manual "what-if" analysis under Case 1, we can perform directed counterfactual simulation to programmatically determine desired outcomes!

In this simulation, we want the behavior of Bearing 1 on Feb 19 (which is failing then) to be that of a pristine bearing (such as on Feb 12 on Day-1 of the NASA test). From the discussion of the significance of spectral peaks earlier (after figure 2), we know that if the spectrum shows only a peak at "running" frequency and NO "bearing" frequencies at higher frequencies, the bearing is behaving in a pristine condition.

Objective of this counterfactual experiment is to determine the Fence Graph link strengths that will render B1 vibration data appear pristine (only "running frequency" peaks) on D-day.

We proceed as follows:

- Create a modified dataset where Bearing 2, 3 and 4 data are identical to the original D-day data but replace B1 data with "running frequency" sine was plus noise.
- Redo SVAR estimation.
- The resulting Fence Graph gives the link connections and strengths that will render B1 data pristine. The result is shown in figure 4.

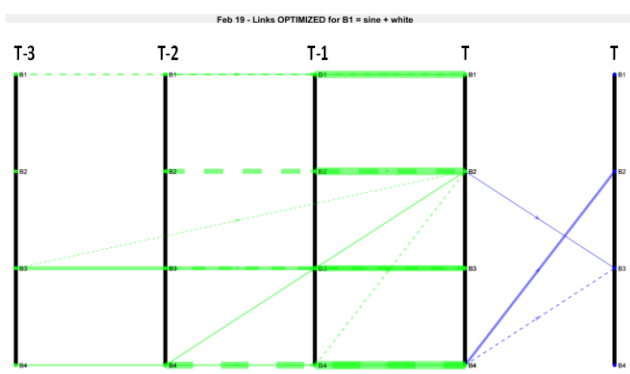


Figure 4. Fence graph required for B1 data have only "running frequency" on D-day

Comparing figure 4 to figure 2 which showed the actual Fence Graph on D-day, we can see that many links have vanished, link patterns and link strengths have changed.

The spectrum of the simulation result is shown in figure 5 and compared to actual spectrum from Day-1. They are almost identical.

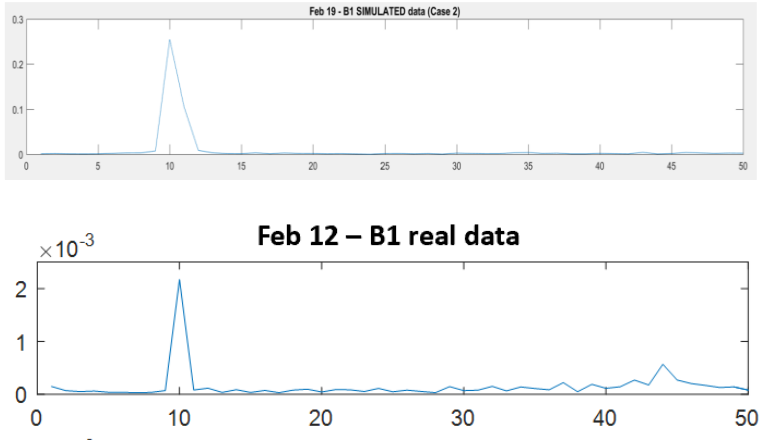


Figure 5. D-day B1 simulated data spectrum on top; actual B1 data spectrum on Day-1 on the bottom

Our counterfactual simulation objective of determining the Fence Graph link strengths that will render B1 operation pristine on the day it failed (D-day) has been met.

We need domain experts to convert these powerful insights into actionable steps such as what are the physical changes needed corresponding to the new link configuration? *Fixes such as shaft alignment or equipment redesign or . . .*

With Fence Causal Graph in hand the ability to perform sophisticated simulations such as Case 2, domain experts can suggest various changes to prevent failures.

Analytical configuration determination of the Fence Graph as demonstrated in Case 2 will lead to speedy and reliable optimization of the connected asset's overall performance leading to increasing productivity.

Dr. PG Madhavan
https://www.linkedin.com/in/pgmad/

#IoT #Simulation #Multichannel #Digitaltwin #Causaldigitaltwin #Fencegraph #Causality