Supplementary Information of Dimension Reduction Approach for Understanding Resource-Flow Resilience to of Climate Change

Ariel Salgado[1,*], Yiyi He[1], John Radke[1], Auroop Ganguly[2], and Marta C. Gonzalez[1,+]

[1]Department of City and Regional Planning, University of California, Berkeley, CA 94720, USA
[2]Sustainability and Data Sciences Laboratory, Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA
[3]Pacific Northwest National Laboratory, Richland, WA, USA
[4]Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, USA
[5]Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, USA
*arielolafsalgado@gmail.com
+martag@berkeley.edu

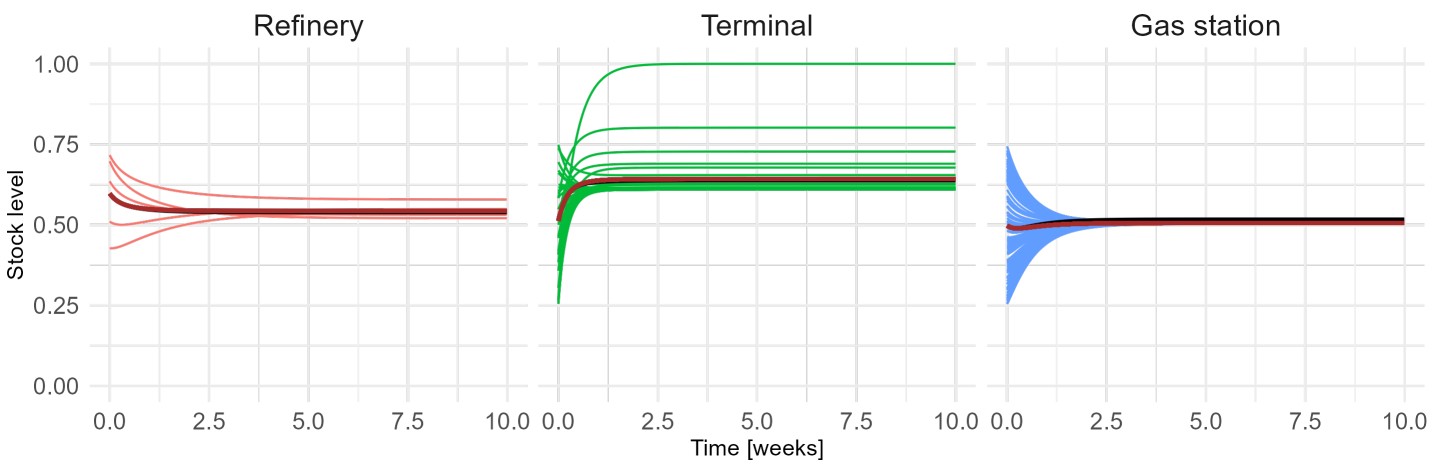
Supplementary Note 1: Numerical testing in San Francisco Fuel Transportation network under different connection strategies

Similarly to the approximation in [1], it is not easy to obtain a general upper bound for the $|x_i^q - y^q|$, as it is heavily dependent on the network structure. However, we note that comparison should be made between y^q (calculated using the approximation) and $\frac{\sum_{i=1}^{N_q} C_i^q x_i^q}{\sum_{i=1}^{N_q} C_i^q}$ (calculated using the exact equations), which is the exact stock level at layer q. To test the approximation, we consider its application under different network structures based on the original geographical network presented in [2] and discussed in the current manuscript at the start of the Results section. The network includes $N_1 = 5$ refineries, $N_2 = 29$ terminals and $N_3 = 3422$ gas stations. Refineries and terminals link through a heavily interconnected system of pipelines, that on the scale of data we have (recall that our flow and production capacities are in $Mgal/week$) allow transferring resource between any pair of them. Thus, we always consider the layers $q = 1$ (refineries) and $q = 2$ (terminals) as fully connected. Layer $q = 3$ (gas stations) connects to refineries and terminals through the road network only. Thus, we explore the impact on the approximation of changes in the number of links to refineries a gas station has. As the cost of transportation is heavily associated to distance travelled, we choose to connect gas stations to terminals and refineries based on the road network distance (i.e. taking into account the length of the route in the road network). We construct networks with constant in-degree in the gas station layer, having $n_r = 1, \ldots, 5$ connections to refineries and $n_t = 1, \ldots, 29$ connections to terminals. For each gas station, we connect it to the n_r nearest refineries and n_t nearest terminals. As our data is on the layer level, we construct the networks to be consistent with our reference values of W^{12}, W^{13}, W^{23} and $N_1 P = N_3 D$ (layer-level flow, production and demand capacities). Thus, on average the production capacity of a refinery is P, the demand capacity at gas station is D and the average flow capacity is W^{qr}/M_{qr}, where M_{qr} is the number of links between layer q and r. We consider two situations:

- In case (1) we assign a value equal to the average for each parameter: $P_i = P$, $D_i = D$ and $W_{ij}^{qr} = W^{qr}/M_{qr}$ if x_i^q is linked to x_j^r and zero otherwise.

- In case (2), we sample the values from a normal distribution with deviation equal to a ten percent of the average value: $P_i \sim \mathcal{N}(P, 0.1P)$, $D_i =\sim \mathcal{N}(D, 0.1D)$ and $W_{ij}^{qr} \sim \mathcal{N}(W^{qr}/M_{qr}, 0.1W^{qr}/M_{qr})$ if x_i^q is linked to x_j^r and zero otherwise.

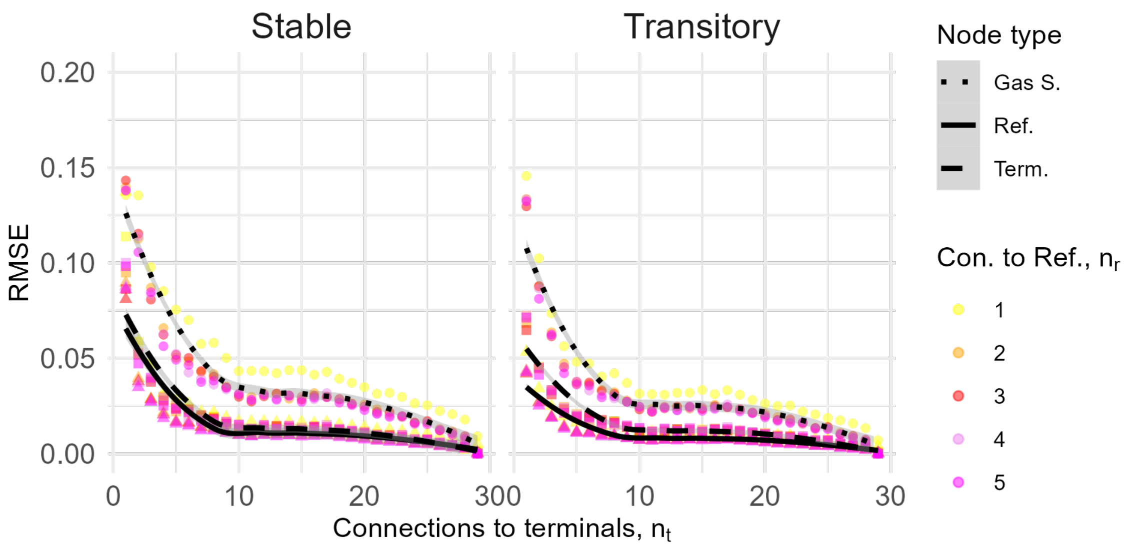
Using these networks, we simulate the evolution of facilities and use the simulation to compute the observed layer stock level as $\sum_{i=1}^{N_q} C_i^q x_i^q / \sum_{i=1}^{N_q} C_i^q$. Then, by calculating the parameters p, d and s_{qr} we simulate the approximated layer stock level y^q. Supplementary Figure 1 shows an example of how this simulation looks like. We can see that while intra-layer variance can be very high, the approximation captures reasonably well the dynamics of the layer stock level. To compare approximation and exact result, we sample initial conditions for every variable $x_i^q(0) \sim U[0.25, 0.75]$ ten times, and calculate the average error over them. We consider two measures of error: how well the stable level is determined and how well is the transition from initial condition to stable state captured. The results are shown in Supplementary Figures 2 and 3 for cases (1) and (2) respectively. Notice that in both cases, we use the same parameters for the layer stock level. In both Supplementary Figures we see that as the number of connections between gas stations and terminals increases, the error reduces until reaching an average error below 0.05 for the gas station layer and 0.025 for the terminals and refineries in both cases.



Supplementary Figure 1: Simulation example with $N_1 = 5$ refineries, $N_2 = 29$ terminals and $N_3 = 3422$ gas stations. Refinery-Terminal layer is fully connected, and gas stations connect to 4 refineries and 26 terminals. The broad brown line represents the exact layer stock level $\sum_{i=1}^{N_q} C_i^q x_i^q / \sum_{i=1}^{N_q} C_i^q$, while the black line corresponds to the approximation of y^q through Eq. 2 in the main article. The simulation setting considers that every parameter is equal to the average value for the corresponding capacity (case 1 described in the main text).

Supplementary Note 2: Numerical testing of the reduced-dimension representation in random tripartite networks

To test the approximation in other structures, we construct random tripartite networks (i.e., networks where there are three layers of nodes, and connections only exists between nodes in different layers), and assign one layer to production (e.g. refineries), other to momentary storage (e.g., terminals), and other to consumption (e.g., gas stations). We consider different sizes for the layers, and different values for the flow capacities. Our testing procedure is similar to one used in [3], in the sense that we consider the Stochastic Block Model (SBM) as benchmark network model for testing. We use the same functions Π, Ψ and Δ used to represent the San Francisco Fuel Transportation Network:

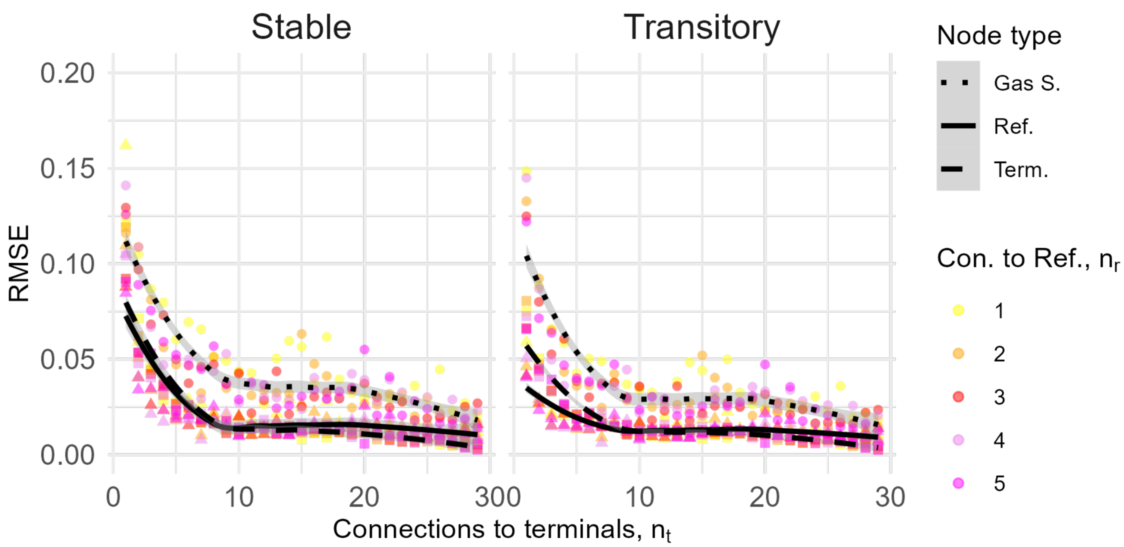


Supplementary Figure 2: Approximation error $\sqrt{\sum_t (y^q(t) - \sum_{i=1}^{N_q} C_i^q x_i^q(t) / \sum_{i=1}^{N_q} C_i^q)^2}$ for case (1), all parameters are equal to their average values. Stable: the only time considered is 20 weeks, where the system has already stabilized. Transitory: the error is considered only in the first part of the evolution, for times between 0 and 5 weeks. Circles correspond to the gas station layer, squares correspond to terminals and triangles to refineries.

$$\begin{aligned} \Pi(x) &= (1+10^{-6})(1-x)/(1+10^{-6}-x) \\ \Delta(x) &= (1+10^{-6})x/(x+10^{-6}) \\ \Psi(x_1,x_2) &= x_1(1-x_2) \end{aligned} \quad (1)$$

By exploring different layer sizes, variability in flow capacities and number of connections, we observe that the layer average provides a good estimator of the true average stock level in most of the cases. As the number of nodes in a given layer increases, the estimation error decreases. We observe that variability in weights induces variability in the stable stock levels of the nodes, but the impact on the average is small. The biggest effect is due to the number of connections per layer: in more sparse networks the difference between the true average and the estimation increases. However, the difference is smaller than the actual variability between the nodes within the same layer (represented through their standard deviation). We find that the error increases due to the increase in the number of nodes that either are disconnected from refineries, or disconnected from gas stations. This nodes are unable to transfer the resource they have (or don't receive any incoming resource), and thus are stuck outside the dynamics. Thus, from the perspective of representing a fully functional network, the approximation captures well the behavior of the system under Eq. 1. The model to construct the networks to test the approximation can be summarised as follows:

- Select layer sizes N_1, N_2 and N_3.
- Sample each possible connection between layers q and r ($q \neq r$) with probability u. The resulting network will have, on average, uN_qN_r connections between layers q and r.
- Production capacity P is set to 1 at every node in layer 1, and demand capacity D is set equal to N_1/N_3.



Supplementary Figure 3: Approximation error $\sqrt{\sum_t (y^q(t) - \sum_{i=1}^{N_q} C_i^q x_i^q(t) / \sum_{i=1}^{N_q} C_i^q)^2}$ for case (2), parameters are sampled from a normal distribution with average equal to the global average and deviation equal to a 10% of the average. Stable: the only time considered is 20 weeks, where the system has already stabilized. Transitory: the error is considered only in the first part of the evolution, for times between 0 and 5 weeks. Circles correspond to the gas station layer, squares correspond to terminals and triangles to refineries.

- Sample flow capacities W_{ij}^{qr} for each layer uniformly from the interval $[1 - dW, 1 + dW]$, with $dW < 1$.

Thus, the total flow capacity for the random network is $W^{qr} \approx uN_qN_r$ on average. Our focus in on comparing the exact layer average $< x^q > (t) = \sum_{i=1}^{N_q} x_i^q / N_q$ with the evolution obtained through the approximation for y^q.

Probability of connection

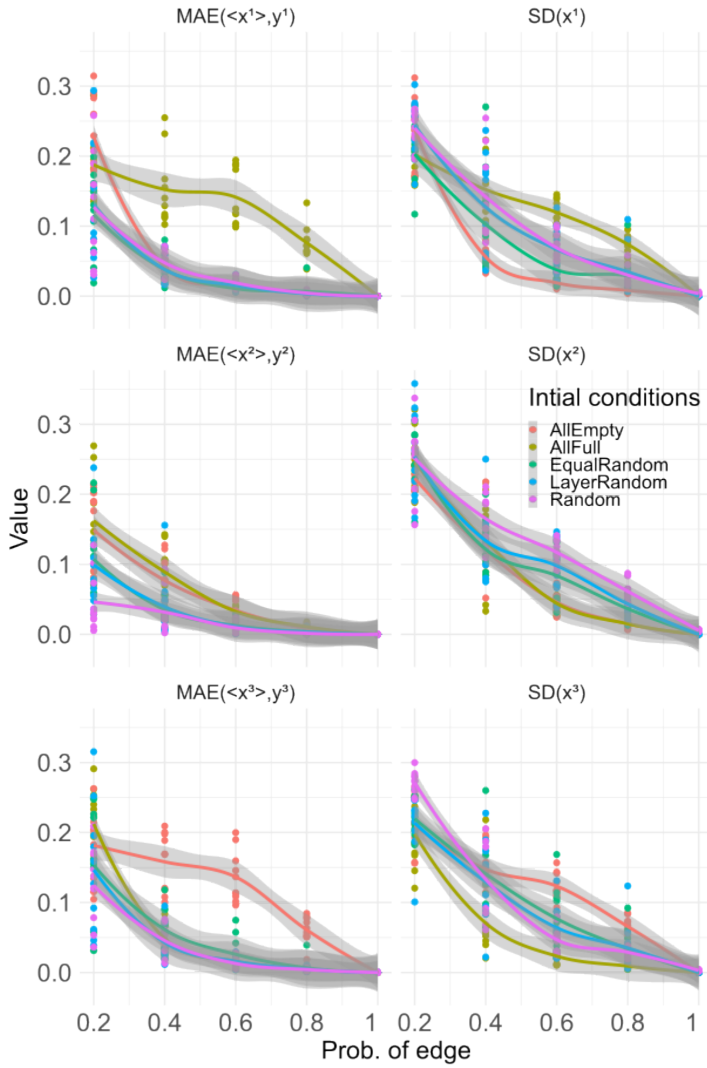
We explore the impact of changing the connection sampling probability u through two magnitudes: the deviation within a single layer, $SD(x^q) = \sqrt{(\sum_{i=1}^{N_q} (x_i^q - < x^q >)^2)}$, and the mean absolute error $MAE(< x^q >, y^q) = | < x^q > - y^q|$. In every simulation, we select the initial condition of the layer average to be consistent with the true average, i.e. $y^q(t = 0) =< x^q >$ $(t = 0)$. We consider $N_q = 10$ and $W_{ij}^{qr} = 1$. This way, we isolate the impact of changes in the connectivity. The results are presented in Supplementary Figure 4 for different initial conditions. In all cases, we see that $MAE(< x^q >, y^q)$ increases for lower levels of connectivity, going from ≈ 0 to 0.2 on average. However, when comparing the MAE with the natural variability in the stock levels x_i^q, $SD(x^q)$, we see that the difference between the $< x^q >$ and y^q is at worst on the same order than the standard deviation of the original nodes. Thus, we conclude that while the error of the approximation increases for lower values of the sampling probability u, the variability within the layer also increases, being greater or within a similar order. The $MAE(< x^q >, y^q)$ is also dependent on the initial condition considered. Notoriously, setting all starting values at random produces the highest variability in stock levels $SD(x^q)$, but the lowest error in the estimator $MAE(< x^q >, y^q)$ (on average).

Layer sizes and probability of edge

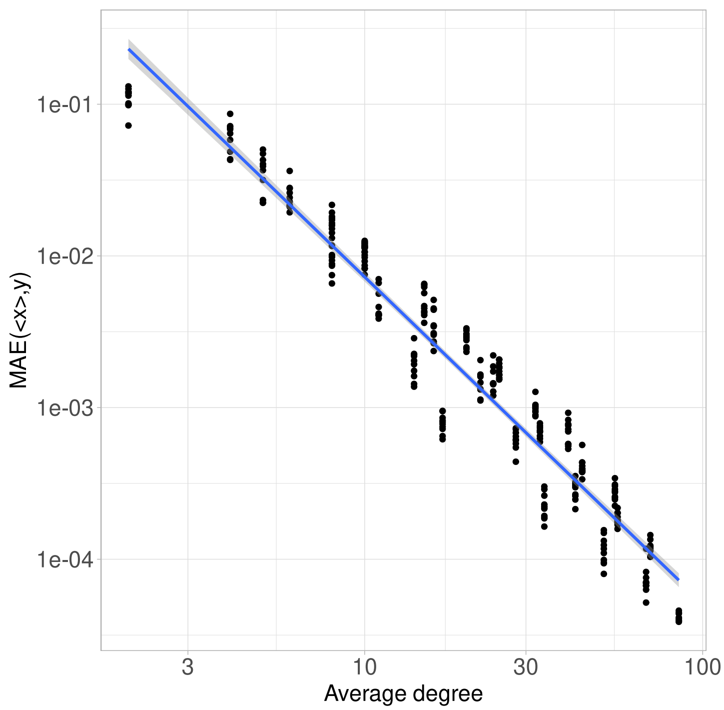
We find that by increasing the layer size N_q, the error of the approximation reduces. This means that as more nodes are considered, the approximation becomes better. We combine the change in layer size with the probability of connection u to inspect the error in terms of the average degree of a layer. Thus, we consider systems with equal layer size ($N^1 = N^2 = N^3 = N$) and different probabilities u. We calculate the global $MAE(< x >, y) = \frac{1}{3} \sum_{q=1}^{3} MAE(< x^q >, y^q)$ and plot it in function of the inter-layer degree uN. We found that the error grows as a power law of the average degree with exponent ≈ -2.15. The plot is presented in Supplementary Figure 5

Flow capacity variability

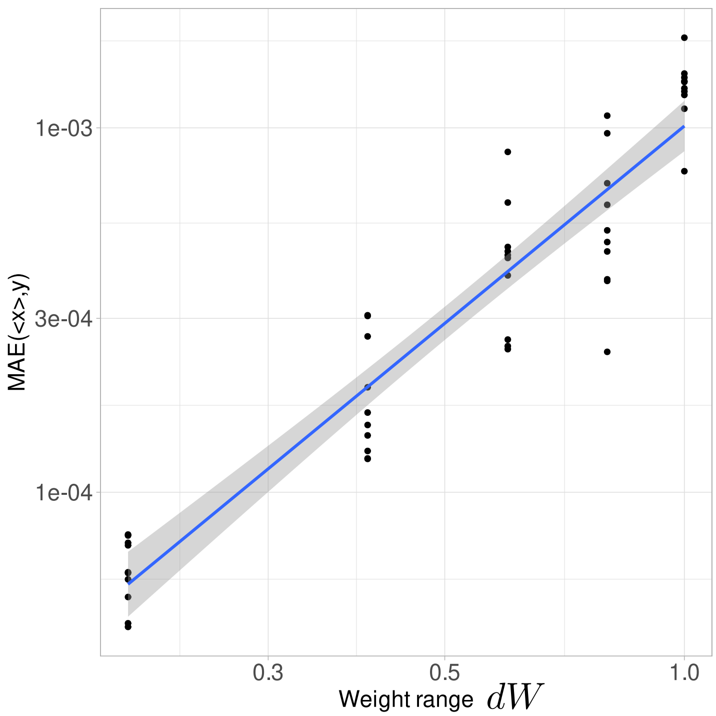
We explore the effect of flow capacity W_{ij}^{qr} variability by keeping the network structure fixed (fully connected layers with $N_q = 10$), and assign flow capacities W_{ij}^{qr} uniformly at random between $[1 - dW, 1 + dW]$, with $dW \in [0, 1]$. We plot $\frac{1}{3} \sum_{q=1}^{3} MAE(< x^q >, y^q)$ as a function of dW (Supplementary Figure 6). We find that the error increases as a power law of dW (exponent 1.8), but it is still low when compared to the impact of the sampling probability u.



Supplementary Figure 4: **Testing of the approximation for different values of the sampling probability of edges.** Each different curve corresponds to a different initial condition. *AllEmpty* corresponds to $x_i^q(t=0) = 0\ \forall i, q$. *AllFull* corresponds to $x_i^q(t=0) = 1\ \forall i, q$. *EqualRandom* corresponds to $x_i^q(t=0) = x_0\ \forall i, q$, with x_0 a random number between 0 and 1. *LayerRandom* corresponds to $x_i^q(t=0) = x_0^q\ \forall i$ with x_0^q a different random number for each layer. *Random* corresponds to $x_i^q(t=0)$ sampled independently within $[0, 1]$.



Supplementary Figure 5: **Testing of the approximation for different values of the average degree between layers**. The relationship found follows a power law with exponent -2.15



Supplementary Figure 6: **Approximation error as function of flow capacity variability.** We found the difference to increase as a power law of the size interval from where flow capacities are sampled. The exponent obtained is 1.8.

Supplementary References

[1] Jianxi Gao, Baruch Barzel, and Albert-László Barabási. Universal resilience patterns in complex networks. *Nature*, 530(7590):307–312, 2016.

[2] Yiyi He, Sarah Lindbergh, Yang Ju, Marta Gonzalez, and John Radke. Towards resilient critical infrastructures: Understanding the impact of coastal flooding on the fuel transportation network in the san francisco bay. *ISPRS International Journal of Geo-Information*, 10(9):573, 2021.

[3] Vincent Thibeault, Guillaume St-Onge, Louis J Dubé, and Patrick Desrosiers. Threefold way to the dimension reduction of dynamics on networks: an application to synchronization. *Physical Review Research*, 2(4):043215, 2020.