0.1 General Model Description

These are just some sketch notes on this - this may not make logical flow. These thoughts come from thinking about "Chinese Whispers" or "the telephone game" and are supposed to represent information mutation as it propagates (as a signal) through a network. These thoughts may or may not be used in any project dissertation. This is to demonstrate, in an abstract way, how I may think about a problem. Some, results, under very strict assumptions are deduced.

0.1.1 Mathematical Structure (just to get some mathematical details out the way)

Suppose we have a time varying graph $G(t) = (V(t), E(t))$ where $V(t)$ represents agents on the graph, and $E(t)$ represents connections between the agents. Here time is continuous, $t \in \mathbb{R}^+$. The manipulation of information between agents v_i and v_j as it diffuses through a network is described by transfer functions ψ_i that operates on some information θ_i .

Information should encompass any data or signals being transmitted and processed by an agent. Here, the information θ_i and θ_j is contained in a set Θ, the set of all possible information on G.

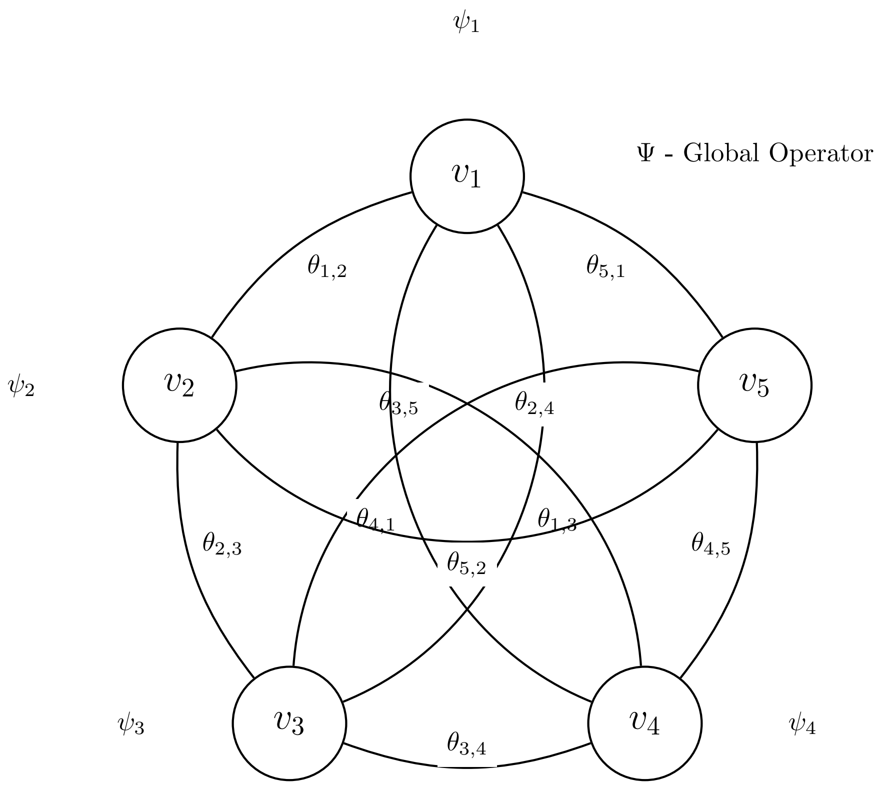
A transfer function ψ_i describes the change of information from θ_i to θ_j contained in the agents $v_i \in V$ to other connected agents $v_j \in V$.

Each transfer function ψ_i is an element an of a vertex function $\psi \in \Theta^{|V|}$ on the graph G. The vertex function shall correspond to a vector function $\psi(\boldsymbol{\theta}) = \big(\psi_1(\theta_1), ..., \psi_{|V|}(\theta_{|V|})\big)^\mathsf{T}$. This essentially projects information exchange into a higher dimensional space $\Theta^{|V|}$. This function should tell us how information itself changes over time.

An operator Ψ acts as a unifying law to control information transformations given collectively by the vertex function ψ and provides values for all information states in a network over time. The operator Ψ lives in a function space and maps between other function spaces

$$\Psi : \Theta^{|V|} \times G \to \Theta^{|V'|}$$

this operator will operate on the vertex function to tell us how the agents change in processing and distributing information. The reason $\Theta^{|V'|}$ is chosen here is to emphasise that the networks structure may change over time. $\Theta^{|V'|}$ is a Hilbert or Banach space, and functional mappings between them are the focus - The graph topology can be time-varying due to factors like node mobility, varying transmission power, or environmental changes affecting connectivity.



This diagram visualises what a process being discussed here might look like. After the overall goal of the project has been discussed a diagram detailing how the mathematical description shall be decode this diagram.

0.1.2 Overarching Goal

The aim is to build a framework to find a solution operator Φ for the nonlinear-vector-differential equation,

$$\begin{aligned}\Psi(\psi(\boldsymbol{\theta}), G; t) &= \frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta}(t), t)}{\partial t} \\ &\stackrel{?}{=} \frac{\partial^2 \boldsymbol{\theta}}{\partial t^2}\end{aligned}$$

Where each transfer function ψ_i itself is a dynamical system living on a network G and $\Psi : \Theta^{|V|} \to \Theta^{|V'|}$. Furthermore, Ψ is a non-autonomous system i.e. non-relativistic mechanical systems subject to time-dependent transformations.

The time derivative of the the vertex function is to highlight that way each transfer function processes information will change under successive iterations of either Ψ or $\boldsymbol{\psi}$ since

$$\theta_j \leftarrow (\psi_j \circ \psi_i)(\theta_i)\,.$$

What we need to work towards: When data is entered into the network in the form θ_1, we may not have sequentially updating of the data in the form:

$$\begin{aligned}\theta_j &= \psi_j\left((\psi_1 \circ \psi_2 \circ \cdots \circ \psi_{i-1} \circ \psi_i)(\theta_{i-1})\right) \\ &= \psi_j\left(\bigcirc_{k=1}^{i} \psi_k(\theta_{k-1})\right), \quad i \geq 2\end{aligned}$$

since this is just simple message passing. We have ψ_1 applied to ψ_2 etc...however we may have ψ_1 applied to ψ_2 applied to ψ_1 and then maybe ψ_{i-1} and then sent back to ψ_1. How do write that down? To capture the more complex dynamics where transfer functions can be applied in a non-sequential and potentially recursive manner, we need a more flexible notation that allows for arbitrary compositions of transfer functions.

The operator Φ generates a trajectory of $\boldsymbol{\theta}$ over time $t \in \mathbb{R}^+$ based on dynamics given by the vertex function ψ. Φ should act like an integral .

If we work with this idea then this framework would achieve this by learning a solution operator Φ. Ideally, Φ should integrate the effect of the information processing throughout the network, which symbolically relates to the integral formulation,

$$\begin{aligned}\Phi\left(\Psi\left(\psi(\boldsymbol{\theta})\right)\right) &= \int\left(\int \Psi(\psi(\boldsymbol{\theta}), G; t)dt\right)dt \\ &= \int\left(\int \frac{\partial \boldsymbol{\psi}}{\partial t}dt\right)dt \\ &\stackrel{?!}{=} \boldsymbol{\theta}(t)\end{aligned}$$

since

$$\boldsymbol{\psi}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\theta}}{\partial t}$$

then

$$\int \boldsymbol{\psi}(\boldsymbol{\theta})dt = \int \frac{\partial \boldsymbol{\theta}}{\partial t}dt$$
$$\stackrel{?}{=} \boldsymbol{\theta}(t)$$

If the information processing rules are allowed to vary but we do not consider the graph structure changing itself, then Ψ no longer depends on G and the mapping becomes $\Psi : \Theta^{|V|} \to \Theta^{|V|}$ and

$$\Psi(\psi(\boldsymbol{\theta}); t) = \frac{\partial \boldsymbol{\psi}}{\partial t}$$

We want to form some sort of basis for approximate solutions here:

0.1.3 Special Case: $\Psi = \mathbf{I}$ and G Non-Time Varying

If the information processing rules remains constant over time and the network structure is fixed, so $\Psi = \mathbf{I}$, the identity operator, i.e. the agents do not change how they process information over time and graph structure is stationary, then,

$$\begin{aligned}\Psi(\psi(\boldsymbol{\theta}); t) &= \Psi(\psi(\boldsymbol{\theta}))\\ &= \boldsymbol{\psi}(\boldsymbol{\theta})\end{aligned}$$

which simplifies the problem a bit. We can define this new problem as

$$\begin{aligned}\boldsymbol{\theta}(t) &= \boldsymbol{\theta}(0) + \int_0^t \Psi(\psi(\boldsymbol{\theta}(s), s); s)ds\\ &= \boldsymbol{\theta}(0) + \int_0^t \boldsymbol{\psi}(\boldsymbol{\theta}(s), s)ds\end{aligned}$$

which is an integral equation.

$$\Phi\left(\psi(\boldsymbol{\theta})\right) := \boldsymbol{\theta}(t) = \boldsymbol{\theta}(0) + \int_0^t \left(\boldsymbol{\psi} \circ \boldsymbol{\theta}\right)(s)ds$$

and possibly nonlinear Volterra type, of the second kind; where $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ are unknown functions. This should tell us how information changes with interacting transfer functions. Somehow apply the Banach Fixed Point Theorem? BFP Theorem might be applicable in cases where a contraction mapping can be established, facilitating convergence to a fixed point that represents a stead-state of $\boldsymbol{\theta}$ change across the network.

- $\Theta^{|V|}$ is a complete-metric space
- Establish Contraction Mapping on $\boldsymbol{\psi}$ and conditions under which this would be true
- There exists a fixed point $\boldsymbol{\psi}\left(\boldsymbol{\theta}^*\right) = \boldsymbol{\theta}^*$

0.1.4 Special Case: ψ changes over time, G is non-time varying:

We know that the structure of the network does not change over time so we can assume $\Psi : \Theta^{|V|} \to \Theta^{|V|}$ and

$$\boldsymbol{\psi}\left(\boldsymbol{\theta}(t), t\right) = \boldsymbol{\psi}\left(\boldsymbol{\theta}(t), 0\right) + \int_0^t \Psi(\psi(\boldsymbol{\theta}(s), s); s) ds$$

We can make a step towards analysing this object by attempting to open up the integral, so we have:

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}(0) + \int_0^t \left(\boldsymbol{\psi} \circ \boldsymbol{\theta}\right)(s) ds + \int_0^t \left(\Psi \circ \boldsymbol{\psi} \circ \boldsymbol{\theta}\right)(s) ds,$$

where $\dim(\psi) = |V|$ since G is not time dependent.

However, the processing rules $(\psi_1(\theta_1, t), ..., \psi_{|V|}(\theta_{|V|}, t))^\mathsf{T}$ are now variable. The operator Ψ would change the transfer functions:

$$\left(\Psi \circ \boldsymbol{\psi} \circ \boldsymbol{\theta}\right)(t) = \begin{pmatrix} \Psi\left(\psi_1(\theta_1), t\right) \\ \Psi\left(\psi_2(\theta_2), t\right) \\ \vdots \\ \Psi\left(\psi_{|V|}(\theta_{|V|}), t\right) \end{pmatrix}$$

We should analyse what this is doing over some fixed period. Consider a iterated applications of the mapping Ψ up to some terminal time t^*, for simplicity we assume that Ψ is a uniform change across all transfer functions. A simplified analogy might be a chemical entering a system of particles, each molecule might react differently, however, the overall change is the same at any time t. However, if we then add another chemical to the mix on top of this, and keep doing it up until some terminal time, then what is going on? This is saying that we want to collect each iteration in a set, so lets define this formally.

Let the iterated composite mapping of the vertex up until a time t^* be defined as:

$$\Psi_k^{(t^*)}\left(\boldsymbol{\psi}(\boldsymbol{\theta})\right) = \begin{pmatrix} \left(\Psi_1 \circ \cdots \circ \Psi_{t^*}\right)\left(\psi_1(\theta_1)\right) \\ \left(\Psi_1 \circ \cdots \circ \Psi_{t^*}\right)\left(\psi_2(\theta_2)\right) \\ \vdots \\ \left(\Psi_1 \circ \cdots \circ \Psi_{t^*}\right)\left(\psi_{|V|}\left(\theta_{|V|}\right)\right) \end{pmatrix}$$

another way to denote this would be

$$\Psi^{(t^*)}\left(\boldsymbol{\psi}(\boldsymbol{\theta})\right) = \bigcirc_{k=1}^{t^*} \Psi_k\left(\boldsymbol{\psi}(\boldsymbol{\theta})\right)$$

we would now like collect these iterated maps into a set

$$\left\{\boldsymbol{\psi}^{(1)}, ..., \boldsymbol{\psi}^{(t^*)}\right\} = \left\{ \begin{pmatrix} \Psi_1\left(\psi_1(\theta_1)\right) \\ \Psi_1\left(\psi_2(\theta_2)\right) \\ \vdots \\ \Psi_{|V|}\left(\psi_{|V|}(\theta_{|V|})\right) \end{pmatrix}, \begin{pmatrix} \left(\Psi_1 \circ \Psi_2\right)\left(\psi_1(\theta_1)\right) \\ \left(\Psi_1 \circ \Psi_2\right)\left(\psi_2(\theta_2)\right) \\ \vdots \\ \left(\Psi_1 \circ \Psi_2\right)\left(\psi_{|V|}(\theta_{|V|})\right) \end{pmatrix}, ..., \begin{pmatrix} \left(\Psi_1 \circ \cdots \circ \Psi_{t^*}\right)\left(\psi_1(\theta_1)\right) \\ \left(\Psi_1 \circ \cdots \circ \Psi_{t^*}\right)\left(\psi_2(\theta_2)\right) \\ \vdots \\ \left(\Psi_1 \circ \cdots \circ \Psi_{t^*}\right)\left(\psi_{|V|}(\theta_{|V|})\right) \end{pmatrix} \right\}$$

what can we say about this? If each Ψ_i is a contraction mapping, then by the Banach Fixed Point Theorem, the iterated application of Ψ might converge to a fixed point. This means there exists a steady state $\boldsymbol{\psi}^*$ such that $\Psi\left(\boldsymbol{\psi}^*\right) = \boldsymbol{\psi}^*$. This would allow for machine learning to look for such state. How does this help us solve the integral equation

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}(0) + \int_0^t (\boldsymbol{\psi} \circ \boldsymbol{\theta})(s)ds + \int_0^t (\Psi \circ \boldsymbol{\psi} \circ \boldsymbol{\theta})(s)ds?$$

Let's focus on the last state given by $\Psi^{(t^*)}(\boldsymbol{\psi}(\boldsymbol{\theta}))$, let $t^* = k$ for clarity, then

$$\boldsymbol{\psi}^{(k+1)} = \Psi^{(k)}\left(\boldsymbol{\psi}^{(k)}\right)$$

and so

$$\boldsymbol{\theta}^{(k+1)}(t) = \boldsymbol{\theta}(0) + \int_0^t \left(\boldsymbol{\psi}^{(k)} \circ \boldsymbol{\theta}^{(k)}\right)(s)ds + \int_0^t \left(\Psi \circ \boldsymbol{\psi}^{(k)} \circ \boldsymbol{\theta}^{(k)}\right)(s)ds,$$

with convergence criteria:

$$\left\|\boldsymbol{\psi}^{(k+1)} - \boldsymbol{\psi}^{(k)}\right\| < \epsilon$$

and

$$\left\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\right\| < \epsilon.$$

From this if we want to apply operator learning, input/output pairs can be taken, the outputs would be the steady states of the system.

Is this even a different problem to special case one. Since, in special case one, iterated composition of ψ is given. This is because

$$\begin{aligned}\theta_{i+1} &= (\psi_{i+1} \circ \psi_i \circ \psi_{i-1} \circ \cdots \circ \psi_2 \circ \psi_1)(\theta_1) \\ &\implies \theta_{i+3} = \psi_{i+2}(\theta_{i+1})\end{aligned}$$

since as initial information θ_1 travels through the network then out pops θ_i. Now, if we just enforce that $\boldsymbol{\psi}$ is a contraction mapping, then, if $\boldsymbol{\psi}$ changes, to, say $\boldsymbol{\psi}^{(k)}$ then this implies $\boldsymbol{\psi}^{(k)}$ must also be a contraction mapping. Therefore, if Ψ changes $\boldsymbol{\psi}$ somehow, it just needs to change $\boldsymbol{\psi}$ in a way such that it is a contraction mapping. So this is just a restatement of special case one with more sub-compositions.

0.1.5 Another form of 1.4 - without enforcing BFPT

If the network G has a fixed topology then $\Psi : \Theta^{|V|} \to \Theta^{|V|}$ we can leverage this property and transform the problem into one of learning the interaction kernel.

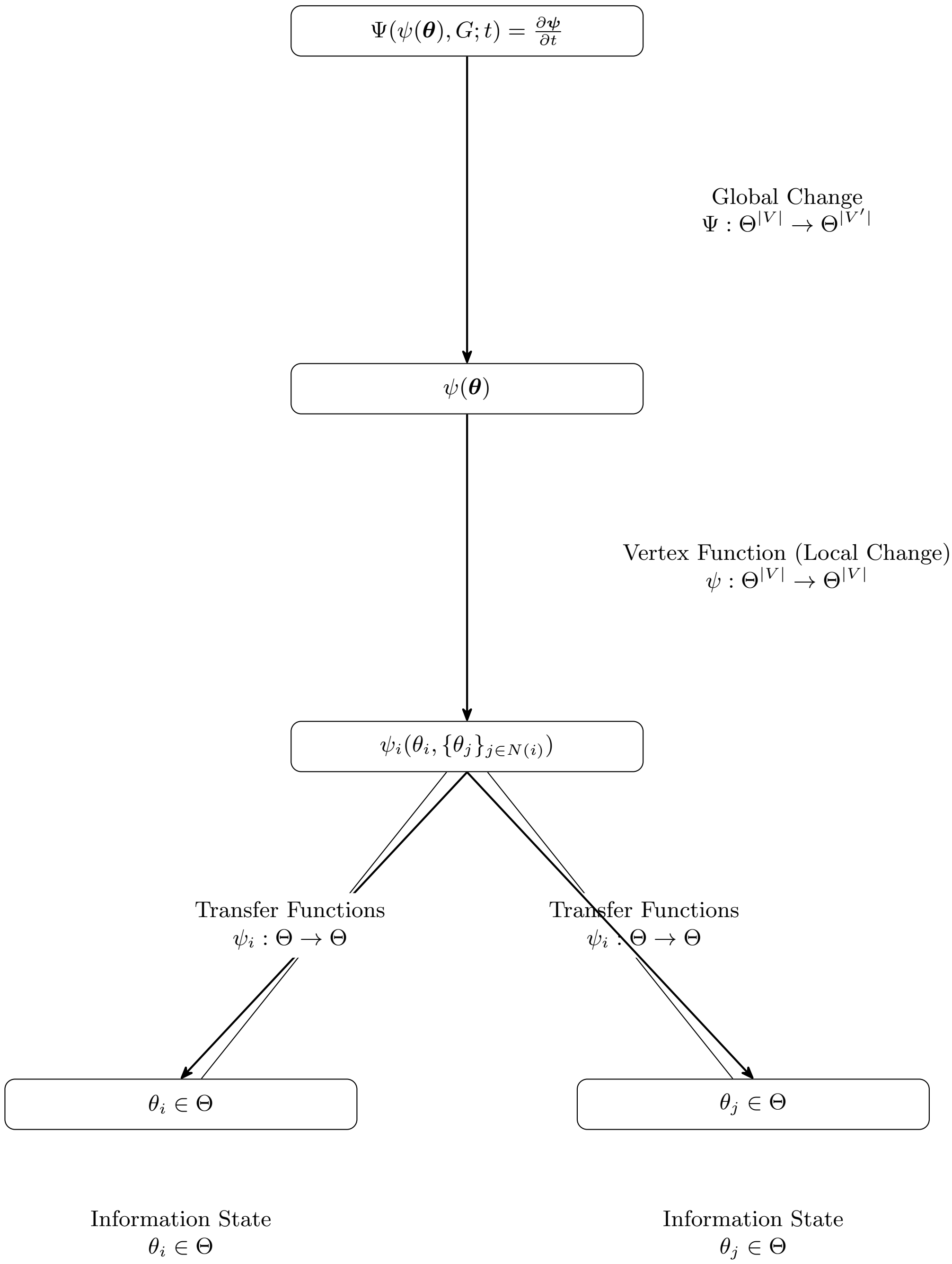
0.1.6 General Case: $\boldsymbol{\psi}$ and G vary in time - complete temporal dependency

so we have $\dim(\psi) = |V(t)|$ which means that the shape of the vector changes over time...

???

Really hard.

A simplified diagram of the three level model is provided below and it should help with visualising the high level structure used throughout this document and the way I think about this problem. The assumption on the existence of a global operator like Ψ is a strong one and this is accepted, however, to truly decode what's going on and apply machine learning techniques i.e. applying a universal approximation theorem insists that some function like Ψ must exist.



The diagram shows how this framework breaks down network information change $\partial_t \theta$ through a three-level structure with distinct mathematical structures. Individual agents possess information states θ_i residing in an information space Θ (this could be $\mathbb{R}, \mathbb{C}$ or some suitable space). Local dynamics are dictated by transfer functions ψ_i, governing state changes at each node in Θ. However, to capture the collective network dynamics, a vertex function ψ is constructed, which coordinates the state transitions across the network. This transformation bridges the gap between the transfer functions $\psi_i : \Theta \to \Theta$ and the vertex function $\psi : \Theta^{|V|} \to \Theta^{|V|}$, so a vector of transfer functions is created $\psi = (\psi_1, ..., \psi_{|V|})$.

Finally, the operator Ψ oversees the evolution of the network's state vector. It is given by the differential equation,

$$\Psi(\psi(\boldsymbol{\theta}), G; t) = \frac{\partial \boldsymbol{\psi}}{\partial t},$$

Realisations of $\psi(\boldsymbol{\theta}(t))$ for $t \in \mathbb{R}^+$ determines the state of each node over time, $\theta_i \in \Theta$ for all t.

0.2 Research Problem Statement

Problem 1. We represent Ψ as a **General Operator** acting on G. From a collection of observations $\{\theta_1, ..., \theta_n\}$ how do we recover the operator Φ?

This is an inverse problem, and an ill-posed one. The goal is to learn a parametrised neural operator $\hat{\Phi}(\cdot, t; \omega)$ to give a family of solutions such that $\hat{\Phi} \approx \Phi$, where $\omega \in \Omega$ are parameters of the neural network. The neural network model $\hat{\Phi}$ is a Graph Neural Operator or a variant of a Deep Operator Network known as a Deep Graph Operator. These models belong to a class of machine learning models designed to learn mappings between infinite-dimensional function spaces and generalise classical neural networks. The idea is to focus on level 3 (the operator) and work our way down to the information level at the bottom. This is a top-down approach.

Problem 2. (Supervised Learning)
We determine an underlying mapping $\Phi : \Theta^{|V|} \to \Theta^{|V'|}$ from samples of input/output pairs representing information change on a network G, the I/O pairs are given as

$$\{\theta_n, \Psi(\psi_n(\theta_n))\}_{n=1}^N, \quad \theta_n \sim \mu$$

where, the input is the information states of each agent at some time t and the output, the information state after the operator Ψ has been applied to the input θ. The function μ is a probability measure supported on the space of all possible information states Θ for any agent θ_n. In this set up we note that the size of vertex set $|V|$ is the dimension of the data.

Problem 3. (Generalisation of Problem 2) Let $\mathcal{U}$ and $\mathcal{V}$ be a Separable Banach Spaces of functions:

$$\mathcal{U} = \left\{ u_n : \Theta^{|V|} \to \mathbb{R}^{d_{\text{in}}} \right\}$$
$$\mathcal{V} = \left\{ \nu_n : \Theta^{|V'|} \to \mathbb{R}^{d_{\text{out}}} \right\}$$

Given $\{\theta_n, \Psi(\psi_n(\theta_n))\}_{n=1}^N$ we seek to determine an approximate operator $\hat{\Phi} : \mathcal{U} \to \mathcal{V}$ from within a family of parametrised functions such that

$$\hat{\Phi} : \mathcal{U} \times \Omega \to \mathcal{V}, \quad \Omega \subseteq \mathbb{R}^p$$

where Ω is the parameter space.

Learning Framework

- Represent Φ as a parameterized neural network (Operator Learning).
- Train the network on data of initial information states and their evolved states after diffusion.
- Use the trained network to predict future information states.
- Control Predictions and state estimates

Research Project Flow

- Finalise Theoretic Framework.
- Find Application where Theoretical Framework fits.
- Recovering Φ from network observations (inverse problem).
- Learning the underlying mapping Φ from input/output pairs representing information change.
- Generalization to different data and network dynamics.
- Extend Framework, prove some propositions, dynamical systems analysis etc (if time permits).

0.3 Some considerations

While interesting, the approach requires significant data to accurately train the model and substantial computational resources to handle real-time network changes. Further research will focus on optimising these aspects, sampling from some high-dimensional dynamical system, and some thought to exactly what mathematical structures should be focussed on in regards to Φ (probably a Banach space) - to enhance feasibility and extend applicability to networks with different data and dynamics to leverage $\Psi(\cdot, G; t)$.

If Θ is not well-defined or if the mapping ψ_i does not preserve certain necessary properties of information states (like continuity, if required), then the application of Ψ might not behave as expected. The stability, convergence, or even the well-posedness of problems involving $\hat{\Phi}$ might not be guaranteed without additional constraints or properties, this would need to be explored in future research.

Alternatively, the invertibility of the operator Ψ under Φ is an interesting problem in itself.

- Clearly define information space Θ: The properties and limitations of the information space Θ need to be well-defined for the framework to work effectively.- highlighted in email
- Transfer function properties: Explore what properties the transfer functions ψ_i must have to ensure meaningful information flow through the network. - highlighted in email
- Alternative approaches: Consider if there are alternative ways to model information flow that might be computationally less expensive.
- Some assumptions, like the existence of a global operator Ψ, need further justification.
- A specific application to guide the development of such a framework would be best.