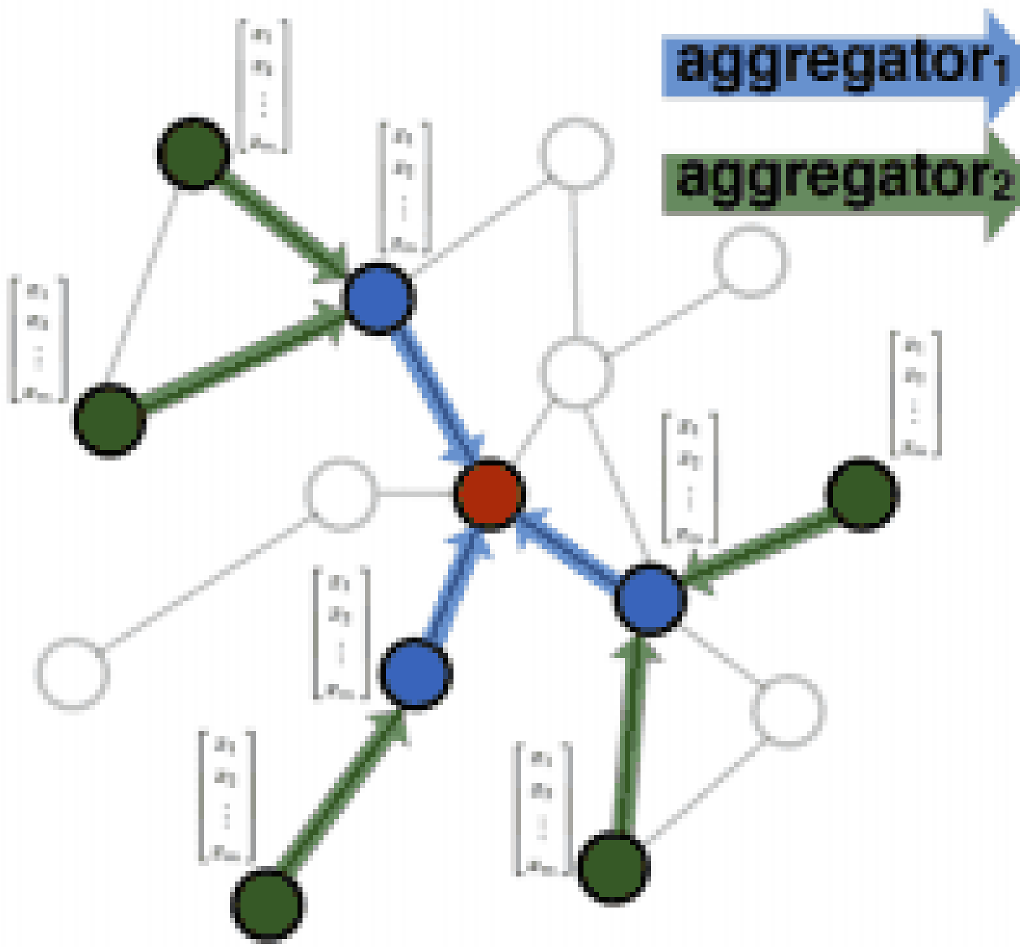
Graph Neural Networks 🧠

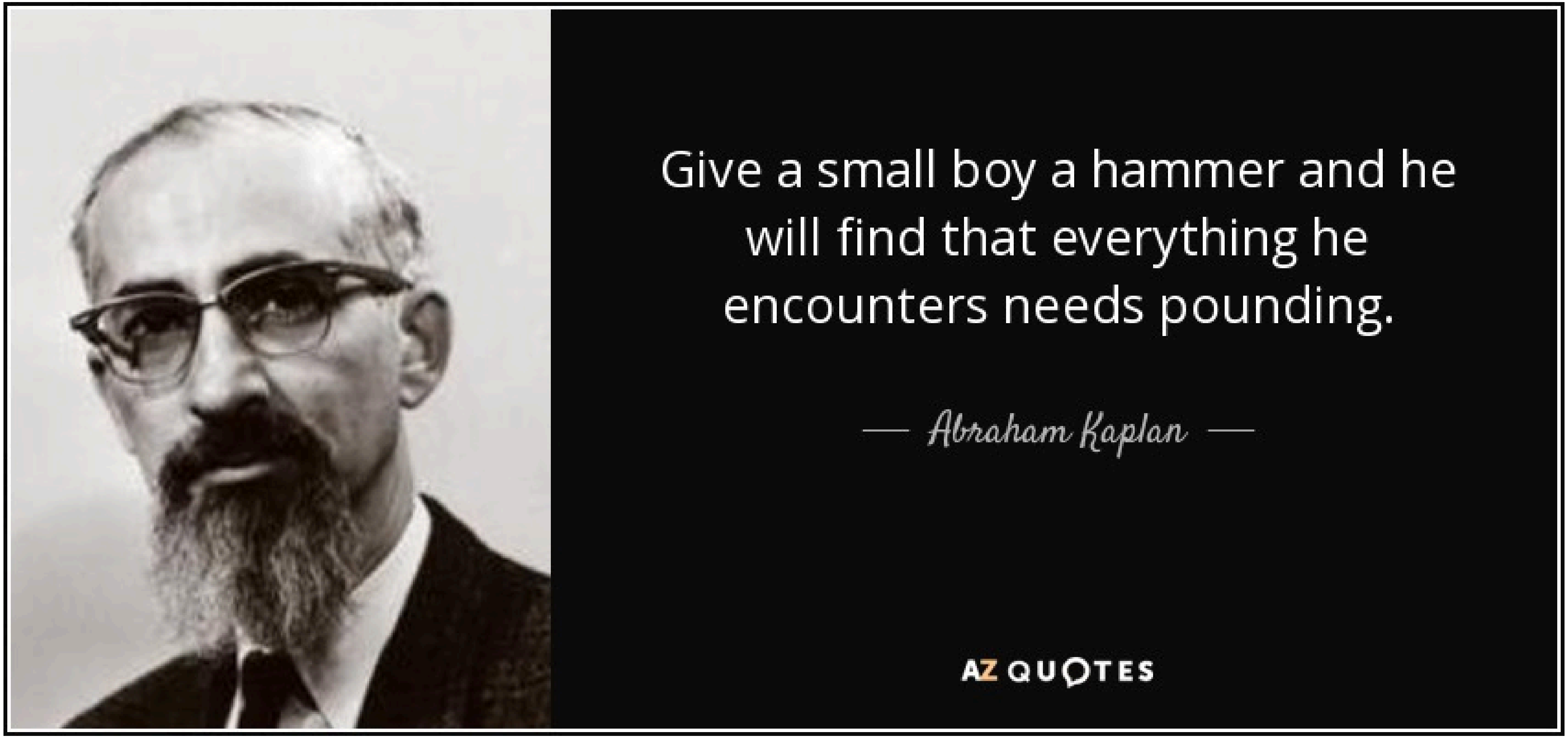
From Images to Graphs



Aggregate feature information from neighbors

What is a graph?

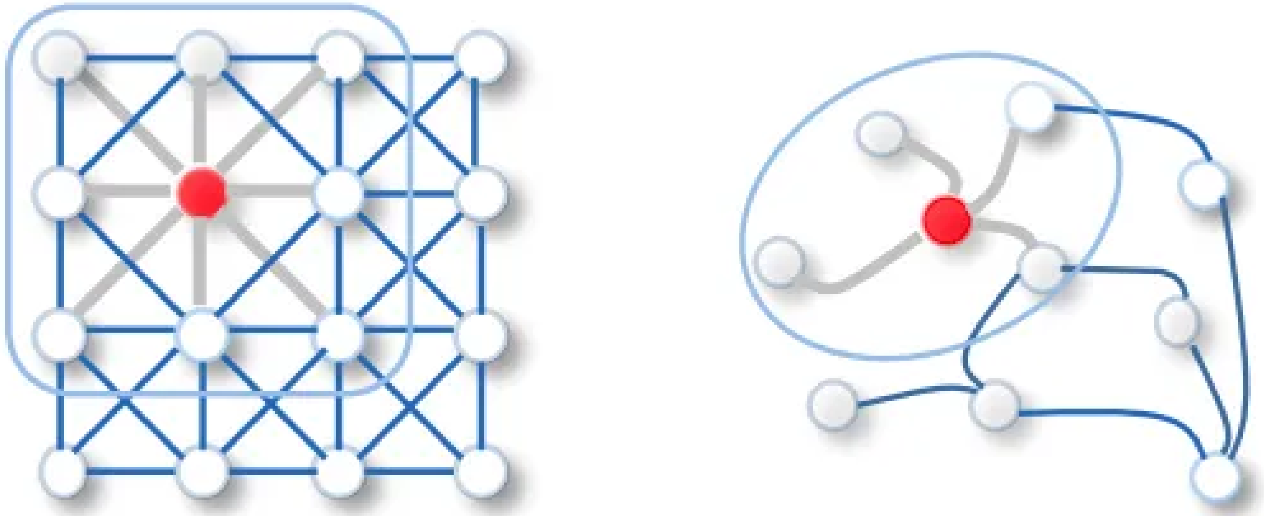
- A set of nodes connected by edges
- Simple representation of a relationship between objects
- Zoo of graphs!


Give a small boy a hammer and he will find that everything he encounters needs pounding.
— Abraham Kaplan —
AZ QUOTES

Neural Networks ~ Our hammer!

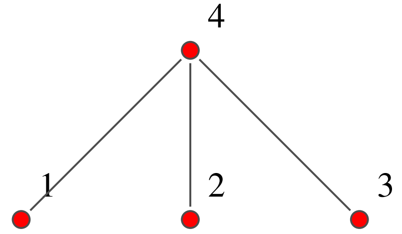
From Images to Graphs

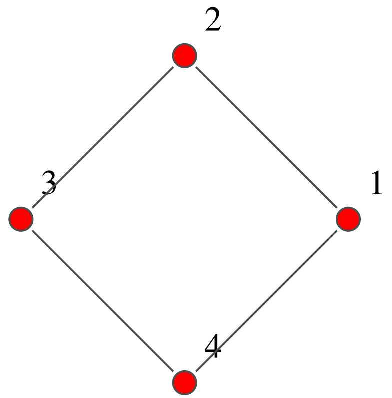
- Image = 2D grid of pixels
- Through a convolution, a pixel value is influenced by its neighbors
- We can represent this neighborhood structure using a graph and define **convolutions on graphs**!

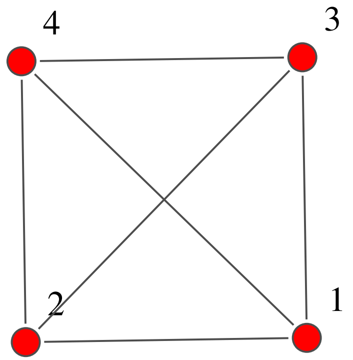


Let's represent a graph mathematically

- Adjacency matrix A
- $A_{ij} = 1$ if there is an edge between node i and node j, otherwise $A_{ij} = 0$







$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Convolution on images ~ Fourier Transform

Convolution on graph ~ ?

Convolution on images ~ Fourier Transform

Convolution on graph ~ Eigenvalues of Laplacian

Key idea

1D signal

- Suppose a 1D signal $x(t)$ as a function of time t.
- The frequency of the signal is essentially *the speed of variation*.
- High frequency signal ~ rapid variations
- Low frequency signal ~ slow variations

2D signal ~ Same as 1D signal

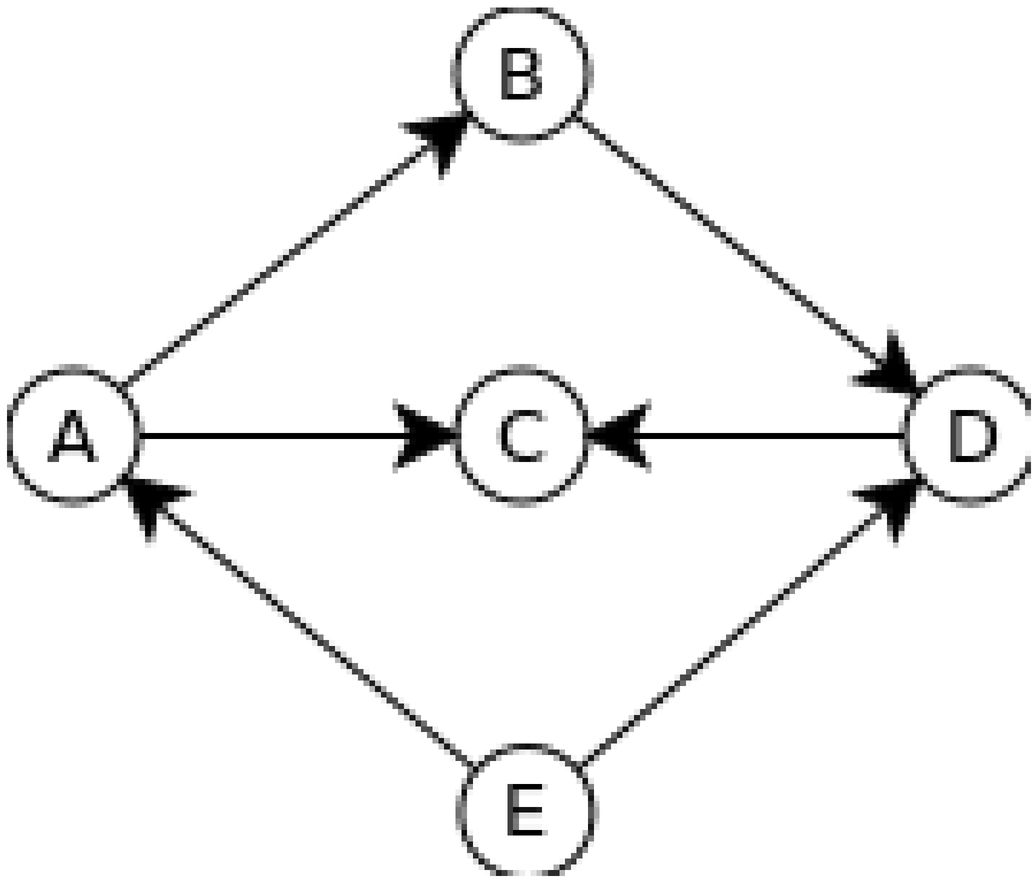
1D signal

- Suppose a 1D signal $x(t)$ as a function of time t.
- The frequency of the signal is essentially *the speed of variation*.
- High frequency signal ~ rapid variations
- Low frequency signal ~ slow variations

2D signal ~ Same as 1D signal

What about graph?

- Graph is non-trivial since it does not have an inherent order of nodes! (like time dimension in 1D signal and spatial dimension in 2D signal)
- But we can still define the variation as the sum of differences between neighboring nodes.



Total variation

- Suppose we have a graph of N nodes, each node has a feature x_i.
- **The total variation** measures the smoothness of the node features:

$$J = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} (x_i - x_j)^2 = \mathbf{x}^\top \mathbf{L} \mathbf{x}$$

where $\mathbf{L}$: Graph Laplacian, x_i: Node features, A_{ij}: Adjacency matrix.

Q: What x makes the total variation smallest (most smooth) and largest (most varying), provided that the norm of x is fixed? 🤔

The eigendecomposition of the Laplacian:

$$\mathbf{L}\mathbf{x} = \lambda\mathbf{x}$$

By multiplying both sides by $\mathbf{x}^\top$, we get

$$\mathbf{x}^\top\mathbf{L}\mathbf{x} = \lambda$$

This tells us that:

1. The eigenvectors with small eigenvalues represent **low-frequency** signals.
2. The eigenvectors with large eigenvalues represent **high-frequency** signals.

Decomposing the Total Variation

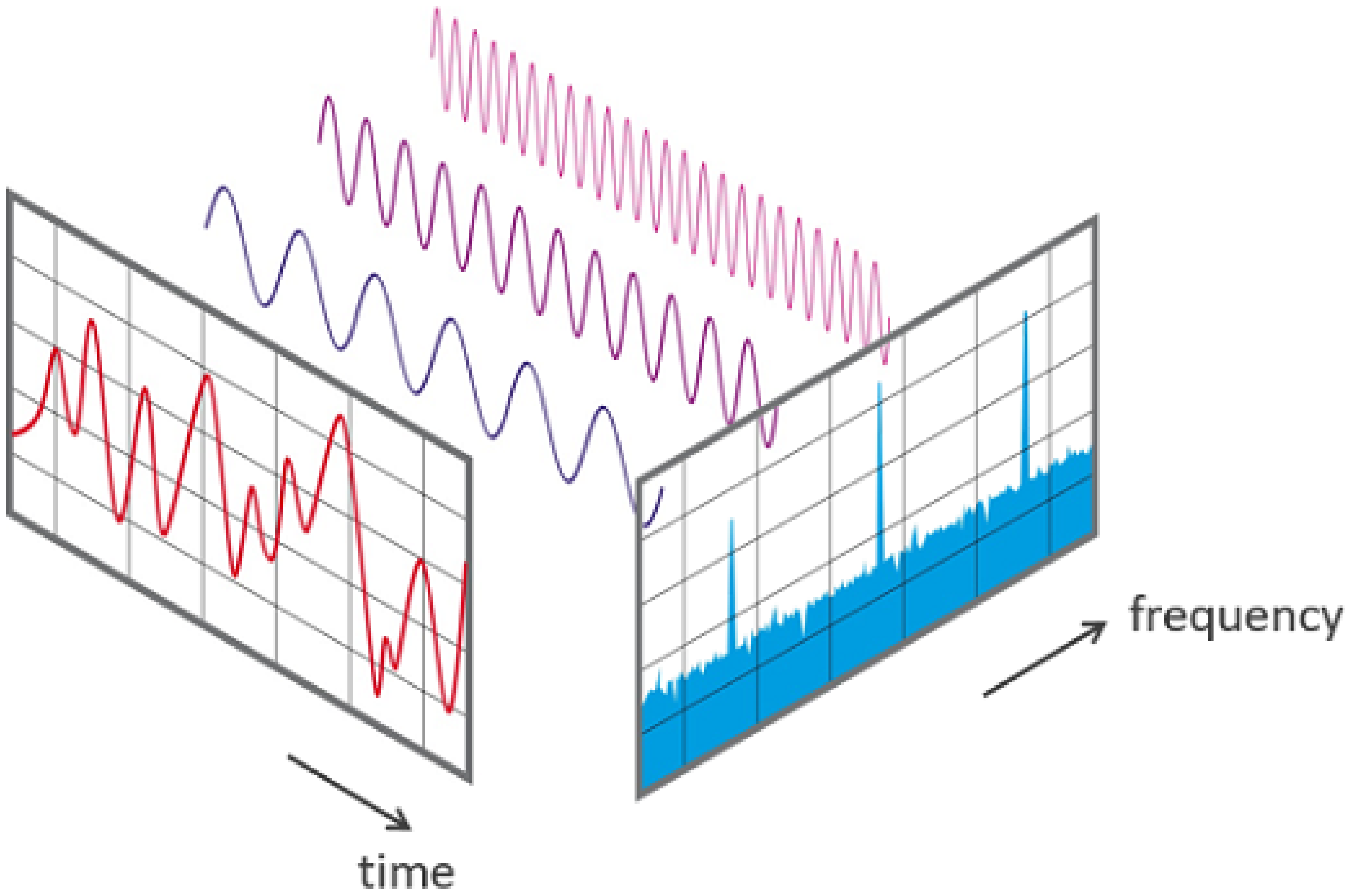
The total variation can be decomposed as follows ($\mathbf{u}_i$ is the eigenvector of the Laplacian):

$$J = \mathbf{x}^\top \mathbf{L}\mathbf{x} = \mathbf{x}^\top \left(\sum_{i=1}^{N} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top \right) \mathbf{x} = \sum_{i=1}^{N} \lambda_i (\mathbf{x}^\top \mathbf{u}_i)(\mathbf{u}_i^\top \mathbf{x})$$

$$= \sum_{i=1}^{N} \lambda_i \underbrace{||\mathbf{x}^\top \mathbf{u}_i||^2}_{\text{alignment between } \mathbf{x} \text{ and } \mathbf{u}_i}$$

Key Insight:

- The total variation is now decomposed into the sum of different frequency components $\lambda_i \cdot ||\mathbf{x}^\top \mathbf{u}_i||^2$.
- λ_i acts as *a filter (kernel)* that reinforces or passes the signal $\mathbf{x}^\top \mathbf{u}_i$.


frequency
time