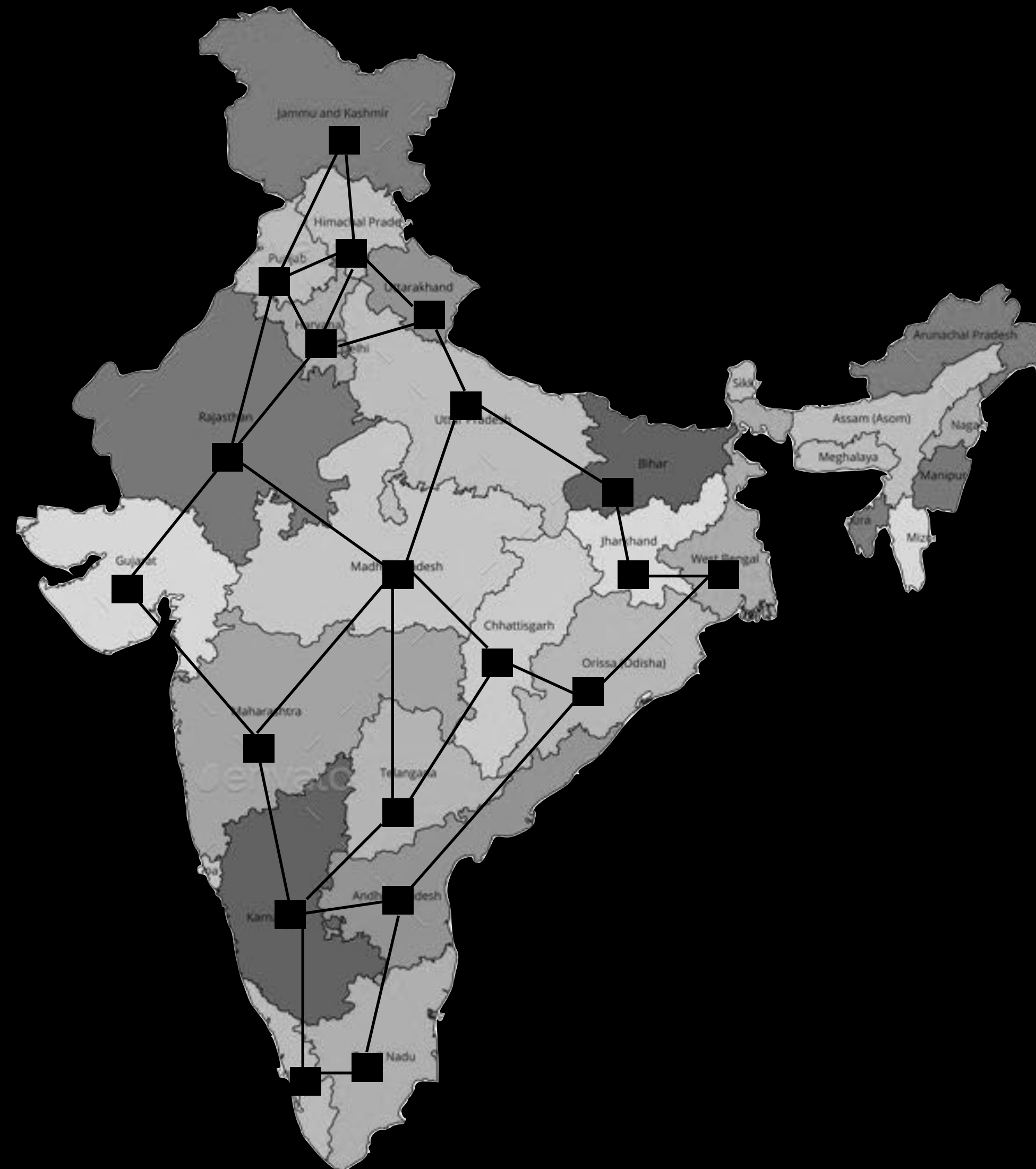


Text Analytics

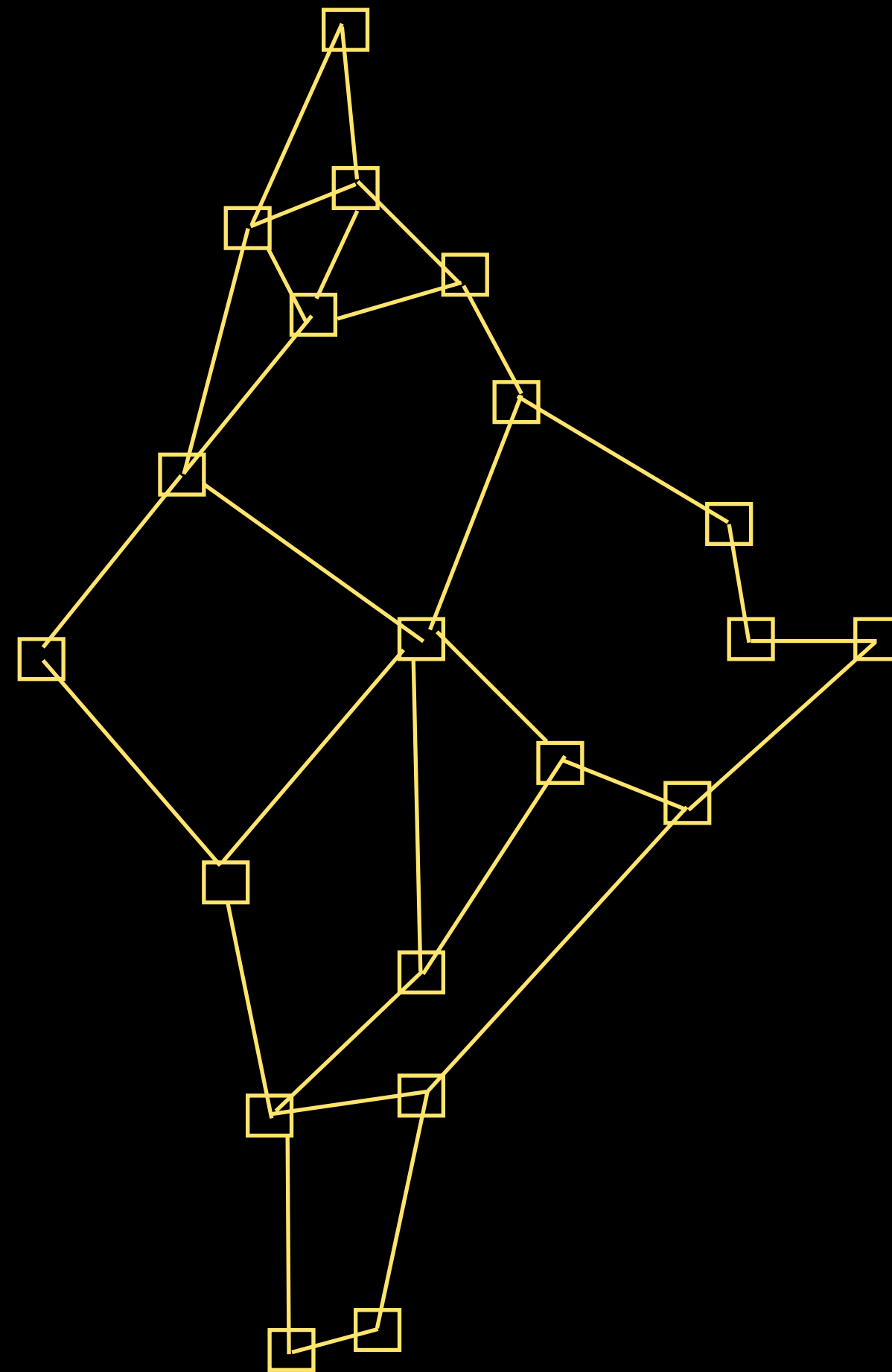
Knowledge Graphs

Ramaseshan Ramachandran

COLOURING A MAP



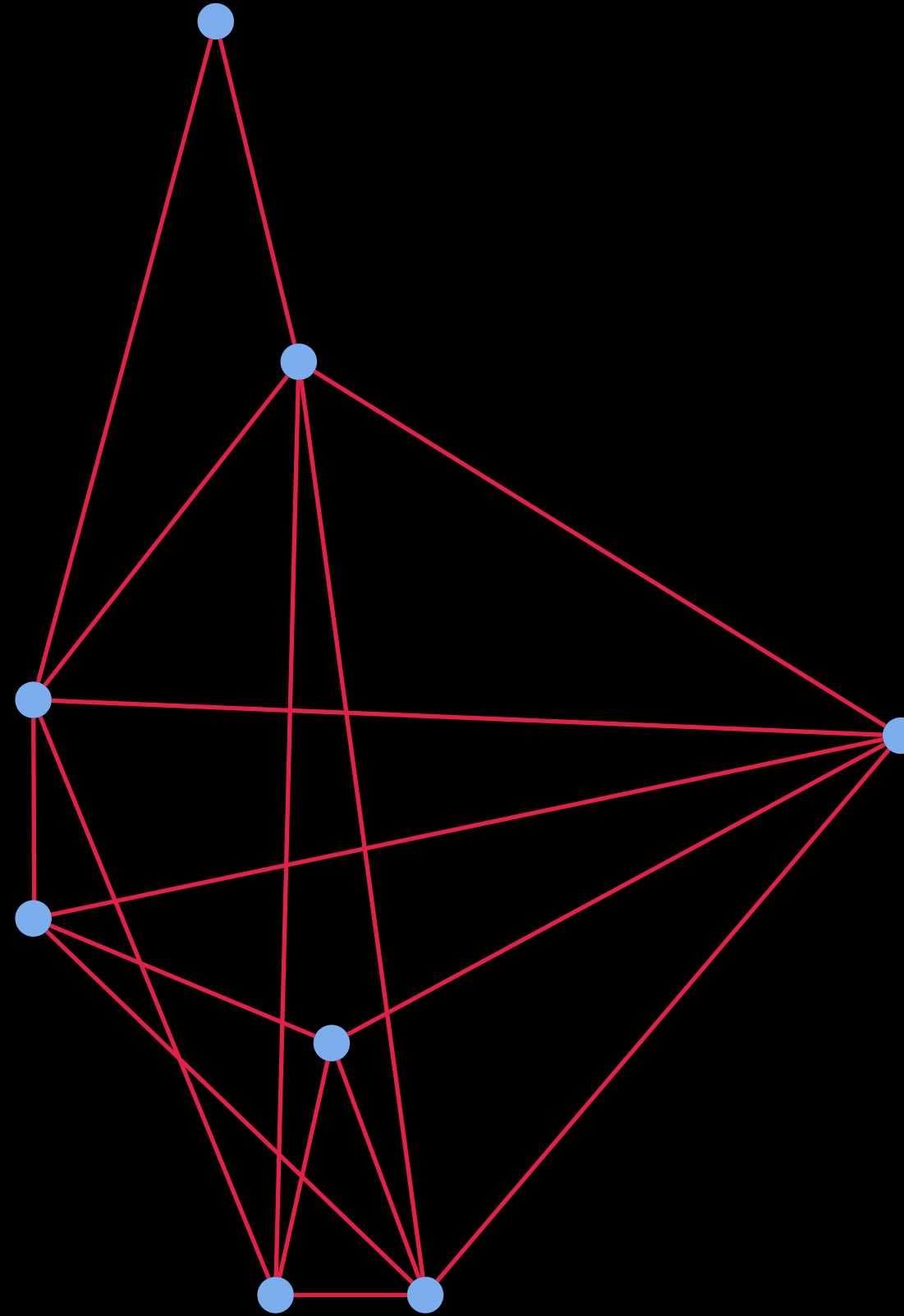
COLOURING A MAP



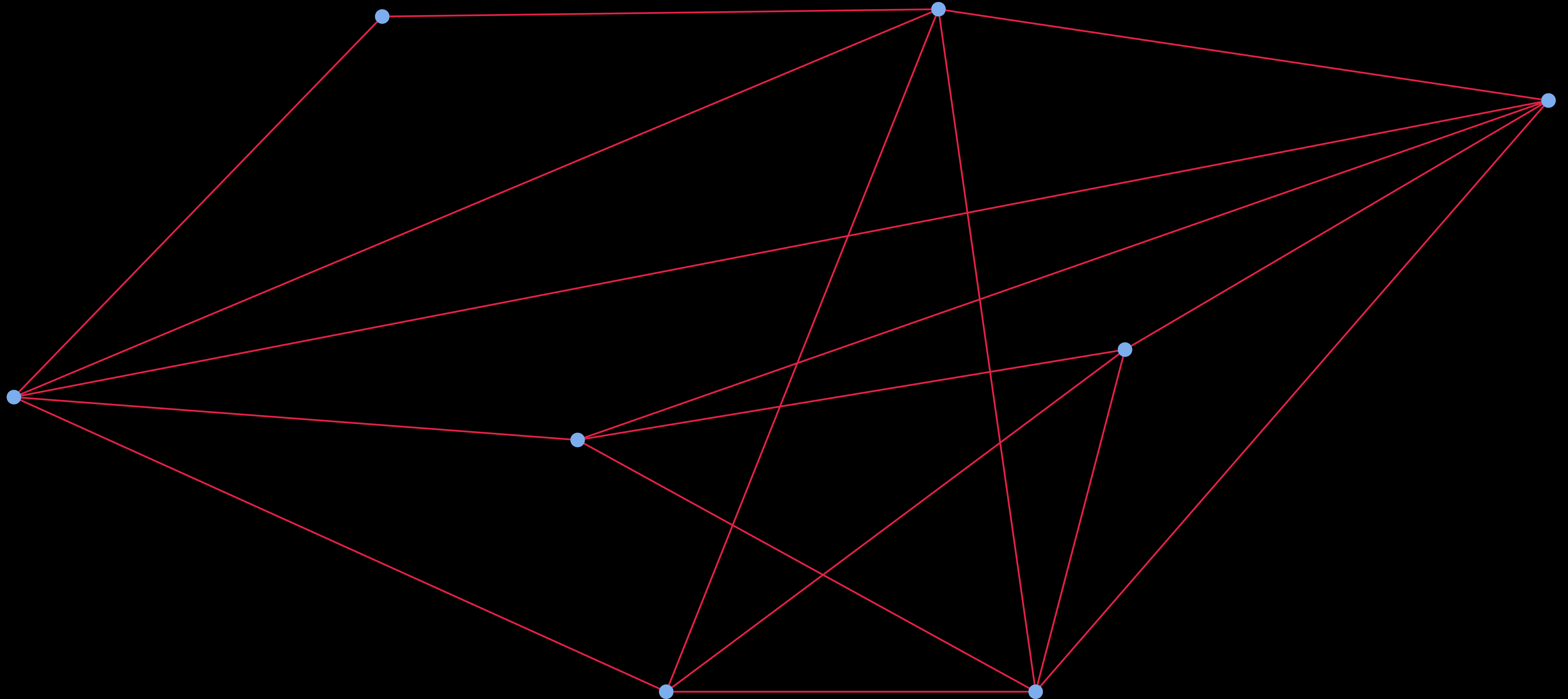
FLIGHT CONNECTIONS



CITIES AS A GRAPH

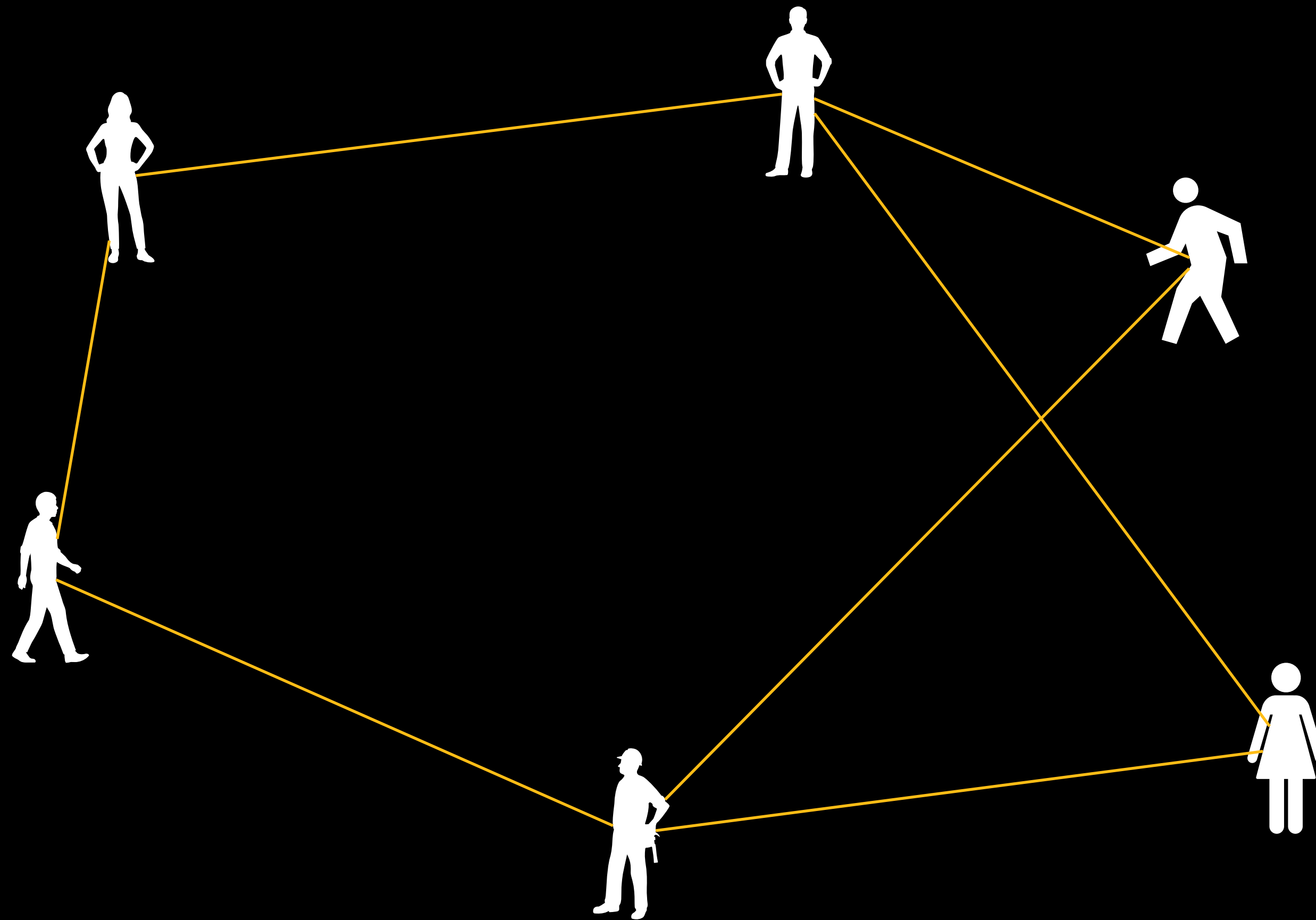


CITIES AS A GRAPH



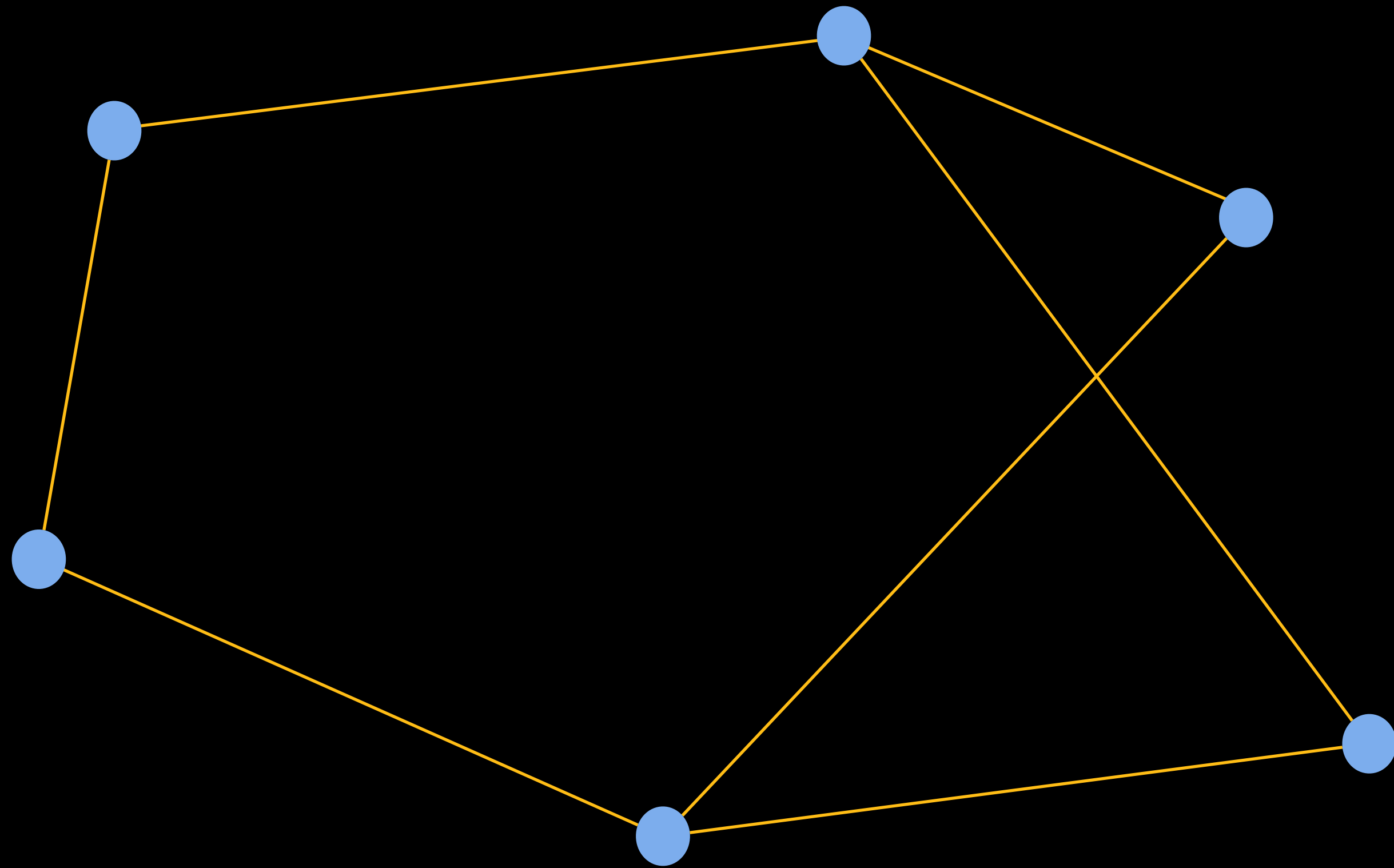
FRIENDSHIP

A relation is defined as a subset of the Cartesian product of two sets



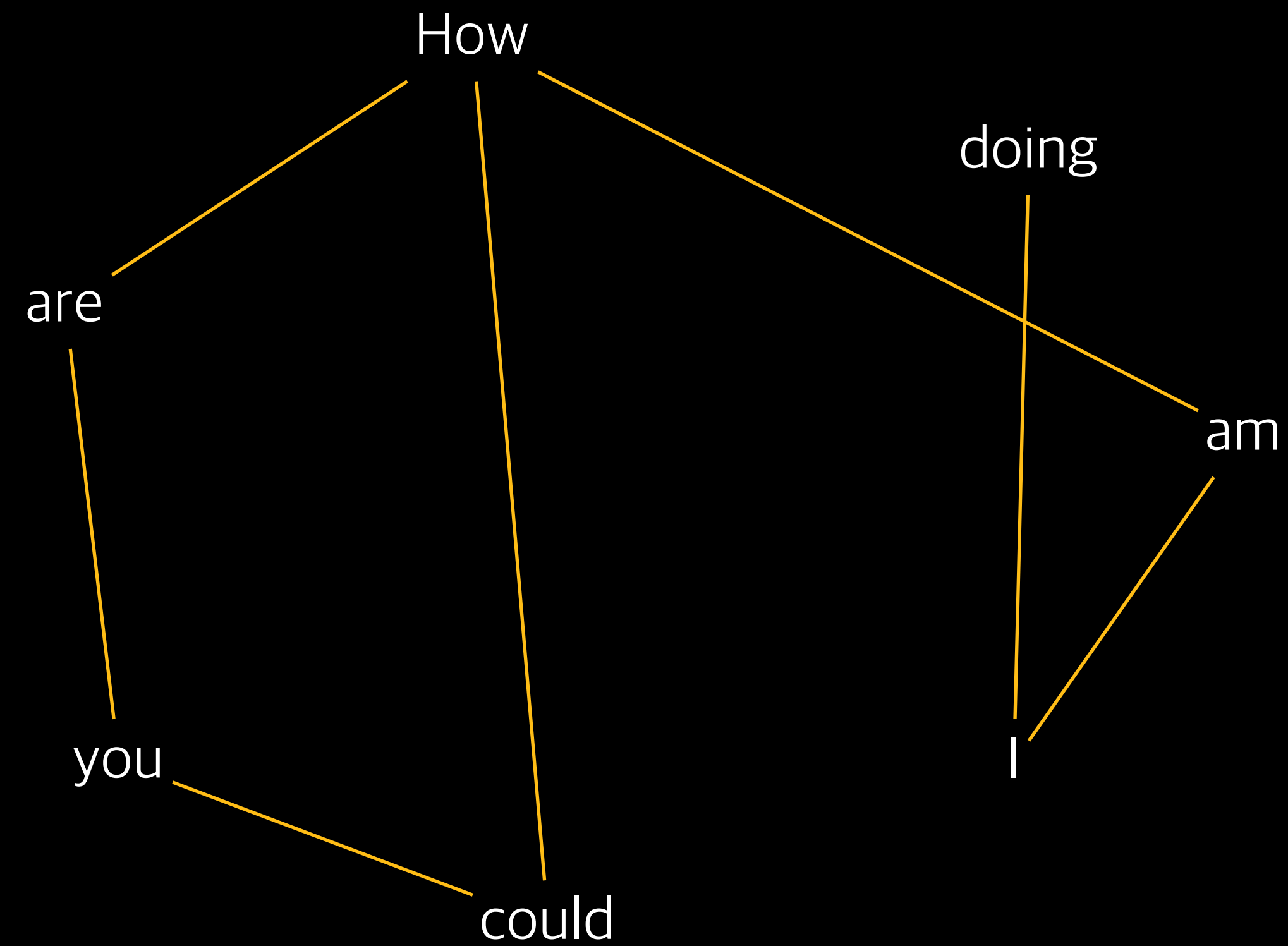
RELATIONSHIP AS A GRAPH

A relation is defined as a subset of the Cartesian product of two sets



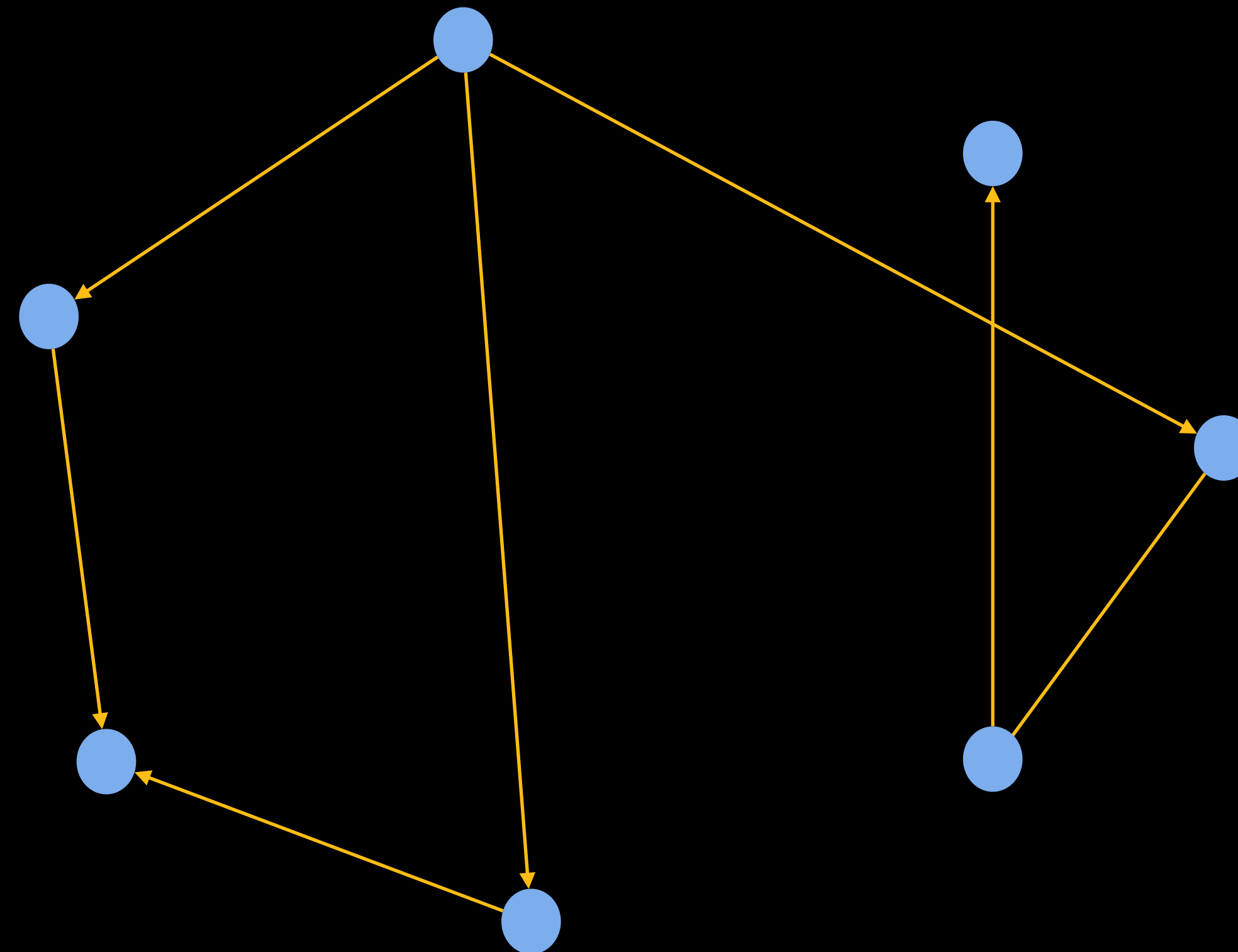
SENTENCES AS A GRAPH

How are you? How could you? How am I doing?



SENTENCES

How are you? How could you? How am I doing?

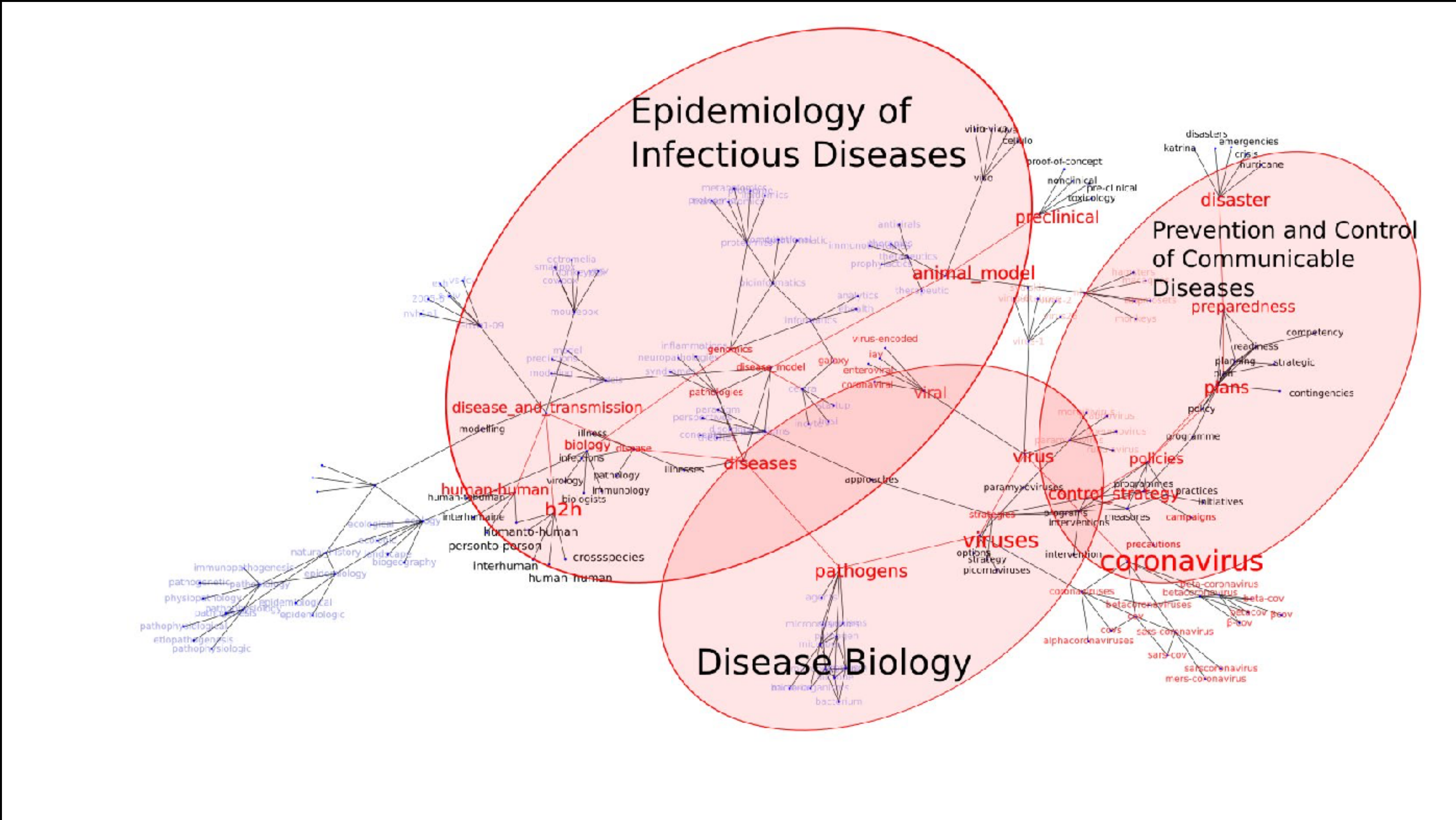


GRAPHS

Most impressive and immersive representation of combinatorics is Graph

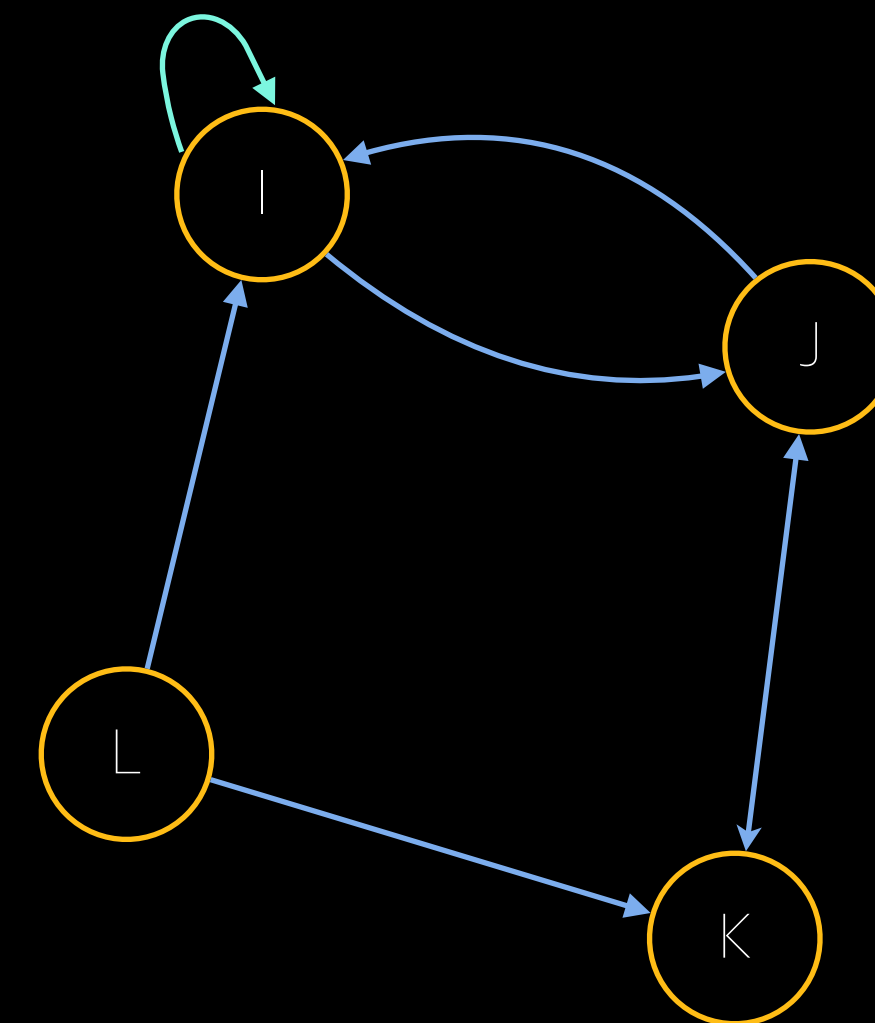
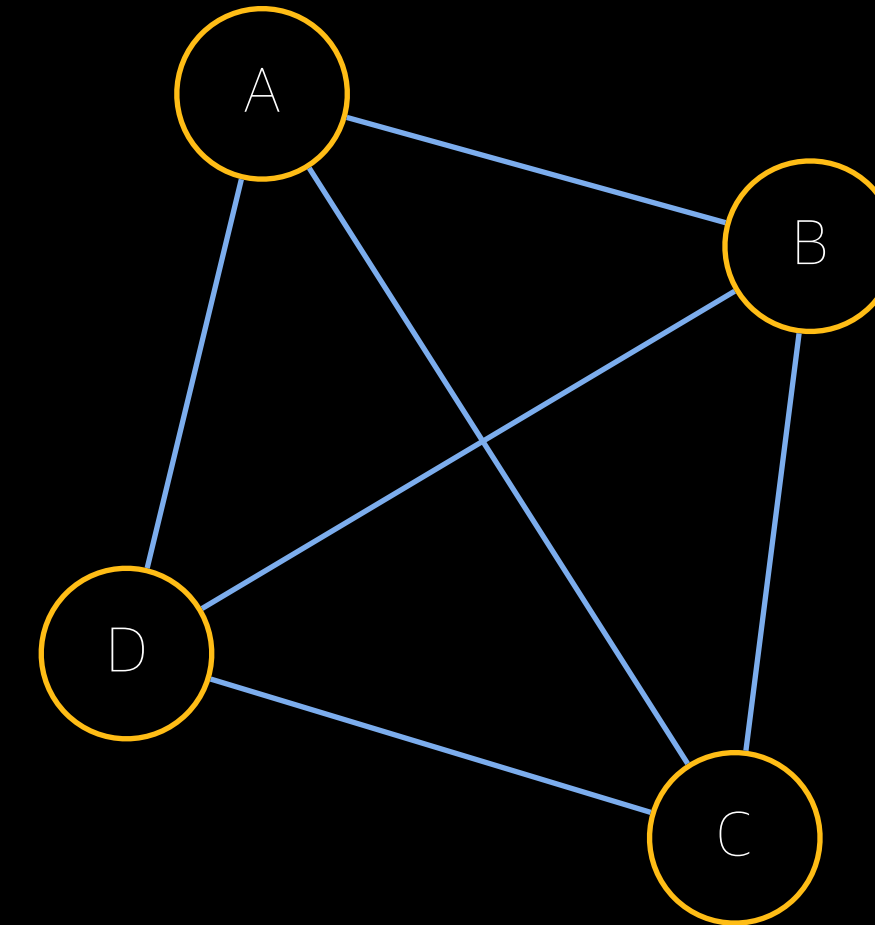
A graph can have cycles, and multiple paths between two nodes.

Non-linear data structure



FORMAL DEFINITION

- A graph is a pair of vertices (or nodes) connected by edges
- Represented as $G = (V, E)$
- $V = \{v_1, v_2, \dots, v_n\}$ or $v \in V$
- and $E = \{e_1, e_2, \dots, e_m\}$ or $e \in E$
- $E \subseteq \{(u, v) \mid u, v \in V\}$
- $n = |V|$ - number of vertices
- $m = |E|$ - number of edges



KNOWLEDGE GRAPH

A knowledge graph is represented as $G = \{E, R, F\}$, where E, R , and F are sets of entities, relations, and facts, respectively

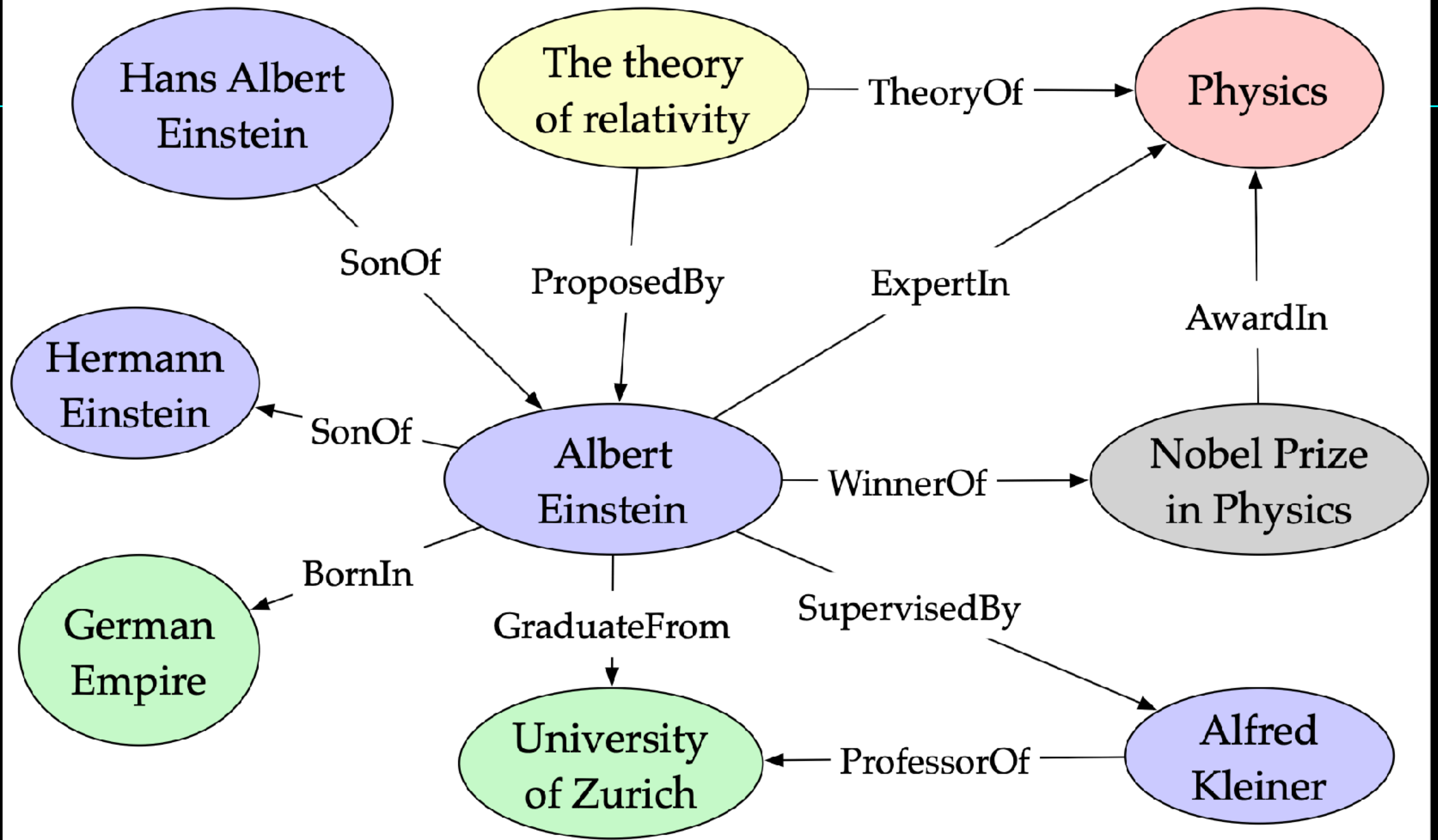
Knowledge can be expressed in a factual triple in the form as $(h, r, t) \in F$

h – head

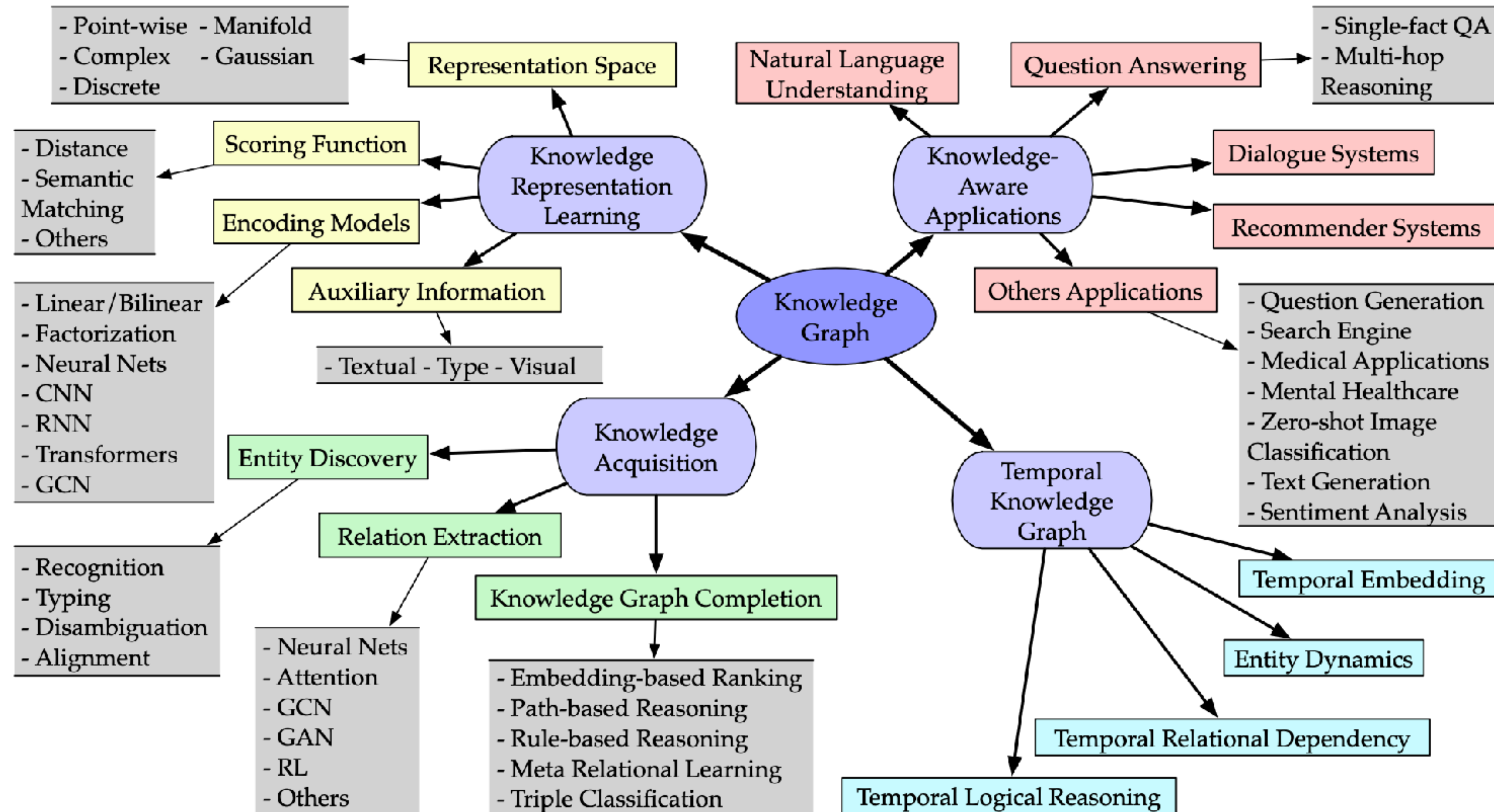
r – relation

t – tail

(Albert Einstein, **BornIn**, German Empire)
(Albert Einstein, **SonOf**, Hermann Einstein)
(Albert Einstein, **GraduateFrom**, University of Zurich)
(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
(Albert Einstein, **ExpertIn**, Physics)
(Nobel Prize in Physics, **AwardIn**, Physics)
(The theory of relativity, **TheoryOf**, Physics)
(Albert Einstein, **SupervisedBy**, Alfred Kleiner)
(Alfred Kleiner, **ProfessorOf**, University of Zurich)
(The theory of relativity, **ProposedBy**, Albert Einstein)
(Hans Albert Einstein, **SonOf**, Albert Einstein)



RESEARCH ON KNOWLEDGE GRAPH



KNOWLEDGE CONNECTION

Dr. Swanson, an information scientist, proved that two distinct knowledge sources together contain implications that cannot be seen within either of the two sets by using an independent lens

Reference: <https://www.ncbi.nlm.nih.gov/pubmed/3075738>

KNOWLEDGE REPRESENTATION LEARNING

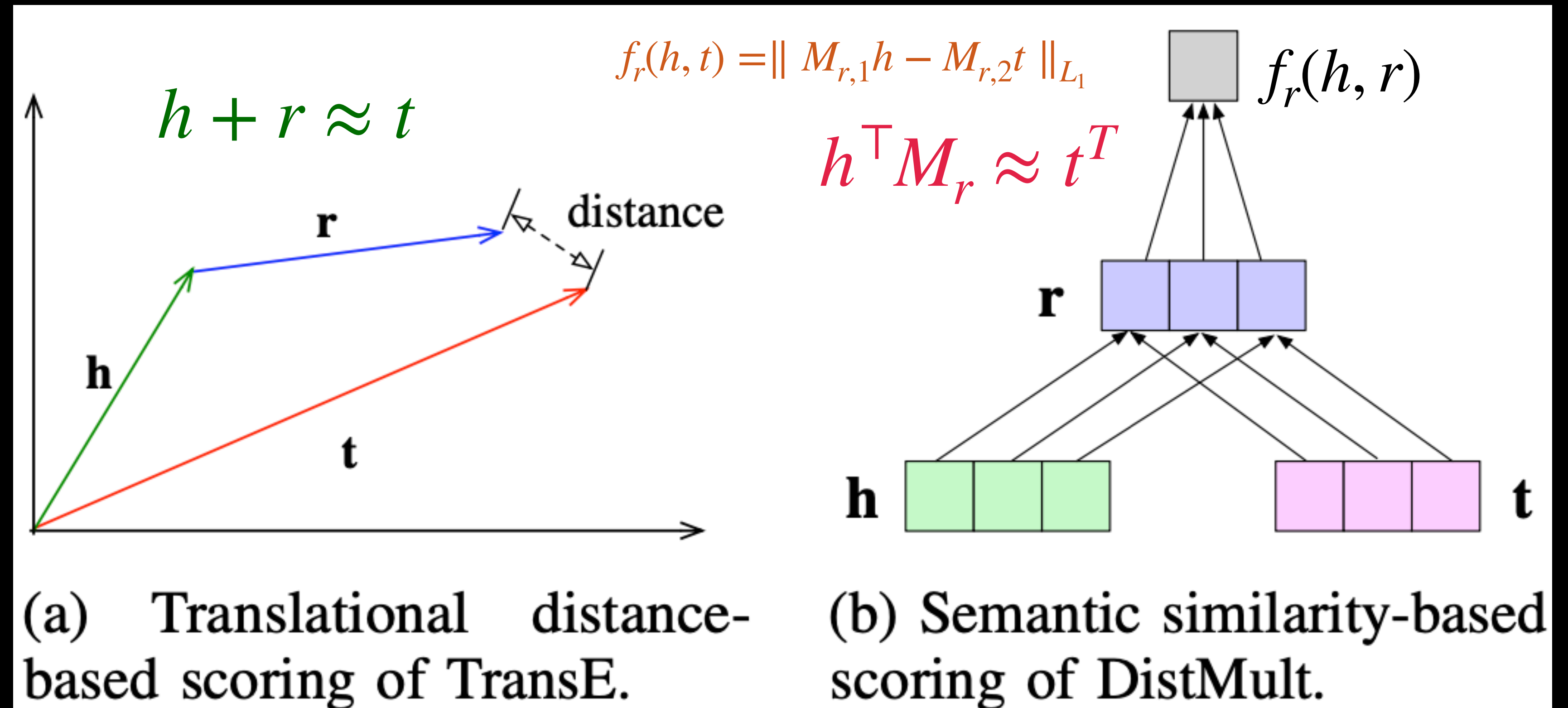
- ✦ Knowledge Representation Learning
 - ✦ representation space in which the relations and entities are represented;
- ✦ Scoring function
 - ✦ for measuring the plausibility of factual triples;
- ✦ Encoding models
 - ✦ for representing and learning relational interactions
- ✦ Auxiliary information
 - ✦ to be incorporated into the embedding methods

KRL

- ✦ Learn to represent the knowledge in the low-dimensional distributed embedding of entities and relations
- ✦ Represent entities and relations in a d -dimension vector space,
 - ✦ $(h, t, r) \in R^d$ and makes embeddings follow the translational principle $h + r \approx t$

SCORING FUNCTION

- ✦ The scoring/energy function is used to measure the plausibility of facts



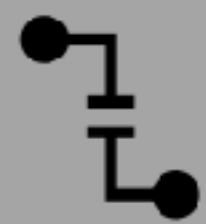
LITERATURE-BASED DISCOVERY

- ✦ Dr. Swanson believed that (early 1980) Computer systems could automate and aid
- ✦ information retrieval
- ✦ Volume of documents/research papers was rapidly increasing
- ✦ Excessive specialization could inhibit scientific creativity
- ✦ The current (early 1980s) information services and systems ability was inadequate to meet this new state of information needs

SEEMINGLY UNCONNECTED PUBLIC KNOWLEDGE



Dr. Swanson believed that unearthing unseen links between two distinct areas of study could yield new discoveries



He wanted to demonstrate the presence of these “undiscovered connections” between two different sets of disjoint literatures



He approached this problem by asking if two separately published articles one stating

"A may cause or influence B"

"B may cause or influence C"

Is it possible to conclude "A may cause or influence C?"

MIGRAINE AND MAGNESIUM: NEGLECTED CONNECTIONS

He had shown that how seemingly unconnected resources could be combined to form a new hypothesis, though he was not an expert in all the fields that he chose to converge

Two literatures "logically" related if the arguments they advance about the phenomena to which they respectively refer are related in some interesting way

Reference: <https://www.ncbi.nlm.nih.gov/pubmed/3075738>

FISH OIL ↔ RAYNAUD DISEASE



Raynaud's Syndrome is a circulatory disorder commonly noted by platelet aggregability, high blood viscosity, and vasoconstriction



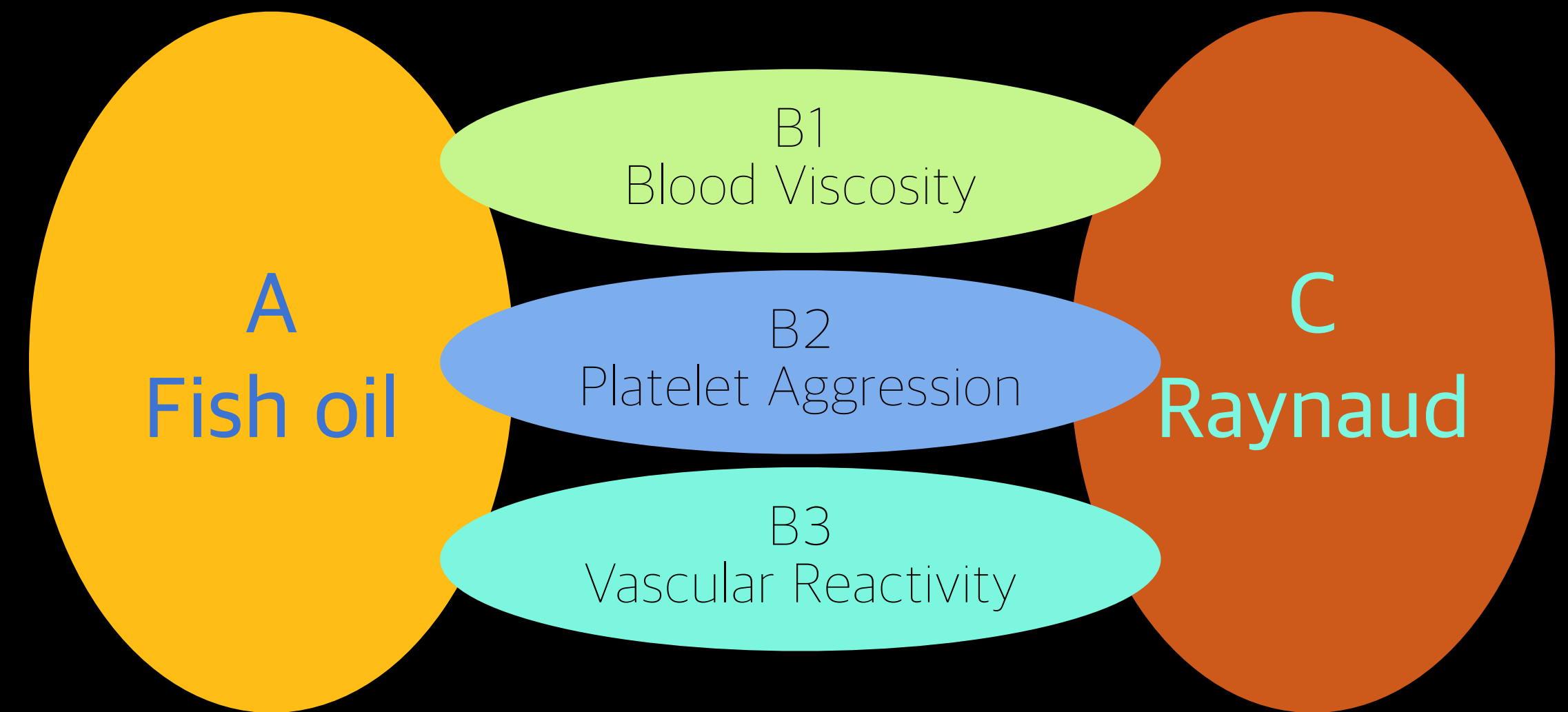
He found these connections by reviewing 2000 papers on Reynaud's Syndrome and 1000 on dietary fish oil



Dietary fish oil has been shown to reduce platelet aggregability, blood viscosity, and vascular reactivity



Both set of documents were disjoint with no citations, co-authorships, or connections between them

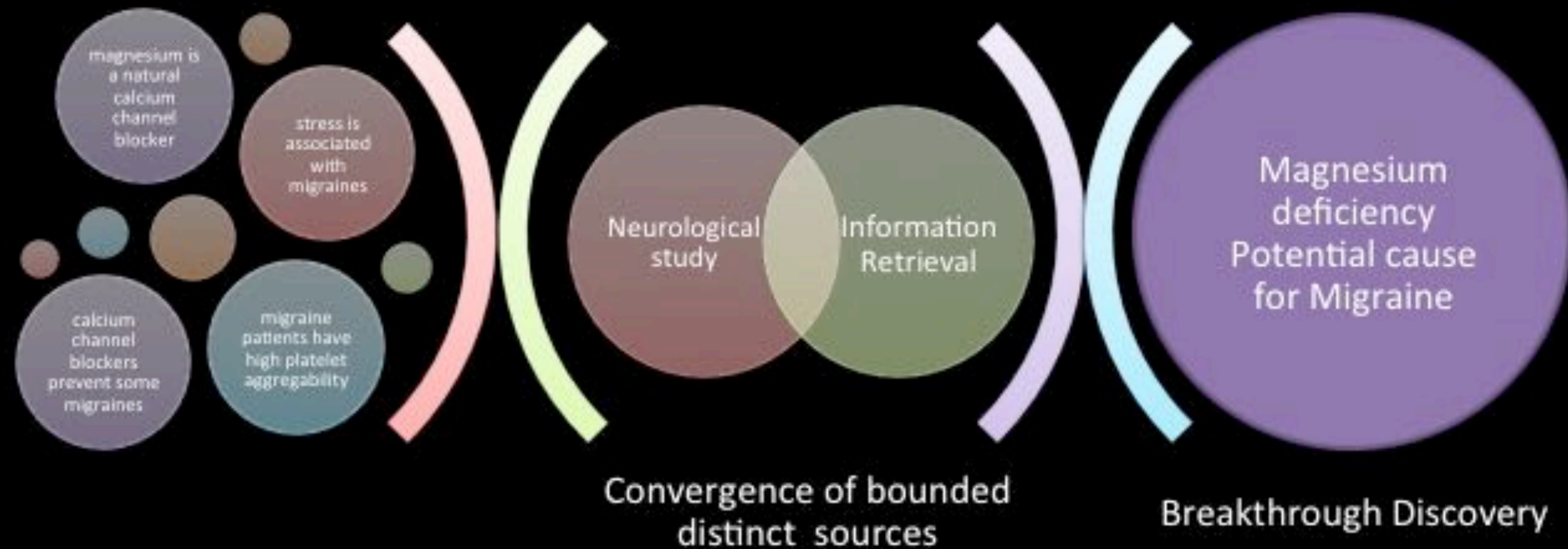


Source: https://www.researchgate.net/figure/Venn-diagram-representing-Swansons-first-discovery-Fish-oil-is-beneficial-for-patients_fig1_224890539

Three years after publishing his groundbreaking work (Swanson, 1986a), a clinical trial validated the use of fish oil for patients with Raynaud's disease (DiGiacomo, 1989).

Knowledge Convergence

Dr. Swanson, an Information Scientist has shown how chains of causal implication within the medical literature can lead to hypotheses for causes of rare diseases



MIGRAINE ↔ MAGNESIUM DEFICIENCY

- ✦ Stress is associated with migraines
- ✦ Stress can lead to loss of magnesium
- ✦ Calcium channel blockers prevent some migraines
- ✦ Magnesium is a natural calcium channel blocker
- ✦ Spreading cortical depression (SCD) is implicated in some migraines
- ✦ High levels of magnesium inhibit SCD
- ✦ Migraine patients have high platelet aggregation
- ✦ Magnesium can suppress platelet aggregation

