**Literature Summary**

1. **PrelimProposal**

\_ Objective: develop a new KG, assess validity of the data, use ontology to enrich the graph, gain insights from MS patients (having Covid vaccination).

\_ Quality control dimension:

* Accuracy
* Trustworthiness
* Timeliness
* Availability
* Completeness
* Consistency

\_ Resource: ImmunoGlobe, Bioteque, donors However, this is gene centric (while we want to focus on cells).

**Question:**

* Why choosing MS patients?
* How to use ontology on KG? (Create new connections by using machine learning, ...)
* How to evaluate the validity when the data is huge?

1. **Systematic integration of biomedical knowledge prioritizes drugs for repurposing**

\_ Objective: construct a Hetionet which is used for drug repurposing, identify patterns that distinguish treatments and non-treatments.

\_ Hetionet v1.0 consists of 47,031 nodes of 11 types and 2,250,197 relationships of 24 types.

\_ Performance measure: AUROC (area under the receiver operating curve). If it’s positive, paths of the given type tended to occur more frequently between treatments than non-treatments, after accounting for different levels of connectivity (node degrees) in the Hetnet.

**Question:**

* Treatments and non-treatments play what role in drug-repurposing?
* Graph explanation and many jargons

1. **Constructing knowledge graphs and their biomedical applications**

\_ Objective: Understand how machine learning helps create KGs and how KGs are applied to biomedical situations.

\_ Two methods:

* Manual: recall is low (publications rate is so high) need future approaches
* Semi-automatic: faster, but still miss lesser-known relationships (ambiguous).

\_ Text-mining: high recall, low precision (sometimes miss relevant information). Full text provided better prediction power than using abstracts alone.

* Unsupervised learning: better but suffer from technical issues of the algorithm. Hence, sentences should be simplified in the initial step.
* Supervised learning: SVM is used, but deep learning techniques are focused (for non-linear mappings) such as RNN, CNN, ...

**Question:**

* If sentences are all simplified, would it be much easier to make nodes and edges (no need for advanced technique)?
* Where do scientists publish the data on (a website, ...)?

1. **ImmunoGlobe: enabling systems immunology with a manually curated intercellular immune interaction network**

\_ Objective: create a network with interactions and detailed annotations.

\_ Immune systems are complex a barrier to scientists.

\_ From text manually create the network.

\_ Accuracy rate: 99.7%