APPLICATION-DRIVEN GENERATIVE MODELS FOR GRAPH AND TEXT

In this thesis, we mainly focus on designing \textit{generative models} for two types of non-euclidean data, namely, graph and text, with specific applications.
We primarily focus on the \textit{property-driven generative models} that generate samples bearing an essential property of interest from the learned representation.
Specifically, we address two property-driven generation problems related to graph and text: (i) Generative model for graphs to discover new molecular graph structures with specified properties like solubility or drug-likeliness, and (ii) Generative model for text to generate sentences with specified language distribution such as code-switched data (a sentence comprising of words from more than one language) and any other specified style feature such as sentiment, formality, etc.

To achieve a property-driven generation, we need to learn the data representation and the relation between the representation and the specified attribute of interest.
Other than the complexity of inherent non-euclidean structures that make the representation learning challenging, property tagged training data is rare for graphs and text. For instance, the actual octanol partition coefficient i.e. logP estimation for a given molecular structure, is time-consuming and expensive. Similarly, language-span tagged real code-switched data or sentiment tagged code-switched data and even monolingual data tagged with complicated attributes like formality are rare and difficult to curate. Data scarcity makes it difficult to learn the relation between representation and property.