Project Proposal

Project Title: Implementation of the Research Paper: “Delay Tolerant Network (DTN) Routing as a Machine Learning Classification Problem”

Student Name(s): Yilin Wang (yw4127), Yicheng Feng (yf2667)

Objective (what you propose to do):

- **Investigate the DTNs**: Conduct thorough research into the paper “Delay Tolerant Network (DTN) Routing as a Machine Learning Classification Problem” and its relevant references to further understand the basic principles of DTNs and the challenges they face.

- **Simulate the Real-World Scenarios for DTNs:** Investigates and applies a more comprehensive training dataset which can make use of real-world location traces and network traffic information. Additionally, selecting a suitable testing environment for simulation is also significant for enabling our solution to be evaluated in environments that are similar to actual operating settings of DTNs. Reaching this goal will guarantee that the solution we developed is adjusted to the subtleties and uncertainties of actual DTN settings, resulting in more precise, practical, and efficient routing choices.

- **Implement Machine Learning-Based Routing Approach**: Selects appropriate machine learning-based approaches for the routing strategies of DTNs with the goal of outperforming conventional routing methods in terms of flexibility and effectiveness.

- **Analyze Solution Performance**: To confirm the model's efficacy, carry out in-depth performance analyses using a variety of metrics. And compare the results under different learning strategies in the space of machine learning.

Significance (why the proposed project is interesting/important):

The proposed project holds significant importance due to the critical role DTNs play in various challenging communication environments, such as deep space satellite networks, rural areas, disaster-hit regions, and underwater communication systems. Traditional communication networks rely on stable connections and predictable network topology, conditions that DTNs do not meet due to their inherent operational challenges. By integrating machine learning into DTN routing, this project tries to reproduce the elegant results of a novel approach that adapts to the dynamic nature of DTNs, thereby improving data transmission reliability and efficiency.

Methodology (how do you plan to achieve the objective, i.e., machine learning model, data, simulation, analysis, etc.):

For this project, we aim to apply several machine learning classification approaches to predict a group of nearby nodes that are most probable to transmit a message to a specified location by utilizing historical data on message delivery. These machine learning models involve Decision Tree, K-Nearest Neighbors (KNN) classifier, and Naive Bayes. Additionally, we might also try to use the Adaptive Boosting (AdaBoost) and Extreme Gradient Boosting (XGBoost) approaches for further comparisons if time permits.

For the original paper, dataset was generated by a variety of tools such as Common Open Research Emulator (CORE), mobility scenario generation tool BonnMotion. However, for our project, we can directly use these datasets since they are public now instead of generating by ourselves for the purpose of implementation and reproduction, but it should be noted that we will still conduct careful investigation and develop thorough understanding into these datasets and their relationship with the actual operating scenarios of DTNs and select those that are still up-to-date for our project.

To evaluate our models, we will place them in four distinct multi-label classification settings: Independent Classifiers, Classifier Chains, Ensemble Methods, and Label

Powerset. To validate their performance, we will utilize two metrics: Hamming Loss and Zero-One Loss. Additionally, we will employ the Micro-averaged F1 Score and the Jaccard Similarity Score to further assess the results.

Delivery (what you expect to provide in the final report, e.g. code, graphs, analysis, etc.):

For the final report, it is expected that:

1. A complete code for implementing and reproducing the results in the paper can be developed and demonstrated.
2. The results of the performance metrics mentioned in **Methodology** can also be visualized for all the machine learning-based classification models under the four multi-label approach settings.
3. Compare the final results and select the most suitable machine learning based model for improving the performance of DTNs.
4. Conduct analysis and discussions on possible future work for improving the reliability and adaptability of DTNs.