Semester Summary

Huan Gui

Dec. 13
Project Description

• Information Diffusion Prediction
  – Data Sources: DBLP, Twitter
  – Goal: Given a topic, predict the future trend of this topic
    • How many authors would work on this topic in the future
Current Progress & Plan

• Apply Rare Logistic Regression
  – Unbalanced Data
  – Relatively accurate prediction on the next year [DBLP data].

• Plan
  – Information Propagates from the network effect
  – The sources of diffusion: Authors mentioned this topic before.
Courses

• Fall
  – CS 412 Introduction to Data Mining
  – CS 598 Health Infrastructure

• Spring
  – CS 512 Data Mining: Principles and Algorithms
  – CS 410 Text Information System
Question

Thank you!