

# 情報科学総合演習・実験

## 第2ラウンド



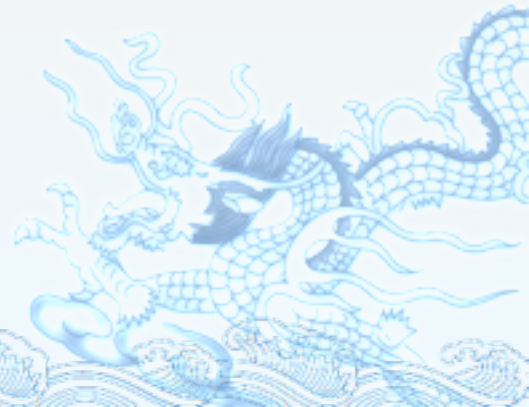
# Self-Introduction

- ◆ 岩崎謙汰 (いわさきけんた)
- ◆ 首藤研究室 修士1年
- ◆ 千葉県出身
- ◆ 特技 Unity30分制作
- ◆ 研究テーマ  
グラフサンプリング



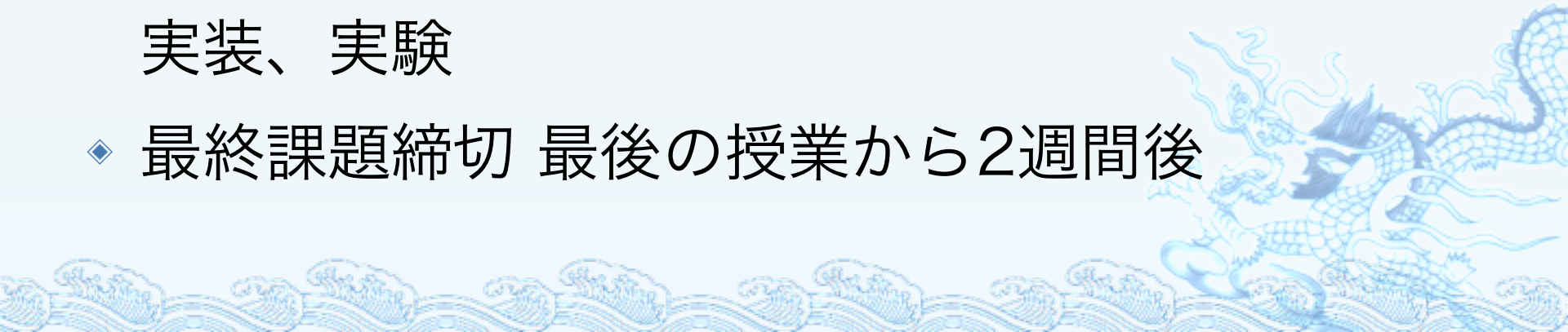
# Goal

- ◆ 英語の論文を輪講
  - ◆ Understanding Graph Sampling Algorithms for Social Network Analysis [IEEE 2011]
- ◆ グラフサンプリングを使ってソーシャルグラフの構造的特徴を推定する。  
(PythonのNetworkXを使用します)



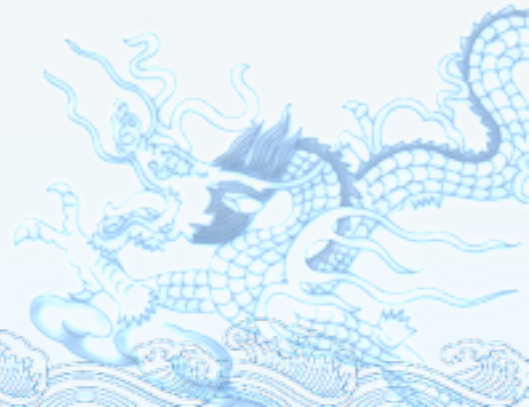
# Suchedule

- ◆ 第1週  
ガイダンス、基礎勉強、輪講担当割り振り、  
環境構築
- ◆ 第2週  
輪講
- ◆ 第3週以降  
実装、実験
- ◆ 最終課題締切 最後の授業から2週間後



# Lecture Style

- ◆ ミニ研究の体験
  - ◆ 英語の論文で既存手法を調べる
  - ◆ 議論 (輪講) によって理解を深める
  - ◆ インパクトのある主張や実験を報告する
- ◆ 大切にしたい雰囲気
  - ◆ 質問、議論しやすい雰囲気



# Question

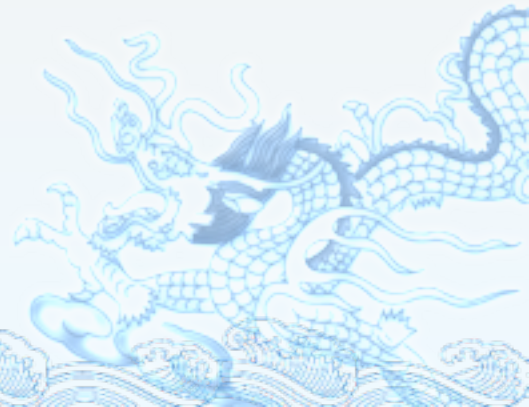
- ◆ 火曜4コマでいい？
- ◆ パソコン持ってますか？
- ◆ 連絡先は何がいいですか？





# 今回の演習の問題設定

- ◆ ソーシャルグラフの特徴を調べたい。
  - ◆ ソーシャルグラフの例  
Facebook, Twitter, 論文共著者ネットワーク
- ◆ しかし問題点が
  - ◆ 空間的and時間的コストがかかる
  - ◆ 取得できるデータに制限がある
- ◆ 解決策
  - ◆ サンプルングを行い、特徴を推定する。



# 今回の演習の問題設定

- ◆ ソーシャルグラフの特徴を調べたい。

- ◆ ソーシャルグラフの例

Facebook, Twitter, 論文共著

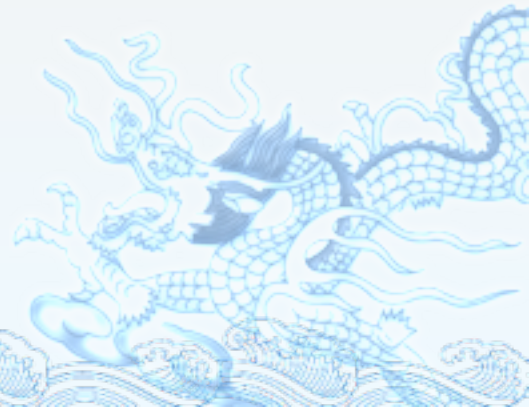
ソーシャルグラフの  
特徴って例えば何があ  
るの？

- ◆ ~~ソーシャルグラフの例~~  
サンプリングって既存手法  
ではどんなものがあるの？

かた  
ある

- ◆ 解決策

- ◆ サンプルングを行い、特徴を推定する。





# 今回輪講する論文

- ◆ Understanding Graph Sampling Algorithms for Social Network Analysis[IEEE 2011]
- ◆ 既存のグラフサンプリング手法を紹介し、比較実験を行っている論文。



# 論文の前に勉強

- ◆ 複雑ネットワークの基礎



# 複雑ネットワークのまとめ

- ◆ 複雑ネットワークの3つの特徴
  - ◆ スケールフリー性(次数分布がべき乗)
  - ◆ スモールワールド性(Lが大きい)
  - ◆ クラスタ性が高い(Cが大きい)



# そこで論文を読もう

- ◆ Understanding Graph Sampling Algorithms for Social Network Analysis[IEEE 2011]
- ◆ 英語なんて読みたくない…
- ◆ 一人で読むのは大変…
  - 日本語の論文なんてどうせ無いんだ
  - みんなで分担して読もう
  - 輪講しよう！



# Introduction

- ◆ 現在、FacebookやTwitterなどのOSNs(Online Social Networks)を分析する需要が高まってきている。



# Introduction

- ◆ しかし研究者がネットワークの全データを分析に使用するのは不可能である。
  - ◆ アクセス制限 (API制限)
  - ◆ プライバシーの問題
- ◆ **グラフサンプリングが有効**  
グラフの一部の情報から全体の特徴を推定する。  
少ないサンプル数で  
バイアスとバリエーションが小さい手法が望ましい。



# Introduction

- ◆ サンプリングアルゴリズム 既存手法
  - ◆ Breadth-First-Sampling(BFS)
  - ◆ Random Walk(RW)
  - ◆ Metropolis-Hasting Random Walk(MHRW)



# Introduction

- ◆ サンプリングの性能を計る指標
  - ◆ node degree distribution(NDD)  
次数分布
  - ◆ clustering coefficient(CC)  
クラスタ係数

