

# 首藤研究室総合演習：グラフ分野

## 1 概要

この総合演習では、グラフサンプリングについて取り扱います。1ヶ月弱と短い期間ですが、研究というものを体験していただきたいと思います。グラフ分野の総合演習は以下の日程と内容で行います。時間は毎回 7,8 限 (15:05-16:25) です。

**第 1 回** 授業の流れの説明, 複雑ネットワークとグラフサンプリングについて説明,

**第 2 回** グラフサンプリングの論文 [?] の輪講

**第 3 回 - 第 5 回** ソーシャルグラフのノード数とクラスタ係数の推定の論文 [?] の輪講

## 2 最終レポート (〆切 1/1 23:59)

最終レポートには以下の 2 つの内容を含めてください。

1. 演習で学んだこと (グラフの統計量やグラフサンプリングのモチベーションやサンプリング手法など)
2. 入力をグラフとサンプル数, 出力をそのグラフの統計量の推定値とするアルゴリズムの内容と評価実験.

推定する統計量は以下の 2 つから 1 つ選んでください。

- グラフの全ノード数  $N$  ( [?] の Neighbor collision 手法)
- グラフのローカルクラスタ係数  $c_l$  とグローバルクラスタ係数  $c_g$  ( [?] のランダムウォークベースの手法)

評価実験は以下の項目 (追加で考察してもよいです。 ) について自分なりの考えを述べ、行ってください。

- 推定値の誤差を評価するのに適した評価指標は何か？
- サンプル数はどのくらいまでが現実的か？
- 初期ノードはどのように決めるか？初期ノードによる精度の誤差はどの程度か？
- mixing time はどのように設定するか？ (グラフのノード数を推定する場合)

考察では、こうすれば [?] の手法よりさらに推定精度が良くなるんじゃないかというアイデアや、逆にここは現実の SNS に適用する上で厳しいんじゃないかなど、何か考えがあれば書いてください。