# Анализ сложных сетей.
# Домашнее задание 1.

Анвар Курмуков

13 Октября

## 1 Прежде чем начать..

1. Всего баллов за дз: 101.
2. Необходимо баллов для
получения максимальной оценки: 75
3. Дата выдачи дз: 13.10.2019.
4. Мягкий делайн: 28.10.2019.
5. Жесткий дедлайн: 4.11.2019
(-5 баллов из 101 за каждый день просрочки).
6. **Любые** формы плагиата будут наказы **полным** обнулением баллов за домашнее задание для **всех** участников действия.
7. 4.11.2019 это самый последний дедлайн (5.11.2019 новые работы приниматься не будут).
8. Все даты указаны по UTC+3.

### 1.1 Куда и что присылать

1. Сюда → kurmukovai@gmail.com.
2. **Тема письма** → Имя-Фамилия-complex-networks-HW1.
3. Присылать сформированный из .ipynb файла **.pdf документ**, рядом прикладывать сам .ipynb.

# 2 Степенное распределение

1. Сгенерируйте выборку со степенным законом распределения (4)
2. Изобразите гистограммы функции распределения и функции плотности (CDF, PDF) (4)
3. Изобразите тоже в log-log шкале (4)
4. Используя линейную регрессию определите угол наклона $\alpha$ (4)
5. Сравните PDF и CDF вашей выборки и теоретический с найденным $\alpha$ (4)

## 2.1 Power Law

$$p(x) = Cx^{-\alpha}$$

Из соображений нормировки (а также при $\alpha > 1$), получаем

$$1 = \int_{x_{min}}^{\infty} p(x) = \int_{x_{min}}^{\infty} Cx^{-\alpha} = C\frac{x_{min}^{(1-\alpha)}}{\alpha - 1}$$

Откуда

$$C = \frac{\alpha - 1}{x_{min}^{(1-\alpha)}}$$

Выпишем $F(x) = 1 - \int_x^{\infty} p(t) = 1 - \left(\frac{x}{x_{min}}\right)^{1-\alpha}$.

Для сэмплирования $x$ найдем $F^{-1}(x)$:

$$1 - y = \left(\frac{x}{x_{min}}\right)^{1-\alpha}$$

$$(1-y)^{\frac{1}{1-\alpha}} = \frac{x}{x_{min}}$$

$$x = (1-y)^{\frac{1}{1-\alpha}} x_{min}$$

Теперь взяв $y$ из равномерного распределения мы можем сэмплировать $x$ из power law.

## 2.2 CDF, PDF

CDF - Cumulative distribution function - Функция распределения $F(x)$
PDF - Probability density function - Функция плотности $p(x)$

1. Используйте matplotlib.pyplot.hist для изображения гистограмм распределения

2. Округлите значения $x$ до 2 знака после запятой и воспользуйтесь numpy.unique(x,return_counts=True) для построения распределений в log-log шкале (для этого воспользуйтесь matplotlib.pyplot.scatter).

3. для оценки параметра $\alpha$ вы можете воспользоваться функцией из sklearn или Normal equation.

# 3 Анализ сети соавторств ИТиС

1. Проведите анализ базовых свойст сети соавторств ИТиС (4)
   (a) Число вершин
   (b) Число ребер, плотность графа
   (c) Число связных компонент
2. Изобразите распределение степеней (degree distribution) (4)
3. Выполняется ли степенной закон распределения для сети ИТиС? Изобразите распределение степеней в log-log шкале, проведите расчет наклона $\alpha$ (4)
4. Для главной связной компоненты (**G**iant **C**onnected **C**omponnent) проведите расчет радиуса, диаметра и распределения кратчайших путей (radius, diameter, shortest path distribution) (4)
5. Проведите расчет основных метрик центральности вершин. Проведите сравнительный анализ результатов. (4)
   (a) Степенной (degree centrality)
   (b) По близости (closenness centrality)
   (c) Betweenness centrality
   (d) Eigenvector centrality

### 3.0.1 Уточнение

1. *Распределение кратчайших путей* следует воспринимать буквально - посчитали кратчайшие пути между всеми возможными парами вершин и построили эмпирическую функцию распределения этих путей.
2. Провести сравнительный анализ метрик центральности означает сравнить ранжирование вершин в соответствии с разными метриками центральности. Можно в ручную отсмотреть top/bot $k$ вершин в соответствие с разными центральностями, можно в придачу еще посчитать какой нибудь простой ранговый критерий.

## 3.1 NetworkX

Для анализа графов рекомундуется воспользоваться готовыми функциями из питоновского пакета NetworkX, он не слишком резвый, но на небольших графах работает достаточно быстро.

Для анализа сети соавторств ИТиСа предоставляется набор данных (название статьи, имя автора, год написания) и готовая питоновская функция get_coauthors (файл utils.py). Функция принимает на вход пандасовский дата фрейм, а на выходе выплевывает матрицу смежности (а также лист с именами авторов, в том же порядке что и строки в матрице смежности и общее число работ каждого автора). Сеть может быть построена для разного промежутка лет см. пример в докстринге функции.

# 4 Анализ эгоцентрической сети VK

## 4.1 Базовый анализ

1. Проведите анализ базовых свойст
   эгоцентрической сети друзей VK (4)
   (a) Число вершин
   (b) Число ребер, плотность графа
   (c) Число связных компонент
2. Изобразите распределение степеней (degree distribution).
   Выполняется ли степенной закон распределения для
   Вашей эгоцентрической сети? (4)
3. Проведите расчет основных метрик центральности вершин. Проинтерпретируйте результаты ранжирования вершин в соответствии с различными метриками центральности (4)
   (a) Степенной (degree centrality)
   (b) По близости (closenness centrality)
   (c) Betweenness centrality
   (d) Eigenvector centrality
4. Изобразите граф Вашей эгоцентрической сети. Используйте сетевые (центральности) и не сетевые (пол, возраст, город, место учебы) характеристики вершин в качестве цвета/размера для вершин. (4)
5. Воспользуйтесь алгоритмом поиска сообществ (из пакетов igraph или python-louvain), проанализируйте и прокомментируйте полученные результаты. (10)

## 4.2 Сравнение с разными моделями порождения графов

Заимплементируйте различные модели порождения сетей (Для выполнения этого задания **нельзя** использовать готовые функции NetworkX для порождения графов). Сгенерируйте из них сети на основе параметров своей эгоцентрической сети. Проведите сравнительный анализ, какая модель "наилучшим"образом описывает Вашу эгоцентрическую сеть? В каком смысле "наилучшим"?

1. Модель предпочтительного присоединения
   (Preferential attachment) (5)

2. Конфигурационная модель (Configuration model) (10)

3. Стохастически-блочная модель (Stochastic-block model) (10)

4. Воспользуйтесь оцененными параметрами Вашей эгоцентрической сети из пункта 4.1 для генерации модельных сетей. Какая из моделей "наилучшим"образом описывает Вашу эгоцентрическую сеть?

## 4.3 VK API

Для получения доступа к VK API вам нужно сперва получить access token, см. детали в .ipynb файле. Для сбора данных предоставляется функция get_friends_ids которая для заданного user_id скачивает информацию о друзьях пользователя с user_id.

Для получения дополнительной информации о пользователях (пол, возраст, город и пр.), воспользуйтесь методом users.get, см. пример в .ipynb файле.