

## Курсовой проект по курсу Лингвистические Сети

### *Модуль 1*

1. Построить лингвистическую сеть по одному из текстов (или коллекции текстов). На выбор, можно рассматривать co-occurrence (collocation) network, syntactic dependency network, semantic network
2. Исследовать полученную сеть методами теории социальных сетей:
  - a. Количество узлов, ребер, плотность графа
  - b. Найти функцию распределения степеней узлов (нарисовать распределение  $P(k)$  в log-log координатах, ). Вычислить среднюю степень узла
  - c. Предполагая, что распределение носит степенной характер power law, вычислить экспоненту распределения  $\alpha$ , указать  $p$ -value. Использовать MLE оценку.
  - d. Найти связанные компоненты графа (далее работать с самой большой).
  - e. Диаметр графа и среднюю длину пути
  - f. Коэффициент ассортативности (по степени узла)
  - g. Кластерный коэффициент (транзитивность) графа
  - h. Найти топ 10 узлов сети с максимальными значениями метрик центральности:
    - i. Degree centrality
    - ii. Closeness centrality
    - iii. Betweenness centrality
    - iv. PageRank
3. Доп. Задание. Сгенерировать случайный граф Erdos-Renyi с тем же числом узлов и ребер, что и лингвистический. Вычислить в нем среднюю степень узла, кластерный коэффициент, среднюю длину пути. Сравнить с результатами в пункте 2.

## Модуль 2

1. Найти core structure сети, выделить узлы наибольшего core
2. Найти сетевые сообщества, вычислить их модулярность, нарисовать дендрограмму в случае hierarchical clustering.
3. Найти наиболее часто встречаемые network motifs, вычислить diad и triad census сети.
4. Вычислить степень структурной схожести узлов, найти наиболее схожее, вычислить и визуализировать similarity matrix.
5. Визуализировать сеть, а также найденные сетевые сообщества используя средства iGraph/Python/R или программы yEd или Gephi.

Для выполнения проекта рекомендуется использовать библиотеку iGraph в Python или R.

По результатам курсовой работы необходимо предоставить:

1. письменный отчет
2. презентацию

Отчеты и презентации будут размещены на веб странице класса.

Отчеты и презентации необходимо прислать на адрес [lzhukov@hse.ru](mailto:lzhukov@hse.ru) в формате PDF .

Окончательный срок сдачи заданий 26/12/2013 в 23.59