Practicum 3

18 апреля 2025 г.

1 Базовые характеристики графов

1.1 Распределение степеней вершин

Если распределение степеней вершин имеет вид степенного закона $f(k) \sim k^{-\gamma}$ (распределение Парето), то говорят, что сеть является scale-free. Это название идет от свойства масштабной инвариантности степенного закона (инвариантность формы закона относительно $k \to \alpha$) — ни один из масштабов (имеется в виду масштаб связности, в нашем случае степени вершин) не доминирует в сети, и чем больше масштаб, тем больше вероятность его обнаружить. Это приводит к характерной особенности scale-free сетей — наличию в них хабов.

В файле `degree_distributions.py` мы рассматриваем два графа: сеть наиболее загруженных аэропортов США и граф Эрдеша-Реньи. Первый является характерным примером scale-free сети — действительно, хорошо известно, что некоторые аэропорты выполняют функцию хабов. Далеко не все графы обладают таким свойством. На самом деле интерес к scale-free сетям появился только к концу 90-х, когда стало очевидно, что WWW сеть содержит в себе много хабов и, следовательно, ее свойства принципиально отличаются от других случайных сетей, таких, как, например, граф Эрдеша-Реньи.

2 Меры центральности

Центральность вершины характеризует ее важность в данном графе. Мера центральности, следовательно, является вещественнозначным выражением центральности. Так как «важность» может пониматься по-разному в различных задачах, существует несколько конкурирующих мер:

- степень вершины,
- степень близости (closeness centrality),
- степень посредничества (betweenness centrality),
- степень влиятельности (eigenvector centrality),
- PageRank.

В файле `centrality_measures.py` мы реализуем и анализируем степени близости, посредничества и влиятельности.

2.1 Степень близости (closeness centrality)

Для вершины $v \in V$ графа $G = (V,E)$:

$$C(v) = \frac{1}{\sum_{u \in V} \mathrm{dist}(v,u)},$$

где $\mathrm{dist}(\cdot,\cdot)$ является функцией кратчайшего расстояния. При сравнении степеней близости в различных графах эту меру нормализуют коэффициентом $1/|V|$.

2.2 Степень посредничества (betweenness centrality)

Эта мера характеризует то, насколько данная вершина часто участвует во взаимодействии двух других вершин. Более конкретно, насколько часто она является мостом в кратчайшем пути между этими вершинами:

$$C(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)},$$

где $\sigma(s,t|v)$ есть количество кратчайших путей из s в t, проходящих через v, а $\sigma(s,t)$ есть общее количество кратчайших путей из s в t.

2.3 Степень влиятельности (eigenvector centrality)

Меры, характеризующие влиятельность, определяют ее рекурсивно через влиятельность соседей. Подобные рекурсивные формулировки часто приводят к задаче о собственных значениях.

Например, мы можем задать степень влиятельности данного узла как сумму степеней влиятельности соседей с некоторым коэффициентом λ:

$$C(v) = \frac{1}{\lambda} \sum_{u \in \mathrm{Nei}(v)} C(u) = \frac{1}{\lambda} \sum_{u \in V} a_{v,u} C(u),$$

где $a_{v,u}$ — элементы матрицы смежности A.

Сложив все значения $C(v)$ в вектор c, мы получаем задачу о собственных значениях матрицы смежности:

$$Ax = \lambda x.$$

Для данного собственного числа собственный вектор будет соответствовать степеням влиятельности каждой вершины. Так как степени влиятельности должны быть неотрицательными, то по теореме Фробениуса-Перрона единственное подходящее собственное число является наибольшим по модулю собственным числом (оно же по этой теореме будет вещественным и строго положительным).