1 Лекция 12

1.1 Сложные сети (complex networks)

Основой для количественного рассмотрения свойств сложных сетей является теория случайных графов. Простейшим объектом такого рода является семейство (случайных) графов с n вершинами $G_{n,p}$ таких, что вероятность наличия связи между любыми двумя вершинами равна p.

1.1.1 Основные характеристики

1. **Степень вершины**

   Степенью вершины называется количество ребер k, присоединенных к вершине. Простейшей характеристикой случайного графа является средняя степень вершины $z = \langle k \rangle$. Вычислим среднюю степень вершины для семейства графов $G_{n,p}$:

   - Среднее число ребер:
   $$\frac{1}{2}n(n-1)p$$
   - Среднее число узлов, которыми кончаются ребра:
   $$2 \cdot \frac{1}{2}n(n-1)p = n(n-1)p$$
   - Средняя степень вершины:
   $$z = \frac{1}{n}n(n-1)p = (n-1)p \sim np \quad [n \gg 1]$$

   Полный набор возможных степеней вершины описывается соответствующим вероятностным распределением $P(k) \equiv \{p_k\}$. Для семейства $G_{n,p}$ распределение по степеням вершины является биномиальным:
   $$p_k = \binom{n-1}{k} p^k (1-p)^{n-k-1} \tag{1}$$

   Действительно, для заданной вершины степень k есть результат наличия ребер, соединяющих эту вершину с k другими и отсутствия ребер, соединяющих его с $n-k-1$ оставшимися узлами. $n \to \infty$, $z = np \sim \text{const}$ распределение (1) переходит в распределение Пуассона
   $$p_k \simeq \frac{z^k \mathrm{e}^{-z}}{k!} \tag{2}$$

2. **Коэффициент кластеризации**

   Наличию нетривиальной кластеризации отвечает ситуация, в которой вероятность наличия ребра между узлами, имеющими общего соседа, выше, чем вероятность наличия ребра между произвольными узлами. Пусть C - средняя вероятность наличия ребра, соединяющего двух соседей некоторого узла. Очевидно, что для простейшей модели $G_{n,p}$

$$C = p = \frac{z}{n} \tag{3}$$

3. **Гигантская компонента**

   Гигантской компонентой называется кластер узлов, соединенных вершинами, размер которого (число узлов) пропорционально числу узлов связного графа n. Ниже мы убедимся, что для семейства $G_{n,p}$ порог образования гигантской компоненты определяется соотношением $z = np = 1$.

Свойства реальных сложных сетей отличаются от свойств простейшей модели $G_{n,p}$:

- Распределение по числу ребер в узле существенно непуассоново. При $k \gg 1$

$$p_k \propto \frac{1}{k^\alpha}, \quad \alpha < 1 \tag{4}$$

- Коэффициент кластеризации реальных сложных сетей гораздо больше его значения для $G_{n,p}$: $C \gg p$.

1.2 Среднее число соседей

Проще всего определить среднее число непосредственных соседей вершины

$$z_1 \equiv z = \langle k \rangle = \sum_k k p_k \tag{5}$$

Интереснее ситуация со вторыми соседями. Дело в том, что при "взгляде"со стороны потенциального второго соседа на одного из соседей первого уровня вероятность появления ребра, связывающего эти узлы, пропорциональна степени узла первого уровня. Для остальных ребер, выходящих из узла второго уровня, ничего не меняется. Для распределения q_{k-1} по числу ребер, выходящих из узла второго уровня (кроме ребра, соединяющего этот узел с узлом первого уровня) имеем поэтому

$$q_{k-1} = \frac{k p_k}{\sum k p_k} \quad \Rightarrow \quad q_k = \frac{(k+1) p_{k+1}}{\sum k p_k} \tag{6}$$

Для средней степени вершины, в которую мы пришли из одного из узлов первого уровня, имеем соответственно

$$\langle q \rangle = \sum_{k=0}^{\infty} k q_k = \frac{\sum k(k+1) p_{k+1}}{\sum k p_k} = \frac{\sum (k-1) k p_k}{\sum k p_k} = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \tag{7}$$

Полное среднее число вторых соседей пропорционально среднему числу первых, так что

$$z_2 = \langle k \rangle \langle q \rangle = \langle k^2 \rangle - \langle k \rangle \tag{8}$$

Очевидно, что переход от $m-1$-го уровня до m-го происходит с тем же коэффициентом размножения $\langle q \rangle$:

$$z_m = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} z_{m-1} \equiv \frac{z_2}{z_1} z_{m-1}, \tag{9}$$

так что

$$z_m = \left[\frac{z_2}{z_1} \right]^{m-1} z_1 \tag{10}$$

Естественным определением взрывного роста кластера (образования гигантской компоненты) является условие $z_2 = z_1$, т.е.

$$\sum k(k-2)p_k = 0 \tag{11}$$

1.3 Диаметр графа

Рассмотрим случай, когда гигантская компонента имеет место. Оценим число итераций l, необходимых для достижения границ кластера при старте в произвольном узле. Простая оценка возникает в предположении, что в результате итераций окажутся охваченными все узлы кластера:

$$z_l = \left[\frac{z_2}{z_1} \right]^{l-1} z_1 = n \tag{12}$$

Определенное таким образом число l называется диаметром графа. Из (12) получаем

$$(l-1) \ln \left(\frac{z_2}{z_1} \right) = \ln \left(\frac{n}{z_1} \right) \tag{13}$$

т.е.

$$l = 1 + \frac{\ln \left(\frac{n}{z_1} \right)}{\ln \left(\frac{z_2}{z_1} \right)} \tag{14}$$

1.4 Эмпирический коэффициент кластеризации

Рассмотрим ребро, выходящее из одного из соседей k_i первого уровня. Кластеризация связана с вероятностью того, что это ребро соединено с другим соседом k_j первого уровня, так что

$$C = \frac{\langle k_i k_j \rangle}{nz} = \frac{1}{nz} \left[\sum k q_k \right]^2 \tag{15}$$

т.е.

$$C = \frac{z}{n} \left[\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right]^2 \equiv \frac{z}{n} \left[c_v^2 + \frac{z-1}{z} \right]^2 \tag{16}$$

Очевидно, что если для реального графа $c_v > 1$, коэффициент кластеризации может сильно превышать свое значение для полностью случайног графа $C = z/n = p$.