Часть VII

Случайные графы

Разделы

Обозначения I

$\lfloor x \rfloor$ — *пол*: $\arg\max\limits_{n \in \mathbb{N}} \{n \leqslant x\}$; $\quad \lceil x \rceil$ — *потолок*: $\arg\min\limits_{n \in \mathbb{N}} \{x \leqslant n\}$;

K_n — полный граф на n вершинах;
$K_k \leqslant K_n$ — подграф K_k графа K_k, $k \leqslant n$.

$K_{p,q}$ — полный двудольный граф на двух подмножествах вершин (долях) по p и q вершин в каждой доле соответственно.

2^M — булеан множества M.
$\mathbf{2}^n$ — конечная булева алгебра с n атомами.

Случайное событие — подмножество пространства элементарных исходов.

$\mathsf{P}[A]$ — вероятность *события* $A \in \{0, 1\}$.

$$\mathsf{P}[A + B] = \mathsf{P}[A] + \mathsf{P}[B] - \mathsf{P}[A \cdot B]$$

$$\mathsf{P}[A \cdot B] = \mathsf{P}[A] \cdot \mathsf{P}[B], \quad \text{если события независимы}$$

Обозначения II

Если проводится n испытаний по схеме Бернулли с вероятностью единичного успеха p, то вероятность $P_{\geqslant 1}$ наблюдать хотя бы один успех —

$$P_{\geqslant 1} = 1 - (1-p)^n < np,$$

$$\mathsf{P}\left[\sum_i A_i\right] \leqslant \sum_i \mathsf{P}[A].$$

$\mathrm{E}[X]$ — математическое ожидание случайной величины X.

$\mathrm{D}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$ — дисперсия случайной величины X.

Неравенство Чебышёва для неотрицательной случайной величины X с конечным математическим ожиданием $\mathrm{E}[X]$:

$$\mathsf{P}[X \geqslant \varepsilon] \leqslant \frac{\mathrm{E}[X]}{\varepsilon}.$$

Обозначения III

Асимптотика при $n \to \infty$.

O: $f(n) = O(g(n)) \Leftrightarrow \left|\frac{f(n)}{g(n)}\right| \leqslant const$ — ограниченность сверху;

Ω: $f(n) = \Omega(g(n)) \Leftrightarrow \left|\frac{f(n)}{g(n)}\right| \geqslant const$ — ограниченность снизу;

Θ: $f(n) = \Theta(g(n)) \Leftrightarrow$ — ограниченность и сверху, и снизу;

o: $f(n) = o(g(n)) \Leftrightarrow \frac{f(n)}{g(n)} \to 0$.

Дискретная случайная величина: определение и примеры

Определение

Случайная величина называется *дискретной*, если пространство её элементарных исходов (носитель) $\Omega = \{\omega_0, \omega_1, \ldots\}$ не более, чем счётно.

Последовательность $p_0, p_1, \ldots$, где $p_i = \mathsf{P}[\omega_i]$, $i = 0, 1, \ldots$, называется *дискретным распределением* или *функцией вероятности* (ясно, что $p_i \geqslant 0$, $\sum_i p_i = 1$).

Примеры

- *Распределение Бернулли* $B(p, q)$: $\Omega = \{0, 1\}$, $p_0 = p$, $p_1 = q$.
- *Распределение Пуассона* $Po(\lambda)$: $\Omega = \{0, 1, \ldots\}$, $\lambda > 0$.

$$1 = e^{-\lambda}e^{\lambda} = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \ldots\right) = \sum_{i \geqslant 0} \underbrace{\frac{\lambda^i e^{-\lambda}}{i!}}_{p_i}$$

Дискретная случайная величина: примеры

- *Отрицательное биномиальное распределение* $NBi(k, p)$: $\Omega = \{0, 1, \ldots\}$, $0 < q < 1$, $p = 1 - q$:

$$1 = \left(\frac{1-q}{1-q}\right)^n = (1-q)^n \cdot \frac{1}{(1-q)^n} =$$

$$= (1-q)^n \cdot \sum_{k \geqslant 0} \overline{C}_n^k q^k = \sum_{k \geqslant 0} \underbrace{\binom{n+k-1}{n-1} p^n (1-p)^k}_{p_k}$$

— распределение дискретной случайной величины равной количеству произошедших неудач в последовательности испытаний Бернулли с вероятностью успеха p, проводимой до n-го успеха.

Дискретная случайная величина: примеры...

- *Биномиальное распределение* $Bi(n,k)$: результат n испытаний по схеме Бернулли с вероятностью $0 < p < 1$ каждое, $q = 1 - p$, $\Omega = \{0, 1, \ldots, n\}$:

$$1 = (p+q)^n = \sum_{i=0}^{n} \underbrace{\binom{n}{i} p^i q^{n-i}}_{p_i}$$

— вероятность наблюдения k успехов, $0 \leqslant k \leqslant n$.

Дискретные случайные события

Вероятность $\mathsf{P}[A]$ случайного события $A \subseteq \Omega$ есть $\sum\limits_{\omega \in A} p(\omega)$.

Классический способ задания вероятностей: количество элементарных исходов конечно и все они имеют одинаковую вероятность.

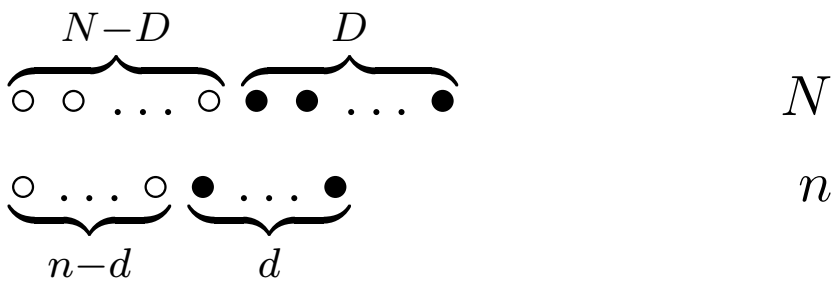
Тогда вероятность любого события при конечном $|\Omega|$ определяется как отношение его мощности (т.е. количества элементарных исходов, благоприятствующих данному событию) к общему числу элементарных исходов:

$$\mathsf{P}[A] = \frac{|A|}{|\Omega|}, \qquad A \subseteq \Omega.$$

Классический способ задания вероятностей: пример

Пример

- *Гипергеометрическое распределение* $HGe(k; N, D, n)$: имеется множество из N объектов, из которых D дефектных; если из данного множества делается случайная n-выборка, то какова вероятность, что в ней окажется ровно d дефектных объектов?



$$p_d = \frac{\binom{N-D}{n-d}\binom{D}{d}}{\binom{N}{n}}.$$

Разделы

Вероятностный метод: основная идея очень проста

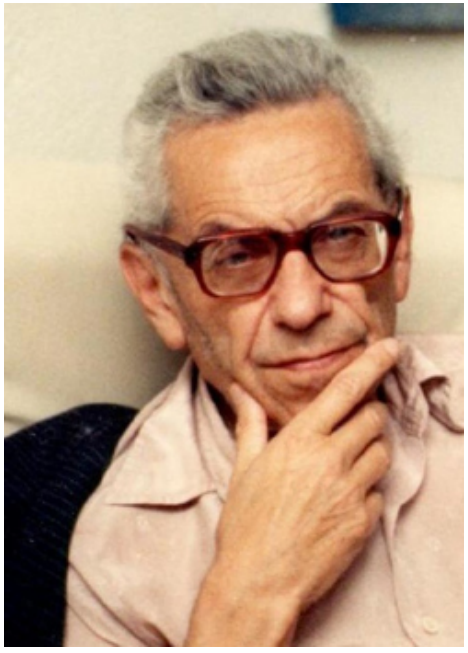
Для доказательства существования объекта с данными свойствами определяют подходящее вероятностное пространство объектов, а затем показывают, что при случайном выборе объекта вероятность наличия у него интересующих свойств строго положительна.

Доказательства в математике:

- *в конструктивных* доказательствах проводится предъявление требуемого объекта или алгоритм его построения;
- *в экзистенциальных* доказывается, лишь что объект существует, поскольку является элементом некоторого непустого множества.

Вероятностный метод, предложенный П.Эрдёшем в середине XX в., стал мощным инструментом для решения многих задач дискретной математики.

П. Эрдёш



Пал Эрдёш (венг. Erdős Pál, 1913–1996) — венгерский «странствующий математик», один из самых знаменитых учёных XX века. Работал в самых разных областях современной математики: комбинаторика, теория графов, теория чисел, математический анализ, теория множеств, теория вероятностей...

Его имя носит несколько десятков теорем, гипотез (некоторые из которых были впоследствии доказаны), констант, неравенств, графов, пространств, моделей...

Лауреат множества математических наград и основатель премии Эрдёша.

Количество написанных им научных статей, так же как и число соавторов не имеет аналогов среди современных ему математиков.

Число Рамсея

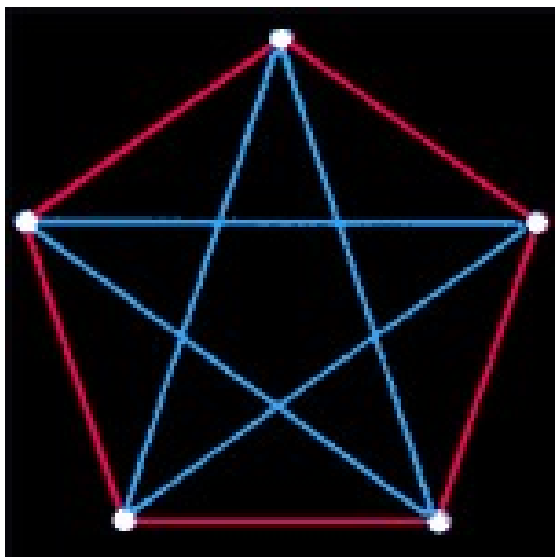
Определение

Число Рамсея $R(k,l)$ есть наименьшее целое n, такое, что *при любой раскраске ребер* полного n-вершинного графа $K_n(V,E)$ в синий и красный цвета либо существует красный подграф K_k, либо существует синий подграф K_l.

Рамсей показал, что число $R(k,l)$ конечно для любых k и l.

Гипотеза: $R(3,3) \stackrel{?}{=} 5$. Можно ли рёбра полного 5-вершинного графа раскрасить в синий и красный цвета так, чтобы не оказалось ни синего, ни красного треугольника?

Можно: ——————————————————→



Поэтому $6 \leqslant R(3,3)$.

«Лирическое отступление»: доказательство $R(3,3) \leqslant 6$

① Ясно, что любой раскраске рёбер графа K_6 в синий и красный цвета каждой его вершине инцидентно либо не менее 3 синих рёбер, либо не менее 3 красных.

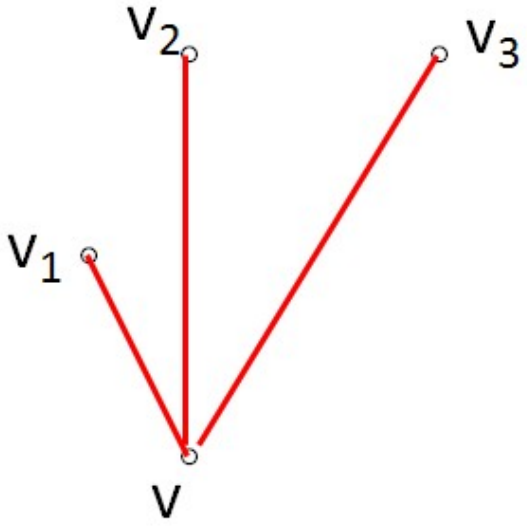
② Рассмотрим произвольную вершину $v \in V(K_6)$ и пусть ей инцидентно 3 красных ребра (v, v_1), (v, v_2), (v, v_3).
Тогда для рёбер (v_1, v_2), (v_2, v_3) и (v_3, v_1) возможны 2 случая:

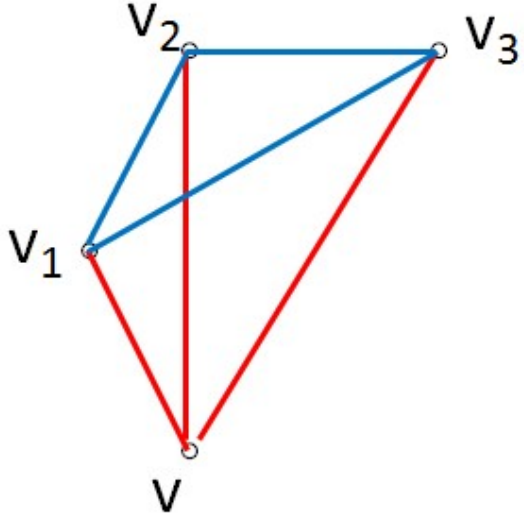
1) все они синие, и тогда они образуют синий треугольник;
2) хотя бы одно из этих рёбер красное, например, (v_1, v_2), и тогда рёбра (v, v_1), (v_1, v_2), (v_2, v) образуют красный треугольник.

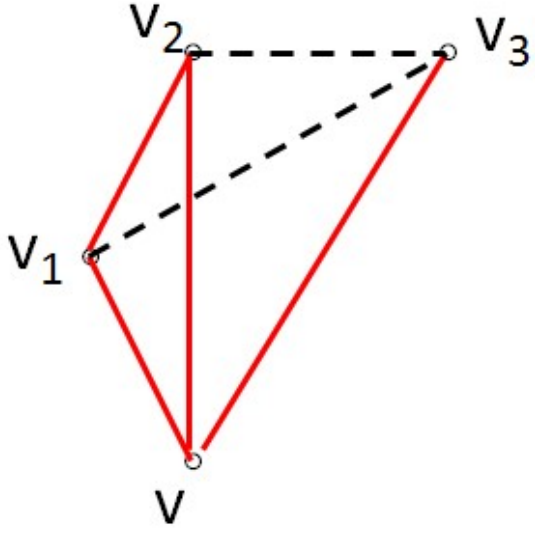
③ Если вершине $v \in V(K_6)$ ей инцидентно 3 синих ребра, рассуждения аналогичны.

Отсюда следует, что $R(3,3) = 6$.

Доказательство $R(3,3) \leqslant 6$: иллюстрация







Числа Рамсея трудновычислимы

Ясно, что $R(1,l)=1$ и $R(2,l)=l$.

Некоторые известные значения $R(k,l)$ и их оценки:

$R(k,l)$	3	4	5	6	7	8	9	10
3	6	9	14	18	23	28	36	40...42
4	9	18	25	36...41	49...61	58...84	73...15	92...149

Числа Рамсея чрезвычайно трудно вычислять: число C_n всевозможных 2-раскрасок рёбер полного n-вершинного графа равно $2^{\binom{n}{2}}$, и, например

$$|V(K_{40})| = \tfrac{40\cdot 39}{2} = 780 \text{ и } C_{40} = 2^{780} \approx 6{,}36\cdot 10^{234}.$$

Эрдёш полагал, что в случае крайней необходимости человечество ещё способно найти $R(5,5)$, но не $R(6,6)$:

$$43 \leqslant R(5,5) \leqslant 49, \qquad 102 \leqslant R(6,6) \leqslant 165.$$

Он нашёл способ получить нижнюю оценку диагональных чисел Рамсея, используя вероятностный метод.

Нижняя оценка диагонального числа Рамсея

— пример применения вероятностного метода.

Утверждение

Если $C_n^k \cdot 2^{1-C_k^2} < 1$, *то* $R(k,k) > n$.
Таким образом, $R(k,k) > \lfloor 2^{k/2} \rfloor$ *для всех* $k \geqslant 3$.

Доказательство

Рассмотрим случайную раскраску ребер графа K_n в красный и синий цвета равновероятно и независимо друг от друга.

Определим для каждого подграфа $K_k \leqslant K_n$, $k \leqslant n$ событие

$$M_k = 1 \Leftrightarrow \textit{подграф } K_k - \textit{ монохроматический}$$

(или одноцветный, все его ребра являются либо являются красными, либо синими).

Событие M_k — монохроматичность k-вершинного подграфа

Ясно, что $\mathsf{P}[M_k] = \frac{2}{C_k^2} = 2^{1-C_k^2}$

(два варианта из $2^{C_k^2}$ возможных 2-цветных раскрасок всех C_k^2 рёбер подграфа).

Т.к. существует C_n^k вариантов выбора K_k, то вероятность P того, что по крайней мере одно из событий M_k произойдет —

$$P = 1 - (1 - \mathsf{P}[M_k])^{C_n^k} < C_n^k 2^{1-C_k^2} < 1,$$

а что не произойдёт — $1 - P > 0 \Rightarrow$ существует 2-раскраска графа K_n без одноцветных подграфов, т.е. $R(k,k) > n$.

Если $k \geqslant 3$ и $n = \lfloor 2^{k/2} \rfloor$, то

$$C_n^k 2^{1-C_k^2} = \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k!} \cdot 2^{1-k^2/2+k/2} < \frac{n^k}{k!} \cdot \frac{2^{1+k/2}}{2^{k^2/2}} < 1,$$

и, следовательно, $R(k,k) > n = \lfloor 2^{k/2} \rfloor$ для всех $k \geqslant 3$.

Вероятностный метод и принцип Дирихле

В данном примере можно обойтись без вероятности, получив доказательство комбинаторным путем: ясно, что число $2^{C_k^2}$ 2-раскрасок графа K_n больше числа тех, что содержат монохроматический подграф K_k.

Принцип Дирихле: если n кроликов рассажены в k клеток, то гарантировать, что в одной из клеток находится более одного кролика, можно если $n > k$, а если $k > n$, то как минимум одна клетка пуста.

Вариант — «принцип голубей и ящиков» (Pigeonhole principle).

Теория Рамсея обобщает этот принцип.

Ф.П. Рамсей

Фрэнк Плампton Рамсей
(Frank Plumpton Ramsey, 1903–1930)
— английский математик, успевший внести также значительный вклад в философию и экономическую науку.

«Он доказал, что упорядоченные конфигурации неизбежно присутствуют в любой большой структуре.

...

Если взять пример со звёздами, то всегда можно найти в ней группу, которая с очень большой точностью образует какую-нибудь заданную конфигурацию: прямую линию, прямоугольник и др. Фактически теория Рамсея утверждает, что любая большая структура обязательно содержит упорядоченную подструктуру».

= полная неупорядоченность невозможна.

Вероятностный метод: алгоритмический аспект

Из доказательства $R(k,k) > \lfloor 2^{k/2} \rfloor$, $k \geqslant 3$ следует, что существует реберная 2-раскраска графа K_n *без одноцветных клик* $K_{2\log_2 n}$.

Как найти такую раскраску явно? Полная проверка *одного* подграфа K_k — $2^{C_k^2} = 2^{k^2/2-k/2}$ вариантов.

Для больших k при $n = \lfloor 2^{k/2} \rfloor$ выполнено

$$C_n^k \cdot 2^{1-C_k^2} < \frac{2^{1+k/2}}{k!} \left(\frac{n}{2^{k^2/2}} \right)^k \leqslant \frac{2^{1+k/2}}{k!} \ll 1.$$

$\Rightarrow$ случайная раскраска графа K_n с большой вероятностью не содержит одноцветных подграфов подграфов $K_{2\log_2 n}$.

Чтобы представить явно 2-раскраску ребер графа K_{1024} без одноцветных подграфов K_{20}, то можно подбросить правильную монету C_{1024}^2 раз и получить требуемую раскраску: вероятность того, что раскрашенный полный граф содержит одноцветный подграф K_{1024} меньше $\frac{2^{11}}{20!} \approx 8,4 \cdot 10^{-16}$.

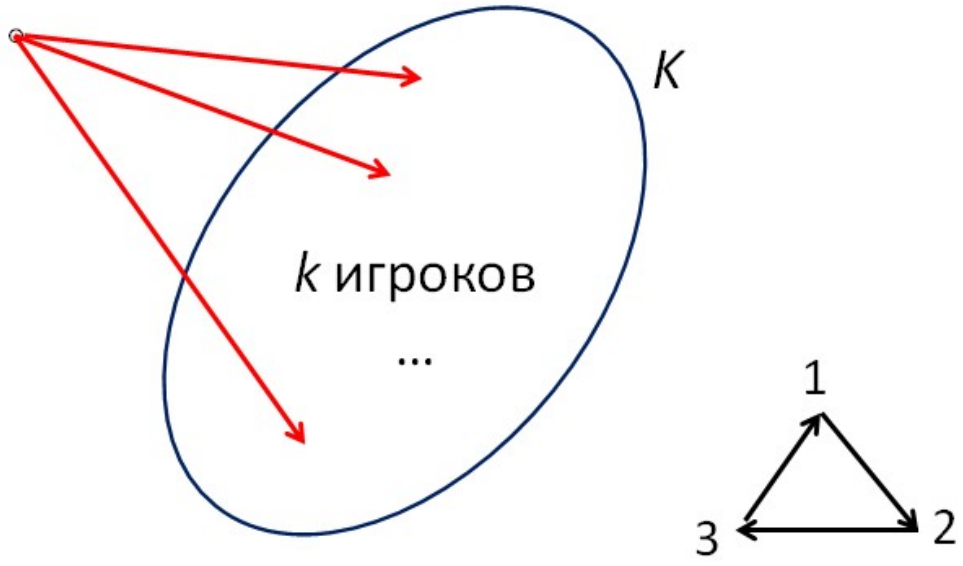
Турниры

Определение

Турнир T на множестве из n игроков есть результат ориентации ребер K_n (= полный n-вершинный орграф).

Турнир T обладает свойством S_k, *если для каждого* k*-элементного подмножества* игроков найдется хотя бы один игрок, который побеждает их всех — *победитель*.

Например, ориентированный треугольник с $V = \{1, 2, 3\}$ и $E = \{(1, 2), (2, 3), (3, 1)\}$ обладает свойством S_1.



Проблема Шютте

— верно ли, что для любого конечного k существует турнир T (с более чем k вершинами), обладающий свойством S_k?

Случайный турнир — турнир, получаемый из полного графа выбором для каждой пары его вершин u и v независимо и равновероятно либо дуги (u, v), либо дуги (v, u).

При этом все $2^{C_n^2}$ возможных турниров на множестве V равновероятны.

Идея: если n достаточно велико по сравнению с k, то случайный турнир на множестве $V = \{1, \ldots, n\}$ с большой вероятностью обладает свойством S_k.

K. Шютте

Курт Шютте
(нем. Kurt Schüette, 1909–1998)
— немецкий математик.
Последний аспирант Д. Гильберта.
В 1961-63 гг. по приглашению К. Гёделя работал в Институте перспективных исследований (г. Принстон, США).

Заведовал кафедрой Математической логики в университете Людвига-Максимилиана (г. Мюнхен).

Работал в области теории доказательств и порядкового анализа.

Автор фундаментальной монографии «Теория доказательств».
В 1966 г. был пленарным докладчиком на Международном конгрессе математиков в Москве.

Проблема Шютте: решение

Теорема (Эрдёш, 1963)
Если $C_n^k \left(1 - 2^{-k}\right)^{n-k} < 1$, *то существует турнир на* n *вершинах, обладающий свойством* S_k.

Доказательство

Рассмотрим случайный турнир на множестве $V = \{1, \ldots, n\}$.

Для каждого k-элементного подмножества вершин $K \subseteq V$ определим событие N_k, состоящее в том, что *не существует* победителя K.

Поскольку для каждой вершины $v \in V \setminus K$ вероятность того, что v *не есть* победитель K равна $1 - 2^{-k}$, и все $n - k$ событий, соответствующих различным выборам вершин v, независимы, то

$$\mathsf{P}[N_k] = \left(1 - 2^{-k}\right)^{n-k}.$$

Проблема Шютте: решение...

Оценим вероятность P того, что *хотя бы одно* событие N_k произошло:

$$P = \mathsf{P}\left[\sum_{\substack{K\subseteq V\\|K|=k}} N_k\right] \leqslant \sum_{\substack{K\subseteq V\\|K|=k}} \mathsf{P}[N_k] = C_n^k \left(1-2^{-k}\right)^{n-k} < 1.$$

Следовательно, с положительной вероятностью $1-P$ ни одно из событий N_k *не происходит* $\Rightarrow$ существует турнир на n вершинах, обладающий свойством S_k.

Разделы

Модель Эрдёша-Реньи

— исторически первая модель случайного графа.

На множестве $V_n = \{1, \ldots, n\}$ вершин вводим множество рёбер E, соединяя пару вершин ребром с заданной вероятностью $p \in (0, 1)$ независимо от всех остальных пар вершин.

Т.о. ребра появляются в соответствии со схемой Бернулли, из C_n^2 испытаний с вероятностью единичного успеха p.

Ясно, что E — случайное множество.

$G(n, p) = (V_n, E)$ — случайный граф в модели Эрдёша-Реньи.

Модель Эрдёша-Реньи: вероятностное пространство

В аксиоматике Колмогорова получим дискретное вероятностное пространство $G(n,p) = (\Omega_n, \mathcal{F}_n, \Pr_{n,p})$, в котором

$$\Omega_n = \{G = (V_n, E)\}, \quad \mathcal{F}_n = \mathbf{2}^{\Omega_n}, \quad \Pr_{n,p}[G] = p^{|E|}(1-p)^{C_n^2 - |E|}.$$

При $p = \frac{1}{2}$ вероятность выбора любого графа равна $2^{-C_n^2}$.

Элементы сигма-алгебры $\mathcal{F}_n$ — наборы графов.

Вероятность, с которой граф на n вершинах обладает данным свойством A:

$$\Pr_{n,p}[\mathcal{A}] = \sum_{G \in \mathcal{A}} \Pr_{n,p}[G],$$

где множество $\mathcal{A} \in \mathcal{F}_n$ состоит из всех графов, для которых выполнено свойство A.

Свойство A *выполнено почти всегда*, если

$$\Pr_{n,p}[\mathcal{A}] \to 1 \text{ при } n \to \infty.$$

Модель Эрдёша-Реньи: транспортная интерпретация

Представим себе, что в некоторой стране есть n городов, которые попарно соединены дорогами (сильное предположение, однако).

Допустим, каждая из дорог за определенный срок изнашивается (т.е. становится непроезжей) с известной вероятностью q и износ данной дороги никак не зависит от износа остальных дорог.

Вопрос: какова максимальная вероятность q, при которой с вероятностью больше $1/2$ ещё не исчезнет возможность перемещения между любыми двумя городами?

Модель Эрдёша-Реньи: транспортная интерпретация...

Сопоставим

- каждому из n городов сопоставим вершину $i \in V_n$,
- дороге между городами i и j — соответствующее ребро,
- а износ дороги — исчезновение ребра.

Тогда утверждение «дорога изнашивается с вероятностью q» $=$ утверждению «ребро появляется с вероятностью $p = 1 - q$».

Таким образом, нас интересует, какова минимальная вероятность p, при которой в модели Эрдёша-Реньи $G(n, p)$ вероятность связности графа больше половины.

Поскольку предположение о том, что все города попарно связаны дорогами — чересчур сильное $\Rightarrow$ далее приводится модификация модели Эрдёша-Реньи, в которой это предположение можно будет адекватно ослабить.

Надежность сети

Теорема

Пусть $p = \frac{c \ln n}{n}$ *в модели* $G(n, p)$.
Если $c > 1$, *то почти всегда случайный граф связен,*
если $c < 1$, *то почти всегда случайный граф не является связным.*

Доказательство

1. Случай $c > 1$.

Введем случайную величину на пространстве $G(n, p)$:

$$X_n = X_n(G) = \begin{cases} 0, & \text{если } G \text{ связен,} \\ k, & \text{если у } G \text{ ровно } k \text{ компонент.} \end{cases}$$

Т.о. X_n принимает значения $0, 1, 2, \ldots$.
Покажем, что $\Pr_{n,p}[X_n = 0] \to 1$ при $n \to \infty$, что равносильно $\Pr_{n,p}[X_n > 1] \to 0$.

Надежность сети: доказательство для случая $c > 1$

По неравенству Чебышёва: $\Pr_{n,p}[X_n \geqslant 1] \leqslant \mathrm{E}[X_n]$, и нам остается обосновать стремление к нулю $\mathrm{E}[X_n]$.

Представим X_n в виде суммы

$$X_n = X_{n,1} + \ldots + X_{n,n-1},$$

где $X_{n,k} = X_{n,k}(G)$ — число k-вершинных компонент графа G.

Занумеруем все k-элементные подмножества множества вершин V_n случайного графа в некотором (произвольном) порядке: $K_1, \ldots, K_{C_n^k}$.

Тогда в свою очередь $X_{n,k} = X_{n,k,1} + \ldots + X_{n,k,C_n^k}$, поскольку

$$X_{n,k,i} = X_{n,k,i}(G) = \begin{cases} 1, & \text{если } K_i \text{ образует компоненту } G, \\ 0, & \text{иначе.} \end{cases}$$

Надежность сети: доказательство для случая $c > 1$...

В итоге

$$\mathrm{E}[X_n] = \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} \mathrm{E}[X_{n,k,i}].$$

Очевидно,

$$\mathrm{E}[X_{n,k,i}] = \Pr_{n,p}[K_i \text{ образует компоненту в } G] \leqslant$$
$$\leqslant \Pr_{n,p}[\text{из } K_i \text{ в } V_n \setminus K_i \text{ нет рёбер в } G].$$

Получая последнее неравенство, мы просто пренебрегли условием связности той части графа G, которая «сидит» на множестве вершин K_i (такую часть принято называть *индуцированным подграфом*, символически $G|_{K_i}$).

Надежность сети: доказательство для случая $c > 1$...

Далее,

$$\Pr_{n,p}[\text{из } K_i \text{ в } V \setminus K_i \text{ нет ребер в } G] = (1-p)^{k(n-k)},$$

и, значит,

$$\mathrm{E}[X_n] \leqslant \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} (1-p)^{k(n-k)} = \sum_{k=1}^{n-1} C_n^k (1-p)^{k(n-k)}.$$

Последняя сумма симметрична в том смысле, что её слагаемые при k и $n-k$ равны. Рассмотрим $k = 1$:

$$n(1-p)^{n-1} \leqslant ne^{-p(n-1)} = ne^{-\frac{c(\ln n)(n-1)}{n}} =$$

$$= n\left(\frac{1}{n}\right)^{c(1+o(1))} = o(1),$$

поскольку $c > 1$.

Надежность сети: доказательство для случая $c > 1$...

Оставшаяся часть рассуждения состоит в доказательстве того, что слагаемые с $k > 1$ и $k < n - 1$ пренебрежимо малы по сравнению с первым слагаемым.
Соответствующую выкладку мы пропустим.
Если же поверить в ее справедливость, то получится, что вся сумма доминируется первым и последним слагаемыми, а стало быть, и она стремится к нулю.

Теорема для случая $c > 1$ доказана.

Надежность сети: доказательство для случая $c < 1$

2. Случай $c < 1$.
Обозначим через X_n количество изолированных вершин в случайном графе. Запишем

$$X_n = X_{n,1} + \ldots + X_{n,n},$$

где

$$X_{n,k} = X_{n,k}(G) = \begin{cases} 1, & \text{если вершина } k \in V_n \\ & \text{— изолированная в } G; \\ 0, & \text{иначе.} \end{cases}$$

Тогда

$$\mathrm{E}[X_n] = \mathrm{E}[X_{n,1}] + \ldots + \mathrm{E}[X_{n,n}]$$

В свою очередь

$$\mathrm{E}[X + n, k] = \mathrm{Pr}_{n,p}[k \text{ — изолированная в } G] = (1-p)^{n-1}.$$

Надежность сети: доказательство для случая $c < 1$...

Таким образом,

$$\mathrm{E}[X_n] = n(1-p)^{n-1} = n\,(1-p)^n(1+o(1)) = \\ = (1+o(1))\,ne^{-c\ln n} = (1+o(1))\,n^{1-c}.$$

Заметим, что ввиду неравенства $c < 1$ выполнено $\mathrm{E}[X_n] \to \infty$.

Посчитаем дисперсию случайной величины X_n:

$$D[X_n] = \mathrm{E}[X_n^2] - (\mathrm{E}[X_n])^2 = \mathrm{E}[X_{n,1} + \ldots + X_{n,n}]^2 - (\mathrm{E}[X_n])^2 = \\ = \mathrm{E}[X_{n,1}^2 + \ldots + \mathrm{E}[X_{n,n}^2] + \sum_{i \neq j} \mathrm{E}[X_{n,i}X_{n,j}] - (\mathrm{E}[X_n])^2 = \\ = \mathrm{E}[X_{n,1}] + \ldots + \mathrm{E}[X_{n,n}] + \sum_{i \neq j} \mathrm{E}[X_{n,i}X_{n,j}] - (\mathrm{E}[X_n])^2 = \\ = \mathrm{E}[X_n] + \sum_{i \neq j} \mathrm{E}[X_{n,i}X_{n,j}] - (\mathrm{E}[X_n])^2.$$

Надежность сети: доказательство для случая $c < 1$...

Далее,

$$\mathrm{E}[X_{n,i}X_{n,j}] = \mathsf{P}[i \text{ и } j \text{ изолированы в } G] = \\ = (1-p)^{2n-1} = (1+o(1))(1-p)^{2n},$$

$$\sum_{i \neq j} \mathrm{E}[X_{n,i}X_{n,j}] = n(n-1)(1+o(1))(1-p)^{2n} = \\ = (1+o(1))n^{2-2c} = (1+o(1))(\mathrm{E}[X_n])^2.$$

В итоге

$$\mathrm{D}[X_n] = \mathrm{E}[X_n] + (1+o(1))(\mathrm{E}[X_n])^2 - (\mathrm{E}[X_n])^2 = o\left((\mathrm{E}[X_n])^2\right).$$

Надежность сети: доказательство для случая $c < 1$...

По неравенству Чебышёва

$$\begin{aligned}\Pr_{n,p}[G \text{ связен}] \leqslant \Pr_{n,p}[X_n = 0] = \Pr_{n,p}[X_{\leqslant}0] = \\ = \Pr_{n,p}[-X_n \geqslant 0] = \\ = \Pr_{n,p}\left[\mathrm{E}[X_n] - X_n \geqslant \mathrm{E}[X_n]\right] \leqslant \frac{\mathrm{D}[X_n]}{(\mathrm{E}[X_n])^2} = o(1).\end{aligned}$$

и вторая часть теоремы доказана.

Надежность сети: обсуждение результатов

1. Сохранение связанности графа при $p \to 0$.

При $n \to \infty$, $p = \frac{c \ln n}{n} \to 0$ (довольно быстро), но при $c > 1$ *граф остаётся связным*.

Например, для 1000 городов мы можем позволить дорогам разрушаться с вероятностью 0.993 и при этом с вероятностью, близкой к единице, перемещение между любыми двумя городами всё ещё останется возможным.

2. Резкий скачок связаности.

Функция $p(n) = \frac{\ln n}{n}$ служит границей перехода от «почти всегда связности» к «почти всегда несвязности».

Такой переход принято называть *фазовым*, а соответствующую функцию $p(n)$ — *пороговой*

Конкретный результат

Теорема
Пусть $p = \frac{c \ln n}{n}$ *в модели* $G(n, p)$.
Тогда если $c > 3$, *то при* $n > 100$

$$\Pr_{n,p}[G \text{ связен}] \geqslant 1 - \frac{1}{n}.$$

Из этой теоремы следует, что, например, при $n = 1000$ городов и вероятности износа дороги $1 - \frac{3 \ln 1000}{1000} \approx 0,98$ вероятность сохранения связанности не менее $0{,}999$!

Компоненты связности графа

Теорема (Эрдёш и Реньи)
Пусть $p = \frac{c}{n}$ *в модели* $G(n, p)$. *Тогда если*

- $c < 1$, *то найдется такая константа* $\beta = \beta(c)$, *что почти всегда размер каждой связной компоненты случайного графа не превосходит* $\beta \ln n$;
- $c > 1$, *то найдется такая константа* $\gamma = \gamma(c)$, *что почти всегда в случайном графе есть ровно одна компонента размера* $\geqslant \gamma n$.

Снова фазовый переход — с пороговой функцией $p = \frac{1}{n}$: если вероятность ребра в $c > 1$ раз

- ниже порога, то все связные компоненты графа, скорее всего, крошечные (имеющие логарифмический размер от общего числа вершин n);
- выше порога, то, скорее всего, найдется компонента с числом вершин порядка n. Такая компонента называется *гигантской*.

Эволюция графа

Уточнения теоремы Эрдёша–Реньи:

- что при $c > 1$, помимо единственной гигантской компоненты, в случайном графе ничего сколь-нибудь крупного почти никогда не возникает: все остальные компоненты снова логарифмические;
- верны не только неравенство $\geqslant \gamma n$, но и асимптотика $\sim \gamma n$.

Изменение свойств случайного графа при изменении вероятности ребра p — *эволюция графа*.

Эволюция графа

Терминология А.М.Райгородского: *история мира*.

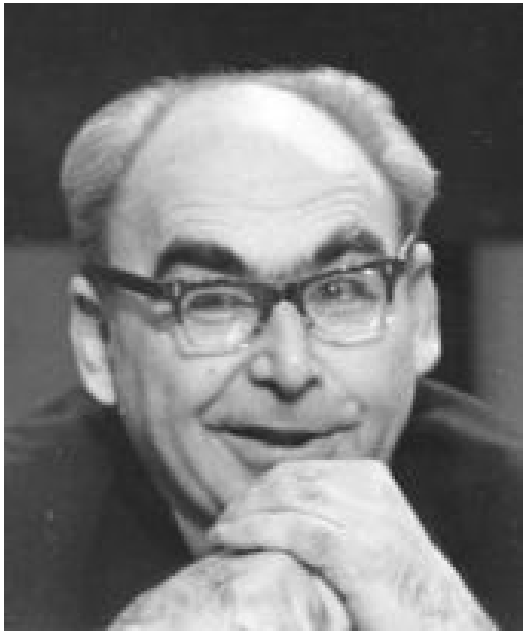
$p \ll \frac{1}{n}$ («феодализм») — весь граф поделен на несвязанные между собой логарифмические кусочки;

$p \gg \frac{1}{n}$ («империя») — гигантская компонента;

$p \gg \frac{\ln n}{n}$ («мировое господство») — «империя» уничтожает «окраины» и добивается всеобщей связности.

Смысл теоремы в терминах надежности: можно ещё в $\ln n$ раз уменьшить вероятность сохранности отдельной дороги, а возможность сообщения между любыми двумя городами останется в значительной части страны.

А. Реньи



Альфред Реньи
(венг. Renyi Alfred, 1921–1970)
— венгерский математик, основатель Математического института в Будапеште, теперь носящего его имя.
Основные труды по теории вероятностей, теории информации, комбинаторике и теории графов.

Написал 32 статьи совместно с П.Эрдёшем, в наиболее известных из которых вводится модель Эрдёша–Реньи случайных графов.

Автор популярных книг по математике.
Ему принадлежит известная фраза:
«*Математик — это автомат по переработке кофе в теоремы*».

Устройство «мира» «внутри фазовых переходов»

т.е. при:

$p \sim \frac{1}{n}$ — всё сложно;

$p \sim \frac{\ln n}{n}$ —

Теорема
Пусть $p = \frac{\ln n + c + o(1)}{n}$.
Тогда $\Pr_{n,p}[G \text{ связен}] \to e^{-e^{-c}}$.
В частности, при $p = \frac{\ln n}{n}$ *вероятность стремится к* e^{-1}.

Здесь уже речь не идёт о «почти всегда связности» или «почти всегда несвязности»: асимптотическая вероятность связности есть, но она лежит в строгих пределах от 0 до 1.

Обобщения модели Эрдёша-Реньи

① Пусть на вершинах $V_n = \{1, \ldots, n\}$ графа вероятность ребра между вершинами i и j есть p_{ij}, т.е. ребра появляются независимо друг от друга, но с разными вероятностями.

В формате аксиоматики Колмогорова получаем дискретное вероятностное пространство, элементы которого суть

$$\Omega_n = \{G = (V_n, E)\}, \quad \mathcal{F}_n = \mathbf{2}^{\Omega_n},$$

$$\Pr_{n,p_{ij}}[G] = \prod_{(i,j)\in E} p_{ij} \cdot \prod_{(i,j)\notin E} (1 - p_{ij}).$$

Обобщения модели Эрдёша-Реньи: важный частный случай

② Фиксируем некоторый граф $H_n = (V_n, E_n)$ и полагаем

$$p_{ij} = \begin{cases} p, & (i,j) \in E, \\ 0, & (i,j) \notin E \end{cases}$$

(ребра графа H_n возникают в случайном графе независимо друг от друга с одной и той же вероятностью $p = p(n) \in [0,1]$, а ребра, которых в нём нет, не возникают вовсе).

Этот вариант модели принято обозначать $G(H_n, p)$, в ней

$$\Pr_{n,n_{ij}}[G] = p^{|E|}(1-p)^{|E_n|-|E|}.$$

Модель $G(H_n, p)$ вполне адекватна вопросу о надежности транспортной сети: с самого начала можно зафиксировать граф дорог H_n и следить за износом его ребер.

Модель $G(H_n, p)$

— более адекватная реальности и более сложная для изучения, чем $G(n, p)$.

Главный результат относительно этой модели — нетривиальная

Теорема (Г.А. Маргулис)
Пусть $\{H_n\}$ *— последовательность графов, реберная связность которых стремится к бесконечности при* $n \to \infty$. *Тогда существует пороговая функция p для свойства связности случайного графа в модели* $G(H_n, p)$.

Поиск пороговой функции, существование которой доказывается в теореме — всякий раз сложная задача, повязанная на специфику графов из последовательности $\{H_n\}$.

Практический смысл теоремы: надо строить дороги так, чтобы связность получающегося графа неуклонно росла.

Г.А. Маргулис

Григорий Александрович Маргулис (1946) — российский и американский математик, д.ф.-м.н., научный сотрудник ИППИ РАН, с 1991 г. — профессор Йельского университета (США).

Лауреат премии Дж. Филдса (1978, на церемонии вручения не присутствовал, т.к. ему было отказано в выездной визе) и премии Вольфа (2005).

Член Национальной академии наук США, действительный членом Американского математического общества.

Законы 0 и 1: язык первого порядка для описания графов

— содержит символы:

- предметных переменных — x, y, ..., обозначают вершины графа;
- отношения $\sim$ смежности вершин;
- логических связок — $\neg$, $\vee$, $\&$, $\supset$, $\equiv$;
- кванторов — $\forall$, $\exists$;
- отношения $=$ равенства выражений;
- скобок $(,)$ — вспомогательные символы, обеспечивающие правильное чтение выражений.

Правила образования (конечных) выражений (формул языка) — обычные.

Пример формулы «граф содержит треугольник»:

$$\exists x\, \exists y\, \exists z : (x \sim y)\&(y \sim z)\&(z \sim x).$$

Законы 0 и 1: язык первого порядка для описания графов

Не все свойства графа выразимы на языке первого порядка, например «граф связен»:

$$\forall\, x\, \forall\, y\, \exists\, n \in \mathbb{N}_0\, \exists\, x_1, \ldots, x_n : (x \sim x_1) \& (x_1 \sim x_2) \& \ldots \& (x_n \sim y)$$

— утверждение $\exists\, n \in \mathbb{N}_0$ невыразимо на введённом языке первого порядка.

Законы 0 и 1: результаты

Свойство случайного графа подчиняется закону 0 и 1, если оно почти наверное либо выполнено, либо не выполнено.

Результаты для модели Эрдёша–Реньи.

Теорема
При $p = const$ свойство графа, которое возможно записать на языке первого порядка, подчиняется закону 0 и 1.

Теорема
При $p = n^{\alpha}$, $\alpha \in \mathbb{Q}$ свойство графа, которое возможно записать на языке первого порядка, подчиняется закону 0 и 1.

Теоремы доказываются с помощью игры Эренфойхта (Andrzej Ehrenfeucht, 1932) на графах.

Разделы

Каким законам подчиняется рост Интернета?

— вопрос естественно возник в 1990-е, когда Интернет только зарождался и возникла необходимость построить адекватную модель его «жизни».

А.-Л. Барабаши и Р. Альберт нашли ряд важных эмпирических закономерностей в поведении Интернета. Основная из них: *новые сайты чаще ссылаются на уже существующие популярные сайты*.

Эти идеи впоследствии формализовывали многие авторы.

Модели Барабаши-Альберт применяют для описания также социальных, биологических, транспортных и т.д. сетей.

А.-Л. Барабаши и Р. Альберт

Альберт-Ласло Бараба́ши
(Albert-László Barabási, 1967)
— физик, работает в Университетах США.

Член Американского физического общества, иностранный член венгерской Академии наук.

Рика́ Альбе́рт (Réka Albert, 1972)
— профессор физики и биологии в Пенсильванском университете (США).

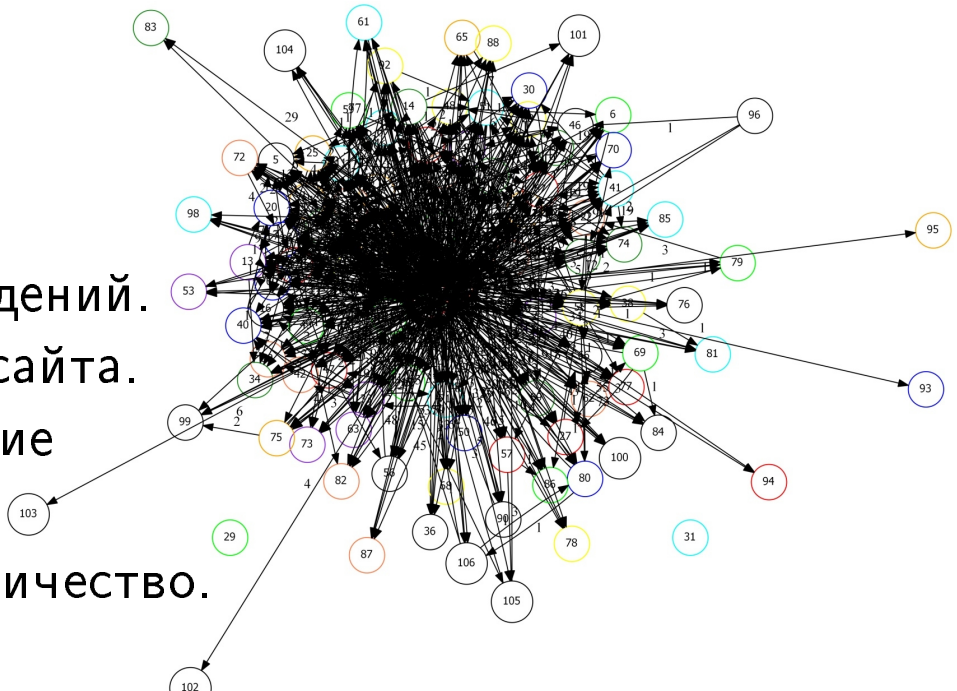
А.-Л. Барабаши и Р. Альберт — этнические венгры, родившиеся в Румынии.

На момент публикации совместной работы Emergence of scaling in random networks (Science, 1999), Р. Альберт — аспирантка А.-Л. Барабаши.

Графы Интернета

Вершины — структурные единицы в Интернете: сайты (Интернет-граф), хосты (хост-графы), статические-html страницы (HyperText Markup Language, веб-графы), владельцы и пр.;

Рёбра — соединяют вершины, между которыми имеются ссылки, т.е. имеются кратные и ребра, и петли.



Веб-граф связей сайтов СО РАН и учебных учреждений. Число в кружке — номер сайта. Стрелка обозначает наличие гиперссылок, цифра у стрелки — их количество.

Характеристики графа I

Расстояние в графе — число рёбер в кратчайшем рёберном пути.

Диаметр $\mathrm{diam}(G)$ ***графа*** — максимальное расстояние между вершинами (неопределён для несвязанного графа).
Если диаметр мал, то граф *тесный*.

Независимое множество — совокупность вершин графа никакие две из которых не соединены ребром (индуцированный этим множеством подграф состоит из изолированных вершин).
Максимальный размер независимого множества графа G обозначают $\alpha(G)$.

Характеристики графа II

Хроматическое число $\chi(G)$ — минимальное число цветов, в которые можно раскрасить вершины графа так, чтобы концы любого ребра имели разные цвета.

Обхват $girth(G)$ — размер кратчайшего цикла в G.

Плотный граф — граф, в котором число рёбер близко к максимальному.

Разреженный граф (противоположное свойство) — граф, имеющий малое число рёбер.

Характеристики графа III

Гигантская компонента Пусть дана последовательность графов $\{ G_n = (V_n E_n) \}$ и $\lim\limits_{n\to\infty} |V_n| = +\infty$. Говорят, что графы G_n содержат *гигантскую компоненту связности*, если
$\exists \gamma > 0 \, \forall n :$ |наибольшая компонента связности $G_n| \geqslant \gamma \, |V_n|$.

Степени вершин — $\mathrm{indeg} \;\; v$, $\mathrm{outdeg} \;\; v$, $deg \; v$ — входящая, исходящая и полная степень вершины v.

Вторые степени вершин — можно определить по-разному:

- число вершин на расстоянии 2 от v;
- число вершин на расстоянии $\leqslant 2$ от v;
- сумма степеней вершин на расстоянии 1 от v;

— значения данных величин могут сильно различаться.

Характеристики графа IV

Пейджранк — ...

Наблюдения Барабаши-Альберт

① Веб-граф весьма *разрежен*: у него в t-вершином подграфе примерно kt ребер, где $k \geqslant 1$ — константа, что отличается по порядку от числа $C_t^2 = \Theta(t^2)$ рёбер у полного t-вершинного графа.

② Диаметр веб-графа невелик: в 1999 г. он имел величину $5...7$ — граф *тесен*.
Это — известное свойство социальных сетей — «мир тесен» (small-word phenomenon):

- любые два человека в мире знакомы через $5...6$ рукопожатий;
- с любого сайта можно перейти на любой другой за $5...7$ переходов (если находимся в гигантской компоненте).

Если учитывать ориентацию рёбер, то диаметр $\approx 10...20$.

Эти параметры не меняются долгие годы.

Наблюдения Барабаши-Альберт...

③ Веб-граф характеризуется степенным законом распределения степеней вершин: вероятность того, что вершина веб-графа имеет степень d, оценивается как c/d^{λ}, где $\lambda = const$, а c — нормирующий множитель, вычисляемый из условия $\sum \Pr = 1$.

Реальные биологические, социальные, транспортные и т.д. сети подчиняются такому же степенному закону с разными $\lambda \in (2, 3)$:

$$\text{веб-графы} — \ \lambda \approx 2{,}1;$$
$$\text{хост-графы} — \ \lambda \approx 2{,}3;$$
$$\ldots$$

Наблюдения Барабаши-Альберт: следствия и вывод

Попытаемся примененить модель Эрдёша-Реньи для описания роста Интернета и подобных сетей.

1. Подбором вероятности p *можно добиться разреженности и тесноты* (при $p = \Theta(1/n)$, хотя и не с наблюдаемыми параметрами).
2. Степенной закон законом распределения степеней вершин *не может быть получен* в схеме Бернулли: в модели $G(n, p)$ степень каждой вершины случайного графа биномиальна с параметрами $n - 1$ и p, и при $p = \Theta(1/n)$ данное биномиальное распределение аппроксимируется пуассоновским, а не степенным.

Вывод: модель Эрдёша-Реньи *не применима* для описания роста Интернета и подобных сетей.

Модели предпочтительного присоединения

Предложения Л.-А. Барабаши и Р.Альберт по моделированию процесса формирования Интернета:

- считаем, что в каждый момент времени появляется новый сайт, который ставит фиксированное количество ссылок на своих предшественников;
- вероятность, с которой новый сайт поставит ссылку на один из прежних сайтов, пропорциональна числу уже имевшихся на тот сайт ссылок.

— модели *предпочтительного присоединения* (preferential attachment). "Имущему дастся, а у неимущего отнимется".

Барабаши и Альберт никак не конкретизировали, какую именно из таких моделей они предлагают рассматривать.

Адекватную формализацию модели Барабаши-Альберт предложили в начале 2000-х Б. Боллобаш и О. Риордан.

Б. Боллобаш и О. Риордан

Бела Боллоба́ш (Béla Bollobás, 1943) — венгерский математик, работающий в Великобритании.
В юности принял участие в двух международных математических Олимпиадах, завоевав две золотые медали.
Руководителем его Ph.D. был Пал Эрдёш.
Стажировался в Москве у И.М. Гельфанда.

Оливер Риордан (Oliver Maxim Riordan, ?) — профессор дискретной математики Оксфордского университета (Великобритания).
Учился в Кембриджском университете у Б. Боллобаша.

Модель Боллобаша-Риордана (динамическая модификация)

- Сначала строится последовательность случайных графов $\{G_1^n\}_{n\geqslant 1} = G_1^1, G_1^2, \ldots$, в которой у графа с номером n число вершин и ребер равно n,
- затем из неё формируется последовательность $\{G_k^n\}_{n\geqslant 1}$, в которой у графа с номером n число вершин равно n, а число ребер — kn, $k \in \mathbb{N}$.

① Начинаем с графа G_1^1 с одной вершиной и одной петлёй.

② Когда граф G_1^{n-1} с $n-1$ вершиной $n-1$ петлями построен, добавим к нему вершину n и ребро (n,i), $i \in \{1,\ldots,n\}$, при этом

- петля (n,n) возникнет с вероятностью $\frac{1}{2n-1}$;
- ребро (n,i) возникнет с вероятностью $\frac{\deg i}{2n-1}$, где $\deg i$ — степень вершины i в графе G_1^{n-1}.

Модель Боллобаша-Риордана (динамическая модификация)...

Распределение вероятностей задано корректно:

$$\sum_{i=1}^{n-1} \frac{\deg i}{2n-1} + \frac{1i}{2n-1} = \frac{2n-1}{2n-1} + \frac{1}{2n-1} = 1,$$

Случайный граф G_1^n построен, и он удовлетворяет принципу предпочтительного присоединения.

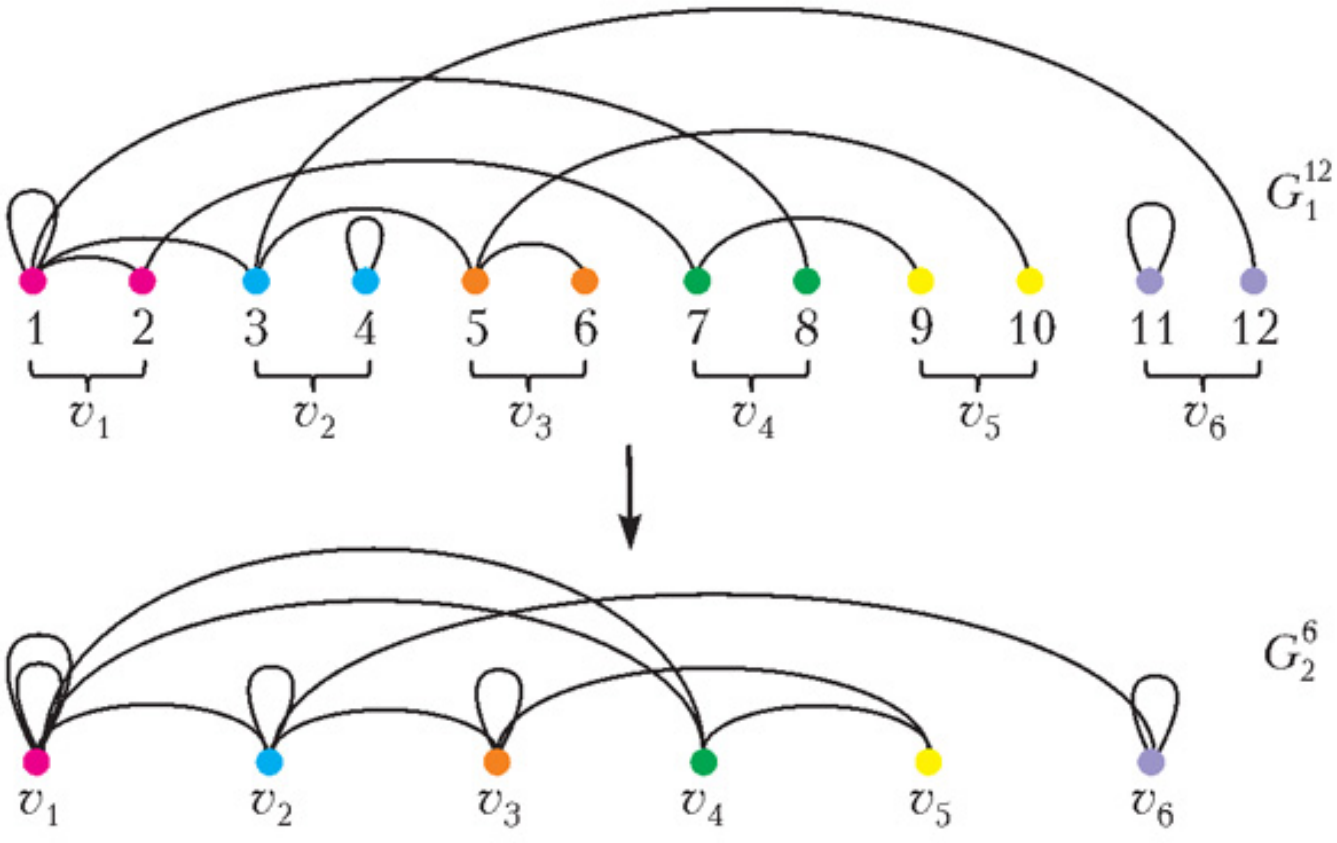
③ Строим G_k^n:

- Берем граф G_1^{kn} с kn вершинами и kn рёбрами.
- Делим множество его вершин на последовательные части размера k:
  $\{1, \ldots, k\}, \{k+1, \ldots, 2k\}, \ldots, \{k(n-1)+1, \ldots, kn\}.$
- Объявляем каждую часть вершиной, а ребра сохраняем: ребра внутри части образуют кратные петли, а между двумя различными частями — кратные ребра.
  Вершин стало n, а ребер — по-прежнему kn.

Цель достигнута.

Модель Боллобаша-Риордана (динамическая модификация)

Иллюстрация к последнему шагу



Модель Боллобаша-Риордана (статическая модификация)

— по своим вероятностным характеристикам практически неотличима от динамической.

Построим *линейную хордовую диаграмму* (linearized chord diagram, LCD):

1) зафиксируем на оси абсцисс на плоскости $2n$ точек $1, 2, 3, \ldots, 2n$;
2) разобьем эти точки на пары, а элементы каждой пары соединим дугой, лежащей в верхней полуплоскости.

Дуги в LCD могут пересекаться, лежать друг под дружкой, но не могут иметь общих вершин.

Количество различных LCD: $l_n = (2n-1)(2n-3)\ldots 1 = \dfrac{(2n)!}{2n(2n-2)(2n-4)\ldots 2} = \dfrac{(2n)!}{2^n n(n-1)(n-2)\ldots 1} = \dfrac{(2n)!}{2^n n!}$.

Если LCD случайная, то вероятность появления каждой — $\frac{1}{l_n}$.

Модель Боллобаша-Риордана (статическая модификация)...

По каждой LCD построим граф:

- идем слева направо по оси абсцисс, пока не встретим впервые *правый* конец какой-либо дуги; пусть его номер i_1;
- объявляем набор $\{1, \ldots, i_1\}$ первой вершиной будущего графа (склеиваем их);
- снова идем от $i_1 + 1$ направо до первого правого конца i_2 какой-либо дуги и обявляем второй вершиной графа набор $i_1 + 1, \ldots, i_2$; и т.д.

Получаем n вершин, а ребра порождаем дугами: вершины соединяем ребром, если между соответствующими наборами есть дуга; ребра ориентируем *справа налево*, петли возникают аналогично: граф с n вершинами и n ребрами построен.

Граф с n вершинами и kn ребрами получаем как в динамической модели.

Модель Боллобаша-Риордана: диаметр графа

Теорема (Боллобаш и Риордан)
Обозначим $\widehat{D}_n = \frac{\ln n}{\ln \ln n}$; *тогда для любого* $k \geqslant 2$ *и любого* $\varepsilon > 0$

$$\mathsf{P}\left[\left|\operatorname{diam} G_k^n - \widehat{D}_n\right| \leqslant \varepsilon\right] \to 1, \quad n \to \infty.$$

Вывод: диаметр графа G_k^n *плотно сконцентрирован* около $\widehat{D}_n$.

У веб-графа порядка 10^7–10^8 вершин; подставляя эти значения, получим

$$5{,}8 \leqslant \widehat{D}_n \leqslant 6{,}2$$

— *фантастическое попадание* (даже при $n = 10^9$: $\widehat{D}_n < 7$)!

Модель Боллобаша-Риордана: распределение степеней вершин

Теорема (Боллобаш, Риордан, Спесер, Тушнади (2001))
Пусть d_n *— доля вершин степени* n *в графе* G_k^n *и* $r_{d,k} = \frac{2k(k+1)}{(d+k+1)(d+k+2)(d+k+3)}$.
Тогда для любых $k \geqslant 1$ *и* $d \leqslant n^{1/15}$

$$\mathrm{E}\,[\,d_n\,] \sim r_{d,k},$$

Т.к. k — константа, $r_{d,k} \approx const/d^3$, т.е. имеем степенной закон с $\lambda = 3$.

Два недостатка теоремы:

1) ограничение $d \leqslant n^{1/15}$: теорема практически неприменима (даже при $n \approx 10^{12}$, имеем $d \leqslant 10^{4/5} \approx 6,31$) — этот недостаток устранён;
2) $\lambda = 3$ — необходимо уточнение модели.

Модель Боллобаша-Риордана: развитие

В рамках модели Боллобаша-Риордана получены оценки распределений (при разных её модификациях/упрощениях и ограничениях на параметры)

- вторых степеней вершин;
- количества рёбер между вершинами заданных степеней;
- кластерных коэффициентов;
- числа копий фиксированного графа: треугольников (в модели — $\sim \Theta(\ln^3 n)$, а реально — $\sim n^\alpha$), тетраэдров и т.д.; наличие клик объясняется действиями спамеров, которые искусственно расставляют ссылки, желая повысить рейтинги сайтов, заплативших за раскрутку;
- пейджранка.

Теоретические распределения в разной степени совпадают с эмпирически наблюдаемыми.

Спам в модели Боллобаша-Риордана не учтён, что также её минус.

Модель Бакли-Остгауза (2004)

— уточнение модели Барабаши-Альберт.

Строится случйный граф $H^n_{a,m}$, a — *начальная притягательность вершины*, параметр модели.

① Начинаем с графа $H^n_{a,1}$ с одной вершиной и одной петлёй.

② Когда граф $H^n_{a,1}$ с вершинами $\{1, \ldots, n\}$ и n петлями построен, добавим к нему вершину $n+1$ и ребро $(n+1, i)$, $i \in \{1, \ldots, n+1\}$, при этом

- петля $(n+1, n+1)$ возникнет с вероятностью $\frac{a}{(a+1)n+1}$;
- ребро $(n+1, i)$ возникнет с вероятностью $\frac{\deg i - 1 + a}{(a+1)n+1}$, где $\deg i$ — степень вершины i в графе $H^n_{a,1}$.

Распределение вероятностей задано корректно $(\sum \mathsf{P} = 1)$ и при $a = 1$ имеем модель Боллобаша-Риордана.

③ Граф $H^n_{a,m}$ строится, как и G^n_m.

Модель Мори (2005) —

— почти совпадает с моделью Бакли-Остгауза.

Строится случйный граф $H^{(n)}_{\beta,m}$, за притягательность вершины отвечает параметр β.

① В начальный момент времени имеем граф $H^{(2)}_{\beta,1}$ с двумя вершинами и ребром между ними.

② Когда граф $H^{(n)}_{\beta,1}$ с вершинами $\{1,\ldots,n\}$ и n петлями построен, добавим к нему вершину $n+1$ и ребро $(n+1,i)$, $i\in\{1,\ldots,n\}$ (*петель нет*), при этом ребро $(n+1,i)$ возникнет с вероятностью $\frac{\deg i+\beta}{(\beta+2)n-2}$, где $\deg i$ — степень вершины i в графе $H^{(n)}_{\beta,1}$.

Легко проверяется, что распределение вероятностей задано корректно и при $\beta=0$ имеем модель Боллобаша-Риордана.

③ Граф $H^{(n)}_{\beta,m}$ строится, как и G^n_m, при этом появляются петли.

Модели Бакли-Остгауза и Мори. Ещё модели...

В этих моделях оценки некоторых параметров улучшаются, но для кластерных коэффициентов и числа треугольников улучшений нет.

Предложены и другие модели Интернета:

- Боллобаша-Боогса-Риодана-Чайес (2003);
- модель копирования (2000);
- Купера-Фриза (2003);
- Холма-Кима (2002).

Разделы

Пейджранк $PR(i)$ — характеристика вершины i веб-графа (V, E), уточняющая понятие 1-й степени, 2-й степени и т.д.

$$PR(i) = d\sum_{j\to i}\frac{PR(j)}{\text{outdeg } j} + \frac{d}{|V|}\sum_{j\,\in\,\{v\in V\,|\,\text{outdeg } v=0\}} PR(j) + \frac{1-d}{|V|},$$

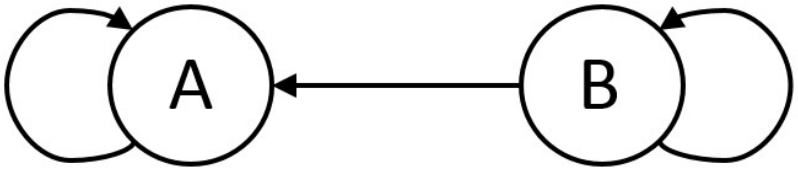
где $d \in (0,1)$ — *демпфирующий фактор* (константа).

Пейджранк (PageRank) страницы —

- учитывает не только количество ссылок на неё, но и их *качество*;
- = сумме качеств страниц, ссылающихся на данную, нормированная числом исходящих из них ссылок.

$(1-d)$ — *вероятность телепортации*: считаем, что пользователь при блуждании по сети с вероятностью d переходит по ссылке, а с вероятностью $1-d$ — на случайную страницу $\Rightarrow$ *возможен переход на любую страницу*.

Пейджранк: пример 1

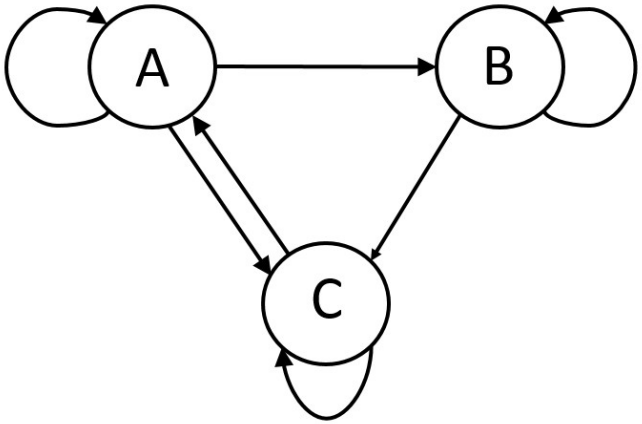


$$\begin{cases} PR(A) = d \cdot PR(A) + \dfrac{d}{2} \cdot PR(B) + \dfrac{1-d}{2}, \\ \\ PR(B) = \dfrac{d}{2} \cdot PR(B) + \dfrac{1-d}{2}. \end{cases}$$

$$\begin{cases} PR(A) = \dfrac{1}{2-d}, \\ \\ PR(B) = \dfrac{1-d}{2-d}. \end{cases} \qquad d \in (0,1) \Rightarrow PR(A) > PR(B).$$

Например, при $d = 0{,}85$: $PR(A) = 0{,}87$, $PR(B) = 0{,}13$.

Пейджранк: пример 2



$$\begin{aligned}\text{outdeg } A &= 3\\ \text{outdeg } B &= 2\\ \text{outdeg } C &= 2\end{aligned}$$

$$\begin{cases} PR(A) = d\left(\dfrac{PR(A)}{3} + \dfrac{PR(C)}{2}\right) + \dfrac{1-d}{3},\\ PR(B) = d\left(\dfrac{PR(A)}{3} + \dfrac{PR(B)}{2}\right) + \dfrac{1-d}{3},\\ PR(C) = d\left(\dfrac{PR(A)}{3} + \dfrac{PR(B)}{2} + \dfrac{PR(C)}{2}\right) + \dfrac{1-d}{3}. \end{cases}$$

При $d \in (0,1)$: $PR(C) > PR(A) > PR(B)$,
$d = 0{,}85 \Rightarrow PR(C) \approx 0{,}45,\ PR(A) \approx 0{,}29,\ PR(B) \approx 0{,}26.$

Пейджранк придумали Л. Пейдж и С. Брин

— основатели поисковой системы Google (1998).

Лоуренс «Ларри» Пейдж
(Lawrence «Larry» Page, 1973)
— получил степени бакалавра в Мичиганском и магистра — в Стэнфордском университете.
С 2011 г. — главный исполнительный директор Google.

Сергей Михайлович Брин
(англ. Sergey Brin, 1973)
— досрочно получил диплом бакалавра в Мэрилендском университете.
Родители Сергея Брина — выпускники Мехмата МГУ (1970 и 1971 годов).

Л. Пейдж и С. Брин входят в число самых богатых людей планеты (17-е и 20-е места в *Forbes-2014*).

Пейджранк: вычисление

Ещё одна конкретизация модели Боллобаши-Альберт:

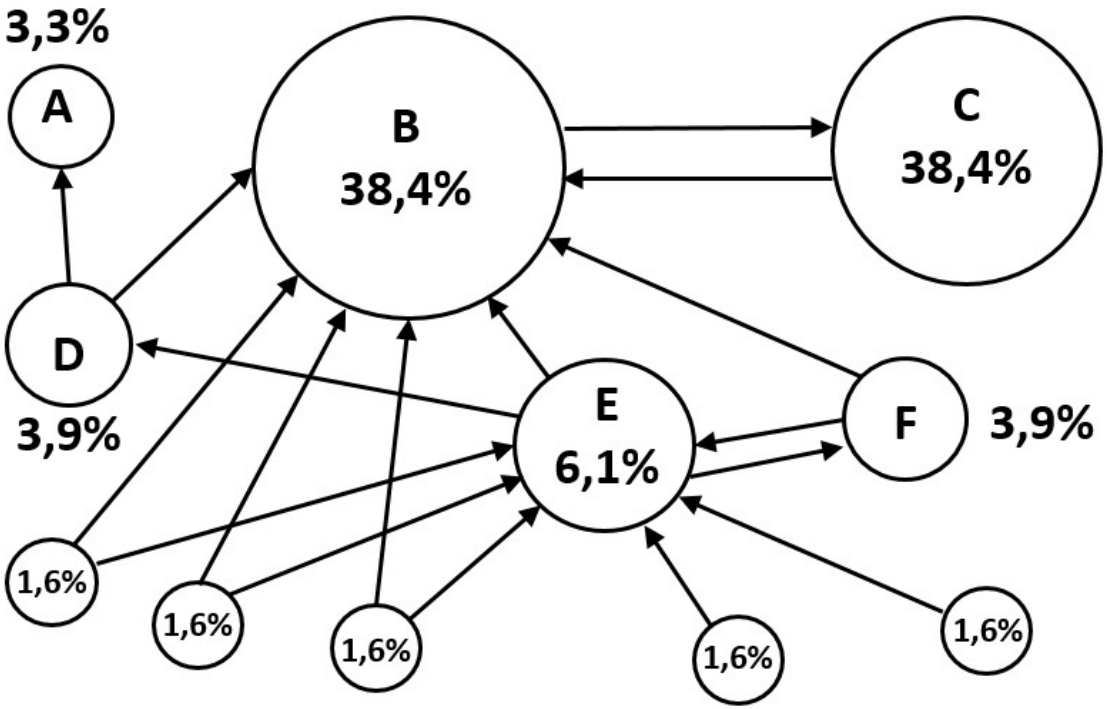
- начинаем с вершины 0 без петель, но полагаем её вес m;
- добавляем вершину 1 и из неё ставим m ссылок на 0, полагаем вес вершины 0 равным $2m$, а вес 1 — m;
- дальнейшие вершины, появляясь на свет, выставляют свои m ссылок независимо, каждую с вероятностью, пропорциональной весам существующих вершин, которые равны сумме их входящих степеней и m:

$$\mathsf{P}[n+1 \to i] = \frac{\text{indeg} \quad i+m}{\sum\limits_{k=0}^{n} (\text{indeg} \quad k+m)} = \frac{\text{indeg} \quad i+m}{2mn+m}.$$

Показано, что в данной модели плотность пейджранка $p_n(x) \approx \Theta\left(x^{-\frac{3+d}{1+d}}\right)$ — степенной закон.

При $d = 0,85$ $\lambda \approx 2,1$, что близко к эмпирическим данным.

Пример: PageRank для простой сети в процентах



$PR(C) > PR(E)$,
хотя
$\text{indeg } C < \text{indeg } E$,
но ссылка на C
исходит из очень
важной страницы B
$\Rightarrow$ она имеет
большой вес.

Без телепортации
все пользователи
в конечном итоге попадают на страницы A, B или C.

При наличии телепортации из A можно попасть на любую страницу в этой Сети, даже при $\text{indeg } A = 0$.

Модель поиска в интернет

- Страницы Интернета неэквивалентны и эта неэквивалентность описывается величиной $PR(v)$, $v \in V_n$.
- Вероятности перехода между страницами содержат две компоненты: обусловленную реальными связями между ними и возможностью попадания на любую страницу (телепортация).
- Если по запросу пользователя найдено множество страниц $W \subset V_n$, то релевантную информацию нужно искать на страницах, имеющих наибольшее значение $PR(v)$, $v \in W$.
- В настоящее время классические пейджранки уже не употребляются.

Надстройка для браузера Google Toolbar

- PageRank каждой веб-страницы — целое число от 0 до 10 (важность этой страницы с точки зрения Google).
- Механизм расчёта PageRank и что в точности обозначает это значение, не раскрывается.
- По некоторым данным, эти значения обновляются лишь несколько раз в год и показывают значения PageRank страниц на логарифмической шкале.
  Замечена особенность: Page Rank выше 5 могут получить сайты только довольно старые или очень большие проекты (сайты) с большим количеством посещений.
- Поисковая система Google использует более 200 ранжирующих сигналов, лишь одним из которых является PageRank, но он до сих пор играет существенную роль.