УДК 514.765

М.А. Львова, В.В. Славский

Число компонент связности случайного графа*

M.A. Lvova, V.V. Slavsky

The Number of Components of Connectivity of the Random Graph

Изучаются случайные толерантные бинарные отношения или случайные графы, еще более точнее – модель Эрдеша-Реньи случайных графов.

Ключевые слова: случайные отношения, случайные графы.

This article focuses on random tolerant binary relations or random graphs. Specifically, the study aimed to investigate the Erdös- Rényi random graph model.

Key words: random relations, random graphs.

В естественных науках (астрономия, биология, география, геология, физика, химия и др.), а также в общественных науках (археология, экономика, история, лингвистика, психология и др.) часто возникает задача классификации или кластеризации объектов по их характеристикам. Исходными понятиями в подобных задачах являются толерантное бинарное отношение между объектами и мера несходства объектов.

На практике матрица толерантного отношения θ на конечном множестве объектов X определяется эмпирически и, следовательно, неточно, то в связи с этим возникают интересные проблемы.

Рассмотрим для примера случай, когда множество X состоит из 4 объектов.

Теорема 1. Пусть толерантное отношение θ имеет вероятностное распределение Бернулли, т.е. $\theta(x_i, x_i) = 1$ и при $i \neq j$

$$\theta(x_i, x_j) = \begin{cases} 1 & \text{с вероятностью } p \\ 0 & \text{с вероятностью } q = 1 - p. \end{cases}$$

Случайные величины $\theta(x_i, x_j)$ независимы. Тогда вероятности P_k того, что транзитивно замыкание $\overline{\theta}$ имеет k классов эквивалентности, равны:

$P_1 = p^3\left(-6p^3 + 24p^2 - 33p + 16\right)$
$P_2 = (1-p)^3 p^2 (15 - 11p)$
$P_3 = 6(1-p)^5 p$
$P_4 = (1-p)^6$

Доказательство. Другими словами, рассмотрим случайный граф в смысле определений Пола Эрдоса и Альфреда Реньи [2].

1. Вероятность того, что $\overline{\theta}$ состоит из 1 класса эквивалентности.

Далее будем рассматривать отношения толерантности как графы, в нашем случае с четырьмя вершинами. Чтобы транзитивно замыкание имело 1 класс эквивалентности, необходимо и достаточно, чтобы граф отношения толерантности был связным. Связный граф с минимальным количеством ребер – это дерево. На t вершинах можно построить t^{t-2} различных деревьев. На четырех вершинах – 4^2, при этом такие деревья будут иметь 3 ребра. Итого получается вероятность появления такого отношения $4^2 \cdot p^3 \cdot (1-p)^3$.

Графы с 4-мя вершинами с 4-мя или более ребрами все будут связными. Вероятность появления соответствующих отношений толерантности

$$C_6^4 \cdot p^4 \cdot (1-p)^2 + C_6^5 \cdot p^5 \cdot (1-p) + C_6^6 \cdot p^6.$$

Соответственно, вероятность того, что $\overline{\theta}$ состоит из 1 класса эквивалентности

$$P_1 = 4^2 \cdot p^3 \cdot (1-p)^3 + C_6^4 \cdot p^4 \cdot (1-p)^2 + \\ + C_6^5 \cdot p^5 \cdot (1-p) + C_6^6 \cdot p^6.$$

2. Вероятность того, что $\overline{\theta}$ распадется на два класса.

Графы с 3-мя ребрами и 2-мя связными компонентами, их число $C_6^3 - 4^2$.

Графы с 2-мя ребрами и 2-мя связными компонентами. $\frac{1}{2}C_4^2$ – число графов, где одна компонента состоит из 2-х вершин, вторая – тоже из двух, $C_4^1 \cdot 3^1$ – число графов, где одна компонента состоит из одной вершины, вторая – из трех.

Соответственно, вероятность того, что $\overline{\theta}$ распадется на два класса, равна

$$P_2 = \left(C_6^3 - 4^2\right) \cdot p^3 \cdot (1-p)^3 + \\ + \frac{1}{2} C_4^2 \cdot p^2 \cdot (1-p)^4 + 3C_4^1 p^2 (1-p)^4.$$

*Работа выполнена при частичной финансовой поддержке РФФИ (проект № 08-01-98001), а так же ФЦПК «Научные и научно-педагогические кадры инновационной России» на 2009 – 2013 годы, тема «Фундаментальные проблемы анализа и геометрии» (номер заявки в информационной компьютеризированной системе «2012-1.1-12-000-1003-014».

3. Вероятность того, что распадется на три класса.

$$P_3 = C_6^1 p(1-p)^5.$$

4. Вероятность того, что $\overline{\theta}$ распадется на четыре класса. Это возможно только, если все недиагональные элементы нули. оответственно, эта вероятность равна

$$P_4 = (1-p)^{\frac{(4-1)4}{2}} = (1-p)^6.$$

Теорема доказана.

Теорема 2. Пусть толерантное отношение θ на множестве X из n элементов имеет вероятностное распределение Бернулли, т.е. $\theta(x_i, x_i) = 1$ и при $i \neq j$

$$\theta(x_i, x_j) = \begin{cases} 1 & \text{с вероятностью } p \\ 0 & \text{с вероятностью } q = 1-p. \end{cases}$$

Случайные величины $\theta(x_i, x_j)$ независимы. Тогда вероятности $P(k,n)$ того, что транзитивно замыкание $\overline{\theta}$ имеет k классов эквивалентности, находятся с помощью рекуррентных равенств:

- $P(1,1) = 1$;
- при $2 \leq k \leq n$

$$P(k,n) = \sum_{\substack{1 \leq r_1 \leq \cdots \leq r_k \\ r_1 + \cdots + r_k = n}} \frac{\binom{n}{r_1,\ldots,r_k}}{\delta(r_1,\ldots,r_k)} P(1,r_1) \cdot \ldots \cdot P(1,r_k) \cdot (1-p)^{\frac{n^2 - r_1^2 - \cdots - r_k^2}{2}}. \quad (1)$$

где число $\delta(r_1,\ldots,r_k) = s_1! \cdots s_t!$, здесь $s_1,\ldots,s_t$ – длины отрезков в неубывающей последовательности $r_1 \leq \cdots \leq r_k$, состоящие из равных между собой чисел (пример: $\delta(2,2,3,4,4,5) = 2!1!2!1!$), $\binom{n}{r_1,\ldots,r_k} = \frac{n!}{r_1!\cdots r_k!}$ – биномиальные коэффициенты.

- при $k = 1$;

$$P(1,n) = 1 - P(2,n) - P(3,n) - \cdots - P(n,n).$$

Доказательство. $P(1,1) = 1$ означает, что на одной вершине можно построить только один граф, который состоит из одной компоненты связности.

Граф с n вершинами содержит k компонент связности тогда и только тогда, когда у него существует k связных подграфов таких, что не существует пути из вершины одного подграфа в вершину другого, причем общее число вершин таких подграфов равно числу вершин графа. Пусть $G = (V,E), |V| = n$ – граф с n вершинами и k компонентами связности. Занумеруем эти компоненты (которые также являются подграфами) по возрастанию числа вершин: $G_1 = (V_1, E_1), \ldots, G_k = (V_k, E_k)$. Обозначим количество вершин i-го подграфа, как $r_i = |V_i|$. Учитывая способ нумирации, имеем $r_1 \leq \cdots \leq r_k$ и $r_1 + \cdots + r_k = n$.

Таким образом, чтобы посчитать вероятность того, что граф с n вершинами состоит из k компонент связности, нужно учесть всевозможные разбиения множества из n элементов на k подмножеств $V_1, \ldots, V_k$. Мощности этих множеств должны удовлетворять условиям $r_1 \leq \cdots \leq r_k$ и $r_1 + \cdots + r_k = n$. Для каждого набора чисел $r_1, \ldots, r_k$ количество способов выбора соответствующих подмножеств вершин равно

$$C_n^{r_1} \cdot C_{n-r_1}^{r_2} \cdot \ldots \cdot C_{n-r_1-\ldots-r_{k-1}}^{r_k} = \frac{n!}{r_1! \cdot \ldots \cdot r_k!} = \binom{n}{r_1,\ldots,r_k}.$$

Заметим, что если в неубывающей последовательности $r_1 \leq \cdots \leq r_k$ встречаются s одинаковых чисел, то при подсчете количества способов выбора подмножеств вершин, у нас $s!$ раз рассматривается один и тот же способ при каждом наборе оставшихся подграфов, т.е число раз равное количеству перестановок порядка s. Если исключить повторение такого рода, то количество способов разбить граф на подграфы равно

$$\frac{\binom{n}{r_1,\ldots,r_k}}{\delta(r_1,\ldots,r_k)},$$

где число $\delta(r_1,\ldots,r_k) = s_1! \cdots s_t!$, здесь $s_1, \ldots, s_t$ есть длины отрезков в неубывающей последовательности $r_1 \leq \cdots \leq r_k$, состоящие из равных между собой чисел.

Также необходимо учесть,что не существует пути из вершины одного подграфа в вершину другого. Всего $\frac{n(n-1)}{2}$ пар вершин. Для каждого подграфа $\frac{r_i(r_i-1)}{2}$. Итого, остается $\frac{n(n-1)}{2} - \sum_{i=1}^k \frac{r_i(r_i-1)}{2}$ пар вершин, которые не должны быть связаны ребрами. Преобразуем это выражение

$$\frac{n(n-1)}{2} - \sum_{i=1}^k \frac{r_i(r_i-1)}{2} = \frac{n^2-n}{2} - \sum_{i=1}^k \frac{r_i^2 - r_i}{2} = \frac{n^2 - \sum_{i=1}^k r_i^2 - (n - \sum_{i=1}^k r_i)}{2} = \frac{n^2 - \sum_{i=1}^k r_i^2}{2}.$$

Из вышеприведенных рассуждений следует, что при $2 \leq k \leq n$

$$P(k,n) = \sum_{\substack{1 \leq r_1 \leq \cdots \leq r_k \\ r_1 + \cdots + r_k = n}} \frac{\binom{n}{r_1,\ldots,r_k}}{\delta(r_1,\ldots,r_k)} P(1,r_1) \cdot \ldots \cdot P(1,r_k) \cdot (1-p)^{\frac{n^2 - r_1^2 - \cdots - r_k^2}{2}}.$$

Для графа с n вершинами справедливо

$$1 = P(1,n) + P(2,n) + P(3,n) + \ldots + P(n,n).$$

Из чего следует третье выражение из условия теоремы

$$P(1,n) = 1 - P(2,n) - P(3,n) - \dots - P(n,n).$$

Теорема доказана.

Пример. Выпишем рекуррентные формулы для $n = 5$.

$$P(2,5) = \binom{5}{1,4} P(1,1)P(1,4)(1-p)^4 + \\ + \binom{5}{2,3} P(1,2)P(1,3)(1-p)^6;$$
$$P(3,5) = \frac{\binom{5}{1,1,3}}{2!} P(1,1)P(1,1)P(1,3)(1-p)^7 + \\ + \frac{\binom{5}{1,2,2}}{2!} P(1,1)P(1,2)P(1,2)(1-p)^8;$$
$$P(4,5) = \frac{\binom{5}{1,1,1,2}}{3!} P(1,1)^4 P(1,2)(1-p)^9;$$
$$P(5,5) = \frac{\binom{5}{1,1,1,1,1}}{5!} P(1,1)^5(1-p)^{10};$$
$$P(1,5) = 1 - P(2,5) - P(3,5) - P(4,5) - P(5,5).$$

С увеличением n полиномы $P(k,n)$ быстро усложняются. Приведем результат окончательных вычислений для $n = 8$.

$$P(1,8) = p^7 \left(-5040p^{21} + 120960p^{20} - 1386000p^{19} + \\ +10086720p^{18} - 52319190p^{17} + 205732800p^{16} - \\ -636845160p^{15} + 1590501640p^{14} - 3258291120p^{13} + \\ +5536123600p^{12} - 7856193296p^{11} + 9345271992p^{10} - \\ -9324568001p^9 + 7786027816p^8 - 5410382880p^7 + \\ +3098951072p^6 - 1441519296p^5 + 532354536p^4 - \\ -150657080p^3 + 30802240p^2 - 4068456p + 262144\right);$$

$$P(2,8) = (p-1)^7 p^6 (13068p^{15} - 225036p^{14} + \\ +1813980p^{13} - 9083564p^{12} + 31615787p^{11} - \\ -81060259p^{10} + 158254628p^9 - 239732560p^8 + \\ +284345642p^7 - 264338326p^6 + 191274720p^5 - \\ -105964740p^4 + 43595055p^3 - 12608323p^2 + \\ +2300624p - 200704);$$
$$P(3,8) = -14(p-1)^{13} p^5 (938p^{10} - 10822p^9 + \\ +56393p^8 - 174947p^7 + 358193p^6 - \\ -506371p^5 + 501339p^4 - 343951p^3 + \\ +156929p^2 - 43171p + 5472);$$

$$P(4,8) = 7(p-1)^{18} p^4 \left(967p^6 - 6730p^5 + 19605p^4 - \\ -30680p^3 + 27285p^2 - 13134p + 2695\right);$$

$$P(5,8) = -70(p-1)^{22} p^3 \left(28p^3 - 98p^2 + 115p - 46\right);$$
$$P(6,8) = 14(p-1)^{25} p^2 (23p - 27);$$
$$P(7,8) = -28(p-1)^{27} p;$$
$$P(8,8) = (1-p)^{28}.$$

Замечание. Известны, также, другие рекуррентные формулы для $P(1,n)$ [1, 2]:

$$1 - P(1,n) = \sum_{k=1}^{n-1} P(1,k) C_{n-1}^{k-1} (1-p)^{k(n-k)}.$$

В данной работе в отличии от [1], во-первых получены новые рекуррентные формулы, во-вторых определены величины:

$$Q(r_1,\dots,r_k) = \frac{\binom{n}{r_1,\dots,r_k}}{\delta(r_1,\dots,r_k)} \times \\ \times P(1,r_1)\cdots P(1,r_k)(1-p)^{\frac{n^2 - r_1^2 - \dots - r_k^2}{2}},$$

где $1 \le r_1 \le \dots \le r_k$, $n = r_1 + \dots + r_k$. Данные величины представляют независимый интерес и равны вероятностям того, что случайный граф с n вершинами в модели Эрдеша-Реньи имеет в точности k компонент связности объемов соответственно $r_1 \le \dots \le r_k$.

Библиографический список

1. Gilbert, E.N. Random graphs, Annls Math. Statist. 30, 1141-1144, 1959.

2. Bela Bollobas. Random Graphs. Cambridge studies in advanced mathematics, 73. Cambridge University Press 2001.