УДК 519.214, 519.179.4

ЭКСТРЕМУМЫ АКТИВНОСТИ В НЕКОТОРЫХ МОДЕЛЯХ ИНФОРМАЦИОННЫХ СЕТЕЙ

А.В. Лебедев

Московский государственный университет им. М.В. Ломоносова

Россия, 119991, Москва, Ленинские горы, 1, ГЗ МГУ, 16-02

E-mail: avlebed@yandex.ru

Ключевые слова: информационные сети, случайные графы, гиперграфы, максимумы, экстремумы, степенные законы, тяжелые хвосты, правильное изменение, распределение Фреше.

Аннотация: В работе дан обзор результатов автора по моделям информационных сетей, описываемых случайными графами и гиперграфами, где каждый узел обладает случайной информационной активностью, распределение которой имеет тяжелый (правильно меняющийся) хвост. Выведены достаточные условия, при которых максимум суммарных активностей (по узлу и его соседям либо по сообществам) растет асимптотически так же, как и максимум индивидуальных активностей, и для них имеет место предельный закон Фреше.

1. Введение

В работах автора [1–3] изучалось поведение максимумов суммарной активности в информационных сетях, описываемых случайными графами, с точки зрения асимптотической эквивалентности максимумам индивидуальных активностей.

Проверка наличия подобного эффекта в реальных сетях, разумеется, требует экспериментального исследования, выходящего за рамки работ автора, которые имеют теоретический характер.

Пусть каждый узел сети обладает случайной информационной активностью (интенсивностью производства информации). Предположим, что активности узлов независимы и одинаково распределены, причем их распределение F имеет тяжелый (правильно меняющийся) хвост, т.е. $\bar{F}(x) \sim x^{-a}L(x)$, $x \to \infty$, $a > 0$, где $L(x)$ — медленно меняющаяся функция [4, §8.8]. Такое предположение находится в русле современных представлений о распространенности степенных законов в природе, технике и человеческой деятельности.

Рассмотрим также суммарную активность в узле (т.е. сумму его собственной и тех, от кого он получает информацию). Например, в Живом Журнале (livejournal.com) [5] каждый пользователь может оставлять свои записи и читать записи друзей, объединенные для удобства в общую «ленту друзей» (френдленту). Для простоты предполагаем, что индивидуальная активность и число соседей независимы. Нас интересует вопрос, когда максимум суммарных активностей растет асимптотически так же, как и максимум индивидуальных активностей узлов. Это означает, что большие значения обычно достигаются не за счет активности большого числа узлов, а за счет отдельных узлов с большой активностью. В этом случае для максимумов легко выводится предельный закон Фреше $\Phi_a(x) = \exp\{-x^{-a}\}$, $x > 0$ [4, §8.8].

При изучении случайных графов используются различные модели: классические, исследование которых восходит к работе [6], и степенные (power law, scale-free), активное исследование которых в последние десятилетия было инициировано работой [7] (и которые в отечественной литературе также называют графами Интернет-типа или Интернет-графами).

В степенных графах предельное распределение степени вершины имеет вид $p_k \sim ck^{-\beta}$, $\beta > 0$. Оказалось, что подобные модели хорошо описывают многие информационные, технические и биологические системы. Однако следует отметить, что к одному и тому же предельному распределению могут приводить самые разные алгоритмы построения, в то время как другие асимптотические свойства графов могут оказаться зависимы от выбора модели.

Будем рассматривать сети из n узлов, затем устремляя n к бесконечности. Обозначим через $M(n)$ максимум суммарных активностей, а через $M_0(n)$ — максимум индивидуальных активностей узлов. Нашей задачей будет определить условия, при которых верно

$$M(n)/M_0(n) \xrightarrow{P} 1, \quad n \to \infty. \tag{1}$$

Введем положительную функцию $v(r)$, такую, что $r\bar{F}(v(r)) \to 1$, $r \to \infty$. Заметим, что $v(r)$ заведомо существует и правильно меняется с показателем $1/a$, т.е. $v(r) \sim r^{1/a}L_2(r)$, $r \to \infty$, где $L_2(r)$ — медленно меняющаяся функция [8, §1.5].

Тогда имеет место предельный закон для максимумов независимых случайных величин в случае правильно меняющихся хвостов [4, §8.8]:

$$\lim_{n\to\infty} \mathbf{P}(M_0(n)/v(n) \leqslant x) = \Phi_a(x), \quad x > 0,$$

что в сочетании с (1) дает

$$\lim_{n\to\infty} \mathbf{P}(M(n)/v(n) \leqslant x) = \Phi_a(x), \quad x > 0. \tag{2}$$

В этом и заключается польза соотношения (1).

Как уже было отмечено, описанию информационных сетей с помощью случайных графов посвящена обширная литература, однако задача изучения максимумов суммарных активностей является новой.

Дальнейшие теоремы посвящены установлению верхних границ для хвостового индекса a в различных моделях информационных сетей, в зависимости от их параметров. Все модели, кроме $P_{\alpha,\beta}$, являются авторскими, созданными по мотивам [9,10].

2. Основные результаты

Рассмотрим сначала модель, описываемую степенным случайным графом $P_{\alpha,\beta}$, предложенным в [11]. Сначала предполагается, что число вершин степени k составляет $\lfloor e^\alpha/k^\beta \rfloor$, $\alpha, \beta > 0$, $k \geqslant 1$. Далее вводится равномерное распределение на множестве графов, удовлетворяющих этому условию. При этом допускаются петли и кратные ребра. При $\beta > 1$ число вершин n связано с параметрами соотношением $n \sim \zeta(\beta)e^{\alpha}$, $\alpha \to \infty$, где $\zeta(r) = 1/\sum_{k=1}^{\infty} k^{-r}$ — дзета-функция Римана. Асимптотические свойства графа не меняются, если параметризовать его числом вершин при $n \to \infty$.

Теорема 1. *Если сеть описывается случайным графом $P_{\alpha,\beta}$ с $\beta > 3/2$, то* (1) *выполняется при $a < \beta - 3/2$, если $3/2 < \beta < 3$, и при $a < \beta/2$, если $\beta \geqslant 3$.*

Рассмотрим теперь модель ориентированного случайного графа, где направления ребер соответствуют направлениям передачи информации. Пусть имеется n вершин и заданы независимые неотрицательные целочисленные случайные величины $K_1, \ldots, K_n$, имеющие одинаковое распределение, заданное вероятностями $p_k \sim ck^{-\beta}$, $k \to \infty$, $\beta > 1$. Положим $D_i = \min\{K_i, n-1\}$. Для i-й вершины выберем случайным образом (равновероятно и независимо от выбора для других вершин) D_i различных вершин из числа остальных (кроме i-й) и выпустим из них ребра, направленные в i-ю вершину. Полученный в результате граф можно отнести к степенным в том смысле, что входящие степени вершин распределены асимптотически по степенному закону. Суммарной активностью в узле в данном случае будем считать сумму собственной активности узла и всех узлов, из которых в него поступает информация (его входящих соседей).

Теорема 2. *Соотношение* (1) *выполняется при $a < \beta - 2$, если $2 < \beta < 3$, и при $a < (\beta-1)/2$, если $\beta \geqslant 3$.*

Рассмотрим далее модели информационных сетей со случайными весами.

Модель 1 представляет собой неориентированный случайный граф, являющийся обобщением классической модели Эрдеша-Реньи. Предполагается, что узлы имеют независимые и одинаково распределенные веса, и связи между узлами устанавливаются с вероятностями, асимптотически пропорциональными произведению соответствующих весов. Похожие модели рассматривались в [10, гл. 6] как обобщенные случайные графы со случайными весами (см. там же обзор литературы). Применительно к социальным сетям веса могут иметь смысл меры общительности пользователей. При этом выбором распределения весов можно широко варьировать предельное распределение степеней вершин, от классического пуассоновского до степенного закона.

Модель 2 представляет собой ориентированный случайный граф. Предполагается, что узлы имеют независимо и одинаково распределенные веса, определяющие вероятности присоединения данного узла к другим для получения информации. Такие веса могут иметь смысл меры любознательности пользователей. При этом обратная связь (присоединение в ответ) не учитывается.

В этом случае может иметь место односторонняя передача информации. Практически, например, в ЖЖ наблюдается промежуточная ситуация между моделями 1 и 2: часть информации доступна всем, а часть только друзьям (входящим соседям).

Модель 3 представляет собой случайный гиперграф. Предполагается, что в сети есть узлы и сообщества. Сообщества имеют независимо и одинаково распределенные веса, определяющие вероятности присоединения к ним каждого узла, независимо от других соединений. Такие веса могут иметь смысл меры популярности сообщества. Здесь мы, в отличие от предыдущих моделей, изучаем максимумы суммарных активностей по сообществам и индивидуальных активностей по узлам. Таким образом, в качестве ребер гиперграфа берутся множества вершин, вошедших в сообщества, и еще множества, содержащие по одной вершине.

В данном случае обозначим через $M(n)$ максимум суммарных активностей (в моделях 1 и 2) или максимум по сообществам и узлам (в модели 3). Нашей задачей будет определить условия, при которых верно (1).

В модели 1 предполагаем, что заданы случайные величины w_i, $1 \leqslant i \leqslant n$, независимые и одинаково распределенные как $W \geqslant 0$ (и не зависящие от индивидуальных активностей). Пусть существует $\beta \geqslant 1$ такое, что $\mathbf{E}W^\beta < \infty$ (заметим, что здесь смысл параметра β иной, чем в предыдущих разделах).

Обозначим $p_i = \varphi(w_i n^{-s/2})$, где $0 < s \leqslant 1$, и для функции φ на $\mathbf{R}_+$ верно $0 \leqslant \varphi(x) \leqslant \min\{1, x\}$, $\varphi(x) \sim x$, $x \to 0$. В качестве φ можно взять, например, $\min\{1, x\}$, $x/(1+x)$ или $1 - e^{-x}$.

Пусть при известных w_i, $1 \leqslant i \leqslant n$, каждая пара вершин i и j соединяется ребром с вероятностью $p_i p_j$ независимо от других пар.

В частности, при $W \equiv c > 0$ получаем классическую модель случайного графа Эрдеша-Реньи $G(n,p)$ с $p \sim c^2 n^{-s}$, $n \to \infty$.

Теорема 3. *В модели 1 верно* (1) *при* $\beta \geqslant 2$*, если*

$$a < \frac{2s-1}{2(1-(s-1/\beta)_+)}, \quad s > 1/2,$$

и при $1 \leqslant \beta < 2$*, если*

$$a < \frac{(1+\beta/2)s-1}{2(1-(s-1/\beta)_+)}, \quad s > \frac{1}{1+\beta/2}.$$

В модели 2 мы делаем те же предположения о весах w_i, $1 \leqslant i \leqslant n$, но теперь $p_i = \varphi(w_i n^{-s})$, $0 < s \leqslant 1$. Пусть при известных w_i, $1 \leqslant i \leqslant n$, в i-ую вершину входит ребро из любой другой вершины с вероятностью p_i независимо от других ребер.

Теорема 4. *В модели 2 верно* (1) *при* $\beta \geqslant 2$*, если*

$$a < \frac{2s-1}{2(1-(s-1/\beta)_+)}, \quad s > 1/2,$$

и при $1 \leqslant \beta < 2$*, если*

$$a < \frac{\beta s-1}{2(1-(s-1/\beta)_+)}, \quad s > 1/\beta,$$

В модели 3 при тех же предположениях о w_i и p_i, $1 \leqslant i \leqslant n$, что и в модели 2 (только эти величины уже относятся не к узлам, а к сообществам), полагаем, что существуют m сообществ, и $m = O(n^r)$, $n \to \infty$, $0 \leqslant r \leqslant 1$. Каждая вершина входит в i-ое сообщество (гиперребро) с вероятностью p_i независимо от членства в других сообществах и поведения других вершин. Кроме того, введем дополнительные гиперребра $\{1\}$, $\{2\}$, …, $\{n\}$ для описания отдельных вершин.

Теорема 5. *В модели 3 верно* (1) *при* $\beta \geqslant 2$*, если*

$$a < \frac{2s-r}{2(1-(s-r/\beta)_+)}, \quad s > r/2,$$

и при $1 \leqslant \beta < 2$*, если*

$$a < \frac{\beta s-r}{2(1-(s-r/\beta))}, \quad s > r/\beta.$$

Заметим, что во всех моделях при $s = 1$ и наличии у W моментов любого порядка ограничения на параметр a вообще снимаются.

Обозначим через $\mathrm{MPois}(\xi)$ смешанное пуассоновское распределение, которое возникает при рандомизации обычного пуассоновского распределения по его случайному среднему $\xi \geqslant 0$. Тогда для $X \sim \mathrm{MPois}(\xi)$ получаем [10, §6.2, определение 6.8]:

$$\mathbf{P}(X = k) = \mathbf{E}\left(\frac{\xi^k}{k!}e^{-\xi}\right), \quad k \geqslant 0.$$

В частности, если ξ имеет степенной хвост, то X также имеет степенной хвост, с тем же показателем степени [10, §6.2, упражнение 6.12]. Если $\xi \equiv \lambda > 0$, то получаем просто пуассоновское распределение с параметром λ. Возможны и промежуточные случаи. Например, если ξ имеет гамма-распределение, то X имеет отрицательное биномиальное распределение.

Важно понять, какие предельные распределения степеней вершин или размеров сообществ возникают в описанных моделях. Пусть D — степень отдельно взятой вершины (в модели 1), входящая степень (в модели 2) или размер сообщества (в модели 3).

Теорема 6. *В модели 1 при* $0 < s < 1$ *верно* $D/n^{1-s} \xrightarrow{d} W\mathbf{E}W$, *а при* $s = 1$ *верно* $D \xrightarrow{d} \mathrm{MPois}(W\mathbf{E}W)$. *В моделях 2 и 3 при* $0 < s < 1$ *верно* $D/n^{1-s} \xrightarrow{d} W$, *а при* $s = 1$ *верно* $D \xrightarrow{d} \mathrm{MPois}(W)$.

Поскольку в теоремах 1–5 мы пользуемся лишь достаточными условиями, не исключено, что эти ограничения в будущем могут быть ослаблены.

Как показано в [12], в ряде случаев полученные результаты можно интерпретировать в терминах экстремального индекса в схеме серий для системы зависимых суммарных активностей.

Список литературы

1. Лебедев А.В. Максимумы активности в случайных сетях в случае тяжелых хвостов // Проблемы передачи информации. 2008. Т. 44, № 2. С. 96-100.
2. Лебедев А.В. Максимумы активности в безмасштабных случайных сетях с тяжелыми хвостами / Информатика и ее применения. 2011. Т. 5, № 4. С. 13-16.
3. Лебедев А.В. Максимумы активности в некоторых моделях информационных сетей со случайными весами и тяжелыми хвостами // Проблемы передачи информации. 2015. Т. 51, № 1. С. 72-81.
4. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 2. М.: Мир, 1984. 752 с.
5. Захаров П. Народ-блогоносец // Компьютерра. 2007. № 27-28. С. 36-39.
6. Erdös P., Rényi A. On random graphs // Publ. Math. Debrecen. 1959. Vol. 6. P. 290-297.
7. Barabási A., Albert R. Emergence of scaling in random networks // Science. 1999. Vol. 286. P. 509-512.
8. Сенета Е. Правильно меняющиеся функции. М.: Наука, 1985. 144 с.
9. Райгородский А.М. Модели случайных графов и их применения // Труды МФТИ. 2010. Т. 2, № 4. С. 130-140.
10. van der Hofstad R. Random graphs and complex networks. Vol. 1. Eindhoven Univ. of Technology, 2014. 338 p. http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf
11. Aiello W., Chung F., Lu L. A random graph model for power law graphs // Exp. Math. 2001. Vol. 10, No. 1. P. 53-66.
12. Лебедев А.В. Экстремальные индексы в схеме серий и их приложения // Информатика и ее применения. 2015. Т. 9, № 3. С. 39-54.