ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 140 за 2019 г.

ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Подлазов А.В.

Модель роста социальной
сети с локальными
правилами

Рекомендуемая форма библиографической ссылки: Подлазов А.В. Модель роста социальной сети с локальными правилами // Препринты ИПМ им. М.В.Келдыша. 2019. № 140. 8 с.
<http://doi.org/10.20948/prepr-2019-140>
URL: <http://library.keldysh.ru/preprint.asp?id=2019-140>

О р д е н а Л е н и н а
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Р о с с и й с к о й а к а д е м и и н а у к

А.В. Подлазов

**Модель роста социальной сети
с локальными правилами**

Москва — 2019

А.В. Подлазов

Модель роста социальной сети с локальными правилами

В работе предлагается подход к описанию роста социальной сети с опосредствованным связыванием, позволяющим полностью исключить из правил преимущественный выбор узлов. Модель, построенная на основе этого подхода, порождает разреженную сеть со степенным распределением узлов по валентности.

Формулы для показателя распределения, получаемые из среднеполевого квазилинейного приближения, выполняются с хорошей точностью. Их эмпирические характеристики определяются из результатов моделирования. Однако согласовать данное приближение с величиной сдвига распределения не удастся, что ограничивает применимость предлагаемого подхода.

Ключевые слова: социальные сети, масштабная инвариантность, степенные распределения, конечно-размерный скейлинг, конкурентный рост, преимущественное присоединение, опосредованное связывание

A.V. Podlazov

Social network growth model with local rules

I propose a description of a social network with indirect binding. This allows you to exclude completely the preferred choice of nodes from the rules. A model built on this basis generates a sparse network with a power-law distribution of nodes degree.

The formulas for the distribution index obtained from the mean-field quasi-linear approximation satisfied with good accuracy. Their empirical characteristics are determined from the simulation results. However, this approximation doesn't agree with the value of the distribution shift. That limits the applicability of the proposed approach.

Key words: social networks, scale invariance, power law distributions, finite-size scaling, competitive growth, preferential accession, mediated binding

Работа выполнена при поддержке РФФИ (проекты 18-01-00619 и 19-01-00602).

Содержание

Введение	3
Правила модели	3
Приближенная теория	4
Анализ результатов	7
Выводы	8

ВВЕДЕНИЕ

Существует два принципиально разных механизма возникновения степенных распределений вероятностей. Один из них основывается на самоорганизации в критическое состояние, которая происходит в открытых далеких от равновесия нелинейных системах [1,2]. Другой связан с опорой на изначальную целостность системы и в общем случае не требует ни нелинейности, ни взаимодействия элементов системы между собой [3,4].

Крайним выражением целостного подхода являются модели конкурентного роста, в которых изменение размера элемента прямо пропорционально его текущему размеру. Целостность при этом используется лишь на уровне знания суммарного размера системы, необходимого для того, чтобы распределить общий его прирост между ее элементами [3]. Без этого было бы невозможно сбалансировать процессы появления в системе новых элементов и роста уже присутствующих в ней. Зато введение данного нефизичного предположения позволяет стоять модели, рождающие масштабно-инвариантные системы на основе очень простых правил.

Модели роста социальных сетей обычно строятся по этой же схеме, используя принцип *преимущественного присоединения* [5,6,7,8,9], согласно которому узел сети выбирается для образования новой связи с вероятностью, пропорциональной его *валентности* (числу уже имеющихся у него связей). Однако знание структуры графа позволяет выбирать узлы с такой вероятностью на основе локальных правил, не требующих никакой информации об интегральных характеристиках системы. Для простоты положим, что мы имеем дело с ненаправленным графом. Тогда, выбрав в нем случайный узел, а у того – случайного соседа, мы получим узел, выбранный преимущественным образом [10].

ПРАВИЛА МОДЕЛИ

На описанной идейной основе были сформулированы следующие правила модели роста социальной сети, в которой узлы интерпретируются как люди, а связи – как их знакомства. На очередном шаге времени t в сети выбирается случайный узел – *хозяин*, к которому присоединяется новый узел – *новичок*. По поводу появления в сети новичка хозяин созывает всех своих знакомых на *вечеринку*, где делает a попыток их попарно познакомить друг с другом. Если двое случайно выбранных различных знакомых хозяина уже связаны, попытка считается потраченной впустую. Если же нет – между ними образуется связь.

Рост сети стартует с затравки, состоящей из 2 связанных узлов. Поскольку на каждом шаге добавляется по одному узлу, удобно принять этот момент за $t = 2$, чтобы номер шага времени задавал число узлов в сети: $n(t) = t$.

Здесь предполагается, что *активность* хозяина a не зависит от его валентности x , т.е. рассматривается простейший вариант правил. Результат их компьютерной реализации представлен на рис. 1. Как можно видеть, эти прави-

Рис. 1. Плотность распределения узлов сети $u(x)$ по валентности x при различных a

По мере увеличения активности хозяина a у распределения появляется всё более выраженный степенной участок, представляемый в двойном логарифмическом масштабе прямой линией. Показатель распределения при этом падает, а максимальная валентность растёт.

ла порождают социальную сеть со степенным распределением узлов по валентности, показатель которого зависит от a . Однако при его малых и больших значениях вид графиков в их правых частях оказывается качественно различным, что затрудняет единообразную обработку результатов моделирования, ставя ее в зависимость от используемой теоретической модели. Без такой модели сколь-либо точно определить показатель распределения в случае малых a на доступной статистике не представляется возможным.

ПРИБЛИЖЕННАЯ ТЕОРИЯ

Пары узлов, которые пытается связать хозяин в ходе вечеринки, не независимы, т.е. если выбор одного из них является преимущественным, то второго — уже нет. Более того, коль скоро эти узлы имеют как минимум одного общего знакомого, вполне возможно, что их уже кто-то пытался познакомить между собой, что уменьшает шансы на успех попытки образования связи. Для описания вероятности выбора и связывания таких узлов здесь используется квазилинейное среднеполевое приближение, введенное в работах [8,9].

Будем считать, что узлов, успешно образующих новую связь в ходе вечеринки, случайным образом выбирается $b + k \cdot a$ штук, а преимущественным —

$l \cdot a$ штук, где b , k и l – эмпирические характеристики модели, полагаемые не зависящими ни от размера сети n , ни от валентности хозяина x , ни от его активности a . Необходимость введения члена b связана с тем, что изначально ни с кем не знакомый новичок успешнее образует связи со знакомыми хозяина, чем они между собой. Мерой успеха связывания старожилов служат величины k и l .

В рамках рассматриваемого приближения рост валентности узла описывается дифференциальным уравнением

$$\frac{dx}{dt} = \frac{1+b+ka}{n(t)} + \frac{la}{n(t)} \cdot \frac{x}{m}, \quad (1)$$

где слагаемое 1 в числителе первого слагаемого учитывает связь, образуемую новичком с хозяином, а m – средняя валентность узла. Она вычисляется как сумма скоростей роста всех n узлов, увеличенная на 1 для учета и второго конца связи, ведущей к новичку. Таким образом,

$$m = (2+b) + (k+l)a. \quad (2)$$

В предположении стационарности распределения $u(x)$ количество узлов валентности x представимо как $N(x,t) = n(t) \cdot u(x)$. Оно удовлетворяет уравнению неразрывности

$$\frac{\partial N}{\partial t} + \frac{\partial(N\dot{x})}{\partial x} = 0,$$

из которого находится вид распределения

$$u(x) \sim (x+\delta)^{-(1+\alpha)}, \quad (3)$$

$$\text{где } \alpha = m/la \quad (4)$$

$$\text{и } \delta = \alpha(1+b+ka). \quad (5)$$

На конечных временах область применимости формулы (3) ограничена максимальной валентностью x^* , динамика которой получается интегрированием уравнения (1). Без потери общности можно считать, что узел максимальной валентности входил в изначальную заправку, что дает $y^*(t) = x^*(t) + \delta \sim t^{1/\alpha} = n^{1/\alpha}$.

В свете сказанного, ожидая скейлингового поведения плотности, его можно переписать в виде

$$u(y) = y^{-(1+\alpha)} g(y/y^*), \quad (6)$$

где обозначено $y = x+\delta$, а $g(z) = \text{const}$ при $z \ll 1$ и $g(z) \rightarrow 0$ при $z \rightarrow \infty$.

Используя представление (6) для анализа эмпирических плотностей, определим параметры распределения, как показано на рис. 2.

Рис. 2. Скейлинг плотности распределения

Даже не зная вида функции g , можно подобрать параметры α и δ , добиваясь совмещения графиков плотности, полученных при различных n , в области средних и больших аргументов.

Обращает на себя внимание различная форма несовмещения графиков в области малых аргументов, возникающая при малых и больших a .

АНАЛИЗ РЕЗУЛЬТАТОВ

Знание значений показателя α при различных a позволяет оценить эмпирические характеристики модели с помощью формулы (4). Если переписать ее в виде $m/\alpha = la$, то удастся определить величину $l = 0,467(1)$ как угловой коэффициент соответствующего графика, представленного на рис. 3а. А если с помощью выражения (2) представить формулу (4) в виде

$$\alpha = \frac{m}{la} = 1 + \frac{k}{l} + \frac{2+b}{la}, \quad (7)$$

то из параметров графика с рис. 3б находятся $b = 1,74(4)$ и $k = 0,128(13)$.

В подобном анализе есть одна тонкость, связанная с тем, что в реальности m возрастает с течением времени, постепенно приближаясь к асимптотическому пределу. Он находится экстраполяцией значений, полученных при конечных n . Предельное значение m превосходит ее величину при $n = 2^{26}$ на 0,33 для $a = 50$, на 0,047 для $a = 35$ и менее чем на 0,001 для меньших a .

Средняя валентность m является третьей характеристикой модели, независимо определяемой из эксперимента. Ее величину невозможно даже оценить с помощью распределения (6), т.к. при $\alpha > 1$ его первый момент набирается в области малых x , где скейлинговая запись неточна. Более того, как следует из формулы (7), $\alpha > \alpha_{\min} = 1,275(3) > 1$ даже при $a \rightarrow \infty$. А если $a < 11,1(5)$, то $\alpha > 2$, т.е. уже и второй момент распределения набирается в области малых x . Этим, по всей видимости, и обуславливаются различия в поведении графиков на рис. 1 и 2 при малых и больших a . Принципиальным выражением этих различий оказывается роль сдвига δ в анализе данных. При совмещении графиков с $\alpha < 2$ можно приближенно принять $\delta = 0$, сохраняя скейлинговое поведение плотности, хотя и ухудшая точность определения показателя. А для графиков с $\alpha > 2$ плотность, если не сдвигать ее аргумент, демонстрирует уже нескейлинговое поведение.

Рис. 3. Определение эмпирических характеристик модели

Использование на одном рисунке в качестве аргумента a , а на другом – $1/a$ позволяет равно наглядно представить малые и большие активности.

Рис. 4. Сравнение эмпирического и расчетного сдвигов валентности

Графики схожи качественно при полном количественном несовпадении.

Завершением анализа модели должно было бы стать подтверждение формулы (5). Но, как показывает рис. 4, на котором приведена зависимость от средней валентности m ее общего сдвига δ , эта формула выполняется из рук вон плохо.

На самом деле, ситуация еще хуже. Для рис. 3 выбрано представление данных, дающее наибольшие R^2 . Однако при непосредственной проверке формулы (4) отклонение точек от прямой, связанное с высокими погрешностями α при малых a , уже заметно глазом, несмотря на $R^2 = 99,2\%$.

Выводы

Предложенная модель порождает масштабно-инвариантную социальную сеть на основе строго локальных правил. Тем не менее их рассмотренный вариант следует признать неудачным. Модель оказалась сложна с точки зрения обработки результатов моделирования, а ее простейшее теоретическое описание приводит к расхождениям с экспериментальными данными. Вместе с тем сама идея отказа от преимущественного выбора узлов представляется перспективной. Она требует исследования моделей с другими вариантами правил.

Литература

1. Бак П. Как работает природа: Теория самоорганизованной критичности. – М.: Либроком, 2015. – 276 с.
2. Подлазов А.В. Теория самоорганизованной критичности – наука о сложности // Будущее прикладной математики. Лекции для молодых исследователей. – М.: УРСС, 2005. С.404-426.
3. Подлазов А.В. Закон Ципфа и модели конкурентного роста // Новое в синергетике. Нелинейность в современном естествознании. – М.: Либроком, 2009. С.229-256.
4. Подлазов А.В. (2002) Ветвящийся процесс с зависимыми частицами как модель катастрофического поведения // Математическое моделирование 14(9), 53-58.
5. Barabási A.-L., Albert R. (1999) Emergence of scaling in random networks // Science 286, 509-512.
6. Barabási A.-L., Albert R., Jeong H. (2000) Scale-free characteristics of random networks: The topology of the world-wide web // Physica A 281, 69-77.
7. Albert R., Barabási A.-L. (2002) Statistical mechanics of complex networks // Rev. Mod. Phys. 74(1), 47-97.
8. Подлазов А.В., Щетинина Д.П. Модель роста социальной сети // Препринты ИПМ им. М.В. Келдыша, №95, 2013, 16 с.
9. Подлазов А.В. Модифицированная модель роста социальной сети // Препринты ИПМ им. М.В. Келдыша, №111, 2017, 16 с.
10. Vázquez A. (2003) Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Phys. Rev E 67, 056104.