

# МИКРО- И МАКРОМОДЕЛИ СОЦИАЛЬНЫХ СЕТЕЙ.

## Ч. 2. Идентификация и имитационные эксперименты

А.В. Батов, В.В. Бреер, Д.А. Новиков, А.Д. Рогаткин

Рассмотрены вопросы идентификации введенных в первой части работы микро- и макрохарактеристик социальных сетей по данным из реальных онлайн-социальных сетей — Facebook, LiveJournal и Twitter. Представлены результаты ряда соответствующих имитационных экспериментов, дано их сравнение.

**Ключевые слова:** пороговое поведение, социальная сеть, теория графов.

### ВВЕДЕНИЕ

В статье [1] были введены понятия *микро- и макромоделей социальных сетей с единым относительным порогом*  $\theta \in [0, 1]$  (см. также в указанной статье соответствующий обзор близких исследований).

В *микромоделе* поведение агента в рамках игрового подхода описывается с помощью его наилучшего ответа  $BR_i$  (Best Response) — агенты одновременно и независимо принимают решение в каждом периоде времени  $k$ :

$$x_i^{(k)} = BR_i(x_{-i}^{(k-1)}) = \begin{cases} 1, & \sum_{j \in D_i} x_j^{(k-1)} > \theta d_i \\ 0, & \sum_{j \in D_i} x_j^{(k-1)} \leq \theta d_i \end{cases},$$

$$i \in N = \{1, 2, \dots, n\}, \quad (1)$$

где  $d_i = |D_i|$  — число соседей агента  $i$ ,  $D_i$  — множество его соседей,  $x_{-i}^{(k-1)} = \{x_1^{(k-1)}, x_2^{(k-1)}, \dots, x_{i-1}^{(k-1)}, x_{i+1}^{(k-1)}, \dots, x_n^{(k-1)}\}$  — обстановка для агента  $i$ . Здесь  $x_i^{(k)} \in \{0; 1\}$ , где  $x_i^{(k)} = 1$  означает, что агент действует, а  $x_i^{(k)} = 0$  — бездействует в период времени  $k$ .

*Макромодель* описывает динамику доли действующих агентов  $p_k = \frac{1}{n} \sum_i x_i^{(k)} \in [0, 1]$ :

$$p_{k+1} = F_n(p_k, \theta), \quad (2)$$

где

$$F_n(p, \theta) = \sum_{d=1}^{n-1} B(p, d, \theta) M(d), \quad (3)$$

$B(p, d, \theta) = 1 - \sum_{k=0}^{[\theta d]} C_d^k p^k (1-p)^{d-k}$  — функция от

биномиального распределения,  $M(d)$  — вероятностное распределение числа соседей  $d$  в графе социальной сети (см. подробнее в работе [1]).

Для проверки адекватности и сравнения между собой теоретических моделей (1) и (2) могут быть использованы данные из реальных онлайн-социальных сетей (СС). В настоящей работе для микромоделей (1) применяются методы имитационного моделирования, для макромоделей (2) — различные виды аппроксимаций функции распределения (3).

### 1. ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЯ

Связи между агентами реальной СС можно представить в виде ориентированного графа  $G$ . Ориентация ребра графа от одного агента (узла) к другому интерпретируется как влияние одного агента на другого. В микромоделах явно используется граф влияния, а в макромоделах — его макрохарактеристика: распределение  $M(\cdot)$  числа соседей. Исходя из условий возможности применения макромоделей (2), а именно из того, что число агентов должно быть достаточно велико, были исследованы данные о подобных связях влияния в трех социальных сетях: русскоязычные сегменты СС Facebook ( $F$ ), LiveJournal ( $L$ ) и Twitter ( $T$ ).

Макропоказатели социальных сетей

CC	MaxFrnds	Users	Nonzero users	Links	AvgFriends
Facebook	4199	3 250 580	3 084 017	77 639 757	50,35
Live Journal	2499	5 758 706	3 586 959	124 729 288	34,77
Twitter	759 313	~41 700 000	35 427 738	1 418 363 662	40,04

Так, в сети Facebook определенный агент связан со своими друзьями, что может интерпретироваться как связи влияния этих друзей на агента. В социальных сетях LiveJournal и Twitter в качестве ориентированных связей влияния использовались подписки агента на просмотр и комментирование информации других агентов. Во всех трех социальных сетях агентов, которые влияют на данного агента, будем считать его *соседями* (см. выражение (1)).

В табл. 1 приведены макропоказатели трех исследуемых СС: максимальное число соседей (MaxFrnds), число агентов (Users), число агентов с ненулевым количеством соседей (Nonzero users), общее число связей (Links) и среднее число соседей у агентов с ненулевым числом соседей ( $AvgFriends = Links/Nonzero\ users$ ).

Как видно из табл. 1, в рассматриваемых СС количество агентов велико, что позволяет выдвинуть гипотезу о применимости для них макромоделей (2).

В функции распределения (3) макромоделей (2) присутствуют две составляющие: вероятность  $V(p, d, \theta)$  того, что доля  $p$  из числа  $d$  агентов будут действовать, и распределение  $M(\cdot)$  числа соседей в СС. Определять эти функции можно по-разному, что порождает следующий набор задач исследования.

**Задача 1.** Идентификация функций распределения  $M(\cdot)$  в указанных трех СС. В рамках этой задачи строятся эмпирические распределения  $M_F(\cdot)$ ,  $M_L(\cdot)$  и  $M_T(\cdot)$  и ищутся аппроксимирующие их в аналитическом виде функции  $\bar{M}_F(\cdot)$ ,  $\bar{M}_L(\cdot)$  и  $\bar{M}_T(\cdot)$ .

**Задача 2.** Построение и исследование имитационных моделей порогового поведения, которое задается наилучшим ответом (1). Так, действующими в начальный момент времени считались случайно выбранные агенты и вычислялось, согласно выражению (1), число действующих агентов на следующем шаге. Затем результат усреднялся по случайным множествам первоначально выбранных агентов. В результате было построено семейство функций (зависящее от параметра  $\theta$ ), которые сравнивались с другими функциями распределения, полученными в результате решения других задач настоящего исследования (см. табл. 5).

**Задача 3.** Аппроксимация сигмовидными функциями зависимостей, полученных при имитационном моделировании в рамках решения задачи 2.

**Задача 4.** Нахождение семейства функций распределения (3) (зависящего от параметра  $\theta$ ), в которые вместо  $M(\cdot)$  подставляются эмпирические функции распределения степеней узлов графов связей для трех рассматриваемых реальных СС —  $M_F(\cdot)$ ,  $M_L(\cdot)$  и  $M_T(\cdot)$ .

**Задача 5** аналогична четвертой, но вместо эмпирических функций распределения СС используются их аппроксимации  $\bar{M}_F(\cdot)$ ,  $\bar{M}_L(\cdot)$  и  $\bar{M}_T(\cdot)$ , найденные в результате решения задачи 1.

Для решения задач 2—5 применялись два метода: анализ эмпирических данных и их аналитическая аппроксимация. Схему исследования можно представить в виде табл. 2.

**Задача 7.** Исследование зависимости от единого относительного порога  $\theta$  положений равновесия коллективного поведения в рамках модели Грановеттера [1, 2], которую можно построить для рассматриваемых СС.

Перейдем к последовательному рассмотрению и описанию результатов решения перечисленных задач.

## 2. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

### 2.1. Идентификация функций распределения $M(\cdot)$ числа соседей агентов в социальных сетях (задача 1)

Существует множество исследований онлайн-новых СС, свидетельствующих, что распределение числа соседей или, другими словами, распределенные степени узлов больших СС хорошо аппроксимируются степенным распределением (см., на-

Таблица 2

Модели и методы для решения задач 2—5

Модель	Метод	
	Анализ эмпирических данных	Аналитическая аппроксимация
Микромодель СС	Задача 2	Задача 3
Макромодель СС	Задача 4	Задача 5

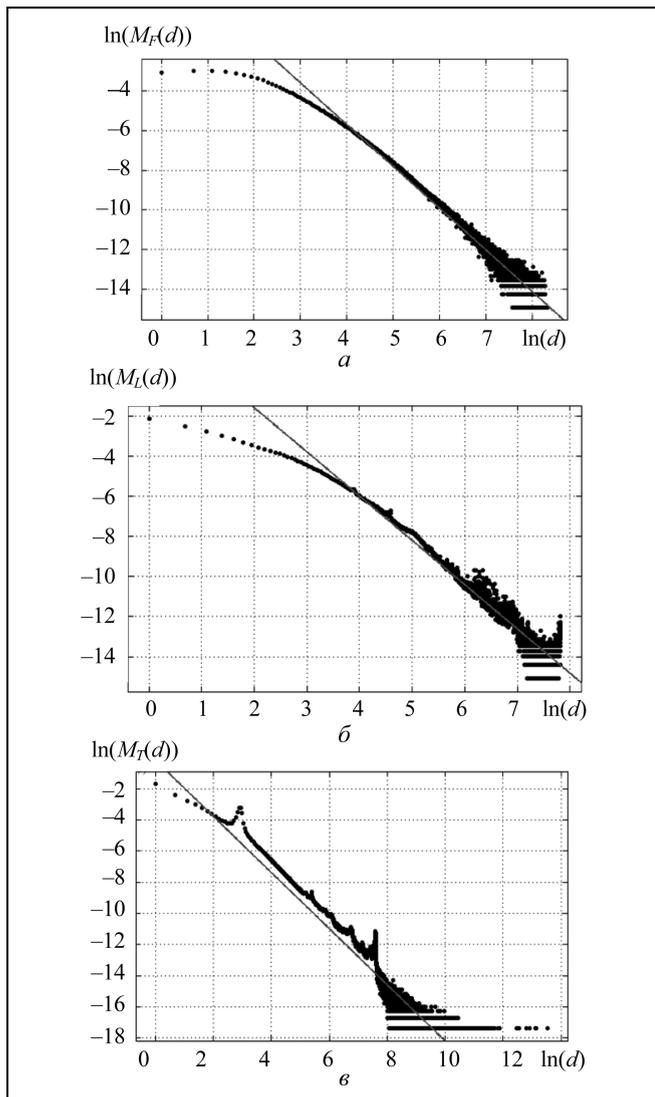


Рис. 1. Графики функций распределений числа соседей агентов в социальных сетях и ее линейной аппроксимации:  $a$  — CC Facebook;  $b$  — CC LiveJournal;  $c$  — CC Twitter

пример, работы [3–5]). Результаты исследований этой зависимости для рассматриваемых СС приведены на рис. 1, где в двойном логарифмическом масштабе (когда по обеим осям отложены логарифмы соответствующих величин) построены графики эмпирических распределений числа соседей  $\bar{M}_F(\cdot)$ ,  $\bar{M}_L(\cdot)$  и  $\bar{M}_T(\cdot)$ . Так как степенная функция в двойном логарифмическом масштабе имеет вид прямой с наклоном  $a$  и значением в нуле  $b$ , то была найдена наилучшая линейная аппроксимация. Полученные значения коэффициентов аппроксимации для различных СС приведены в табл. 3. Остальные обозначения будут описаны далее.

Кривая распределения степеней узлов графа  $G$  для малого числа соседей может быть приближена горизонтальной прямой, что приводит к «срезан-

ным» линейным аппроксимациям  $\bar{M}_F(\cdot)$ ,  $\bar{M}_L(\cdot)$  и  $\bar{M}_T(\cdot)$ , которые представлены на рис. 2 (см. также табл. 4).

Была выбрана горизонтальная аппроксимационная прямая для малых значений числа соседей, поскольку:

— необходимо обеспечить выполнение условия нормировки распределения — площадь под графиком должна быть равна единице; если осуществлять нормировку непосредственно, то изменяются коэффициенты степенного распределения и тем самым изменяется точность аппроксимации;

— «срезанная» линейная аппроксимация дает меньшую ошибку  $R$ -квадрат в сравнении с другими аппроксимациями: как альтернатива рассматривалась аппроксимация кривой распределения Парето (с параметром  $a_{pareto}$ , см. табл. 3), которая в двойном логарифмическом масштабе имеет вид прямой. Как видно из табл. 4, «срезанная» линейная аппроксимация приближает эмпирическое распределение лучше.

Значение  $c$  аргумента, при котором линейная горизонтальная аппроксимация переходит в «наклонную»:

$$\bar{M}(d) = \begin{cases} \exp[b] \times c^a, & d \leq c, \\ \exp[b] \times d^a, & d > c, \end{cases}$$

вычислялось (см. значения  $c_{real}$  в табл. 3) исходя из условия нормировки распределения:

$$\sum_{d=1}^n \bar{M}(d) = 1.$$

После построения аппроксимаций распределения степеней графов реальных СС, которые будут

Таблица 3

Коэффициенты аппроксимации функций  $M_F(\cdot)$ ,  $M_L(\cdot)$  и  $M_T(\cdot)$

Функция	$a$	$b$	$c_{real}$	$a_{pareto}$
$M_F(\cdot)$	-2,181	3,274	26,628	0,688
$M_L(\cdot)$	-2,208	2,828	16,878	0,765
$M_T(\cdot)$	-1,802	-0,196	1,8233	0,799

Таблица 4

Точность «срезанной» линейной и Парето-аппроксимаций ( $R$ -квадрат)

Аппроксимация СС	«Срезанная» линейная	Парето
Facebook	0,962	0,916
LiveJournal	0,929	0,884
Twitter	0,849	0,849

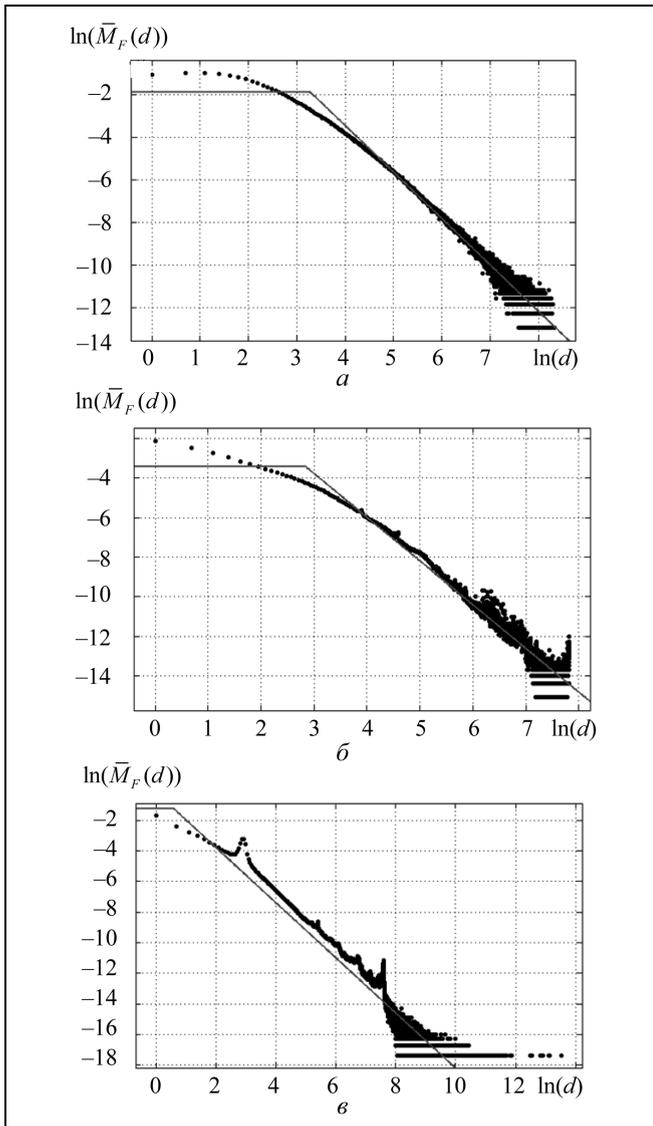


Рис. 2. Графики «срезанной» линейной аппроксимации функции  $M(\bullet)$ :  $a$  — CC Facebook;  $b$  — CC LiveJournal;  $v$  — CC Twitter

использованы далее в рамках решения задачи 4, перейдем к описанию результатов имитационного моделирования поведения агентов в микромодели (1).

## 2.2. Построение и исследование имитационных моделей порогового поведения с наилучшим ответом (1) (задача 2)

Имитационное моделирование состояло в следующем. В CC Facebook и LiveJournal, заданных своими графами связей, случайным образом «возбуждалось» некоторое число  $q$  агентов (доля  $q/n \in [0; 1]$ ). Далее по формуле (1) для каждого агента вычислялся его наилучший ответ (действовать или бездействовать). Получающаяся в результате доля действующих агентов, согласно форму-

ле (2) — это значение функции  $F_n(q/n, \theta)$ . Эксперимент повторялся многократно для каждого из различных значений  $q$ , лежащих на отрезке  $[0, 1]$ . Относительное отклонение значения функции  $F_n(q/n, \theta)$  составляло порядка 0,001 во всех экспериментах (разброс значений связан со случайностью выбора множества первоначально возбужденных агентов). Кривые  $F_n(q/n, \theta)$  при различных значениях параметра  $\theta$  изображены на рис. 3.

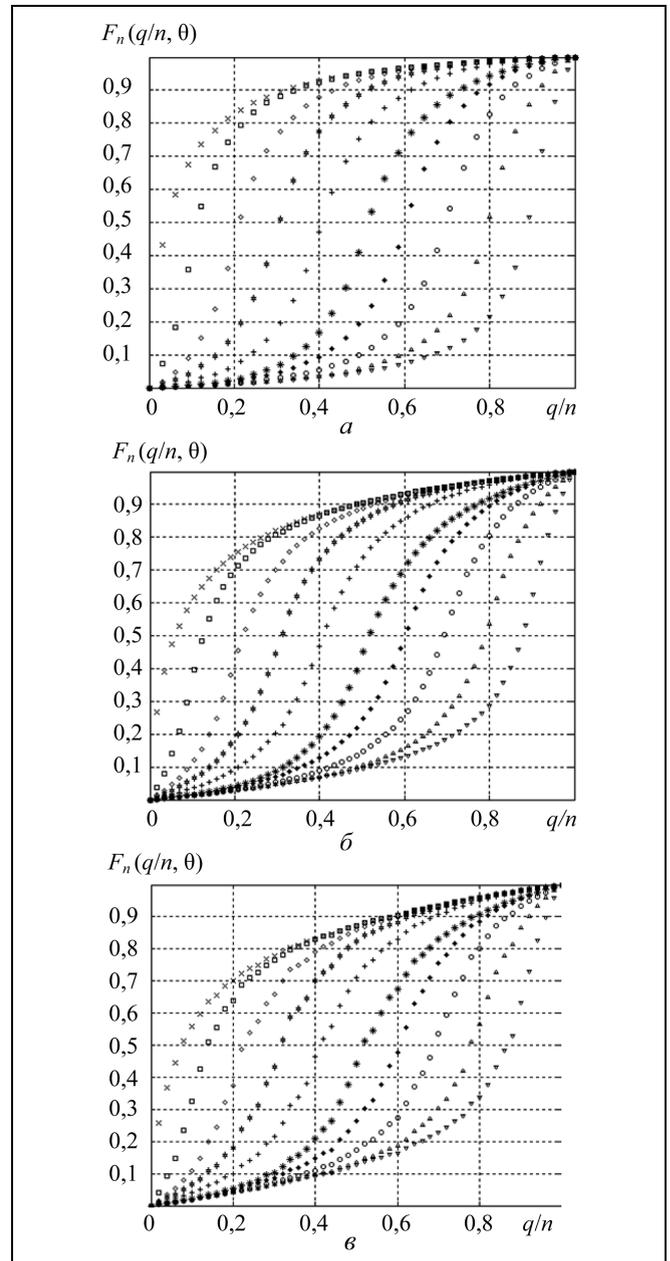


Рис. 3. Результаты имитационного моделирования порогового поведения:  $a$  — CC Facebook;  $b$  — CC LiveJournal;  $v$  — CC Twitter;  $\times$  —  $\theta = 0$ ;  $\square$  —  $\theta = 0,1$ ;  $\diamond$  —  $\theta = 0,2$ ;  $+$  —  $\theta = 0,3$ ;  $+$  —  $\theta = 0,4$ ;  $*$  —  $\theta = 0,5$ ;  $\blacklozenge$  —  $\theta = 0,6$ ;  $\circ$  —  $\theta = 0,7$ ;  $\Delta$  —  $\theta = 0,8$ ;  $\nabla$  —  $\theta = 0,9$

Получив результаты имитационного моделирования, перейдем к их аппроксимации.

### 2.3. Аналитическое приближение функций $F_n(p, \theta)$ , полученных при имитационном моделировании (задача 3)

Необходимо найти аналитическое приближение семейства функций  $F_n(p, \theta)$  для каждой из рассматриваемых СС. На основании проведенных исследований были сделаны следующие наблюдения:

- полученные кривые относятся к классу сигмоид (рис. 4);
- кривые  $F_n(p, \theta)$  имеют перегиб при  $p \approx \theta$ .

С учетом этих наблюдений, в качестве кандидатов для аппроксимации были выбраны следующие параметрические семейства функций:

$$f(p, \theta, \alpha, \lambda, \gamma) = \alpha \cdot \operatorname{arctg} \lambda(p - \theta) + \gamma$$

и

$$g(p, \theta, \alpha, \lambda, \gamma) = \frac{\alpha}{1 + e^{-\lambda(p-\theta)}} + \gamma.$$

Из условия, что функции  $f(\cdot)$  и  $g(\cdot)$  должны являться функциями распределения, получаем параметрические (зависящие от параметра  $\lambda$ ) семейства функций:

$$f(p, \theta, \lambda) = \frac{\operatorname{arctg} \lambda(p - \theta) + \operatorname{arctg}(\lambda \theta)}{\operatorname{arctg} \lambda(1 - \theta) + \operatorname{arctg}(\lambda \theta)} \quad (4)$$

и

$$g(p, \theta, \lambda) = \left( \frac{1 - e^{-\lambda p}}{1 + e^{-\lambda(p-\theta)}} \right) \left( \frac{1 + e^{-\lambda(1-\theta)}}{1 - e^{-\lambda}} \right).$$

Таким образом, задача 3 сводится к нахождению одного неизвестного параметра  $\lambda$ , при котором поверхность  $f(p, \theta, \lambda)$  или  $g(p, \theta, \lambda)$  наилучшим образом приближает экспериментальные данные, и выбору семейства функций, которое дает меньшую ошибку приближения.

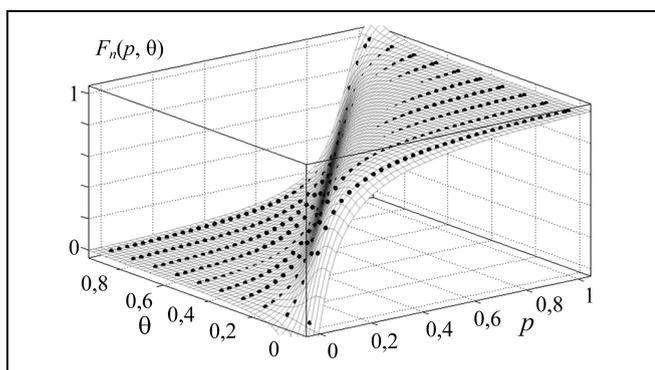


Рис. 4. Приближение экспериментальных данных  $F_n(p, \theta)$  (точки) аналитическим семейством  $f(p, \theta, \lambda_p)$  (сетка) для социальной сети Facebook

Оказалось, что экспериментальные данные для всех социальных сетей лучше приближаются семейством функций (4), при этом минимум ошибки приближения достигается при  $\lambda_F = 13,01$ ,  $\lambda_L = 9,18$ ,  $\lambda_T = 7,34$  (график для СС Facebook приведен на рис. 4).

Наличие аналитического выражения для функции (3) дает возможность построить однопараметрическую модель поведения агентов (при различных значениях единого порога  $\theta$ ), что важно, в частности, для задач управления СС [6].

### 2.4. Макромодель (2) на основе эмпирического распределения числа соседей в графе (задача 4)

Как отмечалось ранее, данная задача состоит в нахождении семейства функций распределения (зависящего от параметра  $\theta$ ) макромодели (3), в которых явно присутствует теоретическая составляющая  $B(p, d, \theta)$ , т. е. в выражении (3) вместо  $M(\cdot)$  подставляются эмпирические функции распределения степеней узлов графов связей для трех СС —  $M_F(\cdot)$ ,  $M_L(\cdot)$  и  $M_T(\cdot)$ . Результаты представлены на рис. 5.

Таким образом, мы построили семейство функций (3). Перейдем к их рассмотрению, используя «срезанные» линейные аппроксимации, полученные в результате решения задачи 1.

### 2.5. Макромодель (2) на основе распределения числа соседей, аппроксимированного аналитической функцией (задача 5)

Данная задача подобна задаче 4, но вместо эмпирических функций распределения СС используются их аппроксимации  $\bar{M}_F(\cdot)$ ,  $\bar{M}_L(\cdot)$  и  $\bar{M}_T(\cdot)$ . Результаты представлены на рис. 6.

Видно, что семейство функций (3) качественно аналогично полученному в результате решения задачи для соответствующих СС. Строгое сравнение осуществляется в рамках решения задачи 6.

### 2.6. Сравнение результатов решения задач 2—5 (задача 6)

Решив задачи 2 и 3 (имитационное моделирование и аппроксимация микромоделей (1)), а также задачи 4 и 5 (различные виды аппроксимаций макромодели (2)), можно сравнить полученные результаты — см. табл. 5, в заголовках столбцов которой содержатся пары номеров сравниваемых задач.

Видно, что макро- и микроописания дают схожие результаты (см. пары задач 2 и 3, 2 и 4, 3 и 4). Близость результатов решений задач 2 и 5, 3 и 5, 4 и 5 меньше, и в перспективе стоит вопрос нахождения функции, которая бы лучше приближала функцию распределения степеней графа  $M(\cdot)$ .

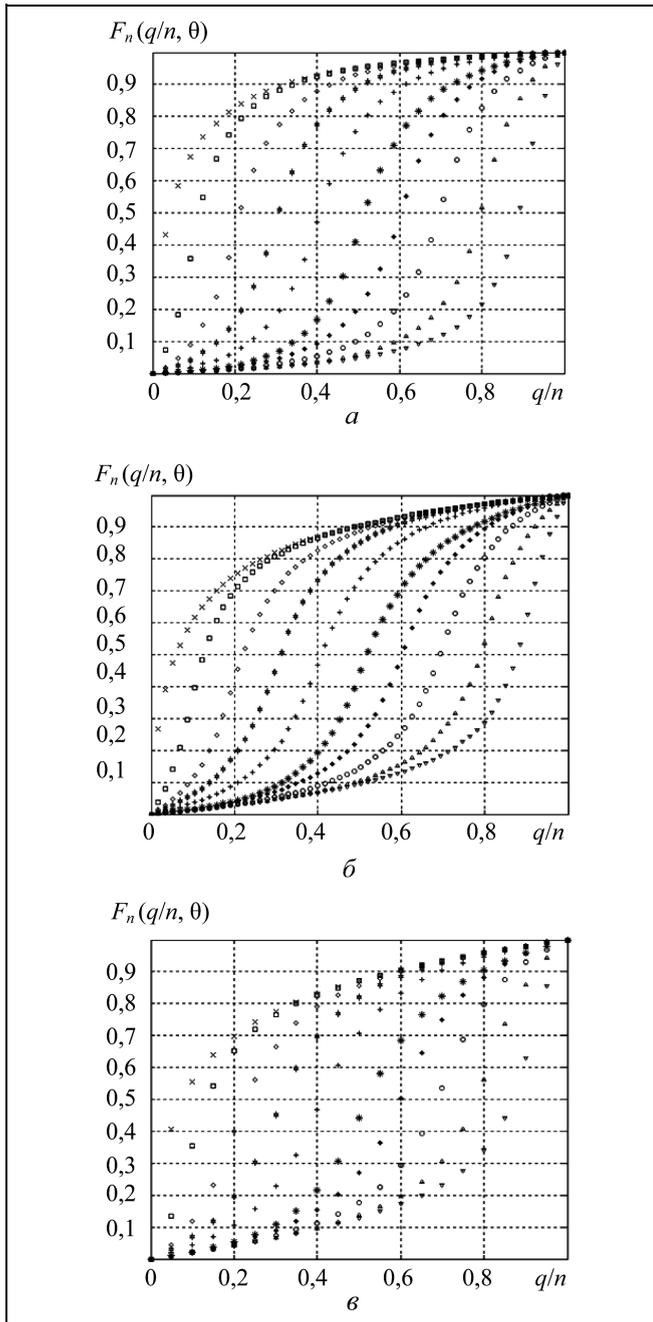


Рис. 5. Макромодель (2) с эмпирическим распределением числа соседей в графе: а – CC Facebook; б – CC LiveJournal; в – CC Twitter; × –  $\theta = 0$ ; □ –  $\theta = 0,1$ ; ◇ –  $\theta = 0,2$ ; † –  $\theta = 0,3$ ; + –  $\theta = 0,4$ ; \* –  $\theta = 0,5$ ; ◆ –  $\theta = 0,6$ ; ○ –  $\theta = 0,7$ ; △ –  $\theta = 0,8$ ; ▽ –  $\theta = 0,9$

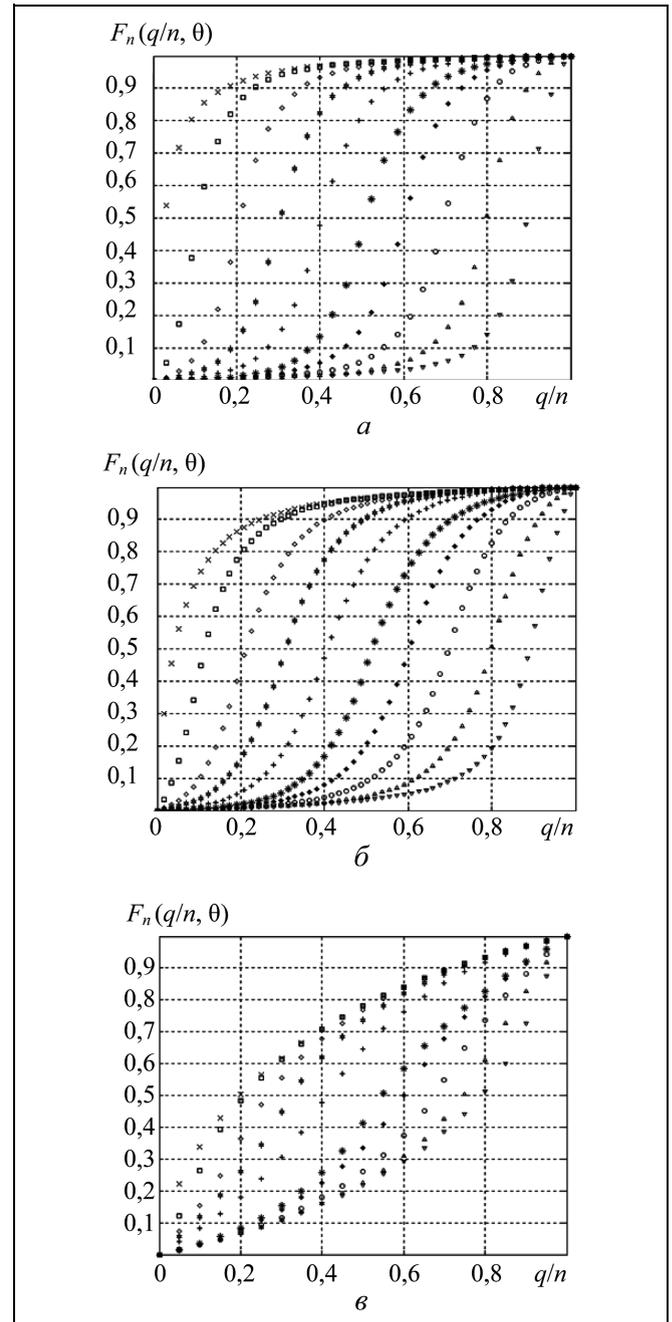


Рис. 6. Макромодель (2) с аппроксимированной функцией распределения: а – CC Facebook; б – CC LiveJournal; в – CC Twitter; × –  $\theta = 0$ ; □ –  $\theta = 0,1$ ; ◇ –  $\theta = 0,2$ ; † –  $\theta = 0,3$ ; + –  $\theta = 0,4$ ; \* –  $\theta = 0,5$ ; ◆ –  $\theta = 0,6$ ; ○ –  $\theta = 0,7$ ; △ –  $\theta = 0,8$ ; ▽ –  $\theta = 0,9$

Таблица 5

Сравнение результатов решения задач (значение R-квадрат)

CC	Задачи 2 и 4	Задачи 2 и 5	Задачи 4 и 5	Задачи 2 и 3	Задачи 3 и 4	Задачи 3 и 5
Facebook	0,9976	0,9932	0,9911	0,9973	0,9973	0,9907
LiveJournal	0,9999	0,9872	0,9872	0,9960	0,9960	0,9855
Twitter	0,9998	0,9631	0,9642	0,9949	0,9950	0,9599

## 2.7. Исследования положений равновесия в социальных сетях (задача 7)

Как было показано в работе [1], поведение агентов в модели с единым относительным порогом эквивалентно пороговому поведению в модели Грановеттера [2]. Для последней важны такие свойства положений равновесия (точки, характеризующиеся уравнением  $F_n(p, \theta) = p$ ), как их число и устойчивость (пересечение биссектрисы первого квадранта в точке равновесия «слева-сверху»). Ответим на эти вопросы для рассматриваемых СС. Из анализа графиков кривых, представленных на рис. 3–6, можно сделать вывод о том, что в зависимости от параметра  $\theta$  возможны различные наборы положений равновесия (их число, устойчивость и неустойчивость и др.).

В самом деле, точка пересечения кривой  $F_n(p, \theta)$  с диагональю единичного квадрата, лежащая внутри интервала  $(0, 1)$ , является неустойчивым положением равновесия, так как кривая  $F_n(p, \theta)$  пересекает диагональ «снизу вверх» (см. работу [2]).

Как видно из рис. 3, это свойство имеет место при  $\theta \in [\sim 0, 1; \sim 0, 9]$  для всех исследуемых социальных сетей. При этом точки  $q = 0$  и  $q = 1$  являются устойчивыми положениями равновесия.

Если же  $\theta \leq 0, 1$ , то в системе имеются два положения равновесия: точка  $q = 0$  является неустойчивым положением равновесия, а точка  $q = 1$  — устойчивым.

Аналогично, при  $\theta \geq 0, 9$ , в системе имеются два положения равновесия: точка  $q = 0$  является устойчивым положением равновесия, а  $q = 1$  — неустойчивым.

Исходя из рис. 7, можно также определять положение равновесия для рекуррентной процедуры (2) макромоделей для различных соотношений между начальным положением  $p_0$  и единым порогом  $\theta$ . Так, если вектор  $(p_0, \theta)$  лежит в области II (см. рис. 7), то процесс (2) завершится в положении равновесия  $p = 0$ . Если вектор  $(p_0, \theta)$  лежит в

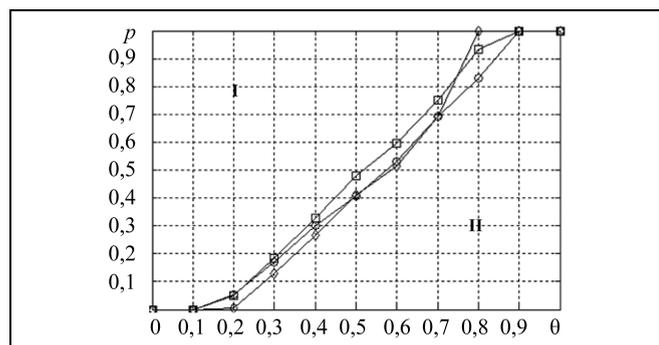


Рис. 7. Зависимость положения точки пересечения кривой  $F_n(p, \theta)$  с диагональю единичного квадрата от единого порога  $\theta$ : —○— Facebook; —□— LiveJournal; —◇— Twitter

области I, то процесс (2) завершится в положении равновесия  $p = 1$ .

## ЗАКЛЮЧЕНИЕ

Предложенные в работе [1] теоретические макромоделей социальных сетей исследовались с помощью имитационного моделирования на реальных данных и аппроксимации его результатов.

Из полученных результатов можно сделать два важных вывода.

Вероятностное описание макромоделей согласуется с микроописанием: результаты имитационного моделирования хорошо совпали с результатами вычисления по вероятностной модели на основании реального распределения степеней графа связей для различных социальных сетей.

Оказалось, что, несмотря на существенные различия масштабов и структуры графов связей реальных социальных сетей, их макромоделей (2) качественно очень похожи: имеют вид сигмиды и хорошо приближаются параметрическим семейством функций (4) с различными значениями коэффициента  $\lambda$ .

Авторы благодарны канд. техн. наук Д.А. Губанову, А.Ш. Яхину и Лаборатории Цифрового Общества — за предоставленные данные по социальным сетям Facebook и LiveJournal, канд. техн. наук А.В. Макаренко — за консультации по методам аппроксимации экспериментальных данных.

## ЛИТЕРАТУРА

1. Бреер В.В., Новиков Д.А., Рогаткин А.Д. Микро- и макромоделей социальных сетей. Ч. 1. Основы теории // Проблемы управления. — 2014. — № 5. — С. 28–33.
2. Granovetter M. Threshold Models of Collective Behavior // The American Journal of Sociology. — 1978. — Vol. 83, N 6. — P. 1420–1443.
3. Albert R., Barabasi A.-L. Statistical mechanics of complex networks // Rev. Mod. Phys. — 2002. — N 74. — P. 47–97.
4. Barabasi A., Albert R. Emergence of scaling in random networks // Science. — 1999. — № 286. — P. 509–512.
5. Barabasi A. Scale-free Networks // Scientific American. — 2003. — № 5. — P. 50–59.
6. Бреер В.В., Новиков Д.А., Рогаткин А.Д. Стохастические модели управления толпой // Управление большими системами. — 2014. — Вып. 52. — С. 85–117.

Статья представлена к публикации членом редколлегии Ф.Т. Алескеровым.

**Батов Алексей Владимирович** — науч. сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ☎ (495) 334-90-30, ✉ batov@ipu.ru,

**Бреер Владимир Валентинович** — канд. техн. наук, бизнес-аналитик, ЗАО «Авиахэлп Групп», г. Москва, ✉ breer@live.ru,

**Новиков Дмитрий Александрович** — чл.-корр. РАН, зам. директора, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ☎ (495) 334-75-69, ✉ novikov@ipu.ru,

**Рогаткин Андрей Дмитриевич** — мл. науч. сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, ✉ andreyrogatkin@gmail.com.