

УДК 021.8 + 025.1
ББК 78.34

ИСПОЛЬЗОВАНИЕ ЯЗЫКОВЫХ ИГР ДЛЯ ИССЛЕДОВАНИЯ СОЦИАЛЬНЫХ СЕТЕЙ НА ПРИМЕРЕ ПОИСКА СООБЩЕСТВ И ВЛИЯТЕЛЬНЫХ АГЕНТОВ

Губанов Д. А.¹, Микулич Л. И.², Наумкина Т. С.³
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Представлен метод исследования социальных сетей, основанный на использовании языковых игр. С помощью языковых игр можно решать такие задачи как выявление неявных сообществ или влиятельных агентов. Приведено описание игры в наименования и ее применение для решения этих задач. Описаны результаты исследования как искусственно сгенерированных графов, так и построенных на основе реальных данных. Также приведен краткий обзор других методов выявления сообществ и сравнение их с представленным методом.

Ключевые слова: социальные сети, сообщества, динамика мнений, игра в наименования.

1. Введение

Социальные сети являются объектом исследования уже давно, однако в последние годы интерес к этой теме возрос. Это связано, с одной стороны, с появлением большого объема дан-

¹ Дмитрий Алексеевич Губанов, кандидат технических наук
(dmitry.a.g@gmail.com).

² Леонид Ильич Микулич, кандидат технических наук
с.н.с. (lmik@ipu.ru).

³ Тамара Сергеевна Наумкина, аспирантка,
(tamara.naumkina@gmail.com)

ных для реальных социальных систем (в основном, онлайн-социальных сетей), а с другой стороны – с доступностью и дешевизной вычислительных ресурсов. Под социальной сетью понимается социальная структура, состоящая из множества агентов и определенного на нем множества отношений (совокупности связей между агентами). Эти отношения можно представить в виде графа (ориентированного или неориентированного), который обычно называется социальным графом.

Большое количество исследований социальных сетей посвящено выявлению неявных сообществ (т.е. не являющихся известными тематическими группами). Под сообществом мы будем понимать группу узлов, тесно связанных друг с другом и слабо связанными с узлами вне сообщества. Неявные сообщества можно выявить, исследуя топологию социального графа.

Другим предметом исследования социальных сетей является динамика мнений агентов. Мнение – это суждение агента по какому-то вопросу, оценка. Каждый агент обладает своим мнением, но может его менять под воздействием других агентов (соседей по графу). Структура социального графа влияет на динамику мнений. Например, можно предположить, что в сообществах мнения агентов будут одинаковы или близки.

Существует целый ряд моделей для исследования динамики мнений. Рассмотрим одну из них – игру в наименования [2, 3, 23]. (Заметим, что под словом «игра» в данном случае подразумевается не игра в теоретико-игровом смысле, а языковая игра, т.е. набор правил взаимодействия). Эта игра используется прежде всего для моделирования возникновения языка в многоагентных системах и различных языковых явлений, однако также находит и более прикладные применения, в том числе и для исследования социальных сетей. Метод применения игры для выявления сообществ был описан в [14]. Ниже также приведено сравнение этого метода с другими.

В следующем разделе статьи представлено описание игры в наименования. В разделе 3.1 приведен краткий обзор методов выявления сообществ в социальных сетях, описано применение игры в наименования для решения этой задачи и произведено

сравнение методов. В п. 3.2. описано применение игры в наименования для решения другой задачи – выявления влиятельных агентов на примере их выявления в популярной социальной сети. В разделе 4 приведены выводы статьи.

2. Игра в наименования

Игра в наименования – это языковая игра [1], т.е. набор правил взаимодействия, следуя которым агенты решают задачу формирования общего лексикона. Под лексиконом подразумевается соответствие между знаками (сигналами, словами и т.п.) и их значениями (объектами окружающей среды). Агенты попарно взаимодействуют друг с другом, вырабатывая общий язык. В общем виде игра в наименования моделирует не столько возникновение общего языка, сколько вообще процесс согласования чего-либо множеством агентов. Например, эта модель может применяться для выработки общего лексикона среди роботов [21, 22], для описания феномена создания совместных тегов пользователями интернета (collaborative tagging) [6], для моделирования распространения и сосуществования диалектов [24].

Общим свойством игры в наименования, характерным для всех ее модификаций и применений, является то, что соглашение в среде агентов возникает «стихийно», т.е. без внешнего управления, а как результат локальных взаимодействий.

Рассмотрим множество агентов A (размера N), помещенных в среду, в которой расположены также некоторые объекты из множества O (размера M). Агенты могут наблюдать (различать) объекты и обмениваться между собой сигналами. Каждый сигнал – это слово, т.е. последовательность символов из заданного алфавита. Лексикон L – это набор связей между объектами и словами, причем в нем возможна полисемия и синонимия, т.е. одному слову может соответствовать несколько объектов, а одному объекту – несколько слов. У каждого агента a из множества A свой собственный лексикон L_a , изначально пустой. Игра в наименования – это взаимодействие между двумя агентами – говорящим и слушающим (обычно они выбираются случайным

образом из множества A). Взаимодействие происходит следующим образом:

1) говорящий (обычно случайным образом) выбирает объект (o) из множества O и каким-то образом «указывает» на него слушающему;

2) говорящий находит в своем лексиконе слово (w), связанное с объектом o , или, если такого слова нет, сам его составляет из символов заданного алфавита;

3) говорящий передает выбранное слово w слушающему;

4) если слушающему известно это слово w и, согласно его словарю, оно означает тот же объект o , который выбрал говорящий, считается, что взаимодействие прошло успешно, и тогда оба игрока выполняют процедуру корректировки своих лексиконов: и говорящий, и слушающий удаляют из лексиконов все слова, кроме w , соответствующего объекту o ;

5) если же слушающему не известно слово w , взаимодействие заканчивается неудачей, и слушатель добавляет в свой лексикон связь между объектом o и словом w .

Такие локальные взаимодействия повторяются много раз между разными агентами. Цель сообщества агентов – выработать общий лексикон, т.е. максимально «сблизить» лексиконы агентов, тем самым увеличив, насколько возможно, долю успешно закончившихся игр.

Игра в наименования имеет множество вариаций, мы будем рассматривать самую популярную из них – минимальную игру в наименования. В ней лексикон L – это для каждого объекта список слов, ему соответствующих. Отметим, что в других вариациях игры в наименования лексиконы могут быть реализованы по-другому, например, как матрицы вероятностей соответствия слов и объектов [10].

На рис. 1 изображен пример взаимодействий агентов в минимальной игре в наименования.

Для наблюдения за ходом игр в момент времени t (время дискретно и соответствует количеству прошедших игр) используются следующие величины:

– $N_w(t)$ – общее количество слов, известное агентам;

- $S(t)$ – средняя доля успешных игр (за последние n игр);
- $CA(t)$ – коммуникативная точность – вероятность, что игра двух случайно выбранных агентов будет успешна.

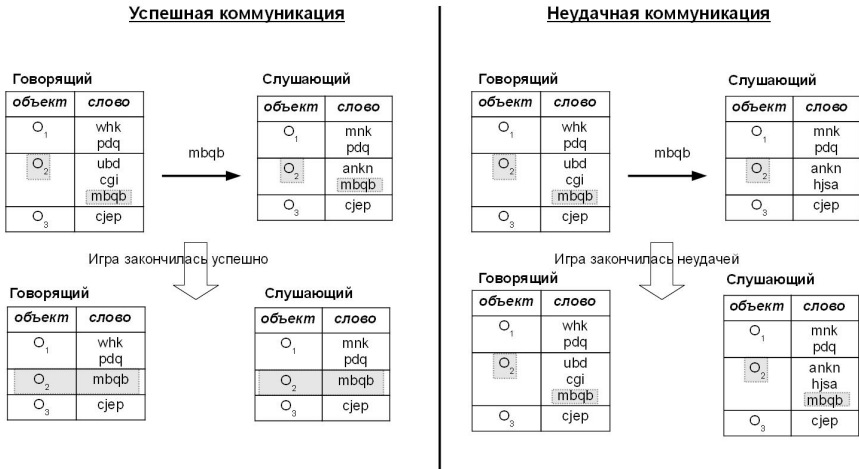


Рис. 1. Пример удачного и неудачного взаимодействия в игре в наименования (в обоих случаях выбран объект O_2)

Лучше всего процесс выработки общего лексикона отражает величина $CA(t)$, но практически ее сложно вычислить (особенно для больших M и N), поэтому чаще используется $S(t)$.

На рис. 2 показаны графики этих величин для нескольких экспериментов. В начальный момент времени лексиконы агентов пусты, поэтому все величины равны 0. Постепенно (после неудачных взаимодействий) агенты вводят слова, и величина N_w растет, пока не достигнет максимума (на рисунке около $t = 100$). Это означает, что у всех агентов сформировались связи для всех объектов. При этом коммуникативная точность (CA) достаточно маленькая, так как чаще всего у взаимодействующих агентов одинаковые объекты обозначают разные слова и игры заканчиваются неудачей. Потом величина N_w идет на убыль – это агенты «договариваются» между собой, какое слово обозначает какой объект (и «забывают» ненужные слова). При этом CA

постепенно растет, пока не достигает 1 (при этом N_w становится равной количеству объектов M). Этот момент времени называется временем сходимости T_c . После достижения времени сходимости все игры заканчиваются удачей и лексикон агентов не меняется.

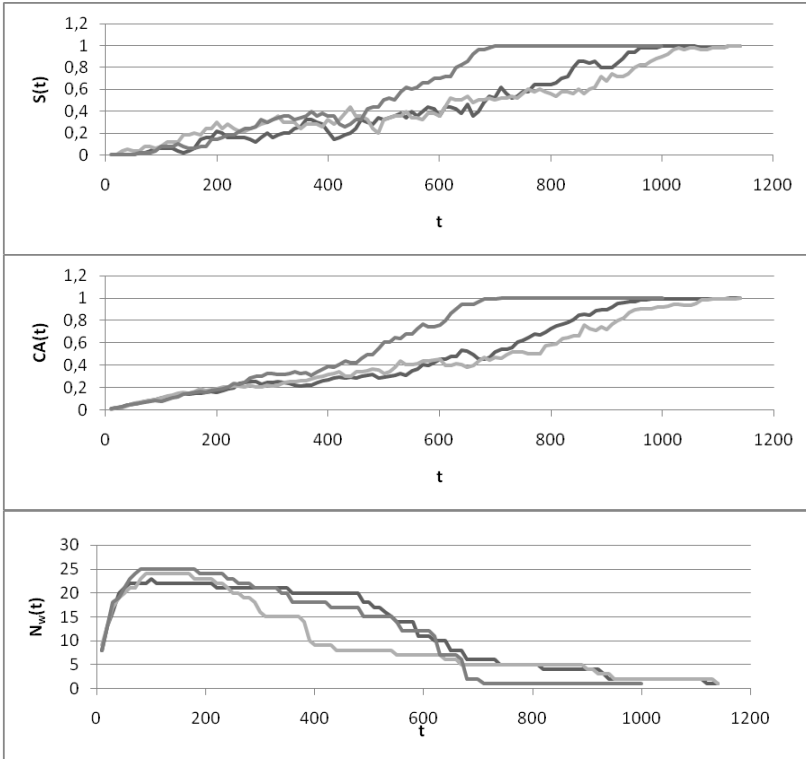


Рис. 2. Средняя доля успешных игр $S(t)$, коммуникативная точность $CA(t)$ и общее число слов $N_w(t)$ для трех экспериментов для $N = 50$ (количество агентов), $M = 1$ (количество объектов)

Игра в наименования имеет множество модификаций: агенты могут быть расположены на графе или на плоскости, количество агентов может меняться в процессе игры, сигналы могут

искажаться при передаче и т.п. Важной особенностью игры является то, что в большинстве случаев (в большом диапазоне исходных данных) последовательность игр приводит к формированию общего лексикона.

3. Применение игры в наименования для исследования социальных сетей

Будем рассматривать социальную сеть как граф, в котором агенты находятся в вершинах, а взаимодействие между ними осуществляется только тогда, когда между соответствующими вершинами есть ребра.

Лексикон, т.е. соответствие «слово – объект», можно интерпретировать как мнения агента по каким-либо вопросам. Тогда результат игр – это достижение (или недостижение) общего мнения (согласование). Обычно количество объектов берется равным одному, т.е. рассматривается мнение агентов по одному вопросу. В таком случае игра в наименования становится моделью динамики мнений.

3.1. ВЫЯВЛЕНИЕ СООБЩЕСТВ

На данный момент существует достаточно большое количество методов выявления сообществ, которые основываются на заданном во введении определении сообщества. Если определение дать более формально, то разбиение $P = \{C_1, \dots, C_k\}$ вершин графа $G = (V, E)$ ($\forall i, C_i \subseteq V$) на сообщества считается хорошим, если доля ребер внутри C_i высока по сравнению с долей ребер между C_i и другими сообществами.

Качество разбиения на сообщества может оцениваться различными показателями качества: нормированным разрезом [25], проводимостью [8], функцией Кернигана–Лина [11] и модулярностью [16]. Наибольшую популярность на сегодня для оценки качества разбиения приобрел показатель модулярности. Модулярность показывает то, насколько при заданном разбиении графа на сообщества плотность связей внутри сообществ больше плотности связей между сообществами. Идея модулярности

состоит в том, что чем больше отличается подграф, соответствующий сообществу, от случайного подграфа, тем лучше разбиение. То есть модулярность Q разбиения графа на k сообществ $\{V_1, \dots, V_k\}$ будет задаваться следующим образом:

$$Q = \sum_{i=1}^k \left[\frac{m(V_i, V_i)}{m} - \left(\frac{\deg(V_i)}{2m} \right)^2 \right],$$

где m – количество ребер в исходном графе; V_i – i -е сообщество; $m(V_i)$ – количество ребер в i -м сообществе.

Соответственно, один из наиболее очевидных методов разбиения сети на сообщества состоит в *жадной оптимизации модулярности* [7]: первоначально сеть разбивается на n сообществ (одна вершина – одно сообщество), затем последовательно осуществляется поиск пары сообществ, объединение которых даст наибольший прирост модулярности, и объединение найденных сообществ в одно сообщество.

Другой подход состоит в использовании так называемой промежуточности ребер графа: если два сообщества соединены малым количеством ребер, то очевидно, что через эти ребра будут проходить все пути между вершинами, принадлежащими различным сообществам (и эти ребра будут обладать наибольшим значением промежуточности); следовательно, удалив ребра с наибольшим значением промежуточности, мы получим искомого разбиение. Соответственно, для получения разбиения исходного графа необходимо последовательно рассчитывать промежуточность ребер и удалять ребро с наибольшим значением промежуточности [16]. Отметим, что пути можно определять различными способами: как кратчайшие пути, как определяемые случайным блужданием пути и как пути распространения тока (см. теорию схем).

Существуют и другие подходы разбиения сети на сообщества: методы, использующие цепи Маркова [26, 17], метод распространения меток [20], методы, использующие собственные вектора матриц [15], метод спинового стекла и имитации отжига [19] и др.

Вышеупомянутые методы выявления сообществ довольно сильно отличаются друг от друга, соответственно, различается их вычислительная сложность и качество работы. Обычно чем выше качество работы метода, тем выше вычислительная сложность. Далее будет предложен метод выявления сообществ на основе языковых игр и проведено сравнение предложенного метода с уже существующими наиболее популярными методами.

Скорость сходимости и вообще поведение системы сильно зависят от свойств графа [4]. Например, если группа узлов тесно связана между собой, то с большой вероятностью агенты из этой группы будут обладать одинаковым мнением. Если граф состоит из несколько таких групп, слабо связанных между собой, то сначала агенты «договорятся» друг с другом внутри групп, а затем группы договорятся между собой. Таким образом, если пронаблюдать за системой в тот момент, когда агенты внутри групп уже договорились между собой, но общий язык еще не возник, и посмотреть, какие агенты какого мнения придерживаются, можно выявить сообщества. Этот метод выявления сообществ был предложен в статье [14]. В ней рассматривался социальный граф (около 1000 вершин), на котором несколько раз запускались серии игр в наименования. Через 10 тыс. игр определялось, какого мнения в итоге придерживается каждый агент. Оказалось, что в большинстве случаев агенты разделялись на 4 группы, в каждой из которых у всех было одинаковое мнение. Эти группы были примерно одинаковы для разных экспериментов, из чего делался вывод, что рассматриваемое сообщество людей делится на 4 группы (эти сообщества были неявными, т.е. их члены напрямую не декларировали свою принадлежность группе). Это разделение объяснялось социальными факторами.

Тот же метод был протестирован нами на нескольких графах, обладающих структурой сообществ. Графы генерировались алгоритмом, описанным в статье [12]. Рассматривались графы разных размеров, разной плотности и с более и менее выраженными сообществами. Степень «выраженности» сообществ отражала величина μ – параметр смешивания (mixing parameter), который для одной вершины выражает долю соседних вершин,

связанных с другими соседями этой вершины. Таким образом, чем больше эта величина, тем более ярко выражены группы вершин. Как оказалось, поведение агентов сильно зависело от этой величины. При маленьких μ (т.е. слабо выраженных сообществах) графики зависимости средней доли успешных игр и количества слов в системе почти не отличались от графиков для полных графов. Средняя доля успешных игр постепенно увеличивалась и становилась равной 1, а количество слов постепенно снижалось до 1.

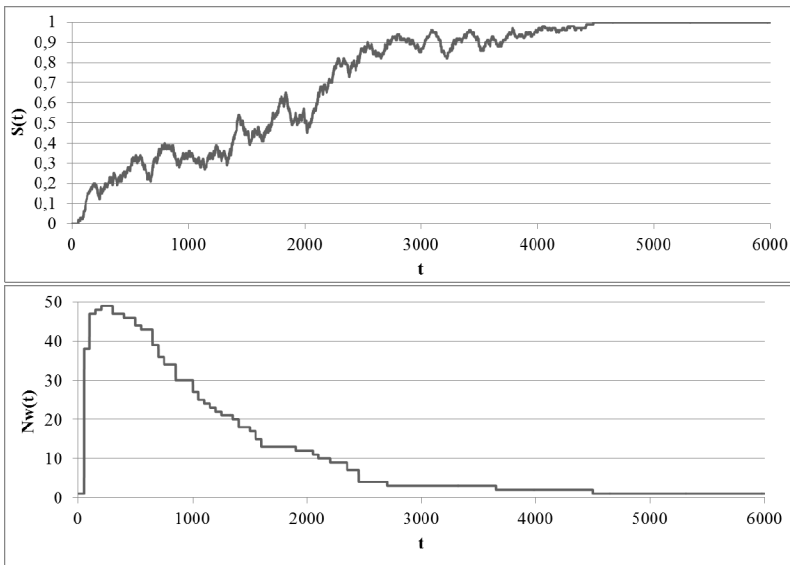


Рис. 3. Средняя доля успешных игр и количество слов для графа со слабо выраженными сообществами ($\mu = 0,2$)

Если граф был с более выраженной структурой сообществ, то количество слов постепенно уменьшалось и выходило на постоянную величину, равную количеству сообществ. (При этом средняя доля успешных игр была достаточно высокой, но меньше 1). В таком метастабильном состоянии система прибывала

достаточно долгое время. Далее количество слов уменьшалось до одного.

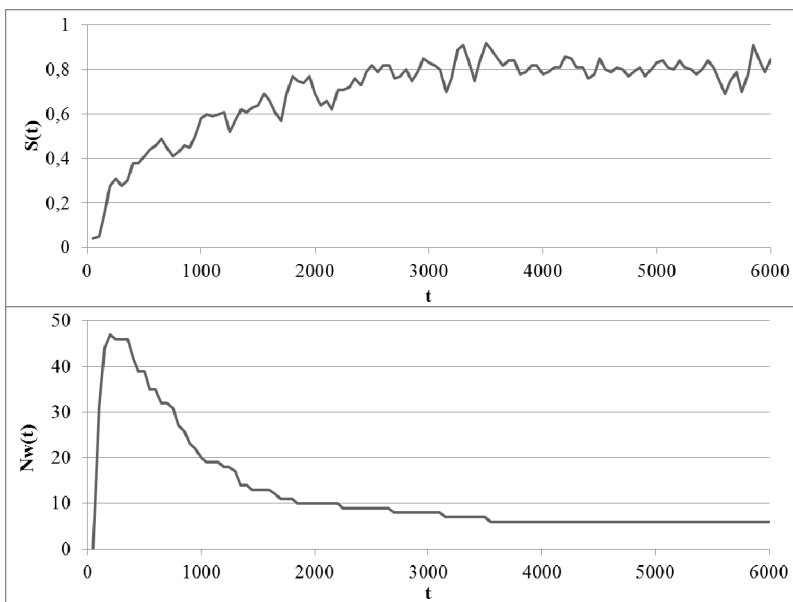


Рис. 4. Средняя доля успешных игр и количество слов для графа с хорошо выраженными сообществами ($\mu = 0,1$)

Таким образом, проследив какие агенты какого мнения придерживаются во время «плато», можно было определить разбиение на сообщества.

Нами были определены сообщества для разных графов и сравнены с эталонными сообществами (по которым генерировался граф) и сообществами, полученными другими методами.

Рассматривались следующие методы:

1. Распространение меток (label propagation) [18].
2. Использование собственных векторов матриц (leading eigenvector) [15].
3. Многоуровневая оптимизация модулярности (multi-level modularity optimization) [5].

4. Infomap [20].
5. Случайное блуждание (random walks) [17].
6. Метод спинового зеркала и имитации отжига (spin-glass) [19].
7. Посредничество ребер (edge betweenness) [16].
8. Жадная оптимизация модулярности (fast greedy modularity optimization) [7].

«Похожесть» разных разбиений графа на сообщества определялось взаимной информацией разбиений I [13].

Нами было исследовано множество графов, сгенерированных с разными параметрами: числом вершин (от 10 до 1000), средней степенью вершин (от 2 до 50) и степенью смешивания (от 0,01 до 0,4). Как оказалось, среди этих трех параметров только степень смешивания оказывает существенное влияние на работу методов определения сообществ. Поэтому в следующих таблицах приводятся результаты, усредненные для нескольких графов и сгруппированные по степени смешивания.

В таблице 1 приведена взаимная информация исходных и полученных разбиений для неориентированных графов нескольких графов.

Таблица 1. Взаимная информация разбиения для неориентированных графов

	$\mu = 0,01$	$\mu = 0,1$	$\mu = 0,2$	$\mu = 0,4$
Игра в наименования	1,00	0,81	0,69	0,16
Посредничество ребер	0,99	1,00	1,00	0,97
Жадная оптимизация модулярности	0,96	0,97	1,00	0,81
Infomap	1,00	1,00	1,00	1,00
Распространение меток	1,00	1,00	1,00	0,98
Использование собственных векторов матриц	0,87	0,85	0,39	0,70
Многоуровневая оптимизация модулярности	0,98	1,00	1,00	1,00
Метод спинового зеркала и имитации отжига	0,99	1,00	1,00	0,91
Случайное блуждание	1,00	1,00	1,00	1,00

Таблица 2. Взаимная информация разбиения для ориентированных графов

	$\mu = 0,01$	$\mu = 0,1$	$\mu = 0,2$	$\mu = 0,4$
Игра в наименования	1,00	0,86	0,59	0,11
Посредничество ребер	1,00	1,00	1,00	0,33
Infomar	1,00	1,00	1,00	1,00
Распространение меток	1,00	1,00	1,00	0,58
Многоуровневая оптимизация модулярности	0,93			
Метод спинового зеркала и имитации отжига	0,92	0,92	0,92	0,92
Случайное блуждание	1,00	1,00	1,00	0,92

Также определялась модулярность разбиения Q [16] – величина, показывающая, насколько при данном разбиении плотность внутригрупповых связей больше плотности межгрупповых связей. Для некоторых графов модулярность указана в таблице 3.

Таблица 3. Модулярность (Q) разбиений для неориентированных графов

	$\mu = 0,01$	$\mu = 0,1$	$\mu = 0,2$	$\mu = 0,4$
Игра в наименования	0,79	0,71	0,40	0,23
Эталонное разбиение	0,79	0,82	0,60	0,09
Посредничество ребер	0,79	0,82	0,60	0,56
Жадная оптимизация модулярности	0,79	0,82	0,60	0,56
Infomar	0,79	0,82	0,60	0,53
Распространение меток	0,79	0,82	0,60	0,56
Использование собственных векторов матриц	0,79	0,71	0,36	0,56
Многоуровневая оптимизация модулярности		0,82	0,60	0,47
Метод спинового зеркала и имитации отжига	0,79	0,82	0,60	0,55
Случайное блуждание	0,79	0,82	0,60	1,00

Как оказалось, ситуация для ориентированных и неориентированных графов не отличается, и при маленьких μ метод работает хорошо. Более подробно зависимость качества разбиений от μ показана на рис. 5. На нем изображен график зависимости качества разбиения (взаимная информация и модулярность) для графов с разным μ (другие параметры были одинаковыми: средняя степень вершин $k = 10$, число вершин $n = 100$). Для каждого графа эксперимент проводился 10 раз, затем вычислялся средний результат (т.е. среднее разбиение).

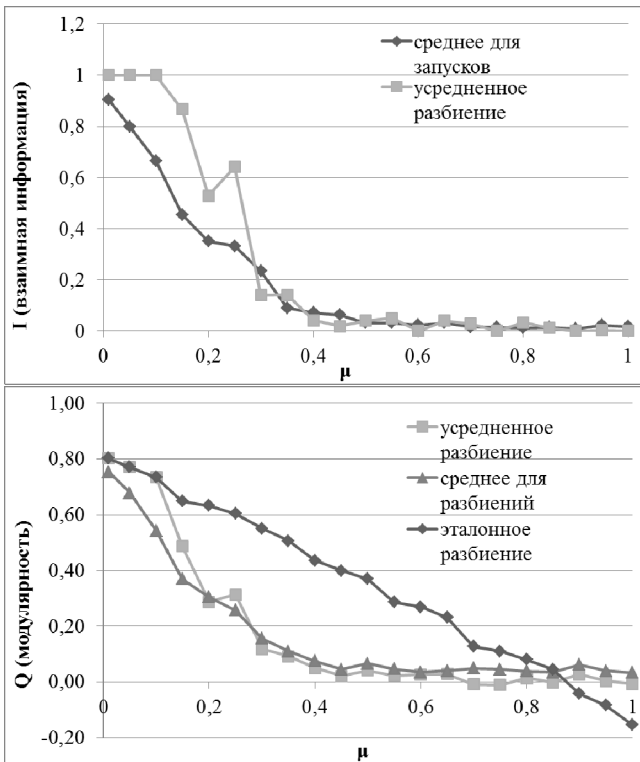


Рис. 5. Сравнение качества разбиений, полученных с помощью игры в наименования для графов с разными μ

На рисунке показаны графики для этих средних разбиений и для среднего, вычисленного для каждого разбиения. Как видно, среднее разбиение оказывается ближе к эталонному, чем разбиения, полученные в отдельном эксперименте. Это связано с тем, что разбиения, полученные с помощью игры в наименования, нестабильны, так как каждый агент легко меняет свое мнение на мнение соседа. Поэтому для более точного результата можно усреднять сообщества за несколько отдельных экспериментов (или наблюдать за динамикой мнений в одном эксперименте).

Основной проблемой при использовании языковых игр для выделения сообществ является определение количества игр, которые необходимо провести. С одной стороны, если взять это число (назовем его t_{opt}) слишком маленьким (см. рис. 6, т. А), то может оказаться, что число слов (N_w) слишком большое, агенты еще не успели их согласовать. С другой стороны, если эта величина слишком большая, то к этому времени агенты уже могут выработать общий язык, то есть окажется, что число слов $N_w = 1$, т.е. сообществ нет (см. рис. 6, т. С). Точно определить это число невозможно, так как игра в наименования допускает некую случайность, и, кроме того, как было сказано выше, не является абсолютно точным методом.

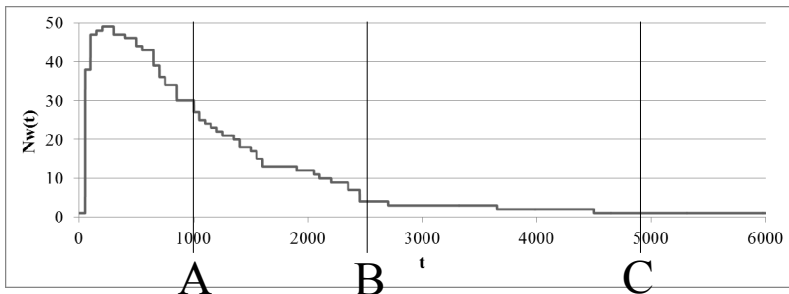


Рис. 6. Пример неудачного (т. А, т. С) и удачного (т. В) определения времени окончания эксперимента

Чтобы примерно определить t_{opt} , нужно запустить серию игр и проанализировать графики $S(t)$ и $N_w(t)$. По этим графикам будет видно, прежде всего, имеет ли граф ярко выраженные сообщества или их границы несколько «размыты» (см. рис. 3 и 4). Если сообщества выражены очень ярко, то график $N_w(t)$ будет иметь длинное «плато» (на котором $N_w > 1$), тогда точку t_{opt} следует взять, лежащей на этом плато.

На рис. 7 показаны графики $S(t)$, $N_w(t)$ и Q (посчитанной для сообществ, выделенных в каждый момент времени) для графа с четко выделенными сообществами ($\mu = 0,01$). По графику Q видно, что действительно, во время «плато» сообщества определяются наиболее точно.

На рис. 8 изображены аналогичные графики для графа с $\mu = 0,1$, а на рис. 9 – для $\mu = 0,2$.

Как видно по рисункам, чем больше μ , тем менее стабилен процесс: графики количества успешных игр и числа слов при больших μ становятся менее гладкими. А модулярность при наибольшем μ принимает гораздо меньшие значения, чем при меньших μ (0,2 против 0,7). Кроме того, при больших μ процесс возникновения общего языка (т.е. начало использования всеми агентами только одного слова) заканчивается быстрее. Таким образом, использование языковых игр в графах с неявно выраженными сообществами становится затруднительным.

Нами был исследован [3] социальный граф, полученный на основе анализа комментариев к записям блогов площадки livejournal.com за несколько месяцев. Вершинами графа были зарегистрированные пользователи и сообщества (рассматривался только русскоязычный сегмент), а связь ставилась, если один из пользователей комментировал пост (запись в блоге) другого. Направление связи шло от автора поста к автору комментария и указывало направление распространения информации (подразумевалось, что если пользователь пишет комментарий к посту, то он прочитал этот пост и воспринял информацию из него). Также дугам были присвоены веса, соответствующие количеству комментариев.

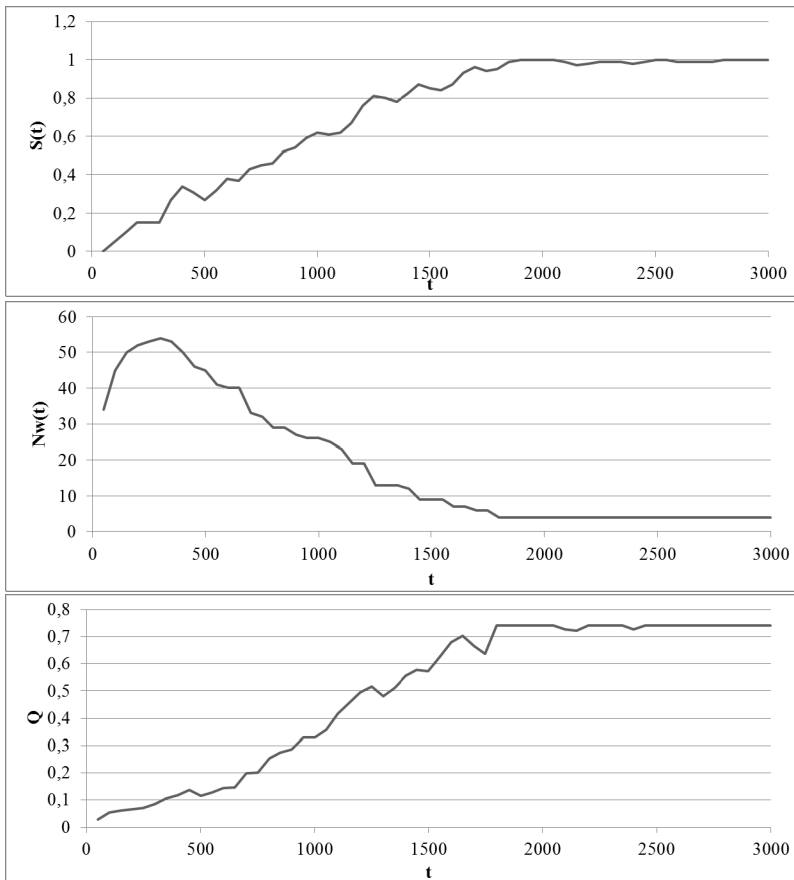


Рис. 7. Средняя доля успешных игр, количество слов и модулярность получившихся разбиений для разных моментов времени для графа с $\mu = 0,01$

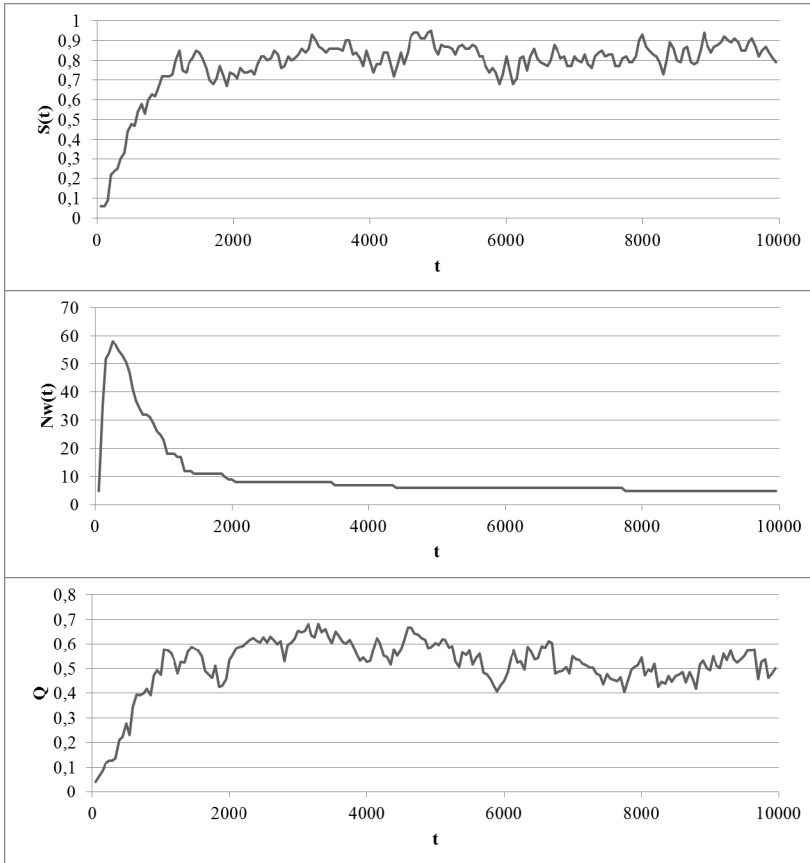


Рис. 8. Средняя доля успешных игр, количество слов и модулярность получившихся разбиений для разных моментов времени для графа с $\mu = 0,1$

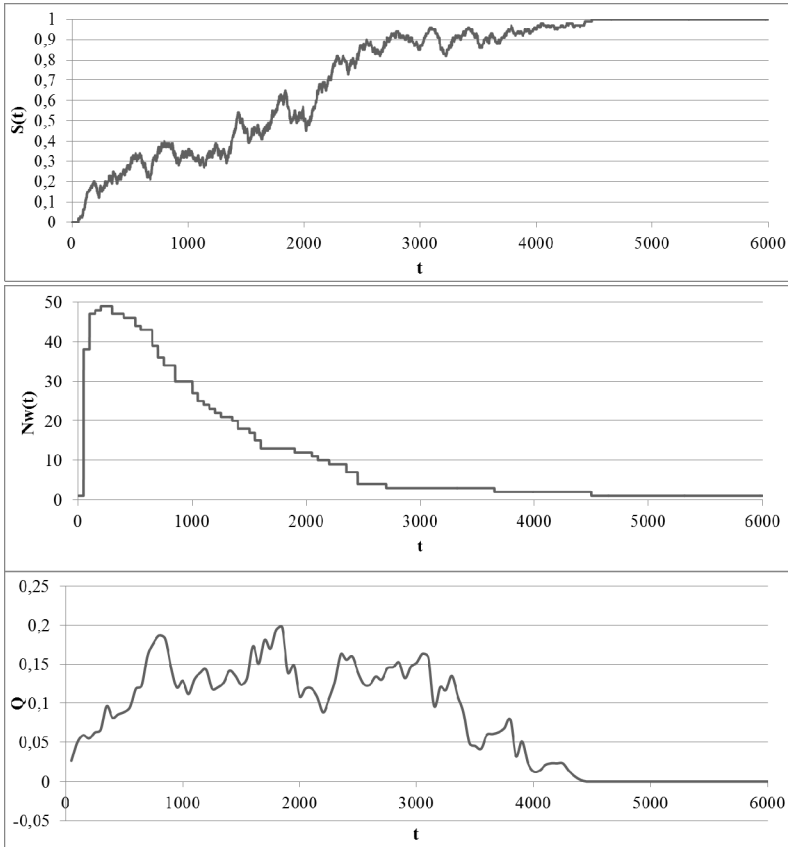


Рис. 9. Средняя доля успешных игр, количество слов и модулярность получившихся разбиений для разных моментов времени для графа с $\mu = 0,2$

Граф состоял из 361 тыс. вершин и около 9 млн. связей. На графе 500 млн. раз запускалась игра в наименования. С помощью игры в наименования удалось выделить около 1500 сообществ (но это не было разбиением, т.е. все вершины входили в сообщества). К сожалению, имеющиеся вычислительные ресурсы не позволили получить разбиения на сообщества другими методами за разумное время. Однако этот факт говорит в пользу использования игры в наименования для поиска сообществ

пусть как не самого точного метода, зато имеющего меньшую вычислительную сложность.

3.2. ВЫЯВЛЕНИЕ ВЛИЯТЕЛЬНЫХ АГЕНТОВ

Другой задачей исследования социальных сетей является выявление влиятельных агентов, т.е. тех агентов, чьи мнения оказываются более значимыми, чем мнения других агентов. Таких агентов можно выявлять и с помощью игры в наименования. Для этого достаточно проследить, какие агенты являются авторами самых популярных мнений в системе.

Обычно в игре в наименования, когда агенту нужно придумать новое слово, он генерирует его как случайную последовательность букв заданного алфавита. В наших экспериментах (на графе пользователей livejournal.com) новые слова соответствовали порядковому номеру агента, что давало возможность проследить, сколько «последователей» (агентов, придерживающегося того же мнения) образуется у каждого агента. Для каждого агента была подсчитана доля его последователей (от всех агентов) и усреднена для 100 экспериментов (по 500 млн. игр в каждом). Полученная величина была названа влиятельностью агентов.

Также было обнаружено, что некоторые агенты все эксперименты заканчивают с одним и тем же мнением. Видимо, эти агенты комментируют только одного пользователя или сообщество (или гораздо чаще, чем других). Большая часть таких групп агентов (а их около 7 тыс.) имеют небольшой размер (до 10 агентов), но есть несколько больших групп в несколько сотен и тысяч агентов. «Лидеров» этих групп (тех, чье мнение все принимают) тоже можно считать влиятельными агентами.

В таблице 4 приведены величины влиятельности некоторых агентов, рассчитанные двумя вышеописанными методами и показатели популярности по меркам livejournal.com (рейтинг и количество друзей).

Таблица 4. Соответствие показателей рассчитанной популярности и популярности в терминах блоговой площадки (псевдонимы пользователей и сообществ скрыты)

Имя	Рейтинг	Друзей	Последователей	Влиятельность
Пользователь №1	3	73017	2563	0,0199
Сообщество №1	15	50000	1044	0,0029
Сообщество №2	84	17842	940	0,0042
Сообщество №3		56903	939	0,0034
Сообщество №4	25	30801	513	0,0051
Сообщество №5	1	58549	486	0,0098
Сообщество №5	188	9911	405	0,0035
Сообщество №6		30716	340	0,0032
Сообщество №7	62	16728	336	0,0019
Пользователь №2	61	9855	315	0,0025
Сообщество №8	172	15282	315	0,0016
Сообщество №9	17	23321	313	0,0023
Сообщество №10		25772	302	0,0031
Пользователь №3		0	300	
Сообщество №11	6	40220	296	0,0033
Сообщество №12		2758	291	0,0019
Сообщество №13		17899	291	0,0012
Сообщество №14	56	34190	277	0,0013
Сообщество №15	237	3326	267	0,0019
Сообщество №16	71	13484	263	0,0015

Как оказалось, самые влиятельные агенты, рассчитанные с помощью игры в наименования, являлись также популярными по меркам блоговой платформы (имеют высокий рейтинг и большое количество друзей). Также выяснилось, что большинство влиятельных агентов – это сообщества, а не отдельные пользователи. Это подтверждает само назначение сообщества – объединение пользователей.

4. Заключение

Итак, мы показали, что игру в наименования можно использовать для исследования социальных графов, в частности для выявления сообществ и определения влиятельных агентов. Также игру в наименования можно использовать и для других исследований в области социальных сетей, в частности для моделирования управления общественным мнением и эффекта масс-медиа [14].

Кроме того, можно усложнить модель, добавив агентам репутацию и учитывая доверие между агентами.

Было проведено сравнение выявления сообществ с помощью игры в наименования с другими популярными методами. Показано, что игра в наименования по точности результата несколько уступает другим методам (особенно для графов со слабо выделенными сообществами). Однако сложность предложенного метода меньше, чем многих других, что позволяет использовать его на больших графах в условиях ограниченных вычислительных ресурсов.

Литература

1. ВИТГЕНШТЕЙН Л. *Философские исследования*. – М.: АСТ, Астрель, 2011. – 352 с.
2. МИКУЛИЧ Л.И., НАУМКИНА Т.С. *Использование языковых игр для моделирования языка в многоагентных системах* // Тринадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2012 (16–20 октября 2012 г., Белгород, Россия): Труды конференции. Т. 3. – Белгород: Изд-во БГТУ, 2012. – С. 34–39.
3. МИКУЛИЧ Л.И., НАУМКИНА Т.С. *Применение языковых игр в многоагентных системах* // Материалы конференции «Управление в технических, эргатических, организационных и сетевых системах» (УТЭОСС-2012). – СПб.: ГНЦ РФ ОАО «Концерн «ЦНИИ «Электроприбор», 2012. – С. 1188–1191.

4. BARONCHELLI A., DALL'ASTA L., BARRAT A. *The role of topology on the dynamics of the Naming Game* // The European Physical Journal: Special Topics. – 2007. – No.143(1) – P. 233–235.
5. BLONDEL V.D., GUILLAUME J.-L. LAMBIOTTE R. et al. *Fast unfolding of communities in large networks* // Journal of Statistical Mechanics: Theory and Experiment. – 2008. – P. 10008.
6. CATTUTO C., LORETO V., PIETRONERO L. *Semiotic dynamics and collaborative tagging* // Proc. National Academy of Sciences of the United States of America. – 2007. – Vol. 104. – P. 1461–1464.
7. CLAUSET A., NEWMAN M., MOORE C. *Finding community structure in very large networks* // Physical review. – 2004. – No. 70 – 066111.
8. DHILLON I.S., GUAN Y., KULIS B. *Weighted Graph Cuts without Eigenvectors: A Multilevel Approach* // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2007. – No. 29(11). – P. 1944–1957.
9. GOSTI G., BATCHELDER W.H. *Naming on a directed graph* // Social Computing, Behavioral-Cultural Modeling. – 2011. – P. 358–365.
10. KE J., MINETT J.W.J., AU C.-P.C. et al. *Self-organization and selection in the emergence of vocabulary* // Complexity. – 2002. – No.7. – P. 1–28.
11. KERNIGAN B., LIN S. *An Efficient Heuristic Procedure for Partitioning Graphs* // The Bell System Technical J. – 1970. – No.49. – P. 291–307.
12. LANCICHINETTI A., FORTUNATO S. *Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities* // Physical Review. – 2009. – No. E80. – 016118.
13. LANCICHINETTI A., FORTUNATO S., KERTÉSZ J. *Detecting the overlapping and hierarchical community structure in complex networks* // New Journal of Physics. – 2009. – No.11(3). – 033015.

14. LU Q., KORNISS G., SZYMANSKI B.K. *The Naming Game in social networks: community formation and consensus engineering* // Journal of Economic Interaction and Coordination. – 2009. – No.4. – P. 221–235.
15. NEWMAN M.E.J. *Finding community structure in networks using the eigenvectors of matrices* // Physical review. – 2006. – No. E70. – 036104.
16. NEWMAN M., GIRVAN M. *Finding and evaluating community structure in networks* // Physical review. –2004. – No. E69. – 026113.
17. PONS P., LATAPY M. *Computing communities in large networks using random walks* // Computer and Information Sciences-ISCIS 2005. – Springer Berlin Heidelberg, 2005. – P. 284–293.
18. RAGHAVAN U., ALBERT R.; KUMARA S. *Near linear time algorithm to detect community structures in large-scale networks* // Physical Review. –2007. – No. E76. – 036106.
19. REICHARDT J., BORNHOLDT S. *Statistical mechanics of community detection* // Physical Review. – 2006. – No.E74. – 016110.
20. ROSVALL M., BERGSTROM C.T. *Maps of random walks on complex networks reveal community structure* // Proc. National Academy of Sciences of the United States of America. – 2008. – Vol. 105. – P. 1118–1123.
21. SCHULZ R., GLOVER A., MILFORD M. *Lingodroids: Studies in spatial cognition and language* // ICRA, 2011 – IEEE 2011. – P. 178–183.
22. STEELS L. *Language games for autonomous robots* // IEEE Intelligent Systems. – 2001. – P.16-22.
23. STEELS L. *Self-organizing vocabularies* // Proceedings of the 5th International Workshop on Artificial Life: Synthesis and Simulation of Living Systems, ALIFE-96. – 1997. – P. 179–184.
24. STEELS L., MCINTYRE A. *Spatially Distributed Naming Games* // Advances in Complex Systems. – 1999. – No.1. – P. 301–323.

25. SHI J., MALIK J. *Normalized Cuts and Image Segmentation* // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2001. – No. 22(8). – P. 888-905.
26. VAN DONGEN S. *Graph Clustering by Flow Stimulation*. – PhD thesis, University of Utrecht, 2000. – 169 p.

LANGUAGE GAMES IN INVESTIGATION OF SOCIAL NETWORKS: FINDING COMMUNITIES AND HIGH-IMPACT AGENTS

Dmitry Gubanov, *Institute of Control Sciences of RAS, Moscow, Cand.Sc. (dmitry.a.g@gmail.com).*

Leonid Mikulich, *Institute of Control Sciences of RAS, Moscow, Cand. Sc. (lmik@ipu.ru).*

Tamara Naumkina, *Institute of Control Sciences of RAS, Moscow (tamara.naumkina@gmail.com).*

Abstract: We suggest using language games in social network studies. This method can be used for finding implicit communities and high-impact agents in social networks. We introduce a Naming game and explain how it is applied in investigation of social networks. Simulation results are presented for several specially designed graphs and for one real-life graph. We also survey alternative approaches and compare them with our method.

Keywords: social networks, communities, opinion dynamics, naming game.

Статья представлена к публикации членом редакционной коллегии Д.А. Новиковым

*Поступила в редакцию 22.07.2014.
Опубликована 30.09.2014.*