

# ИНФОРМАЦИОННЫЕ СООБЩЕСТВА В СОЦИАЛЬНЫХ СЕТЕВЫХ СТРУКТУРАХ. Ч. 3. Прикладные аспекты выявления и анализа сообществ<sup>1</sup>

Л.М. Бойко, Д.А. Губанов, И.В. Петров

**Аннотация.** Представлен обзор эмпирических исследований процессов формирования и выявления информационных сообществ в социальных сетевых структурах. В первой и второй частях обзора кратко изложена концепция информационного сообщества и рассмотрены релевантные математические модели формирования представлений индивидов. Неопределенность относительно механизмов обработки информации индивидами и структуры сетевого взаимодействия делает актуальными задачи идентификации моделей, сбора данных и анализа информационного взаимодействия в социальных сетях. В целях решения задачи идентификации проводятся поведенческие эксперименты и полевые исследования; на практике исследователи анализируют сообщества исходя из доступной им информации, применяя методы анализа на основе структурных свойств сети информационного взаимодействия между индивидами, характеристик индивидов и комбинации структурных и индивидуальных характеристик. В третьей, заключительной части обзора представлены исследования, посвященные вопросам идентификации моделей формирования представлений в сети, рассмотрены прикладные аспекты анализа информационных сообществ в социальных сетях.

**Ключевые слова:** социальные сетевые структуры, информационное сообщество, формирование информационных сообществ, формирование представлений, выявление информационных сообществ.

## ВВЕДЕНИЕ

В первой и второй частях настоящего обзора [1, 2] введена проблематика выявления и исследования информационных сообществ в социальных сетях, представлены математические модели динамики представлений и формирования информационных сообществ в социальных сетях, а также рассмотрены факторы и условия формирования информационных сообществ. Идентификация таких моделей на практике представляет собой непростую задачу: многие параметры являются экзогенными, в то время как в прикладных исследованиях существенная сторона процесса научения остается ненаблюдаемой. Во многих ситуациях лю-

ди не демонстрируют как свои истинные убеждения, так и доступную им для принятия решений информацию или механизмы обработки этой информации.

В процессе социального взаимодействия люди получают информацию от своих оппонентов в неполном объеме — например, в виде информации о результатах действий/активности других людей, но не о том, на основании чего и как эти решения принимались. Это ограничение может быть вызвано многими факторами, например, природой средств социального взаимодействия или высокими затратами на получение и/или передачу информации в полном объеме. Несмотря на возникающие трудности, исследователями проводятся полевые исследования и поведенческие эксперименты, направленные на выявление механизмов обработки информации людьми в реальной жизни; предложено множество методов анализа информационных сообществ на примерах общедоступных данных.

<sup>1</sup> Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научных проектов № 19-17-50225, № 20-07-00190 (Д.А. Губанов) и № 20-31-90133 (И.В. Петров).



Структура третьей части обзора такова: в § 1 представлены работы, посвященные вопросам идентификации моделей формирования представлений в сети, в § 2 рассмотрены прикладные исследования информационных сообществ в социальных сетях.

## 1. ИДЕНТИФИКАЦИЯ МОДЕЛЕЙ ФОРМИРОВАНИЯ ПРЕДСТАВЛЕНИЙ В СЕТИ

Можно привести такую классификацию типов агентов по методам обработки информации (см. первую и вторую части обзора [1, 2]):

- рациональные агенты (например, в рамках концепции байесовской рациональности), которых также можно разделить на «близоруких» (*myopic*) — выбирающих ответ, наилучший в краткосрочной перспективе, и стратегических — выбирающих ответ, оптимальный с точки зрения некоторой теоретико-игровой концепции (например, равновесия Нэша).
- «наивные» (*naive*) — в описание этой категории зачастую включают правило ДеГроота, согласно которому агенты формируют собственные представления, усредняя наблюдаемые мнения других агентов.

Существует ряд ранних работ, посвященных выявлению типов агентов в лабораторных и полевых исследованиях [3–5]. Недостатком этих исследований являются допущения о существовании агентов с различными механизмами научения в рамках одной и той же подвергающейся исследованию группы агентов. В недавней работе [6] была предпринята попытка выявления типов агентов на нескольких наборах реальных данных — показано, что социальные группы состоят из смеси рациональных (байесовских) и наивных (действующих по ДеГрооту) агентов, и что соотношение между типами варьируется для различных наборов данных. Так, в случае серии поведенческих экспериментов с участием жителей 19 индийских деревень выявлено 10 % населения, поведение которых согласуется с концепцией байесовской рациональности, в то время как остальные агенты предпочитают усреднять ответы своих соседей в сети социальных отношений. В том же эксперименте, но уже среди студентов Технологического института Мехико, доля байесовских агентов достигла 50 %. Число экспериментов для каждой из групп составляло 95 и 50 соответственно, число участников — 665 и 350 соответственно.

В этом исследовании авторы идентифицировали четыре паттерна научения, позволяющих различить агентов, обучающихся по правилам Байеса и ДеГроота в модели с неполной информацией и

определили ключевую сетевую характеристику, разделяющую типы обучения, которую они называют кланом (*clan*) — сильно связанной компонентой графа:

1) если клан полностью состоит из агентов, обучающихся по правилу ДеГроота, и если они в некоторый момент времени достигают согласия относительно состояния мира, то во все следующие моменты времени они не изменят своих представлений, даже если они ошибочны;

2) в модели с полной информацией байесовский агент  $i$ , соседи которого принадлежат множеству соседей байесовского агента  $j$ , копирует оценку состояния мира агента  $j$ ;

3) вне зависимости от типа агента  $i$  байесовский агент  $j$  никогда не принимает во внимание его оценку состояния мира (дополнение паттерна 2);

4) даже в случае неполной информации байесовский агент, идентифицирующий простое большинство оценок состояния мира своих соседей, никогда не изменит своей оценки при изменении оценок отдельными соседями.

В работе [7] проведено менее масштабное экспериментальное исследование со схожими результатами: авторы обнаружили, что испытуемые принимают решения, согласующиеся с обновлением по ДеГрооту в 80–98 % случаев, причем ошибки прогноза обусловлены позицией агента в сети. Однако в их исследовании центральное место занимает такое наблюдение: соответствие реального процесса научения агентов «наивному» правилу ДеГроота наблюдается только в сравнительной статике, в то время как динамика достижения консенсуса в лабораторных экспериментах предполагает использование более сложных правил обработки информации. Авторы идентифицируют эти эвристики, тестируют широкий класс других правил научения и модифицируют классическую модель ДеГроота, позволяя агентам корректировать вес своих предшествующих состояний.

Помимо сложности идентификации механизмов обработки информации агентами, зачастую и структура социального взаимодействия также может быть идентифицирована с трудом или вовсе недоступна внешнему наблюдателю, в то время как ее влияние на научение агентов может играть критическую роль [8–10]. С этой точки зрения современные технологии (реализованные, например, в виде социальных онлайн-платформ) кардинально изменили способы взаимодействия и потребления информации. Однако и здесь возникают явления, препятствующие выявлению социальных связей и источников информации. Одним из ключевых аспектов в данной области является политика обработки персональных данных онлайн-платформами [11], когда пользователь вынужден выби-

рать между приватностью и раскрытием большого объема личной информации (такой, как биографические сведения, геолокационные данные или так называемые цифровые следы — история активности в сети интернет) другим пользователям, владельцам платформы или сторонних приложений, принимая решение о доступности сведений о своих социальных связях с другими участниками сети. Также, несмотря на возросшую эффективность в передаче информации, пользователи по-прежнему имеют когнитивные и временные ограничения, что привело к развитию рекомендательных систем и росту интереса к области алгоритмической персонализации. Влияние алгоритмической фильтрации на социальное научение пока мало изучено, однако существует ряд моделей (см., например, статью [12]) в которых показано, что порядок получаемых информационных сообщений может существенно влиять на эффективность научения и достижение консенсуса. Все эти факторы играют решающую роль в задаче идентификации структур информационного взаимодействия и затрудняют наблюдаемость социальных связей.

Таким образом, возникает неопределенность как относительно механизмов обработки информации индивидами, так и относительно структуры взаимодействия, в рамках которой агенты обмениваются полученной информацией, что определяет интерес исследователей к данной области.

## 2. ИССЛЕДОВАНИЕ ИНФОРМАЦИОННЫХ СООБЩЕСТВ В СОЦИАЛЬНЫХ СЕТЯХ

### 2.1. Выявление информационных сообществ

В литературе нет единого мнения относительно формального определения информационных сообществ, и в прикладных исследованиях авторы выбирают достаточно общие определения, разумно отражающие суть явлений, происходящих в сетевых структурах информационного взаимодействия. Существует ряд таких явлений, характеризующих наличие информационных сообществ, которые вместе принято характеризовать как *разногласия (controversy)*:

— *эхо-камера (echo chamber)* — социально-психологический феномен, в рамках которого мнения или убеждения поддерживаются в сообществах, созданных единомышленниками, которые одобряют и укрепляют мнения друг друга;

— *информационный пузырь (filter bubble)* — явление, при котором алгоритмы персонифицированных рекомендательных систем предлагают контент, согласующийся с информацией, поступившей пользователю ранее, тем самым исключая для

пользователя возможность ознакомиться с альтернативной или новой информацией.

Подавляющее большинство работ, направленных на исследование информационных сообществ, связаны с существенными ограничениями. В них рассматриваются представления общества по политическим вопросам, в частности, они сосредоточены на масштабных долгосрочных событиях, таких как выборы. Во многих странах граждане активно участвуют в обсуждениях общественно значимых вопросов в онлайн-социальных сетях (таких как Twitter, Facebook и др.), вследствие чего для анализа становятся доступными огромные тематические массивы данных, содержащие информацию о пользователях и их действиях. Это приводит к тому, что многие работы можно охарактеризовать как тематические исследования (case study), в которых информационные сообщества исследуются на специфическом наборе данных, связанных с отдельным социальным явлением.

В этих исследованиях информационных сообществ, как правило, изучаются процессы распространения информации и их свойства, при этом не идентифицируются механизмы формирования представлений участников сети (рассматриваемые в соответствующих моделях первой и второй частей обзора [1, 2]). Помимо сложности идентификации правил научения, объясняется это еще и тем, что в большей части теоретических моделей сетевая структура задается экзогенно и не зависит от результатов научения: результаты научения не приводят к изменению взаимного влияния участников информационного процесса. Однако эмпирические исследования явлений, характеризующих информационные сообщества, обнаруживают свидетельства взаимосвязи между научением и структурой взаимодействий. Идентификация этих явлений и их формализация в теоретических моделях могли бы существенно сократить разрыв между теорией и практикой, однако в данный момент прикладные исследования ограничены разработкой методов идентификации состояния индивидов (оценкой внутренних представлений на основе наблюдаемой информации) и анализом сравнительной статистики.

Попытки решения задачи идентификации состояний участников информационного взаимодействия зачастую базируются на следующем наблюдении. В общем случае процесс формирования информационного сообщества может быть представлен в виде диффузионного процесса на сетевой структуре (известного как процесс распространения инноваций, идей или информации), в котором присоединение к новому сообществу является аналогом принятия идеи/представлений. Верно и обратное — любой процесс распространения на


**Краткая характеристика наиболее цитируемых работ, посвященных идентификации информационных сообществ**

Работа	Понятие сообщества	Тип графа	Тип показателя	Тип данных	Источник данных
Testing Models of Social Learning on Networks: Evidence from Two Experiments [6]	Набор узлов, которые больше связаны между собой, чем с теми, кто находится вне группы	Искусственно созданная сеть связей между участниками эксперимента	Структурный (clan)	Offline	Лабораторные эксперименты
Ideological Segregation Online and Offline [14]	Сообщество с эквивалентными характеристиками участников	Связи между участниками взаимодействия не учитываются	Индивидуальные характеристики (isolation index)	Online/Offline	Интернет-новости/офлайн-медиа/личное взаимодействие
Quantifying Controversy in Social Media — [15]	Мнения или убеждения поддерживаются в сообществах, созданных единомышленниками, которые укрепляют и одобряют мнения друг друга	Граф диалогов — граф, соответствующий тематическим дискуссиям, где связи между участниками формируются в случае ответов пользователей на сообщения друг друга	Структурный (Random Walk Controversy, Betweenness Centrality Controversy, Embedding Controversy)	Online	Twitter
Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship [16]	Предпочтения к контенту, получаемому пользователями в сети, совпадают с предпочтениями к контенту, который они распространяют	Граф подписчиков — направленная связь между участниками возникает в случае, если один участник следит за обновлениями информации от другого участника	Индивидуальные характеристики (production polarity, consumption polarity)	Online	Twitter
Community Interaction and Conflict on the Web [17]	Участники сообщества преимущественно взаимодействуют с другими членами своего сообщества	Двудольный мультиграф между пользователями и сообществами. Связи возникают в случае общения между пользователями в рамках отдельного сообщества.	Смешанный	Online, временной ряд	Reddit
Quantifying Echo Chamber Effects in Information Spreading over Political Communication Networks [18]	Убеждения укрепляются в результате неоднократного взаимодействия с лицами, разделяющими одни и те же точки зрения	Граф подписчиков — направленная связь между участниками — возникает в случае, если один участник следит за обновлениями информации от другого участника	Смешанный	Online, временной ряд	Twitter
An Empirical Examination of Echo Chambers in US Climate Policy Networks [19]	Сообщество характеризуется двумя признаками: информацией, которая совпадает с уже устоявшимися убеждениями, и кластеризованной структурой взаимодействия	Сеть взаимодействия между экспертами	Смешанный	Online	Опросы
Echo Chambers: Emotional Contagion and Group Polarization on Facebook [20]	Группы единомышленников, убеждения которых приобретают экстраординарные значения	Связи между участниками взаимодействия не учитываются	Индивидуальные характеристики (user sentiment polarization)	Online	Facebook

Работа	Понятие сообщества	Тип графа	Тип показателя	Тип данных	Источник данных
Exposure to Ideologically Diverse News and Opinion on Facebook [21]	Два типа сообществ: — множество участников, подверженных только воздействию информации от единомышленников, — соответствие предлагаемой алгоритмами информации предыстории действий пользователя	Граф связей дружбы между участниками социальной сети	Смешанный (Alignment score)	Online	Facebook
Filter Bubbles, Echo Chambers, and Online News Consumption [22]	—"—"—	Связи между участниками взаимодействия не учитываются	Индивидуальные характеристики (audience-based measure of outlet slant)	Online	История веб-серфинга

сети можно рассматривать с точки зрения формирования сообщества в сетевой структуре, в которой элементы сети сгруппированы по своему состоянию. Одним из примеров таких процессов является процесс распространения информации — так называемый информационный каскад (рис. 1, см. третью страницу обложки).

Естественно предположить, что такие информационные процессы/каскады должны коррелировать с представлениями участвующих в них индивидов, влиять на их представления. Данная аналогия с диффузионными процессами зачастую выступает отправной точкой в исследовании формирования информационных сообществ, в которых авторы применяют методы на основе структурных свойств сети информационного взаимодействия, свойств элементов сети или комбинации структурных и индивидуальных характеристик.

Тем самым возникает задача идентификации двух основных характеристик информационного взаимодействия — структуры связей между участниками и индивидуальных характеристик участников. В таблице приведен список наиболее цитируемых работ, посвященных идентификации информационных сообществ: краткие характеристики используемых данных, предлагаемые показатели и применяемые методы. Как было сказано выше, обе задачи идентификации являются сложными и выбор метода выявления информационных сообществ исследователями во многом обусловлен тем, каким набором реальных данных обладают авторы.

В п. 2.2 приводятся некоторые наиболее общие методы и показатели, позволяющие идентифицировать информационные сообщества.

## 2.2. Методы идентификации информационных сообществ, основанные на свойствах элементов структуры сети

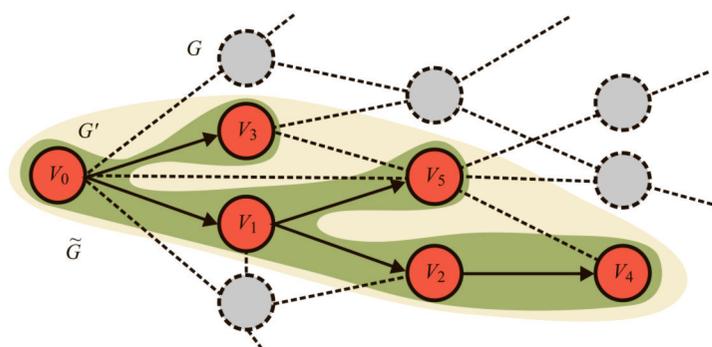
В прикладных исследованиях центральную роль в измерении эффектов, характеризующих наличие информационных сообществ, играет контент. Наиболее существенной характеристикой информационного сообщества является степень соответствия между потребляемым и производимым участниками сети контентом. В связи с этим в работе [16] вводится разделение процесса информационного взаимодействия на процессы потребления и производства информации. Авторы работы определили принадлежность каждого сообщения  $t$  социальной сети одному из двух подклассов  $l(t) = l_n \in \{0, 1\}$  (в исследовании использовались данные социальной сети Twitter;  $P_u$  — информация, опубликованная пользователем на своей личной странице;  $C_u$  — публикации, которые пользователь получает от других авторов; классификация информации проводилась на основании политических взглядов пользователей, где  $l_n = 1$  для представителей консервативных взглядов и 0 — для либеральных). На основе множества всей потребляемой (произведенной) пользователем  $u$  информации  $P_u(C_u)$  определяется степень разнообразия потребляемого/произведенного пользователями контента (production/consumption polarity) как отношение количества информации из одного класса к общему количеству информации, потребленной/произведенной пользователем:

$$p(u) = \frac{\sum_{t \in P_u} l(t)}{|P_u|}, \quad c(u) = \frac{\sum_{t \in C_u} l(t)}{|C_u|}.$$

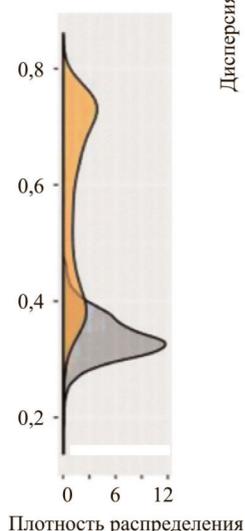
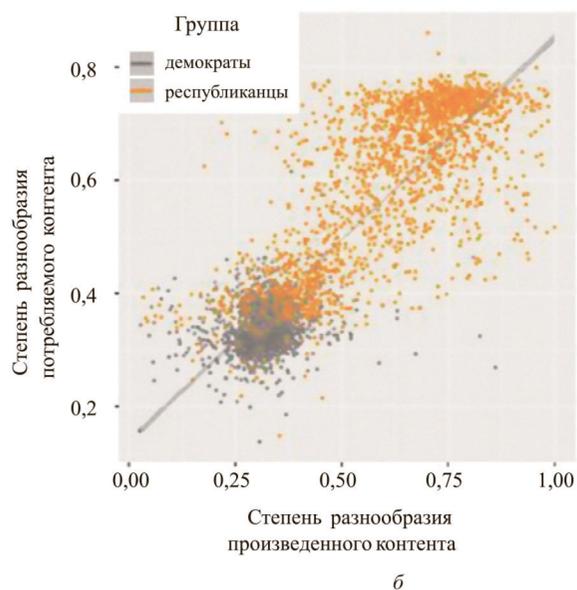
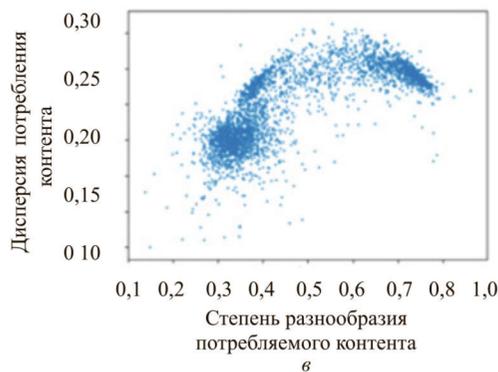
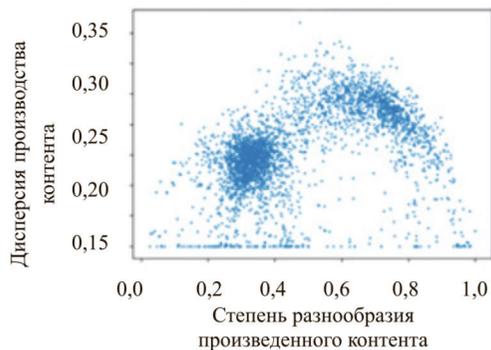
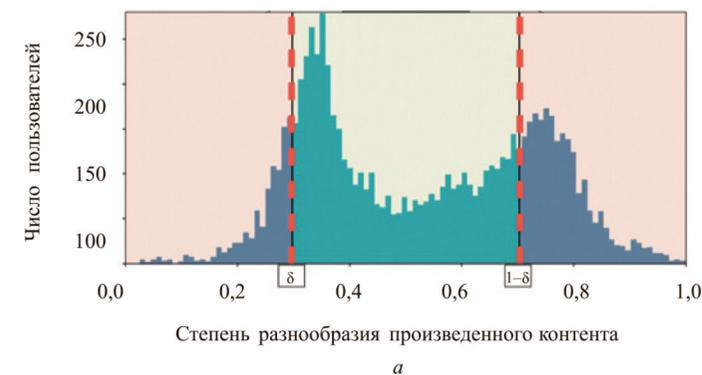
## «ИНФОРМАЦИОННЫЕ СООБЩЕСТВА В СОЦИАЛЬНЫХ СЕТЕВЫХ СТРУКТУРАХ.

### Ч. 3. ПРИКЛАДНЫЕ АСПЕКТЫ ВЫЯВЛЕНИЯ И АНАЛИЗА СООБЩЕСТВ»

(с. 16–24)



**Рис. 1.** Интерпретация процесса формирования сообщества в виде диффузионного процесса на сети [13]. Утолщенными направленными ребрами в графе  $G$  изображен процесс распространения информации, начинающийся с вершины  $v_0$  и охватывающий вершины подграфа  $G'$ . Ненаправленные ребра — связи социального взаимодействия узлов сети (например, связи дружбы). Вместе эти связи порождают подграф друзей  $G'$



**Рис. 2.** Оценка статистических характеристик информационного взаимодействия пользователей online-платформы Twitter по теме законодательного регулирования оборота оружия [16]



Существуют различные вариации этих показателей (где вариативность обусловлена спецификой реальных данных), которые позволяют делать содержательные выводы о процессе распространения информации среди пользователей. В частности, вычисляя статистические характеристики (такие как дисперсия, корреляция или меры близости распределений) полученных оценок разнообразия потребляемого/производимого контента, исследователи демонстрируют наличие информационных сообществ (рис. 2, см. третью страницу обложки).

Рис. 2, а демонстрирует распределение пользователей по степени разнообразия производимого контента: двупиковость данного распределения является одним из признаков наличия информационных сообществ (эхо-камер). График взаимосвязи между разнообразием потребляемого и производимого контента (рис. 2, б) демонстрирует высокую степень кластеризации значений для представителей различных групп пользователей. Графики справа призваны оценить зависимость между разнообразием потребляемого/производимого контента и его дисперсией.

### 2.3. Методы анализа информационных сообществ, основанные на структурных свойствах информационного взаимодействия

Анализируя структурные свойства информационных сообществ, исследователи делают акцент на сопоставление процессов взаимодействия между узлами различных сообществ. В качестве инструмента анализа зачастую выступают свойства случайных процессов на графах или меры центральности, отражающие эффективность узлов в процессе распространения информации.

Показатель *Random Walk Controversy* (RWC [19]) — основанная на случайном блуждании мера, которая определяется следующим образом. Предположим, что уже произведено разбиение графа по некоторому признаку на два подграфа, множества вершин которых не пересекаются. Пусть теперь имеется два случайных блуждания, одно из которых заканчивается в подграфе  $X$ , а другое — в подграфе  $Y$ ; RWC — это разница вероятностей двух событий (первое — оба случайных блуждания начались в том же подграфе, в котором закончились, второе — оба случайных блуждания начались в подграфе, отличном от того, в котором они закончились):

$$RWC = P_{XX}P_{YY} - P_{YX}P_{XY},$$

где  $P_{AB} = P$  (процесс начинается в  $A$  | процесс оканчивается в  $B$ ) — условная вероятность,  $A, B \in \{X, Y\}$ .

Другим методом выявления информационных сообществ, основанном на случайных блужданиях, является персонализированная версия алгоритма PageRank [23], в котором коэффициент затухания (damping factor) изменяется в зависимости от того, к какой группе относится вершина графа, с которой начинается процесс случайного блуждания [17]. В классической версии алгоритма переходы могут осуществляться либо к связанным узлам, либо к любому другому узлу, выбранному равновероятно (что интерпретируется как завершение процесса перехода по ссылкам и начало нового). В персонализированной версии распределение вероятностей на множестве вершин различно для вершин из разных сообществ. Таким образом, данный метод также позволяет оценить степень разногласий между сообществами на основе сравнения вероятностей взаимодействия между участниками различных сообществ [17].

*Мера, основанная на центральности по посредничеству* [15]. Центральность по посредничеству  $bc(e)$  ребра сети  $e$  определяется как

$$bc(e) = \sum_{s \neq t \in V} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}},$$

где  $\sigma_{s,t}$  — общее число кратчайших путей между вершинами  $s$  и  $t$  в графе и  $\sigma_{s,t}(e)$  — число кратчайших путей, проходящих через ребро  $e$ . Авторы работы [15] предлагают анализировать различия в центральностях вершин двух множеств, образующих разбиение графа (в оригинальной работе авторы применяли алгоритм METIS [24]). Идея предлагаемого показателя заключается в сопоставлении центральностей ребер, входящих в разрез графа (т. е. ребер, соединяющих вершины из разных подмножеств вершин графа), и центральностей ребер в остальной части графа. Если получен «хороший» разрез графа, то большинство кратчайших путей из одной части графа в другую будут пролегать через ребра разреза и центральность этих ребер будет иметь более высокие значения по сравнению с центральностью ребер в остальном графе. Сравнивая два распределения центральностей — в разрезе и вне его — например, с помощью KL-дивергенции  $d_{KL}$  и проводя нормировку, получаем выражение для расчета меры разногласий *Betweenness Centrality Controversy* (BCC)

$$BCC = 1 - e^{-d_{KL}}.$$

Помимо указанных методов, применяются классические методы кластеризации, не связанные с диффузионными процессами на сетях или расчетом путей между вершинами: полученные таким образом структурные характеристики исследовате-

ли связывают с индивидуальными характеристиками отдельных узлов графа, тем самым комбинируя продемонстрированные выше методы.

#### 2.4. Методы анализа информационных сообществ, основанные на комбинации структурных и индивидуальных характеристик

Комбинирование индивидуальных характеристик участников и структурных характеристик информационного взаимодействия представляет собой непростую задачу и редко встречается в работах исследователей. Одним из вариантов для анализа является применение методов машинного обучения, позволяющих учесть все имеющиеся у исследователей характеристики процесса информационного взаимодействия (множество всей потребляемой/произведенной информации, содержание информации, структурные характеристики сети и отдельных участников и т. д.) и погрузить их в единое пространство признаков. При этом для выявления информационных сообществ применяются методы классификации.

Перспективным направлением в задаче объединения индивидуальных и структурных характеристик представляется использование различных преобразований исходных данных — вложений (embedding) и, в частном случае, вложений вершин/ребер/графов (node/edge/graph embedding) [25].

В общем случае данная операция состоит в преобразовании исходного пространства признаков в другое пространство, как правило, меньшей размерности. С этой точки зрения все перечисленные выше методы можно понимать как частные случаи таких преобразований. Решение задачи кластеризации классическими методами в новом пространстве [26, 27] и использование, например, меры *Embedding Controversy* (EC)

$$EC = 1 - \frac{d_X + d_Y}{2d_{XY}},$$

где  $d_X(d_Y)$  — среднее расстояние между парами элементов множества  $X(Y)$ , а  $d_{XY}$  — среднее расстояние между парами элементов, принадлежащих различным множествам, приводит к еще одному способу идентификации информационных сообществ [15]. Значение показателя, близкое к единице, характеризует наличие информационных сообществ и высокую степень кластеризации графа; значение показателя, близкое к нулю, говорит об обратном.

Метод вложения графов более сложен, однако позволяет анализировать не отдельные узлы графа, а графы целиком. Метод заключается в использовании ядер графов (graph kernel) — преобразований, позволяющих производить попарное срав-

нение структур между собой, и может применяться как для сравнительного анализа отдельных групп вершин графа [28], так и информационных процессов, протекающих на сетевых структурах [29—31]. Такой подход открывает возможность исследования последовательностей информационных потоков, сравнительного анализа характеристик информационных каскадов (таких как размеры, скорость и др.), протекающих в информационных сообществах, и предсказания этих характеристик.

#### ЗАКЛЮЧЕНИЕ

В работе проведен обзор исследований, связанных с информационными сообществами во взаимодополняющих направлениях: модели формирования информационных сообществ в социальных сетевых структурах, которые имеют как микроэкономические, так и когнитивные и социально-психологические основания, а также методы выявления информационных сообществ и прикладные исследования, посвященные анализу информационных сообществ в социальных сетевых структурах.

В первой и второй частях обзора кратко изложена концепция информационного сообщества и рассмотрены релевантные модели формирования представлений индивидов, которые стремятся устранить неопределенность относительно заданного(-ых) вопроса(-ов), формируя в итоге информационные сообщества. Описаны подходы к моделированию обновления представлений индивидов и влияние различных факторов на достижение истинных представлений, а также формирование в сети различных или одинаковых устойчивых представлений, приводящих к возникновению информационных сообществ.

Эмпирическим исследованиям, изучающим вопросы существования информационных сообществ в реальных социальных сетях, посвящена настоящая часть обзора. Неопределенность относительно механизмов обработки информации индивидами и структуры взаимодействия, а также обилие наборов реальных данных (преимущественно из онлайн-социальных сетей) приводит к разнообразию эмпирических методов идентификации информационных сообществ и специализации исследований на источниках реальных данных. Отметим, что в силу специфики доступных данных рассмотренные методы характеризуют не столько представления самих пользователей социальных сетей, сколько информацию, которая производится и потребляется пользователями. Отсутствие предпосылок о механизмах формирования представлений является существенным недостатком указанных методов и позволяет делать только косвенные выводы как об истинных убеждениях участников информационных взаи-



модействий, так и о динамике формирования информационных сообществ. Переход от анализа взаимодействия на уровне отдельных узлов к анализу взаимодействия структур более высокого порядка, характеризующих развитие информационного процесса, является перспективным для решения задачи идентификации сообществ в структурах информационного взаимодействия. Исследования в данном направлении могут существенно расширить понимание взаимосвязи между информационными процессами и формированием информационных сообществ.

## ЛИТЕРАТУРА

1. Губанов Д.А., Петров И.В. Информационные сообщества в социальных сетевых структурах. Ч. 1. От основного понятия к математическим моделям формирования // Проблемы управления. — 2021. — № 1. — С. 15–23. [Gubanov, D.A., Petrov, I.V. Information communities in social network structures. part 1. from concept to mathematical models // Control Sciences. — 2021. — No. 1. — P. 15–23. (In Russian)]
2. Губанов Д.А., Петров И.В. Информационные сообщества в социальных сетевых структурах. Ч. 2. Математические сетевые модели формирования сообществ // Проблемы управления. — 2021. — № 2. — С. 18–32. [Gubanov, D.A., Petrov, I.V. Information communities in social networks. Part II: Networked models of formation // Control Sciences. — 2021. — No. 2. — P. 18–32. (In Russian)]
3. Hung, A.A., Plott, C.R. Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions // American Economic Review. — 2001. — Vol. 91, no. 5. — P. 1508–1520.
4. Kübler, D., Weizsäcker, G. Limited Depth of Reasoning and Failure of Cascade Formation in the Laboratory // The Review of Economic Studies. — 2004. — Vol. 71, no. 2. — P. 425–441.
5. Choi, S., Gale, D., Kariv, S. Behavioral Aspects of Learning in Social Networks: An Experimental Study. — Emerald Group Publishing Limited, 2005.
6. Chandrasekhar, A.G., Larreguy, H., Xandri, J.P. Testing Models of Social Learning on Networks: Evidence From Two Experiments // Econometrica. — 2020. — No. 1 (88). — P. 1–32.
7. Grimm, V., Mengel, F. Experiments on Belief Formation in Networks // Journal of the European Economic Association. — 2020. — Vol. 18, no. 1. — P. 49–82.
8. Acemoglu, D., Ozdaglar, A. Opinion Dynamics and Learning in Social Networks // Dynamic Games and Applications. — 2011. — No. 1 (1). — P. 3–49.
9. Golub, B., Sadler, E. Learning in Social Networks. — 2017. — URL: <http://dx.doi.org/10.2139/ssrn.2919146>.
10. Young, J.G., Cantwell, G.T., Newman, M.E.J. Robust Bayesian Inference of Network Structure from Unreliable Data. — arXiv preprint arXiv:2008.03334. — 2020.
11. Lovato, J., Allard, A., Harp, R., Hébert-Dufresne, L. Distributed Consent and Its Impact on Privacy and Observability in Social Networks. — arXiv preprint arXiv:2006.16140. — 2020.
12. Perra, N., Rocha, L.E.C. Modelling Opinion Dynamics in the Age of Algorithmic Personalisation // Scientific Reports. — 2019. — Vol. 9, no. 1. — P. 1–11.
13. Cheng, J., Adamic, L., Dow, P.A., et al. Can Cascades Be Predicted? // Proceedings of the 23rd International Conference on World Wide Web. — 2014. — P. 925–936.
14. Gentzkow, M., Shapiro, J.M. Ideological Segregation Online and Offline // The Quarterly Journal of Economics. — 2011. — Vol. 126, no. 4. — P. 1799–1839.
15. Garimella, K., Morales, G.D.F., Gionis, A., Mathioudakis, M. Quantifying Controversy on Social Media // ACM Transactions on Social Computing. — 2018. — Vol. 1, no. 1. — P. 1–27.
16. Garimella, K., Morales, G.D.F., Gionis, A., Mathioudakis, M. Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship // ArXiv180101665 Cs. — 2018.
17. Kumar, S., Hamilton, W.L., Leskovec, J., Jurafsky, D. Community Interaction and Conflict on the Web // Proceedings of the 2018 World Wide Web Conference. — 2018. — P. 933–943.
18. Cota, W., Ferreira, S.C., Pastor-Satorras, R., Starnini, M. Quantifying Echo Chamber Effects in Information Spreading over Political Communication Networks // EPJ Data Science. — 2019. — Vol. 8, no. 1. — P. 35.
19. Jasny, L., Waggle, J., Fisher, D.R. An Empirical Examination of Echo Chambers in US Climate Policy Networks // Nature Climate Change. — 2015. — Vol. 5, no. 8. — P. 782–786.
20. Del Vicario, M., Vivaldo, G., Bessi, A., et al. Echo Chambers: Emotional Contagion and Group Polarization on Facebook // Scientific Reports. — 2016. — Vol. 6. — P. 37825.
21. Bakshy, E., Messing, S., Adamic, L.A. Exposure to Ideologically Diverse News and Opinion on Facebook // Science. — 2015. — Vol. 348, no. 6239. — P. 1130–1132.
22. Flaxman, S., Goel, S., Rao, J.M. Filter Bubbles, Echo Chambers, and Online News Consumption // Public Opinion Quarterly. — 2016. — Vol. 80, no. S1. — P. 298–320.
23. Page, L., Brin, S., Motwani, R., Winograd, T. The PageRank Citation Ranking: Bringing Order to the Web. — Stanford InfoLab. — 1999.
24. Karypis, G. METIS: Unstructured Graph Partitioning and Sparse Matrix Ordering System // Technical Report. — 1997.
25. Rossi, R.A., Jin, D., Kim, S., et al. On Proximity and Structural Role-Based Embeddings in Networks: Misconceptions, Techniques, and Applications // ACM Transactions on Knowledge Discovery from Data (TKDD). — 2020. — Vol. 14, no. 5. — P. 1–37.
26. Chattopadhyay, S., Ganguly, D. Community Structure Aware Embedding of Nodes in a Network // arXiv preprint arXiv:2006.15313. — 2020.
27. Billings, J.C.W., Hu, M., Lerda, G., et al. Simplex2Vec Embeddings for Community Detection in Simplicial Complexes // arXiv preprint arXiv:1906.09068. — 2019.
28. Rossi, R.A., Jin, D., Kim, S., et al. From Community to Role-Based Graph Embeddings // arXiv preprint arXiv:1908.08572. — 2019.
29. Zhou, X., Zafarani, R. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities // arXiv preprint arXiv:1812.00315. — 2018.
30. Ma, J., Gao, W., Wong, K.F. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. — Association for Computational Linguistics. — 2017.
31. Wu, K., Yang, S., Zhu, K.Q. False Rumors Detection on Sina Weibo by Propagation Structures // 2015 IEEE 31st International Conference on Data Engineering. — IEEE, 2015. — P. 651–662.

Статья представлена к публикации членом редколлегии чл.-корр. РАН Д.А. Новиковым.

Поступила в редакцию 26.08.2020, после доработки 09.11.2020.  
Принята к публикации 24.11.2020.

**Бойко Лилия Михайловна** — мл. науч. сотрудник,  
✉ boiko.lilia@gmail.com,

**Губанов Дмитрий Алексеевич** — канд. техн. наук,  
✉ dmitry.a.g@gmail.com,

**Петров Илья Владимирович** — аспирант,  
✉ zyxzy@protonmail.ch,

Институт проблем управления им. В.А. Трапезникова РАН,  
г. Москва.

## INFORMATION COMMUNITIES IN SOCIAL NETWORKS. PART III: Applied Aspects of Detection and Analysis

L.M. Boiko<sup>1</sup>, D.A. Gubanov<sup>2</sup>, and I.V. Petrov<sup>3</sup>

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

<sup>1</sup>✉ boiko.lilia@gmail.com, <sup>2</sup>✉ dmitry.a.g@gmail.com, <sup>3</sup>✉ zyxzy@protonmail.ch

**Abstract.** This paper overviews the empirical studies of the formation and detection of information communities in social networks. In parts I and II of the survey, we outlined the concept of an information community and considered the relevant mathematical models describing the formation of private beliefs. Model identification, data gathering, and data analysis become highlighted areas of current research due to the uncertainty about social learning mechanisms and networked interaction structure. To solve the identification problem, researchers carry out behavioral experiments and field investigations. In practice, researchers analyze communities on available real-world data, applying methods based on the structural properties of the network of information interactions between agents, the individual characteristics of agents, and a combination of structural and individual characteristics. Part III of the survey presents studies on identifying learning models and discusses some practical aspects of analyzing information communities in social networks.

**Keywords:** social networks, information community, formation of information communities, belief formation, identification of information communities.

**Funding:** This research was supported by the Russian Foundation for Basic Research, projects nos. 19-17-50225, 20-07-00190 (Gubanov D.A.), and 20-31-90133 (Petrov I.V.).



*Новая книга*

**Сидельников Ю.В. Технология создания новаций: теоретические и прикладные аспекты: монография** / Ю.В. Сидельников; Институт проблем управления им. В.А. Трапезникова [и др.], Минобрнауки РФ. — Москва: ИПУ РАН, 2021. — 244 с. — ISBN 978-5-91450-253-6. — Текст: непосредственный.

В научном издании рассмотрена гипероболочка технологии, позволяющая повышать эффективность получения новых знаний от экспертов. Можно ли создать технологию создания новых знаний с помощью экспертов, которая позволит решать любые задачи? Вряд ли. Но можно ли создать некую «размытую» технологию, как гипероболочку, которая в процессе как постановки задачи, так и в ходе подстройки самой технологии, будет адаптироваться специалистами на каждом этапе под нужный вид, тип и класс задач, и тем самым способствовать их решению? Не исключено, что это возможно. Важно подчеркнуть, что ни одна технология или креативный метод не сможет гарантированно привести к новации. Создав предлагаемую здесь гипероболочку технологии, мы сможем организовать процесс создания «подсказок» или новых идей, которые преобразовывались бы в конструктивные приемы решения научной задачи создания новаций, что в конечном итоге повысило бы эффективность мыслительных процессов исследователей с учетом их индивидуального опыта, знаний, личных склонностей и взглядов.

В рамках монографии анализируется сам процесс создания нового с помощью экспертов. Для этой цели выделяем важнейшие процессы рассматриваемой технологии.

- Постановка задачи ЛППР и (или) специальной группой и ее корректировка.
- Генерация идей группой людей или отдельным исследователем.
- Анализ идей группой людей или отдельным исследователем.
- Принятие группового решения и (или) ЛППР.

*Рецензенты:* д-р экон. наук, проф. А.И. Агеев, д-р техн. наук, проф. В.Н. Бурков, д-р геогр. наук, проф. В.Б. Коробов; д-р экон. наук, проф. Р.М. Нижегородцев.