

С. В. Сивуха

МЕТОДЫ АНАЛИЗА РЕЛЯЦИОННЫХ АСПЕКТОВ КОММУНИКАЦИИ

Мотивация. Коммуникацию можно определить как процессы создания, сохранения, изменения и передачи смыслов. Эти процессы основаны на взаимодействиях участников (акторов). Сегодня сложность содержательных моделей взаимодействия превышает возможности традиционных количественных методов, ориентированных на анализ атрибутивных данных. Как следствие, процессы коммуникации редуцируются к дихотомическим отношениям (передатчик–приемник, СМИ–аудитория), что затрудняет эмпирическую проверку теорий. Методология анализа социальных сетей позволяет изучать связи и взаимодействия и потому выглядит особенно перспективной для исследований коммуникативной реальности. Эта статья посвящена обзору возможностей анализа социальных сетей (АСС) для моделирования процессов коммуникации, обсуждению практических приложений этой методологии, а также ее ограничений. Особое внимание уделяется публикациям последнего десятилетия. Формальный понятийный аппарат АСС основан на алгебре матриц и теории графов. Простое изложение этих понятий можно найти, например, в работе [4].

Для того чтобы последующее обсуждение было понятным, введем ряд определений. Графом называется множество вершин и связей между ними. Связи могут быть ненаправленными (ребра) и направленными (дуги). Граф с заданными на нем дугами называется ориентированным, или орграфом. Вершины, соединенные ребром, являются смежными. Степенью вершины называется число ребер, соединенных с ней. Исходящей степенью (степенью исхода) вершины называется число дуг, исходящих из вершины; заходящей степенью (степень захода) – число дуг, входящих в вершину. Последовательность вершин, соединенных ребрами, называется маршрутом. Если все ребра в маршруте различны, он называется цепью. В простой цепи ни одна из вершин и ни одно из ребер не повторяются. Число ребер в маршруте называется его длиной. Длина самой короткой цепи, связывающей две вершины, называется расстоянием между вершинами, а сама цепь называется геодезической. Последовательность вершин, соеди-

ненных дугами, называется направленным маршрутом. Если все дуги направленного маршрута различны, он называется путем. В простом пути ни одна из вершин и ни одна из дуг не повторяются. Число дуг, составляющих путь, называется его длиной. Длину самого короткого пути, связывающего две вершины, называют расстоянием между ними (с учетом направления связей). Орграф, в котором из каждой вершины существует направленный маршрут к любой другой вершине, называется сильно связным (двигаться можно лишь по направлению дуг). Орграф, в котором существует маршрут из каждой вершины к любой другой, называется слабо связным. Плотность графа есть отношение числа наличных связей к потенциально возможному.

К преимуществам АСС можно причислить следующие. Во-первых, все его методы приурочены для работы с реляционными данными – данными об отношениях между акторами. Другие подходы в социальных науках ориентированы на анализ атрибутивных данных. В традиционных атрибутивных методах делается важное предположение о независимости наблюдений. Их нельзя использовать для исследования взаимодействий, поскольку последние порождают взаимное сходство и зависимость. Это означает принципиальную недостаточность традиционных количественных методов для исследования социального, ведь «событие или процесс представляют собой *социальный* феномен именно потому, что индивидуальные действия взаимосвязаны» [31, р. 192], а «...зависимости между наблюдениями предполагают сущностные процессы, поддерживающие социальные события» [31, р. 193].

Это уникальное преимущество АСС можно выразить иначе. Атрибутивные модели атомизируют индивида, рассматривая его либо как рационально действующего и автономного актора («недостаточно социализированный» взгляд), либо как актора, жестко следующего усвоенным нормам и ценностям («сверхсоциализированный» взгляд). АСС уникален тем, что позволяет операционализировать теоретический принцип социальной «укорененности» акторов [2] – погруженности индивидов в непосредственный социальный контекст, в системы продолжительных социальных отношений (социальные сети). В коммуникативной науке принцип укорененности позволяет избежать редукции к дихотомическим взаимодействиям абстрактных акторов (СМИ–аудитория) и перейти к более реалистичным моделям, в которых множество передатчиков, посредников и получателей информации взаимодействуют на разных уровнях.

Во-вторых, методы АСС позволяют идентифицировать особые целостности – структуры связей, повторяющиеся, неслучайные паттерны отношений. Вместо редукции к произвольным, пусть и важным, аспектам взаи-

модействий, типичной для других методологий, АСС выделяет устойчивые «узлы» связей и отношений на нескольких уровнях: индивидуальном (центральность актора, или его положение в сети), локальном (ближайшая окрестность), групповом (сплоченные подгруппы), сетевом (глобальные структурные паттерны) [12]. В последние годы разработаны модели эволюции социальных сетей, а также модели, связывающие локальные и глобальные структурные свойства. Тем самым понятие структуры приобретает сложность и гибкость, несвойственную другим формализованным концепциям. С позиций современного АСС структуры отношений эмерджентны, они возникают вследствие локальных и глобальных социальных процессов [22].

В-третьих, АСС обладает богатыми средствами графического представления связей. Визуализация не только репрезентирует реальность, она выступает в качестве мощного аналитического средства, обнаруживая структурные свойства и делая невидимое видимым. Часть этих методов является общей для количественного анализа, например, многомерное шкалирование и декомпозиция сингулярной матрицы [16], другие используют аналогии между графами и физическими системами, например, минимизируют потенциальную энергию или оптимизируют напряжения между вершинами графа (подробное обсуждение достоинств и недостатков двух упомянутых подходов можно найти в [13]). Эксплораторные процедуры позволяют отражать в графическом виде полученную ранее информацию об атрибутивных и структурных свойствах акторов и тем самым становятся мощной аналитической техникой [10].

Измерение социальных сетей. Информация об отношениях между акторами может быть получена из разных источников: из данных обычного социологического опроса, социометрического опроса, из материалов наблюдения, официальных документов, архивных данных, дневников, технической регистрации взаимодействия (например, в службах электронной почты). Анализ могут быть подвергнуты данные о наличии связей между акторами, о силе, частоте, знаке, направлении, длительности и временном порядке этих связей, о совместном участии акторов в одних мероприятиях.

По принципу организации акторов выделяют социоцентрические (полные), эгоцентрические и открытые сети. Первые хорошо известны в социологии, психологии и организационных науках и отражают зиммелевскую традицию исследования сообществ, имеющих естественные границы. Примерами служат организации и их подразделения. Эгоцентрические сети (эго- или личные сети) организованы вокруг центрального индивида, или эго. Идея личных сетей получила развитие в работах британских

антропологов Дж. Барнса, Э. Ботт и К. Митчела, которые увидели в них средство изучения естественных и неформальных социальных организованностей. Критериями связей в эго-сети могут быть родственные и дружеские отношения, знакомства, источники информационной или эмоциональной поддержки. В открытых сетях границы отсутствуют, а акторы трудно исчислимы; измерение продолжается до тех пор, пока представляющее интерес сообщество не будет исчерпано, охвачено с разумной полнотой, или у исследователя закончатся ресурсы. В качестве примеров можно назвать финансовую или политическую элиты страны, экономические связи между организациями, интернет-сообщества, сети наркоманов или ВИЧ-инфицированных индивидов.

Различение упомянутых типов сетей имеет ключевое значение для выбора процедур получения информации, оценки ее валидности, а также для выбора уместных методов анализа данных. Тем не менее во многих случаях различия между ними условны, а границы полной сети – проницаемы или неясны [20].

Предположим, мы хотим изучить достаточно большую сеть, например, связи между маркетинговыми исследователями Минска. Первая, позиционная, стратегия – составить полный список маркетинговых агентств и список сотрудников в каждом из них. Проблема границ здесь проявляется в том, что в список агентств могут не попасть небольшие группы или отдельные исследователи, работающие в непрофильных организациях. Напротив, формальный список сотрудников может включать специалистов в других областях знания или обслуживающий персонал. Кроме того, этот способ требует серьезной подготовительной работы по переписи агентств и сотрудников.

Другая стратегия (событийная) ограничивает исследование участниками некоторого «события», например конференции маркетингологов. Хотя критерий отбора ясен и недвусмыслен, речь идет в лучшем случае о смещенной выборке из генеральной совокупности. Третья стратегия основана на изучении эго-сетей. Отбирают некоторое количество акторов, например, участников конференции, и задают вопросы о том, с кем из коллег они поддерживают связи и с кем связаны коллеги, перечисленные на предыдущем шаге. Полную сеть можно реконструировать, собрав информацию о заметной доле эго-сетей входящих в нее индивидов. Четвертая стратегия (расширенный отбор) состоит в том, чтобы задавать вопросы об отношениях с коллегами членам первоначально сконструированной выборки и добавлять в сеть тех, кого называли, скажем, не менее трех специалистов [14]. Наконец, мы можем использовать метод «случайного маршрута»: случайным образом отобрать несколько маркетингологов и опросить их о

профессиональных контактах; из названных ими коллег случайно выбрать следующих респондентов и попросить их назвать свои профессиональные контакты и т. д. Первые две стратегии выстраивают полные сети с произвольными границами. Третий метод связывает личные и полные сети. Две последние стратегии реконструируют открытые сети.

Для построения полных сетей используются различные варианты социометрической процедуры. Респондентам предъявляют список членов организации и просят указать тех, с кем они связаны определенными отношениями. В больших группах используется свободное припоминание контактов, однако ошибки, вызванные забыванием, могут быть довольно заметными [11].

Процедура реконструкции эго-сетей включает два шага – «генератор имен» и «интерпретатор имен» [20]. На первом шаге исследователи строят список участников эго-сети, на втором – получают информацию о связях между участниками, а также о релевантных характеристиках участников. Наибольшие процедурные и методические различия касаются первого шага. Критериями генерирования имен управляет исследователь. Например, респондента просят назвать людей, которые помогли ему или ей найти работу, с кем он или она живет по соседству, вместе работает, связан или связана дружескими отношениями и т. д. Наиболее объемлющий критерий – назвать всех, кого респондент знает, «кого он может узнать в лицо или по имени, кто может узнать его в лицо или по имени, тех, с кем он может вступить в контакт, если это потребуется» [21]. Валидность измерения повышается, если используются несколько разных критериев.

Другой метод перечисления участников личной сети («генератор позиций») основан на измерении их профессионального положения [19]. Респонденты указывают, есть ли среди их знакомых врачи, юристы, строители, полицейские, секретари. Список позиций представляет собой репрезентативную выборку из генеральной совокупности профессий. Предполагается, что разнообразные знакомства обеспечивают индивида важными и уникальными ресурсами. Голландские исследователи [37] разработали для измерения индивидуального социального капитала «генератор ресурсов». Респонденты называют тех знакомых и друзей, кто имеет автомобиль, знает иностранные языки, может дать рекомендацию для желаемого места работы, дать денег в долг и т. д. В отличие от генератора имен эта методика легка в применении и измеряет реальный доступ к ресурсам. Обе процедуры генерируют безличные, но высокоинструментальные сети.

При измерении отношений между акторами с помощью опросных методов следует иметь в виду, что рейтинговые оценки тесноты связей (например, по пятибалльной шкале) более надежны и валидны, чем дихотомии-

ческие (присутствие либо отсутствие связи) оценки или ранжирование акторов [20, р. 11]. Среди причин снижения валидности измерения отношений в полной сети – неверное определение границ и устанавливаемое исследователями ограничение на количество возможных выборов (последняя ошибка подкреплена авторитетом Я. Морено).

Серьезной проблемой для описательного и особенно для статистического анализа социальных сетей является неполнота измерения отношений между акторами. Удовлетворительного решения эта проблема в АСС не получила. Методы оценки случайности отсутствующих данных и их реконструкции, аналогичные статистическим методам для нереляционных данных, не разработаны. В случае, когда отсутствие информации о связях обусловлено отказом индивидов от участия в исследовании, данные обычно воссоздаются по ответам других участников сети [20, 36]. Этот прием позволяет восстановить лишь часть отсутствующих данных.

Методы анализа социальных сетей. Подавляющее большинство методов АСС применимо для полных и (при определенных предположениях) открытых социальных сетей. Ф. Паттисон [24, р. 15] свела методы АСС к шести группам. К первой отнесены различные индексы для всего графа (сети), например, плотность, иерархизация, централизация, связность, средняя достижимость, средняя длина цепи (пути), баланс, транзитивность, перепись диад и триад. Подобно мерам центральной тенденции в классической статистике, большинство показателей в силу обобщающего характера мало информативны. Тем не менее сочетания некоторых индексов дают представление о структуре сети. Так, определенное сочетание транзитивности (которая в статистической физике называется кластеризацией) и средней длины пути при условии, что размер сети велик, а плотность мала, указывает на то, что сеть принадлежит к классу сетей «тесного мира», имеющих ряд интересных свойств [3, 41]. Д. Уоттс получил сети «тесного мира», добавляя к гомогенной и регулярной структуре локальные процессы. Перепись триад – это распределение частот определенных конфигураций связей между тройками акторов. В орграфе насчитывается 16 типов триад, и некоторые из них отражают чрезвычайно важные локальные свойства сети. Например, транзитивная триада соответствует закономерности «друг моего друга – мой друг» (если актор i связан с актором j и i связан с k , то k должен быть также связан с j). Как показал Дж. Дэвис, транзитивность является самой заметной чертой, отличающей реальные социальные сети от случайных связей (см. [35]). В совокупности перепись триад отражает основные (но не все) глобальные свойства сети.

Вторая группа методов включает индексы для акторов и связей: центральность, статус, престиж, достижимость и др. Это самые понятные ха-

рактеристики социальных сетей, поскольку они ближе к атрибутивным способам описания акторов. Простейшие индексы центральности были введены Дж. Морено и дополнены советскими социометристами (Я. Л. Коломинским, И. П. Волковым, Ю. А. Орном). Существенный вклад в типологизацию и классификацию показателей центральности внесли недавние работы С. Боргатти и его соавторов. Они идентифицировали в социальных сетях разные типы потоков и связали их с традиционными и новыми показателями центральности. Так, оптимальное движение ресурсов предполагает самые короткие пути (цепи), однако вирусная инфекция может распространяться по длинным и неоптимальным путям, а денежные банкноты в экономических обменах – двигаться в любом направлении, в том числе попадать к одним и тем же акторам и перемещаться по одним и тем же «дугам». В одной из работ С. Боргатти выделил пять характеристик маршрутов движения ресурсов и свел их к двум основаниям классификации: типу траектории и методу распространения [9]. Как показали С. Боргатти и М. Эверет, все меры центральности «оценивают включенность вершины в структуру цепей в сети. <...> Они оценивают количество или длину маршрутов определенного вида, начинающихся, заканчивающихся или проходящих через вершину» [8, р. 476]. Показатели центральности можно представить в терминах свойств маршрутов (их количества и длины), позиции вершин в маршрутах (радиальное или медиальное) и маршрутов (кратчайшие цепи/пути, простые цепи/пути, сложные цепи/пути). Все показатели можно получить с помощью операций над матрицей цепей. С этим связан другой смысл структурной природы мер центральности. Матрица цепей (например, матрица расстояний между вершинами) отражает сплоченность вершин в графе. В свою очередь, «меры центральности являются показателями общей диадической сплоченности, присущей каждой вершине» [8, р. 477], т. е. вкладу вершин в общую сплоченность.

Третья группа методов по классификации Ф. Паттисон включает средства пространственного представления сети по критерию сходства связей между акторами и методы классификации акторов: многомерное шкалирование, структурная, автоморфная и регулярная эквивалентность. В четвертую группу отнесены методы идентификации сплоченных подгрупп, таких как клики, кланы, клубы, плексы и *LS*-множества. Пятая группа включает блочные матрицы и реляционные модели, например, модели баланса, кластеризуемости и транзитивности, блочные и стохастические блочные модели. Общая цель всех этих подходов – выделение подгрупп, гомогенных по некоторому критерию. Например, члены подгрупп имеют либо сходный характер связей со всеми другими акторами, согласно принятому определению сходства (структурная и регулярная эквивалент-

ность), либо более высокую долю связей с членами внутри данных подгрупп, чем с «внешними» акторами (клики, кланы, клубы, блочные матрицы). Почти все методы, перечисленные в этом абзаце, имеют детерминистскую природу. Шестая группа в классификации Ф. Паттисон включает вероятностные методы – стохастические блочные модели и модели для случайных графов. Эти подходы стали развиваться в социальных науках лишь в последнее десятилетие.

Исчерпывающее обсуждение традиционных методов анализа можно найти в [40]. Изложение наиболее популярных методов доступно на русском языке [1, 5, 6]. По этой причине в этом разделе рассматривается лишь современный и самый перспективный подход – статистический анализ экспоненциальных случайных графов (систематическое изложение основных понятий вероятностного подхода к социальным сетям см. в [26]).

Первые вероятностные методы анализа социальных сетей. Статистическое моделирование социальных сетей, сдерживавшееся жесткими предположениями традиционных статистических моделей, имеет очень короткую историю. Г. Робинс с соавторами [27, р. 174–175] называют пять основных мотивов для развития стохастической методологии в АСС. Во-первых, вероятностные модели учитывают как регулярные влияния в сети, так и случайную изменчивость. Это важно по той причине, что добавление небольшой стохастичности к детерминистским (регулярным) структурам драматически меняет поведение последних [41]. Во-вторых, эти модели позволяют делать выводы о неслучайных конфигурациях в сети и статистическими средствами проверять содержательные гипотезы о социальных процессах. В-третьих, хорошо специфицированные модели помогают дифференцировать похожие эффекты, создаваемые социальными процессами разной природы. В-четвертых, статистические модели незаменимы для изучения сложных структур, для которых адекватных детерминистских методов не разработано, особенно динамичных и множественных сетей. Наконец, статистические модели проверяемым образом связывают локальные процессы в сети с глобальными, т. е. сочетают достоинства микро- и макроанализа.

Первая простая модель для случайных графов была разработана венгерскими математиками П. Эрдошем и А. Реньи в 1959 г. Эти так называемые бернуллиевы графы основывались на нереалистичном предположении о независимости связей (ребер графа) и их равновероятности. Более сложные модели математически обоснованы Дж. Безагом в 1974 г. в работе, посвященной пространственной статистике. Между климатом или урожайностью земли в соседних регионах больше сходства, чем в удаленных регионах, а в традиционной статистике постулируется взаимозависимость

наблюдений. Описать простыми средствами распределение вероятностей взаимозависимых наблюдений не удастся. Это ограничение в полной мере относится и к социальным сетям, акторы в которых взаимосвязаны.

Дж. Безаг использовал теорему Хаммерсли – Клиффорда для построения распределения случайных графов через экспоненциальную функцию нескольких параметров. Эти идеи были применены О. Франком и Д. Страуссом [15] к марковским случайным графам и в упрощенном изложении выглядят следующим образом. Любая наблюдаемая сеть отношений рассматривается как реализация y двумерного случайного множества сетей Y . Теорема Хаммерсли – Клиффорда связывает вероятность реализации случайного графа или орграфа (как элемента случайного множества графов с данным количеством вершин) с постулируемой структурой зависимости, а именно с полными подграфами графа зависимости. Вершинами графа зависимости являются все возможные связи между связями оригинального графа. Например, для графа, содержащего вершины i, j, k, l , соответствующий граф зависимости содержит вершины $y_{ij}, y_{ik}, y_{il}, y_{jk}, y_{jl}, y_{kl}$. Если в модели предполагается, что наличие связи между i и j зависит от наличия связи между k и l , между переменными y_{ij} и y_{kl} в графе зависимости проводится связь. Исследователь постулирует структуру графа зависимости так, чтобы она отражала природу изучаемых социальных процессов. Теорема Хаммерсли – Клиффорда представляет вероятность реализации случайного графа в виде экспоненциальной функции статистик (описательных характеристик) максимальных полных подграфов (клик) графа зависимости:

$$P(Y = y) = \frac{1}{K} \exp\left(\sum_{A \subseteq N_D} \lambda_A z_A(y)\right).$$

В этой формуле N_D – множество вершин графа зависимости D ; A – конфигурация вершин графа зависимости D ; λ_A – параметр модели, принимающий ненулевые значения только в том случае, если A является максимальным полным подграфом графа зависимости; $z_A(y)$ – состоятельная статистика, равная произведению всех связей в конфигурации A (т. е. $\prod_{y_{ij} \in A} y_{ij}$) и показывающая, все ли связи y_{ij} в A принимают значение 1; K – нор-

мализующая величина, обеспечивающая выполнение условия, что вероятность лежит в диапазоне между 0 и 1, и равная сумме экспоненциальных компонентов уравнения при всех значениях y .

Приведенная выше формула описывает распределение плотности вероятностей для широкого семейства экспоненциальных случайных графов, включая графы Бернулли и марковские графы. Ключевой проблемой для этих моделей является обоснование структуры графа зависимости – фор-

мального представления *условной* независимости связей между элементами социальной сети (т. е. независимости, обусловленной состоянием всех других переменных в моделируемой сети). Так, в бернуллиевом графе постулируется отсутствие условных зависимостей между связями графа (или вершинами графа зависимости). Вероятности реализации всех связей в сети равны. Единственным оцениваемым параметром служит плотность сети. Более сложный случай – модель диадической зависимости (т. н. p_1 -модель) устанавливает условную зависимость между связями (i, j) и (j, i) . В дополнение к плотности модель оценивает параметр взаимности. В марковском случайном графе предполагается условная зависимость ребер (или дуг в ориентированном графе), инцидентных одной и той же вершине, например, условно зависимы y_{ij} и y_{ik} или y_{ij} и y_{kj} . Иными словами, условно зависима связь между вершинами, связанными с одной и той же вершиной. Или по-другому: между вершинами существует путь длиной 2.

Важнейший вопрос моделирования социальных сетей касается включения в модель тех или иных параметров. Теорема Хаммерсли – Клиффорда связывает вероятность реализации случайного графа с параметрами, соответствующими кликам графа зависимости, а также подграфами этих кликов, включая отдельные вершины. О. Франк и Д. Страусс предложили в правую часть приведенного выше уравнения включать параметры плотности, звезд 2-го и 3-го порядков* и треугольников (полных триад). Звезды более высоких порядков в модель не включались. Однако и в этом случае модель включает слишком много параметров и часто не идентифицируема. Необходимо уменьшать количество параметров в правой части уравнения за счет установления эквивалентностей между некоторыми членами уравнения. Например, все диады с взаимными связями между акторами понимаются как изоморфные, и им ставится в соответствие один параметр.

Результаты, полученные О. Франком и Д. Страуссом, были адаптированы С. Вассерманом и Ф. Паттисон [39] к анализу социальных сетей. Эти авторы назвали модель экспоненциальных случайных графов p^* -моделью, и это название закрепилось среди исследователей социальных сетей. С. Вассерман и П. Паттисон представили модель как автологистическую регрессию. (Она не является логистической в силу взаимозависимостей между единицами анализа.)

Для простейшей социальной сети с единственным бинарным отношением между акторами p^* -модель в логистической форме имеет вид [39, р. 407]:

* Звездой k -го порядка называется конфигурация, в которой вершина связана дугами или ребрами с k другими вершинами.

$$\varpi_{ij} = \log \left(\frac{P(Y_{ij} = 1 | \mathbf{Y}_{ij}^C)}{P(Y_{ij} = 0 | \mathbf{Y}_{ij}^C)} \right) = \mathbf{Q}' \mathbf{d}(y_{ij}),$$

где ϖ_{ij} – обозначение логита (логарифма отношения вероятностей), $\mathbf{Y}_{ij}^C$ – социоматрица, дополнительная к дуге $i \rightarrow j$, т. е. все связи в орграфе без данной дуги, $P(Y_{ij} = 1 | \mathbf{Y}_{ij}^C)$ – вероятность существования связи $i \rightarrow j$ при условии, что остальные связи таковы, как в $\mathbf{Y}_{ij}^C$, $\mathbf{Q}'$ – транспонированный вектор параметров уравнения логистической регрессии, и $\mathbf{d}(y_{ij})$ – вектор изменений достаточных статистик графа (плотности, взаимности, транзитивности и др.) при изменении связи $i \rightarrow j$ с 1 на 0.

Логарифм отношения шансов связи между акторами рассматривается как зависимая переменная, а структурные свойства социальной сети (например, экспансивность актора, популярность актора, взаимность, баланс, транзитивность) – как независимые. При такой параметризации p^* -модели могут оцениваться с помощью любой программы логистической регрессии, в том числе в программах SPSS, Statistica, SAS, но перед этим социоматрицу следует преобразовать так, чтобы единицами анализа («случаями») были не N акторов, а $N(N-1)/2$ связей между акторами, а также рассчитать статистики графа. Результаты анализа легко понятны всем исследователям, имеющим минимальный опыт работы с логистической регрессией.

Статистическое оценивание параметров в экспоненциальных моделях – непростая задача. Дело в том, что обычные алгоритмы теоретически обоснованного метода максимального правдоподобия неприменимы для разумно больших сетей ($N > 30$) из-за трудностей в расчетах. В качестве альтернативы Дж. Безаг предложил функцию максимального псевдоправдоподобия (см. [38]). Этот подход широко использовался для оценки p^* -моделей в параметризации С. Вассермана, Ф. Паттисон [39] в 1990-е гг. Однако статистические свойства функции псевдоправдоподобия неизвестны, что фактически не позволяет использовать p^* -модель для проверки статистических гипотез о структурных процессах в социальных сетях. В лучшем случае полученные параметры используются как эвристические оценки, а для приблизительной оценки их значимости параметров сравнивают значения функций псевдоправдоподобия для вложенных (иерархических) моделей, отличающихся одним или несколькими параметрами. Известные критерии, такие как хи-квадрат или критерий Уолда, не могут использоваться для проверки статистических гипотез.

Одним из недостатков работы [39] стало увлечение авторов формальным определением параметров, полезных для репрезентации структурных эффектов в социальных сетях. Связи этих параметров с характеристиками гра-

фа зависимости и, следовательно, – с содержательными социальными процессами при этом не всегда были очевидными. Традиционными статистиками ске;bkб количества пятнадцати изоморфных классов триад [40, ch. 14].

В последующих статьях Ф. Паттисон, С. Вассермана и Г. Робинса были предложены параметризации для логистического регрессионного моделирования взвешенных сетей [29] и множественных сетей [27]. В докторских диссертациях Г. Робинса и П. Элиота разработаны методы для статистического моделирования влияния характеристик акторов на образование связей в сетях (модель социального отбора, см. [33]) и влияния социальной сети на приобретение акторами определенных характеристик (модель социального влияния, см. [32]). В этих работах ключевая роль графа зависимости и формулируемых исследователями предположений о структурных характеристиках сетей была более очевидной.

Сообщество социальных исследователей познакомилось с методами статистического анализа социальных сетей по триаде статей о p^* -моделях [27, 29, 39]. Этот ранний подход к моделированию экспоненциальных случайных графов отражен в двух работах А. Н. Чуракова [6, 7]. Простота оценивания этих моделей, представленных в форме автологистической регрессии, породила энтузиазм, который вскоре сменился унынием.

Разочарование в методах статического анализа. Первые эксперименты с оцениванием параметров p^* -моделей и особенно марковских графов выявили серьезные проблемы. Для многих реальных сетей полученные значения параметров уравнения были неправдоподобно большими. То же касалось стандартных ошибок параметров, а это указывало на проблемы как в формулировании моделей, так и в методах оценивания параметров. Введение в уравнение достаточно произвольных параметров (подход, восходящий к [39]), высоко коррелирующих между собой, или исключение из модели параметров с неправдоподобными значениями превратило статистический анализ в набор технических трюков, единственная цель которых состояла в том, чтобы получить модель с «хорошими» формальными характеристиками.

В начале нашего столетия в качестве альтернативы максимальному псевдоправдоподобию было применено моделирование случайных графов методом Монте-Карло для марковских цепей [17, 34, 35]. Метод Монте-Карло для марковских цепей генерирует распределение случайных графов, основываясь на заданном наборе значений параметров. Характеристики полученного распределения сравниваются с характеристиками наблюдаемого графа и итеративно используются для уточнения значений параметров – до тех пор, пока не будет получено оптимальное (максимально правдоподобное) решение. Параметры и стандартные отклонения, оцененные методом макси-

мального правдоподобия, имеют хорошо известные статистические свойства и могут использоваться для статистической проверки гипотез.

Проблема в том, что марковские случайные графы оказались неустойчивыми к небольшим изменениям значений параметров, и нередко получаются вырожденные модели (генерируются пустые или полные графы). М. Хэндкок предложил более мягкое понятие – «близкие к вырожденным» (near degenerate) – для моделей, в которых плотности вероятности непропорционально распределялись в окрестности ограниченного набора значений параметров (генерируются графы с определенными характеристиками). М. Хэндкок показал существование участков стабильности и нестабильности параметров [17]. Т. Снийдерс изучил похожие ситуации, в которых распределения значений параметров принимало форму бимодального [34]. В этих ситуациях модель часто «не сходилась», т. е. при разных начальных значениях параметров их итоговые значения могли значительно варьировать и отличаться от характеристик наблюдаемого графа. Одна из интерпретаций состояла в том, что в регионах нестабильности распределения параметров действовал нестохастический, детерминистский механизм генерирования социальных сетей [38].

В любом случае первой реакцией на «близость к вырожденности» марковских графов стали попытки подобрать такие параметры модели, которые исключили бы эту ситуацию. Например, было обнаружено, что включение в марковские модели звезд, как минимум, 3-го порядка существенно снижали риск вырождения (см. [28]), однако их содержательная интерпретация неочевидна. Проблемы с оцениванием модели мотивировали поиск других параметризаций и технических трюков, например, фиксация в модели количества ребер в наблюдаемом графе [34].

Другая мотивация была связана с разочарованием содержательной интерпретацией параметров марковских графов, предложенных в работах [15, 39], например, пятнадцати типов триад, и нереалистичностью марковских предположений.

Новые модели: содержательные и технические новшества. Переломной в развитии статистической методологии анализа социальных сетей стала работа Ф. Паттисон и Г. Робинса [25], опубликованная в 2002 г., но оцененная по достоинству сообществом исследователей несколько лет спустя. Авторы исходили из того, что граф зависимости случайного графа отражает предположения о локальных процессах, генерирующих распределение наблюдаемых связей и объединение этих локальных конфигураций в глобальные свойства сети, и поэтому структура графа зависимости должна определяться адекватно сложными содержательными теориями. Статья содержала две важных новации. Во-первых, Ф. Паттисон и Г. Ро-

бинс ввели понятие «социального окружения» (social setting), расширяющего понятие «соседства» в марковском графе (т. е. условно зависимы вершины, имеющие общего соседа). В реальных разумно больших социальных сетях акторы могут не знать друг друга, но тем не менее иметь шансы быть связанными друг с другом, поскольку работают в одной организации, связаны похожими интересами, состоят в одной партии. По Паттисон и Робинсу, «социальное окружение» создается пространственно-временной близостью акторов, социокультурной близостью (сходством интересов) или внешними границами (организационными структурами, техническими каналами коммуникаций). Возможные связи между вершинами графа условно зависимы, если они принадлежат одному «социальному окружению». При формулировании модели в расчет берутся только те конфигурации графа зависимости, которые разрешены структурой «социального окружения». Таким образом, этот подход не противоречит предположениям марковских случайных графов. Структура «социального окружения» подобна структуре пересекающихся групп в организации [31].

Вторая новация выходит за рамки марковских предположений. Ф. Паттисон и Г. Робинс [25] вводят понятие частичной условной зависимости (partial conditional dependence). Связи между несмежными ребрами Y_{ij} и Y_{kl} частично условно зависимы, если наблюдается хотя бы одна связь $-Y_{ik}, Y_{il}, Y_{jk}, Y_{jl}$. Частично условно зависимые связи могут быть представлены как противоположные стороны прямоугольника (тетрады вершин). Между двумя любыми вершинами в структуре частичной зависимости существует путь (без учета направления связей) длиной 3 и менее («знакомый знакомого знакомого»). Паттисон и Робинс [25, 31] обсуждают некоторые содержательные интерпретации такого рода зависимостей [31]. Новое, немарковское представление графа зависимости открывает возможности для приложения широкого круга социальных теорий к анализу сетей.

Математические модели для новых структур графа зависимости были разработаны в 2005 – 2006 гг. Они явно мотивированы техническими соображениями – желанием сформулировать модель случайных экспоненциальных графов с низкими шансами вырождения. Тем не менее статья Т. Снийдерса с соавторами [35] начинается с обсуждения транзитивности как важного свойства реальных социальных сетей, без адекватной репрезентации которого невозможно говорить об успешном моделировании. Ситуация осложняется тем, что транзитивность сети, измеряемая количеством транзитивных триад в орграфах или треугольников в графах, может быть результатом трех локальных социальных процессов. Первый состоит в самоорганизации сети, или стремлении акторов к структурно сбалансированным отношениям (дружить с друзьями друзей). Второй процесс состоит в стремлении индивидов

устанавливать связи с популярными акторами, имеющими большое количество связей. Как следствие, сеть приобретает структуру плотного ядра и периферии с редкими связями. Большое количество транзитивных триад или треугольников возникает в ядре в силу того, что все связаны со всеми. Оба процесса носят структурный характер, т. е. связаны с эндогенной динамикой сети. Третий процесс отражает гомофилию, или стремление акторов устанавливать контакты на основании сходства характеристик (пола, возраста, образования, профессии, интересов и ценностей). Гомофилия представляет собой универсальный принцип социальных отношений [22]. Образование транзитивных триад является побочным результатом установления связей по принципу сходства характеристик. Три упомянутых процесса концептуально обосновываются в разных социальных теориях, поэтому исследователи социальных сетей должны иметь адекватные средства репрезентации этих процессов в формальных моделях.

Т. Снийдерс с соавторами [35] предлагают математически элегантные формулировки трех новых параметров для экспоненциальных случайных графов. Каждый параметр заменяет целый набор традиционных параметров, используемых в марковских моделях, что позволяет избежать чрезмерно большого количества членов уравнения. Это можно пояснить на примере первого параметра – знакопеременной звезды порядка k (*alternating k -stars parameter*). Соответствующая статистика содержит информацию о количестве k -звезд. Положительный вес для звезды 2-го порядка уравнивается отрицательным весом для звезды 3-го порядка, и так далее до звезды k -го порядка (этот прием был опробован в моделировании марковских графов. Кроме того, вес конфигураций все более высоких порядков геометрически уменьшается, ограничивая появления сверхпопулярных акторов, доминирующих над всей сетью. Эта идея может быть сформулирована как геометрически убывающее распределение степеней вершин.

Параметр знакопеременных треугольников порядка k позволяет дать довольно тонкое описание тенденции сети к транзитивности. Графически ситуацию можно представить так: треугольники порядка k строятся на одной основе (связи между акторами i и j) и множеством вершин от m_1 до m_k . Чем больше общих контактов у i и j с другими акторами, тем выше вероятность образования связи между ними. Вклад каждого следующего общего партнера геометрически уменьшается. Можно показать, что такое представление транзитивности соответствует описанной выше частичной условной зависимости между связями [30, 35].

Третий параметр – знакопеременные пути длиной 2 – описывает условие транзитивности, давление к транзитивному замыканию. Графически это треугольники с вершин от m_1 до m_k , но без основы. Утверждается, что два пос-

ледних параметра, включенных в модель, позволяют оценить «чистый» эффект транзитивности. Если знакопеременные пути длиной 2 принимают положительное значение, а знакопеременные треугольники порядка k – отрицательное, речь может идти о серьезных причинах для отсутствия связи между акторами i и j , несмотря на структурное давление, например, о взаимной неприязни. Как можно понять, введение знакопеременных путей длиной 2 мотивировано желанием избежать вырождения моделей [30, р. 200].

Другие формулировки тех же параметров предложены Д. Хантером [18]. Исследования показали перспективность новых параметров. Так, для 20 социальных сетей, включенных в базу данных программы UCINET, марковские модели были вырожденными в 8 случаях, а новые модели показали прекрасные показатели пригодности во всех случаях. Статья [35] и более поздние публикации дают примеры интересных содержательных интерпретаций сочетаний параметров, проясняющих законы функционирования социальных сетей. Однако новые модели, как признают эти авторы, не являются панацеей и не полностью исключают риск вырожденности. Смысл ряда параметров не совсем ясен. Это особенно касается знакопеременных путей длиной 2: «необходима дальнейшая работа, чтобы лучше понять их эффект при включении с другими параметрами» [30, р. 200]. Позиция Т. Снийдерса с соавторами такова: «Большая трудность в интерпретации новых статистик по сравнению с традиционными спецификациями кажется нам неизбежной, учитывая сложность эмпирически наблюдаемых социальных сетей» [35, р. 149], а эта сложность обусловлена тем, что «структура сети, наблюдаемая в данный момент <...> есть результат многих сил и механизмов, которые действовали – часто в течение длительного времени – до момента наблюдения» [35, р. 149].

Вполне вероятно, что нынешний энтузиазм сетевых аналитиков сменится периодом разочарования, но в данный момент наблюдается всплеск публикаций и исследований с использованием преимуществ нового подхода к моделированию экспоненциальных случайных графов.

Практические приложения анализа социальных сетей. Подробное и подкрепленное примерами обсуждение возможностей методологии АСС для решения практических задач в социальных науках и науках о коммуникации заслуживает отдельной публикации. Здесь эти задачи могут быть обозначены лишь пунктирно. Разумеется, АСС лишь ограниченно полезен для исследования самых интимных процессов коммуникации, связанных с порождением смыслов. Однако представленная методология обладает уникальными возможностями для изучения и моделирования эффектов коммуникации – организованностей и упорядоченностей, например, паттернов формальных и неформальных связей во внутри- и межорганизационных струк-

турах. АСС позволяет изучать и, следовательно, контролировать эмерджентные процессы в социоцентрических системах. Методология уместна для изучения и оптимизации информационных и коммуникационных структур. Открываются новые возможности для изучения и управления диффузией инноваций, исследования процессов распространения опыта и знаний. Методы АСС перспективны в исследованиях процессов социального влияния и социального отбора. Моделирование открытых эмерджентных сетей позволяет поддерживать коммуникации в просоциальных сообществах. Популярной темой последних лет стало моделирование асоциальных и террористических сетей – здесь решается противоположная задача идентификации и разрушения каналов коммуникации, например, через изоляцию «ключевых игроков». Относительно новая тема связана с моделированием влияния пространственных параметров на топологию сети. Это важно потому, что процессы коммуникации протекают во времени и пространстве, и характеристики последних одновременно облегчают и ограничивают коммуникации. Все больше публикаций посвящено эволюции сетей. Одним из важных приложений стала оценка т. н. скрытых популяций, недоступных изучению традиционными методами. АСС помогает понять локальные и глобальные закономерности коммуникаций в сверхбольших сетях, таких как Всемирная паутина. Сетевая методология полезна для семантических и семиотических исследований текстов. Операциональное наполнение получают красивые метафоры социальных наук, такие как социальный капитал. Приведенный список задач, решаемых с помощью АСС, далеко не полон. Общим знаменателем этих исследований является перенос акцента с атрибутов на связи, с индивидуального на социальное, т. е. коммуникативное.

ЛИТЕРАТУРА

1. Градосельская, Г. В. Сетевые измерения в социологии: учеб. пособие / Г. В. Градосельская. – М.: Издат. дом «Новый учебник», 2004.
2. Грановеттер, М. Экономическое действие и социальная структура: проблема укорененности / М. Грановеттер // Экономическая социология [Электронный ресурс]. – 2002. – Т. 3. – № 3. – С. 42–58. – www.escos.msses.ru.
3. Милграм, С. Эксперимент в социальной психологии / С. Милграм – СПб.: Питер, 2000.
4. Новиков, Ф. А. Дискретная математика для программистов / Ф. А. Новиков; 2-е изд. – СПб.: Питер, 2007.
5. Сивуха, С. В. Анализ социальных сетей / С. В. Сивуха // Социология: Энциклопедия / сост. А. А. Грицанов [и др.]. – Минск: Книжный дом, 2003. – С. 42–45.
6. Чураков, А. Н. Анализ социальных сетей / А. Н. Чураков // Социологические исследования. – 2001. – № 1. – С. 109–121.

7. Чураков, А. Н. Вероятностные модели социальных сетей / А. Н. Чураков // Социологические исследования. – 2001. – № 9. – С. 99–111.
8. Borgatti, S. P. A graph-theoretical perspective on centrality / Stephen P. Borgatti, Martin G. Everett // Social networks. – 2006. – Vol. 28, № 4. – P. 466–484.
9. Borgatti, S. P. Centrality and network flow / Stephen P. Borgatti // Social networks. – 2005. – Vol. 27, № 1. – P. 55–71.
10. Brandes, U. Exploratory network visualization: simultaneous display of actor status and connections / Ulrik Brandes, Jorg Raab, Dorothea Wagner // Journal of social structure [Electronic resource]. – 2001. – Vol. 2, № 4. – www.cmu.edu/joss/ <<http://www.cmu.edu/joss/>>.
11. Brewer, D. D. Forgetting in the recall-based elicitation of personal networks / Devon D. Brewer // Social networks. – 2000. – Vol. 22, № 1. – P. 29–43.
12. Burt, R. S. Models of network structure / Ronald S. Burt // Annual review of sociology. – 1980. – Vol. 6. – P. 79–141.
13. DeJordy, R. Visualizing Proximity Data / Rich DeJordy, Stephen P. Borgatti, Chris Roussin, Daniel S. Halgin // Field Methods. – 2007. – Vol. 19, № 3. – P. 239–263.
14. Dorean, P. Defining and locating cores and boundaries of social networks / Patrick Dorean, Katherine L. Woodard // Social networks. – 1994. – Vol. 16, № 2. – P. 267–293.
15. Frank, O. Markov graphs / Ove Frank, David Strauss // Journal of the American statistical association. – 1986. – Vol. 81, № 395. – P. 832–842.
16. Freeman, L. C. Graphic techniques for exploring social network data / Linton C. Freeman // Models and methods in social network analysis / Ed. by P.J. Carrington, J. Scott, S. Wasserman. Cambridge: Cambridge University Press, 2005. – P. 248–269.
17. Handcock, M. S. Statistical models for social networks: inference and degeneracy / Mark S. Handcock // Dynamic social network modeling and analysis / Ed. by R. Breiger, K. Carley, P. Pattison. – Washington: The National Academies Press, 2003. – P. 229–240.
18. Hunter, D. Curved exponential family models for social networks / David Hunter // Social networks. – 2007. – Vol. 29, № 2. – P. 216–230.
19. Lin, N. The position generator: measurement techniques for investigations of social capital / Nan Lin, Yang-Chih Fu, Ray-May Hsung // Social capital: theory and research / Ed. by N. Lin, K. Cook, R. S. Burt. – New Jersey: Transaction Publishers, 2001. – P. 57–81.
20. Marsden, P. V. Recent developments in network measurement / Peter V. Marsden // Models and methods in social network analysis / Ed. by P. J. Carrington, J. Scott, S. Wasserman. – Cambridge: Cambridge University Press, 2005. – P. 8–30.
21. McCarty, C. Structure in personal networks / Christopher McCarty // Journal of social structure [Electronic resource]. – 2002. – www.library.cmu.edu:7850/JoSS/McCarty.htm <<http://www.library.cmu.edu:7850/JoSS/McCarty.htm>>.
22. McPherson, J. M. Birds of feather: Homophily in social networks / J. M. McPherson, L. Smith-Lovin, J. Cook // Annual Review of Sociology, 2001. – Vol. 27. – P. 415–444.
23. Morris, M. Local rules and global properties: modeling the emergence of network structure / Martina Morris // Dynamic social network modeling and analysis / Ed. by R. Breiger, K. Carley, P. Pattison. – Washington: The National Academies Press, 2003. – P. 174–186.
24. Pattison, P. Algebraic models for social networks / Philippa Pattison. – Cambridge: Cambridge University Press, 1993.
25. Pattison, P. E. Neighbourhood-based models for social networks / Philippa E. Pattison, Garry L. Robins // Sociological methodology, 2002. – Vol. 32, № 1. – P. 301–337.

26. *Pattison, P. E.* Probabilistic network theory / Philippa E. Pattison, Garry L. Robins // Handbook of probability theory with applications / Ed. by T. Rudas – Thousand Oaks [etc.]: Sage, 2008. – P. 291–312.
27. *Pattison, P.* Logit models and logistic regressions for social networks: II. Multivariate relations / Philippa Pattison, Stanley Wasserman // British journal of mathematical and statistical psychology. – 1999. – Vol. 52. – P. 169–193.
28. *Robins, G.* An introduction to exponential random graph (p^*) models for social networks / Garry Robins, Philippa Pattison, Yuval Kalish, Dean Lusher // Social networks. – 2007. – Vol. 29, № 2. – P. 173–191.
29. *Robins, G.* Logit models and logistic regressions for social networks: III. Valued relations / Garry Robins, Stanley Wasserman, Philippa Pattison // Psychometrika. – 1999. – Vol. 64, № 3. – P. 371–394.
30. *Robins, G.* Recent developments in exponential random graph (p^*) models for social networks / Garry Robins, Tom Snijders, Peng Wang, Mark Handcock, Philippa Pattison // Social networks. – 2007. – Vol. 29, № 2. – P. 192–215.
31. *Robins, R.* Interdependencies and social processes: dependence graphs and generalized dependence structures / Garry Robins, Philippa Pattison // Models and methods in social network analysis / Ed. by P. Carrington, J. Scott, S. Wasserman. Cambridge: Cambridge University Press, 2005. – P. 192–214.
32. *Robins, R. L.* Network models for social influence processes / Garry L. Robins, Philippa E. Pattison, Peter Elliot // Psychometrika. – 2001. – Vol. 66, № 1. – P. 161–190.
33. *Robins, R. L.* Network models for social selection processes / Garry L. Robins, Peter Elliot, Philippa E. Pattison // Social networks. – 2001. – Vol. 23, № 1. – P. 1–30.
34. *Snijders, T. A. B.* Markov chain Monte Carlo estimation of exponential random graph models / Tom A. B. Snijders // Journal of social structure [Electronic resource]. – 2002. – Vol. 3, № 2. – www.cmu.edu/joss/content/articles/volume3/Snijders.pdf <<http://www.cmu.edu/joss/content/articles/volume3/Snijders.pdf>>.
35. *Snijders, T.* New specifications for exponential random graph models / Tom Snijders, Philippa Pattison, Garry Robins, Mark Handcock // Sociological methodology. – 2006. – Vol. 36. – P. 99–153.
36. *Stork, D.* Nonrespondents in communication network studies: problems and possibilities / Diana Stork, William D. Richards // Group and organization management. – 1992. – Vol. 2, № 2. – P. 193–209.
37. *Van der Gaag, M.* The recourse generator: Social capital quantifications with concrete items / Martin Van Der Gaag, Tom A. B. Snijders // Social networks. – 2005. – Vol. 27, № 1. – P. 1–29.
38. *Wasserman, S.* An introduction to random graphs, dependence graphs, and p^* / Stanley Wasserman, Garry Robins // Models and methods in social network analysis / Ed. by P. Carrington, J. Scott, S. Wasserman. – Cambridge: Cambridge University Press, 2005. – P. 148–161.
39. *Wasserman, S.* Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p^* / Stanley Wasserman, Philippa Pattison // Psychometrika. – 1996. – Vol. 61, № 3. – P. 401–425.
40. *Wasserman, S.* Social network analysis: methods and applications / Stanley Wasserman, Katherine Faust. – London [etc]: Cambridge University Press, 1994.
41. *Watts, D.* Networks, dynamics, and small-world phenomenon / Duncan Watts // American journal of sociology. – 1999. – Vol. 105, № 2. – P. 493–527.