

**Г.А. Сатаров\***

## **НОВЫЙ ПОДХОД К ЭМПИРИЧЕСКОМУ АНАЛИЗУ МАССОВОГО ПОЛИТИЧЕСКОГО СОЗНАНИЯ**

*Аннотация.* Описывается новый подход к анализу массового (группового) политического сознания, основанный на представлении данных социологических опросов как когнитивных сетей, узлами которых являются вектора ответов на вопросы анкеты, а связями – статистически значимые зависимости между этими векторами. В качестве объектов анализа рассматриваются не частотные характеристики ответов, а свойства сети, ее отдельных фрагментов и узлов. Приводятся примеры применения данного подхода в социологических исследованиях Фонда ИНДЕМ. Продемонстрировано, что подход позволяет получать нетривиальную информацию о состоянии массового (группового) политического сознания, которая не может быть получена в рамках традиционного подхода.

*Ключевые слова:* общественное мнение; массовое политическое сознание; социологические опросы; когнитивные сети.

**G.A. Satarov**

### **A new approach to the empirical analysis of mass political consciousness**

*Abstract.* The article describes a new approach to the analysis of mass (group) political consciousness. The approach represents data of sociological surveys as cognitive networks: vectors of answers to the questionnaire are regarded as nodes of the network, statistically significant relationship between these vectors – as its links. Objects of analysis are not frequencies of answers but properties of the network (or of its separate fragments and components). The article gives examples of the approach from so-

---

\* **Сатаров Георгий Александрович**, кандидат технических наук, президент Фонда развития исследовательских программ «Информатика для демократии», e-mail: moskars@mail.ru

**Satarov Georgii**, Foundation for the development of research programs «Informatics for Democracy», e-mail: moskars@mail.ru

ciological studies of the INDEM Foundation. It shows that the approach allows to obtain a non-trivial information about the state of the mass political consciousness which cannot be obtained by traditional methods.

*Keywords:* public opinion; mass political consciousness; sociological surveys; cognitive networks.

### Постановка проблемы

Науке свойственно разделять непосредственно наблюдаемые явления и скрытые (латентные) сущности, которые обычно становятся объектами теорий, объясняющих эти явления. Если под лист фанеры поднести магнит, то лежащие на нем металлические опилки расположатся определенным образом: это непосредственно наблюдаемое явление. Магнитное поле – латентная сущность и объект теории. Очень важно, что прорыв в этой сфере научной деятельности обеспечивает именно придуманное человеком понятие магнитного поля, а не опилки и фанерка. Существенно также, что теории устанавливают регулярную взаимосвязь между наблюдаемыми явлениями и латентными сущностями. Это позволяет верифицировать вытекающие из теории следствия и тем самым подкреплять теорию.

Все эти общие места теряют свою банальность, когда дело касается эмпирической социологии общественного мнения, опирающейся на результаты анкетных опросов. Эта сфера интересна тем, что здесь явным образом взаимодействуют наблюдаемые явления и латентные сущности. Первые – это результаты анкетных опросов, выраженные в таблицах частот ответов на вопросы. Вторые – рассуждения социологов о состоянии общества, опирающиеся на таблицы или временные ряды частот. Существующая традиция неформально и несистематически обозначает эти две сферы разными терминами. Часто полагают, что частоты ответов на вопросы фиксируют общественное мнение. Когда же социолог начинает теоретизировать, опираясь на результаты опросов, он, как часто пишут, попадает в сферу массового (группового) сознания. Однако такое теоретизирование не опирается на верифицируемые теории взаимосвязи между общественным мнением и массовым сознанием.

Здесь самое время обратиться к классикам. Когда У. Липпман писал свой труд «Общественное мнение» (1922)

[Липпман, 2004], опросы общественного мнения в их современном понимании еще не проводились. Любопытно также, что для Липпмана общественное мнение ближе к тому, что в отечественной традиции называют массовым сознанием, а в англоязычной – общепринятой картиной мира, существующей в сознании людей. Но именно с тех пор под общественным мнением принято понимать также совокупность вербализованных индивидуальных установок. Ведь именно под влиянием идей Липпмана в 1930–1940-х годах зарождалась практика опросов общественного мнения.

У Дж. Цаллера (1992) [Цаллер, 2004] общественное мнение уже неотделимо от социологических опросов и потому не требует определения. Цаллер – один из немногих, кто пытается последовательно разграничивать наблюдаемые явления и латентные сущности и устанавливает взаимосвязь между ними. Он занимается именно политической социологией, которая обладает постоянным свойством: большая часть респондентов не обладает личным опытом в сфере политики (за исключением участия в выборах). Цаллер строит модель порождения **индивидуального** мнения в ходе опроса, основываясь на предположениях о различиях между респондентами. Этому противоречит имплицитная установка, рассматривающая общественное мнение как сумму индивидуальных мнений, что выражается в использовании частот ответа в качестве индикатора общественного мнения. При этом Цаллер помимо индивидуальных различий в осведомленности и пр. упоминает и другие скрытые сущности. Например, он неоднократно без особого пояснения использует понятие «структура общественного мнения», относя к нему неким образом связанные друг с другом убеждения, ценности, установки [Цаллер, 2004, с. 69]. Он метафорически говорит о «кристаллизации» общественного мнения [Цаллер, 2004, с. 77], о «кристаллизованной структуре общественного мнения» [Цаллер, 2004, с. 197]. Существование структурных свойств общественного мнения подразумевается формулировкой второй аксиомы, вводимой автором для его описания [Цаллер, 2004, с. 94]. Ближе к концу книги автор признает, что если его модели порождения общественного мнения чего-то и не хватает, так это представлений о его структуре [Цаллер, 2004, с. 416–417]. При этом Цаллер ссылается почему-то на социальную психологию, что не совсем правомерно.

Скорее всего, следовало бы начинать с лингвистики. Еще Ф. де Соссюр писал: «Значимость любого термина определяется его окружением» [Соссюр, 2006, с. 115]. Значимость, в отличие от значения, обозначает у него то, что позже стали именовать смыслом. Фактически речь идет о сетях терминов с их смысловыми взаимосвязями, открывших огромное русло лингвистических исследований. Эти идеи перешли и в теоретическую социологию. К ним часто обращался Н. Луман, у которого социальность (а значит и язык) конституируется коммуникацией.

Основательность такого подхода подтверждается эмпирическими данными наук о человеке. Тем более он актуален в социологии общественного мнения. Ведь социологический опрос – одна из разновидностей коммуникации, в которой участвуют две стороны: общество и социолог. Социолог хочет что-то узнать про общество, а общество (если оно здорово) хочет что-то узнать про себя, глядясь в результаты опроса как в зеркало. Уже на этом общем уровне обращение к коммуникации приносит свои плоды. В развитых демократических странах осознание несимметричности коммуникации между следователем и подозреваемым привело к тому, что признание больше не считается доказательством вины подозреваемого. В некоторых странах несимметричны отношения между властью и обществом: власть играет роль следователя, общество – подозреваемого. И как следователь может воспользоваться несимметричностью отношений для получения нужных ему показаний, так и власть может заставить социологические службы выдать нужные ей результаты опросов, а общество – увидеть в них нужную ей картину. Здесь кроется вторая причина того, почему речь в статье пойдет именно о политической социологии.

Любопытно, что западная социология не подвергает сомнению ни существование общественного мнения как социального явления, ни стабильность его функционирования. Если у кого-нибудь (например, у П. Бурдые [Бурдые, 2005]) сомнения возникают, то они касаются не общественного мнения в целом, а его неоднородности и несоответствия между сложностью этого явления и примитивностью методов его изучения. О том же писал еще О. Нейрат [Neurath, 1920], критиковавший простое суммирование ответов. Было бы странно, если бы тотальная госпропаганда и запугивание людей не влияли на состояние общественного мнения и не искажали бы мотивы, определяющие генерацию ответов. По-

дозреваю, что это понятно и российским социологам. И если они не обсуждают эту проблему, то скорее всего потому, что их традиционный инструментарий не позволяет доказательно проверить такую гипотезу. В данной статье это сделано.

Во «Введении в системную теорию» Н. Луман пишет [Луман, 2007, с. 126–127], что язык осуществляет функцию структурного сопряжения между индивидуальным сознанием и коммуникацией как универсальным и конституирующим социальность явлением. При этом мы помним, что естественным языкам свойственно многообразие значений и смыслов, приписываемых словам, словосочетаниям, предложениям и законченным текстам. Если бы этим все и ограничивалось, то разнообразие значений, хаотически распределенных по индивидуальным сознаниям, делало бы невозможными плодотворную коммуникацию и кооперацию наших предков при решении общих задач. Однако коммуникация используется не только для обмена информацией. Она способствует редукции сложности в хаосе значений. Иными словами, существенная часть коммуникации «расходуется» на установление общих значений и взаимосвязей между ними. Это приводит к тому, что индивидуальные сознания, не будучи тождественными, становятся достаточно близкими для того, чтобы коммуникация была результативной. Хотя Луман не формулирует это в явном виде, но некоторые его замечания позволяют трактовать редукцию сложности как образованную из медиума значений форм, являющихся смыслами. Можно сказать, что смыслы здесь – это комплексы взаимосвязанных значений<sup>1</sup>.

Поддерживая точку зрения о том, что «проблему сознания» нельзя решить редукционистски, я бы позволил себе еще более дерзкое высказывание: индивидуальное сознание бессмысленно рассматривать вне категории массового (группового) сознания. В связи с этим я буду использовать термин «со-знание», подчеркивая его способность быть предметом совместного пользования в процессе **коммуникации**. Последнее слово подчеркнуто не случайно. Аутопойетическая (самопорождающая) коммуникация подразуме-

---

<sup>1</sup>На самом деле у Лумана несколько иначе: его медиум коммуникации образован смыслами. Но я не цитирую его буквально, а использую его подход. Не претендую на теоретическую чистоту, а скорее считаю важным выстроить теоретическую цепочку – своего рода мост в сферу эмпирических исследований. В употреблении терминов я следую де Соссюру.

вает, что ее участники явно или неявно стремятся осознавать общность их индивидуальных сознаний или принадлежность к общему сознанию. Но я не берусь утверждать, что социологи, спрашивая, в какой мере респонденты доверяют правительству, часто осознают собственное понимание одного из многих значений этого понятия, не говоря уже о том, чтобы задуматься о значении, которое используется респондентами (не важно – всеми или отдельными группами). И проблема опять не в том, что социологи не знакомы с этим разнообразием значений и смыслов. Просто стандартный инструментальный частотного анализа ответов не пригоден для изучения разнообразия стоящих за ними значений и смыслов, в результате чего эта задача остается на откуп свободному, не подверженному критической проверке теоретизированию. Необходимо же сделать эти значения и смыслы объектом строгого анализа, допускающего возможность подтверждения или опровержения. Но чтобы выбраться из этих ограничений, нужен «новый подход» к анализу результатов опросов, который способен различать общественное мнение как непосредственно наблюдаемое явление и массовое сознание (групповое сознание, со-знание) как латентную сущность.

С операциональной точки зрения в качестве общественного мнения предлагается рассматривать совокупность ответов респондентов на группу вопросов (мы понимаем, что речь всегда идет о фрагменте общественного мнения). Каждая совокупность ответов на тот или иной вопрос, с одной стороны, порождена неким утверждением («я – мужчина»; «я считаю, что достаток моей семьи ниже среднего»; «я полностью доверяю президенту страны»; «я считаю, что страна движется не в том направлении»; «одной из главных проблем страны является безработица»), а с другой стороны, задается подмножеством респондентов, выразивших согласие с соответствующим утверждением (как правило, в условиях выбора между возможными вариантами ответа). Традиционный подход приписывает общественному мнению простые характеристики, исходящие из частоты выбора этих ответов.

Возможной моделью массового сознания как результата социологического опроса мы предлагаем считать целостный объект, образованный всеми совокупностями ответов и взаимосвязями между этими совокупностями. В этом случае массовому сознанию будут приписываться сетевые характеристики нового объекта исследования. Если переходить на статистический язык, то узлами

такой сети могут быть бинарные векторы одинаковой длины, равной объему выборки. Единицы соответствуют респондентам, выбравшим конкретный ответ на один из вопросов анкеты. Связи между узлами образованы значениями некоторой меры близости между бинарными векторами, превосходящими заданный порог, который отсекает статистически незначимые значения. Связям могут приписываться значения мер близости<sup>1</sup>. Для подобных объектов более 15 лет используется понятие «когнитивные сети». Его применяют к структурам нейронных сетей, моделям представления знаний и некоторым другим объектам. В одной из отечественных работ [Когнитивные сети, 2011] это понятие применено к сетям, моделирующим структуру произведений искусства. Мой поиск не выявил примеров использования когнитивных сетей для анализа результатов массовых социологических опросов<sup>2</sup>.

Ниже будут рассмотрены три примера применения «нового подхода» к результатам социологических опросов, проводившихся в разное время фондом ИНДЕМ. Задача состоит в том, чтобы продемонстрировать способность такого подхода получать нетривиальные результаты и выявлять, что стоит за ответами респондентов, каково состояние общественного мнения и как можно определять, описывать и объяснять разнообразие значений и смыслов, приписываемых респондентами своим ответам.

## 1. Пример первый. Оценка проблем страны

В течение долгого времени социологи регулярно включали в свои анкеты вопрос: «Какие из внутренних проблем нашего общества беспокоят вас больше всего? (не более трех ответов)». Число возможных ответов составляло около дюжины.

---

<sup>1</sup>В одном из приведенных ниже примеров сеть будет строиться иначе.

<sup>2</sup>Поскольку это направление еще слишком молодо, я не рискнул приводить какое-либо определение когнитивной сети. Есть понятное определение сети как математического объекта, и есть множество эмпирических реализаций этой модели. Когнитивные модели – одна из таких реализаций, отличающихся от сетей, в которых связи означают возможность перемещения по ним потоков электронов, или социальных сетей, в которых происходит циркуляция информации между людьми или группами людей. Это тоже сеть, но она не относится ни к первой, ни ко второй группе [Когнитивные сети, 2011].

Проблемы образуют тематически осмысленную сеть, в которой два узла связаны или независимы друг от друга. Эта сеть с большими или меньшими искажениями порождает общий для респондентов смысловой контекст «проблемы страны», формирующийся под влиянием социальной практики респондентов, их обыденной коммуникации и средств массовой информации. Мы предполагаем, что эта сеть структурирует проблемы, проявляясь в осмыслении респондентами повседневных проблем и влияя на их жизненные планы (эмиграция, смена работы, участие в политическом протесте и т.п.), а также на выбор ответов на вопросы анкеты. Наиболее распространенная консистентная стратегия выбора ответов состоит в том, что респонденты отдают предпочтение проблемам одного класса (типа). Это делает классификацию респондентов сопряженной с классификацией проблем: к одному классу относятся те, кто черпает большинство ответов из одного класса проблем. Однако некоторые респонденты ограничиваются одним ответом и в результате не подлежат классификации. То же самое относится и к тем, кто берет по одному ответу из разных классов – к ним неприменимо понятие «большинство ответов из одного класса». И те и другие образуют класс неклассифицированных респондентов. С точки зрения сформулированной выше процедуры ответы таких респондентов не признаются консистентными. Этот подход описан в работах фонда ИНДЕМ [Сатаров, 2010; Российская коррупция, 2013]. На первом этапе классификация ответов строится с помощью иерархического кластерного анализа, когда в качестве меры близости между бинарными векторами – индикаторами выбора респондентами каждого из ответов – используется коэффициент ассоциации Q Юла [Кендалл, 1973, с. 723].

В 1998 г. ИНДЕМ провел большое исследование массового политического сознания населения, базировавшееся на репрезентативной общероссийской выборке объемом в 2200 респондентов. Анкета включала в том числе вопрос о проблемах страны. Получившаяся классификация проблем представлена в таблице 1. В названиях классов учтены терминология и особенности политического дискурса того периода (демократами называли сторонников реформ, консерваторами – сторонников возврата к советской власти).

Таблица 1

**Классификация вариантов ответа на вопрос  
о проблемах страны, по данным опроса 1998 г.**

<b>1. Проблемы выживания</b>	
1	Нехватка продуктов питания, товаров первой необходимости
2	Рост цен
3	Угроза безработицы
12	Невыплата зарплаты работникам государственных предприятий и учреждений
<b>2. Проблемы, беспокоящие демократов</b>	
7	Обострение национальных конфликтов
11	Угроза диктатуры
4	Рост преступности
6	Ухудшение состояния окружающей среды
<b>3. Проблемы, беспокоящие консерваторов</b>	
5	Кризис морали, культуры, нравственности
8	Уход от идеалов социализма
10	Отсутствие ясных целей, ориентиров развития страны
9	Слабость, беспомощность государственной власти

На основании этой классификации проблем в соответствии с описанным выше алгоритмом была построена классификация респондентов (рис. 1). Как видим, более 40% респондентов отвечают неконсистентно (непоследовательно). Из оставшихся более половины озабочены проблемами выживания.

Любопытно, что эта классификация тесно коррелирует с другими политическими переменными. Например, среди респондентов, выбирающих преимущественно из класса 2, в три раза больше считающих, что «дела в стране идут в правильном направлении», чем среди респондентов, обеспокоенных проблемами из класса 3, и в четыре раза больше, чем среди тех, кто обеспокоен проблемами выживания. Подобные зависимости наблюдаются и с другими переменными.



Рис. 1.

**Частоты распределения респондентов по классам, определяемым вариантами ответа на вопрос о наиболее важных проблемах страны (%). Опрос 1998 г.**

Чтобы убедиться, что данная структура существует сама по себе, а не порождена особенностями метода, воспользуемся другим эвристическим приемом. Рассмотрим ту же самую матрицу – коэффициентов близости между векторами – индикаторами выбора ответов на основе Q Юла – и используем технику, которая несколько десятилетий назад называлась методом корреляционных плеяд (точнее, была одной из ее разновидностей)<sup>1</sup>. Метод предельно прост: зададим некоторое положительное пороговое значение для коэффициента Q. После этого от матрицы корреляций перехо-

<sup>1</sup> Чтобы еще больше отдалить эксперимент от исходного статистического расчета, мы проводим его не на всей выборке, а только на классифицированных респондентах.

дим к матрице инцидентий графа, в котором вершины соответствуют проблемам, а ребра соединяют те пары вершин, которым соответствует пара проблем со значением  $Q$ , большим заданного порогового значения<sup>1</sup>. Получившийся граф (рис. 2) распадается на три связанных подграфа, между которыми нет ребер. Эти подграфы по своему составу полностью совпадают с классами, полученными методом кластерного анализа (табл. 1). Тем самым подтверждается устойчивость классификации.

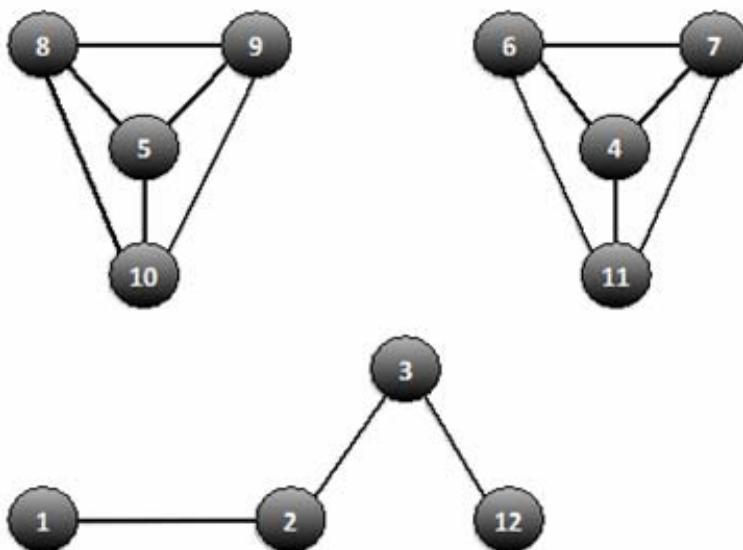


Рис. 2.

**Граф проблем, вершины которого совпадают с проблемами, а ребра образованы связями, превосходящими заданный порог**

Теперь подсчитаем для вершины графа среднее значение коэффициентов  $Q$  между соответствующей проблемой и проблемами, с которыми она связана в построенном графе. Подобные характери-

---

<sup>1</sup>Автор осведомлен о нетривиальности проблемы выбора порога, но формат статьи не позволяет остановиться на ней более подробно.

стики узла сети (или вершины графа) обычно называют *центральностью*. В нашем случае она имеет важный смысл. Коэффициенты, подобные  $Q$  Юла, в принципе обладают общим вероятностным смыслом и в условиях нашей задачи указывают на то, что если произвольный респондент выбрал в качестве ответа проблему под номером  $i$ , то он с высокой вероятностью выберет и проблему под номером  $j$ . Если для двух этих проблем значение  $Q_{(ij)}$  достаточно велико, то выбор одной из проблем скорее всего означает и выбор другой. Это значит, что чем «центральнее» (в терминах нашей задачи) проблема, тем увереннее ее выбор предсказывает выбор других связанных с ней проблем<sup>1</sup>. Можно перефразировать это свойство так: проблемы с высокой центральностью занимают в сети проблем более важное место, чем остальные. В классе «проблем, беспокоящих демократов», к примеру, наибольшей центральностью обладает проблема «рост преступности», а в классе «проблем, беспокоящих консерваторов» – проблема «кризис морали, культуры, нравственности». В своих классах это центральные, «прототипические» проблемы. Прототипичность проблемы в нашем случае задается ее центральностью.

А теперь главное, ради чего затронут этот сюжет. Рассмотрим взаимосвязь между центральностью (прототипичностью) проблем и их «популярностью» – частотой выбора респондентами. Оказывается, никакой зависимости нет. Это ярко демонстрирует диаграмма рассеяния между двумя переменными (рис. 3).

Мораль. Традиционные статистические методы анализа опросных данных сосредоточены на определении частот выбора ответа – в целом по выборке или в отдельных группах респондентов. Частота – это результат подсчета числа респондентов, выбравших один и тот же ответ. Однако такая частота осмысленна в той степени, в какой выбор респондентами одного и того же ответа имеет сходные смысл, причины, последствия. В частности, подсчитывая долю респондентов, выбирающих из списка партию, за которую они намерены голосовать, мы вправе рассчитывать на возможность прогноза результатов выборов. Но наш расчет оправдан, если выборы – это рутинная практика респондентов; если партии

---

<sup>1</sup> Вот еще одна «техническая проблема». Далеко не всегда, как в данном контексте, можно установить функциональную взаимосвязь между критерием и вероятностью, но само существование монотонной зависимости доказуемо. Часто этого достаточно для решения практических задач.

равноправны в политическом и информационном пространствах; если выбор в пользу той или иной партии в результате пропагандистских усилий не рассматривается общественным мнением как нечто неприличное или опасное.

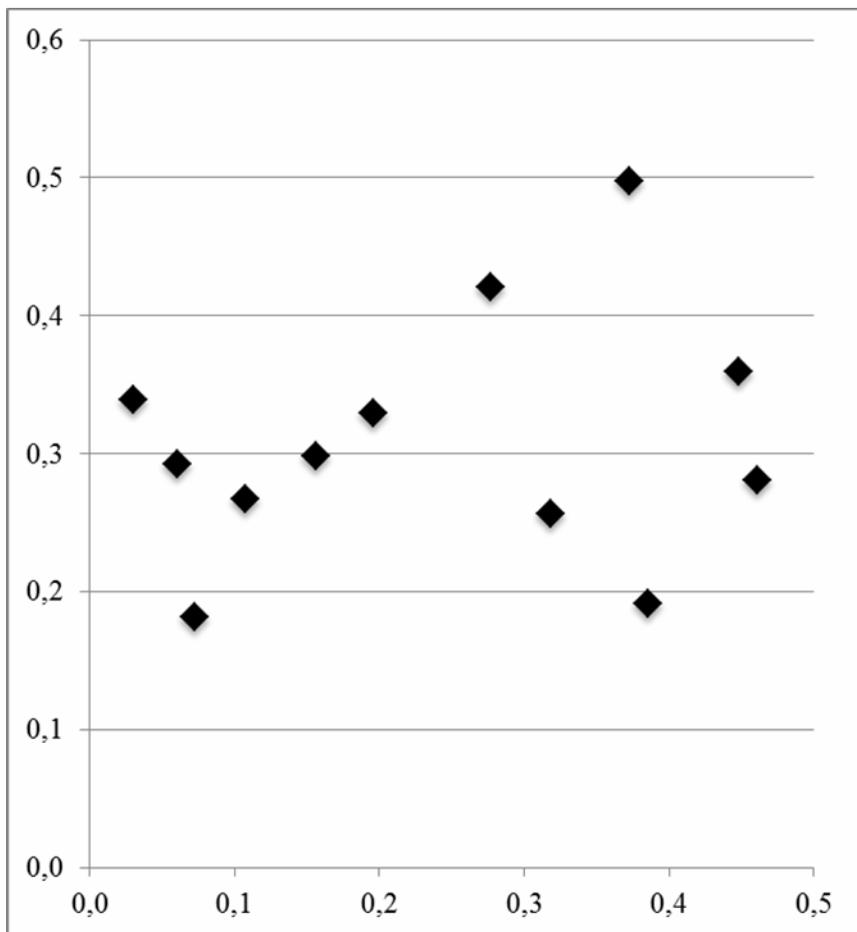


Рис. 3.

**Диаграмма рассеяния частоты выбора проблем (горизонтальная ось) и центральности проблем в сети взаимосвязей (вертикальная ось)**

Все становится гораздо сложнее, когда мы подсчитываем частоту выбора ответов на вопрос, который изучается в данном примере. Рассмотрим частоту выбора проблемы «рост преступности» в целом по выборке и внутри каждого класса респондентов. В целом она равна 40,1%. Внутри класса «проблемы, беспокоящие демократов» она гораздо выше – 92,1%. Но ее выбирают также респонденты из классов «проблемы выживания» (26,2%) и «проблемы, беспокоящие консерваторов» (24,6%). В двух последних случаях респонденты выбирают две проблемы из своего класса и присоединяют к ней одну из чужого. Более чем вероятно, что респонденты из разных классов вкладывают в выбор одной и той же проблемы разный смысл. Вот пример возможной (именно возможной) интерпретации. Респонденты из класса «проблемы выживания» трактуют рост преступности как одну из проблем, мешающих им жить (выживать). Для респондентов из класса «проблемы, беспокоящие демократов» рост преступности – помеха проводимым реформам. А респонденты из класса «проблемы, беспокоящие консерваторов» рассматривают его как свидетельство слабости власти. Еще раз: все перечисленное – не подтвержденные социальные факты, а лишь иллюстрация неоднородности причин выбора одной и той же проблемы. Но когда мы, как обычно, получаем общую по выборке частоту выбора этой проблемы, мы не задумываемся о возможном разнообразии смыслов, обуславливающих такой выбор; мы не принимаем во внимание это разнообразие, лишая эту частоту реального социального содержания и нагружая полученное число неким приписываемым нами, но не получившим обоснования смыслом (причинами, последствиями).

Конечно, люди всегда вкладывают в используемые понятия разные смыслы, руководствуются в своем социальном поведении разными мотивами и преследуют разные цели. Но как это обстоятельство отражается в эмпирических социологических исследованиях – их методах, структуре решаемых задач, способах интерпретации и использования получаемых результатов? За редчайшими исключениями – никак. Предложенный фрагмент исследования – один из вариантов выхода из клубка проблем, порождаемого традицией эмпирической политической социологии.

Далее. Повсеместно используемая социологами частота выбора ответа на вопрос может не только быть неинформативной, но и не исчерпывать всего, что можно узнать с помощью опросов об

изучаемом фрагменте социальной реальности. Можно же, во-первых, увидеть за мнениями (ответами на вопросы анкеты) механизмы их порождения, опирающиеся на то, что мы вправе назвать массовым сознанием; во-вторых, представить массовое сознание (групповое сознание) как сеть, вершины которой – ответы на вопросы анкеты, а связи между вершинами – эмпирически вычисляемые значения взаимосвязей между ответами. Тогда объектом изучения становятся структурные свойства этой сети, а ответы обретают структурные свойства – такие как центральность. Есть веские основания полагать, что эти структурные свойства позволяют описывать состояние массового сознания и объяснять наблюдаемое общественное мнение, выражаемое в том числе в частоте ответов на вопросы. В следующем параграфе этот тезис будет подтвержден результатами другого исследования, в котором в роли объекта статистического анализа выступали не сами ответы на вопросы анкеты, а некоторые производные от них переменные.

## **2. Пример второй. Групповое правосознание**

В недавнем исследовании фонда ИНДЕМ сравнивалось правосознание трех социальных групп: судей, населения (стандартная общероссийская выборка) и предпринимателей [Сатаров, Римский, Благовещенский, 2016 b]. Сравнение осуществлялось по пяти переменным: 1) поддержка независимости судебной власти; 2) поддержка концепции естественного права; 3) законопослушание; 4) поддержка независимости судебных решений; 5) нетерпимость к нарушению законов. Каждой из переменных соответствовал вопрос, содержащий 8–10 суждений, с которыми респондент оценивал степень своего согласия. На основании этих оценок респонденту приписывалось число – координата на шкале соответствующей переменной. На рис. 4 приведены средние значения всех респондентов в каждой из трех групп по всем пяти переменным. Из диаграммы можно заключить, что предпринимателям присуще более развитое правосознание, чем простым гражданам, а судьи превосходят и тех и других. Впрочем, по параметрам поддержки концепции естественного права (2) и нетерпимости к нарушению законов (5) все три выборки статистически неразличимы.

Все становится гораздо интереснее, если изучить взаимосвязи между переменными внутри каждой выборки с помощью нетрадиционных методов многомерной линейной статистики. В описываемом исследовании использовалась множественная линейная регрессия: каждая из пяти переменных по очереди выступала в качестве зависимой переменной, а остальные – в качестве независимых. В результате были получены пять векторов коэффициентов регрессии (в стандартизированном виде), каждый из которых показывал, в какой мере данная переменная прогнозирует значения остальных переменных.

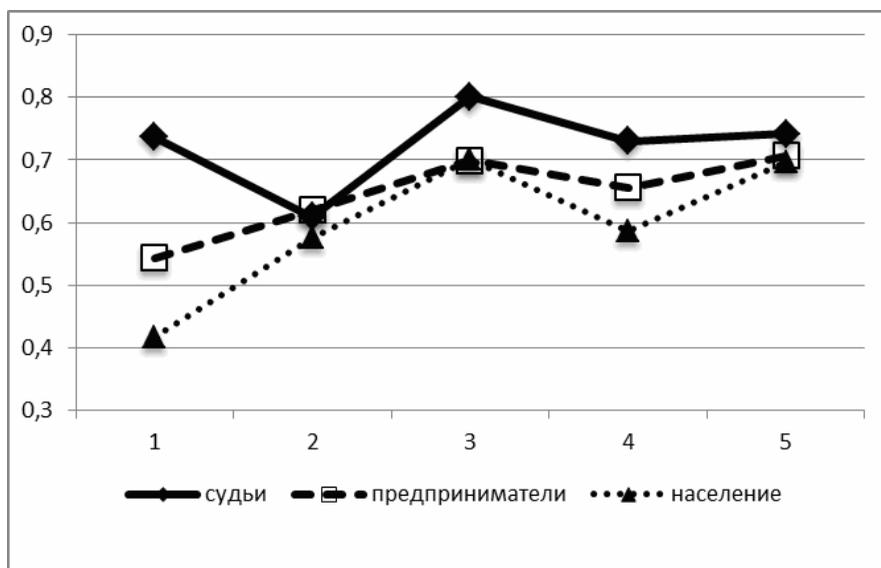


Рис. 4.

**Профили средних значений (ось Y) для выборок судей, предпринимателей и граждан по пяти шкалам правосознания (ось X)**

*Примечание:* 1) поддержка независимости судебной власти; 2) поддержка концепции естественного права; 3) законопослушание; 4) поддержка независимости судебных решений; 5) нетерпимость к нарушению законов.

Сведенные в матрицу коэффициенты (табл. 2) можно трактовать как меру влияния одних переменных на другие. В отличие от коэффициента корреляции, эта мера имеет несимметричный

характер. В частности, поддержка независимости судебных решений почти в два раза убедительнее предсказывает поддержку независимости судебной власти, чем наоборот. Влияние четвертой переменной на остальные в среднем больше, чем в других строках, и это «среднее» имеет смысл центральности как узла сети. Если отбросить коэффициенты с большими значениями доверительной вероятности, то получится сеть (граф) значимых влияний (рис. 5).

Таблица 2

**Матрица взаимных регрессионных коэффициентов  $\beta_{ij}$  для пяти рейтингов правосознания в случае выборки судей ( $i$  – независимая переменная,  $j$  – зависимая переменная)**

	1	2	3	4	5
1		-0,074	-0,025	0,103	0,067
2	-0,102		-0,021	0,283	0,245
3	-0,036	-0,022		0,342	0,211
4	0,190	0,379	0,439		0,295
5	0,104	0,274	0,226	0,246	

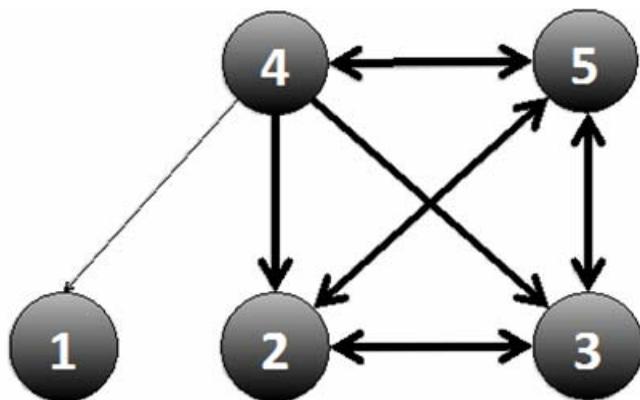


Рис. 5.

**Граф структуры взаимозависимостей (взаимовлияний), определяемой регрессионными коэффициентами  $\beta_{ij}$ , между пятью переменными правосознания в выборке судей**

*Примечание:* Смысл номеров вершин графа: 1) поддержка независимости судебной власти; 2) поддержка концепции естественного права; 3) законопослушание; 4) поддержка независимости судебных решений; 5) нетерпимость к нарушению законов.

Теперь рассмотрим «влиятельность» каждой переменной правосознания на остальные для всех трех выборок (рис. 6).

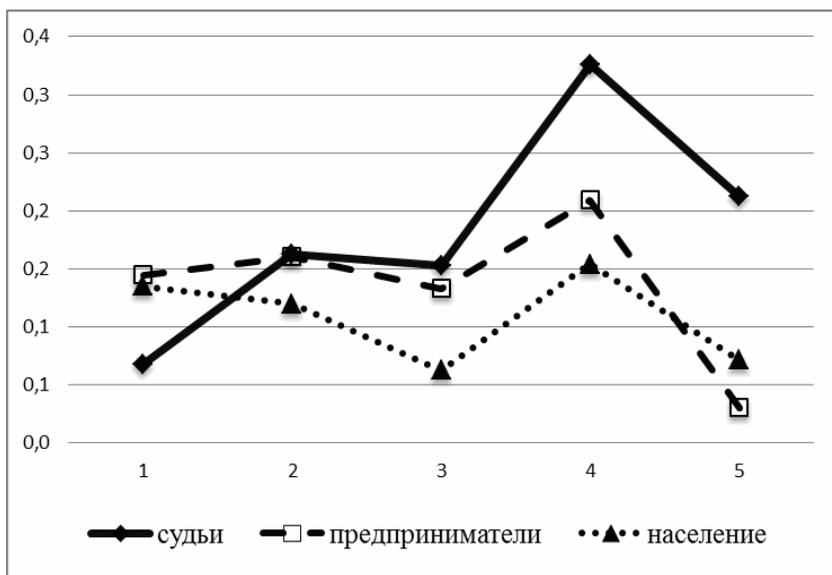


Рис. 6.

**Сравнение профилей влияния (ось Y) переменных правосознания на остальные переменные (ось X) для выборок граждан, предпринимателей и судей**

Если сравнить профили на данной диаграмме с приведенными на рис. 4, то разница очевидна. Но самое главное: во всех трех профилях явно лидирует поддержка независимости судебных решений. Именно эта переменная является прототипической для структур правосознания. Этот факт нельзя было обнаружить, оставаясь в рамках традиционного анализа данных, причем самое разительное лидерство выявляется в группе судей. Именно это отличие судей от двух других групп – наиболее важное, определяющее их правосознание. На рис. 4 обнаруживается лидерство судей (с высоким разрывом) и для другой переменной – поддержки независимости судебной власти. Однако более тщательный анализ показывает, что судьи в своих суждениях на этот счет руководствуются не конституционно-правовыми соображениями, а корпоративными

интересами. Этот факт иллюстрирует то мало осознаваемое обстоятельство, насколько обманчивы могут быть обычные частоты.

Завершая сюжет, снова обратимся к соотношению частот и центральных в случае переменных правосознания (рис. 7).

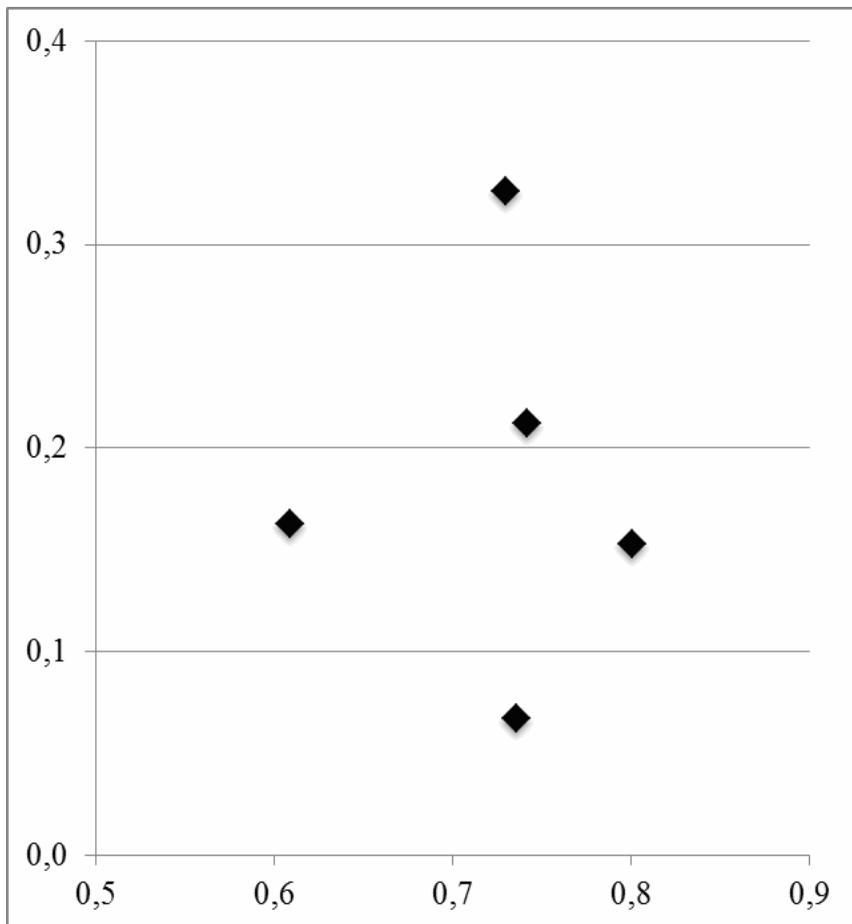


Рис. 7.

**Диаграмма рассеяния среднего значения переменных правосознания (ось X) и центральности этих переменных в сети взаимовлияний (ось Y) для выборки судей**

Диаграмма еще раз демонстрирует факт независимости между средними как традиционными характеристиками<sup>1</sup> и центральностями как структурными характеристиками (и это повторяется для двух других выборов). Воспроизведение этого факта указывает на их несводимость друг к другу. Перейдем теперь к центральностям как таковым и на следующем примере покажем, какую нетривиальную информацию может давать их изучение.

### 3. Пример третий. Изучение доверия

Третий сюжет почерпнут из двух статей автора [Сатаров, 2016 a; Сатаров, 2016 b], в которых элементы «нового подхода» реализуются наиболее последовательно. Здесь статистический «сетевой» анализ также применяется не к сырым ответам на вопросы анкеты, а к результатам их предварительной обработки. Суть нововведений в следующем. Мы мыслим сеть как образованную совокупностью бинарных переменных – конечных векторов, компоненты которых принимают значения 0 или 1; длина вектора равна объему выборки (или, если потребуется, некой ее подвыборки). Каждая переменная соответствует одному из вариантов ответа на все вопросы анкеты. Эти переменные образуют узлы сети. Связи между узлами – это некоторые меры близости (зависимости, сопряженности) между бинарными переменными. Автор уже давно использует в качестве такой меры величину, известную в анализе таблиц сопряженности под названием приведенных стандартизированных остатков. Есть два бинарных вектора  $x$  и  $y$  длины  $n$ ;  $xy$  – вектор, образованный из первых двух путем перемножения соответствующих координат. Через  $|x|$  обозначается число единиц в векторе  $x$  (аналогичным образом это обозначение используется для других векторов). Если вектору  $x$  соответствует некий ответ на некий вопрос, то  $|x|$  – число респондентов, выбравших этот ответ на этот вопрос. Величина

$$e_{xy} = |xy| - |x||y|/n$$

известна под названием «остаток». Он равен нулю или близка к нему, если (как в нашем примере) выбор респондентами отве-

---

<sup>1</sup> Напомним, что частота – это обычное среднее для бинарной переменной.

тов  $x$  и  $y$  есть события независимые. Когда остаток  $e_{xy}$  значимо отклоняется от нуля, мы имеем основания говорить о положительной ( $e_{xy}$  больше нуля) или отрицательной ( $e_{xy}$  меньше нуля) зависимости между выбором респондентами двух ответов. Чтобы оценить статистическую значимость отклонения, от величины  $e_{xy}$  переходят к связанной с ней статистике<sup>1</sup>:

$$A_{xy} = \frac{e_{xy}}{\sqrt{\frac{|x|(n - |x|)|y|(n - |y|)}{n^2(n - 1)}}}$$

которая имеет стандартное нормальное асимптотическое распределение. Это позволяет дальше действовать следующим образом. Мы задаемся некоторой устраивающей нас доверительной вероятностью  $P$  (0,05; 0,01, 0,005 ...) и используем соответствующее ей значение  $z$  – такое, что  $(1 - N(z)) = P$ , где  $N$  – функция стандартного нормального распределения в качестве порогового значения. Далее мы исключаем из сети все связи между узлами (бинарными векторами)  $x$  и  $y$ , которым соответствуют абсолютные значения статистики  $A_{xy}$ , меньшие  $z$ . Таким образом, в сети остаются только связи между узлами со статистически значимыми зависимостями.

На данном этапе развития «нового подхода» мы не изучали всю сеть, которую можно построить на основе ответов на все вопросы некой анкеты. Описанное выше сетевое представление данных опроса использовалось для изучения локальных свойств отдельных узлов сети. Изучались все значимые связи некоего узла и характеристики этой совокупности связей (например, центральность). В двух упомянутых выше статьях объектами изучения были ответы респондентов на вопросы о степени доверия к четырем институтам: к президенту, правительству, Федеральному собранию и милиции (полиции). Респонденты могли выбрать один из вариантов ответа: «Полностью доверяю», «Скорее доверяю», «Скорее не доверяю», «Совсем не доверяю». Вопросы о доверии этим институтам задавались в разные годы. В описываемом исследовании сравнивались данные 1998, 2001 и 2015 гг. В качестве меры центральности использовалась доля высоких по абсолютному

<sup>1</sup> Это есть не что иное, как статистика  $t$ , определяемая формулой (33.52) в монографии [Кендалл, с. 739]. Она же известна под названием «приведенный стандартизированный остаток».

значению связей во всей совокупности рассматриваемых связей. Сопоставлялись центральности (и другие характеристики) для разных вариантов ответа, разных институтов и разных годов. Этот весьма поверхностный анализ дал немало важных результатов, часть из которых будет представлена ниже.

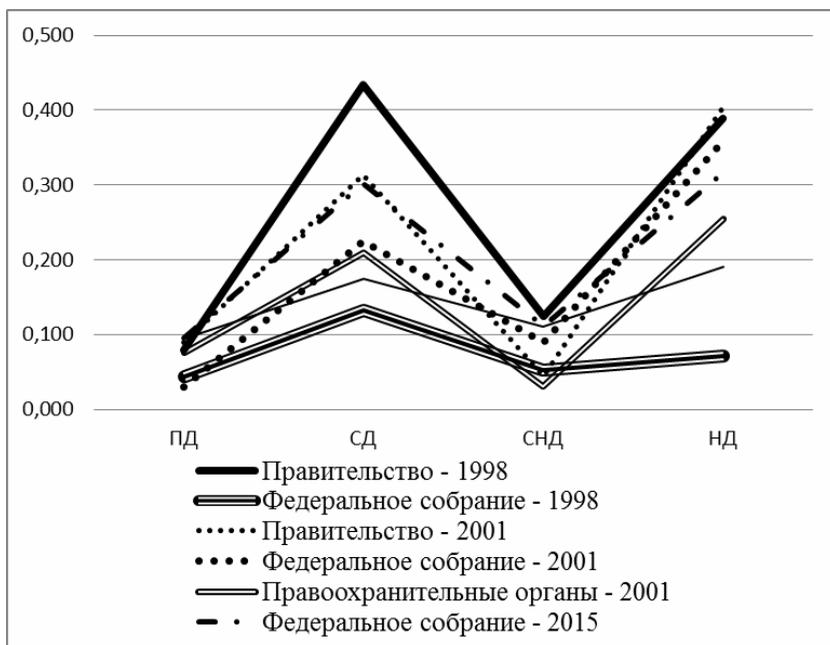


Рис. 8.

**Профили центральностей (долей больших по абсолютной величине остатков) для семи властных институтов в разные годы для трех исследований, образованные четырьмя вариантами ответов**

*Примечание:* ПД – «Полностью доверяю», СД – «Скорее доверяю», СНД – «Скорее не доверяю», НД – «Совсем не доверяю». Не представлены профили центральностей для президента (приводятся ниже), правоохранительных органов 1998 г. и правительства 2015 г. (причина будет объяснена ниже).

Все приведенные на диаграмме профили (рис. 8) объединяет общая форма – единство максимумов и минимумов. Смысл данного единства становится ясен из социологической интерпретации центральности в сети, сконструированной указанным выше способом.

Каждому узлу сети соответствует некоторое подмножество выборки респондентов, выбравших определенный ответ на вопрос, соответствующий этому узлу. Чем выше центральность узла и соответственно доля больших по абсолютной величине остатков, тем больше других переменных описывают это подмножество респондентов с высокой степенью статистической достоверности. Учитывая это соображение и возвращаясь к профилям на рис. 8, мы можем утверждать, что группы респондентов, выбирающих варианты ответов «Скорее доверяю» или «Совсем не доверяю», характеризуются более богатыми и контрастными «социальными портретами», чем группы респондентов, выбирающих два оставшихся ответа. В исследовании объяснялось, почему это свойственно нормальному состоянию общественного мнения в демократической стране.

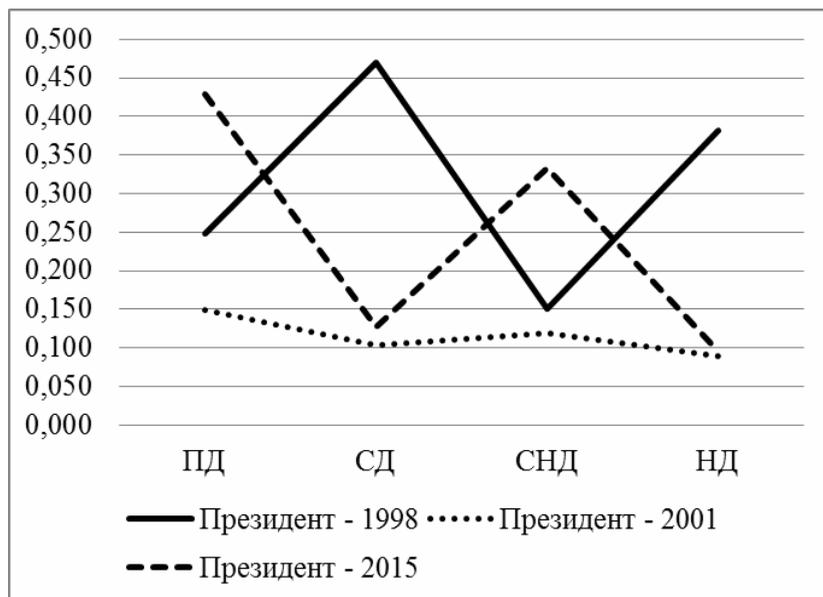


Рис. 9.

**Профили центральностей (долей больших по абсолютной величине остатков) по данным трех исследований о доверии президенту**

В случае с доверием президенту (рис. 9) мы видим уникальную метаморфозу формы профиля между двумя временными точ-

ками – 1998 и 2015 гг. Профиль срединной точки (2001) сглажен и соответствует переходному состоянию. Ни одна из групп респондентов, какой бы ответ она ни выбирала, не имеет отчетливого социального лица, что свойственно переходным хаотическим процессам. В исследовании объяснялось, почему профиль центральных в 2015 г. соответствует аномальному состоянию общественного мнения. Любопытно также, что сопоставление двух приведенных выше диаграмм позволяет предположить, что внешнее политическое воздействие на общественное мнение приводит к его аномальному состоянию в локальных зонах и не обязательно поражает всю сферу общественного мнения (если оно однажды сформировалось).

Когда мы строим сеть описанным способом, бессмысленно рассматривать зависимость между узлами сети, которым соответствуют два разных ответа на один и тот же вопрос: их большая отрицательная зависимость обусловлена конструкцией вопроса, а именно необходимостью выбирать единственный ответ из заданного списка. Но мы можем рассматривать структурную близость (обобщение структурной эквивалентности) двух таких узлов. Мы оцениваем структурную близость двух узлов сети как высокую, если они сходным (не обязательно тождественным) образом связаны с остальными узлами. В описываемом исследовании эта идея была реализована следующим образом. При сопоставлении группы узлов сети, образованных полным набором ответов на вопрос о доверии, скажем, президенту (таких узлов в группе 4), рассматривались (почти) все статистически значимые зависимости с другими вопросами анкеты. Для каждого из четырех ответов строились вектора значений величин  $A_{xy}$  одинаковой длины, где  $x$  соответствует одному из ответов на вопрос о доверии, а  $y$  – произвольному ответу на другой вопрос анкеты, находящийся в значимой сопряженности с этим вопросом. Эти вектора назывались «векторами остатков».

В нашем исследовании структурная близость индцировалась обычным линейным коэффициентом корреляции между векторами остатков четырех ответов на вопрос о степени доверия (например, к президенту). Общепринятая практика состоит в том, что, оповещая о результатах своих опросов, социологи складывают частоты «смежных» ответов – например, «полностью доверяю» и «скорее доверяю», утверждая затем, что президенту в целом доверяют столько-то (сумма двух частот) процентов респондентов. При определенных условиях этот ход вполне правомерен. Лет 60–70 на-

зад его обосновывали бы следующим образом. Мы предполагаем, что существует некоторая латентная, недоступная нам при непосредственном наблюдении, непрерывная шкала степени доверия (кого-либо чему-либо или кому-либо). В социологическом опросе мы огрубляем эту шкалу, разбивая ее на несколько (в нашем случае четыре) смежных интервалов – от «полностью доверяю» (крайний правый интервал) до «совсем не доверяю» (крайний левый интервал). Важно, что вся эта шкала (и интервалы на ней) объединены единым смыслом, т.е. трактовкой доверия / недоверия. При социологическом опросе в эти интервалы попадают конкретные респонденты, среди которых очень непохожие личности, отличающиеся друг от друга самыми разнообразными качествами, которые фиксируются в том числе и анкетой. Тем не менее это разнообразие существенно меньше, чем разнообразие всех респондентов в выборке. Поэтому мы можем описать социальный портрет типичного респондента, попавшего в некий интервал шкалы (к примеру, «скорее доверяющего»). В нашем случае этот портрет задан совокупностью вершин в построенной нами сети, обладающих высоким значением остатков  $A_{xy}$ . Понятно, что социальные портреты «полностью доверяющих» и «скорее доверяющих» будут не тождественны, но близки. А портреты «полностью доверяющих» и «совсем не доверяющих» будут далеки, т.е. совсем не похожи.

Если вернуться к векторам остатков, то сходство / различие социальных портретов представителей четырех интервалов на непрерывной шкале доверия / недоверия будет выражаться коэффициентами корреляции между векторами остатков. Если описанные предположения верны, то матрица этих коэффициентов может выглядеть примерно так (см. табл. 3 и рис. 10).

Таблица 3

**Искусственный пример матрицы линейных корреляций между векторами остатков четырех вариантов ответов на вопрос о доверии произвольному властному институту**

	Полностью доверяю	Скорее доверяю	Скорее не доверяю	Совсем не доверяю
Полностью доверяю	1	0,75	-0,55	-0,85
Скорее доверяю	0,75	1	-0,40	-0,55
Скорее не доверяю	-0,55	-0,40	1	0,75
Совсем не доверяю	-0,85	-0,55	0,75	1

Предполагается, что все корреляции статистически значимы, но в разной степени. На рисунке это отражено толщиной линий: сплошные линии соответствуют положительным зависимостям, пунктирные – отрицательным.

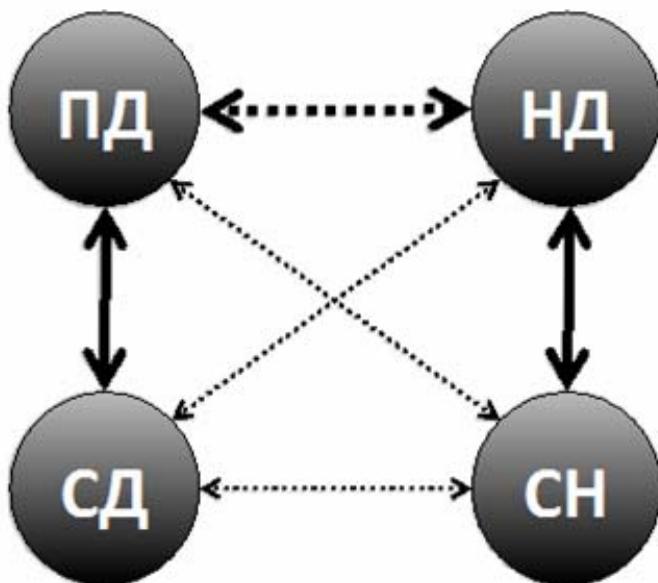


Рис. 10.

**Граф, соответствующий искусственному примеру взаимосвязей между векторами остатков для четырех ответов на вопрос о доверии**

Приведенный выше идеальный портрет структуры зависимостей можно сравнить с составленным для доверия президенту, по данным 1998 г. (рис. 11), на котором малозначимые корреляции не видны.

Как видим, реальная диаграмма похожа на «идеальную» за двумя исключениями. Первое незначительно: отрицательная связь для ответов «скорее доверяю» и «скорее не доверяю» стала статистически более значимой. Второе существеннее: связь между векторами остатков ответов «скорее не доверяю» и «совсем не доверяю» потеряла статистическую значимость. В статье приводится одно из объяснений этого факта.

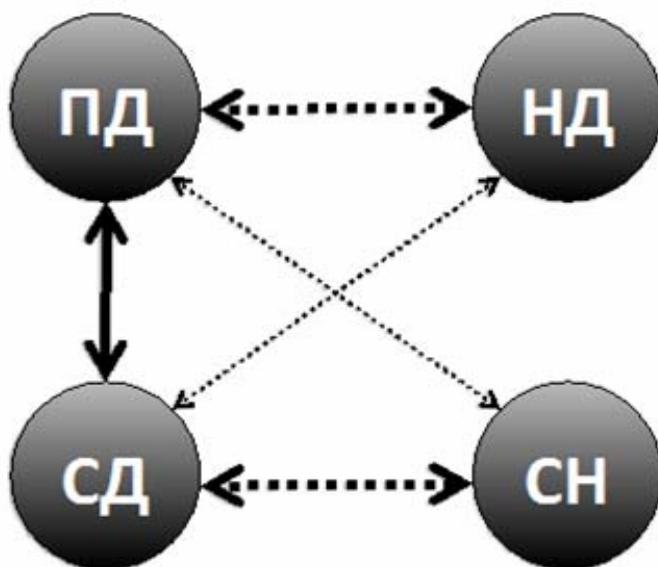


Рис. 11.

**Граф взаимосвязей между векторами остатков для четырех ответов на вопрос о доверии президенту, по данным опроса 1998 г.**

Теперь рассмотрим картину с данными 2015 г. (рис. 12).

Здесь мы фактически имеем два совершенно разных портрета «полностью доверяющих» и «скорее доверяющих», и оба они не похожи на портреты «скорее не доверяющих» и «совсем не доверяющих». Революция в структуре доверия президентам в промежутке между 1998 и 2015 гг. становится еще более очевидной, когда переходишь к изучению самих портретов, а не структурных соотношений между ними [Сатаров, 2016 б]. Содержательный анализ подтверждает вывод, который можно было сделать на основании последней диаграммы: частоты выбора ответов «полностью доверяю» и «скорее доверяю» нельзя суммировать не только потому, что в данном случае не работает модель смежных интервалов на непрерывной шкале доверия, но и в силу принципиального различия смыслов, вкладываемых респондентами в эти два ответа.

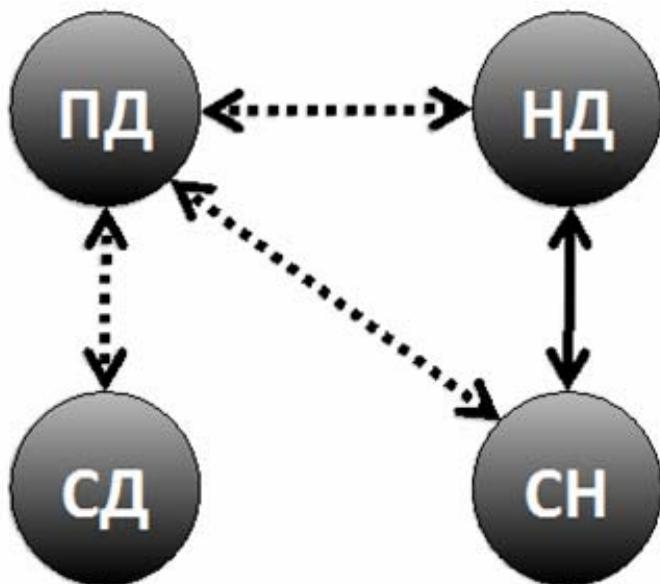


Рис. 12.

**Граф взаимосвязей между векторами остатков для четырех ответов на вопрос о доверии президенту по данным опроса 2015 г.**

#### 4. Заключение

Три рассмотренных примера применения нового подхода к данным массовых опросов были подобраны так, чтобы они различались и по тематике, и по технике анализа. Во всех трех случаях получены новые результаты, выходящие за рамки традиционной задачи – узнать, что думают люди, и каково содержание общественного мнения. Акцент переносится на решение другой задачи: понять, как устроено общественное мнение и почему оно устроено именно так. Такое смещение неслучайно, оно вызвано тем, что мы очень мало знаем об устройстве общественного мнения. Это не значит, что первый тип вопросов неактуален или что новыми методами нельзя найти ответы на традиционные вопросы. Просто эти ответы теперь выглядят по-другому. В частности, на первом примере показано, что, переходя на структурный (сетевой) уровень

анализа, мы не только определяем частоту выбора ответа, но и выявляем группы респондентов, в которых эти частоты различаются – не потому, что это разные группы (например, мужчины и женщины), а потому, что представители этих групп приписывают своим ответам разный смысл. Это сложнее, но зато вместо частоты того или иного ответа на вопрос мы получаем картину распределения респондентов, придающих этому ответу разный смысл<sup>1</sup>. Одновременно подтверждается то, о чем предупреждал О. Нейрат, опираясь на свою социологическую интуицию, и о чем писал П. Бурдые, анализируя ответы респондентов: частота ответа на вопрос может быть бессмысленной<sup>2</sup>.

Во всех трех примерах легко усмотреть нечто общее. Переходя на сетевой уровень анализа, мы начинаем видеть нечто, что раньше было скрыто. Например, существует утверждение: «Рост преступности – одна из важнейших проблем страны». Рассматриваемое изолированно, оно может обладать различными смыслами. Но когда оно, как в примере 1, сопрягается с другими утверждениями, оно участвует в конструировании более конкретного смысла. В 1998 г. мы проводили опрос почти по такой же анкете в Эстонии. Там проблема «рост преступности» попала в один класс с «ростом цен» и «слабостью, беспомощностью государственной власти». Совершенно очевидно, что она оказалась там вместе со своим другим смыслом, отличным от того, с которым она вошла в свой класс в России. При традиционном подходе мы просто зафиксировали бы различие в частоте выбора этой проблемы в двух странах и придумали бы, как это интерпретировать, не заботясь о верификации объяснения. Но при использовании «нового подхода» к разнице частот добавляются различия в принадлежности проблемы к некоторым классам в структуре массового сознания и в значениях проблемы, которые, сопрягаясь с другими проблемами, приобретали различные смыслы<sup>3</sup>.

---

<sup>1</sup> Надо напомнить, что рассмотренный в первом примере набор смыслов, которые можно приписать одному и тому же ответу в трех разных классах респондентов, совершенно умозрительны, хотя и правдоподобны. Но довольно легко сконструировать схему социологической верификации этого умозрительного построения и подтвердить его.

<sup>2</sup> Из этого, правда, вопреки мнению П. Бурдые, не следует, что общественного мнения не существует.

<sup>3</sup> Может возникнуть вопрос, почему в России, в отличие от Эстонии, не образовалось класса проблем с общим смыслом «неэффективность власти». Представляется, что в обеих странах респонденты невысоко оценивали эффек-

Еще более поразительные результаты дает третий пример. Выясняется, что может быть несостоятельной сама практика сложения частот двух близких ответов, и причина тому – принципиальное различие в мотивах выбора ответа, один из которых не имеет отношения к категории доверия. Это, в свою очередь, объясняется деформацией общественного мнения, обусловленной политическими причинами. Причем в каких-то зонах общественного мнения эта деформация происходит, а в каких-то – нет.

Рушится привычная, простая и плоская, картина общественного мнения – вместо нее проступает нагромождение неожиданных фактов, которые хором кричат, что та картина не имела никакого отношения к реальности. Это если и преувеличение, то довольно близкое к истине. Мне, к примеру, трудно объяснить, почему потоки коммуникаций в обществе устроены проще, чем потоки электрических сигналов в проводах компьютера. Наоборот, я считаю общество объектом несопоставимо более сложным, чем компьютер.

Утешить может только одно: «новый подход» открывает зрелище, более интересное, чем то, с которым мы имели дело раньше. Общество, социальная жизнь усложняются на наших глазах. Одновременно возрастают риски, порождаемые неадекватностью наших представлений о самих себе. Социология, и в первую очередь политическая социология, не может оставаться такой же, какой она была 70 лет назад.

Если принять предлагаемую в статье позицию, описывать общественное мнение можно только анализируя массовое сознание. При этом меняется само «устройство» описания общественного мнения. Представленное в статье – лишь поверхностный взгляд на проблему. Ее решение потребует гораздо более серьезных усилий. Подспорьем может оказаться опыт, уже накопленный при анализе сетевых структур и в сфере *datamining* [Степанов, 2008; Чураков, 2001; McCulloh, 2013; Tan Pang-Ning, Steinbach, Kumar, 2006]. Не удивлюсь, если при анализе массового сознания и общественного мнения эти методы дадут самые интересные результаты.

## Список литературы

- Бурдые П. Общественное мнение не существует // Бурдые П. Социальное пространство: Поля и практики / Пер. с франц.; Отв. ред. перевода, сост. и послесл. Н.А. Шматко. – М.: Алетей, 2005. – С. 272–285.
- Кендалл М., Стьюарт А. Статистические выводы и связи / Пер. с англ. – М.: Гл. ред. физ.-мат. лит.-ры: Наука, 1973. – 899 с.
- Когнитивные сети / Евин И.А., Кобляков А.А., Савриков Д.В., Шувалов Н.Д. // Компьютерные исследования и моделирование. – М., 2011. – Т. 3, № 3. – С. 231–239.
- Липпман У. Общественное мнение / Пер. с англ. – М.: Институт Фонда «Общественное мнение», 2004. – 384 с.
- Луман Н. Введение в системную теорию / Пер. с нем. – М.: Логос, 2007. – 360 с.
- Российская коррупция: Уровень, структура, динамика. Опыт социологического анализа / Под ред. Г.А. Сатарова. – М.: Фонд «Либеральная миссия», 2013. – 752 с.
- Сатаров Г.А. Доверие как объект политической социологии. Часть 1 // Полис. Политические исследования. – М., 2016 а. – № 1. – С. 121–138.
- Сатаров Г.А. Доверие как объект политической социологии. Часть 2 // Полис. Политические исследования. – М., 2016 б. – № 2. – С. 109–130.
- Сатаров Г.А., Благовещенский Ю.Н., Римский В.Л. Социологический анализ правосознания судей, населения и предпринимателей. – М.: Фонд «Либеральная миссия», 2016 а. – 176 с.
- Сатаров Г.А., Римский В.Л., Благовещенский Ю.Н. Социологическое исследование российской судебной власти. – СПб.: Норма, 2010 б. – 536 с.
- Соссюр Ф. де. Курс общей лингвистики / Пер. с франц. – М.: КомКнига, 2006. – 256 с.
- Степанов Р.Г. Технология datamining: Интеллектуальный анализ данных. – Казань: КГУ, 2008. – 58 с.
- Цаллер Дж. Происхождение и природа общественного мнения / Пер. с англ. – М.: Институт Фонда «Общественное мнение», 2004. – 559 с.
- Чураков А.Н. Анализ социальных сетей // Социологические исследования. – М., 2001. – № 1. – С. 109–121.
- McCulloh I., Armstrong H. Social network analysis with applications. – Hoboken, New Jersey: Wiley & Sons, 2013. – xxx, 283 p.
- Neurath O. Foundations of the social sciences // International encyclopedia of unified science. – Chicago, IL: Univ. of Chicago press, 1944. – Vol. 2, N 1. – Rep. in: Foundations of the unity of science: Toward an international encyclopedia of united science / Neurath O., Carnap R., Morrus C. (eds.). – Univ. of Chicago press, 1970. – Vol. 2. – P. 1–52.
- Tan Pang-Ning, Steinbach M., Kumar V. Introduction to data mining. – Boston, Mass.: Pearson Addison Wesley, 2006. – 769 p.