



Национальный исследовательский университет «Высшая школа экономики»

Факультет социологии

Российское общество социологов

Сообщество профессиональных социологов

СБОРНИК
СТАТЕЙ



НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
**СОВРЕМЕННАЯ СОЦИОЛОГИЯ -
СОВРЕМЕННОЙ РОССИИ**

посвящается памяти
первого декана факультета социологии
**Александра Олеговича
Крыштановского**



МОСКВА
1 – 3
ФЕВРАЛЯ
2012



ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ



Российское общество социологов

Советская социологическая ассоциация - Российское общество социологов



УДК 3160:167/168(06)

ББК 66.5

С56

Редакционный совет

А. Б. Гофман, Г. В. Градосельская, И. Ф. Девятко, Д. Х. Ибрагимова,
И. М. Козина, Л. Я. Косалс, В. А. Мансуров, В. Г. Николаев, О. А. Оберемко,
Н. Е. Покровский, Ю. Н. Толстова, А. Ю. Чепуренко (председатель), Е. Р. Ярская-Смирнова

С56 **Современная социология — современной России: Сборник статей памяти первого декана факультета социологии НИУ ВШЭ А. О. Крыштановского [Электронный ресурс] / НИУ ВШЭ; РОС; СоПСо. — М.: НИУ ВШЭ, 2012. — 753 с. — 1 электрон. диск (CD-ROM).**

ISBN 978-5-904804-08-4

Сборник содержит статьи, подготовленные на основе докладов, сделанных на VI научно-практической конференции «Современная социология — современной России» (г. Москва, 1–3 февраля 2012), посвященной памяти первого декана факультета социологии НИУ ВШЭ Александра Олеговича Крыштановского.

Сборник рассчитан на тех, кто интересуется методологией, методикой и современной практикой проведения социологических исследований.

УДК 3160:167/168(06)

ББК 66.5

ISBN 978-5-904804-08-4



© Национальный исследовательский университет
«Высшая школа экономики», 2012

© Российское общество социологов, 2012

© Сообщество профессиональных социологов, 2012



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

- Е. И. Князева**
544 **Методологические характеристики и познавательные возможности исследований социальных сетей в социологии**
- Н. А.Найденов**
556 **Автоматический сбор информации из открытых Интернет-источников**
- Д. В. Просянюк**
561 **Содержательные основания выделения границ Интернет-сетей**
- С. В. Бондаренко**
581 **«Электронное государство» как социотехническая система**
- И. С. Хван, А. А. Горгадзе**
589 **Этнические сообщества в социальных сетях**
- И. В. Забаев, Е. В.Пруцкова**
598 **Социальные сети поддержки в православной общине. На примере трех крупных православных приходов г. Москвы**
- Г. В. Градосельская**
609 **Анализ структуры бизнес-конфликтов в российских СМИ при помощи поисковых алгоритмов**
- Е. С.Михеева**
617 **Методологические и методические основания изучения полилога, как формы коммуникативных процессов**
- V. V. Ulyanov**
625 **On asymptotics for power divergence family of statistics**



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Методологические характеристики и познавательные возможности исследований социальных сетей в социологии

Екатерина Ивановна Князева

Белорусский государственный университет (Минск)

В статье обсуждаются место и роль сетевого подхода в социологических исследованиях. Осуществлен сравнительный анализ количественных социальных исследований индивидов, текстов и социальных связей по основным методологическим характеристикам – объект, предмет, выборочные методы и стратегии, методы сбора и анализа данных, уровни анализа. В ходе эмпирического исследования сети поддержки выпускников отделения социологии Белорусского государственного университета (Минск, Беларусь), выявлены познавательные возможности атрибутивных, реляционных и ресурсных данных в сетевых исследованиях.

В настоящее время сетевой подход широко распространен в исследованиях социальных связей и их роли в различных сферах деятельности и повседневной жизни людей. Синтетически объединяя в себе сетевую теорию и анализ социальных сетей, сетевой подход представляет собой одно из наиболее мощных и влиятельных направлений в современной западной социальной теории, о чем свидетельствует множество выпущенных статей и книг, посвященных данной проблематике, создание специализированных научных журналов и ассоциаций.

Популярность сетевого подхода в целом и сетевых исследований в частности можно объяснить большой гибкостью при изучении коммуникативных практик социальных взаимодействий и возможностью структурировать взаимоотношения между социальными акторами в наиболее ясной визуальной и количественной форме.

Цель данной статьи состоит в репрезентации методологических характеристик и познавательных возможностей исследований социальных сетей на основе собственного опыта изучения социальной сети поддержки выпускников отделения социологии БГУ, работающих в рекламном бизнесе и маркетинге. В данном исследовании мы опирались на концептуальный подход М. Грановеттера, который определяет

социальную сеть поддержки как устойчивую совокупность взаимосвязей и отношений между ее участниками по обмену различными ресурсами.

Методологические характеристики сетевых исследований

Единицей анализа в сетевых исследованиях является *связь* между двумя акторами сети (или ее отсутствие). Специфика единицы анализа и соответствующих ей методов сбора и анализа данных позволяет рассмотреть сетевые исследования в качестве самостоятельного вида социологических исследований, наряду с видами исследований, в которых единицами анализа являются индивиды (группы индивидов, организации и т. п.), и тексты (фрагменты текстов).

Результаты сравнительного анализа трех видов социальных исследований, выделенных по критерию единицы анализа: исследования индивидов, исследования текстов, исследования социальных связей (сетевые исследования) для удобства восприятия сведены в таблицу, представленную ниже:

Таблица 1

Основные методологические характеристики количественных социальных исследований индивидов, текстов и социальных связей

	Классификация количественных социальных исследований (СИ) по критерию единицы анализа		
	Исследования индивидов	Исследования текстов	Исследования социальных связей
Объект	Генеральная совокупность (ГС) индивидов (коллективов), относящихся к теме СИ	ГС текстов, относящихся к теме СИ	Социоцентрическая или эгоцентрическая сеть
Выборочные методы и стратегии	Вероятностные и целевые (теоретические) методы отбора		Реконструкция сети на основе характеристик или отношений акторов (номиналистическая и реалистическая стратегия)
Предмет	Атрибутивные характеристики ГС, нормы, убеждения, мотивы поведения...	Формальные и содержательные характеристики текстов, коммуникативные намерения авторов...	Структура связей актора или акторов социальной сети
Методы сбора данных	Методы опроса, наблюдения, эксперимента (структурированные и неструктурированные)	Контент-аналитическое кодирование, измерение параметров текста,	Социометрический опрос (наблюдение, анализ дневников, биографических материалов...)
Основные методы анализа данных	Статистический анализ		Анализ социальных сетей
Уровень анализа	Мезо- и макроуровень		Микро-, мезо- и макроуровень

Рассмотрим подробнее методологические характеристики сетевых исследований. *Объектом* сетевых исследований являются социальные сети, рассматриваемые как конкретные совокупности акторов и их социальных связей. При этом сети могут быть самые разные – личные или эгоцентричные сети, сети цитирования, электронные социальные сети, сети поддержки, сети распространения информации и др.

Мы исходим из определения социальной сети как совокупности *трех* составляющих – *акторов* (индивиды или коллективы), *социальных связей*, возникающих между ними на основе определенных отношений, и *ресурсов* (материальных и нематериальных), доступных членам сети через социальные связи (прямые или опосредованные) с теми акторами, которые ими владеют или распоряжаются. В соответствии с предложенным определением специфика проведения эмпирических исследований социальных сетей заключается в сборе, обработке, анализа и интеграции *трех* видов данных – *атрибутивных, реляционных и ресурсных*, – каждый из которых позволяет решать собственный класс задач. При этом основным *предметом* исследования все же выступает структура связей, поэтому сеть рассматривается в качестве генеральной совокупности исследования.

Выборочный метод, в традиционном его понимании, в сетевых исследованиях не применяется. Для реконструкции социальной сети используется одна из двух стратегий: первая определяется границами организации и /или социальной общности (номиналистическая стратегия), вторая – отношениями между акторами (реалистическая стратегия). В последнем случае часто применяется метод «снежного кома», который в сетевых исследованиях получил название «сетевой выборки».

В соответствии с единицей анализа, основным *методом сбора данных* в сетевых исследованиях является социометрический опрос. Могут также применяться наблюдение, анализ текстов и другие методы сбора данных. Но в любом случае они используются для того, чтобы получить *реляционные данные* – информацию о *связях* между акторами сети, и реконструировать социоматрицу – основу для дальнейшей обработки и визуализации данных.

Для *обработки реляционных данных* используются методы анализа социальных сетей (АСС). Основными методами АСС являются описательный сетевой анализ (анализ сплоченных подгрупп, индивидуальные и групповые показатели центральности, позиционный и ролевой анализ); методы визуализации сетевой структуры; стохастический сетевой анализ, включающий в себя методы проверки гипотез. С их помощью можно реконструировать изучаемую сеть, визуализировать ее структуру, формализовать положение индивида в сети посредством различных показателей центральности, изучать сплоченные подгруппы и их основные характеристики и мн. др. Данные методы реализованы в многочисленном программном обеспечении анализа социальных сетей – UCINET, Pajek, STOCNET, GRADAP, NEGOPY, MultiNet и др.

Сетевые исследования дают возможность понять через социальные сети актора (акторов) как его (их) частные способы взаимодействия с социумом (микроуровень), так и некоторые механизмы существования самого социума (мезо- и макроуровень). Другими словами, сетевые

исследования являются *многоуровневыми*. С одной стороны, природа сетей определяется входящими в них акторами. С другой стороны, сами акторы и их поведение во многом зависят от сетевых структур, в которые они входят.

Таким образом, роль сетевого подхода в социальных исследованиях определяется спектром задач, которые он позволяет решать. Типичными задачами сетевого подхода являются реконструкция структуры связей, определение ее характеристик, выявление принципов и механизмов функционирования. Базируясь на количественных исследованиях социальных сетей, сетевой подход позволяет выходить за рамки традиционных методов и объяснительных теорий, представляя новые возможности для изучения социальной реальности.

Познавательные возможности сетевых исследований

Для того, чтобы проиллюстрировать методологические характеристики сетевых исследований, которые мы обсуждали выше, а также для более четкой репрезентации их познавательных возможностей приведем результаты собственного эмпирического исследования.

Объектом нашего исследования стали выпускники отделения социологии Белорусского государственного университета, работающие в сфере рекламы и маркетинга. Отделение социологии БГУ было создано в 1989 году. К моменту исследования (2006 г.) состоялось тринадцать выпусков, всего было аттестовано около 400 специалистов-социологов. Довольно значительная доля выпускников отделения работает в сфере рекламы и маркетинга, при этом большинство из них довольно активно взаимодействуют между собой как по рабочим, так и по личным вопросам. В частности, было известно, что многие из них оказали и продолжают оказывать помощь друг другу в трудоустройстве.

Предметом исследования являются социальные связи социологов, работающих в сфере рекламного бизнеса и маркетинга, и связанные с ними ресурсы, способствующие трудоустройству, а также подбору квалифицированного персонала, такие как рекомендация, информация о вакансии, содействие в обучении, создание рабочего места.

Социальные связи выпускников в данном исследовании рассматриваются как неформальный канал трудоустройства. В условиях реструктуризации экономики и трансформационных процессов в постсоветских странах в целом и Беларуси в частности для экономически активного населения вопросы поиска высокооплачиваемой работы являются достаточно актуальными. В данных условиях личные связи, с одной стороны, являются наиболее распространенным способом поиска работы для многих, в том числе и для выпускников вузов, с другой стороны, распространенной стратегией подбора работников.

Совокупность социальных связей и доступных через них ресурсов мы рассматриваем как форму социального капитала, предоставляющую уникальные возможности и облегчающую взаимодействия индивидуальных (социологи) и корпоративных (организации, в которых они работают) акторов в сфере рекламного бизнеса и маркетинга.

Определение границ сети является важной задачей сетевого исследования, так как упущение действующих акторов сети, произвольное очерчивание границ могут ввести исследователя в заблуждение и профанировать результаты.

В силу того что объем и состав исследуемой нами сети выпускников отделения социологии БГУ, работающих в сфере рекламного бизнеса и маркетинга, не поддается объективной рефлексии, сеть была реконструирована методом «снежного кома». Список членов сети начинался с выпускников, работающих в крупнейших рекламных и маркетинговых агентствах, и продолжался до тех пор, пока респонденты не перестали называть новые имена. В итоге в списке оказалось 45 человек. Выбор исходных источников информации позволил избежать систематической ошибки, т.е. возможности пропустить значительный «фрагмент» сети. Полученная совокупность акторов сети рассматривается нами как генеральная, поскольку сеть поддержки включает в себя тех, кто поддерживает дружеские и / или профессиональные связи, и вероятность того, что кто-то из действующих членов сети не был назван ни одним из включенных в нее социологов, очень мала.

Сбор данных. Специфика сбора данных определялась необходимостью получения данных трех видов – атрибутивных, структурных и ресурсных. *Атрибутивные данные* – характеристики респондентов (пол, возраст, год окончания университета, место работы, должность) и данные о динамике профессиональной карьеры (число мест работы, длительность работы в каждой указанной организации) – были получены методом *анкетного опроса*. *Структурные данные* – частота контактов респондентов между собой, год знакомства, тип взаимодействия (профессиональный, дружеский, смешанный), и *ресурсные данные* – информация о связях, задействованных выпускниками при трудоустройстве / поиске работников, и характере полученной помощи – собраны методом *социометрического опроса*. Респондентам были заданы вопросы относительно их трудового опыта, начиная с первого места работы, в частности, какие действия они предпринимали, чтобы попасть на работу, кто и чем им помогал в этом и кому они сами способствовали в трудоустройстве.

Результаты обработки и анализа данных. Атрибутивные данные нашего исследования были проанализированы на основе стандартных статистических процедур с помощью программы SPSS 13.0. Структурные и ресурсные данные обрабатывались методами сетевого анализа, реализованными в программе UCINET 6.0, визуализированы с помощью программы Netdraw.

Атрибутивные характеристики социологов, входящих в сеть. Количество опрошенных выпускников отделения социологии составило 45 человек; из них 25 человек работают в сфере рекламы, 20 – в сфере маркетинга. Занимаемые должности: специалист по размещению рекламы – 17; руководитель в сфере размещения рекламы (директор рекламного агентства, начальник медиа-отдела в агентстве) – 8; специалист по маркетингу – 15; руководитель в сфере маркетинга (начальник отдела маркетинга, заместитель начальника отдела маркетинга) – 5 (см. таблицу 2). Распределение акторов сети по полу – 13 мужчин и 32 женщины – отражает гендерный дисбаланс, который существует на отделении социологии в БГУ с момента его создания.

Распределение респондентов по полу и сферам деятельности

	Муж.	Жен.	Всего
Должность			
Специалист по рекламе	4	13	17
Руководитель в сфере рекламы	2	6	8
Специалист по маркетингу	4	11	15
Руководитель в сфере маркетинга	3	2	5
Всего:	13	32	45

Структурные характеристики сети. В результате обработки данных о связях выпускников между собой нами была получена социометрическая симметричная матрица профессиональных, дружеских и смешанных связей между выпускниками размерности 4545, по числу респондентов. На основании данной матрицы с помощью программы визуализации NetDraw нами был построен граф связей и тем самым реконструирована сеть социальных связей социологов в сфере рекламного бизнеса и маркетинга (рис. 1). Для того чтобы проще было визуально анализировать полученную сетевую структуру, мы ввели следующие условные обозначения для ее вершин: ▲ – специалист в сфере маркетинга (их идентификационные номера начинаются с буквы *m*); ▼ – руководитель в сфере маркетинга (*rm*); – специалист по рекламе (*r*); □ – руководитель в сфере рекламы (*dr*). Последние две цифры в идентификационном номере респондента соответствуют году выпуска. В результате мы можем сделать предварительное заключение о том, что выпускники, работающие и специалистами и руководителями в сфере рекламы, занимают в данной сети наиболее выгодные позиции, так как располагаются преимущественно в ее центре.

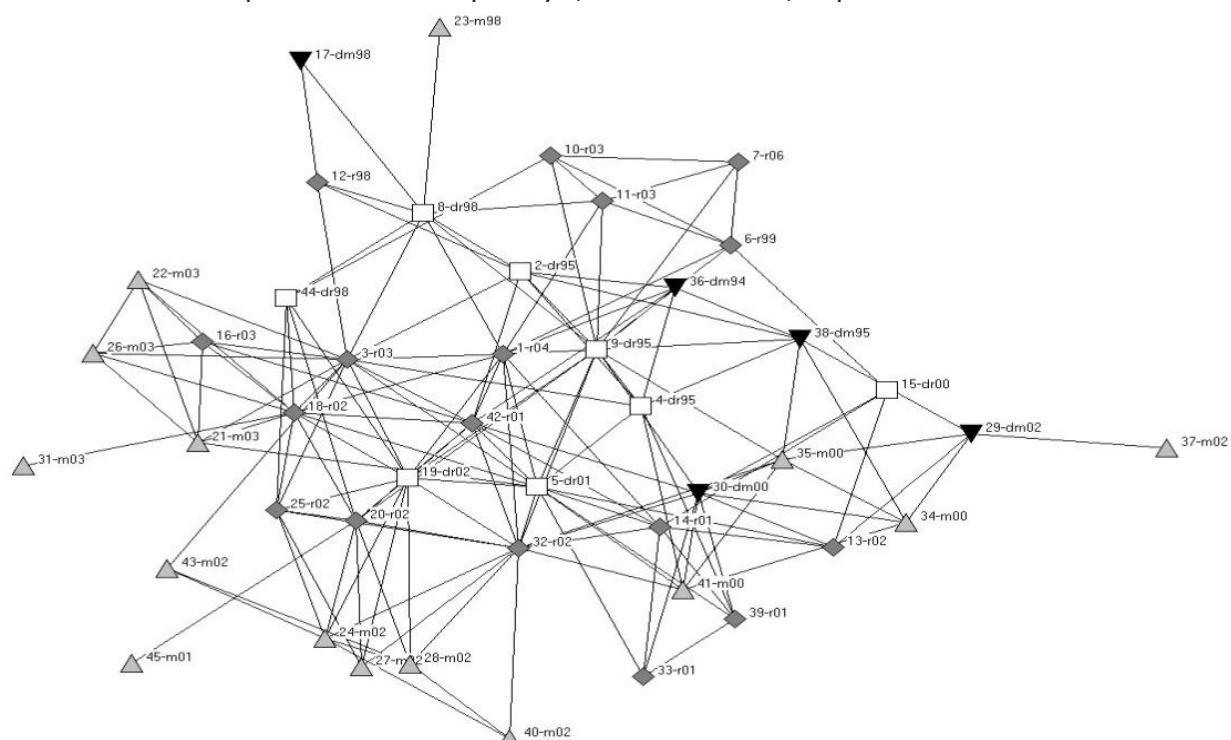


Рисунок 1. Сеть социальных связей социологов

Для анализа сети социальных связей социологов, работающих в сфере рекламы и маркетинга, мы использовали такие количественные структурные показатели как расстояние, диаметр, плотность и центральность (табл. 3).

Таблица 3

Структурные характеристики сети социальных связей социологов

Показатель	Значение
Размер	44
Расстояние	2,4
Диаметр сети	4
Плотность	0,16
Центральность по степени (Degree centrality)	21%
Центральность по близости (Closeness centrality)	28%
Центральность по посредничеству (Betweenness centrality)	15%

Расстояние является одним из самых значимых показателей сетевого анализа. То, как «далеко» акторы сети находятся друг от друга, в значительной степени определяет скорость передачи информации в данной сети. В нашей сети нет изолированных акторов, поэтому кратчайшее расстояние или геодезик можно определить для каждого актора сети, в среднем оно равно 2,4. Диаметр сети, представляющий наибольшее значение геодезика, равен 4. Это говорит о том, что, несмотря на относительно большой размер, сеть является достаточно компактной. Любой актор сети находится от другого на расстоянии не более четырех «шагов». На основании этого, мы можем утверждать, что вероятность быстрого и эффективного распространения информации в изучаемой сети социологов довольно велика.

Плотность сети – показатель, позволяющий оценить отношение числа возможных связей в сети к числу связей, реально в ней присутствующих. Данный показатель позволяет оценить степень сплоченности акторов сети. Показатель плотности сети может принимать значения из интервала [0,1] – чем больше связей присутствует в сети, тем ближе значения показателя к 1. В нашем случае плотность равна 0,16. Таким образом, изучаемую сеть никак нельзя назвать плотной или сплоченной, скорее разряженной.

Центральность – показатель распределения власти и влияния в сети, имеющий три основных разновидности – центральность по степени, по близости и по посредничеству. Показатель центральности по степени для нашей сети равен 21%. Это означает, что социальная сеть социологов на 21% соответствует идеально централизованной сети, в которой все ее члены связаны только с одним актором. В данном случае сеть довольно далека от идеально централизованной сети, и следовательно в данной сети присутствует значительное количество ключевых акторов, которые выполняют посреднические функции. Обобщенные показатели центральности по близости (28%) и посредничеству (15%) так же не являются высокими. Это говорит о том, что и на макро-уровне в данной сети степень концентрации власти и влияния не высока или, другими

словами, в данной сети относительно небольшое количество акторов, пользующихся влиянием на других.

Анализ индивидуальных показателей центральности показал, что в социальной сети связей социологов можно выделить два ярких типа акторов, обладающих специфическими отличиями. К первому типу акторов относятся те, кто работает в сфере маркетинга – у них не много контактов, вследствие чего при взаимодействии с большинством сетевых акторов они вынуждены прибегать к помощи посредников. Ко второму типу акторов относятся те, кто работает в сфере рекламы – они обладают наибольшим количеством контактов и выполняют связующие функции в сети, наиболее активно участвуя в создании и воспроизводстве своего социального капитала. С одной стороны, наличие у них большего количества контактов и плотной сети взаимодействий между собой дает им явные преимущества при поиске необходимой информации и естественно делает их гораздо более значимыми в данной сетевой структуре. С другой стороны, тот факт, что они являются во многих случаях посредниками между замкнутыми группами в сети, дает им возможность контролировать поведение других в случае возникновения у них различных проблем, связанных с исполнением служебных обязанностей либо с принятием решения о смене работы и т. д. В связи с этим наиболее вероятно, что помощь при трудоустройстве или при поиске работника на вакантное место они смогут оказать наиболее эффективно.

Ресурсные данные: анализ графа направленных связей, используемых при трудоустройстве. Согласно полученным данным, 36 из 45 выпускников получили свою первую работу благодаря личным связям, при этом примерно половине из них (23 человекам) вообще не пришлось предпринимать никаких усилий – друзья, родственники, знакомые, преподаватели кафедры обратились к ним с предложением о работе сами. Формальные каналы поиска работы (служба занятости, объявления о вакансиях в СМИ) использовали только 9 выпускников, прекрасно осознавая их весьма низкую эффективность.

Для того чтобы зафиксировать все личные связи, которые были использованы выпускниками при трудоустройстве, мы просили респондентов для каждого места работы, начиная с первого, указать название организации, должность, период работы, кто помог при трудоустройстве в данную организацию, чем, и помог ли сам респондент кому-нибудь в нее устроиться. Помощь в трудоустройстве, которая могла быть получена через личные связи, включала информацию о вакансии, содействие в обучении, рекомендация работодателю и даже создание рабочего места.

Полученные ответы мы зафиксировали в матрице несимметричных направленных связей трудоустройства. При этом следует отметить, что в список личных контактов, используемых при трудоустройстве, вошли не только акторы изучаемой социальной сети, но и выпускники отделения социологии, которые уже выбыли из сети, а также некоторые преподаватели кафедры социологии. На основе полученной матрицы связей с помощью программы NetDraw мы построили направленный граф связей трудоустройства (рис. 2), вершины которого получили следующие условные обозначения: – выпускники, работающие в сфере рекламы (их идентификационные номера начинаются с букв *r* и *rm*); – выпускники, работающие в настоящий момент в сфере маркетинга (*m* и *dm*); +–

преподаватели кафедры, а также выпускники, которые работали в сфере рекламы и маркетинга, но выбыли из сети по различным причинам (P). Последние две цифры идентификационного номера соответствуют году выпуска.

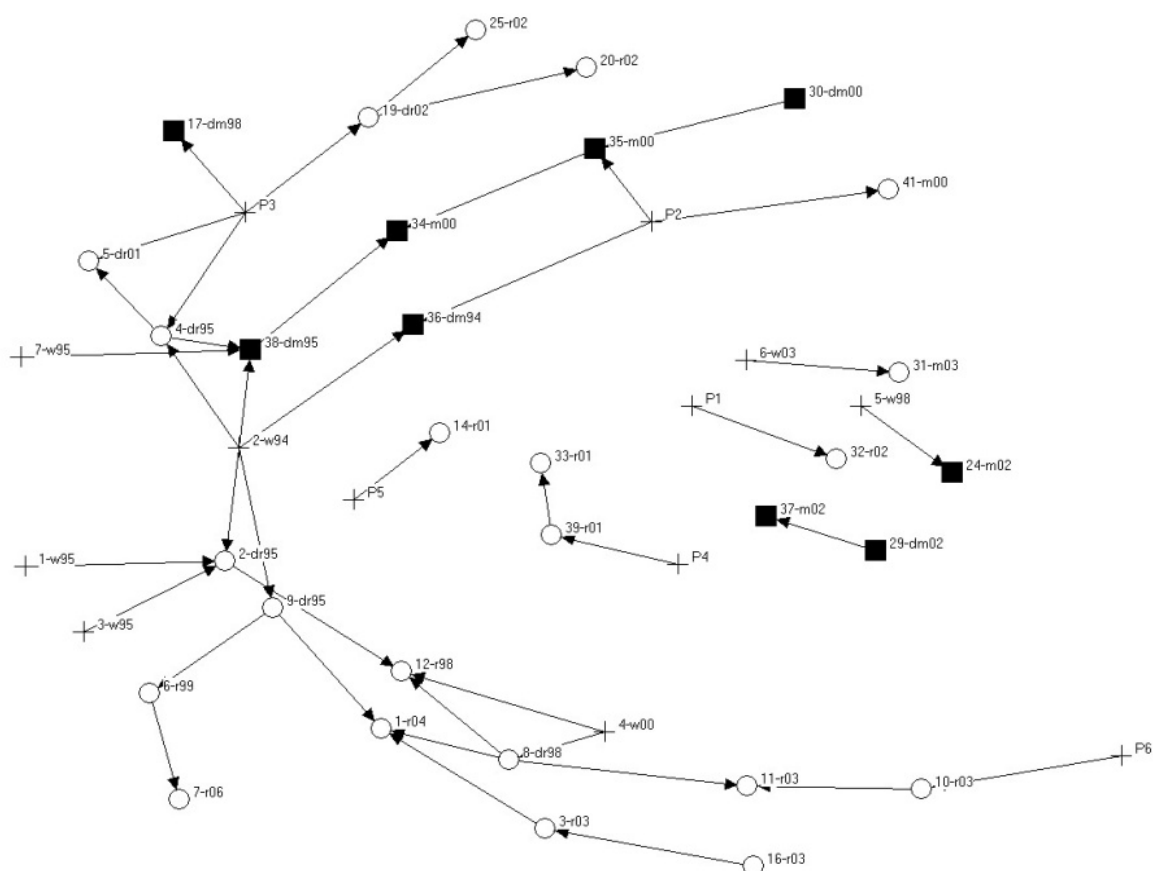


Рисунок 2. Граф направленных связей, которые были задействованы социологами при трудоустройстве

Граф включает в себя 45 вершин. Из них 6 соответствуют преподавателям кафедры социологии, 7 – бывшим участникам изучаемой сети, работавшим определенный период времени в сфере рекламы и маркетинга. Остальные 32 вершины представляют активных акторов сети – социологов, которые на момент исследования активно участвовали в ее создании и воспроизводстве путем принятия и оказания помощи в процессе трудоустройства, причем 19 из них использовали свои личные связи при трудоустройстве однократно, 10 человек – два раза и более, 3 не воспользовались возможностями личных связей сами, но помогли это сделать другим.

Для того чтобы облегчить восприятие и анализ графа связей трудоустройства, были выделены «пути» – последовательности вершин, соединенных дугами, в которых ни одна из вершин и ни одна из дуг не повторяется. Число дуг, составляющих путь, определяет его длину. В результате было обнаружено, что в данном графе преобладают пути с длиной равной двум (17 из 31), связывающие трех акторов, один из которых только оказывает помощь при трудоустройстве, второй принимает эту помощь и оказывает ее сам третьему актору, который помощь только принимает. Немного меньше в данном графе и путей с длиной равной единице – диад (11 из 31). Путь с длиной равной трем не так уж и много по сравнению с другими (3 из 31), но они являются самыми

значимыми для данного графа, и свидетельствуют о наличии не просто поддержки в трудоустройстве, а преемственности в оказании данной помощи. Социальная сеть социологов, работающих в сфере рекламы и маркетинга, выстраивалась в течение тринадцати лет и естественно, что ее состав в течение этого времени постоянно изменялся, но мы видим, что принцип взаимопомощи и поддержки остается неизменным. Показателем этого служит тот факт, что за время существования данной сети образовались пути, связывающие два, три и даже четыре поколения социологов. Наиболее часто старшие коллеги (как выпускники, так и преподаватели) помогают младшим. Особенно четко данную тенденцию можно проследить в путях с длиной равной трем. Вместе с тем, достаточно часто социологи одного выпуска, очевидно нашедшие работу несколько раньше или лучше чем сокурсники, помогают в трудоустройстве друг другу.

Характерной особенностью изучаемых путей является значительное преобладание в них акторов, работающих в сфере рекламы, особенно это касается путей с длиной два и три. Практически все акторы, которые имеют наибольшие индивидуальные показатели центральности, присутствуют в цепочках связей трудоустройства. Благодаря значимости своей позиции они либо сами помогли найти кому-то работу, либо получили помощь. Вполне очевидно, что исследуемая социальная сеть социологов не может воплотить в себе полностью весь социальный капитал ее акторов, но на основании анализа путей графа связей трудоустройства можно сделать вывод о том, что данная сеть является его важной и неотъемлемой частью, выполняя важную функцию установления соответствия между спросом и предложением на рынке труда в сфере рекламы и маркетинга, удачно дополняя тем самым деятельность в данном направлении официальных структур (центров занятости, кадровых агентств и др.).

Реконструкция и анализ сети организаций, в которых работают социологи. Социальную сеть организаций, в которых работают социологи, являющиеся участниками социальной сети, мы получили в результате наложения профессиональных и смешанных связей отдельных акторов, работающих в одной и той же организации (рис. 3). Если в определенной организации работает только один социолог, то мы рассматривали только его контакты как связи данной организации. Условные обозначения: – организации, в которых социологи работают как специалисты по маркетингу; – рекламные агентства.

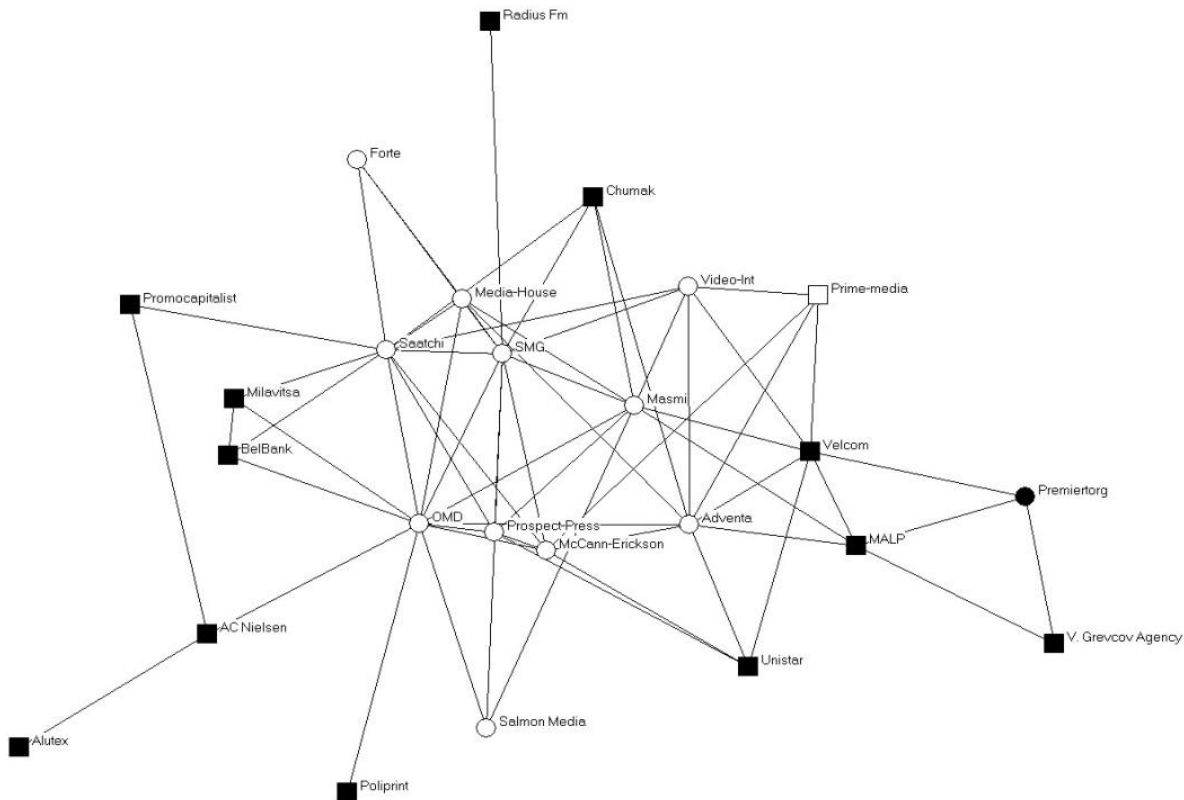


Рисунок 3. Граф ненаправленных связей между организациями, в которых работают социологи

Следуя логике сетевого анализа, позволяющего осуществить многоуровневое исследование социальных структур взаимодействия, можно определить, как социальные связи социологов на микроуровне отражаются на мезоуровне взаимодействий тех организаций, в которых они работают. Вследствие того что в различных организациях работают сотрудники, входящие в одну сетевую структуру, данные организации уже не являются изолированными друг от друга субъектами рынка рекламных и маркетинговых услуг, они имеют реальные возможности эффективного взаимодействия в своей профессиональной сфере, позволяющие делиться важной информацией и снижать риски экономических транзакций.

Таким образом, активно взаимодействуя между собой, специалисты рекламной сферы являются связующими звеньями не только для своей профессиональной сети, но и для сети организаций, в которых работают. Значение этих связей для организаций связано с уменьшением издержек на координацию совместной деятельности, поскольку контракты, формальные правила и бюрократические процедуры заменяются здесь отношениями доверия, усвоенными профессиональными стандартами, этикой общения, то есть теми неформальными общественными нормами, которые передаются традициями и образованием.

Интересно отметить тот факт, что в рамках сети социальных связей организаций происходит постепенное перемещение сотрудников компаний. Многие социологи меняют место работы, переходя из одной организации в другую, обеспечивая тем самым динамику этой сетевой структуры и увеличение количества связей в ней. По данным исследования, такие перемещения осуществили 15 из 45 опрошенных

социологов, все эти перемещения способствовали либо карьерному росту, либо росту их материального благосостояния, так как они переходили на более высокую должность или зарплату, а чаще – и то и другое вместе.

Вместо заключения

Социальные взаимодействия, человеческие отношения составляют основу общественной жизни, которая не видна невооруженному взгляду, но определяет ее функционирование. Исследования социальных сетей позволяют зафиксировать в наиболее четкой форме эту основу и процессы, происходящие в социальной структуре общества, позволяя при этом решать широкий круг социологических задач, связанных с реконструкцией структуры связей, определением ее характеристик, выявлением принципов и механизмов функционирования и мн. др. Так, определение возможностей использования ресурсов социальной сети при трудоустройстве социологов в сфере рекламы и маркетинга при помощи комбинации трех типов данных – атрибутивных, структурных и ресурсных, на наш взгляд, представляет неоспоримые преимущества. Во-первых, позволяет учесть субъективный потенциал действий акторов исследуемой социальной сети, не ограничиваясь объяснением поведения социальных субъектов только их атрибутивными характеристиками. Во-вторых, такого рода перспектива исследования помогает понять, объяснить и социальные механизмы существования сети, и частные способы функционирования каждого актора в данной структуре взаимодействий, благодаря возможности перехода от микро- и мезо- к макроуровню исследования.



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Автоматический сбор информации из открытых Интернет - источников

Никита Анатольевич Найденов

Вычислительный центр
им. А. А. Дородницына РАН (Москва)

Данная работа посвящена созданию инструментов автоматической обработки информационных интернет ресурсов и несет в себе практическую значимость в задачах анализа текста. Во введении обосновывается актуальность выбранной темы, формулируются цель и задачи исследования, указывается объект и предмет исследования. Рассматривается такая задача, как сбор и первичная обработка текстовых данных из новостных источников. Автор предоставляет механизм автоматической обработки большого количества HTML страниц, приводит практическое применение данного подхода на реальных данных открытых ресурсов Государственной телерадиокомпании. Приводится сравнение последовательной и параллельной обработки источников, перечисление достоинств и недостатков таких методов и представление результатов, полученных в ходе выполнения работы.

Введение

Все данные, представленные в глобальной сети Интернет, можно назвать неструктурированными, ввиду индивидуальности и специфичности архитектуры каждого ресурса. В основном, такие данные – это HTML страницы, т.е. текстовые структуры. В настоящее время, в связи с постоянным ростом информации во всемирной паутине, необходимо развитие технологий, позволяющих использовать ее для решения различных производственных задач предприятий и организаций, вследствие чего активно развивается область анализа текстовых данных и неструктурированной информации. Общепринятым является разбиение задачи предварительной обработки данных на 3 этапа: консолидация, трансформация и очистка [1]. Самым трудоёмким этапом является консолидация данных, которая включает в себя сбор данных. Если исследования ведутся с большой выборкой, то для того, чтобы накопить достаточное количество материала, могут уйти недели или месяцы кропотливого труда. Данная работа посвящена разработке методов автоматического сбора информации из открытых интернет источников.

Описание данных

Данные, которые требовалось собрать и обработать, представляют собой статьи в новостных ресурсах. Объектом данных является статья (или новость), у которой есть заголовок, дата появления в СМИ и сам текст сообщения. Таким образом, данные представлены в формате HTML. Общий объем извлеченных данных составил около 220 Мб. Источником

данных являются 43 региональных сайта Государственной Телерадиокомпании.

Постановка задачи

Задачу предварительной обработки данных, как упоминалось выше, можно разложить на 3 этапа: консолидация, трансформация и очистка.

Консолидация – это комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразование к единому формату, в котором они могут быть загружены в хранилище данных или аналитическую систему.

Трансформация – комплекс методов и алгоритмов, направленных на оптимизацию представления и форматов данных с точки зрения решаемых задач и целей анализа. Трансформация данных не ставит целью изменить информационное содержание данных. Ее задача – представить эту информацию в таком виде, чтобы она могла быть использована наиболее эффективно.

Очистка – процесс выявления и исключения различных факторов, мешающих корректному анализу данных: аномальных и фиктивных значений, пропусков, дубликатов и противоречий, шумов и т.д.

В рамках данной работы рассматривается задача консолидации данных. Опишем постановку реальной задачи, которая решалась в процессе работы: требуется собрать новостные материалы из открытых источников за последние полтора года. Как говорилось выше, нас интересуют сайты Государственной Телерадиокомпании. Результатом является отсортированный список новостей с датами их появления и заголовками. Сортировка осуществляется по дате появления новости в СМИ.

Обзор существующих решений

Существует множество систем для выполнения задач извлечения текстовых данных из интернет-источников [2]. Вот некоторые из них: TSIMMIS, WebOQL, FLORID, XWRAP, RoadRunner, Lixto, RAPIER, SRV, WHISK. Последние три направлены на работу с относительно грамматически связными текстами, некоторые привязаны к определенной однотипной структуре данных. В основном, они узко специализированы под конкретные задачи, поэтому использование данных технологий не является уместным в рамках данной работы.

Обработка ресурсов

Для ручной обработки ресурса, в котором за год накопилось около 4000 новостей, занимает огромное количество времени. Задача извлечения информации из интернет-источников была бы гораздо проще, если бы существовал единый стандарт построения сайтов. Но, к сожалению, такие стандарты отсутствуют – все многообразие сайтов и web-страниц объясняется фантазией веб-дизайнеров. Единственное, что их объединяет, – это язык HTML, который определяет внешний вид Интернет-ресурсов, но не может описать его содержание.

Большинство новостных сайтов имеют возможность предоставлять новостную ленту в формате RSS, что позволяет легко использовать существующие обработчики данных в таком виде. Однако в конкретной задаче могут встречаться источники, в которых не предусмотрен такой новостной формат. Таковыми являются ресурсы в нашей работе, поэтому применение существующих RSS-технологий неприемлемо. Несмотря на то, что данные сайты работают независимо, некоторое сходство у них все же есть. Все они информационные источники, которые предоставляют данные в виде новостей. «Новость» в нашем понимании – это объект данных, представляющий собой информацию о заголовке, дате создания и о самом тексте сообщения.

В процессе анализа новостных ресурсов удалось выделить два типа структуры этих источников:

1. Ресурсы с календарем

Многие информационные источники предоставляют доступ к данным при помощи календаря. Т.е. можно выбрать интересующую дату и просмотреть новости за этот день

2. Ресурсы со списком

Такой тип источников не предполагает использование календаря. В таком случае имеется просто список всех страниц с новостями, упорядоченный по дате добавления сообщений.

Несмотря на свое различие в предоставлении данных, у всех ресурсов можно выделить трехуровневую структуру – это календарь с датами (или страница), где каждой дате (странице) соответствует список новостей за этот день. Этот список содержит краткую информацию о событии и ссылку на него. Третий уровень представляет собой полноценную новость, т.е. интересующие нас данные плюс некоторое оформление. У конкретных источников некоторых полей может не быть, или вместо них могут появиться другие. Также блок содержит некоторый «информационный» мусор, т.е. данные, не имеющие отношения к конкретной извлекаемой новости или ее реквизитам, который при извлечении сообщения надо постараться отфильтровать

Необходимо для каждого ресурса обойти все такие новости за интересующий период и собрать их в один список. Основной задачей извлечения данных из сети интернет является: получение определенных фрагментов информации (поля) из указанных HTML документов. В данной работе были использованы известные методы по извлечению полезной информации из текста [4,5].

Введем определение обработчика страниц:

Поисковый робот (web crawler, веб-краулер) - специальная программа, основная задача которой является сканирование веб-страниц с последующей обработкой данных. В дальнейшем будем использовать слово «краулер» для программы-сборщика данных.

Для того чтобы краулер смог обработать информационный источник, ему необходимо предоставить шаблоны ссылок на список с новостями и на сами новости. Разбирая HTML страницу, краулер извлекает ссылочные теги [3] и сопоставляет их шаблону, после чего переходит на нужные страницы и собирает новости. Таким образом, для каждого источника необходимо указать набор параметров для его обработки или, другими

словами, показать краулеру, как обходить массив с новостями и какую полезную информацию извлекать из страницы.

Автоматизация системы

Такой подход имеет несколько недостатков:

- Новому ресурсу систему надо обучать
- Изменение старого ресурса необходимо обнаруживать
- При изменении старого источника систему необходимо переобучать

Избавиться от этих недостатков помогает автоматизация системы. Достигается она путем формализации структуры источников, что позволяет краулеру автоматически находить шаблоны ссылок на страницы. Процесс автоматизации состоит из следующих этапов:

1. Поиск календаря (списка страниц) на сайте
2. Формирование шаблона ссылок на страницы со списком новостей
3. Формирование шаблона ссылок на новости
4. Извлечение даты, заголовка и текста
5. Определение изменения ресурса

Результаты экспериментов

Как говорилось выше, исходные данные представляют собой 43 региональных новостных ресурса Государственной телерадиокомпании. Из них 35 были обработаны автоматически, что показывает применимость данной системы для реальных данных.

Достоинства и недостатки

Главное достоинство такого подхода заключается в том, что скорость обработки одного ресурса практически оптимальна. Также автоматическая настройка системы позволяет применять ее на новостных источниках, которые имеют схожую структуру. Однако есть несколько недостатков: во-первых, необходимо четко вылавливать все исключения и ошибки, потому что при такой обработке можно очень легко потерять данные или пропустить важное исключение в структуре источников; также такой метод очень сильно зависит от характеристик сервера – часто мы имеем дело с серверами, у которых стоят ограничение на количество подключений с одного адреса. В таком случае, необходимо правильно настроить данный инструмент, используя оптимальные параметры.

Объем данных от источников варьировался от 1 Мб до 22 Мб (рис.), время обработки одного источника от 1 до 7 минут (не принимая в расчет сайт с ограничением).

Параллельность обработки одного интернет ресурса позволила сократить среднее время с 20 до 3 минут. К сожалению, ввиду вышесказанных недостатков, не ко всем источникам удалось применить такой подход. Например, сайт ГТРК Ставропольского края имеет достаточно жесткое ограничение на подключения к серверу, поэтому его пришлось обрабатывать последовательно.

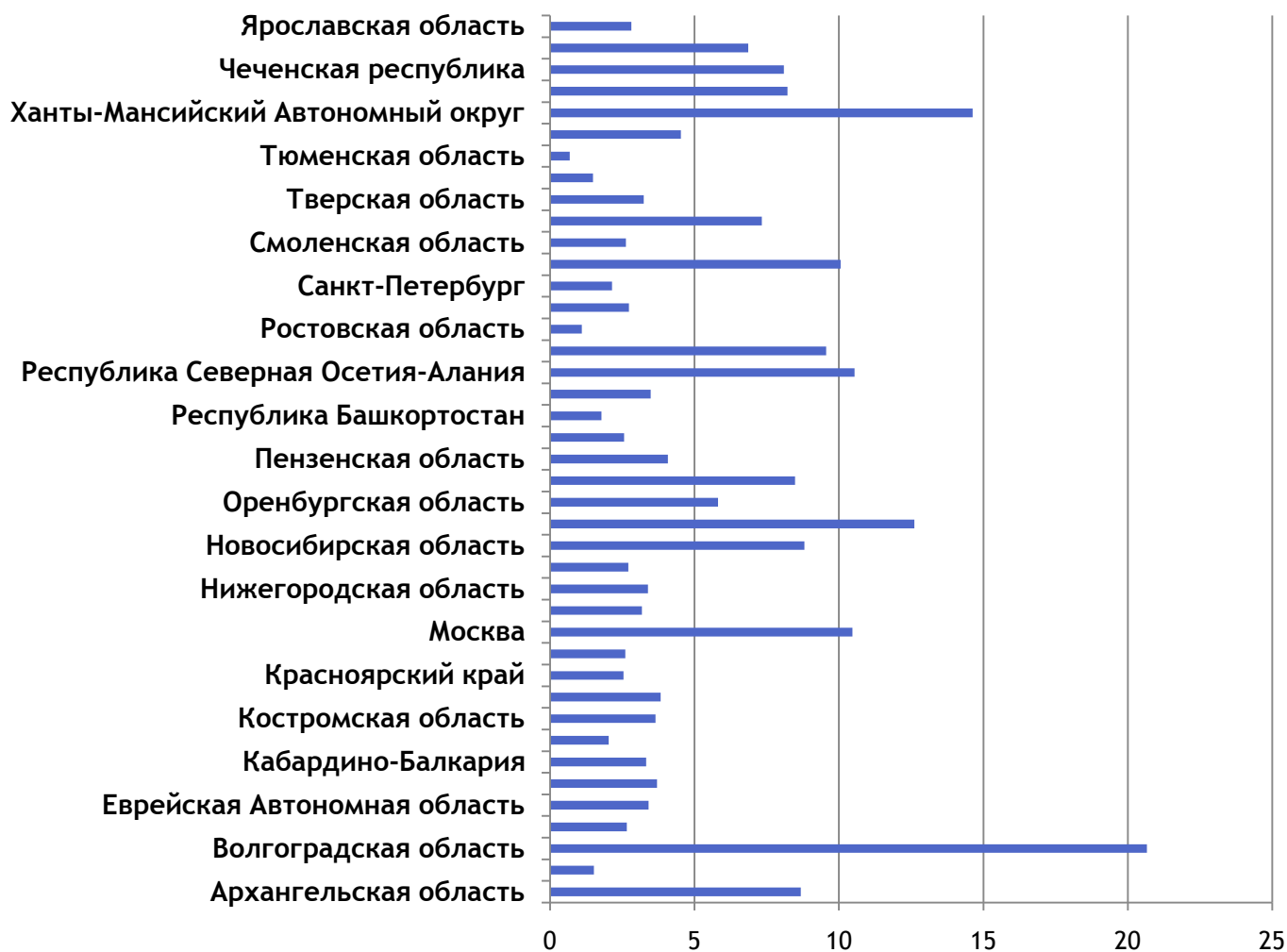


Рисунок. Объем данных от источников в Мб¹

Выводы

- Создана система загрузки отдельных новостей с различных новостных ресурсов
- Предложен формат описания новостного ресурса, подходящий для большинства сайтов
- Для большинства ресурсов происходит автоматическое определение параметров
- Сокращено время добавления в систему нового ресурса
- Сокращены простои в результате изменения ресурсов

Источники:

1. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям // Спб.: Питер, 2009 - 624 с
2. A.H.F. Laender, B. A. Ribeiro-Neto, Juliana S.Teixeria. A brief survey of web data extraction tools // ACM SIGMOD Record 31(2), pp 84-93. 2002
3. W3Schools. – <http://www.w3schools.com>
4. <http://ai-depot.com/articles/the-easy-way-to-extract-useful-text-from-arbitrary-html>
5. <http://sujitpal.blogspot.com/2009/11/extracting-useful-text-from-html.html>

¹ На рисунке подписаны не все регионы, ввиду их большого количества



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Содержательные основания выделения границ Интернет - сетей

Дарья Вячеславовна Просянюк

Национальный исследовательский университет
«Высшая школа экономики» (Москва)

Сегодня Интернет как среда общения, получения и распространения информации, обмена ресурсами и услугами, а также самовыражения стремительно растет, и практика посылает запрос автоматической обработки больших массивов информации. Для поиска информации и принятия решения эксперту, аналитику или простому пользователю уже недостаточно использовать поисковые машины. Для наиболее эффективного решения поставленных задач и поиска необходимой информации нужны новые методы, которые позволили бы пользователю (в широком смысле этого слова) при минимальных усилиях получить необходимую информацию.

Одной из ярких иллюстраций воплощения данной проблемы на практике может быть проблема построения экспертных сетей в Интернете.

Очевидно, что для принятия обоснованных решений необходимо опираться на опыт, знания и интуицию специалистов. В настоящее время все шире применяются различные методы экспертных оценок. Они незаменимы при решении сложных задач оценивания и выбора технических объектов, в том числе специального назначения, при анализе и прогнозировании ситуаций с большим числом значимых факторов - всюду, когда необходимо привлечение знаний, интуиции и опыта многих высококвалифицированных специалистов-экспертов.

Методы экспертных оценок - это методы организации работы со специалистами-экспертами и обработки мнений экспертов, выраженных в количественной и/или качественной форме с целью подготовки информации для принятия решений¹.

Интернет-среда содержит огромное количество хорошо известных, эксплицитно-идентифицируемых сообществ – групп индивидов, объединившихся в формальные группы и разделяющих сходные интересы. Их нахождение и описание не составляет труда. Вместе с тем, существует и не меньшее количество имплицитных сообществ – людей, имеющих сходные интересы, но не объединяющихся вместе, и, возможно, даже незнакомых. Выявление и анализ таких сообществ представляет особый интерес по ряду причин. Сообщества в социальных сетях могут отражать реальные социальные группы; сообщества в сетях цитирования могут представлять связанные статьи на схожие темы; сообщества в метаболических сетях могут представлять циклы или другие функциональные группировки; сообщества во всемирной паутине – страницы на связанные темы. Возможность грамотно идентифицировать

¹ Орлов А.И. Теория принятия решений . Учебное пособие. - М.: Издательство "Март", 2004.

такие объединения поможет понять их свойства и использовать их более эффективно. Следующая причина обусловлена научным интересом. Сообщества репрезентируют социологию сети: их изучение дает представление об интеллектуальной эволюции сети. Другая причина заключается в том, что такого рода сообщества чаще всего концентрируют в себе наиболее ценные и современные информационные ресурсы, необходимые пользователям, интересующимся той или иной тематикой. Четвертая причина состоит в том, что наличие информации о сообществах и их контурах дает возможность для распространения той или иной информации (например, рекламной или идеологической). Наконец, это очень удобный способ работы с экспертным сообществом. Данная проблема особенно актуальна в России, где на сегодняшний день практически невозможно получить независимую информацию от экспертов, не находящихся под влиянием лоббирующих группировок.

Что касается методов выделения границ сетевых сообществ, то традиционным является выделение сообществ в сети путем иерархической кластеризации. Этот метод основан на вычислении силы связей между объектами и поэтапном присоединении (агломеративная кластеризация) или отсоединении (дивизимная кластеризация) объектов. Для вычисления силы связей существует ряд подходов (например, вычисляется общее количество путей между вершинами, или количество независимых путей и пр.).

С другой стороны, существует целый ряд методов, основанных на нечеткой кластеризации объектов. Нечеткая кластеризация зарекомендовала себя как весьма эффективный инструмент анализа данных во многих областях. Например, в биоинформатике она предоставляет возможности для исследования особенностей генной структуры¹.

Наконец, говоря о виртуальных сетях невозможно не упомянуть сравнительно молодой, но динамично развивающийся социолингвистический подход. Одно направление методов данного подхода анализирует синтаксическое содержание файлов, определяя степень их близости².

Другое направление центром своего внимания видит не столько содержание файлов, сколько их взаимные связи посредством гиперссылок³.

Недостатком существующих подходов является тот факт, что они развивались на стационарных данных. Интернет же является динамической средой, поэтому требует особых подходов и методов анализа. Более того, многие подходы подразумевают, что исследователь изначально осведомлен о границах/совокупности исходных объектов. В реальности же зачастую стоят задачи в определении этих границ.

В связи с этим особое внимание следует уделить двум направлениям исследований – поисково-разведывательному (поиск, описание возможных методов решения задач) и их сравнительному

¹ M. Sato, Y. Sato and L.C. Jain "Fuzzy Clustering Models and Applications," Physica Verlag, Heidelberg, New York 1997.

² S. Brin, J. Davis, H. Garcia-Molina. Copy Detection Mechanisms for Digital Documents. Proceedings of the ACM SIGMOD Annual Conference, San Francisco, CA, May 1995.

³ K. Bharat, Monika R. Henzinger Improved Algorithms for Topic Distillation in a Hyperlinked Environment, 1998.

анализу (подробный анализ методов, сопоставление и подбор к конкретным задачам).

Поисково-разведывательное направление

Образование «сгущений» или сообществ (то есть совокупностей узлов с большой плотностью связей друг с другом и с низкой с остальной частью сети) – это естественная характеристика сетевых структур. Конкретные причины образования сообществ могут зависеть от типа сети, но само это свойство является неотъемлемой чертой любой сети, будь то сеть социальная, биологическая или компьютерная. Обнаружение и определение границ таких сообществ является главным шагом на пути к пониманию топологии сети.

Анализ существующих подходов к данной проблематике показал наличие достаточно большого количества методов и способов определения границ сетевых сообществ.

Нами была произведена классификация и подробный разбор основных алгоритмов (см. Таблица 1).

Как видно, все алгоритмы выделения сетевых сообществ могут быть разделены на два большие класса – математические и социолингвистические.

Математические алгоритмы представляет собой широкий класс алгоритмов, разрабатывавшихся не только на виртуальных (шире – текстовых данных), а на более широком круге данных – офф-лайн социальные сети, сети цитирования и соавторства, метаболические и пищевые сети.

Социолингвистические алгоритмы имеют более узкую направленность – будучи разработанными и тестируемыми на текстовых документах, они применимы для анализа текстовой информации (офф-лайн и он-лайн тексты, сайты, гиперссылки и пр.), а также другой информации, теоретически разложимой на совокупности символов: музыка, графика, видео, аудио, базы данных.

Таблица 1

Классификация алгоритмов выделения сетевых сообществ

Алгоритмы выделения сетевых сообществ					
Математические				Социолингвистические	
Иерархические:	Основанные на четкой кластеризации (представители)	Основанные на нечеткой кластеризации (представители)	Основанные на принципах термодинамики (представители)	Анализ синтаксического содержания файлов (представители)	Анализ связей файлов (представители)
Агломеративные	Besagni D. Boudourides A. Broder A. Polanco X. Garton L. Kumar R. Law J. Rip A. Roche I.	Bezdek J. C. Deva A. Gath I. Gustafson D. E. Jain L. C. Kessel W. C. Palm R. Rusrini E. Sato M. Valdes J.	Zakharov P.	Baker B. S. Brin S. Broader A. Davis J. Garcia_Molina H. Glassman S. C. Heintze N. Manasse M. S. Manber U. Shivakumar N. Zweig G.	Bharat Carrière Henzinger M. R. Kleinberg
Дивизимные	Djidjev H. N. Girvan M., Newman M., Osorio	...			

Начнем анализ с представителей математического подхода – агрегативных алгоритмов, основанных на четкой кластеризации.

Применение кластеризации и картографирования веб-сайтов для обнаружения неявных сообществ¹

Задача: выявление, анализ и визуализация скрытых ассоциаций сайтов.

Возможности применения: выявление имплицитных он-лайн сообществ, связанных веб-документов и сайтов.

Статья описывает создание метода анализа ассоциаций веб-сайтов. Этот метод использует связи веб-сайтов для получения представления о структуре ассоциаций. Цель заключается в проведении анализа неявных ассоциаций. Для реализации этого авторы предлагают перейти с совместного анализа слов на совместный анализ сайтов. Сайты считаются связанными, если они оба связаны с третьим сайтом. Авторы предлагают разработанный ими коэффициент значимости ассоциаций, а также картовые и кластерные техники.

Данные, на которых основана эта работа, были собраны в январе 2001 года в Институте компьютерных технологий в г. Патры, Греция, в рамках проекта EICSTES. Первоначально набор данных состоял из 1064 веб-сайтов университетов 22 европейских стран-участниц Европейского Союза. Для каждого научного сайта поисковая система AltaVista определяла количество ссылок с сайта на другие сайты, а также количество внутренних ссылок (то есть ссылок сайта на самого себя). Таким образом была получена квадратная матрица $N \times N$ (где N – количество сайтов) и $D(i;j)$ – количество ссылок сайта i на сайт j и $D(i;i)$ – количество внутренних ссылок сайта i . Но в конечном варианте первоначальный набор данных был снижен до 791 университетского веб-сайта из 15 европейских стран.

Алгоритм. Задача web mining, как называют свой подход авторы, состоит из четырех основных этапов: нахождение ресурсов, отбор информации, обобщение и анализ. Два последних этапа описаны в данной работе.

Представленный в статье метод может быть разложен на четыре основных этапа:

1. Трансформация исходной матрицы данных в ассоциативную матрицу;
2. Расчет коэффициента ассоциации;
3. Переформирование сети ассоциаций в кластеры;
4. Размещение полученных кластеров на двумерной карте.
5. Программное обеспечение, используемое в данном исследовании называется SDOC.

¹ X. Polanco, Moses A. Boudourides, Dominique Besagni, Ivana Roche Clustering and Mapping Web Sites. For Displaying Implicit Associations and Visualizing Networks, 2001.

Масштабируемый многоуровневый алгоритм для кластеризации графов и обнаружения структуры сообществ¹

Задача: идентификация сетевых сообществ посредством кластеризации.

Возможности применения: распознавание сообществ в сетевых структурах различной природы: социальные он-лайн и офф-лайн сети.

Проблема идентификации сетевых сообществ обычно рассматривается с позиции теории графов и кластеров (graph clustering, GC), где вершины представляют отдельные объекты, а ребра описывают отношения между ними. Сообществами тогда считаются подграфы, с наибольшим количеством ребер внутри, и наименьшим – вне (с другими подграфами). Данное исследование представляет новый подход к решению данной проблемы.

Разработанный алгоритм тестировался на нескольких наборах данных. Для сравнения эффективности алгоритма с эффективностью уже известных алгоритмов Ньюмана-Гирвана² и Клозета-Ньюмана-Мура³ использовались случайно сгенерированные графы. Для оценки степени работоспособности и эффективности алгоритма использовались реальные данные домена nd.edu, данные о связях в футбольной команде одного из американских колледжей и в карате клубе Zachary.

Алгоритм. Предлагаемый алгоритм разделения графа состоит из следующих фаз:

1 фаза – грубое разделение – coarsening phase.

Исходный граф G разделяется на подграфы и каждый из них заменяется одной вершиной, а множество ребер, связывающих эти подграфы – одним ребром. Вес каждой новой вершины (и, соответственно, ребра) равен сумме весов вершин (и, соответственно, ребер), которые они заменили. Таким образом деление графа повторяется несколько раз до тех пор, пока его размер не станет сравнительно малым. Пусть $G_0 = G, G_1, \dots, G_l$ – итоговая последовательность графов.

2 фаза – разделение – partitioning phase.

Граф G_l разделяется на две части при помощи любого доступного метода (например, спектральное разделение по алгоритму Кернигана – Лина⁴).

3 фаза – деликатное разделение и очищение – uncoarsening and refinement phase.

Данная фаза является наиболее важной в алгоритме, так она в большей степени именно она определяет уровень его точности и эффективности, поэтому она будет описана несколько более подробно. Планируемое разделение G_l равно G_{l-1} . Так как вес каждой вершины G_l равен сумме весов соответствующих вершин G_{l-1} , то часть G_{l-1} будет уравновешена, если часть G_l и отрезки обеих частей будут иметь одинаковые веса.

¹ H. N. Djidjev A scalable multilevel algorithm for graph clustering and community structure detection, Los Alamos National Laboratory, Los Alamos, NM 87545.

² M. Newman and M. Girvan. Finding and evaluating community structure in networks, Phys. Rev. E 69, 026113, 2004.

³ A. Clauset, M. Newman and C. Moore. Finding community structure in very largenetworks, Phys. Rev. E 70, 066111 (2004).

⁴ Kernighan B. W. and Lin S. An efficient heuristic procedure for partitioninggraphs, The Bell System Technical Journal, 1970).

Однако, поскольку в G1-1 входило большее количество вершин, чем в G1, она имела большее количество степеней свободы и, следовательно, возможно отделить часть G1-1 с тем, чтобы снизить её размер (длину кратчайшего пути).

Алгоритм Кернигана – Лина описывается набором итераций, каждая из которых состоит в перемещении вершин из одного подграфа в другой.

Реализация данного алгоритма возможна в программном пакете METIS.

Для сравнения эффективности данного алгоритма и алгоритма Ньюмана-Гирвана был сгенерирован произвольный граф с 128 вершинами и 4 сообществами из 32 вершин каждое. Предполагаемая степень каждой вершины равна 16, но внешняя степень (то есть ожидаемое количество смежных вершин, принадлежащих к другому сообществу) изменяется от 1 до 8. Следовательно, наибольшие значения внешних степеней присущи вершинам, принадлежащим слабосвязанным кластерам.

Данный алгоритм оказался эффективнее алгоритма Ньюмана-Гирвана при любых значениях внешних степеней вершин.

При сравнении с алгоритмом Клозета-Ньюмана-Мура (который отличается от алгоритма Ньюмана-Гирвана более короткими сроками реализации), метод также показал превосходные результаты.

Также алгоритм был протестирован на нескольких наборах реальных данных. Во всех случаях алгоритм продемонстрировал практически безошибочные результаты распознавания сообществ.

Итак, исследование представляет собой разработку нового подхода к кластеризации графов. Алгоритм показал свою эффективность как при тестировании на данных с уже известной структурой, так и при сравнении с уже известными алгоритмами.

Недостатки: Основным недостатком этого подхода является отсутствие описания реализации подстадий (фаза 1 – грубое разделение – «Исходный граф G разделяется на подграфы ...») Как? Очевидно, что для решения данной задачи может быть предложено несколько вариантов (которые, вероятнее всего, способны составить отдельную стадию исследования). Очевидно, что от выбора конкретного способа зависят результаты, а, следовательно, и дальнейшая реализация алгоритма. Авторам следовало бы указать возможные способы разделения графа на подграфы, или, как минимум, способ, использованный ими.

«Быстрый» алгоритм Ньюмана для распознавания структуры сообществ в сетях¹

Задача: выявление кластеров в сетевой структуре.

Возможности применения: распознавание сообществ в сетевых структурах различной природы: социальные он-лайн и офф-лайн сети.

Авторы предлагают алгоритм, основанный на итеративном отсоединении элементов с наибольшей центральностью. Данный подход был опробован на различных видах сетей, таких как e-mail сообщения, социальные сети животных, объединения джазовых музыкантов, геномные разработки и пр.

1 M. E. J. Newman Fast algorithm for detecting community structure in networks, 2003.

Алгоритм. Алгоритм основан на идее «модулярности». В любой сети алгоритм формирует некоторые объединения вершин, вне зависимости от того существуют ли они на самом деле. Для определения качества решения используется специальная функция – «модулярность», рассчитываемая по формуле:

$$Q = \sum_i (e_{ii} - a_i^2)$$

e_{ii} – это все ребра, соединяющие каждую вершину

кластера i с другими вершинами этого же кластера,

a_i – это все ребра, выходящие из каждой вершины кластера i и связывающие её с вершинами кластера j .

Таким образом, в ситуации случайного объединения вершин $Q=0$ и в этом случае есть основания говорить об отсутствии выделенных сообществ. Величина Q изменяется от 0 до 1 и на практике её значения, превышающие 0,3 свидетельствуют о наличии сообществ.

Однако проблема максимизации значения Q обычно является слишком затратной для решения и на практике пригодна лишь для случая наличия 20-30 вершин. Для решения этой проблемы авторы предлагают следующий оптимизирующий алгоритм. По своей принадлежности он попадает в категорию аггломеративных иерархических кластерных методов. На первом этапе каждая вершина является отдельным сообществом. На следующем этапе вершины объединяются в пары. Цель – наибольшее увеличение (или наименьшее снижение) значения Q . Процесс реализации алгоритма может быть визуализирован в помощью дендрограммы. «Срезая» дендрограмму на разных уровнях мы имеем возможность варьировать количество сообществ.

Очевидно, что для увеличения значения Q необходимо рассматривать только те сообщества (вершины), между которыми существуют ребра. Изменение функции Q рассчитывается по следующей формуле:

$$\Delta Q = e_{ij} - e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j), \text{ где}$$

e_{ij} – половина ребер, соединяющих кластеры i и j ,

e_{ji} – другая половина ребер, соединяющая кластеры j и i .

Недостатки: одним из значимых недостатков алгоритма авторы называют значительные временные затраты для его реализации.

Структуры сообществ в биологических и социальных сетях¹

Задача: выявление границ сообществ в сетевых структурах.

Возможности применения: идентификация сообществ в сетевых структурах различной природы: социальные сети, сети цитирования и соавторства, пищевые сети животных, нейронные сети, компьютерные и он-лайн сети.

Традиционным методом распознавания сообществ в сети является иерархическая кластеризация – пошаговое объединение наиболее сильно связанных узлов. Для расчета силы связей предлагаются различные методики. Однако такой способ имеет ряд недостатков – игнорирование наиболее периферийных элементов. Также, при тестировании этого метода на сетях с уже известной структурой и границами сообществ он показал весьма скромные результаты. Поэтому авторами предлагается альтернативный подход. Вместо определения наиболее близких узлов,

1 M. Girvan, M. E. J. Newman. Community structure in social and biological networks, 2002.

метод предлагает выявлять наименее близкие элементы сети. То есть вместо того, чтобы конструировать сообщества, присоединяя узлы, метод конструирует их, пошагово отсоединяя от исходного графа.

В исследовании используются 2 типа данных: сетевые структуры с уже известными границами сообществ и сети, в которых сообщества еще не были распознаны. В первую группу входили искусственно сгенерированное множество графов, данные о социальных связях членов каркте клуба, а также социальная сеть американской футбольной команды. Во вторую группу входили данные о соавторстве членов исследовательского центра в Санта Фе (Нью Мексико).

Алгоритм. Предложенный алгоритм выглядит следующим образом:

1. Расчет посредничества центральности (betweenness) для всех ребер сети (Авторы вводят понятие посредничества центральности ребра - количество кратчайших путей между парами вершин, которые проходят параллельно (рядом, в одном направлении) этому ребру.);
2. Удаление ребер с наибольшим значением посредничества центральности;
3. Пересчет посредничества центральности для ребер, связанных с удаленными;
4. Возвращаемся ко второму шагу и продолжаем до тех пор, пока не останется ни одного ребра.

Для расчета посредничества центральности ребер авторы предлагают использовать быстрый алгоритм Ньюмана, описанный выше.

Итак, вышеописанный метод выделения сетевых сообществ опирается на алгоритм Ньюмана и предлагает изначально определять периферию сети. При тестировании на сети с изначально известной структурой метод показал блестящие результаты, практически безошибочно её распознав. В случае с неизвестной топологией сети результаты метода подтвердили изначальные предположения и послужили толчком для дальнейших изысканий. Так, например, при исследовании научного сообщества были выделены 2 сообщества ученых: по сходной исследовательской тематике и по сходной методологии исследований.

В дальнейшем авторы предлагают усовершенствовать метод, введя в анализ веса и направленные графы. Также возможно применить разработанный метод в исследовании других типов сетей, например, нейронных или виртуальных.

Недостатки: Недостатком данного алгоритма являются (по мнению самих авторов) значительные временные затраты на его реализацию.

Интерпретация результатов нечеткой кластеризации с помощью их виртуального представления с использованием генных микрочипов¹

Задача: нечеткая кластеризация объектов, визуализация и анализ результатов кластеризации.

Возможности применения: нечеткая кластеризация и анализ структуры комплексных, неполных и неточных данных: медицинских, биологических, он-лайн.

1 Valdes, J. Interpreting Fuzzy Clustering Results With Virtual Reality-based Visual Data Mining: Application to Microarray Gene Expression Data. March 2004.

Как показывают многие исследования, нечеткая кластеризация является весьма эффективным инструментом анализа данных. Данная работа предлагает алгоритм нечеткой кластеризации разнородных структур и демонстрацию его реализации на виртуальных данных.

Виртуальная реальность представляет достаточно простые, но в то же время мощные возможности для понимания и интерпретации комплексных, неполных и неточных данных. Подход, описанный в данной статье, апробировался на базах медицинских данных о генах болезни Альцгеймера и лейкемии.

В изучении болезни Альцгеймера было взято 23 пробы больных и здоровых людей. Эти пробы были описаны 9600 генами. В результате работы алгоритма были получены 2 четких и 2 нечетких кластера, $m = 4$, коэффициент разделения равен 0,571556, множество геометрий визуального пространства $G = \{\text{сфера, конус, куб}\}$.

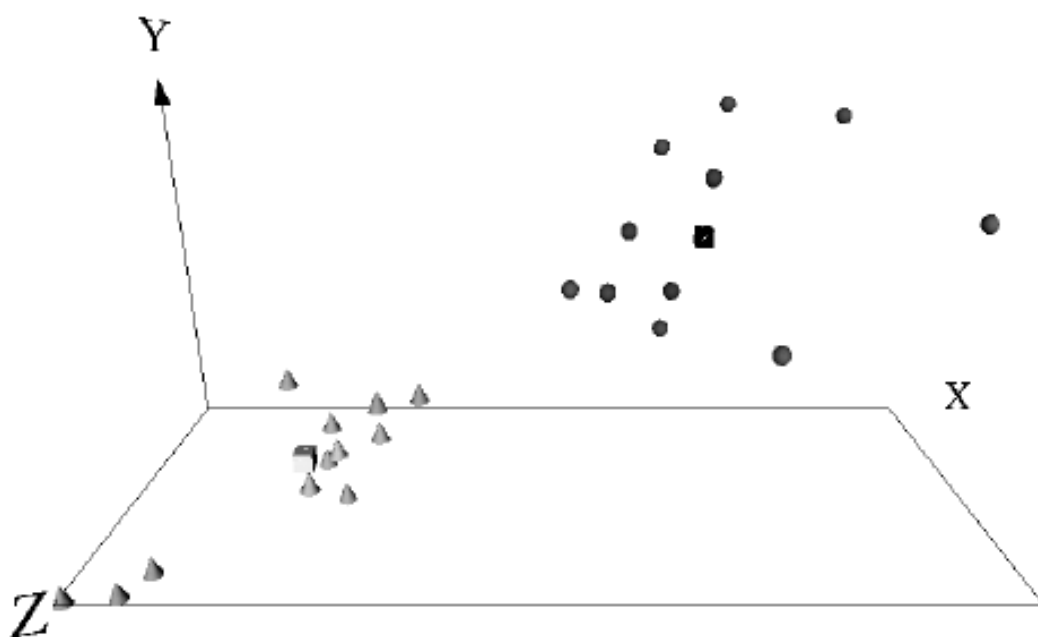


Рисунок. Визуальное представление получившихся кластеров.

Сферы и конусы представляют четкие кластеры (гены болезни Альцгеймера (белый цвет) и нет (черный цвет), кубы представляют расположение центров нечетких кластеров (серый цвет).

Подобные результаты были получены и в случае исследования лейкемии.

Алгоритм. Цель нечеткой кластеризации – формирование кластеров объектов, основанное на их сходстве, которое определяется по их атрибутам, а также по заданному формальному критерию оценки сходства (различия). При нечетком разделении n объектов на K кластеров кластерная структура может быть описана матрицей $U = (u_{ik})$ размера $n \times K$, где u_{ik} принадлежит $[0,1]$, $i = 1, \dots, n$; $k = 1, \dots, K$ и необходимо, чтобы $\sum_{k=1}^K u_{ik} = 1$. u_{ik} представляет принадлежность каждого объекта к кластеру. Значение близкое к 1 свидетельствует о сильном сходстве между объектами, близкое к 0 – о слабом сходстве. Такой подход демонстрирует классические основы нечеткой кластеризации, так как один объект может всецело принадлежать одному кластеру или же –

частично нескольким. Мерой оценки качества кластерного решения служит сумма внутрикластерных дисперсий. Достижение хорошего кластерного решения подразумевает ее минимизацию.

Классический алгоритм продолжается последующей аппроксимацией изначально оцененных центров:

Затем членство каждого объекта определяется по специальной формуле.

Проблема оптимизации J_m довольно сложна для решения. Обычно, для гарантии проверки кластерного решения в дополнение к J_m используются дополнительные коэффициенты – коэффициент разделения F_c и коэффициент энтропии H_c .

Итак, данный подход позволяет отображать в виртуальной реальности одновременно четкие и нечеткие кластеры. Возможности, предоставляемые данным подходом, практически бескрайние, поскольку существует неисчислимо множество различных мер близостей, разностей и расстояний.

Данный подход является единственным представителем подкласса математических алгоритмов и предполагает использование законов физики (явление диффузии, закон сохранения массы) для решения социологических задач - распознавания сообществ в сети.

Термодинамический подход для распознавания сообществ в комплексных сетях: Исследование Живого Журнала¹

Задача: выявление (очерчивание границ) крупных сетевых сообществ

Возможности применения: выявление сообществ в социальных (в частности, он-лайн) социальных сетях.

Как известно, одной из основных характеристик Живого Журнала является возможность «дружбы» пользователей. Каждый пользователь Живого Журнала имеет свой список друзей, куда он добавляет других пользователей (исходящие связи), а также он может быть добавлен в список друзей другого пользователя (входящие связи). Если пользователь А присутствует в списке друзей пользователя Б, и имеет пользователя Б в списке своих друзей, то такая связь является взаимной. Для данного исследования собирались данные о пользователях Живого Журнала, включающие их Ники (виртуальные имена), входящие, исходящие и взаимные связи. Время сбора данных составило 14 дней. Количество исследуемых пользователей – 3746264. Среднее количество исходящих связей $K_{исх} = 15,91$, входящих $K_{вх} = 16,07$, среднее соотношение входящих и исходящих связей ($K_{вх}/K_{исх} = 1,157$). Средний коэффициент кластеризации (который представляет собой отношение связей друзей пользователя к максимально возможному количеству связей между ними) на данных равен 0,3302, что свидетельствует о высокой степени пользовательской кластеризации (для сравнения, коэффициент кластеризации случайного направленного графа с теми же параметрами, что и исследуемый равен $4,24 \times 10^{-6}$).

1 Zakharov P. Thermodynamic approach for community discovering within the complex networks: LiveJournal Study, 2008.

Алгоритм. Метод обнаружения сетевых сообществ, предлагаемый в данном исследовании, основан на принципах термодинамики. Основная идея – имитация процесса диффузии в комплексных сетях как в многомерных «пористых» системах с направленными связями, подчиняющихся физическим законам. Процесс диффузии инициируется в одном из узлов путем добавления в него «виртуальных чернил», распространяющихся затем по другим узлам. Сильно связанные между собой узлы (другими словами – сообщества) идентифицируются по одинаковому количеству чернил в них. В этом смысле данный метод может быть отнесен к категории аггломеративных методов.

Итак, прежде всего, предположим, что узел имеет неограниченную емкость. Прямой поток из узла А в узел Б возможен только если между А и Б есть направленная связь и $\phi_A > \phi_B$ (где ϕ_A и ϕ_B – это значения концентраций «чернил» в узлах А и Б). Сетевые связи в этом случае представляются трубками, направленные связи – трубки, распространяющие «чернила» только в одном направлении.

Теперь представим себе кластер узлов, плотно связанных друг с другом, но имеющий всего несколько внешних связей с остальной частью сети. Тогда диффузия чернил будет происходить сравнительно быстро внутри этого кластера. Небольшое же количество внешних связей ведет к эффекту «бутылочного горлышка» - сравнительно медленной скорости обмена чернилами узлов кластера с внешними узлами. Таким образом, члены одного кластера будут иметь одинаковое количество «чернил», измерение которого позволит идентифицировать эти кластеры.

Данный метод не предназначен для разбиения сети на маленькие совокупности, он помогает распознать сравнительно большие и значимые кластеры. Используя логику и терминологию Ньюмана, авторы предлагают для оценки качества решения следующий коэффициент:

$$K_i = \frac{e_{ii}}{\sum_j e_{ij}} = \frac{e_{ii}}{b_i}, \text{ где}$$

e_{ij} – доля ребер сети, связывающих кластеры i и j ,

$\sum_j e_{ij} = b_i$ - доля ребер сети, имеющих начало в узлах сообщества i .

Таким образом, этот параметр оценивает долю внутренних связей сообщества i среди всех связей, исходящих из сообщества i .

Для тестирования метода было решено использовать несколько разных начальных узлов (юзеры doctor_livsy – писатель Сергей Лукьяненко и future_visions). В результате было обнаружено 2 большие сообщества, сравнительно слабо связанные между собой – русскоговорящий сегмент Живого Журнала и остальная часть сети.

Недостатки: Что такое «виртуальные чернила»? Из статьи этого не ясно. Сам автор так отвечает на этот вопрос: «виртуальные "чернила" - это просто абстрактная субстанция, необходимая для моделирования диффузии в подобной сети. Однако, информацией в моей конкретной модели она быть не может, поскольку для нее не действует закон сохранения массы. В моем блоге я давал объяснение принципа работы алгоритма: "Объясняю на пальцах, то есть на деньгах, что есть эта самая термодинамическая дистанция: дал я [info]doctor_livsy некую сумму денег

в рублях и сказал: распределяй каждый день (допустим) деньги между своими друзьями так, чтобы у вас у всех было поровну. Потом пусть друзья тоже присоединяются и раздают своим друзьям, а дальше - их друзья и т.д. То есть каждый день все у кого есть деньги делятся с теми друзьями, кто в них нуждается. В определенный момент я говорю стоп и проверяю сколько у кого денег. Так вот, разница в накоплениях и определяет расстояние на карте. Допустим теперь [info]future_visions одновременно раздает другую валюту - доллары, например. Тогда их сумма определяет вторую координату. Посмотрите теперь, что получается - поскольку российский сектор хорошо связан внутри, но слабо соединен с остальным и лишь малая часть рублевой массы уходит вовне, в то время как среди РЖЖ юзеров рубли распределены почти равномерно, а с долларами получается обратно - они слабо приходят в российский сектор, но быстро распределяются внутри него. Из-за этого и получается отдельный остров».

Следующим недостатком является слишком высокая степень вовлеченности исследователя: для определения границ сети ему самому необходимо предпринимать определенные весьма затратные действия: распространять «чернила». Более того, непонятно по какому принципу определять инициатора (начальный узел добавления «чернил»), и что зависит от этого инициатора?

Следующие статьи представляют социолингвистическое направление анализа.

Обнаружение сходных файлов в крупной файловой системе¹

Задача: распознавание текстовых файлов из одного ресурса или содержащих одинаковые фрагменты текста.

Возможности применения: Основная цель метода – распознавание текстовых файлов из одного ресурса или содержащих одинаковые фрагменты текста. Два файла считаются сходными, если они содержат значительное количество общих цепочек знаков. Файлы не обязательно должны быть одинакового размера, один может включать другой.

Метод предлагает новый подход, основанный на понятии так называемых «примерных отпечатков». «Примерный отпечаток» представляет собой компактную репрезентацию файла такую, что «отпечатки» двух схожих файлов будут схожими (не обязательно одинаковыми), а «отпечатки» двух разных файлов будут разными. Алгоритм (исходное название которого *sif* – Прим. авт.) может быть реализован двумя способами: все-против-всех и один-против-всех). В первом случае алгоритм выбирает все сходные файлы из заданного множества и выявляет степень их схожести. Второй способ сравнивает один заданный файл с предварительно сформированным каталогом всех других файлов и очень быстро (напр., 3 сек. для 4000 файлов размером 60 Mb) определяет те файлы, которые похожи на заданный. В обоих случаях схожесть может быть установлена, даже если сходный фрагмент составляет всего 25% размера меньшего файла.

1 U. Manber. Finding similar files in a large file system. Proceedings of the 1994 USENIX Conference, pp. 1-10, January 1994

Авторы предсказывают несколько возможных применений разработанного ими алгоритма *sif*. Это значительно облегчает выполнение задач так называемому файл-менеджменту (обнаружение разных версий одних и тех же программ и статей), помощь в системном администрировании организаций, разработка поисковых механизмов в Интернете, сжатие сходных файлов. Также данный метод может быть полезен простым пользователям, которые, например, используют несколько компьютеров (положим, домашний, рабочий и портативный). Преподавателям и писателям он может помочь выявить плагиат при написании работ, политическим деятелям – сходные фрагменты официальных писем и деловой корреспонденции и пр.

Альтернативный подход выявления схожести файлов был предложен Б. Бейкер. Она предлагала считать два файла схожими, если один может быть получен из другого изменением параметров слов. Автор называла такой способ проверки на схожесть параметрическим согласованием и предлагала несколько алгоритмов для выявления схожести файлов.

Алгоритм. Для выявления степени схожести файлов предлагается формировать «отпечатки» файлов, причем несколько для каждого (однако не стоит в качестве оных использовать части текста).

Для создания «отпечатков» текстов предлагается 2 способа. Первый из них – якорный. Якорем называется последовательность символов. При начале каждой процедуры сравнения файлов используется постоянное множество якорей. Если два файла содержат идентичный фрагмент, и этот фрагмент содержит якорь, тогда последовательность символов около якоря также идентична. Например, предположим, что якорем мы выбрали последовательность символов *арак*. Находим все файлы, содержащие последовательность символов *арак*. Например, это файлы, содержащие слово *характеристика*. Затем мы внимательно изучаем каждый текст на предмет содержания в нем последовательности символов, включающих *арак*. Выбираем фиксированное число этих последовательностей, скажем, 50. Эта контрольная сумма и будет «отпечатком» файла. Такие «отпечатки» будут сформированы для всех файлов, содержащих эти 50 единиц последовательностей. В случае, если в файле нет 50-ти последовательностей символов, содержащих *арак*, «отпечаток» для него не будет сформирован. Тонкость такого подхода состоит в использовании нескольких якорей для описания файлов однообразным способом.

Второй способ создания «отпечатков» текстов представляется авторам более простым и удобным. Его идея состоит в формировании «отпечатков» из всех возможных подстрок определенной длины и дальнейшем выборе подмножества этих «отпечатков» на основании их весов.

Задача выбора «отпечатка» имеет несколько решений. Простейшее из них состоит в выборе того, чьи последние k бит равны 0. Приблизительно 1 «отпечаток» будет выбран из 2^k последовательностей.

Сравнение один-против-всех. Этот способ, как уже было обозначено выше, сравнивает один файл (назовем его запрашиваемый файл) с множеством файлов, для которых заблаговременно уже были сформированы «отпечатки». Набор всех «отпечатков» составляет отдельный файл. Для запрашиваемого файла также составляются

«отпечатки» (также отдельным файлом). Следующий шаг заключается в сравнении данных двух файлов, содержащих «отпечатки».

Сравнение все-против-всех. Этот способ представляется более сложным. На первом этапе для всех файлов (как и в предыдущем способе) составляются «отпечатки». Затем для каждого «отпечатка» составляется список файлов, содержащих его. Далее представление информации немного видоизменяется – для каждого из выбранных множеств файлов указывается количество общих «отпечатков». Пользователю необходимо задать порог количества общих «отпечатков» по достижении которого файлы будут считаться сходными.

Возможные направления будущих разработок.

Области применения представленного метода могут быть расширены как минимум в четырех направлениях:

- Расширение возможных типов файлов – комментарии (пояснения, сноски и пр.), рабочие и сжатые файлы;
- Дробление и исправление файлов различных размеров;
- Усовершенствование механизмов сравнения двух файлов;
- Улучшение характеристик выходных данных.

Недостатки: Алгоритм не работает с содержательным составом файлов. Он сосредоточен на синтаксическом составе. То есть файлы, содержащие сходную информацию, но написанные разными словами не считаются сходными. Другими словами данный подход не может использоваться для выявления семантической близости текстов.

Синтаксическая кластеризация всемирной паутины¹.

Задача: анализ схожести символического состава файлов, степени вмещения одного файла другим. Объединение схожих файлов в кластеры.

Возможности применения: Алгоритм разрабатывался и тестировался на текстовых документах. Также он может применяться для анализа:

- Аудио-сообщений, содержащих человеческую речь;
- Документов на иностранных языках;
- Музыкальных композиций;
- Картинок, видео, баз данных.

Более того, этот метод может использоваться для распознавания измененных URL-страниц, кластеризации результатов поиска в поисковых системах для более удобного их использования, анализа изменения сайтов, а также для распознавания плагиата.

Предложенный метод позволяет определять синтаксическую схожесть файлов в Интернете. Разработанный алгоритм основан на установлении степени совпадения символического состава файлов, а также степени вмещения одного файла другим.

Исходные данные исследования содержали 30000000 HTML и текстовых документов из сети Интернет. Парное сравнение содержало 10^{15} (квадриллион) сравнений. Окончательная кластеризация происходила на основе 50%-й степени схожести. Было обнаружено 3,6 миллиона кластеров, содержащих 12,3 миллионов документов (из них 2,1

1 Andrei Z. Broader, Steven C. Glassman, Mark S. Manasse, Geoffrey Zweig. Syntactic Clustering of the web. 1997

миллион кластеров содержали идентичные документы (5,3 миллиона документов), 1,5 миллиона кластеров содержали 7 миллионов документов).

Алгоритм. Каждый документ рассматривается в первую очередь как последовательность слов. Назовем смежную последовательность символов во множестве всех символов документа D «шинглом» (shingle). Каждому документу D ставится в соответствие множество последовательностей символов $S(D, w)$. То есть каждый исходный документ D мы определяем как w -шингловый как множество всех уникальных шинглов размера w . Так, например, 4-шингловое множество для документа {эта, роза, красива, эта, роза, красива, эта, роза} будет множество {эта, роза, красива, эта}, {роза, красива, эта, роза} и {красива, эта, роза, красива}.

Для определения степени схожести документов A и B использовались две метрики: *сходство* (resemblance) и *вместимость* (containment).

Для данного шингл-размера сходство рассчитывается по следующей формуле:

$$r(A; B) = |S(A) \cap S(B)| / |S(A) \cup S(B)|$$

Показатель сходства документов изменяется в интервале $[0;1]$ – чем ближе эта величина к единице – тем более документы похожи.

Вместимость документов рассчитывается как

$$c(A; B) = |S(A) \cap S(B)| / |S(A)|$$

Подобным образом измеряется и вместимость – от 0 до 1. Чем ближе она к единице, тем более один документ включен в другой.

Для расчета данных показателей удобно иметь *эскиз* (sketch) (то есть некоторую «выжимку») каждого документа на несколько сотен байт. Имея эскизы двух документов по ним становится возможным рассчитать вышеописанные метрики.

Пусть U – это множество всех «шинглов» зафиксированного размера w . Параметр s определим следующим образом: для множества $W \subseteq U$ определим $\text{MINs}(W)$ как

$$\left\{ \begin{array}{l} \text{MINs}(W) = \text{множество минимальных } s \text{ элементов в } W, \text{ если } |W| \geq s; \\ W, \text{ иначе.} \end{array} \right.$$

где «минимальный» относится к цифровому порядку и определяется $\text{MOD}_m(W)$ = множество элементов W , от 0 до W .

Пусть $\pi : U \rightarrow U$ это случайное равномерное изменение U . Пусть $F(A) = \text{MINs}(\pi(S(A)))$ и $V(A) = \text{MOD}_m(\pi(S(A)))$. $F(B)$ и $F(A)$ определим аналогично.

Тогда

$$\text{Полезность} = \frac{|\text{MINs}(F(A) \cup F(B)) \cap F(A) \cap F(B)|}{|\text{MINs}(F(A) \cup F(B))|} \text{ - и}$$

$$\text{Полезность} = \frac{|V(A) \cap V(B)|}{|V(A) \cup V(B)|}$$

являются объективными оценками схожести документов A и B .

$\text{Полезность} = \frac{|V(A) \cap V(B)|}{|V(A)|}$ - объективная оценка вместимости документов A и B .

Алгоритм имел следующий вид:

- 1) Извлечь информацию из каждого документа;
- 2) Сформировать эскиз каждого документа;
- 3) Сравнить эскизы каждой пары документов на предмет превышения порога схожести;
- 4) Сформировать пары схожих документов с целью их дальнейшего объединения в кластеры;

Кластеризующий алгоритм имеет следующий вид:

- 1) На первом этапе формируем эскиз каждого документа;
- 2) На втором этапе создаем список всех «шинглов» и документов, в которых они присутствуют, сортируя их по значимости. Вследствие этого эскиз каждого документа расширяется на пару параметров <шингл, ID документа>;
- 3) На третьем этапе мы формируем список всех пар документов, содержащих идентичные «шинглы», указывая какое количество общих «шинглов» они имеют. Таким образом, список <шингл, ID документа> расширяется до <ID, ID, количество общих шинглов>.
- 4) На финальной стадии формируем кластеры схожих документов. Анализируя триады <ID, ID, количество общих шинглов> мы принимаем решения о том, превышает ли степень схожести документов установленный нами порог. Если да – устанавливаем связь между двумя документами и относим их к одному кластеру.

Итак, статья содержит описание метода, позволяющего выделять схожие/идентичные множества элементов (символов) и измерять степень их схожести, объединяя затем в кластеры.

Сходная проблематика описана в статье Nevin Heintze. *Scalable Document Fingerprinting*¹, где акцент делался на распознавании плагиата. Также данной проблематикой занимались U. Manber², N. Shivakumar и H. Garcia-Molina.³, S. Brin, J. Davis⁴.

Следующий алгоритм своей основой имеет связи (гиперссылки) между документами.

¹ Nevin Heintze. *Scalable Document Fingerprinting*. Proceedings of the Second USENIX Workshop on Electronic Commerce, Oakland, California, November 18-21, 1996

² U. Manber. Finding similar files in a large file system. Proceedings of the 1994 USENIX Conference, pp. 1-10, January 1994.

³ N. Shivakumar, H. Garcia-Molina. *SCAM: A Copy Detection Mechanism for Digital Documents*. Proceedings of the 2nd International Conference on Theory and Practice of Digital Libraries, Austin, Texas, 1995

⁴ S. Brin, J. Davis, H. Garcia-Molina. *Copy Detection Mechanisms for Digital Documents*. Proceedings of the ACM SIGMOD Annual Conference, San Francisco, CA, May 1995.

Улучшенный алгоритм

«отсеивания» тем в он-лайн среде¹

Задача: поиск документов наиболее релевантных теме запроса.

Возможности применения: выявление публикаций и авторов со сходными интересами, заказных публикаций, плагиата

Статья описывает проблему «отсеивания» тем в Интернет – поиск документов, наиболее релевантных запросам пользователей. Подход представляет собой объединение анализа связей документов и контент-анализа. Авторы описывают три основные проблемы в данном поле - взаимное усиление позиций связанных документов, автоматическое генерирование ссылок и выдача нерелевантных документов – и предлагают методы их решения.

Алгоритм. Анализ связей. Основная цель анализа связей документов – использование гиперссылки между документами для отбора необходимых документов (основываясь на предположениях, что документы, связанные гиперссылками имеют сходное содержание и что если один автор в своей работе дает ссылку на работу другого, то он считает эту работу ценной). Рассмотрим подробнее алгоритм Клейнберга, представляющий данный метод².

Улучшенный алгоритм Клейнберга.

Введем две метрики документа: *центральность* и *авторитетность*. Документы с большим значением авторитетности хорошо соответствуют теме запроса, в то время как документы с большим значением центральности содержит большое количество ссылок на документы, релевантных теме. Далее вытекает следующая логика: документ, который ссылается на многие документы имеет высокую центральность, документ, на который ссылаются многие документы – высокую авторитетность. Транзитивно, документ, который ссылается на многие авторитетные документы еще более централен, так же, как и еще более авторитетен тот, на который ссылаются многие центральные. В контексте запроса пользователя алгоритм вначале формирует определенный граф, где вершинами являются документы. Затем узлам итеративно назначаются центральности и авторитетности. Граф формируется следующим образом: стартовое множество документов, релевантных запросу, определяется поисковиком (скажем, 200 первых выданных документов). Затем это множество увеличивается на «соседей» этих документов – то есть документы которые ссылаются на и на которые ссылаются выданные документы (автор советует рассматривать не более 50 предшественников каждого документа). Эти два множества формируют соседский граф. Ребра графа - гиперссылки между документами.

Если документы одного источника связаны k связями с одним документом второго источника, то каждой связи присваивается вес авторитетности (authority weight) $1/k$. Если один документ из первого источника связан l связями с документами второго источника, то для каждой связи вес центральности (hub weight) – $1/l$. Изолированные узлы не рассматриваются.

Расчет центральностей и авторитетностей документов происходит по следующей схеме:

- 1) Пусть N – это множество вершин соседского графа;

1 K. Bharat, Monika R. Henzinger Improved Algorithms for Topic Distillation in a Hyperlinked Environment, 1998.

2 Kleinberg, J. 1998. "Authoritative sources in a hyperlinked environment." Proc. of 9th ACM/SIAM Symposium on Discrete Algorithms. Also appeared as IBM Research Report RJ 10076, May 1997.

- 2) Для каждой вершины n из множества N , $H[n]$ – ее центральность, $A[n]$ – ее авторитетность;
- 3) Назначим $H[n]$ и $A[n]$ равными 1 для всех n из множества N ;
- 4) До тех пор пока $H[n]$ и $A[n]$ не сойдутся в одну точку:
- 5) Для всех n из множества N , $A[n] = \sum_{(n',n) \in N} H[n'] \times auth_wt(n',n)$;
- 6) Для всех n из множества N , $H[n] = \sum_{(n,n') \in N} A[n'] \times hub_wt(n,n')$;
- 7) Нормируем вектора $H[n]$ и $A[n]$.

По уверениям авторов, алгоритм сходится после примерно 10 итераций.

Далее, для определения релевантности документов, авторы предлагают совместно использовать анализ связей и контент-анализ.

- 1) Очевидно, что тема запроса шире, чем сам запрос. Поэтому авторы предлагают использовать документы стартового множества как расширенный запрос и сопоставлять каждый узел (документ) графа с ним (точнее, с первыми 1000 словами каждого документа).
- 2) Для удаления нерелевантных документов предлагается три методики:
 1. Медиана. Порог – медиана всех значимых весов;
 2. Медиана стартового множества. Порог – медиана значимых весов узлов стартового множества;
 3. Доля максимальных весов. Порог – фиксированная доля максимальных весов. В данном исследовании использовался порог макс/10.

Алгоритм ARC Чаркабартти при анализе текстов также опирается на алгоритм Клейнберга. Но ARC-алгоритм, в отличие от данного, использует соседство второй очереди.

Анализ связей используется при исследовании соавторства в библиометрике.

Финальное сравнение алгоритмов

В таблице 2 представлены результаты финального сравнения алгоритмов. Как видно, нами был выделен ряд критериев для сравнения. Отдельной частью являются недостатки алгоритмов, которые по своей природе не являются параметрами для сравнения, а представляют собой характеристики алгоритмов.

Таблица 2

Финальная сравнительная таблица

Решаемая задача	Выявление степени схожести текстов (схожих, вложенных файлов)		Поиск документов, релевантных теме запроса	Выявление сообществ				
				Четкие границы, любой размер			Нечеткие границы, крупные размеры	
Алгоритм (фамилия автора)	Broader	Manber	Bharat	Girvan, Newman	Djidjev	Polanco	Zakharov	Valdes
Декларируемые области применения	Анализ текстов, аудио, видео, графики	Файл-менеджмент, поиск, преобразование данных	Выявление авторов со сходными интересами, заказных публикаций, плагиата	Сети: социальные, сети цитирования и соавторства, пищевые сети животных, нейронные сети, компьютерные и он-лайн сети	Сетевые структуры различной природы: социальные он-лайн и офф-лайн сети	Выявление имплицитных он-лайн сообществ, связанных веб-документов и сайтов	Выявление сообществ в социальных (в частности, он-лайн) социальных сетях	Анализ структуры комплексных, неполных и неточных данных: медицинских, биологических, он-лайн
Реальные области применения	Выявление степени совпадения символического состава информации		Выявление плагиата, связанных публикаций	Выявление сообществ в сетях различной природы		Выявление связанных сайтов	Распространение информации	-
Время выполнения	30000000 (150 Гб) 10,5 дней	500 Мб – 1 Гб 1 час	2000 документов 3 минуты	56276 вершин 42 минуты	15000 вершин 6295801 ребер 15,18 с			
Метод	Анализ символического состава информации, кластеризация		Анализ связей, контент-анализ	Кластеризация				
Программное обеспечение		sif	TREC		Metis	SDOC		
Необходимые ресурсы	Изначальное множество сравниваемых объектов		Пользовательский запрос, поисковая система	Граф (визуализированные объекты и связи между ними)		Список кластеризуемых объектов и описание их характеристик		
Степень вовлеченности (затратности, участия) исследователя	Выбор размера «шингла»	Выбор размера «якоря», порога совпадения	Определение параметров, формулировка запроса	Построение изначального графа		Описание объектов	Распространение «виртуальных» чернил	Описание объектов
Недостатки	Отсутствие описания промежуточных стадий.	Невозможность содержательного анализа.	Трудоемкость, отсутствие критериев выбора параметров	Значительные временные затраты на реализацию	Отсутствие описания способов реализации стадий алгоритма	Путаница в терминологии	Теоретический	Трудоемкость характер

Заключение

Бесспорно, сегодня Интернет-среда стремительно растет, и практика диктует запрос на всё новые, более эффективные методы обработки больших массивов информации. Для поиска информации и принятия решения эксперту, аналитику или простому пользователю уже недостаточно использовать поисковые машины. Для наиболее эффективного решения поставленных задач и поиска необходимой информации нужны новые методы, которые позволили бы пользователю (в широком смысле этого слова) при минимальных усилиях получить необходимую информацию.

Для настоящей работы были подобраны современные методы анализа и выделения сетевых сообществ, представляющие два различных направления (математическое и социолингвистическое). На первом этапе методы были классифицированы по направлениям принадлежности. Затем были подробно разобраны типичные представители каждого направления. На финальной стадии методы были распределены, исходя из задач, ими решаемых. Такой подход отвечает практическим запросам – ведь на практике исследователь исходит именно из практической задачи и подбирает методы и процедуры для её решения, а не наоборот.

В заключение необходимо отметить, что данная работа является начальным этапом для более глубоких и масштабных теоретико-методологических исследований и их практического применения. В дальнейшем планируется разработать методологию (с подробным описанием алгоритма в целом и конкретных процедур) выделения сообщества (скорее всего, экспертного), учитывающую достоинства, недостатки и целевую направленность проанализированных методов. Продолжением теоретических изысканий может стать онтологическая основа методов, расширение и углубление произведенной классификации.



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

«Электронное государство» как социотехническая система

Сергей Васильевич Бондаренко

Центр прикладных исследований
интеллектуальной собственности
(Ростов-на-Дону)

Начнем рассмотрение заявленной темы с постановки исследовательской проблемы. Формирующееся «общество знаний» («информационное общество») немислимо без создания крупномасштабных инфраструктурных систем, таких как «электронное правительство», «электронные муниципалитеты» и др. В большинстве стран мира, несмотря на значительный объем ресурсов выделяемых для создания сложных информационных систем, их реализация не достигает в полной мере заявленных целей.

Ориентация исключительно на качество техники и программного обеспечения не оправдывает себя, поскольку не учитывает человеческий фактор. Если в быту компьютеризированные артефакты и новейшие технологии достаточно быстро становятся частью повседневности, то в практике функционирования государства наблюдаются перманентные сбои. То, что под силу процессам самоорганизации социума, оказывается сложнейшей задачей для бюрократической машины.

Вопрос не столько в сопротивлении чиновников (хотя и этот фактор имеет место), сколько в просчетах, связанных с игнорированием социокультурных особенностей человеко-машинных систем. Чтобы стать частью повседневности эти системы должны отражать особенности среды обитания.

Еще в 1866 году немецкий биолог Эрнст Геккель ввел научный оборот дефиницию «экология» (от греч. *Óikos* – местопребывание и *logos* – понятие, учение). Уже несколько десятилетий термин применяется не только к природной среде, но и в расширенном смысле практически ко всем видам человеческой деятельности, отражая влияние среды на функционирование различных социальных и социотехнических систем.

Системы, о которых мы поведем речь, не изолированы от среды, а встроены в нее, поэтому не учитывать этот фактор значит бессмысленно тратить ресурсы. Казалось бы, этот тезис очевиден, но на практике все далеко не так просто. Трудно формализуемые социальные отношения и психологические реакции достаточно сложно учитывать при проектировании систем, поскольку это качественно иной уровень культуры разработки, тем более отражать их при бюджетировании проектов.

Вот и получается, что сложные информационные системы, предназначенные для повышения эффективности взаимодействий государства и граждан, становятся «вещью в себе». Говоря иными словами, инструменты гуманитарной коммуникации, отторгая благодаря некавалифицированным проектировщикам социальную составляющую, становятся самоценными, тем самым утрачивая изначальный смысл

своего существования. Примеров тому множество, в частности, достаточно показательны, что в России системы «электронного правительства» внедряются с начала 2000 года и до сих пор не достигли эксплуатационной готовности, хотя в других странах уже не один год успешно применяются.

Виновна в таком положении и отечественная наука, значительная часть представителей которой воспринимают компьютеризированные системы как исключительно технические, а значит – недостойные гуманитарного знания... В настоящей статье, в силу объективного ограничения на ее объем мы попытаемся показать абсурдность этого заблуждения, представив контуры разрабатываемой нами модели системных понятий, связанных с феноменом «электронного государства», являющегося частью цифровой экосреды.

Осмысление особенностей формирования и развития экосистем стало частью научного дискурса, ориентированного на формализацию междисциплинарных усилий по изучению потоков знаний и применения их как для воспроизводства, так и развития сложных социально-технических систем. Основываясь на знаниях, полученных в процессе изучения биологических экосистем¹, исследователи пытаются перенести оправдавшие себя в природе принципы для оптимизации взаимодействий людей как с техническими артефактами, так и с информационными потоками.

Поэтому термин «экосистема» должен рассматриваться в смысловом отношении гораздо шире, чем просто метафора. На наш взгляд, этот термин имеет парадигмальную коннотацию и описывает видовое разнообразие как естественных (природных), так и сформированных людьми систем. Не случайно его стали применять в различных сферах, следствием чего стали словосочетания «бизнес-экосистема»², «социальная экосистема», «экосистема знаний» и др.

Важно понимать, что как в природе существуют различные экосистемы, так и в обществе происходит взаимодействие самостоятельных экосистем, происходящее благодаря дифференциации различных сторон человеческой деятельности и исходя из возникающих потребностей. Такое возможно в силу *открытого* характера экосистем позволяющего обмен знаниями, идеями и технологическими решениями.

В соответствии с принципом целостности свойства экосистемы как целого невозможно свести к сумме свойств составляющих ее частей. Процессы и особенности взаимодействий зависят от организационного контекста связей между ее частями.

Развитие цифровых технологий привело к появлению зонтичного понятия «цифровая экосистема», под которой понимают цифровые артефакты и инфраструктуру передачи данных их хранения и обработки, пользователей систем, включая социальные, экономические, политические, психологические и иные факторы, влияющие на осуществление взаимодействий. Результатом взаимодействий становится

¹ В природе экосистемы рассматривают с позиций того как биологические сообщества взаимодействуют между собой и другими живыми существами, осуществляют воспроизводство и эволюционное развитие с учетом влияния физической среды.

² Так, под термином «бизнес-экосистема» понимается сеть покупателей, поставщиков и производителей товаров и услуг, а также социально-экономическая среда, в которых указанные акторы действуют на основе институциональной и нормативной базы.

появление сложно организованной цифровой среды функционирующей на принципах самоорганизации (не исключающих нормативного регулирования) и поддерживающей принципы самовоспроизводства и разнообразия.

За рубежом понятие «цифровая экосистема» (англ. - the Digital Ecosystem) вошло в научный оборот относительно недавно¹ и стало применяться, прежде всего, в исследовании бизнеса², изучения возможностей влияния новых технологий на региональное социально-экономическое развитие³, а также при создании моделей функционирующей в телекоммуникационных сетях искусственной жизни⁴. В проводимых исследованиях значительное место занимает поиск аналогий с процессами, идущими в природных экосистемах⁵.

На наш взгляд, проводя параллели с биологическими экосистемами, исследователи нередко впадают в крайности редуccionизма. Безусловно, природа величайший творец, однако созданные людьми социально-технические системы имея много общего с аналогичными биосистемами, тем не менее, обладают и отличиями.

Как мы отметили выше, термин «экосистема» имеет парадигмальную коннотацию, соответственно ведя речь, к примеру, о «цифровой экосистеме» необходимо учитывать, что в ее состав входят самостоятельные «цифровые экосистемы» уже видового уровня. Вдобавок практически каждая видовая экосистема включает в свой состав различные социотехнические системы прикладного назначения, при этом и сама экосистема может быть охарактеризована как социотехническая. Возникает когнитивная путаница, разрешить которую можно используя таксономию.

На наш взгляд, таксономия должна выглядеть следующим образом. Цифровая экосистема является *родовым понятием* и в ее рамках связанные с цифровыми технологиями инфраструктурные платформы позволяют развивать фундаментальные идеи и модели их реализации в различных видовых применениях. К примеру, экосистема «цифровое (электронное) государство» включает в свой состав: «электронное правительство», «электронный парламент», «цифровую судебную систему», «электронные муниципалитеты» и т. д. Мотивацией появления прикладных цифровых экосистем является необходимость решения в процессе использования возможностей предоставляемых цифровыми технологиями актуальных для развития общества проблем.

¹ Briscoe G. Digital Ecosystems. / PhD thesis. London: Imperial College 2009.

² Moore J The death of competition: leadership and strategy in the age of business ecosystems. New York: Harperbusiness, 1996; Corallo A., Passiante G., Principe A.. The Digital Business Ecosystem. Cheltenham: Edward Elgar Publishing, 2007.

³ Dini P., Lombardo G., Mansell R., Razavi A., Moschoyiannis S., Krause P., Nicolai A., Rivera Le'on L. Beyond interoperability to digital ecosystems: regional innovation and socio-economic development led by SMEs // International Journal of Technological Learning, Innovation and Development, 2008, № 1. PP. 410–426.

⁴ Marrow P., Koubarakis M., van Lengen R., Valverde-Albacete F., Bonsma E., Cid-Suerio J., Figueiras-Vidal A., Gallardo-Antolin A., Hoile C., Koutris T., Molina-Bulla H., Navia-Vazquez A., Raftopoulou P., Skarmas N., Tryfonopoulos C., Wang F., Xiruhaki C. Agents in decentralized information ecosystems: The DIET approach / In: Symposium on Information Agents for Electronic Commerce. Society for the Study of Artificial Intelligence and Simulation of Behavior, 2001. PP. 109–117.

⁵ Kurz T., Heistracher T. Simulation of a self-optimising digital ecosystem / In: Digital Ecosystems and Technologies Conference. IEEE Press, 2007. PP. 165–170.

Цифровая экосистема обеспечивает концептуальную структуру для применения связанных с цифровыми технологиями идей и моделей в различных сферах функционирования общества. Соответственно, упомянутые выше подсистемы носят *прикладной характер*, в отличие от онтологического характера цифровой экосистемы.

Онтологическим отличием цифровой экосистемы (цифровой экоподсистемы) от входящих в нее социотехнических систем является учет фактора окружающей среды. К примеру, цифровая экоподсистема «электронного правительства» не только включает в свой состав социотехническую систему с аналогичным названием, но и воздействующие на систему внешние факторы, к примеру, тенденции на рынке цифровых артефактов применяемых в «электронном правительстве», возможные угрозы и риски связанные с делинквентно ориентированными группами хакеров, системными просчетами в сфере обеспечения на национальном уровне информационной безопасности и т. д. и т. п.

Социотехнические системы, хотя по принципу матрешки и входят в состав экосистемы (подсистемы), тем не менее, они в большей мере структурированы и функционально независимы. Эта структурированность, в частности, обеспечена системным единством программного обеспечения. Однако на настоящий момент никакое программное обеспечение не может отразить не прописанные в ее алгоритме реакции на изменение окружающей среды.

Цифровые экосистемы характеризуются как адаптивные открытые нелинейные динамические социотехнические системы, со свойствами самоорганизации и способностями к устойчивому развитию. Такие системы обладают свойствами адаптации к широкому диапазону паттернов пользовательского поведения, за счет ориентации на осуществление сотрудничества, обмена знаниями и навыками, а также развития адаптивных технологий.

Применяемые в прикладной экосистеме технологические решения должны обладать качествами *масштабируемости*, то есть увеличения производительности за счет дополнительных ресурсов без структурных изменений центрального элемента системы. Кроме того система должна обладать диалектическими качествами: с одной стороны быть *стабильной*, а с другой – быстро *адаптироваться* к трансформации среды в которой она функционирует и не допускать, чтобы быстрые изменения нормативной базы или же неудачные технические решения, приводили к потере контроля над ней.

Стабильность в диалектическом понимании отнюдь не означает отсутствие разнообразия, как вариантов развития, так и утрату возможностей динамической адаптации к изменениям среды. Необходимо учитывать, что техническая поддержка большого числа плохо приспособленных к нуждам пользователей решений требует значительных ресурсных затрат.

Даже упомянутые требования, зачастую противоречивые (и при этом далеко не исчерпывающие), свидетельствуют о высокой сложности, как проектирования, так и эксплуатации таких систем. В природе за счет инструментов самоорганизации такие проблемы находят решение, поэтому и при проектировании цифровых экосистем востребованы достижения теоретической экологии, эволюционной теории и многих

иных научных направлений в которых изучаются законы природы и общества.

Согласно теории *социальной экосистемы*, группы живых существ приспосабливаются к окружающей среде в силу необходимости выжить. При этом люди, как верно подметил Хуан Диез Николас (Juan Diez Nicolas)¹, единственные существа способные приспособиться к изменениям среды с использованием возможностей культуры. Поэтому, культуру можно и нужно рассматривать в качестве инструментального ответа со стороны сообществ на проблему лучшей адаптации к окружающей среде². Это, по нашему мнению, в полной мере относится и к цифровой среде.

Поскольку телекоммуникационные технологии стали неотъемлемой частью повседневности значительной части человечества, а их проявления многообразны, то в научном осмыслении происходящего легко впасть в ошибочную простоту идеологии технологического детерминизма. Сторонниками такого взгляда на мир делается упор на технические практики и математический формализм программных артефактов, а необходимость учета поведения пользователей попросту игнорируется.

Цифровая экосистема по своей сути является социотехнической, поскольку в ее состав входят как цифровые артефакты, так и люди, поведение которых является неотъемлемой частью системы. Сложность изучения этой среды усугубляется дефицитом общепризнанного в научном сообществе методологического инструментария.

Основная проблема в выборе методологии состоит в междисциплинарном характере цифровых экосистем. Поскольку цифровые информационные системы абстрактны, существуют сложности в осуществлении сбора эмпирических данных, фрагментарные же данные не позволяют в силу эмерджентности системы делать обобщающие выводы.

Выход видится в использовании различных методологических инструментов теорий социального конструктивизма, функционализма, феноменологии, символического интеракционизма, аутопойезиса и иных существующих в общественных науках наработок, позволяющих исследовать объективную реальность. Диапазон возможного методологического инструментария отражает глубокие онтологические и эпистемологические различия между техническим и социотехническим исследованиями.

Феномен внутрисистемной коммуникации является «стержнем», вокруг которого происходит объединение, на первый взгляд, несовместимых знаний. Тем не менее, коммуникативные барьеры, существующие в когнитивных подходах у представителей различных дисциплинарных областей, а также практиков-эксплуатационников являются препятствием для объективного анализа сложных систем. Кроме того, девальвируется ценность знаний, воспринимаемых как субъективные, релятивистские и плюралистичные.

Внедрение технологий «электронного правительства» во всем мире сегодня является ведущим трендом трансформации взаимодействий государства с гражданами и юридическими лицами. Как мы уже выше

¹ Díez Nicolás J. Postmaterialism and the social ecosystem / In: Beat and Beatrix Sitter-Liver (eds.). Culture Within Nature / Culture dans la Nature. Basel: ASSM, 1995.

² Hawley A. Human ecology: A theoretical essay. Chicago: University of Chicago Press, 1986.

отметили, в России такая технология внедряется с использованием идеологии технологического детерминизма. Считается, что самым важным является создание технологической инфраструктуры, а связанные с социальными и политическими аспектами использования технологий вопросы отходят на второй и даже на третий план.

Такой дискурс представляет нам ошибочным, поскольку «электронное правительство» представляет собой социотехническую систему, в которой человеческий фактор является решающим с точки зрения успешности реализации модернизационных проектов и внедрения инноваций. Социотехническая система «электронного правительства», согласно разработанной автором теоретической модели, включает следующие элементы:

- аппаратную инфраструктуру (компьютеры, компьютерные сети, информационные киоски, системы видеонаблюдения и т. п.);
- программное обеспечение (операционные системы, специализированные программы электронных сервисов, прикладные программные артефакты и т. п.);
- физическую инфраструктуру (здания, системы вентиляции, электропитания, освещения и т. п.);
- индивиды – пользователи системы;
- группы пользователей, влияющих на создание и функционирование виртуальных сетевых сообществ, формирование общественного мнения, культуры применения системы и т. п.;
- коммуникативная подсистема, в рамках которой циркулируют информационные потоки, осуществляется передача, обработка и хранение информационных массивов;
- ресурсная подсистема, включает материальные, финансовые, интеллектуальные и иные ресурсы, в числе которых знания занимают одно из первых мест;
- стратификационная подсистема, в которой зафиксированы права и обязанности всех участников взаимодействий;
- система нормативных актов, в которой отражены как вопросы функционирования «электронного правительства», так и встроенность этой системы в функционирование институтов государства;
- ценностная подсистема, в рамках которой создаются, воспроизводятся и получают развитие ценностные ориентации необходимые для обеспечения функционирования и эволюции модели электронных взаимодействий государства с гражданами и иными акторами.

В рамках системы, упомянутые выше компоненты, могут соединяться различным образом, как и значимость каждого компонента в разных системах может существенно различаться. К примеру, две государственные структуры, имеющие одинаковый набор аппаратных и программных артефактов могут создать различные социотехнические системы, поскольку будут различаться подходы к создаваемому контенту с точки зрения ориентации его на нужды конечного пользователя.

Поэтому ведя речь об эмерджентности социотехнической системы, необходимо учитывать как человеческий фактор, так недокументированные возможности программных и аппаратных артефактов, а также организационную культуру, господствующую в органе

власти и управления. Социотехнические системы не статичные, а динамичные сложно структурированные объекты.

В них перманентно происходит изменение взаимосвязей, а также разрешаются конфликты интересов отдельных элементов. Поэтому на онтологическом уровне необходимо понимание важности перманентной работы по оптимизации технических и социальных подсистем. Вместе с тем исследователи признают, что достичь полной оптимизации в рамках сложной системы невозможно¹, тем не менее, без поддержания баланса интересов и ресурсных возможной в социотехнической системе будут усиливаться дисфункциональные процессы, которые могут привести к выходу ее из строя.

Как утверждает Мария Виммер (Maria A. Wimmer)², использование комплексного подхода при разработке социотехнических систем «электронного правительства», подразумевает примат мультидисциплинарности, в рамках которой общественные науки занимают важное место. Проекты должны рассматриваться исходя из следующих перспектив: технологической, экономической, организационной, процессуальной, юридической, социальной, культурологической, безопасности, политики и этики.

Не учет хотя бы одного из упомянутых выше факторов приводит к деформации социотехнической системы, а в некоторых случаях, при стечении неблагоприятных обстоятельств, проект «электронного правительства» оказывается нежизнеспособным. Несмотря на распространенные неудачи таких проектов, в общественных науках явно недостаточно осмысленны социотехнические и когнитивные проблемы, а также вопросы, связанные с ценностными ориентациями пользователей, находящие отражение в моделях их поведения.

В рамках формирования современной научной парадигмы социотехнических систем акцент делается на понимании:

- особенностей сетевых взаимодействий между индивидами, социальными, культурными и технологическими элементами системы;
- особенностей формирования и разработки систем, способных к самоорганизации и приспособлению к изменяющей социальной, экономической и организационной среде, ориентированных на максимальное использование способностей включенных в систему индивидов³;
- специфики коммуникативных процессов и возможностей ими латентно управлять в интересах создания новых знаний;
- учета интересов пользователей в процессах эволюции систем⁴ и др.

Социотехническая система «электронного правительства» существует не сама по себе, а является составной частью государственной службы, состоящей из людей (социальная система), используемых ими артефактов, методов и специальных знаний (техническая система). Оказываемые государством гражданам услуги, получают оценку

¹ Keating C.B., Fernandez A.A., Jacobs D.A., Kauffman P. A Methodology for analysis of complex Socio Technical Process // Business Process Management Journal, 2001, vol. 7, № 1. PP. 33-50.

² Wimmer M.A. Integrated service modeling for online one-stop Government // Electronic Markets, 2002, vol. 12, № 3. PP. 1-8.

³ Coakes E. Knowledge Management: A Socialtechnical Perspective. / In: Coakes E., Willis D., Clarke S. (eds.). Knowledge Management in the SocioTechnical World. London: Springer-Verlag, 2002.

⁴ Bødker S. Through the Interface: A Human Activity Approach to User Interface Design. Hillsdale, NJ: Erlbaum, 1990.

населения и служат для выполнения взятых государством обязательств перед обществом. В таком контексте не учет при создании общенациональной инфраструктуры человеческого фактора означает игнорирование общественных интересов. При таком подходе становится очевидной опасность ориентации разработчиков на примат технологического детерминизма.

Создание сложных информационных систем подразумевает переход на новый уровень культуры, ориентированной не только на решение конкретных задач проекта, но и учет взаимосвязей и взаимозависимостей, существующих в цифровой экосистеме. Социально-технические принципы подразумевают не только осуществление анализа с учетом поведения человека, но и креативный подход к использованию технологий способами, отличными от изначально предполагавшихся.

Технологии существуют сегодня не сами по себе, а встроены в сложные экосистемы. Пересечение технического с различными слоями бытия человека ставит перед общественными науками новые проблемы, в которых в качестве объекта исследования цифровые экосистемы с каждым годом будут играть все большую роль.



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Этнические сообщества в социальных сетях

Ирина Сергеевна Хван
Алексей Алексеевич Горгадзе

Национальный исследовательский университет
«Высшая школа экономики» (Санкт-Петербург)

В процессе глобализации национальные границы стираются, миграция возрастает, появляются не только реальные, но и виртуальные этнические сообщества. Основной целью данного проекта является изучение этнических сообществ на сайте “vkontakte.ru”, на котором они создают «этнические» группы. Предполагалось, что это самый популярный сайт социальных сетей, однако в ходе работы стало очевидно, что другие социальные сети привлекают внимание других этнических групп, например, среди армян сайт «odnoklassniki.ru» более популярен, а прибалтийские этнические группы и вовсе не объединяются на данном Интернет-ресурсе.

Доклад является изложением результатов коллективной студенческой работы (2010г. октябрь - 2011г. май), которую помимо авторов выполняли: Е. Сукиасян, К. Сухарев, М. Остапенко, М. Маргарян, В. Костенко, а научным руководителем являлся Д.А. Александров. Сначала команда формировалась так, чтобы каждый участвующий анализировал ту этническую группу, с которой он себя идентифицировал. А в ходе работы мы расширили наш анализ на все постсоветские государства, а кроме того выявили новые этнические сообщества, такие как «кавказцы» и «славяне».

Основные этапы проекта соответствовали последовательному выполнению следующих задач: (1) выявление географического распределения образовательной миграции в России, (2) поиск и анализ этнических студенческих виртуальных сообществ и их локализации, (3) определение связности между разными виртуальными сообществами одной этнической группы при помощи метода сетевого анализа. Сетевой анализ, таким образом, применялся в данном случае только на третьем этапе.

Все данные для исследования мы собирали с помощью поиска на сайте vkontakte.ru. Что было с одной стороны очень удобно, а с другой стороны возникали и большие сложности. Дело в том, что администраторы сайта несколько раз за время нашего исследования меняли систему поиска, из-за чего нам не удавалось зафиксировать полностью некоторые данные. Кроме того, мы не сможем провести повторный сбор данных для сравнения, как было запланировано. Но, с другой стороны, мы собрали уникальные данные, потому что теперь vkontakte.ru имеет уже совсем другую систему поиска.

Ниже будет подробно описан каждый этап исследования, после чего мы подведем некоторые выводы о проделанной работе.

Образовательная миграция

На первом этапе было определено географическое распределение образовательных мигрантов из стран бывших советских республик и из северокавказских республик. Поиск приезжих студентов осуществлялся при помощи «подробного поиска», с использованием функций «Университет», «Школа». В качестве места обучения в школе, выбиралась страна, откуда приехал студент и отдельно был организован поиск приехавших из столицы. В итоге в список вошли города-миллионеры и несколько следующих непосредственно за ними по численности (всего в список вошли 19 городов).

В ходе анализа были выделены наиболее и наименее популярные города для абитуриентов из бывших советских республик и подсчитаны отношения, показывающие долю студентов из соседних республик в общем числе студентов данных городов.

Грузия

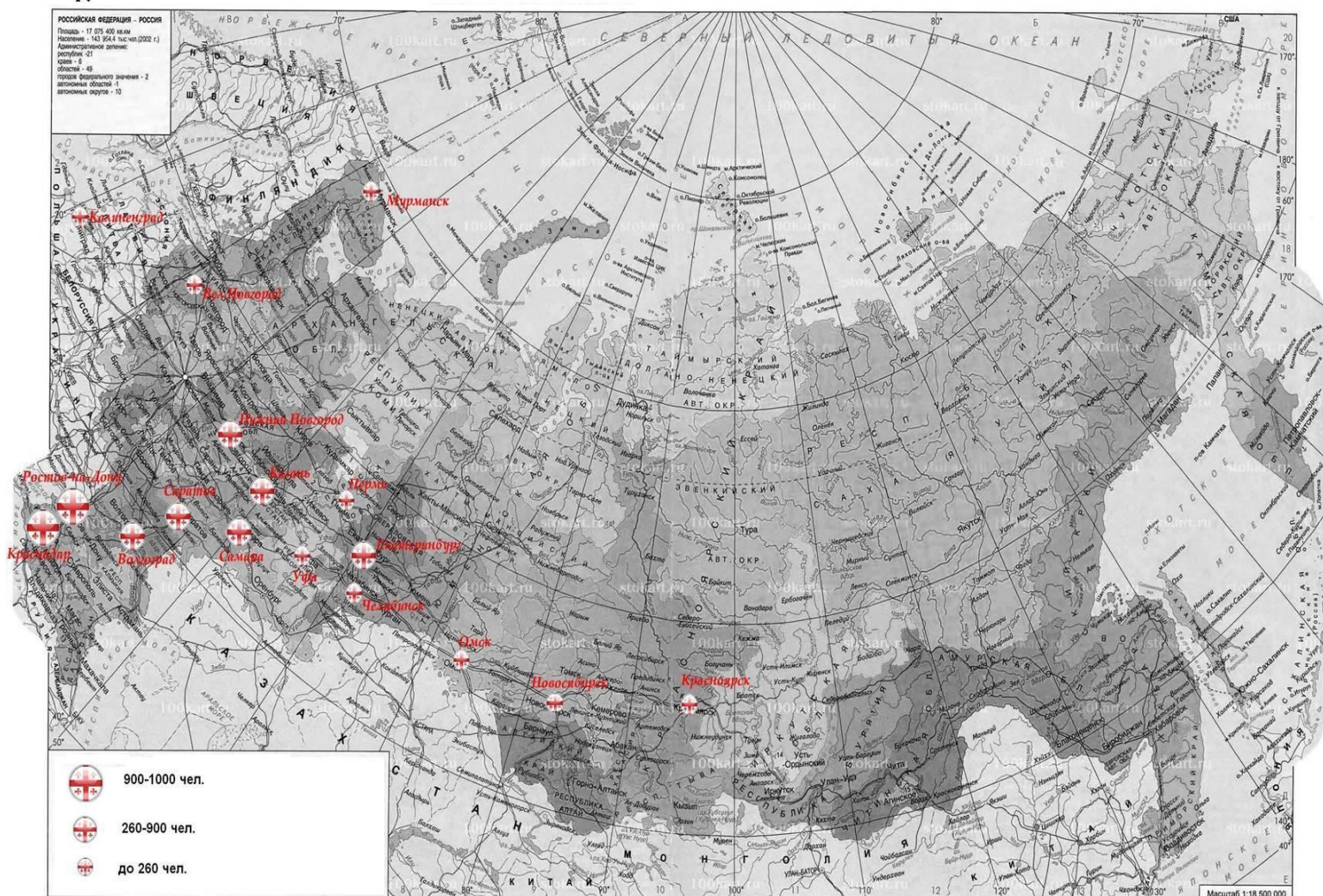


Рисунок 13. Грузинские студенты в российских города

Как видно на рис.1, наибольшее количество приехавших из Грузии студентов учатся в Ростове-на-Дону (933 студента) и Краснодаре (900 студентов), а меньше всего приехало в Мурманск (100 студентов). Из Украины студенты едут учиться в основном в Великий Новгород (3795 студентов), Волгоград (3535 студентов), Екатеринбург (4614 студента) и Ростове-на-Дону (4158 студентов), в то время как в Омск (1471 студента) и Уфу (1372 студента) украинские студенты почти доезжают(Рис.2). Значительная часть студентов из Узбекистана мигрирует в университеты Казани (3109 студентов), и опять же в Мурманске очень мало учатся студенты из Узбекистана (127 студентов)(Рис.3). Большинство литовских студентов проживают в Калининграде (798 студентов), однако меньше всего в Мурманске (51 студента).

Украина

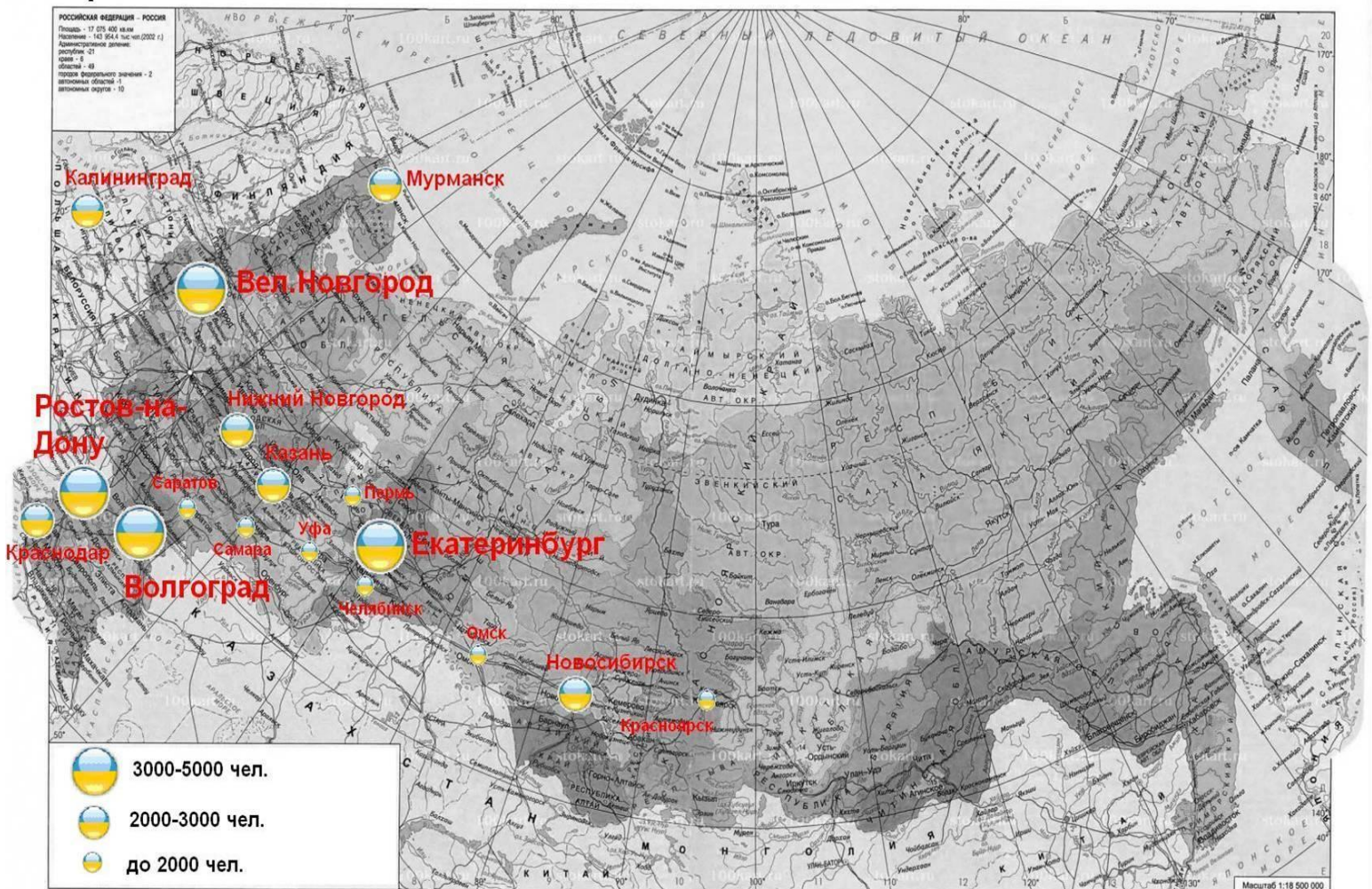


Рисунок 14. Украинские студенты в российских городах

Узбекистан

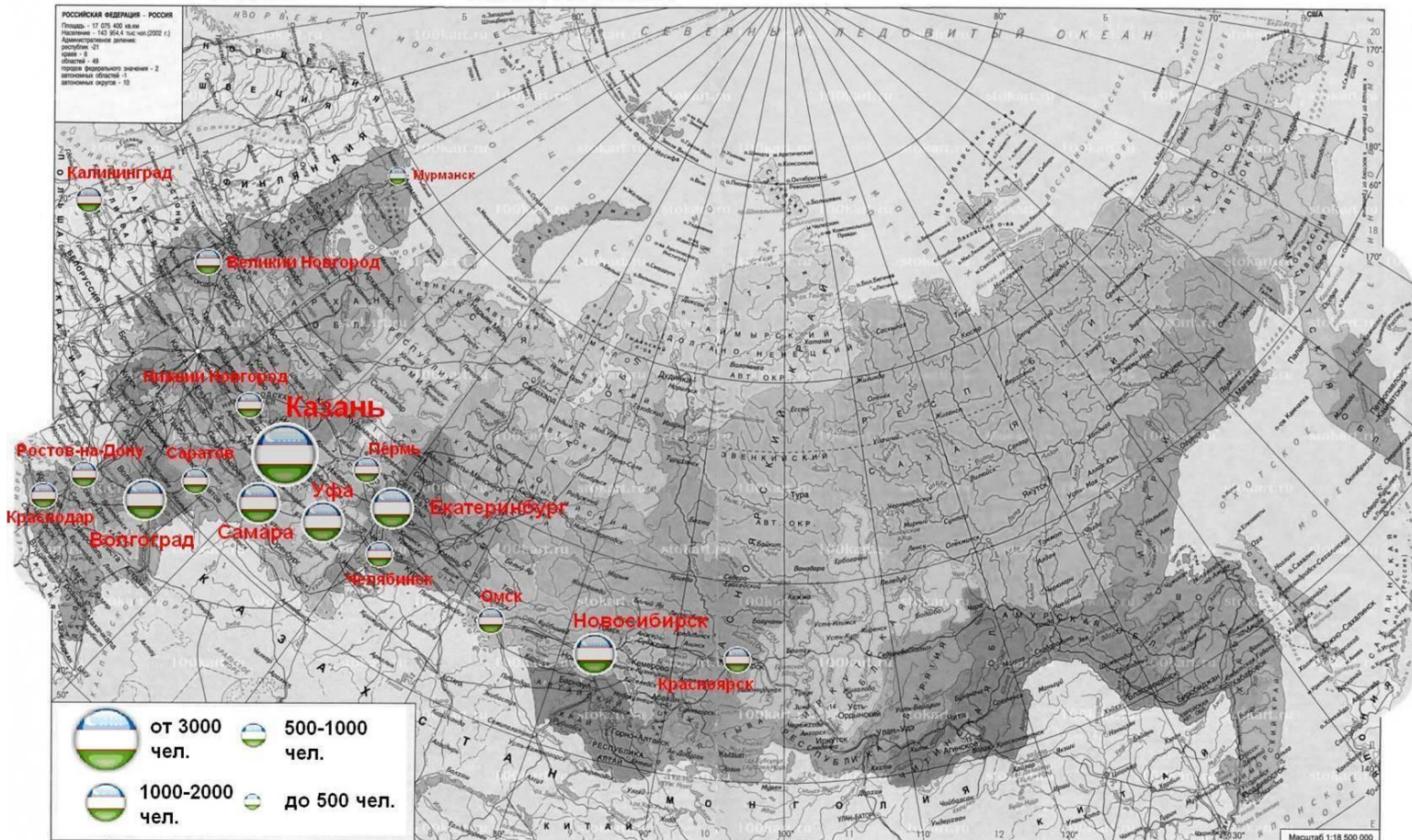


Рисунок 15. Узбекские студенты в российских городах

Стоит отметить, что анализируя полученные данные, мы, во-первых, можем получить не всю точную информацию (потому что люди могут писать неправду), а также мы, конечно, учитываем то, что например, студенты с Украины могут приезжать не только этнические украинцы, но в том числе и другие этнические группы, которые закончили школу на Украине.



Рисунок 16. Этнические подгруппы

Все этнические группы были разделены на четыре подгруппы(Рис.4): представители стран с преимущественно славянским населением (Украина, Белоруссия, Молдавия), республики Кавказа(Рис. 5), как независимых, так и входящие в состав РФ (Грузия, Чечня, Армения, Дагестан и Азербайджан) и среднеазиатские государства (Рис.6)

Кавказ

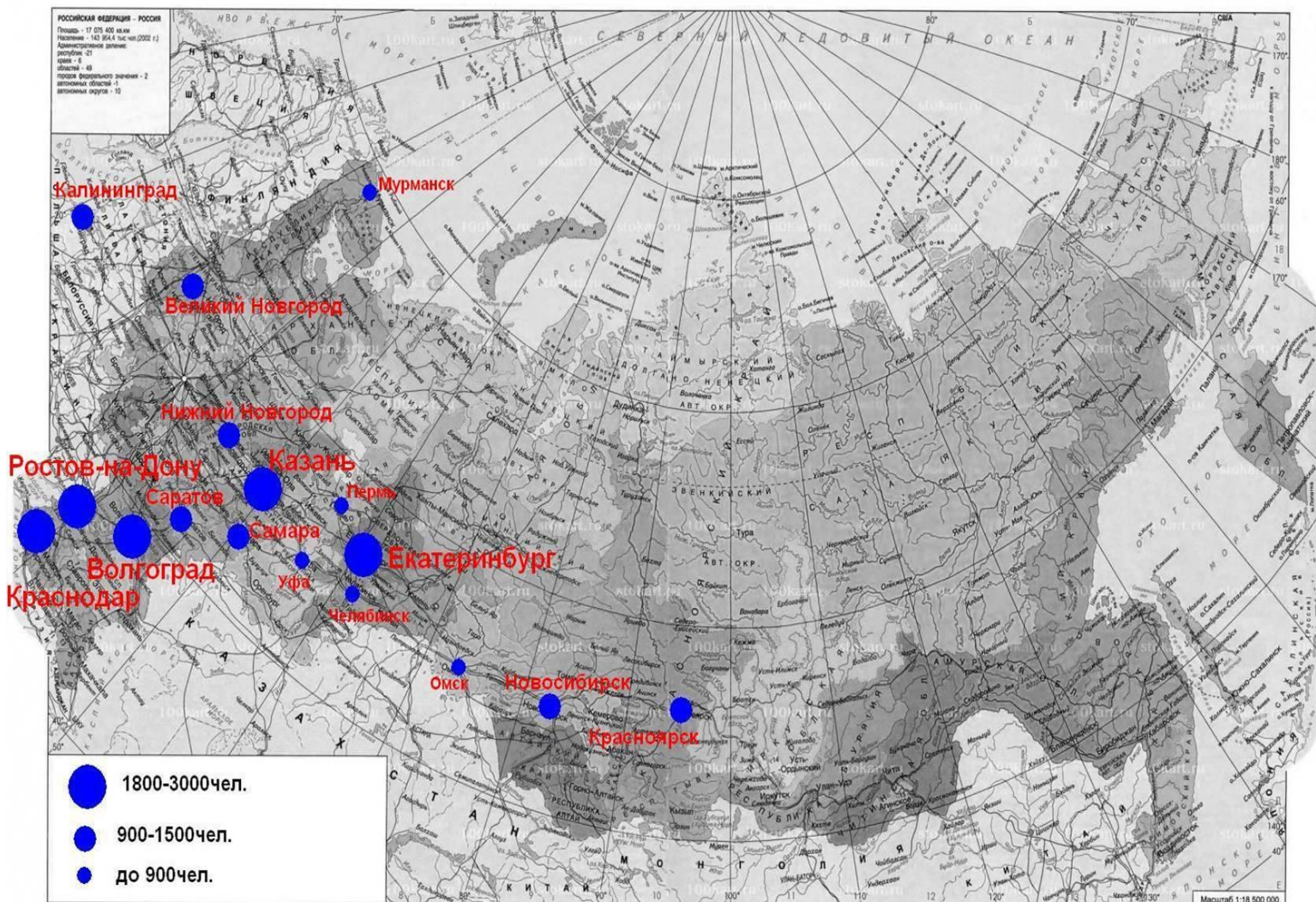


Рисунок 17. Студенты кавказских республик в российских городах

Средняя Азия

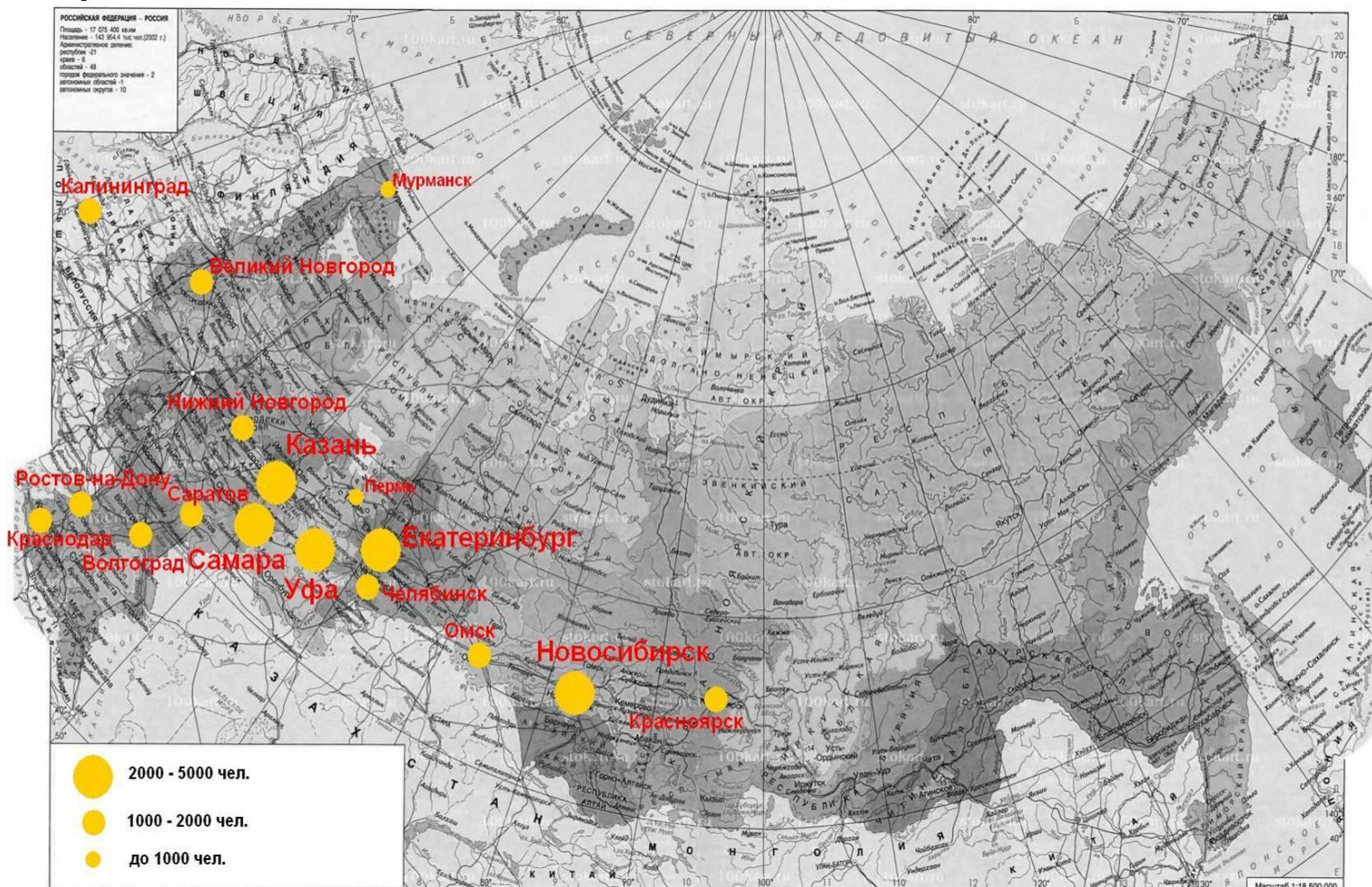


Рисунок 18. Студенты из Средней Азии в российских городах

Таким образом, можно подвести следующие итоги: большинство студентов из первой группы стран предпочитают учиться в Великом Новгороде, и реже всего они приезжают в Уфу; среди представителей кавказских республик большой популярностью пользуется Ростов-на-Дону и Казань, но менее всего привлекает Мурманск; в группе

Среднеазиатских государств популярностью пользуется город Казань, самым непопулярным городом также является Мурманск. Возможно, самая главная причина, по которой студенты выбирают тот или иной город для обучения – это расстояние. Этим можно объяснить низкую популярность таких городов как Мурманск и Уфа. Кроме того, как нам кажется, играет роль престижность вузов. И, наконец, можно предположить, что студенты из кавказских и среднеазиатских республик предпочитают Казань и Ростов-на-Дону в силу культурной особенности этих городов. Ростов даже называют “воротами Кавказа”.

Этнические студенческие сообщества

Анализ географического местоположения и принадлежность к вузу студенческих этнических виртуальных сообществ помог обнаружить самые популярные вузы среди определенных этнических групп.

Благодаря возможностям поиска в «Vkontakte» мы смогли найти студенческие этнические сообщества, которые существуют на сайте.

РОССИЙСКАЯ ФЕДЕРАЦИЯ. ФЕДЕРАТИВНОЕ УСТРОЙСТВО

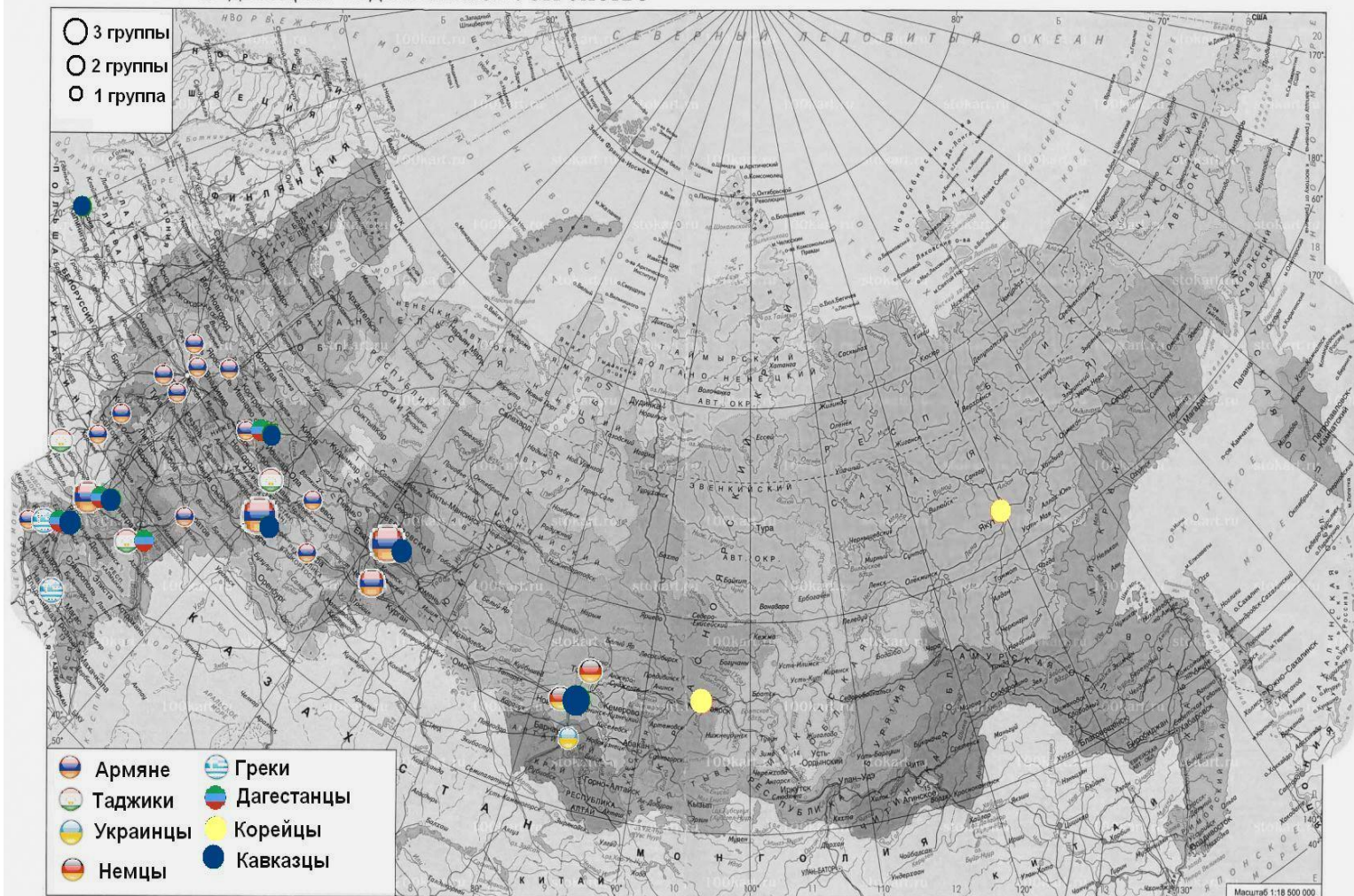


Рисунок 19. Студенческие этнические сообщества в социальной сети «Vkontakte»

На карте (Рис.7), можно увидеть полученные результаты. В качестве символа, мы брали флаг, той или иной этничности. Размер кружка отображает количество групп в одном из городов.

Одними из интересных фактов, которые мы обнаружили, было то, что 1) наибольшее количество этнических студенческих организаций создается в Москве (70 групп), и уже гораздо меньше в Петербурге (19 групп); 2) в таких вузах, как РЭУ имени Плеханова и РУДН, больше всего этнических групп “[vkontakte.ru](https://vk.com/vkontakte.ru)”; 3) наибольшая часть всех этнических студенческих сообществ состоит из армянских объединений.

Определение связности между разными виртуальными сообществами одной этнической группы методом сетевого анализа

Сравнение связности между разными виртуальными сообществами одной этнической группы проводилось при помощи метода сетевого анализа и программы Pajek. Данные для этого подраздела собирались из «групп-друзей» или сейчас это «ссылки» в каждой этнической группе. Таких этнических групп мы брали одинаковое количество для всех этничностей (30 групп).

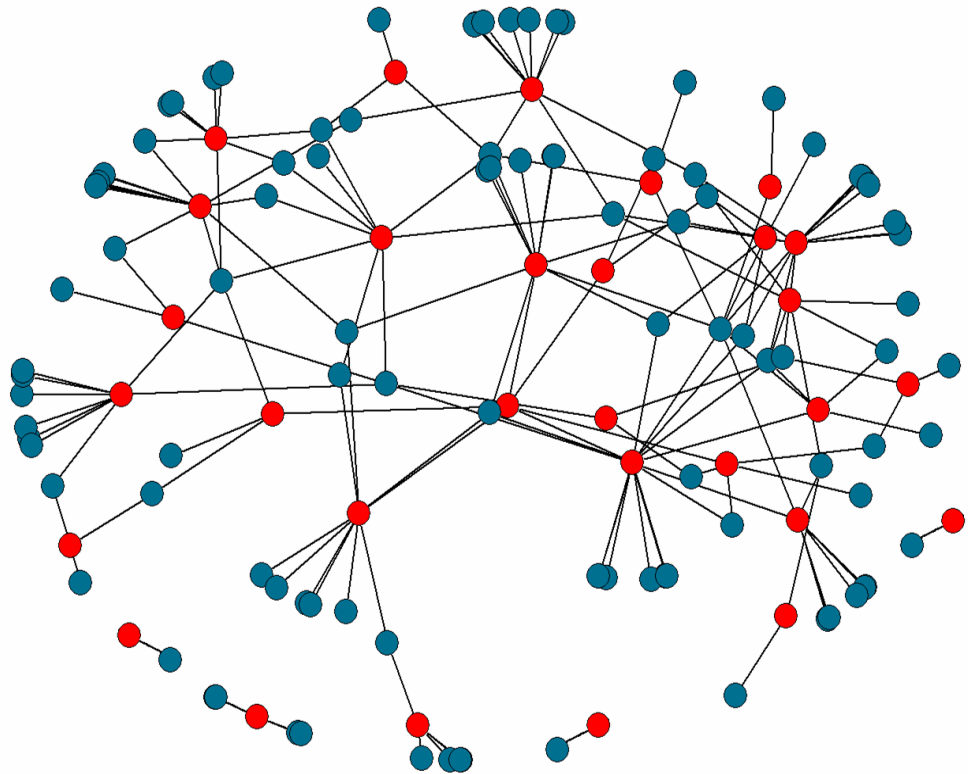


Рисунок 20. Корейские сообщества

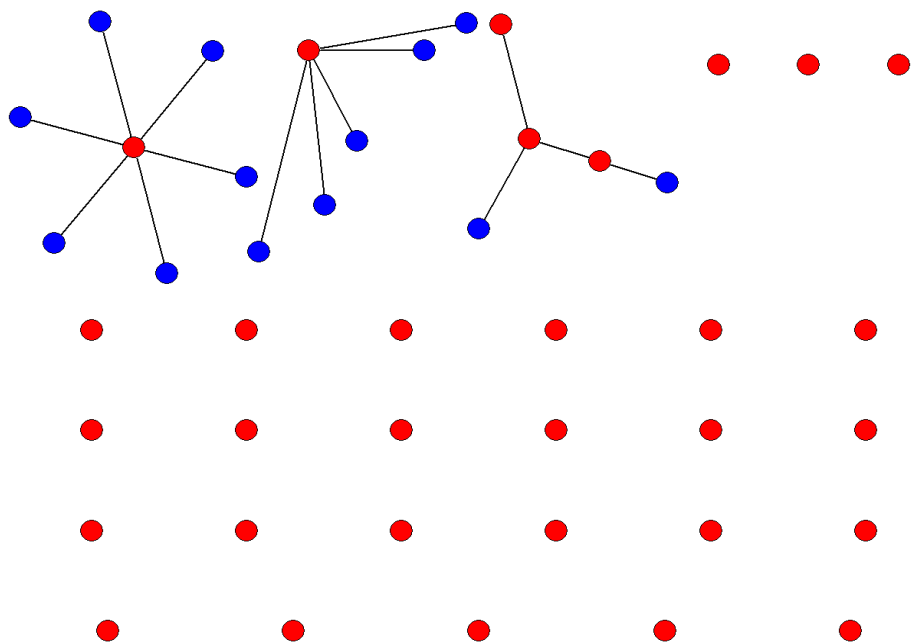


Рисунок 21. Узбекские сообщества

При построении социальных сетей почти все этнические сообщества показали примерно равный уровень связности (Рис.8). Исключением стали узбекские этнические сообщества (Рис.9), которые имеют малое число «групп-друзей» и связей с такими же этническими группами. А также прибалтийские страны (Эстония, Латвия, Литва) абсолютно не поддаются сетевому анализу, потому что число этнических групп всех трех стран в совокупности не превышает и пяти.

Таким образом, нам удалось рассмотреть каждое этническое сообщество с трех сторон: количество приезжих студентов, объединение студентов, принадлежащих одной и той же этнической группе, и, наконец, связанность этнических сообществ между собой. В дальнейшем мы собираемся развивать эти темы исследования, но не только в рамках данного сайта.



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Социальные сети поддержки в православной общине. На примере трех крупных православных приходов г. Москвы

**Иван Владимирович Забаев
Елена Викторовна Пруцкова**

Православный Свято-Тихоновский
гуманитарный университет (Москва)

В статье рассматриваются вопросы исследования городской православной приходской общины с точки зрения сетевого подхода. Описываются данные, полученные в ходе исследования «Три прихода на Покров: основные социально-демографические показатели и установки представителей общин крупных приходов г. Москвы», проведенного авторами в октябре 2011 года¹. Для того, чтобы сравнить данные по приходским общинам с результатами, характеризующими Россию в целом, некоторые вопросы, задававшиеся в ходе указанного исследования, мы также повторили во всероссийском опросе, проведенном Институтом Фонда «Общественное мнение» по заказу Православного Свято-Тихоновского гуманитарного университета в период с 14 по 21 декабря 2011 года. Выборка репрезентирует городское и сельское население России в возрасте от 18 лет и старше².

Постановка проблемы

Современная Россия, несмотря на устойчивый рост ряда показателей экономики и социальной сферы, до сих пор остается страной социальных болезней на фоне большинства развитых и даже развивающихся стран. При отсутствии развитого и законодательно оформленного партнерства государства и Русской Православной Церкви по вопросам социальной деятельности, в России существует множество примеров эффективной

¹ Опрос методом основного массива проводился в Москве 14.10.2011 на приходах трех крупных московских храмов, имеющих опытных духовников и (по экспертным оценкам) — сильную приходскую общину. Для получения данных по общинам приходов, опросные процедуры проводились после литургии в будний день, на который приходился православный праздник «Покров Пресвятой Богородицы». Мы предполагаем, что в данных условиях мы зафиксировали пришедшую в храм общину — ее ядро и периферию. Общий объем выборки составил 1072 человека. Были опрошены пришедшие в храмы: святого благоверного царевича Димитрия (настоятель — епископ Смоленский и Вяземский Пантелеимон (Шатов)), Святителя Митрофана Воронежского (настоятель — прот. Дмитрий Смирнов), Святителя Николая в Кузнецях (настоятель — прот. Владимир Воробьев). Для сравнения с общероссийскими данными в настоящей статье мы рассматриваем подвыборку 18 лет и старше (1028 чел.).

² Более подробная информация об исследованиях, а также некоторые результаты представлены на сайте исследовательского семинара «Социология религии» ПСТГУ. URL:<http://socrel.pstgu.ru>

организации социальной деятельности на уровне приходов (церковных общин), способных задействовать активную часть прихожан и сочувствующих. Ресурсной базой для такой деятельности являются, по большей части, время, рабочие руки, материальные средства прихожан и сочувствующих людей, иногда — организаций-партнеров из числа НКО, государственных и бизнес-структур. На базе такого партнерства реализуются масштабные благотворительные инициативы, способные мобилизовать большую долю населения (например, для сдачи донорской крови, помощи пострадавшим от катаклизмов, и проч.). Кроме того, социальная деятельность церкви на приходском уровне является примером самоорганизации граждан для решения важных социальных задач «снизу».

Наиболее активно социальная деятельность ведется на тех приходах РПЦ, где присутствует сильная община, активные прихожане¹. При этом эффективность мобилизации человеческих ресурсов для реализации каких-либо проектов зависит от характеристик структуры социальных связей в приходской общине и за ее пределами.

Однако характеристики структуры, а также вопросы формирования, организации и мобилизации православной общины для осуществления тех или иных проектов в настоящее время остаются малоизученными. Данное исследование направлено на преодоление данного пробела.

Мы рассматриваем приходскую общину как социальную сеть. В общем виде социальная сеть формально определяется как «набор социально-релевантных узлов, связанных одним или несколькими отношениями»². В качестве узлов могут выступать люди, организации, веб-страницы, публикации, страны, позиции и многое другое. В качестве связей изучаются сотрудничество, дружба, отношения обмена или власти, веб-ссылки, цитирование, потоки информации и различных видов ресурсов и т.д. В разных случаях определения сети в значительной степени различаются, однако на высоком уровне обобщения можно выделить три компонента таких определений, которые в каждом конкретном случае задают определение социальной сети. Чтобы определить социальную сеть, в первую очередь необходимо описать, что представляют собой узлы, которые в нее включены, во-вторых, необходимо ответить на вопрос о границах сети, и, в-третьих, — установить, какие отношения связывают узлы сети. В качестве первого шага определим понятие церковной общины, затем попытаемся определить, кто входит в социальную сеть православной общины, и покажем, как соотносятся эти два понятия.

¹ Об этом свидетельствуют данные опроса «Социальный работник на приходе. Восприятие представителями приходов введения на приходах Московской епархии должности социального работника», проведенного авторами в 2010 году. Наличие на приходе сильной общины на 15-25% повышает вероятность осуществления на нем тех или иных видов социального служения. Опрос социальных работников, священников и настоятелей московских храмов проводился на семинарах Комиссии по церковной социальной деятельности при Епархиальном совете г.Москвы в феврале-декабре 2010 г. Общий объем выборки — 458 человек. URL:<http://socrel.pstgu.ru/socialwork> (дата обращения: 12.02.2012)

² Martin A., Wellman B. Social Network Analysis: An Introduction. *Handbook of Social Network Analysis*. Edited by Peter Carrington and John Scott. Thousand Oaks, CA: Sage, 2011. P.11. URL:<http://homes.chass.utoronto.ca/~wellman/publications/index.html> (дата обращения: 12.02.2012)

Определение общины

В эмпирических социологических исследованиях жизни и деятельности Русской Православной Церкви, в ряде случаев целесообразно разделять понятия «общины» и «прихода». По ряду важных характеристик (репродуктивное поведение, ценности и установки и др.) люди, входящие в общину, отличаются от обычных прихожан.

В официальных документах РПЦ понятия прихода и общины взаимосвязаны. В частности, ныне действующий Устав гласит: «Приходом является община православных христиан, состоящая из клира и мирян, объединенных при храме... Границы приходов устанавливаются епархиальным советом»¹.

Данная формулировка не является исчерпывающей, в ней совмещаются два принципиально различных принципа построения мельчайших единиц христианского (внемонастырского) общежития. Первый принцип — территориальный. Так, в Приходском уставе Поместного Собора 1917-1918 гг. дано следующее определение прихода (закрепляющее существовавшую в досинодальный и синодальный периоды традицию): «Приходом в Православной Церкви называется общество православных христиан, состоящее из клира и мирян, пребывающих на определенной местности и объединенных при храме, составляющее часть епархии и находящееся в каноническом управлении своего епархиального архиерея, под руководством поставленного последним священника-настоятеля»². Этот принцип работал в ситуации, когда практически все население страны принадлежало к православному вероисповеданию³.

Второй принцип — общинный, — фиксирует иную ситуацию. В XX веке в силу усиления пространственной мобильности, территориальный принцип перестает играть значительную роль. Прихожане собираются в приход не только, а часто и не столько с близлежащей территории (в среде современного духовенства распространяется термин «приезжане»).

Устав РПЦ не дает однозначной формулировки того, кто / что является новой общностью православных. Тем не менее, руководствуясь приведенным выше кратким определением прихода, а также исследованиями, выполненными в рамках социологии прихода в Европе и США⁴, мы можем предложить следующие различия и операциональные

¹ Устав Русской Православной Церкви. Глава «Приходы». URL:<http://www.patriarchia.ru/db/text/133141.html> (дата обращения: 12.12.2011).

² Цит. по: Цыпин В.А. Церковное право. Часть III. Органы церковного управления. Приход в Русской Православной Церкви. Приходской устав Поместного Собора 1917-1918 гг. URL:<http://lib.eparhia-saratov.ru/books/22c/cipin/eccllaw/152.html> (дата обращения: 12.12.2011).

³ В католической церкви пространственный принцип приходского устройства был зафиксирован Тридентским собором (1545 — 1563).

⁴ См. подробнее: Donovan J.D. Sociology of Parish. New Catholic Encyclopedia. 1967. Vol.10. P.1019-1020; Sociology of the Parish / C.J. Nuesse, T. J. Harte eds. Milwaukee: Bruce Publishing Company, 1951; Schuyler J.B., s.j. Northern Parish. A Sociological and Pastoral Study, Chicago: Loyola University Press, 1960.; The Emerging Parish. The Notre Dame Study of Catholic Life since Vatican II / J. Gremillion, J. Castelli, eds. Notre Dame: University of Notre Dame Press, 1987; Hornsby-Smith M.P. The Changing Parish. A Study of Parishes, Priests, and Parishioners after Vatican II. London and New York: Routledge, 1989; Fichter J.H. Dynamics in the City Church, Chicago: The University of Chicago Press, 1951; Fichter J.H. Social Relations in the Urban Parish. Chicago: University of Chicago Press, 1954. На русском языке: Орешина Д. Прикладные социологические исследования католического прихода в США до реформ Второго Ватиканского собора (1962-1965 гг.). Материалы семинара «Социология религии» 2010-4; Орешина Д. Католический приход во второй половине XX века: факторы формирования приходской общины. Материалы

определения, позволяющие на основании эмпирических данных оценить распространенность данного феномена.

Приход православного храма в настоящее время (в начале XXI в.) состоит из трех частей: (1) ядро общины, (2) периферия общины, (3) внеобщинные православные.

Для идентификации человека как принадлежащего к той или иной части прихода могут быть использованы следующие три группы критериев: (1) участие в религиозных практиках (причастие, посещение богослужения и т.п.), (2) самоидентификация как члена общины, (3) участие во внебогослужебных видах деятельности (или осведомленность о них).

Для получения данных по общинам приходов, опросные процедуры проводились после литургии в будний день, на который приходился православный праздник «Покров Пресвятой Богородицы». Мы предполагаем, что в данных условиях мы зафиксировали пришедшую в храм общину — ее ядро и периферию (и не зафиксировали внеобщинных православных). Можно логически отделить внеобщинных прихожан от общины (состоящей из клира и мирян), ставящих значительный религиозный праздник (внутрицерковное событие) по значимости выше, нежели мирскую повседневную жизнь и корректирующих свое повседневное поведение в соответствии с церковными событиями.

Ядро общины мы выделяем по степени интенсивности участия в церковной жизни. В данном исследовании мы использовали следующие показатели для выделения ядра общины:

1. *Религиозная практика:*
 - посещение религиозных служб 2-3 раз в месяц или чаще
 - причастие 1 раз в месяц или чаще
2. *Самоидентификация:*

Ответ «да» на вопрос «Ощущаете ли Вы принадлежность к церковной общине, приходу?»
3. *Осведомленность о жизни прихода и прихожан:*

Ответ «да» на 4 или более вопросов из 6:

- На этом приходе много прихожан, которые общаются вне литургии
- На приходе есть прихожане, к которым я могу обратиться за помощью, и знаю, что мне не откажут
- На приходе есть группы прихожан, которые занимаются работой на приходе без участия священника (напр., уборкой храма, работой в трапезной...)
- Я знаю, что на этом приходе многие прихожане молятся друг за друга
- Прихожане этого храма регулярно собираются в трапезной на чаепитие
- Когда Вы приходите на воскресную службу в этот храм, обычно Вы видите людей по большей части Вам знакомых

Попытаемся дать количественную оценку распространенности такого феномена, как православная община, и соотнесем полученные нами данные с данными общероссийских опросов (таблицы 1-3).

По разным оценкам, к православному вероисповеданию относят себя 60-80% россиян. Доля воцерковленных православных при этом значительно меньше. Православные, посещающие церковные службы раз в месяц или чаще составляют, по разным оценкам, до 10% от населения России в возрасте 18 лет и старше. Исповедующиеся и причащающиеся раз в месяц или чаще — 2-3%.

На основании приведенных данных можно оценить долю ядра православных общин по отношению к населению России на уровне 0,5-3%, а долю общины в целом (ядра и периферии) — на уровне порядка 7-10%.

Таблица 1

Общая распространенность православия по данным опросов общественного мнения (%)

	Россия в целом*
православие	78
другие христианские конфессии	1
ислам	6
другие религии	1
не отношу себя ни к какому вероисповеданию	11
убеждённый атеист	2
затрудняюсь ответить	1
База (количество респондентов)	1500

* Вопрос: «Относите ли Вы себя к какому-либо вероисповеданию? Если да, то к какому именно?»

Таблица 2

Частота посещения церковных служб (% по столбцу)

	Община в целом*	Периферия общины	Ядро общины	Россия в целом*
несколько раз в неделю	36	22	51	3
раз в неделю	37	34	41	
два-три раза в месяц	14	19	8	2
раз в месяц	6	13		3
несколько раз в год	5	10		20
раз в год	<1	<1		10
реже, чем раз в год	<1	1		10
очень редко, почти никогда	<1	<1		28
нет ответа / затрудняюсь ответить	<1	1		1
База (количество респондентов)	1028	518	510	1500

* Вопрос «Посещаете ли Вы религиозные службы? Если да, то как часто?». Ответы всех респондентов, относящих себя к православию (78%)

Частота причащения (% по столбцу)

	Община в целом*	Периферия общины	Ядро общины	Россия в целом**
никогда / практически никогда	5	11		50
реже одного раза в год	3	5		11
1-2 раза в год	5	10		9
несколько раз в год, но реже, чем раз в месяц	14	27		3
1 раз в месяц	20	18	21	1
2-3 раза в месяц	29	13	45	<1
1 раз в неделю или чаще	21	8	34	<1
нет ответа	1	2		
затрудняюсь ответить	3	6		3
База (количество респондентов)	1028	518	510	1500

* Вопрос: «Как часто Вы причащаетесь в этом храме?»

** Вопрос: «Как часто Вы исповедуетесь и причащаетесь?»

Ответы всех респондентов, относящих себя к православию (78%)

Социальная сеть православной приходской общины

В отличие от собственно приходской общины, социальная сеть приходской общины включает в себя как воцерковленных, так и невоцерковленных людей, связанных с воцерковленными и периодически включающихся в совместную деятельность, общение, и т.п. Подобное понимание социальной сети церковной общины позволяет получить иную оценку влияния православия в стране, нежели показывают данные опросов, оценивающих только практикующих верующих.

Рассмотрим имеющиеся на этот счет результаты эмпирических исследований.

В ближайшем окружении (среди друзей и родственников) половины россиян есть люди православного вероисповедания, регулярно посещающие церковные службы (таблица 4). Логично, что доля имеющих таких людей в ближайшем окружении значительно изменяется в зависимости от степени воцерковленности¹ человека. У наиболее воцерковленной части православных она составляет 74%, в то время как для тех, кого с православием связывает только культурная самоидентификация, этот показатель снижается до 47%.

Тем не менее, даже среди мусульман и людей, не относящих себя ни к какому вероисповеданию, достаточно высок процент имеющих связи с членами православных общин (36% и 29% соответственно).

¹ На основании вопросов о вере в Бога, частоте посещения церковных служб, а также о частоте исповеди и причащения всех относящих себя к православному вероисповеданию можно разделить на следующие 5 групп: *Воцерковленные (13%)*: верят в Бога, посещают службы несколько раз в год, причащаются 1-2 раза в год. *Полувоцерковленные (18%)*: верят в Бога или колеблются, исповедуются и причащаются хотя бы изредка. *Слабовоцерковленные (32%)*: верят в Бога или колеблются, посещают службы хотя бы изредка, не причащаются или затрудняются ответить о причастии.

Невоцерковленные (19%): верят в Бога, но не посещают службы, не исповедуются и не причащаются.

Культурная самоидентификация с православием (17%): не верят в Бога или колеблются, не посещают службы или затрудняются ответить, не причащаются или затрудняются ответить о причастии.

Таблица 4

Есть ли среди Ваших родственников или друзей люди православного вероисповедания, регулярно посещающие церковные службы? (% по строке)

	есть	нет	затрудняюсь ответить	База (количество респондентов)
Россия в целом	51	38	11	1500
воцерковленные православные	74	16	10	151
полувоцерковленные православные	64	24	12	214
слабовоцерковленные православные	56	31	13	378
невоцерковленные православные	47	47	6	221
культурная самоидентификация с православием	43	47	10	202
ислам	36	51	13	96
без вероисповедания	29	56	15	164

Мы также можем рассмотреть социальную сеть православной общины на уровне эго-сетей прихожан — представителей ядра общины и ее периферии и оценить состав и размер этих эго-сетей.

В социальной сети православной городской общины присутствует много негомологичных (т.е. различных по базовым характеристикам) узлов, поскольку община связывает людей разного социального статуса. Так, например, если мы рассмотрим ответы на вопрос о том, на какой должности работают опрошенные члены приходских общин, то увидим заметное многообразие ответов (таблица 5). Здесь встречаются студенты, пенсионеры, домохозяйки, служащие, специалисты, рабочие, руководители разных уровней, частные предприниматели, собственники предприятий и др. Связи между входящими в церковную общину людьми выходят за рамки их родственных и профессиональных кругов, что с высокой вероятностью влечет за собой существование значительного числа мостов¹ в социальной сети церковной общины. Таким образом, сети взаимопомощи входящих в ядро приходской общины людей оказываются мощнее, чем у других людей: связей (вероятно, слабых) — большее количество, поскольку община связывает людей разного социального статуса, которые в иных ситуациях оказываются не связанными друг с другом.

¹ В одной из наиболее цитируемых в рамках анализа социальных сетей работ М. Грановеттер вводит понятие «моста»: «под мостом понимается ребро в сети, которое обеспечивает единственный путь между двумя точками» (Грановеттер М. Сила слабых связей, 2009, с.35, ссылаясь на Harary F., Norman R., Cartwright D. Structural Models. N. Y.: Wiley, 1965, p.198) и обосновывает важность мостов: «Поскольку в целом у каждого человека огромное множество контактов, то мост между А и В обеспечивает единственный путь, по которому проходит информация или распространяется влияние от любого контакта индивида А к любому контакту индивида В, а следовательно, от любого человека, опосредованно связанного с А, к любому человеку, опосредованно связанному с В. Таким образом, в ходе изучения диффузии мы можем ожидать, что мосты будут играть важную роль» (Грановеттер М. Сила слабых связей, 2009, с.35).

С целью определить состав и размер эго-сетей поддержки в православных общинах, а также по России в целом, мы задали вопрос: «Если Вы попадете в кризисную ситуацию (напр., потеря работы, проблемы в семье или болезнь), к кому Вы можете обратиться за помощью? Как Вам кажется, сколько всего таких людей, к которым Вы могли бы обратиться за помощью в кризисной ситуации?»¹

По числу людей, к которым респондент может обратиться в кризисной ситуации, сети членов ядра общин значительно шире, нежели у представителей периферии (таблица 6)². Так, более 10 человек в своей сети поддержки имеют 32% представителей ядра общины и только 13% периферии. По России в среднем этот показатель еще ниже — всего 5%.

Таблица 5

Работаете ли Вы, и если да, на какой должности? (% по столбцу)

	Община в целом	Периферия общины	Ядро общины
не работаю (домохозяйка и т.д.)	11	10	12
не работаю, ищу работу	2	3	1
пенсионер и не работаю	22	25	19
учусь и не работаю	5	4	5
рабочий, сельскохозяйственный рабочий	2	3	1
служащий	12	11	13
имею свой мелкий бизнес (ПБОЮЛ, ЧП), оказываю частные услуги, самозанятый	3	3	3
специалист	16	13	19
руководитель подразделения	7	7	7
руководитель предприятия	1	2	1
предприниматель, собственник фирмы	1	2	<1
другое	12	10	15
нет ответа	6	8	4
затрудняюсь ответить	0	0	0
База (количество респондентов)	1028	518	510

¹ Вопрос о размере сети поддержки задавался без подсказок, респондентов просили записать ответ числом. Затем все ответы были объединены в категории, представленные в таблице 6. В категорию «более 10» мы также включили вариант ответа «много», в категорию «1-3 человека» — ответы «несколько», «немного».

² $p < 0,001$ (для определения значимости различий использовался критерий Манн-Уитни — непараметрический аналог теста Стьюдента).

Таблица 6

Как Вам кажется, сколько всего таких людей, к которым Вы могли бы обратиться за помощью в кризисной ситуации? (% по столбцу)

	Община в целом	Периферия общины	Ядро общины	Россия в целом
нет таких людей	2	3	1	6
1-3 человека	10	14	6	32
4-6 человек	15	16	13	26
7-10 человек	17	14	20	14
более 10	22	13	32	5
нет ответа	5	8	2	
затрудняюсь ответить	29	31	27	17
База (количество респондентов)	1028	518	510	1500

Следующий важный вопрос, на который необходимо ответить, определяя некоторое явление как сеть, — это какие отношения существуют между узлами сети. Отношения — это «контакты, связи, объединение, принадлежность к группе, встречи — все, что связывает одного агента с другим и поэтому не может быть сведено к свойствам самих агентов»¹. Например, можно выделить следующие широкие категории отношений²: 1) сходство: возникает, когда два или более узла обладают схожими атрибутами (социально-демографические характеристики, установки, членство в определенной группе и т.д.); 2) социальные отношения: отношения родства, ролевые отношения (друг, коллега, студент и т.п.), аффективные связи (нравится / не нравится), когнитивные связи (знание); 3) взаимодействие: связи, основанные на поведении (помощь, общение, визиты и т.п.); 4) потоки: перемещение различного рода ресурсов, информации, влияния и т.д. по сети между узлами.

Мы рассмотрим один из видов связей, а именно — включенность в практики взаимной помощи по трем направлениям: затраты сил и времени на оказание услуг в виде помощи по хозяйству, материальная помощь и моральная поддержка (таблица 7) и попытаемся оценить силу этих связей через их регулярность (интенсивность).

Респондентам задавался вопрос: «В течение последних 12 месяцев, как часто Вы делали что-либо из следующего списка для кого-нибудь из родственников, друзей, соседей или знакомых? ... помогали кому-либо вне дома по хозяйству или с покупками; давали немного денег взаймы; уделяли время, чтобы поговорить с расстроенным / подавленным человеком». Данный вопрос задавался в одной из волн «Международной программы социальных исследований» (ISSP — International Social Survey Program), посвященной теме социальных сетей (2001 г.). По сравнению с указанным исследованием, мы немного расширили список возможных вариантов ответа. Эти вопросы мы также повторили во всероссийском опросе, проведенном Институтом Фонда «Общественное мнение» по заказу Православного Свято-Тихоновского гуманитарного университета в

¹ Scott J. Social network analysis: a handbook, Sage Publications, 2005, p.3.

² Borgatti S., Mehra A., Brass D., Labianca G. Network Analysis in the Social Sciences. *Science*, Vol. 323, No. 5916, 2009, pp. 892-895.

декабре 2011 года. Это позволяет нам сравнить данные по приходским общинам с результатами, характеризующими Россию в целом.

Таблица 7

В течение последних 12 месяцев, как часто Вы делали что-либо из следующего списка для кого-нибудь из родственников, друзей, соседей или знакомых? (% по столбцу)

	Община в целом	Периферия общины	Ядро общины	Россия в целом
<i>... помогли кому-либо вне вашего дома по хозяйству или с покупками</i>				
несколько раз в неделю	9	9	9	12
примерно 1 раз в неделю	11	11	11	11
2-3 раза в месяц	11	9	14	15
примерно 1 раз в месяц	13	12	15	14
не менее 2-3 раз за год	16	15	16	15
1 раз за год	4	3	4	6
ни разу за последний год	5	4	5	20
нет ответа	22	25	19	-
затрудняюсь ответить	9	11	8	8
<i>... давали немного денег взаймы</i>				
несколько раз в неделю	3	2	4	2
примерно 1 раз в неделю	3	3	3	5
2-3 раза в месяц	8	8	8	13
примерно 1 раз в месяц	13	10	16	18
не менее 2-3 раз за год	25	24	27	24
1 раз за год	11	11	10	8
ни разу за последний год	6	6	5	22
нет ответа	24	28	19	-
затрудняюсь ответить	8	8	7	9
<i>... уделяли время, чтобы поговорить с расстроенным / подавленным человеком</i>				
несколько раз в неделю	22	17	27	12
примерно 1 раз в неделю	16	14	18	13
2-3 раза в месяц	16	15	17	17
примерно 1 раз в месяц	14	13	14	16
не менее 2-3 раз за год	8	10	6	15
1 раз за год	3	3	3	6
ни разу за последний год	1	1	1	12
нет ответа	13	17	8	-
затрудняюсь ответить	8	9	6	10
База (количество респондентов)	1028	518	510	1500

В целом по указанному блоку вопросов в приходских общинах оказался очень высок процент не ответивших. При этом многие респонденты не просто пропускали данный вопрос, но и писали в комментариях, что о своих добрых делах нельзя говорить и писать.

По части затрат собственных сил и времени (помощь по хозяйству) между членами ядра общин и периферией нет существенной разницы, от общероссийской выборки представители приходских общин отличаются только по доле не оказывавших этот вид помощи за последний год (20% по России и всего 5% среди представителей православных общин).

Материальную помощь в целом оказывают заметно реже, чем моральную поддержку — как по России в целом, так и представители опрошенных приходских общин. 22% россиян не оказывали материальной помощи деньгами в течение последних 12 месяцев, в то время как среди опрошенных представителей приходских общин эта группа в заметно меньше (6%), при этом ядро и периферия практически не различаются между собой.

Что касается моральной поддержки, то здесь различия между средними россиянами и православной общиной оказались более существенными: в ядре общины 27% респондентов за последний год несколько раз в неделю уделяли время, чтобы поговорить с расстроенным / подавленным человеком, по России в целом таких людей в два раза меньше — 12%. Среди опрошенных представителей общины практически отсутствуют те, кто ни разу за последний год не оказывал моральную поддержку близким (по России — 12%).

Таким образом, можно сделать вывод о том, что связи представителей приходской общины, и особенно — ее ядра, характеризуются более высокой регулярностью и интенсивностью. В ядре общины более выражена взаимопомощь и моральная поддержка, по сравнению с периферией общины и среднестатистическими россиянами.

Заключение

По ряду значимых параметров члены приходских общин отличаются от среднестатистических россиян. Особенно заметны различия в ценностной сфере, что в свою очередь частично обуславливает различия в некоторых практиках (в том числе и взаимной поддержки). Размер эго-сетей поддержки в православных общинах (особенно — в ядре общин) значительно больше, нежели в среднем по России. При этом отношения в эго-сетях поддержки представителей приходской общины, по сравнению со среднестатистическими россиянами, характеризуются более высокой силой связи.

Социальная сеть православной общины, включающая в себя не только воцерковленную часть населения, но и ближайшее окружение воцерковленных людей, представляет собой явление заметно более широкое, чем принято считать в рамках традиционного подхода к изучению религии с точки зрения атрибутов и групп. В то же время, фокусирование внимания на отношениях, связывающих членов церковных приходских общин друг с другом, с людьми невоцерковленными, с представителями других конфессий, с обществом в целом, может принести заметные плоды для социологического осмысления роли религии в жизни общества.



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Анализ структуры бизнес-конфликтов в российских СМИ при помощи поисковых алгоритмов

Галина Витальевна Градосельская

Национальный исследовательский университет
«Высшая школа экономики» (Москва)

Информационное пространство постоянно служит источником пополнения базы для исследования различных социальных явлений: в экономике, культурно-религиозной сфере, политике и т.д.

Смысловая структура текстов при этом приобретает особое значение. Для исследователя информационного пространства структура текстов может быть полезной в двух направлениях – как способ анализа социальных явлений, так и для сбора информации.

Возникает новое направление развития структурного анализа тестов – в поисковых алгоритмах, как целевом механизме отбора информации из широкого ежедневного информационного потока.

Опишем прикладные возможности структурного анализа текстов на примере анализа информационного пространства по бизнес-конфликтам в России.

Практической целью исследования был поиск потенциально проблемных предприятий, где бизнес-конфликт находится в начальной стадии. Необходимо было изучить функционирование социального института бизнес-конфликтов по данным сообщений СМИ, вышедших в открытых источниках в посткризисный период (май 2010 года – октябрь 2011 года), на основе этих данных построить прогнозные сценарии развития конфликтных ситуаций в трех срезах: отраслевом, региональном и по типу конфликта.

Промежуточной методической задачей был отбор публикаций в российских СМИ за определенный период, который помог бы выявить бизнес-конфликты и их акторов.

Хотим оговориться, что данная статья касается в основном методологической стороны реализации данного исследования. Более подробно его результаты будут отражены в других публикациях. Исследование проводилось на русском языке, но в дальнейшем разработанная методология может быть адаптирована и к другим языкам.

Методологическая проблема, которая всплывает в самом начале исследования – **определение бизнес-конфликтов**, поскольку оно является очевидным для многих участников (акторов) этих конфликтов, но в то же время является весьма ситуативным. В подавляющем большинстве публикаций, где описывались бизнес-конфликты, словосочетания «бизнес-конфликт» не встречалось. Как только проблемная ситуация получает статус бизнес-конфликта, круг участников

бизнес-конфликта становится явным, и сценарии развития конфликтов становятся ограниченными.

Социальное явление бизнес-конфликтов реальность последних двух десятилетий российской истории. Особенностью этого явления - оно минимально формализовано, поскольку существует именно благодаря «дырам» в законодательстве. Вырабатываются определенные стандартные ходы по относительно честному отъему или переделу бизнеса между собственниками, акционерами, менеджментом. Широкий круг привлекаемых акторов: это и трудовые коллективы, и экологические организации.

Поэтому принципиально необходимо участие эксперта в формализации определения бизнес-конфликтов.

Во-первых необходимо классифицировать существующие на данный момент конфликты.

Во-вторых, определение конфликтов проводится через подбор публикаций-иллюстраций, которые эксперт может отнести к тому или иному типу.

В-третьих, эксперт поможет определить статьи, которые не подходят для целей исследования, или определить правила для исключения нерелевантных статей.

1. Структурирование текста на этапе заполнения информационной базы данных

Для алгоритмизации и разработки методологии обработки бизнес-конфликтов в информационном пространстве необходимо предпринять следующие аналитические шаги:

- **Типологизация бизнес-конфликтов.** Виды бизнес-конфликтов разнообразны. Это зависит от акторов, участвующих в конфликте, так и от существа взаимных претензий. Типологизация бизнес-конфликтов (во всяком случае на первом этапе) возможна только с помощью экспертов. В дальнейшем возможна коррекция типологии по имеющемуся статистическому материалу.
- **Операционализация понятия «бизнес-конфликт».** Это сложно, поскольку в подавляющем большинстве статей словосочетание «бизнес-конфликт» не упоминается. Необходимо определить набор речевых маркеров, которые косвенно смогут указать на наличие бизнес-конфликта, описываемого в статье, а также на тип этого конфликта.
- **Понимание сопровождающих (социальных, политических) обстоятельств.** Некоторые типы конфликтов могут быть спровоцированы общими системными обстоятельствами – например, мировым финансовым кризисом. Иногда рост конфликтов связан с отраслевым кризисом. Также понимание внешних обстоятельств поможет скорректировать суть запроса, вытаскивающего необходимые нам публикации из общего информационного потока, поможет отсечь лишний информационный шум.

Базовая классификация российских бизнес-конфликтов

На начальной стадии исследования в базовую классификацию российских бизнес-конфликтов были включены 15 типов конфликтов, из которых было выделено 11 потенциально публичных типов (Табл. 1).

Таблица 1

Базовая классификация бизнес-конфликтов*

Конфликты во внутренней среде компании	Внешние конфликты
<ul style="list-style-type: none">■ [A1] конфликты между акционерами, раздел имущества, экстренные продажи активов;■ [A2] конфликты между подразделениями, приводящие к сбоям в работе компании**;■ [A3] конфликты собственников или менеджмента с трудовым коллективом;■ [A4] конфликты между акционерами (имеющими контрольный пакет) и менеджментом: <i>>> [A4.1] попытки захвата собственности менеджментом;</i> <i>>> [A4.2] трения между менеджментом и акционерами, когда акционеры не могут упорядочить отношения (грамотно делегировать полномочия) с нанятыми управленцами*.</i>	<ul style="list-style-type: none">■ [B1] конфликты с властью, в т.ч. с контролирующими и правоохранительными органами;■ [B2] конфликты компании с общественностью и общественными организациями;■ [B3] конфликты компании с захватчиками, в том числе миноритариями, которые хотят больше, чем дивидендов;■ [B4] конфликты миноритариев, чьи права ущемляются, с владельцами контрольного пакета;■ [B5] конфликты с конкурентами;■ [B6] конфликты кредиторов и должников;■ [B7] конфликты с поставщиками и подрядчиками (операционные, т.е. конструктивные: с целью надавить, чтобы те поставили/закупили товар в оговоренных объемах);■ [B8] конфликты с клиентами (в случае массового обслуживания: с абонентами, клиентами банков, покупателями квартир, пассажирами, туристами и т.п.).

* Приносим искреннюю благодарность Р.Е.Бумагину, который разработал типологизацию бизнес-конфликтов

** Конфликты, которые по оценкам экспертов не имеют публичного измерения

Генерирование списка ключевых слов на основе анализа кейсов

По каждому типу конфликтов были подготовлены несколько статей (от 2 до 9), описывающие кейсы, названные экспертами.

На основе этих новостных и аналитических материалов СМИ для каждого типа конфликтов был сформулирован набор ключевых слов, позволяющих идентифицировать появление данного типа конфликтов в новостном потоке.

По каждой, наиболее типичной новостной публикации, описывающей бизнес-конфликт, был проведен дискурс-анализ – реконструирована смысловая структура текста, показывающая ключевые речевые маркеры, во взаимосвязи.

На рисунке, показанном ниже, проведен дискурс-анализ статьи «"Фокус" "слетел" с конвейера», выбранной как пример для иллюстрации типа бизнес-конфликтов АЗ – конфликты собственников или менеджмента с трудовым коллективом.

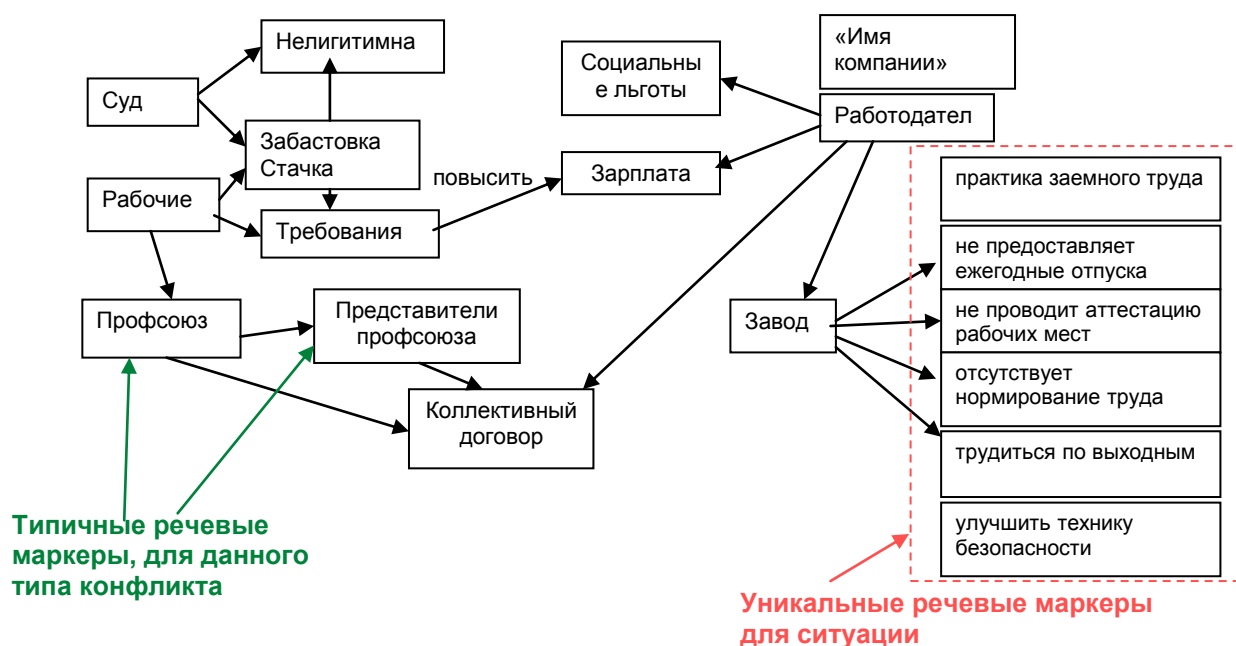


Рисунок 1. Пример дискурс-анализа статьи «"Фокус" "слетел" с конвейера»

После того, как проведен дискурс анализ почти по всем предоставленным статьям, иллюстрирующим каждый тип конфликтов, по каждому типу конфликта был отобран набор речевых маркеров, группированных по степени важности для будущего семантического запроса (пример см. на таблице ниже).

Ключевые речевые маркеры для семантического запроса

1 уровень (ядро запроса)	<ul style="list-style-type: none"> ▪ работодатель, администрация/ представители ▪ рабочие, заводчане, сотрудники, докеры, участники ▪ профсоюз, стачком/ представители ▪ акция, протест, голодовка, забастовка, стачка, вышли, перекрыть (улицу) ▪ (требование) долги, повышение, увеличение, зарплата, выплата
2 уровень (уточнение запроса)	<ul style="list-style-type: none"> ▪ прокуратура, суд, арбитраж ▪ иск ▪ незаконный ▪ банкрот, банкротство, внешнее управление, внешний управляющий ▪ администрация области
3 уровень	<ul style="list-style-type: none"> ▪ ухудшение здоровья, обмороки, хронические заболевания

По завершении формирования списка ключевых слов был приобретен доступ к специализированной базе данных «Медиалогия», которая насчитывает более 6 400 источников на постсоветском пространстве (с учетом требований технического задания число источников дисконтируется до 5000 единиц). Таким образом, объем выборки данного исследования составил при средней частоте выходов 6 раз в неделю и среднем количестве сообщений в одно такое выходе порядка 10 единиц составил для зондажного исследования (период – два месяца) более 380.000 сообщений, а для ретроспективного контент-аналитического исследования за период с 01.05.2010 по 31.10.2011 года – более 3.000.000 сообщений.

После ручной чистки базы данных поисковых ответов осуществлялась конвертация базы из системы «Медиалогия» в пакет SPSS для дальнейшей подготовки данных к статистической обработке.

База данных, получаемых по запросу в мониторинговой системе «Медиалогия» включал в себя следующие поля: дата, СМИ, город (регион), URL статьи

Далее эти данные транслировались в массив SPSS, где добавлялись новые поля, которые заполнялись вручную по контексту статьи (отрасль кодировалась путем обращения по каждому случаю в исходный массив «Медиалогии»): тип конфликтов, дата – для временных рядов, регион, отрасль.

2. Динамика конфликтных ситуаций в России

Конфликтная активность неравномерно распределена во времени. Довольно явно выделяются осенние и весенние пики конфликтов.

Поскольку для наблюдения был отобран не очень длинный ряд, выделить сезонную и трендовую составляющие временного ряда не представляется возможным (каждый месяц должен быть представлен как минимум 3-4 раза – это означает 3 года наблюдений). Кроме того, предварительный анализ данных в SPSS показал скорее кубическую или экспоненциальную зависимость, чем линейную – что также затрудняет

сезонную декомпозицию. На графике также видно, что амплитуда конфликтных проявлений увеличивается все сильнее – поэтому сезонные отклонения также будут неравномерными.

Метод авторегрессии оказался единственным способом, позволяющим спрогнозировать значения конфликтов в России на полгода вперед.

В таблице, представленной ниже, показаны результаты прогнозирования конфликтов в России методом авторегрессии (реализованном в пакете SPSS). Показан значительный рост конфликтных ситуаций к весне следующего года.

Эти же прогнозные значения представлены на рисунке ниже.

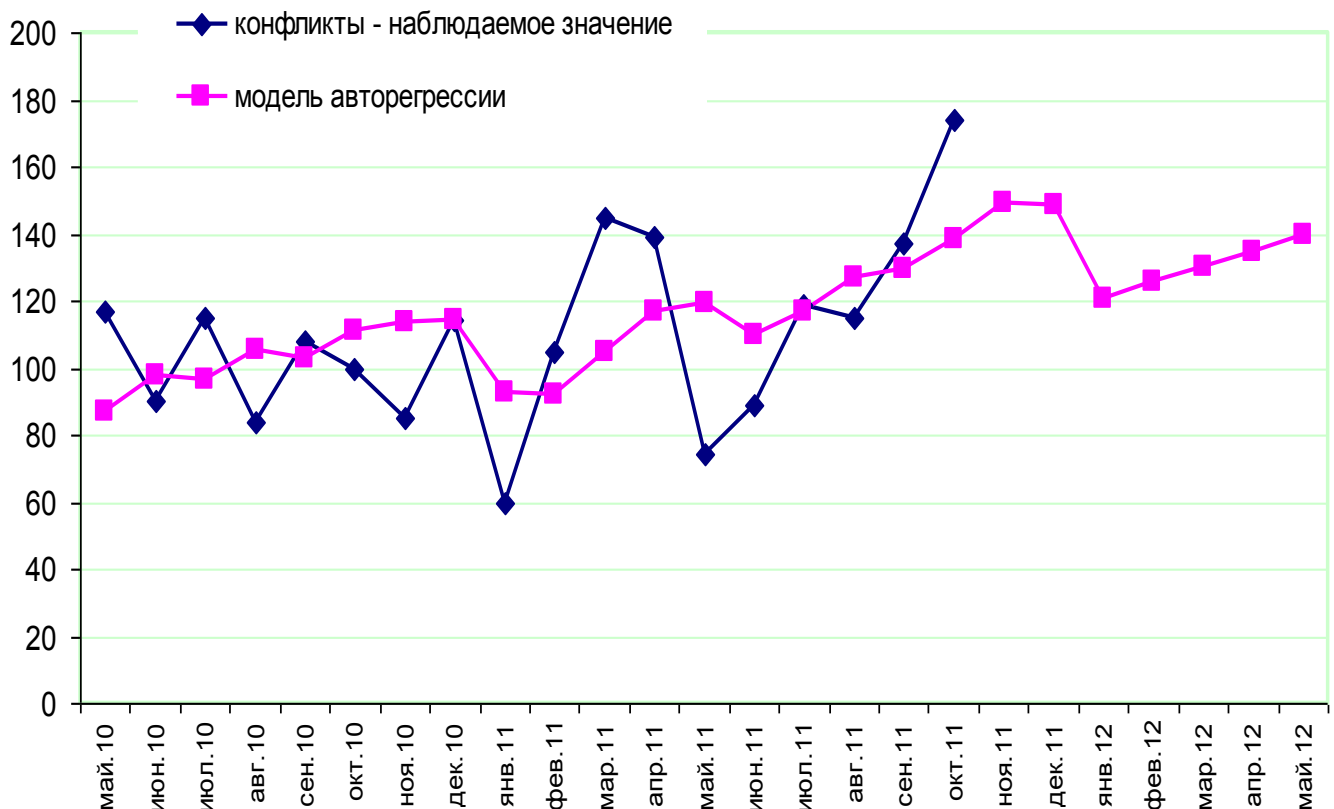


Рисунок 2. Прогнозирование развития конфликтов в России методом авторегрессии

Также сложно было провести прогнозирование по отдельным отраслям или типам конфликтов: короткий временной ряд и малая наполненность по каждому месяцу не позволила провести прогнозирование. Однако эта ситуация может быть исправлена с продолжением наблюдений и сбором информации по бизнес-конфликтам в России.

Ниже показано, как распределяется динамика для конфликтов разного типа. Практически для всех конфликтов пик приходится на март-апрель 2011 года, однако к осени этого года (октябрь 2011) возможен следующий пик конфликтов.

3. Отраслевой фактор в развитии конфликтов

Существенным фактором в распространении и воздействии конфликтов является отрасль. На рисунке ниже показано, что самыми конфликтными являются отрасли: производство, машиностроение (23,5% всех конфликтов), банковская, финансовая сферы (15,3%), строительство (14,5%). Далее следуют транспорт (10,2%), торговля (9,1%), сельское хозяйство (7,6%).

На рисунках ниже показано, что для разных отраслей преобладают разные типы конфликтов.

Конфликты между собственниками/ менеджментом и трудовыми коллективами ([A3]) преобладают в сферах: производство и машиностроение; транспорт, горнодобывающая, металлургия.

Конфликты кредиторов и должников ([B6]) преобладают в сферах банковской и финансовой, в торговле, а так же в производстве и машиностроении.

Конфликты компании с захватчиками распространяются довольно равномерно, но преобладают в производстве и машиностроении, сельском хозяйстве, торговле и банковской сфере.

Конфликты между акционерами и менеджментом чаще всего встречаются на производстве и машиностроении, в банковской сфере, строительстве.

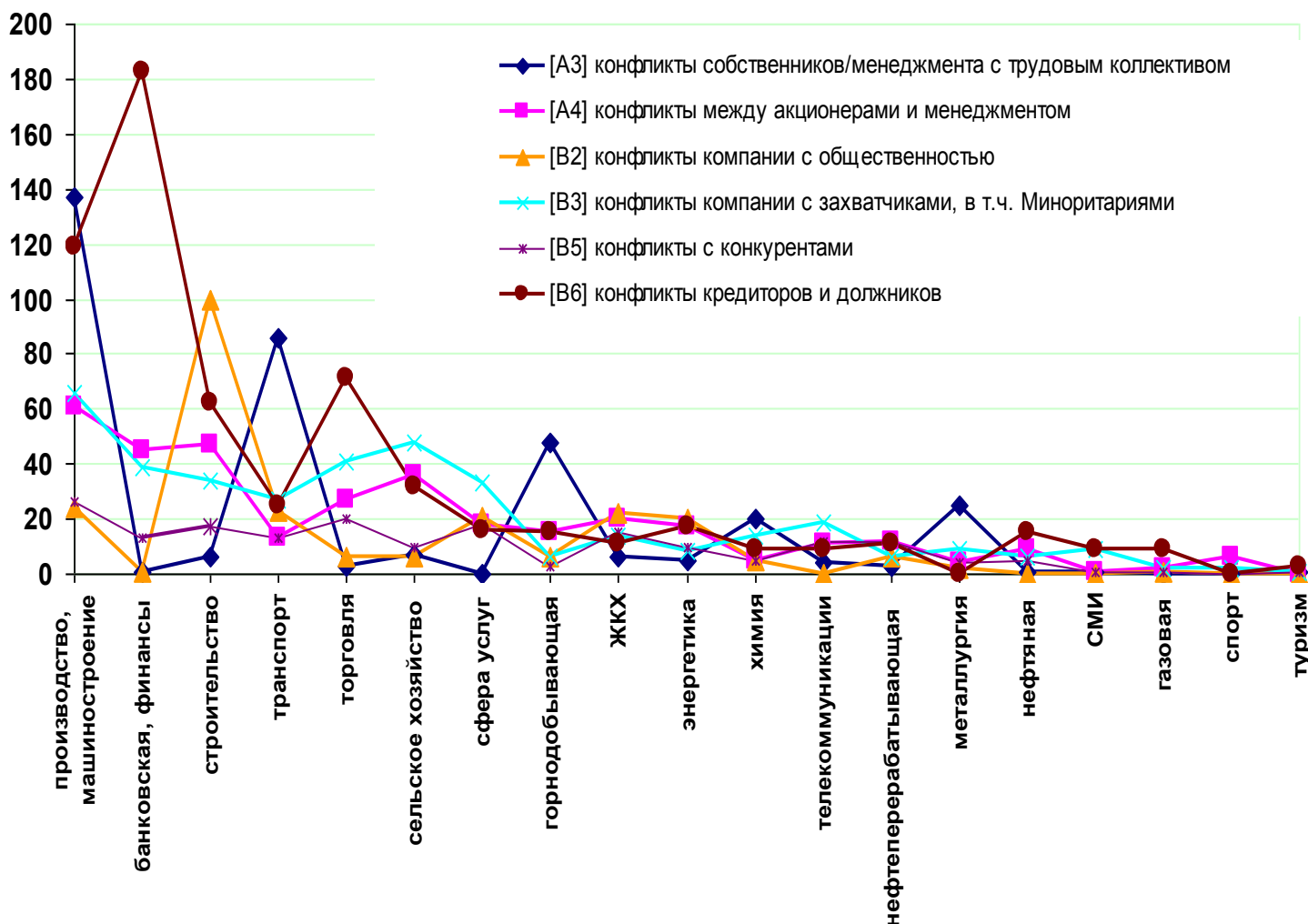


Рисунок 3. Распределение типов конфликтов по отраслям: определение пиков конфликтов

По результатам настоящего исследования была сформирована база данных по конфликтам в России за 1,5 года. Она позволяет проанализировать общую бизнес-конфликтность по регионам (после соответствующей чистки). Также возможно будет уточнить семантический запрос, основываясь уже на статистической информации.

Имеющаяся информация и опыт обработки данных дает возможность развивать проект в трех направлениях:

- методологическом (уточнение структуры информационного поля о бизнес-конфликтах);
- содержательном (приложения анализа информационного пространства в рамках бизнес-интересов клиента);
- технологическом (базы данных, семантика запроса, поисковый-паук, ПО).

Этот проект имеет весьма широкие перспективы практического применения. Содержательные возможности развития проекта:

- Уточнить структуру запросов и запустить их в постоянном мониторинговом режиме в интернет-паука. Возможно обучение соответствующей нейронной сети для сбора информации о бизнес-конflikтах России в режиме мониторинга. Постоянный интеллектуальный мониторинг информационного пространства позволит получить конкурентное преимущество для компаний, использующих разработанный нами софт.
- Накапливать всю информацию по конфликтам в России, что позволит прогнозировать дальнейшее перемещение конфликтной активности между отраслями, регионами и типами конфликтов.
- Собирать досье на компании, их собственников и топ-менеджеров – кто проявляет повышенную конфликтность, а так же повышенные потенциальные проблемы для будущих партнеров, властных элит регионов и т.п..
- Собрать список источников (в т.ч. неформальных), распространяющих информацию о компаниях – определить источники информационных угроз. Это потребует соответствующей корректировки семантических запросов. И расширение поля запросов.
- Изучить взаимосвязь типов конфликтов и наиболее распространенные сценарии их развития. Это позволит проводить более целевой консалтинг, заранее подготовить ответные меры (если мы защищаем клиента), и разрабатывать неожиданные ходы, если мы планируем активно воздействовать.



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

Методологические и методические основания изучения полилога, как формы коммуникативных процессов

Екатерина Сергеевна Михеева

Несомненно, основным инструментом взаимодействия индивидов, в современном обществе, является язык. С помощью языка человек осмысляет недискретную действительность, интерпретируя ее в дискретных терминах языковых категорий. Происходит и другой процесс – недискретный многомерный мир смысла преобразуется в развернутую по временной оси дискретную последовательность языковых выражений. Его изучению и анализу посвящено множество работ, как области специализирующихся на этом наук, таких как филология, лингвистика, семиология и т.п., так и наук, предмет изучения которых, тем или иным образом опосредован языком (психология, история, культурология и т.д.). Социология, чаще всего, обращается к языку как форме коммуникации, позволяющей изучить не только некоторый конкретный социальный феномен, но и описать функционирование общества в целом.

Обращения к данной проблематике вызвано необходимостью изучения речевого взаимодействия и коммуникации, как явления, включающего индивида в группы разного уровня, с учетом членения количества коммуникантов. Вступая во взаимодействие посредством вербальной коммуникации, человек реализует категорию воздействия: т.е. он не просто получает и передает информацию, а оказывает влияние и достигает некоторых целей. Общение посредством языка выступает как определяющий фактор при совершении того или иного действия. Язык, посредством которого человек выражает все свои чувства, потребности, эмоции, желания, является своеобразным «инструментом социального действия». По мнению Хабермаса, для достижения консенсуса коммуниканты должны принять установки, сопряженные с коммуникативными ролями 1,2 и 3 лица, а так же с каждым из трех миров (существующего положения вещей, межличностных отношений и собственных эмоций) [13].

В процессе общения партнеры не только обмениваются имеющейся у них информацией, но и на базе процесса взаимодействия коммуниканты трансформируют старую информацию с учетом ориентировки на партнера, а также формируют новую информацию, получая ответные сигналы о своем поведении от партнера. Так, по мнению М.М. Бахтина, высказывание никогда не является только отражением чего-то вне его уже существующего и данного, а всегда создает нечто до него никогда не бывшее, нечто новое, неповторимое [2, с. 229].

В данной работе нас будут интересовать не все коммуникативные процессы, а только те, которые оказывают воздействие на их участников. Под речевым воздействием мы будем понимать речевое общение, взятое в аспекте его целенаправленности, мотивационной обусловленности [6]. В коммуникации существенную роль играют коммуникативные (или социальные) роли, так как именно они во многом определяют способ использования коммуникативных средств и выбор их вариантов. При исследовании коммуникативных ролей необходимо учитывать, что между ролью и статусом существует двусторонняя связь [3]. Как социальный статус может обуславливать ролевые отношения коммуникантов, так и по распределению коммуникативных ролей можно судить о социальном статусе акторов.

При изучении дискурса, как процесса, важно, в зависимости от исследовательских задач, правильно выделить минимальную единицу анализа. Для этого следует различать такие понятия как речевая тактика, стратегия, коммуникативный акт, коммуникативный ход, речевой акт, минимальная диалогическая единица.

Под речевой тактикой будем понимать ряд коммуникативных действий, направленных на реализацию стратегии, в то время как стратегия осуществляется через использование коммуникативных приемов более низкого порядка – коммуникативных ходов. Стратегия – это некоторая последовательность речевых действий организованных в зависимости от цели взаимодействия. Коммуникативный акт – совокупность высказываний актора между сменами коммуникативной инициативы, в зависимости от реализуемых целей может состоять из одного и более коммуникативных ходов. Упорядочение целей преследуемых коммуникантами помогает обнаружить базисные операции, лежащие в основе речевых манипуляций, и через них выявить связь коммуникативных ходов и тактик.

Коммуникативные формы повседневного общения

Как мы уже говорили, коммуникация является средством воздействия и достижения своих целей индивидом. Чаще всего мы вступаем во взаимодействие с другими индивидами посредством повседневной устной речи. Разговорная речь, по сути, является связующим элементом между индивидами как членами общества, позволяя с той или иной степенью успешности осуществлять взаимодействие. Она имеет существенные особенности на всех языковых уровнях, и поэтому является самостоятельной языковой системой.

Общение не однородно и может различаться по многим параметрам: языковые средства, место, время, ситуация, поведение участников и п.т. Наиболее часто коммуникативные процессы подразделяют на типы исходя из количества участников коммуникации. Традиционно выделяют такие виды разговорной речи как монолог, диалог и полилог.

Под монологом будем понимать некоторое длительное рассуждение одного актора. Диалог является коммуникативным взаимодействием двух активно действующих индивидов. Виноградов С.И.[4] определяет полилог как «разговор нескольких участников общения». Различают следующие виды полилогов: собрания, совещания, переговоры, интервью, конференции. Исходя из того, что человек

большую часть времени проводит как член какой-либо группы, неизбежно взаимодействуя со всеми ее членами, наиболее распространенной формой разговорной речи является полилог. В связи с бурным развитием «виртуальной реальности» за последние 10 лет, особую актуальность приобрело изучение таких форм полилога как форум, чат, конференция и т.п. Часто исследователи проецируют проблемы полилога на диалог, определяя его как разговор более чем одного участника общения. Однако это не всегда верно, особенно при изучении социального аспекта общения, поскольку полилог обладает рядом специфических черт.

Полилог как форма коммуникативного взаимодействия

Ежедневно вступая в повседневное общение с членами социума, человек взаимодействует одновременно с несколькими партнерами. В условиях современного общества такая форма коммуникации (групповая коммуникация) является более распространенной, чем форма общения «тет-а-тет», поскольку большую часть своего времени индивид проводит в публичных местах (работа, транспорт, кафе, места отдыха и др.) принуждающих вступать во взаимодействие сразу с несколькими людьми.

Так, в любом полилоге нужно прилагать усилия для «захвата» коммуникативной инициативы. Здесь нет практически автоматического переключения инициативы типа «говорящий - слушающий», какое характерно для любого естественного диалога. Смена коммуникативных ролей осуществляется по линии наибольшей тематической ценности новой реплики или ее имитации. Полилог состоит из ряда мини-монологов, адресованных всем участникам. Каждая реплика должна быть четко оформлена, «правильно и вовремя начата и правильно и вовремя закончена»[5, с.86]. Реплики должны быть, во-первых, адресованы всем участникам полилога, а не одному избранному собеседнику; во-вторых, недопустимы длительные монологи, то есть полный захват лидерства одним из участников.

Вообще смена реплик коммуникантов в полилоге обладает сложным механизмом. В ней наблюдаются многочисленные наложения одновременных реплик разных коммуникантов [10]. Полилог обладает более сложной структурой взаимодействия и передачи информации по сравнению с диалогом, и потому не может быть сведен к нему. Полилогичность, как специфическая черта развертывания коммуникации, делает ее структуру «ветвистой», т.е. логика смысловых построений может развиваться в нескольких направлениях.

Отсутствие единой четко заданной сюжетной линии, ветвистость и наслоение информационных потоков, позволяет говорить о слабоструктурированности данных в полилогии. Особые методические трудности, в этой связи, приобретает изучение полилогов разворачивающихся в условиях «виртуальной реальности». К рассмотрению данного типа полилогов мы обратимся позже.

Из анализа коммуникативных цепочек на протяжении всего дискурса выявляются речевые стратегии. Как правило, стратегию определяет макроинтенция одного (или всех) участников полилога, обусловленная социальными и психологическими ситуациями [4]. Выбор той или иной стратегии, как правило, обусловлен поиском общего языка и выработкой основ диалогического сотрудничества. На выбор стратегии влияет и окружающая среда, задавая рамки стилистических норм.

Стратегия намечает общее развитие полилога, стилистика диктует сюжетные повороты развития и способы выражения интенций, полностью проявляясь лишь в заключительных репликах. В полилоге речевые стратегии реализуются актерами через речевые тактики, формирующие его отдельные части, модальные оттенки разговора (оценки, мнения, радость, удивление и т.д.). Так, например, стратегия отказа от выполнения просьбы может быть реализована через следующие тактики: а) выдать себя за некомпетентного человека (не способного к выполнению этой просьбы); б) сослаться на невозможность выполнения просьбы в данное время (занятость); в) иронии; г) отказа без мотивировки; д) уклониться от ответа, не обещать ничего неопределенного; е) дать ясно понять, что не желает выполнять просьбу.

Для установления контакта между говорящими используются речевые тактики особого рода, основанные на кооперативных стратегиях и служащие поддержанию заинтересованности собеседников, активизации и переключения внимания, захвата коммуникативного лидерства. При этом используются речевые приемы типа: «А я...», «А у нас...». Для укрепления коммуникативной инициативы используются тактика привлечения внимания и вовлечения в разговор многих собеседников, тактика эпатирования через отрицание привычных схем и культурных норм социума.

Важным коммуникативным приемом осуществления ряда тактик является молчание. Молчание может играть роль речевого акта утверждения, обещания, просьбы, согласия, ожидания, запинки, оценки, прекращения разговора. Функция молчания определяется контекстуально, исходя из конкретной речевой ситуации.

Векторы развития речевого общения задаются регулятивными репликами, имеющих строго упорядоченную структуру и дифференцирующихся в зависимости от социальных и психологических ролей говорящих. К регулятивам относятся вводные слова и предложения, междометия, вопросы, переспросы, слова-предложения да и нет, комментарии, оценочные суждения[4]. Такого рода коммуникативные ходы показывают правильность прогноза говорящего относительно уровня понимания адресата, выявляют тональность общения, намечают повороты в линии развития полилога.

Моделирование коммуникативной ситуации в условиях полилога

На формирование каждой конкретной коммуникативной социальной ситуации полилога оказывают влияние следующие условия: 1) место коммуникации (локальный параметр анализа); 2) временная продолжительность и ориентация коммуникации (временной параметр); 3) участники коммуникации, как непосредственные, так и окружающие, «наблюдатели» (социороловой параметр).

- 1) Локальный параметр. При построении и развитии полилога безусловно очень важно конкретное место и обстановка, в которой находятся актеры (улицы, кафе, поезд, аудитория), как фактор влияющий на актуализацию коммуникативной деятельности индивида. Все вербальные и невербальные коммуникативные средства могут быть четко детерминированы окружающей ситуацией. Поэтому, при изучении полилога вначале следует определить место его протекание, и проводить

дальнейшее изучение с учетом специфических особенностей его локального развития;

- 2) Временной параметр – длительность протекания полилога. Разные по длительности итерации, например, кратковременная беседа в лифте и длительная дискуссия в аудитории, являются по сути разными «формами» коммуникации и характеризуются различными «речевыми формулами» построения и протекания полилога. В зависимости от длительности используются специфические формы, выражающие приветствие, прощание, благодарность, извинение, согласие/несогласие. Чем более кратковременен полилог, тем больше вероятность использования невербальных средств;
- 3) Участники коммуникации влияют на характер социальной ситуации. Состав участников определяет распределение социальных и коммуникативных ролей, социальных статусов, межличностные отношения и ориентацию на окружающих. Эти признаки актуализируются в полилоге, задавая тональность протекания полилога, варианты вербальных и невербальных средств выражения. Состав участников отражается и в используемом словарном запасе, и в конструировании структуры полилога в целом.

Разрабатывая модель коммуникативного процесса, разворачивающегося в полилогии, следует учитывать не только дескриптивные и структурные компоненты, но и семиотические компоненты. В общем виде модель изучения коммуникативной ситуации может выглядеть следующим образом (см. рис.1).



Рисунок 1. Модель коммуникативной ситуации

Полилог как форма «виртуальной» коммуникации

С развитием современных технологий и увеличением уровня доступности интернетресурсов все больший процент общения переносится из реального мира в «виртуальный». Интернет-сообщества и различные социальные сети стали частью повседневного мира. В виду все возрастающих темпов жизни большая часть общения современного человека протекает в пространстве виртуальных сообществ. Наиболее распространенными формами взаимодействия становятся форумы, чаты, сообщества. Коммуникация здесь также имеет форму полилога (участвует более двух коммуникантов), однако, в отличие от взаимодействия в «реальном» мире, коммуникативная ситуация здесь будет обладать рядом специфических характеристик, оказывающих принципиальное влияние на процесс. Можно выделить следующие основные структурные компоненты речеактовых процессов, связанных как с индивидуальными стратегиями и целями индивидов, так и общей структурой протекания и развития процессов взаимодействия, системой сотрудничества и конфликтов акторов.

- 1) **Участники.** Социальные, демографические и личностные качества в форуме размыты и не оказывают сильного влияния на протекание процесса. С другой стороны, акторы принимающие активное участие в коммуникативном процессе, приобретают такие качества как активность (например, количество оставленных сообщений), характер сообщений (например, позитивный или негативный), направленность коммуникации (на донного партнера, на нескольких акторов, на самого себя и т.п.). Оставляемые сообщения не обязательно должны иметь некоторого конкретного адресата, хотя большинство сообщений, как правило, его сохраняют. Число активных коммуникантов, как правило, ограничивается количеством зарегистрированных пользователей. Хотя для просмотра и доступа к информации, она уже не требуется. Если говорить о ролевых характеристиках, то учитывая, что общение носит неформальный характер, очевидно что сохраняются только межличностные связи, хотя и они будут иметь очень слабо выраженный характер. Социально-ролевые связи между акторами, в большинстве случаев, не представлены.
- 2) **Ситуация.** В форуме общение протекает в пространстве виртуальной реальности, особенностью которой является дискретность временных координат. Участники дискуссии могут игнорировать некоторые высказывания, отвечать только на те реплики, которые у них вызывают интерес, реагировать на некоторые реплики несколько раз, отвечать на любые высказывания в ходе дискуссии, не зависимо от их последовательность и времени появления. Таким образом, ход дискуссии приобретает независимость от времени начала и может длиться практически неограниченное время. Важной особенностью является то, что число участников коммуникации не ограничено, так же как и число наблюдателей, присутствующих в виртуальном пространстве, гарантирующим анонимность и деперсонализацию личности. Таким образом, обстоятельства общения здесь обладают следующим набором характеристик: неформальность, публичность, неофициальность, безличность.
- 3) **Речевая форма.** В отличие от повседневного реального общения взаимодействие здесь протекает не в устной, а в письменной форме. В этой связи, при изучении полилога, следует обратить внимание на ряд

существенных особенностей. Тогда как устная речь является спонтанной, неподготовленной, а письменная строго продуманной и кодифицированной, в «виртуальной» реальности речь будет разворачиваться в письменной форме, но с элементами спонтанности и в виде фраз устной разговорной речи, а также ее грамматических особенностей. Устное разговорное общение протекает при личного контакта между коммуникантами. Здесь же общение хоть и сохраняет неформальный характер, всегда опосредованно компьютером.

- 4) **Поведение.** В рамках виртуально общение поведение носит неформальный характер. Однако, как правило, жесткие ограничения накладываются в правилах общения, декларируемых модераторами (например: нельзя отправлять пустые сообщения, употреблять нецензурную лексику, оскорблять других участников и т.п.). Поведение, выходящее за рамки допустимы границ, строго карается – сообщения удаляются, а участники могут быть исключены из группы.

Здесь всю совокупность пользователь можно разделить группы:

- 1) модераторы - задают тему, поддерживают к ней интерес, выступают ведущими, следят за «правовым» порядком взаимодействия;
- 2) активные участники – оставляют сообщения и комментарии по теме;
- 3) пассивные участники – выступают только в роли наблюдателей, читают, но пишут;
- 4) тролли – провоцируют негатив и негативные реакции остальных участников.

Коммуникация является инструментом, обеспечивающим взаимодействие между индивидами, позволяющими им объединяться в группы, передавать и сохранять информацию. Язык, являясь средством общения с другими, благодаря одинаковым символам позволяет поставить себя на место другого и сделать возможным коммуникативное понимание. Язык индивида позволяет сделать выводы не только о внутреннем мире индивида, но и об окружающей его среде. Речевая тактика представляет собой ряд коммуникативных действий, направленных на реализацию стратегии, в то время как тактика осуществляется через использование коммуникативных приемов более низкого порядка – коммуникативных ходов. Упорядочение целей преследуемых коммуникантами помогает обнаружить базисные операции, лежащие в основе речевых манипуляций, и через них выявить связь коммуникативных ходов и тактик.

Одним из наиболее часто используемых в повседневной жизни речевых жанров является полилог – общение между тремя и более партнерами. Полилог обладает специфическими чертами, требующими специального подхода. Как форма общения он поддерживает групповые связи акторов, способствуя консолидации группы как целого. Особое значение приобретает изучение полилогов в пространстве «виртуальной» реальности. Слабоструктурированные данные являются хорошим источником социальной информации, однако требуют методически правильно разработанного инструментария при построении модели коммуникации.

Источники:

1. Агеев В.Н. Семиотика. – М.: «Весь мир», 2002.
2. Бахтин М.М. Эстетика словесного творчества. – М., 1979.
3. Белл Р.Т. Социоллингвистика. Цели, методы, проблемы/ Пер. с англ. – М., Международные отношения, 1980. – 318 с.
4. Виноградов И.С. Культура русской речи. Учебник для вузов. Под ред. проф. Л. К. Граудиной и проф. Е. Н. Ширяева. - М.: Издательская группа НОРМА-ИНФРА М, 1999. - 560 с.
5. Деменьтьев В.В. Лингвистический аспект светскости / Вестник Омского университета, 1999, вып. 4. С. 85-88
6. Иссерс О.С. Коммуникативные стратегии и тактики русской речи. – М., УРСС, 2002. – 284 с.
7. Карасик В.И. Язык социального статуса. – М., 2002. – 333 с.
8. Конецкая В.П. Социология коммуникации - М., 1997
9. Макаров М. Основы теории дискурса. – М., 2003. – 280 с.
10. Полилог в деловом общении/ <http://posyo.narod.ru/polilogue.htm>
11. Русский язык и культура речи. Программа курса./ http://technical.bmstu.ru/metod_kab/programi/hf08.htm
12. Юссерел М. Социоллингвистика / Пер. с нем. – Киев: Вища шк., 1987. – 200 с.
13. Habermas J. What is universal pragmatics. – L., 1979.
14. Lasswell H.D. The Structure and Function of Communication in Society // Mass Communication. W. Schramm (Ed.). – Urbana, 1949. – p. 102 – 115.
15. Luckmann T. The Sociology of Language. — Indianapolis: Bobbs-Merrill, 1975. — 79 p.
16. Mead G. Geist, Identitat und Gesellschaft / Mit einer Einleitung herausgegeben von Ch. Morris. - Frankfurt am Main, 1973.
17. Schutz A., Luckmann T. Strukturen der Lebenswelt. Bd. I. - Neuwied, 1975



РАЗДЕЛ 7

СЕТЕВОЙ АНАЛИЗ КОММУНИКАТИВНО- ИНФОРМАЦИОННЫХ ПРОЦЕССОВ

On asymptotics for power divergence family of statistics

Vladimir V. Ulyanov

National research university
“Higher school of economics” (Moscow)

ABSTRACT. We review the results about asymptotic properties for distributions of statistics from power divergence family of statistics. This family includes such famous goodness-of-fit statistics as Karl Pearson chi-squared test, Freeman-Tukey statistic and log-likelihood ratio statistic. We consider recent results, where for the first time the order of approximation was obtained in the form $O(n^{-c})$ with some positive constant c .

All three gems in probability theory – the law of large numbers, the central limit theorem and the law of the iterated logarithm – concern the asymptotic behavior of the sums of random variables. It would be natural to extend the results to the functionals of the sums, in particular to quadratic forms. Moreover, in mathematical statistics there are numerous asymptotic problems which can be formulated in terms of quadratic or almost quadratic forms. In this article we review the corresponding results with rates of convergence. The review does not pretend to completely illuminate the present state of the area under consideration. It reflects mainly the author’s interests and results.

We consider accuracy of approximations for distributions of sums of independent random elements in $k - 1$ -dimensional Euclidian space. The approximation is considered on the class of sets which are “similar” to ellipsoids. Its appearance is motivated by study of asymptotic behavior of goodness-of-fit test statistics – power divergence family of statistics.

Consider a vector $(Y_1, \dots, Y_k)^T$ with multinomial distribution $M_k(n, \pi)$, i. e.

$$\Pr(Y_1 = n_1, \dots, Y_k = n_k) = \begin{cases} n! \prod_{j=1}^k (\pi_j^{n_j} / n_j!), & n_j = 0, 1, \dots, n \ (j = 1, \dots, k) \\ & \text{and } \sum_{j=1}^k n_j = n, \\ 0, & \text{otherwise,} \end{cases}$$

where $\pi = (\pi_1, \dots, \pi_k)^T$, $\pi_j > 0$, $\sum_{j=1}^k \pi_j = 1$. From this point on, we will assume the validity of the hypothesis $H_0: \pi = \mathbf{p}$. Since the sum of n_i equals n , we can express this multinomial distribution in terms of a vector $\mathbf{Y} = (Y_1, \dots, Y_{k-1})$ and define its covariance matrix Ω . It is known that so defined Ω equals

The author is partly supported by Higher School of Economics grant, No. 12-05-0052.

2000 *Mathematics Subject Classification.* Primary 62E20, 62H10; Secondary 52A20.

Key words and phrases. Accuracy of approximations, goodness-of-fit statistics, power divergence family of statistics.

$(\delta_i^j p_i - p_i p_j) \in \mathbf{R}^{(k-1) \times (k-1)}$. The main object of the current study is the power divergence family of goodness-of-fit test statistics:

$$t_\lambda(\mathbf{Y}) = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^k Y_j \left[\left(\frac{Y_j}{np_j} \right)^\lambda - 1 \right], \quad \lambda \in \mathbf{R},$$

When $\lambda = 0, -1$, this notation should be understood as a result of passage to the limit.

These statistics were first introduced in Cressie and Read (1984) and Read (1984). Putting $\lambda = 1, \lambda = -1/2$ and $\lambda = 0$ we can obtain the chi-squared statistic, the Freeman-Tukey statistic, and the log-likelihood ratio statistic respectively.

We consider transformation

$$X_j = (Y_j - np_j)/\sqrt{n}, \quad j = 1, \dots, k, \quad r = k - 1, \quad \mathbf{X} = (X_1, \dots, X_r)^T.$$

Herein the vector \mathbf{X} is the vector taking values on the lattice,

$$L = \left\{ \mathbf{x} = (x_1, \dots, x_r)^T; \mathbf{x} = \frac{\mathbf{m} - n\mathbf{p}}{\sqrt{n}}, \mathbf{p} = (p_1, \dots, p_r)^T, \mathbf{m} = (n_1, \dots, n_r)^T \right\},$$

where n_j are non-negative integers.

The statistic $t_\lambda(\mathbf{Y})$ can be expressed as a function of \mathbf{X} in the form

$$T_\lambda(\mathbf{X}) = \frac{2n}{\lambda(\lambda+1)} \left[\sum_{j=1}^k p_j \left(\left(1 + \frac{X_j}{\sqrt{np_j}} \right)^{\lambda+1} - 1 \right) \right], \quad (1)$$

and then, via the Taylor's expansion, transformed to the form

$$T_\lambda(\mathbf{X}) = \sum_{i=1}^k \left(\frac{X_i^2}{p_i} + \frac{(\lambda-1)X_i^3}{3\sqrt{np_i^2}} + \frac{(\lambda-1)(\lambda-2)X_i^4}{12p_i^3 n} + O(n^{-3/2}) \right).$$

As we see the statistics $T_\lambda(\mathbf{X})$ is "close" to quadratic form

$$T_1(\mathbf{X}) = \sum_{i=1}^k \frac{X_i^2}{p_i},$$

considered in Section 1.

We call a set $B \subset \mathbf{R}^r$ *extended convex set*, if for for all $l = \overline{1, r}$ it can be expressed in the form:

$$B = \{ \mathbf{x} = (x_1, \dots, x_r)^T : \lambda_l(x^*) < x_l < \theta_l(x^*) \text{ and } x^* = (x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_r)^T \in B_l \},$$

where B_l is some subset of \mathbf{R}^{r-1} and $\lambda_l(x^*), \theta_l(x^*)$ are continuous functions on \mathbf{R}^{r-1} . Additionally, we introduce the following notation

$$\begin{aligned} [h(\mathbf{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)} &= h(x_1, \dots, x_{l-1}, \theta_l(x^*), x_{l+1}, \dots, x_r) \\ &\quad - h(x_1, \dots, x_{l-1}, \lambda_l(x^*), x_{l+1}, \dots, x_r). \end{aligned}$$

It is a known fact that the distributions of all statistics in the family converge to chi-squared distribution with $k - 1$ degrees of freedom (see e.g. Cressie and Read (1984), p. 443). However, more intriguing is the problem to find the rate of convergence to the limiting distribution.

For any bounded extended convex set B in Yarnold (1972) it was obtained an asymptotic expansion, which in Siotani and Fujikoshi (1984) was converted to

$$\Pr(\mathbf{X} \in B) = J_1 + J_2 + O(n^{-1}), \quad (2)$$

with

$$\begin{aligned} J_1 &= \int \cdots \int_B \phi(\mathbf{x}) \left\{ 1 + \frac{1}{\sqrt{n}} h_1(\mathbf{x}) + \frac{1}{n} h_2(\mathbf{x}) \right\} dx, \text{ where} \\ h_1(\mathbf{x}) &= -\frac{1}{2} \sum_{j=1}^k \frac{x_j}{p_j} + \frac{1}{6} \sum_{j=1}^k x_j \left(\frac{x_j}{p_j} \right)^2, \\ h_2(\mathbf{x}) &= \frac{1}{2} h_1(\mathbf{x})^2 + \frac{1}{12} \left(1 - \sum_{j=1}^k \frac{1}{p_j} \right) + \frac{1}{4} \sum_{j=1}^k \left(\frac{x_j}{p_j} \right)^2 - \frac{1}{12} \sum_{j=1}^k x_j \left(\frac{x_j}{p_j} \right)^3; \\ J_2 &= -\frac{1}{\sqrt{n}} \sum_{l=1}^r n^{-(r-l)/2} \sum_{x_{l+1} \in L_{l+1}} \cdots \sum_{x_r \in L_r} \\ &\quad \left[\int \cdots \int_{B_l} [S_1(\sqrt{n}x_l + np_l)\phi(\mathbf{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)} dx_1, \cdots, dx_{l-1} \right]; \quad (3) \end{aligned}$$

$$L_j = \left\{ \mathbf{x}; x_j = \frac{n_j - np_j}{\sqrt{n}}, n_j \text{ and } p_j \text{ defined as before} \right\};$$

$$S_1(x) = x - [x] - 1/2, [x] \text{ is the integer part of } x;$$

$$\phi(\mathbf{x}) = \frac{1}{(2\pi)^{r/2} |\Omega|^{1/2}} \exp \left(-\frac{1}{2} \mathbf{x}^T \Omega^{-1} \mathbf{x} \right).$$

In Yarnold (1972) it was showed that $J_2 = O(n^{-1/2})$.

Using elementary transformations it can be easily shown that the determinant of the matrix Ω equals $\prod_{i=1}^k p_i$.

In Yarnold (1972) it was also examined that expansion for the most known power divergence statistic, which is the chi-squared statistic. Put $B^\lambda = \{\mathbf{x} \mid T_\lambda(\mathbf{x}) < c\}$. It is easy to show that B^1 is an ellipsoid, which is a particular case of a bounded extended convex set. J. Yarnold managed to simplify the item (3) in this simple case and converted the expansion (2) to

$$\begin{aligned} \Pr(\mathbf{X} \in B^1) &= G_r(c) + (N^1 - n^{r/2} V^1) e^{-c/2} / \left((2\pi n)^r \prod_{j=1}^k p_j \right)^{1/2} \\ &\quad + O(n^{-1}), \quad (4) \end{aligned}$$

where $G_r(c)$ is the chi-squared distribution function with r degrees of freedom; N^1 is the number of points of the lattice L in B^1 ; V^1 is the volume of B^1 . Using the result from Esseen (1945), he obtained an estimate of the second item in (4) in the form $O(n^{-(k-1)/k})$. If we estimate second term in (4) taking the result from Götze (2004) instead of Esseen's one from Esseen (1945) we get (see Götze and Ulyanov (2003)) in the case of Karl Pearson chi-squared statistics, i.e. when $\lambda = 1$, that for $r \geq 5$

$$\Pr(\mathbf{X} \in B^1) = G_r(c) + O(n^{-1}).$$

In Shiotani and Fujikoshi (1984) it was showed that, when $\lambda = 0, \lambda = -1/2$, we have

$$\begin{aligned} J_1 &= G_r(c) + O(n^{-1}) \\ J_2 &= (N^\lambda - n^{r/2}V^\lambda) e^{-c/2} / \left((2\pi n)^r \prod_{j=1}^k p_j \right)^{1/2} + o(1), \\ V^\lambda &= V^1 + O(n^{-1}). \end{aligned} \quad (5)$$

These results were expanded by T. Read to the case $\lambda \in \mathbf{R}$. In particular Theorem 3.1 in Read (1984) implies

$$\Pr(T_\lambda < c) = \Pr(\chi_r^2 < c) + J_2 + O(n^{-1}). \quad (6)$$

This reduces the problem to the estimation of the order of J_2 .

It is worth mentioning that in Siotani and Fujikoshi (1984) and in Read (1984) there is no estimate for the residual in (5). Consequently, it is impossible to construct estimates of the rate of convergence of statistics T_λ to the limiting distribution, grounded on the simple representation for J_2 initially suggested by J. Yarnold.

In Ulyanov and Zubov (2009) and in Asylbekov, Zubov and Ulyanov (2011) the rate of convergence in (20) was obtained for any power divergence statistic. Then we construct an estimate for J_2 based on the fundamental number theory results of Hlawka (1950) and Huxley (1993) about approximation of number of integer points in convex sets (more general than ellipsoids) by Lebesgue measure of the set.

Therefore, one of the main point is to investigate the applicability of the afore-mentioned theorems from number theory to the set B^λ .

In Ulyanov and Zubov (2009) it is shown that $B^\lambda = \{\mathbf{x} \mid T_\lambda(\mathbf{x}) < c\}$ is a bounded extended-convex (strictly convex) set. As it has been already mentioned, in accordance with the results of Yarnold (1972)

$$J_2 = O(n^{-1/2}).$$

For the specific case of $r = 2$ this estimate has been considerably refined in Asylbekov, Zubov and Ulyanov (2011):

$$J_2 = O(n^{-50/73}(\log n)^{315/146}), \quad r = 2.$$

In Asylbekov, Zubov and Ulyanov (2011) it was used the following theorem from Huxley (1993):

THEOREM 1. *Let D be a two-dimensional convex set with area A , bounded by a simple closed curve C , divided into a finite number of pieces each of those being 3 times continuously differentiable in the following sense. Namely, on each piece C_i the radius of curvature ρ is positive (and not infinite), continuous, and continuously differentiable with respect to the angle of contingence ψ . Then in a set that is obtained from D by translation and linear expansion of order M , the number of integer points equals*

$$N = AM^2 + O(IM^K(\log M)^\Lambda)$$

$$K = \frac{46}{73}, \quad \Lambda = \frac{315}{146},$$

where I is a number depending only on the properties of the curve C , but not on the parameters M or A .

In Ulyanov and Zubov (2009) the results from Asylbekov, Zubov and Ulyanov (2011) were generalized to any dimension. The main reason why two cases when $r = 2$ and $r \geq 3$ are considered separately consists in the fact that for $r \geq 3$ it is much more difficult than for $r = 2$ to check applicability the number theory results to B^λ . In Ulyanov and Zubov (2009) we used the following result from Hlawka (1950)

THEOREM 2. *Let D be a compact convex set in \mathbf{R}^m with the origin as its inner point. We denote the volume of this set by A . Assume that the boundary of this set is an $(m - 1)$ -dimensional surface of class \mathbf{C}^∞ , the Gaussian curvature being non-zero and finite everywhere on the surface. Also assume that a specially defined "canonical" map from the unit sphere to D is one-to-one and belongs to the class \mathbf{C}^∞ . Then in the set that is obtained from the initial one by translation along an arbitrary vector and by linear expansion with the factor M the number of integer points is*

$$N = AM^m + O\left(IM^{m-2+\frac{2}{m+1}}\right)$$

where the constant I is a number dependent only on the properties of the curve C , but not on the parameters M or A .

Providing that $m = 2$, the statement of theorem 2 is weaker than the result of Huxley.

The above theorem is applicable in Ulyanov and Zubov (2009) with $M = \sqrt{n}$. Therefore, for any fixed λ we have to deal not with a single set, but rather with a sequence of sets $B^\lambda(n)$ converging in some sense to the limiting set B^1 when $n \rightarrow \infty$. It is necessary to emphasize that the constant I in our case, generally speaking, is $I(n)$, i.e. it depends on n . Only having ascertained the fulfillment of the inequality

$$|I(n)| \leq C_0,$$

where C_0 is an absolute constant, we are able to apply Theorem 2 without a change of the overall order of the error with respect to n .

In Ulyanov and Zubov (2009) we prove the following estimate of J_2 in the space of any fixed dimension $r \geq 3$.

THEOREM 3. *For the term J_2 from decomposition (6) the following estimate holds*

$$J_2 = O\left(n^{-r/(r+1)}\right), \quad r \geq 3,$$

The Theorem implies that for the statistic $T_\lambda(\mathbf{X})$ defined by formula (1) it holds that

$$\Pr(T_\lambda(\mathbf{X}) < c) = G_r(c) + O\left(n^{-1+\frac{1}{r+1}}\right), \quad r \geq 3.$$

References

- [1] Zh. A. Asylbekov, V. N. Zubov and V. V. Ulyanov, On approximating some statistics of goodness-of-fit tests in the case of three-dimensional discrete data. *Siberian Mathematical Journal*. 2011, Volume 52, Number 4, Pages 571-584.
- [2] Bentkus, V.: On dependence of Berry–Esseen bounds on dimensionality. *Lithuanian Math. J.* **26**, 205–210 (1986)
- [3] Bentkus, V., Götze, F.: Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. *Probab. Theory Relat. Fields* **109**, 367–416 (1997)
- [4] Bentkus, V.; Götze, F. Optimal bounds in non-Gaussian limit theorems for UU -statistics. *Ann. Probab.* 27 (1999), no. 1, 454521.
- [5] Berry, A.C.: The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49**, 122-136 (1941).
- [6] Bhattacharya, R.N. and Ranga Rao, R. Normal approximation and asymptotic expansions. Robert E. Krieger Publishing Co., Inc., Melbourne, FL, 1986. xiv+291 pp. ISBN: 0-89874-690-6.
- [7] Bogatyrev, S.A., Götze, F., Ulyanov, V.V.: Non-uniform bounds for short asymptotic expansions in the CLT for balls in a Hilbert space. *J. Multivariate Anal.* **97**, 9, 2041–2056 (2006)
- [8] N. A. C. Cressie, T. R. C. Read, Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society, Series B*, **46** (1984), 440–464.
- [9] Esseen, C.G.: On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.* **28A**, no. 9, 19 pp. (1942).
- [10] Esseen, C.G.: Fourier analysis of distribution functions. *Acta Math.* **77**, 1–125 (1945)
- [11] Esseen, C.G.: A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.* **39**, 160-170 (1956).
- [12] Götze, F.: Asymptotic expansion for bivariate von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **50**, 333–355 (1979)
- [13] Götze, F.: Expansions for von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **65**, 599–625 (1984)
- [14] Götze, F.; Prokhorov, Yu. V.; Ulyanov, V. V. Estimates for the characteristic functions of polynomials of asymptotically normal random variables. (Russian) *Uspekhi Mat. Nauk* 51 (1996), no. 2(308), 3–26; translation in *Russian Math. Surveys* 51 (1996), no. 2, 181204
- [15] Götze, F.; Prokhorov, Yu. V.; Ulyanov, V. V. On the smooth behavior of probability distributions under polynomial mappings. (Russian) *Teor. Veroyatnost. i Primenen.* 42 (1997), no. 1, 51–62; translation in *Theory Probab. Appl.* 42 (1997), no. 1, 2838 (1998).

- [16] Götze, F.: Lattice point problems and the central limit theorem in Euclidean spaces. Doc. Math. J.DMV, Extra Vol. ICM, III, 245–255 (1998)
- [17] Götze, F. and Margulis, G.A.: Distribution of values of quadratic forms at integral points, Preprint <http://arxiv.org/abs/1004.5123>, 2010.
- [18] Götze, F., Ulyanov, V.V.: Uniform approximations in the CLT for balls in Euclidian spaces, Preprint 00-034 SFB 343, Univ.Bielefeld (2000).
- [19] Götze, F., Ulyanov, V.V.: Asymptotic distribution of χ^2 -type statistics, Preprint 03-033, Research group "Spectral analysis, asymptotic distributions and stochastic dynamics", 2003.
- [20] F. Götze, Lattice point problems and values of quadratic forms, *Inventiones mathematicae*, **157** 2004, 195–226.
- [21] Götze, F., Zaitsev, A. Yu.: Uniform rates of convergence in the CLT for quadratic forms. Preprint 08119. SFB 701, Univ.Bielefeld (2008).
- [22] Götze, F., Zaitsev, A. Yu.: Explicit rates of approximation in the CLT for quadratic forms. <http://arxiv.org/pdf/1104.0519.pdf> (2011).
- [23] E. Hlawka Über integrale auf konvexen körpern I. *Mh. Math* **54** (1950), 1–36.
- [24] M. N. Huxley, Exponential sums and lattice points II. *Proceedings of London Mathematical Society*, **66** (1993), 279–301.
- [25] Kandelaki, N.P.: On limit theorem in Hilbert space. *Trudy Vychisl. Centra Akad. Nauk Gruzin. SSR* **11**, 46–55 (1965)
- [26] Nagaev, S.V.: On new approach to study of distribution of a norm of a random element in a Hilbert space. Fifth Vilnius conference on probability theory and mathematical statistics. *Abstracts* **4**, 77–78 (1989)
- [27] Nagaev, S.V., Chebotarev, V.I.: A refinement of the error estimate of the normal approximation in a Hilbert space. *Siberian Math. J.* **27**, 434–450 (1986)
- [28] Nagaev, S.V., Chebotarev, V.I.: On the accuracy of Gaussian approximation in Hilbert space. *Acta Applicandae Mathematicae* **58**, 189–215 (1999)
- [29] T. R. C. Read, Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics., *The Annals of Mathematical Statistics, Part A*, **36** (1984), 59–69.
- [30] Sazonov, V.V.: On the multi-dimensional central limit theorem. *Sankhya Ser. A* **30**, no.2, 181–204 (1968)
- [31] Sazonov, V.V.: Normal approximation – some recent advances. *Lecture Notes in Mathematics*, 879, Springer-Verlag, Berlin, NY (1981)
- [32] Sazonov, V.V., Ulyanov, V.V., Zalesskii, B.A.: Normal approximation in a Hilbert space. I, II. *Theory Probab. Appl.* **33**, 207–227, 473–483 (1988a)
- [33] Sazonov, V.V., Ulyanov, V.V., Zalesskii, B.A.: A sharp estimate for the accuracy of the normal approximation in a Hilbert space. *Theory Probab. Appl.* **33**, 700–701 (1988b)
- [34] Sazonov, V.V., Ulyanov, V.V., Zalesskii, B.A.: A precise estimate of the rate of coverage in the CLT in Hilbert space. *Mat.USSR Sbornik* **68**, 453–482 (1991)
- [35] Senatov, V.V.: Four examples of lower bounds in the multidimensional central limit theorem. *Theory Probab.Appl.* **30**, 797–805 (1985)
- [36] Senatov, V.V.: On rate of convergence in the central limit theorem in a Hilbert space. Fifth Vilnius conference on probability theory and mathematical statistics. *Abstracts.* **4**, 222 (1989)
- [37] Senatov, V.V.: Qualitive effects in the estimates of convergence rate in the central limit theorem in multidimensional spaces. *Proceedings of the Steklov Institute of Mathematics. Vol.215*, Moscow, Nauka (1996)
- [38] Shevtsova, I.G.: On the absolute constants in the BerryEsseen type inequalities for identically distributed summands, Preprint <http://arxiv.org/pdf/1111.6554.pdf>, 2011.
- [39] M. Siotani, Y. Fujikoshi, Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics, *Hiroshima Mathematical Journal*, **14** (1984), 115–124.

- [40] Tyurin, I.S.: Sharpening the upper bounds for constants in Lyapunov's theorem. (Russian) *Uspekhi Mat. Nauk* 65 (2010), no. 3(393), 201–201; translation in *Russian Math. Surveys* 65 (2010), no. 3, 586–588.
- [41] Ulyanov, V.V.: Normal approximation for sums of nonidentically distributed random variables in Hilbert spaces. *Acta Sci. Math. (Szeged)* **50**, no. 3-4, 411-419, (1986).
- [42] Ulyanov, Vladimir V.; Zubov, Vasily N. Refinement on the convergence of one family of goodness-of-fit statistics to chi-squared distribution. *Hiroshima Math. J.* 39 (2009), no. 1, 133161.
- [43] Ulyanov, V.V., Götze, F.: Short asymptotic expansions in the CLT in Euclidian spaces: a sharp estimate for its accuracy. *Proceedings 2011 World Congress on Engineering and Technology*. Oct.28-Nov.2, 2011. Shanghai, China, IEEE Press, **1**, 260–262, 2011.
- [44] J. K. Yarnold, Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. *The Annals of Mathematical Statistics*, **43** (1972), 1566–1580.
- [45] Yurinskii, V.V.: On the accuracy of normal approximation of the probability of hitting a ball. *Theory Probab. Appl.* **27**, 280–289 (1982)



ДИЗАЙН, ВЕРСТКА
К. П. ЛАЗЕБНАЯ

