

АКАДЕМИЯ ПРАВОВЫХ НАУК УКРАИНЫ
Научно-исследовательский институт правовой информатики

ФУРАШЕВ В.Н., ЛАНДЭ Д.В., БРАЙЧЕВСКИЙ С.М.

**Моделирование информационно-электоральных
процессов**

Киев – 2007

УДК 681.3+519.8
ББК 22.18, 32.81, 60.54
Ф95

*Рекомендовано к изданию
Ученым советом Научно-исследовательского центра правовой
информатики Академии правовых наук Украины
(протокол № 9 от 24 октября 2007 года)*

Рецензенты:

Заславский В.А. - доктор технических наук, профессор

Даниляк С.Н. - доктор технических наук

Тихий В.П. - доктор юридических наук, профессор, Заслуженный юрист Украины, академик Академии правовых наук Украины

Швец Н.Я. - доктор экономических наук, профессор, Заслуженный деятель науки и техники Украины, член-корреспондент Академии правовых наук Украины

Ф95 **Фурашев В.Н., Ландэ Д.В., Брайчевский С.М.**

Моделирование информационно-электоральных процессов:
Монография. - К.: НИЦПИ АпрН Украины, 2007. – 182 стр.

ISBN 978-966-96927-2-6

В монографии представлены результаты исследований информационных аспектов, присущих электоральным процессам в современном обществе. Главное внимание уделяется вопросам математического моделирования: рассмотрены различные нелинейные и индивидуум-ориентированные модели. Подробно представлены вопросы динамики информационных потоков. Большое внимание уделено таким методам, как корреляционный и фрактальный анализ, формализм клеточных автоматов.

Книга предназначена как для широкого круга специалистов - аналитиков в областях подготовки и проведения выборов, правовых наук и прикладной социологии, так и для студентов старших курсов, аспирантов.

УДК 681.3+519.8
ББК 22.18, 32.81, 60.54

ISBN 978-966-96927-2-6

© **Фурашев В.Н., 2007**

© **Ландэ Д.В., 2007**

© **Брайчевский С.М., 2007**

ОГЛАВЛЕНИЕ

1. Проблемы моделирования информационно-электоральных процессов	4
2. Моделирование электоральных процессов	11
2.1. Нелинейные конкурентные модели	15
2.2. Индивидуум-ориентированные модели	48
2.2.1. Модель, основанная на влиянии окружения	58
2.2.2. Модель, учитывающая концепцию «малых миров».....	64
2.2.3. Модель, учитывающая влияние внешних факторов	78
2.3. Другие подходы	81
2.3.1. Метод анализа иерархий.....	81
2.3.2. Теоретико-игровой подход.....	89
3. Моделирование информационных потоков	96
3.1. Тематические информационные потоки	100
3.2. Корреляционный анализ информационных потоков.....	102
3.3. Моделирование динамики информационных потоков.....	111
3.3.1. Линейная модель	116
3.3.2. Экспоненциальная модель	118
3.3.3. Логистическая модель	119
3.4. Модели диффузии информации.....	129
3.5. Фрактальные свойства информационных потоков	140
3.6. Анализ стабильности информационных источников	159
3.7. Объектно-статистический анализ информационных потоков.....	167
Заключение	174
Литература	175

1. ПРОБЛЕМЫ МОДЕЛИРОВАНИЯ ИНФОРМАЦИОННО-ЭЛЕКТОРАЛЬНЫХ ПРОЦЕССОВ

Избирательные и референдумные процессы играют, несомненно, исключительно важную роль в динамике развития общественных систем с демократической формой правления. Для широких слоев населения они примечательны в первую очередь своей прозрачностью. Мы все так или иначе участвуем в выборах и, что еще важнее, не остаемся равнодушными не только к конечному результату, но и к самому протеканию избирательной кампании. Даже те, кто внешне демонстрирует политическую пассивность, на самом деле обычно реагируют на поведение действующих лиц и возлагают на исход выборов определенные надежды.

Большинство социальных явлений известны и понятны преимущественно социологам, демографам, политологам и т. д. Обычный же человек либо не замечает этих явлений, либо не пытается разобраться в их природе и влиянии на другие сферы жизни. Но выборы чаще всего воспринимаются в полной мере, во всяком случае, большинство полагает, что вполне разбирается в них. И такая точка зрения имеет под собой известные основания, поскольку в основе избирательного процесса лежит простая и доступная общему пониманию мотивация участников.

Тем не менее на самом деле очень редко удастся предсказать исход выборов с точностью, достаточной для оценки реального расклада политических сил, определяющего последующий ход развития событий. Причина состоит в том, что электоральные процессы отнюдь не исчерпываются своей «очевидной» компонентой. Более того, не будет ошибкой сказать, что они относятся к числу наиболее сложных и запутанных общественных явлений. И, к сожалению, приходится констатировать, что сегодня мы не располагаем общепризнанной методикой их анализа. А значит,

изучение этих явлений является не только интересным, но и в высшей степени актуальным.

Что бы ни говорили поклонники «чистого разума», любая реальная система знаний, ориентированная на изучение процессов в окружающем нас мире (как физических, так и общественных), всегда так или иначе стремится к точным выводам, в идеале допускающим количественные оценки, которые можно сравнивать с эмпирическими данными. «Гуманитарные», как их часто называют, методы неизбежно рано или поздно упираются в извечный вопрос о том, похожи ли очки на велосипед. Каждый волен отвечать на него так, как ему вздумается, и экспертному сообществу не остается ничего другого, как драться табуретками (что мы регулярно и наблюдаем, если и не в прямом, то в переносном смысле).

Особенно это касается прогнозирования. Действительно, задним числом при желании можно объяснить все, что угодно. Прогноз же должен содержать вполне конкретные положения, явно коррелирующие с теми событиями, которые мы будем наблюдать на практике уже после того, как он составлен. От предсказаний «возможного некоторого роста тенденции» или еще чего-нибудь подобного, мало пользы. Мы хотим, чтобы в прогнозе говорилось, например, что рейтинг такой-то партии в течение месяца повысится на 3% при возможной погрешности в 0.25%.

Однако на практике программы, предполагающие количественное прогнозирование, неизбежно сталкиваются с серьезными, часто непреодолимыми трудностями. Это касается любой сферы научной (в широком смысле этого слова) мысли, в том числе, например, физики: отнюдь не просто описать поведение многоатомного газа с учетом взаимодействия между его молекулами. Но в области общественных наук, как мы хорошо знаем, ситуация оказывается неизмеримо более сложной. Процессы, протекающие в человеческом обществе, с большим трудом поддаются количественному анализу, в результате чего мы фактически не имеем более или менее надежных законов, описывающих социальные системы. Разумеется,

в какой-то мере к социумам применимы законы статистики, но при этом достоверными, как правило, оказываются лишь общие характеристики на демографическом уровне. Главная проблема, по-видимому, заключается в том, что сам человек обладает настолько сложным поведением, что в общественных процессах крайне слабо выражена повторяемость определенных ситуаций, на основе изучения которой можно было бы установить некие закономерности. Ведь естественнонаучная методология строится именно на выявлении устойчивых повторяемостей явлений, отмечающихся при определенных условиях.

Одним из наиболее эффективных путей преодоления подобных сложностей является применение математического моделирования интересующих нас явлений. Центральная идея заключается в том, чтобы заменить реальную сложную систему, механизмы которой нам известны не в полной мере, воображаемой простой, механизмы которой мы в явном виде закладываем в ее основу. Конечно, заранее не известно, будет ли модель функционировать так, как ожидается, но именно в достижении этой цели и заключается искусство моделирования.

В любом случае, моделирование предполагает постоянное соотнесение теоретических построений с наблюдаемой практикой. И если удастся построить и отладить приемлемую модель, результаты с лихвой оправдывают затраченные усилия.

В настоящее время математическое моделирование широко применяется в естественных науках, однако его применение для решения проблем в социальных науках остается ограниченным. Причина этого заключается, прежде всего, в сложности формализации основных понятий социологии, и в частности, в теории и практике избирательного процесса. Глобальным социальным явлениям присущи многоуровневость, разномасштабность, многопараметричность. При этом параметры зачастую на практике не поддаются формализации. Следует заметить, что, существенные трудности

при моделировании вызывает также необходимость учета социально-психологических факторов.

Попытки детального учета параметров такого типа настолько усложняют модели, что они редко оказываются успешными. Вместе с тем известно, что очень часто небольшие изменения параметров могут привести к настолько значительным изменениям выходных значений, что полностью дискредитируется вся модель.

Таким образом, сегодня при моделировании социальных явлений, каковыми являются электоральные процессы, наибольший интерес представляют модели, не претендующие на детальное описание особенностей каждого конкретного случая, а позволяющие обобщать и одновременно учитывать некоторую социальную конкретику. Конечно, многие процессы, близкие к электоральным, можно моделировать достаточно точно, если четко их параметризовать и установить граничные параметры. К таким процессам можно отнести, например, информационные потоки электронных СМИ (в частности, в Интернете), сопутствующих выборным процессам [6, 24]. Однако, по-видимому, на данном этапе в области моделирования более сложных социальных процессов успех может быть достигнут только путем синтеза достаточно простых алгоритмов и концепций.

Так или иначе, сегодня достигнуты определенные успехи в моделировании социальных процессов, которые базируются на таких уже традиционных в этой области методах, как теория нелинейных дифференциальных уравнений, теория игр и математическая статистика.

В частности, организационно-технические и нормативно-правовые аспекты избирательных (референдумных) процессов, происходящие в Украине, довольно полно рассмотрены в научном плане в работах [48-50, 52].

В работах [28-29, 51] основное внимание уделено информационным аспектам избирательных (референдумных) процессов, которые подталкивают к изучению возможности моделирования электоральных процессов на основе методов математической статистики, теории игр и дискретной математики.

Следует признать перспективными в этой области и теорию клеточных автоматов, впервые предложенную более тридцати лет назад Дж. фон Нейманом [32].

В соответствии с этим подходом, по аналогии с биологическими системами, динамика образующих электорат социальных групп описывается в терминах популяций. С этой целью введем понятие электоральной популяции, связанное с широко используемым понятием электорального поля. Здесь уместно прибегнуть к другой аналогии – с физическими системами. Физические поля порождаются зарядами и, в свою очередь, оказывают на них воздействие. В этом смысле электоральные популяции играют роль зарядов: они порождают электоральное поле и взаимодействуют с ним.

Электоральной популяцией (ЭП) будем называть относительно устойчивую социальную группу, системообразующим признаком которой является приверженность определенной политической силе и, соответственно, готовность голосовать за нее на выборах. Так же, как и в биологии, электоральная популяция представляет собой низшую форму социальной организации. Подчеркнем: популяцию образуют не активные политики, а рядовые члены партий и беспартийные избиратели, определившиеся со своими симпатиями. Симпатии, естественно, могут меняться во времени, что отражается на динамике ЭП. Главное, что в каждый момент времени член каждой из них может однозначно ответить на вопрос о том, какую политическую силу он поддерживает.

Более сложной является проблема природы взаимодействия ЭП. В отличие от биологических видов, здесь одна и та же особь может переходить из одной популяции в другую и обратно. Кроме того, скорость электоральных процессов очень высока, так что смена поколений роли не играет. В конечном счете, речь может идти о перераспределении электоральных ресурсов, понимаемое в том или ином смысле, в зависимости от поставленной задачи. Это, в свою очередь, может означать не только

переход избирателей от одной политической силы к другой, но и перетекание идей, лозунгов, политтехнологий. Очевидно, что реальная динамика ЭП в любом случае не может быть сведена к какому-либо одному набору однотипных механизмов. Мы можем выделить, по крайней мере, две такие группы.

Первая предполагает непосредственное изменение численности данной ЭП. В этом случае она расширяется за счет «захвата на свою орбиту» тех, кто еще не определился, а также тех, кто решил сменить свои симпатии. Это не обязательно может быть сознательная агитация, возможны различные причины. Например, молодежь часто руководствуется позицией взрослых, которые пользуются авторитетом, и с которыми она может быть даже не знакома лично (выдающиеся спортсмены, артисты и т. д.). Такие процессы ограничены только количественными характеристиками данной общественной системы.

Ко второй группе относятся механизмы изменения политического веса популяций, возможно, при неизменности их физической численности. Понятие политического веса (веса политической силы) заслуживает отдельных комментариев.

Отметим, прежде всего, что численность политических популяций как таковая не является четко определенной величиной. В биологических системах видовая принадлежность жестко фиксирована, и всегда есть возможность, пусть даже теоретическая, в точности пересчитать всех щук и карасей. В нашем случае это не так. Действительно, изменения (в том числе и колебания) политических симпатий, особенно непосредственно перед выборами, могут достигать таких масштабов, что говорить об устойчивых параметрах политических сил не приходится. Ситуация усложняется тем, что переходы между политическими силами часто носят транзитный характер (например, через промежуточное состояние «не поддерживаю никого»).

С другой стороны, рейтинги, основанные на социологических опросах, не дают точного представления о реальной численности той или иной

популяции уже потому, что число опрошенных респондентов намного меньше общего числа социально активных граждан. Они скорее показывают относительную меру влияния каждой политической силы на общество в целом. Именно эту меру и выражает политический вес соответствующей силы. Предполагается также, что эти данные свидетельствуют и о численности популяций, однако насколько это соответствует действительности, заранее не известно. Разумеется, определенная связь существует (популяция, обладающая большим политическим весом, скорее всего, будет многочисленной), но на значительных промежутках времени она часто оказывается опосредованной. Явно она проявляется непосредственно перед голосованием, когда взаимные переходы дают пренебрежимо малый вклад. Действительно, вес политической силы не может считаться значительным, если она на выборах набрала мало голосов. Однако политическая сила, стартовавшая при относительно слабой поддержке, вполне может быстро «набрать обороты» и получить заметное представительство. Но это означает, что изначально она обладала заметным политическим весом. Возникает естественный вопрос: как оценить вес политической силы на ранних стадиях избирательной кампании? Помимо теоретических прогнозов, основанных на качестве предвыборной программы и методов работы с электоратом, наиболее эффективным представляется анализ изменений ее поддержки избирателями, то есть динамики соответствующей ЭП.

Таким образом, политический вес силы является не менее важной характеристикой, чем физическая численность ее сторонников. Следовательно, изучение его динамики представляет не меньший интерес. Поэтому, говоря о росте ЭП, будем иметь в виду оба фактора.

2. МОДЕЛИРОВАНИЕ ЭЛЕКТОРАЛЬНЫХ ПРОЦЕССОВ

Электоральные процессы выделяются из общего массива социальных процессов по крайней мере двумя важными особенностями:

- устойчивая пространственно-временная регулярность (выборы происходят через определенный промежуток времени на фиксированной территории, имеющей более или менее сложившуюся структуру избирательных округов);
- одновременное участие в них больших масс населения с широким спектром социологических параметров.

Это обстоятельство в значительной мере упрощает организационный аспект их изучения. Действительно, для того, чтобы проверить построенную модель, не требуется ждать неопределенно долгое время или провоцировать (что, к тому же, часто бывает опасным) требуемую ситуацию. Нужно лишь подождать очередных выборов, которые состоятся в худшем случае через пару лет. Мы также можем в первом приближении фиксировать начало и конец электорального процесса, что бывает исключительно важным при использовании количественных методов. И не вызывает сомнений то, что при изучении этих процессов в силу их массовости (даже при не очень высокой политической активности общества) мы имеем возможность уверенно пользоваться надежными и апробированными средствами статистики.

Говоря о моделировании электоральных процессов, следует сделать несколько предварительных замечаний. Эти модели, по крайней мере, на данном уровне развития теории, мало похожи на модели метеорологических явлений или развития эпидемий, где существуют достаточно точные методы прогнозирования некоторых возможных ситуаций. В нашем случае

возможные модели позволяют лишь качественно описать возможный ход развития событий и прогнозировать общие тенденции. И это уже не мало!

Социальные процессы вообще и электоральные, в частности, обладают крайне сложной природой и потому с трудом поддаются изучению. Это, в числе прочего, связано с тем, что они определяются, по крайней мере, двумя группами факторов:

- субъективные факторы, связанные с сознательной, целенаправленной деятельностью участвующих в данных процессах людей;
- объективные факторы, связанные с тем, что в социальной системе, состоящей из большого числа элементов (отдельных лиц, больших и малых социальных групп, организаций и т. д.) действуют обычные статистические законы, придающие динамике этих процессов дополнительные характеристики.

В отношении субъективных факторов мнения образуют широкий спектр, на краях которого находятся следующие полярные точки зрения:

- здесь все ясно, и, зная мотивацию действующих лиц (а ее при желании всегда можно выяснить тем или иным способом), мы можем точно предсказать ход событий;
- душа человека – потемки, людям свойственно думать одно, говорить другое, делать третье, а последнее, как правило, не получается из-за недостатка личных качеств и внешних обстоятельств, и поэтому о ходе событий сказать ничего нельзя.

Каждое из приведенных представлений содержит долю истины, и понимание этого обстоятельства должно определять стратегию изучения общественных явлений. Очевидно, что невозможно разработать и применять на практике более или менее надежную методику. Единственное, что остается – это доверять информированности и интуиции аналитиков, профессионально занимающихся этими вопросами. Все-таки им обычно удается прогнозировать некоторые общие закономерности процессов, которые отчетливо проявляют себя на уровне общественной практики.

Совершенно иначе обстоит дело с объективными факторами, обладающими статистической природой. Они вполне поддаются точному анализу и допускают количественные оценки, которые могут использоваться для построения обоснованных прогнозов. Тем более, что современная статистика располагает большим арсеналом детально разработанных и апробированных многолетним опытом методов. Однако слабой стороной является то обстоятельство, что статистика сама по себе позволяет описывать лишь формальные аспекты изучаемых явлений, содержательные же аспекты статистическому анализу пока не поддаются в принципе.

Поэтому возникает необходимость расширения набора инструментальных средств, используемых в изучении социальных процессов. Одним из наиболее интересных и перспективных направлений в этом плане является, безусловно, математическое моделирование. Главное его достоинство - возможность одновременного участия как формальных, так и содержательных аспектов, определяющих динамику исследуемого процесса. Действительно, структура модели в той мере, в какой мы ее сознательно строим, отражает смысл процесса, как мы его субъективно понимаем. С другой стороны, то, что модель строится с помощью математических методов, обеспечивает ее формальную строгость и надежность полученных результатов.

Уравнения, образующие модель какого-либо процесса, описывают общий характер его протекания, а связь с реальностью обеспечивается посредством использования набора эмпирических параметров, которые соответствуют конкретным условиям. Определение их значений представляет собой отдельную задачу, решение которой часто наталкивается на серьезные трудности. Однако в ряде случаев для понимания происходящего достаточно сопоставить поведение решений данных уравнений, полученных при разных значениях параметров. Этот метод особенно продуктивен для выявления общих тенденций в той или иной ситуации.

Применительно к социальным явлениям перспективным, на наш взгляд, является моделирование общественных процессов для изучения влияния изменения параметров на поведение системы. В такой постановке задачи не требуются точные значения параметров, просто ведется наблюдение за тем, что произойдет, если изменить, скажем, в два раза, значение того или иного параметра. Возможна также и своего рода обратная задача: по реальному поведению некой зависимости оценить величину интересующего нас параметра.

Знание общего поведения устойчивых решений позволяет прогнозировать развитие общих тенденций в изучаемой системе, даже в том случае, когда мы не имеем точного представления о конкретных механизмах, определяющих ход развития событий, причем такого рода прогнозы могут оказаться более точными, чем полученные традиционными социологическими методами. Если же решения оказываются неустойчивыми, то из этого также может быть получена ценная информация о нашей системе. Во всяком случае, будет известно, что такая система может мгновенно изменить траекторию своего движения, и в ряде случаев предсказать, куда такие изменения могут быть направлены.

Попытки моделирования общественных процессов предпринимались давно, но они тормозились вычислительными трудностями, особенно в случае нелинейных уравнений, описывающих динамику систем с обратными связями. Сейчас в нашем распоряжении есть достаточно возможностей для машинной обработки данных, что позволяет, с одной стороны, подготавливать наборы входных параметров на основании анализа результатов статистических исследований, а с другой стороны – решать требуемые уравнения с хорошей степенью точности и за разумное время. Кроме того, современные пакеты прикладных программ, предназначенных для решения математических задач, позволяют не только оперативно получать решения, но и подавать их в удобной для пользователя форме,

например, в виде графиков, отображающих представляющие интерес тенденции.

Поэтому есть основания полагать, что со временем математическое моделирование общественных процессов станет основным инструментальным средством социологии.

Ниже мы воспользуемся этим методом для анализа некоторых закономерностей в электоральных процессах.

2.1. Нелинейные конкурентные модели

Прежде чем говорить собственно о нелинейных конкурентных моделях, определим некоторые важные моменты, связанные с самим понятием нелинейности.

Термин «нелинейный» употребляется во многих значениях, из которых выберем одно, наиболее часто применяемое в теории систем и имеющее прямое отношение к социальным явлениям. Речь идет о нелинейных системах, в широком смысле слова, и для начала вспомним, что такое система.

Согласно классическому определению Людвиг фон Берталанфи, система – это комплекс взаимодействующих элементов [57].

Это, на первый взгляд, простое определение на самом деле является необычайно глубоким и дает пищу для разнообразных размышлений. На эту тему написано множество толстых книг, и, при желании, можно написать еще столько же. Не будем углубляться в тонкости общей теории систем, ограничившись коротким комментарием, полезным для понимания нижеизложенного.

Главное состоит в том, что система обладает свойствами, которые принципиально не сводятся к свойствам (сумме свойств) образующих ее элементов. Так, например, системой являются часы – прибор, показывающий текущее время, однако ни одна из его деталей время показывать не может.

Она не может показывать, образно выражаясь, даже «часть времени». Свойство показывать время появляется у всех деталей вместе, причем после того, как они будут определенным способом собраны в единый комплекс и, тем самым, вступят друг с другом в определенные взаимодействия. Возникающие таким образом связи между элементами системы образуют ее структуру. В этом смысле не будет ошибкой сказать, что структурой обладают только системы: совокупности невзаимодействующих элементов бесструктурны.

Возможно, что некоторое подмножество элементов системы А также, в свою очередь, образует систему В. Тогда говорят, что В является подсистемой А. Систему содержащую в себе хотя бы одну подсистему, называют сложной.

Если состояние системы не меняется во времени, ее называют статической, в противном случае – динамической. Ясно, что в плане изучения окружающего мира основной интерес представляют именно динамические системы: в первую очередь нас интересуют происходящие вокруг изменения, а статические системы никаких изменений не порождают. Однако существует особый случай, часто встречаемый в биологических сообществах, в том числе и в социуме. Он называется гомеостазом. В гомеостазе значения сущностных параметров, определяющих его как некую целостность, остаются неизменными во времени, но при этом его состояние может изменяться. Примером гомеостаза в мире неживой материи может служить, например, обычный холодильник. В нем постоянно происходят какие-то процессы: по проводам течет электрический ток, мотор включается и выключается и т. д., и т. п. Но температура внутри холодильника, определяющая его сущность, сохраняется постоянной. Более сложный пример гомеостаза – племя индейцев бассейна Амазонки, из поколения в поколение живущих по законам предков под управлением общающихся с духами шаманов. В таком племени тоже что-то постоянно происходит, но, тем не менее, общие его характеристики остаются неизменными. Таким

образом, самая, пожалуй, интересная и важная особенность гомеостаза состоит в том, что его статика описывается его же динамикой.

Но вернемся к нашей теме. Динамические системы подразделяются на два класса: линейные и нелинейные. Линейными называют системы, характеристики которых не зависят от изменения их состояний. Напротив, характеристики нелинейных систем зависят от таких изменений.

Ярким примером, демонстрирующим свойства нелинейных систем, может служить принцип Ле Шателье-Брауна: если система испытывает внешнее возмущение, стремящееся изменить ее состояние, то в ней происходят процессы, направленные на компенсацию данного возмущения. Первоначально этот принцип был сформулирован применительно к нелинейной термодинамике, но в последствии выяснилось, что он справедлив для достаточно широкого класса систем, в том числе и социальных.

С точки зрения стороннего наблюдателя нелинейные системы проявляют себя, прежде всего, несоразмерностью отклика на внешнее воздействие. Хорошо известно, что общественные системы могут поразительно легко и без последствий переносить тяжелые потрясения, и вместе с тем мгновенно «пойти в разнос» от малозначимого события. Именно эта особенность и делает изучение нелинейных систем, с одной стороны интересным, а с другой – трудным.

Модели, используемые применительно к нелинейным системам, также называют нелинейными, при чем не только ради единства терминологии. Дело в том, что модель - это тоже система, и она, естественно, может быть нелинейной.

Смысл построения и дальнейшего применения нелинейных моделей состоит в том, что их нелинейность формально выражена в структуре используемых уравнений, а их решение в некоторых случаях может быть вполне осуществимой задачей. Если нас устраивают численные решения (а

на практике обычно именно так и бывает), то с помощью современных компьютеров эта задача разрешима почти всегда.

Основной акцент при построении нелинейных конкурентных моделей электоральных процессов в настоящее время делается на анализе принципиальных внутренних взаимодействий динамических систем на основании логистических моделей. Моделирование динамики развития на основе дифференциальных логистических уравнений широко используется для моделирования самых разнообразных как естественных, так и социальных процессов.

Естественно, перед применением математических моделей необходимо обосновать их адекватность. Для этого используются хорошо известные методики, в частности, ретроспективный анализ.

Простейшая модель роста $\dot{x} = kx$, была предложена Мальтусом (для роста населения Земли). Она ведет, как хорошо известно, к экспоненциальному (т.е. очень быстрому) росту населения x во времени. Эта модель может быть применима, например, к развитию приверженцев новой политической силы на начальном этапе.

Изучая динамику электоральной популяции, мы будем для краткости говорить о ее численности, имея в виду, что на самом деле речь может идти также и об ее политическом весе. Различать эти два понятия будем при необходимости тогда, когда точная формулировка требуется для понимания механизмов процесса.

Мы уже говорили о существовании отчетливых аналогий между биологическими и электоральными популяциями. Эти аналогии позволят нам воспользоваться достаточно развитым инструментарием для построения интересующих нас математических моделей. В его основе лежит представление о том, что скорость изменения численности популяции в каждый момент времени пропорциональна ее текущей величине. Тогда для численности популяции $n(t)$ имеем уравнение:

$$\begin{aligned}\frac{dn(t)}{dt} &= kn(t), \\ n(0) &= n_0.\end{aligned}\tag{1.1}$$

где n_0 – начальная численность популяции, а коэффициент k соответствует скорости роста.

Его обычно называют законом Мальтуса, хотя это на самом деле не совсем верно (сам Мальтус формулировал свой закон в терминах геометрической прогрессии).

Такая зависимость действительно характерна для ряда динамических систем. Иногда она принимается в качестве разумного допущения.

В динамичной электоральной популяции с полным основанием можем утверждать, что каждая ее особь (будем употреблять этот стандартный в биологии термин без какого-либо подтекста; просто надо же как-то называть образующие ЭП «элементы») оказывает на свое окружение, а также на другие электоральные популяции, воздействие, влияющее на их численности. Таким образом, чем больше в популяции особей, тем сильнее влияние популяции и, поскольку оно приводит к изменению ее численности, выше скорость количественного её изменения. Поэтому мы можем говорить, что скорость изменения численности ЭП, по крайней мере в определенных пределах, действительно пропорциональна ее текущей величине.

Решением (1.1), как известно, является знаменитая экспонента, в недавнем прошлом повергшая в суеверный ужас членов Римского клуба.

$$n(t) = e^{kt}.\tag{1.2}$$

Подобная реакция, впрочем, вполне понятна, так как ни один реальный процесс, ни физический, ни общественный, не может развиваться по экспоненциальному закону неограниченное время. Действительно, начиная с некоего момента, зависимость слишком быстро стремится к бесконечности, которая в природе, по понятным причинам, не реализуется. Следовательно, приходится признать, что рано или поздно, и скорее рано, чем поздно,

произойдет некая катастрофа, которая изменит характер зависимости и вернет ее в рамки допустимого интервала значений.

Вообще, уравнение (1.1) не пригодно для описания равновесных состояний: его решения, в зависимости от знака коэффициента k , либо неограниченно возрастают, либо асимптотически стремятся к нулю.

Таким образом, закон Мальтуса если и описывает реальные процессы, то лишь на начальных их стадиях, где экспоненциальная зависимость не слишком сильно отличается от линейной.

В более или менее стабильных системах, встречающихся на практике, всегда присутствует элемент самосогласованности, в силу чего на значительных промежутках времени зависимость динамики их развития описывается более сложными уравнениями, содержащими обратные связи. Поэтому характер зависимости со временем меняется, причем изменения происходят не тривиальным образом. В результате возникают следующие типичные случаи:

- зависимость достигает насыщения и система переходит в статическое (или, возможно, гомеостатическое) состояние;
- зависимость имеет локальный максимум, за которым следует убывание (в том числе и до нуля);
- устанавливается колебательный режим (обычно затухающий, но возможен и автоколебательный);

Отметим, что случаи неограниченного возрастания одного или более параметров действительно возможны (т. н. резонансные явления), и они в конечном счете приводят к физическому разрушению системы (например, в случае резонанса в акустической системе – разрыв диффузора динамика). Такие ситуации полезно учитывать и при анализе социальных процессов, хотя и встречаются крайне редко.

В реальной жизни, как правило, динамические системы обладают достаточно эффективными обратными связями, позволяющими корректировать характер происходящих в них процессов и тем самым

удерживать их в определенных рамках. Природа и действие таких механизмов далеко не всегда очевидны, более того, часто мы можем лишь догадываться об их наличии по общему характеру поведения системы.

Казалось бы, естественным в данном положении было отказаться от закона Мальтуса и попытаться найти более совершенное уравнение. Сложность, однако, состоит в том, что тогда пришлось бы также отказаться и от линейной зависимости скорости изменения численности ЭП от ее текущей величины, что в ряде случаев действительно имеет место на начальной стадии. Поэтому наиболее приемлемым путем решения проблемы в таких случаях следует считать использование обобщенного закона Мальтуса, который на больших промежутках времени давал бы зависимость, удовлетворительно описывающую поведению интересующей нас системы, а на малых – экспоненциальную.

Наиболее простым обобщением закона Мальтуса (1.1), позволяющим решить (по крайней мере, принципиально) проблему неограниченного возрастания решения, является замена постоянного коэффициента k некоторой функцией времени $k(t)$. Естественно, эта функция должна быть выбрана так, чтобы соблюдались следующие условия:

- решение уравнения обладало бы приемлемым поведением;
- структура функции имела бы определенный смысл с точки зрения изучаемого явления.

Ниже используется широко применяемое обобщение закона Мальтуса, известное как логистическая модель. В рамках данного исследования нет намерения строить рабочие модели конкретных процессов, описывающие реальные ситуации. Цель данной работы состоит в иллюстрации возможностей математического моделирования электоральных процессов на интуитивно понятных примерах. Однако, при желании, методы, приведенные ниже, могут успешно применяться на практике, разумеется при условии, что в распоряжении имеются надежные способы определения численных значений эмпирических параметров. А это требует комплексного подхода с

участием специалистов в различных областях, а также практической возможности проводить требуемые социологические исследования.

Ясно, что в дальнейшем такие темпы роста электората становятся невозможными, и такое резкое развитие прекращается. Аналогичные явления насыщения происходят в любой общественной среде: когда численность приверженцев определенной политической силы становится слишком большой, мальтусова модель с постоянным коэффициентом роста k перестает быть актуальной. Естественно, при слишком больших x конкуренция за политические ресурсы приводит к уменьшению k , и модель Мальтуса должна быть заменена моделью с переменным показателем $k(x)$:

$$\dot{x} = k(x)x,$$

с зависимым от численности электората коэффициентом его роста. Простейшим примером является выбор $k(x) = a - bx$, что приводит к так называемой логистической модели.

Именно такая модель может служить наиболее простым описанием динамики численности приверженцев отдельно взятого кандидата. Логистическая модель была предложена П. Ферхлюстом [83] для описания динамики человеческого населения и Р. Перлом [77] для биологических сообществ. В соответствии с ней, динамика численности электората описывается обыкновенным дифференциальным уравнением:

$$\dot{x} = ax - bx^2,$$

где $a = \text{const}$ представляет собой максимальную удельную скорость роста электората, коэффициент $b = \text{const}$ описывает отрицательные для кандидата процессы, связанные с внутренними факторами. Коэффициент a представляет собой разность между приростом электората C и уменьшением D : $a = C - D$ в бесконкурентной среде.

При малых x (т.е. когда $bx^2 \ll ax$) логистическое уравнение описывает экспоненциальное увеличение численности электората. Кроме того, экспоненциальное увеличение часто наблюдается на отдельных промежутках

времени при стечении благоприятных факторов развития электорального поля.

Очевидно, что неограниченный рост электората невозможен. Начиная с некоторого момента, рост будет тормозиться в результате недостатка ресурсов - количество избирателей ограниченное, конкуренция, внутренние противоречия в электорате и т.п. Поэтому часто электоральное поле, попав в новую социальную среду, обнаруживают так называемое S -подобный рост численности. Т.е. кривая временной зависимости численности имеет ярко выраженную S -образную форму [28]. Нетрудно видеть, что логистическое уравнение описывает все эти основные свойства развития отдельного электората. Его аналитическое решение дается формулой логистической функции:

$$x(t) = \frac{B}{1 + \left(\frac{B}{x(t_0)} - 1 \right) e^{-a(t-t_0)}},$$

при равновесном значении $B = a/b$. График функции x , имеет S -образную форму (рис. 1).

Как правило, максимальная численность электората кандидата, который может стойко существовать в данной избирательной кампании, ограниченная сверху определенной величиной, которую называют емкостью избирательной кампании.

Выбором системы единиц x и t можно превратить коэффициенты a и b в 1 (т.е., к виду $\dot{x} = x - x^2$). Необходимо подчеркнуть, что выводы, которые будут сделаны ниже, остаются (с точностью до числовых значений констант) справедливыми и при любых значениях коэффициентов a и b и даже для широкого класса моделей с разными функциями $k(x)$.

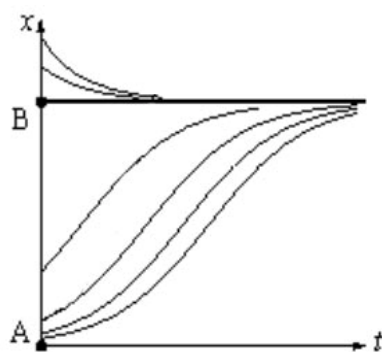


Рис. 1. Логистическая модель

На рис. 1 изображена результирующая зависимость численности электората от времени при разных начальных условиях. В точках A и B скорость изменения численности электората равна нулю: это стационарное состояние. Между A и B скорость положительная (численность электората возрастает), а за точкой B - отрицательное (численность убывает).

Модель предусматривает, что со временем устанавливается стационарный режим B , что выглядит вполне естественно: больший электорат уменьшается, меньший - увеличивается.

Логистическая модель удовлетворительно описывает многочисленные явления насыщения [1], [2]. Вблизи A , когда численность электората малая, она очень близкая к мальтузианской модели. Но при довольно больших x (порядка $1/2$ при нашем выборе коэффициентов) наблюдается резкое отличие от мальтузианского роста: вместо следования x к бесконечности численность электората приближается к стационарному значению B (рис. 2).

Посмотрим, как логистическая модель может применяться при анализе избирательных процессов, а именно определение точной минимальной начальной квоты избирателей (которую можно, например, выделить для утверждения и использование). Можно себе представить, например, что x - это количество приверженцев определенного кандидата. На динамику этого электората оказывают влияние другие политические силы, которые отбирают своих приверженцев, что описывается таким образом:

$$\dot{x} = x - x^2 - c.$$

Вычисления показывают, что ответ резко меняется при некотором критическом значении квоты c . Для приведенной модели это критическое значение есть $c = 1/4$.

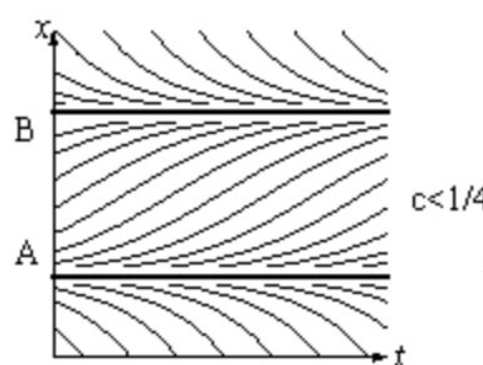


Рис. 2. Две постоянные точки модели

Если квота c мала, то изменения (в сравнении с отсутствием конкуренции, когда $c = 0$) состоят в следующем. Система, модель которой изображена на рис.2, имеет два равновесных состояния, A и B . Состояние B устойчивое: численность электората в этом случае немного меньше, чем при бесконкурентной ситуации, но оно восстанавливается при малых отклонениях x от равновесного значения B .

Состояние A неустойчивое: если вследствие каких-нибудь причин (скажем, политических преследований, незаконного административного влияния и т.п.) численность электората упадет ниже уровня A , то в дальнейшем электоральное поле (хотя и медленно, если отличие от A небольшое) будет сведено на нет за вполне конечное время.

При больших значениях критической квоты c электорат x исчезает за конечное время, каковым бы большим оно не было в начальный момент.

Очевидно, что при наличии благоприятных внешних условий (при некоторой плотности ресурса) численность электората одного кандидата развивается свободно, что оказывает содействие логистическому росту. В этом случае даже более сложные модели должны давать результаты, подобные приведенным. С другой стороны это означает, что основные параметры (электоральный потенциал кандидата, емкость избирательной

компании, избирательная квота) для конкретизации общей модели, могут определяться в результате анализа приведенной упрощенной логистической модели. Как правило, данные о логистическом росте электората, становятся известными довольно точно по результатам наблюдений за избирательной кампанией.

Самый простой способ ограничить рост экспоненциальной зависимости (1.2) решения уравнения (1.1) заключается в том, чтобы ввести для нее предельное значение. Для этого выберем $k(t)$ такого вида:

$$k(t) = k_0[N - n(t)], \quad (1.3)$$

где N – максимально возможное значение численности популяции, а k_0 – коэффициент пропорциональности. Причем предполагается, что всегда $n_0 \leq N$. Тогда вместо (1.1) имеем

$$\begin{aligned} \frac{dn(t)}{dt} &= k_0 n(t)[N - n(t)], \\ n(0) &= n_0. \end{aligned} \quad (1.4)$$

Решение уравнения (1.4) имеет следующий вид:

$$\begin{aligned} n(t) &= \frac{Ne^{at}}{b + e^{at}}, \\ a &= k_0 N, b = \frac{N}{n_0} - 1. \end{aligned} \quad (1.5)$$

Мы видим, что по мере приближения численности популяции $n(t)$ к N , скорость ее изменения стремится к нулю.

Модель, основанная на уравнении (1.4) в литературе называется логистической. При кажущейся простоте, подобное обобщение закона Мальтуса отнюдь не является примитивным. Напротив, оно позволяет явно включить в описание динамики популяций исключительно важную обратную связь, роль которой в окружающем нас мире трудно переоценить. Логистическое уравнение (1.4), по существу, следует считать феноменологическим: мы не знаем, как действуют конкретные механизмы, снижающие по мере роста численности популяции скорость ее изменения. И

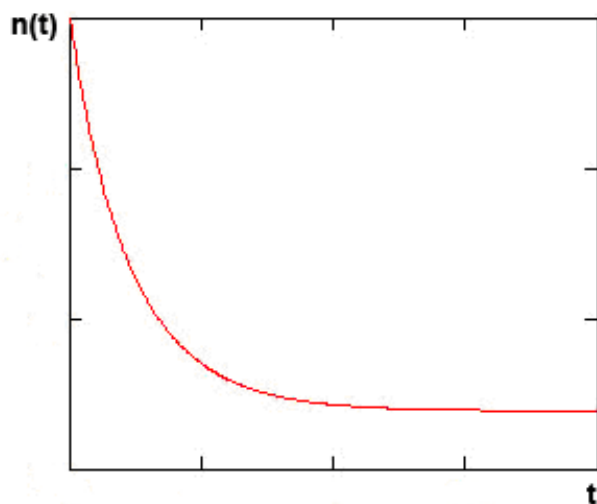
это, в данном случае, серьезное преимущество, поскольку в исторически обозримом будущем мы, скорее всего, даже не приблизимся к пониманию таких механизмов, а раз так, о создании полноценной теории популяционных процессов пока не приходится говорить.

Существует два класса решений уравнения (1.4), которые, в зависимости от значений коэффициента k_0 и n_0 , описывают возрастание и убывание зависимости $n(t)$. Их типичное поведение изображено на рис. 3. Как видно, логистическая модель, в отличие от закона Мальтуса, описывает достижение системой некоторого равновесного состояния.

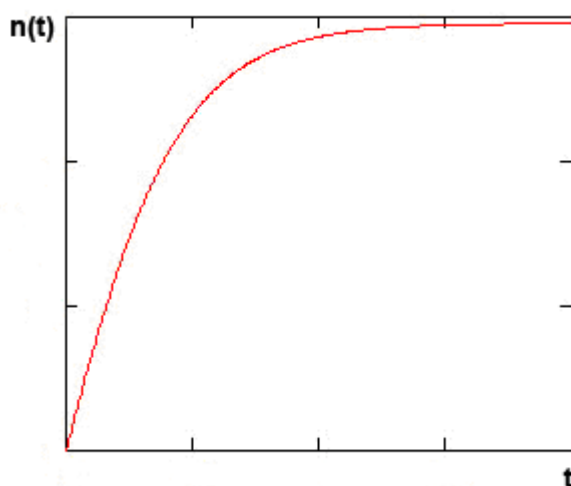
Уравнение (1.4) имеет два равновесных решения: $n(t)=0$ и $n(t)=N$. С формальной точки зрения первое из них неустойчиво, так как при малых значениях $n(t)$ ее отклонение от нуля приводит к росту. Однако в практическом плане это не совсем так. Дело в том, что реальные популяции являются дискретными множествами, и если в какой-то момент $n(t)$ принимает значение, меньшее единицы, то возрасти оно уже не сможет. Поэтому в случае описания того, что происходит в реальности, решение $n(t)=0$ также следует считать равновесным.

Второе же решение $n(t)=N$ является равновесным во всех смыслах. Действительно, при $n(t)>N$ включаются механизмы спада зависимости, а при $n(t)<N$ – соответственно роста.

Итак, логистическая модель успешно описывает достижение популяцией некоторого равновесного состояния.



a)



б)

Рис. 3. Два класса решений логистической модели: а) – снижение популяции, б) – рост популяции

Взаимодействие популяций

Уравнение (1.4) описывает динамику одной популяции, взаимодействующей только с окружающей средой.

В реальном мире подобные ситуации возникают крайне редко и не представляют особого интереса. Как правило, различные популяции сосуществуют друг с другом и при этом активно взаимодействуют между собой.

В теории популяционной динамики разработана классификация различных форм такого взаимодействия [8, 10, 11]. К числу основных относятся следующие:

- нейтрализм (отсутствие прямого воздействия популяций друг на друга);
- конкуренция (взаимное подавление популяций);
- амменсализм (одностороннее подавление одной популяции);
- хищничество (уничтожение особями одной популяции особей другой);
- симбиоз (продуктивное сосуществование популяций).

Каждая из них, в свою очередь, имеет варианты, поэтому общая картина взаимоотношений между популяциями выглядит достаточно сложной и разнообразной. Следует также учитывать, что взаимодействие популяций может быть не только прямым (например, поедание одним видом другого), но и опосредованным (например, совместное потребление ограниченных ресурсов).

В динамике взаимодействующих популяций выделяются две категории воздействий, отличающиеся временным характером:

- фазовые (однократные);
- параметрические (постоянные).

Логистическая модель позволяет вполне удовлетворительно описывать и динамику M взаимодействующих между собой популяций. В общем случае это осуществляется с помощью следующей системы уравнений:

$$\frac{dn_i(t)}{dt} = n_i(t) \left[p_i - \sum_{j=1}^M q_{ij} n_j(t) \right], \quad (1.6)$$

$$n_i(0) = n_{0i}.$$

Тип процесса, описываемого этой системой уравнений (1.6), определяется величиной и знаком коэффициентов p_i и q_{ij} . Следует также иметь ввиду, что в каждом уравнении диагональные члены $n_i(t)n_j(t)$

описывают внутривидовое взаимодействие, а перекрестные $n_i(t)n_j(t)$ – межвидовое.

Другими словами, диагональные члены описывают влияние на популяцию внешней среды, в том числе исчерпание доступных ресурсов, а перекрестные – воздействие одной популяции на другую (положительные значения соответствуют благоприятному влиянию, отрицательные – неблагоприятному). Коэффициенты p_i имеют смысл скоростей роста соответствующих популяций при отсутствии взаимодействия.

Важным моментом является также поведение популяции при заданных значениях параметров и в отсутствии взаимодействия (например, ее рост ограничен сам по себе).

Система уравнений (1.6) в принципе может описывать широкий спектр зависимостей, и это, в определенном смысле, является проблемой, так как при желании из ее решений можно «вытащить» все, что угодно. Поэтому работа с ней требует взвешенного и ответственного отношения.

Однако решения, характеризующие реальные процессы, обычно относятся к одному из следующих режимов:

- стационарный;
- автоколебательный;
- квазистохастический.

Как правило, эти режимы в полной мере проявляют себя на достаточно больших (не обязательно бесконечных) промежутках времени. Но вот переходные процессы, которые предшествуют установлению определенного режима, исключительно полиморфны и с большим трудом поддаются классификации. В биологии это не имеет значение, поскольку там интерес представляют как раз значительные периоды времени. Предполагается, что популяции возникли давно, и нам не важно, как именно. Нас интересует лишь то, как они воспроизводят себя во времени.

Однако в случае электоральных процессов ситуация становится совершенно иной. Поэтому здесь требуется глубокий и всесторонний анализ ситуации.

Система уравнений (6) позволяет описывать динамику любого числа популяций, как взаимодействующих между собой, так и находящиеся в изоляции. Но для понимания основных ее закономерностей бывает достаточно ограничиться небольшим их числом. Даже изучение поведения двух взаимодействующих популяций позволяет проследить общие закономерности их динамики, по крайней мере, качественные. Практика показывает, что для разумных прогнозов этого вполне достаточно.

Ниже исследуется динамика трех ЭП, из которых две считаются основными в том смысле, что их поведение должно иллюстрировать интересующие нас моменты, а третья (дополнительная) введена для демонстрации роли, которую в данном случае может играть политический контекст. При этом изучаются три основных варианта взаимодействия политических сил: конкуренция, хищничество и симбиоз.

Поскольку аналитические решения системы уравнений (6) в случаях, когда они могут быть построены, оказываются громоздкими и плохо поддающимися анализу, изначально применяются численные методы, тем более, что графическая форма представления результатов в данном случае наиболее удобна и наглядна. Поскольку нас интересует качественное поведение зависимостей, результаты представлены в условных единицах. На графиках приняты следующие соглашения: кривая 1 изображает зависимость первой политической силы ($i=1$), кривая 2 – второй ($i=2$) и 3 – дополнительной ($i=3$). Для большей наглядности зависимости основных и дополнительных ЭП представлены на отдельных графиках.

Динамика ЭП – «Конкуренция»

Конкуренция представляет собой форму взаимодействия популяций, при которой они взаимно подавляют друг друга из-за ограниченности общей ресурсной базы. Главной особенностью конкуренции является то, что конкурирующие популяции не оказывают друг на друга непосредственного воздействия. Взаимодействие осуществляется опосредованно, путем вытеснения друг друга из области ограниченных ресурсов. При этом возможно полное подавление одной из популяций, в результате чего она исчезает. Именно конкурентные отношения представляют реальную опасность для ЭП (как, впрочем, и для любых популяций).

Именно конкурентные отношения наиболее характерны для основных участников электоральных процессов.

В зависимости от условий, в которых находятся взаимодействующие популяции, и значений определяющих динамику параметров, возможны как различные равновесные состояния системы, так и механизмы их достижения. В частности, возможны различные варианты сосуществования конкурирующих ЭП, причем в процессе установления этих состояний численность каждой из популяций может как возрастать, так и убывать. Более того, возможны случаи, в которых тенденция может меняться во времени (популяция вначале растет, а затем начинает сокращаться, или наоборот). А отсюда следует, что окончательный результат конкуренции политических сил далеко не всегда совпадает с прогнозами, сделанными на основании интуитивных оценок их начальных потенциалов.

Ниже приведены наиболее характерные, на наш взгляд, случаи динамики популяций при конкурентных отношениях.

Конкуренции соответствует система уравнений (1.6) с положительными значениями коэффициентов q_{ij} .

Равновесное сосуществование сил в собственных политических нишах

Будем считать, что взаимодействие между двумя основными силами, с одной стороны, и третьей силой с другой в основном сводится к взаимному ограничению ресурсной базы. Главная конкуренция имеет место между основными силами.

При достаточно малых значениях коэффициентов q_{ij} , описывающих влияние одной конкурирующей популяции на другую, и достаточно больших значениях скоростей их роста каждая популяция достигает равновесного состояния и стабилизируется в нем. В зависимости от значений других параметров и начальной численности популяция в процессе достижения этого состояния может как возрастать, так и сокращаться.

Как видим на рис. 4, обе основные ЭП достигают равновесного значения, однако при этом одна из них возрастает, а другая сокращается. Разумеется, при выборе других значений параметров обе популяции могут и возрастать, и сокращаться.

На первый взгляд, приведенные результаты выглядят совершенно очевидными, но, на самом деле, если учесть, что речь идет о конкурентных отношениях, они отнюдь не тривиальны. Оказывается, что обе конкурирующие ЭП могут оказаться в равновесном состоянии, из которого они сами по себе не смогут выйти ни при каких условиях.

Полное подавление одной силы другой

При значительной величине q_{ij} численность одной из популяций сокращается до нуля (рис. 5). Если значения коэффициентов q_{12} и q_{21} близки, то ситуация становится неустойчивой в том смысле, что будет ли подавлена популяция, зависит от малых отклонений в значениях других параметров.

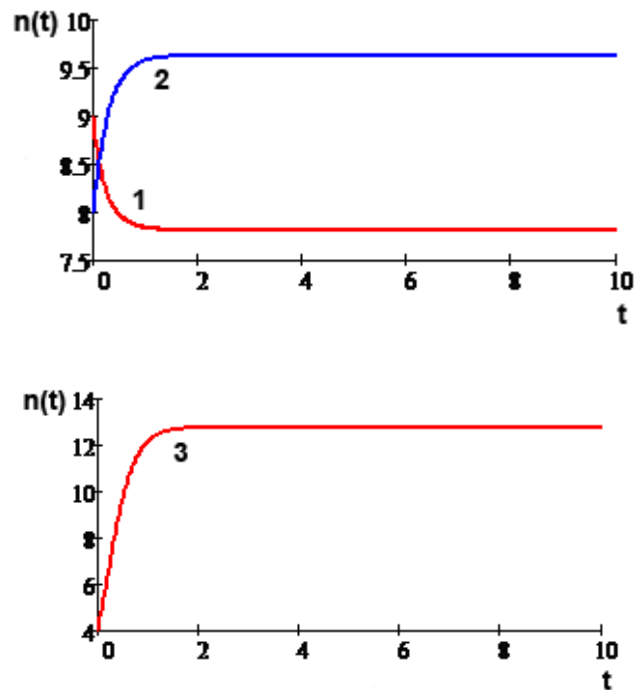


Рис. 4. Равновесное сосуществование сил в собственных политических нишах

Подавление одной из конкурирующих ЭП другой – сценарий, который большинством населения воспринимается как естественный и закономерный. Поэтому приведенные результаты как таковые не представляют особого интереса. Интересно другое: при выбранных условиях победившая популяция не истощала «освободившиеся ресурсы». Ее возрастание незначительно по сравнению с потерями проигравшей ЭП.

Таким образом, мы приходим к важному выводу. Победа в конкуренции не означает автоматической поддержки со стороны электората конкурента.

Равновесное сосуществование за счет воздействия третьей силы

В предыдущих случаях мы пренебрегали непосредственным влиянием третьей силы на основные. Однако если третья сила тем или иным способом

«подпитывает» одну из основных, их динамика может кардинально измениться (рис. 6). Например, возможен сценарий сосуществования.

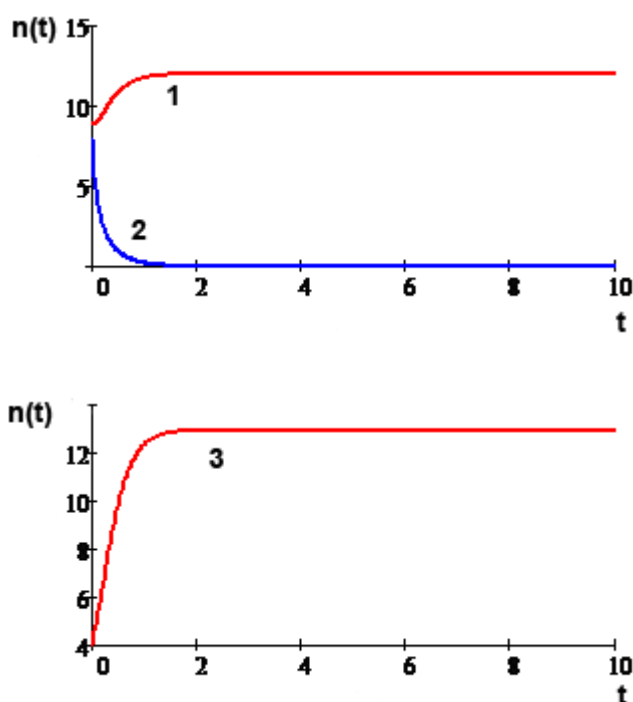


Рис. 5. Полное подавление одной силы другой

Равновесное сосуществование конкурирующих основных ЭП, возникающее за счет положительного воздействия на одну из них дополнительной ЭП иллюстрирует такую крайне важную вещь как влияние на взаимодействующие ЭП политического контекста, которое может быть как положительным, так и отрицательным. И в том, и другом случае, такое влияние способно эффективно компенсировать сильные и слабые стороны основных конкурентов, приводя к, казалось бы, неожиданным сценариям.

Зависимость от скорости роста

Одним из ключевых вопросов является устойчивость решений по отношению к численным значениям параметров. В ряде случаев поведение, в том числе и качественное, решений системы уравнений (6) очень сильно

зависит от них. Проиллюстрируем это на примере зависимости решений от скорости роста (рис. 7).

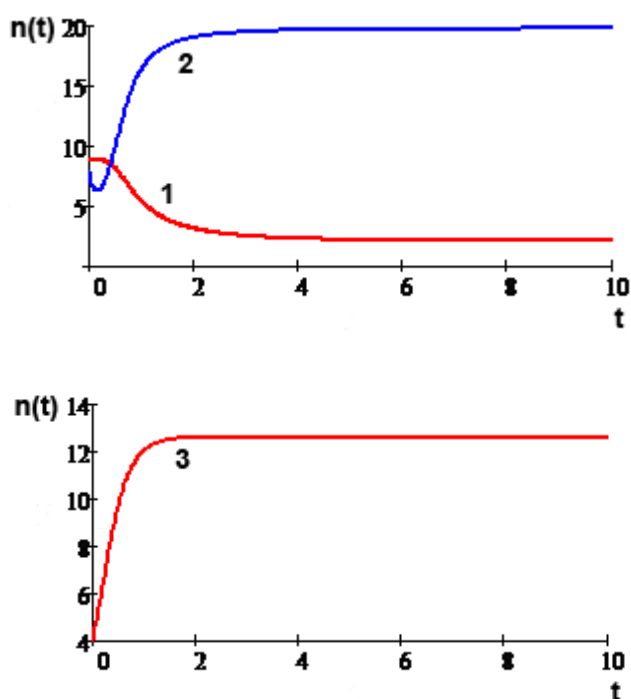


Рис. 6. Равновесное сосуществование за счет воздействия третьей силы

Приведенные графики отличаются значениями коэффициента ρ_2 (1.2985953 в первом случае и 1.2985954). Мы приводим округленные значения с точностью до седьмого знака после запятой, но в действительности различие между ними намного меньше.

Мы видим, что в данном случае поведение приведенных кривых отличаются не только количественно, но и качественно. Более наглядно это видно на соответствующих фазовых портретах (рис. 8).

Таким образом, динамика ЭП может существенно зависеть от малых изменений скорости их роста. Это означает, что политическая сила может победить из-за малого преимущества в этой характеристике.

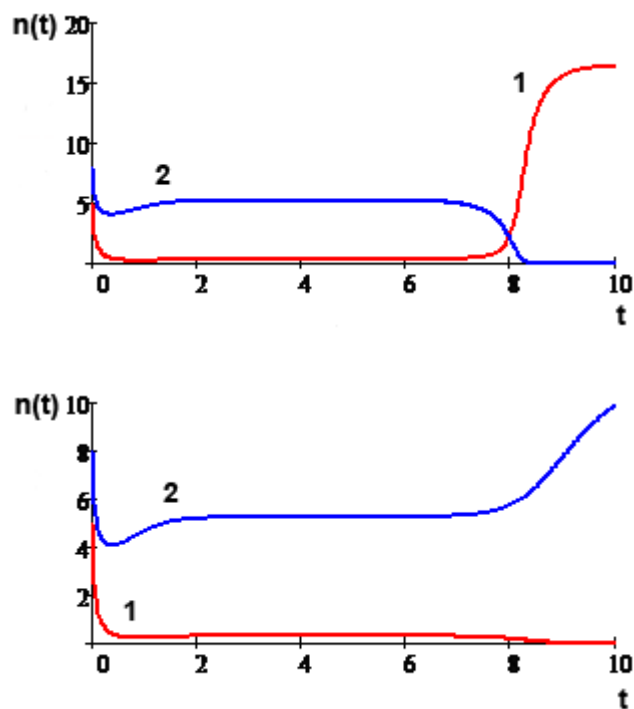


Рис. 7. Зависимость численностей популяций от скорости их роста

Отметим, что нетривиальное поведение первой зависимости: вначале доминирует вторая популяция, а первая - подавлена, но потом они меняются местами. Таким образом, возможны ситуации, когда конкуренция двух политических сил в присутствии третьей приводит к тому, что та из них, которая вначале была бесспорным лидером, со временем вытесняется на периферию.

Динамика ЭП – «Хищничество»

Хищничество часто рассматривают как разновидность конкуренции, при которой особи одной популяции непосредственно поглощают особей другой. В отличие от случая обычной конкуренции, здесь взаимоотношения популяций асимметричны: одна играет роль хищника, другая – жертвы, причем их нельзя поменять местами.

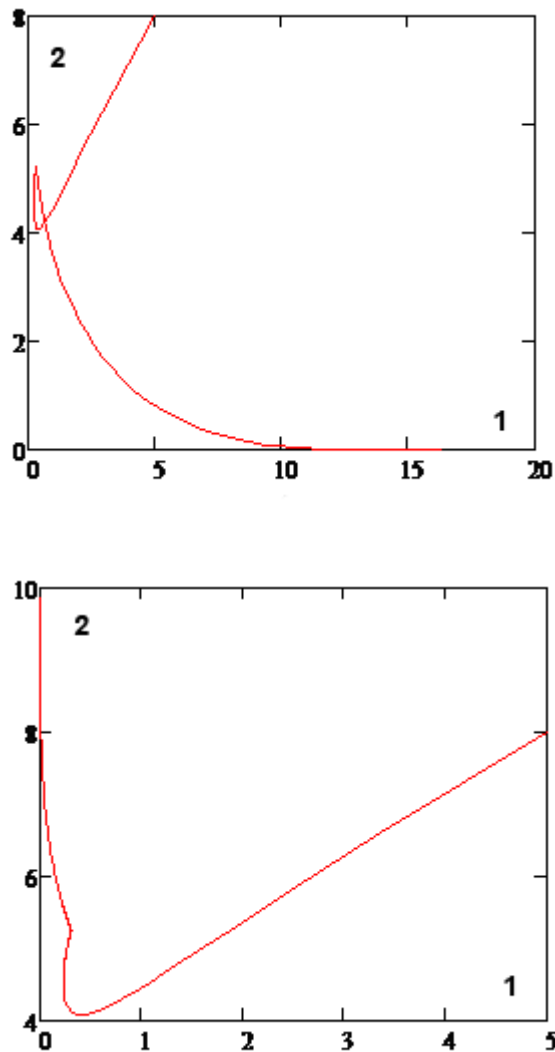


Рис. 8. Фазовые портреты зависимостей численности популяций от скорости их роста

Отношения «хищник-жертва» сами по себе никогда не приводят к подавлению одной популяции другой: хищники не могут воспроизводить свою популяцию без жертв – своей пищи (как в буквальном, так и переносном значении). Сокращение популяции жертв, в свою очередь, вызывает сокращение популяции хищников, так как часть из них остается без средств к существованию. Но, с другой стороны, сокращение числа хищников приводит к снижению внешнего давления на жертв, и они начинают восстанавливаться. Поэтому для данного случая характерны разнообразные колебательные моды.

Естественно, влияние третьей силы может существенно изменить типичную картину.

Этот вид взаимодействия политических сил является наиболее сложным и, вместе с тем, наиболее интересным. Главная причина состоит в том, что здесь основную роль играют механизмы второй группы.

Действительно, в случае взаимодействия биологических видов сокращение численности жертв, начиная с некоторого момента, ведет к сокращению численности хищников, так как тем самым сокращается их ресурсная база (доступные объемы пищи). В политической борьбе на уровне численности популяций ничего похожего не происходит. Напротив, представители одной политической силы, «съеденные» другой политической силой, механически увеличивают численность «хищников», причем без каких-либо ограничений. Партия, поглотившая остальные партии, становится политическим монополистом, как это произошло, например, в 1918 г. с большевиками.

Поэтому механизмы первой группы сами по себе в данном случае могут приводить лишь к тривиальным эффектам механического расширения одной силы за счет других.

С механизмами второй группы ситуация намного сложнее. Часто политический вес одной популяции основан на эксплуатации ресурсов, в первую очередь интеллектуальных, другой. Например, пропаганда конкурирующих сил может быть построена на противоположностях, которые исключают друг друга, но одновременно и придают друг другу смысл. Так, безыдейно пропагандировать синий цвет там, где все предметы синие. Партия, выдвигающая лозунги рыночных отношений не вызовет интереса в обществе, построенном на основе либеральной экономики. В таких случаях для поддержания активности необходим противник, с которым имело бы смысл полемизировать. И реальный «политический вес» определяется превосходством в этой полемике. Другой пример – использование «образа врага». Здесь нет полемики, политическая сила обращается с обещаниями

«каленным железом искоренить скверну» непосредственно к избирателям, но общий принцип тот же.

Также возможна систематическая эксплуатация одной силой идейных наработок, созданных другой. К ним могут относиться идеологемы, программные положения, методы критики общих противников и т. д.

При таких сценариях вполне можно говорить об отношениях «хищник-жертва» между политическими силами.

Проанализируем несколько ситуаций, характерных для отношений хищничества.

Хищничеству соответствует система уравнений (6) с отрицательными значениями коэффициентов p_i и q_{ij} для хищника и положительными для жертвы.

Слабое воздействие третьей силы

Если третья сила оказывает на основные силы слабое воздействие (не участвует в потреблении соответствующих ресурсов), то имеем обычную картину хищничества (рис. 9).

Наблюдаются типичные колебания численностей обеих ЭП, причем затухающие. При достаточно больших промежутках времени обе зависимости стремятся к состоянию некоторого равновесия, но амплитуда колебаний никогда не становится равной нулю.

В зависимости от значений параметров кривые могут иметь различные соотношения фаз, что иногда приводит к интересным эффектам, однако это тема отдельного исследования.

Из соображений наглядности сделаем равными начальными численности обеих основных популяций.

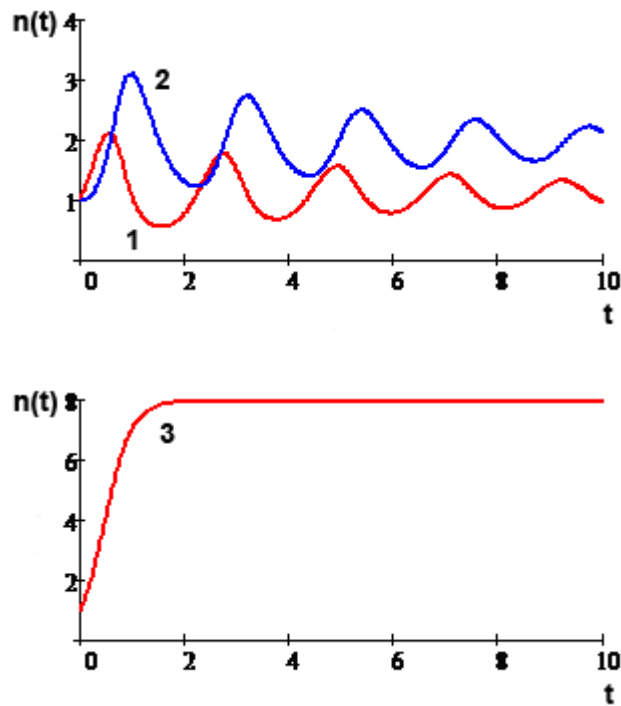


Рис. 9. Слабое воздействие третьей силы

Колебательный режим, характерный для отношений «хищник-жертва», в поведении динамики политических сил действительно наблюдался в целом ряде социологических исследований.

Зависимость от скорости поглощения

Выше мы видели, как поведение наших зависимостей зависит от скорости роста популяций. Теперь посмотрим, как они зависят от скорости поглощения хищниками жертв (рис. 10).

Сама по себе такая зависимость представляется совершенно естественной. Однако определенный интерес представляет вопрос о том, что именно изменится. Также полезно выяснить, насколько решения в этом случае будут устойчивыми.

Приведенные ниже зависимости отличаются от приведенных в предыдущем пункте значением коэффициента p_2 (-2 в первом случае и -4 во втором).

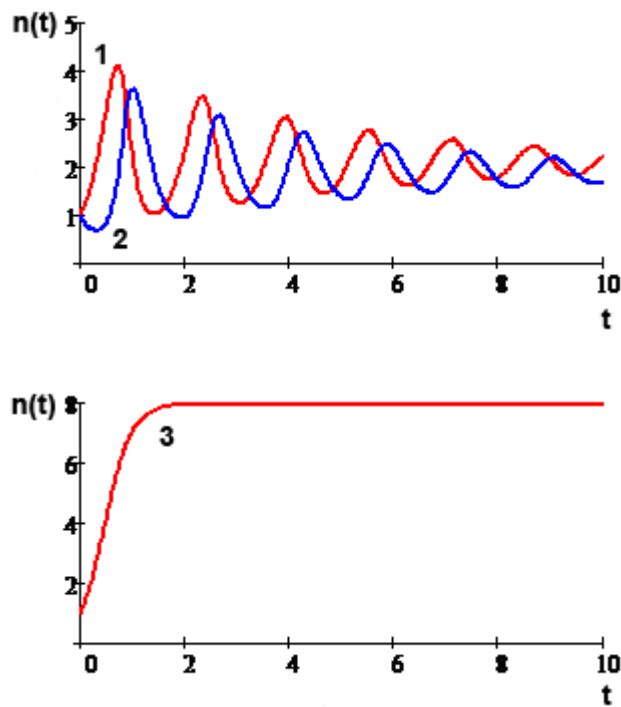


Рис. 10. Зависимость численности популяций от скорости поглощения

Прежде всего мы видим, что изменения в поведении кривых заметны, но вполне соразмерны с изменениями параметра, что свидетельствует об устойчивости данных решений. Кроме того, популяции поменялись местами по значениям амплитуд, и, что еще интереснее, изменилась частота колебаний.

Влияние третьей силы

Следующие две зависимости показывают нам, как на динамику хищника и жертвы может влиять воздействие третьей силы (рис. 11). Приведенные решения отличаются друг от друга значением коэффициента q_{13} . Он равен -1 в первом случае (положительное воздействие на первую из основных популяций) и 1 во втором (отрицательное воздействие на нее).

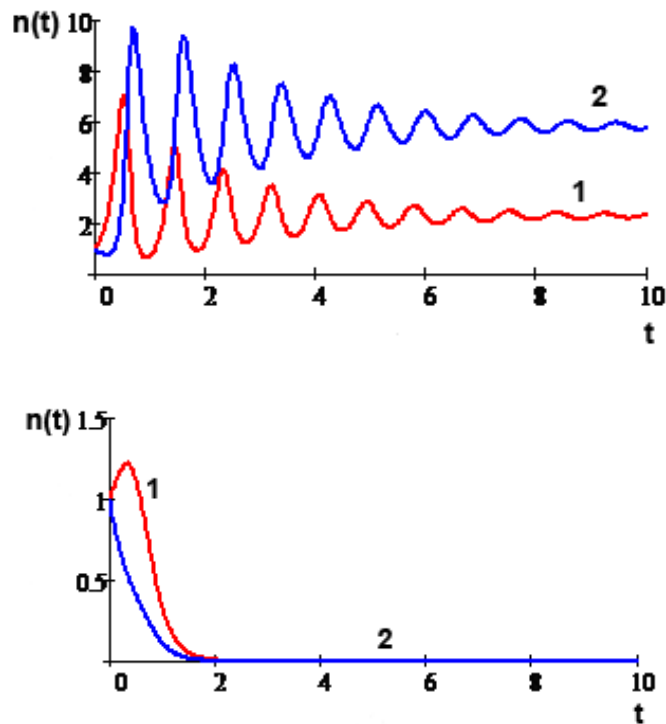


Рис. 11. Влияние третьей силы

При положительном воздействии третьей силы также наблюдаются существенные изменения характера колебательных процессов, в частности, возрастает их частота, но при этом сокращается время затухания (внешнее стабилизирующее влияние). Важным также является то обстоятельство, что популяции могут многократно меняться местами по величине своей численности. С точки зрения стороннего наблюдателя такой процесс может выглядеть странным и тревожным, вызывая мысли о политическом хаосе и «ситуации абсурда». Однако он быстро нормализуется и переходит в (квази)равновесное состояние.

Основные силы поглощают третью

Здесь приведен любопытный случай, когда основные ЭП пытаются полностью поглотить дополнительную, но в результате сами гибнут.

Третья сила вначале как будто действительно сокращает свою численность, но затем наступает период роста, и она становится монополистом (рис. 12).

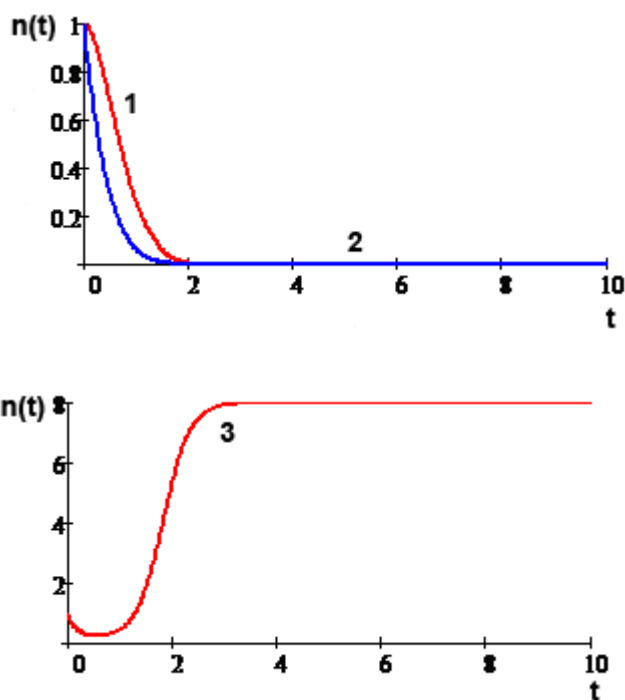


Рис. 12. Поглощение дополнительной силы основными

Ситуация, на наш взгляд, весьма поучительная.

Динамика ЭП – «Симбиоз»

Симбиоз политических сил возникает в том случае, когда по каким-либо причинам они или не мешают друг другу, или друг друга поддерживают. Например, в случае, когда они ориентируются на взаимоисключающие ценности и поэтому рассчитывают на поддержку непересекающихся секторов электората.

В ряде случаев популяции, находящиеся в отношениях симбиоза, оказывают друг на друга положительное влияние, помогая выжить в жесткой борьбе с другими популяциями. Однако не следует думать, что симбиоз всегда представляет собой мирное и благостное сосуществование в

политическом плане. На самом деле отношения могут быть жесткими и даже антогонистическими. Просто такие ЭП не имеют реальной возможности воздействовать друг на друга с помощью электоральных механизмов. Ведь, как уже говорилось, характер отношений между политическими силами определяется не только субъективными факторами, порождаемыми сознательной деятельностью их представителей, но и чисто объективными, связанными с коллективными общественными процессами.

Симбиоз, пожалуй, наименее интересен как в теоретическом, так и в прикладном плане. Действительно, динамика популяций, находящиеся в отношениях симбиоза, мало чем отличается от динамики невзаимодействующих популяций. Однако важен сам факт, что взаимодействующие популяции могут находиться в таком состоянии.

Тем не менее, для полноты изложения он включен в общую картину.

Симбиоз предполагает, что для всех участвующих в нем популяций коэффициенты p_i положительные, а q_{ij} – отрицательные.

Симбиоз основных сил при воздействии их на третью

Чтобы не упрощать анализ возможных ситуаций, рассмотрим, как выглядит влияние третьей силы. То есть, предположим, что симбиоз имеет место только между двумя основными ЭП.

Ниже приведены зависимости для случаев позитивного ($p_{3i} < 0$) и негативного ($p_{3i} > 0$) воздействий основных ЭП на дополнительную (соответственно, рис. 13 и 14).

Оказывается, что поведение популяций, связанных отношениями симбиоза, очень слабо зависит от их влияния на третью силу, а вот ее поведение зависит от такого воздействия кардинально. Что, впрочем, не удивительно.

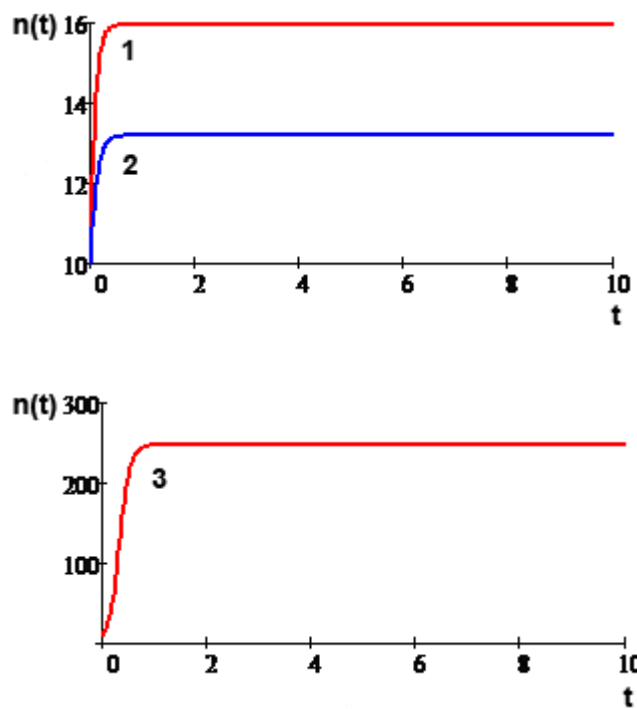


Рис. 13. Симбиоз основных сил при воздействии их на третью (случай позитивного воздействия)

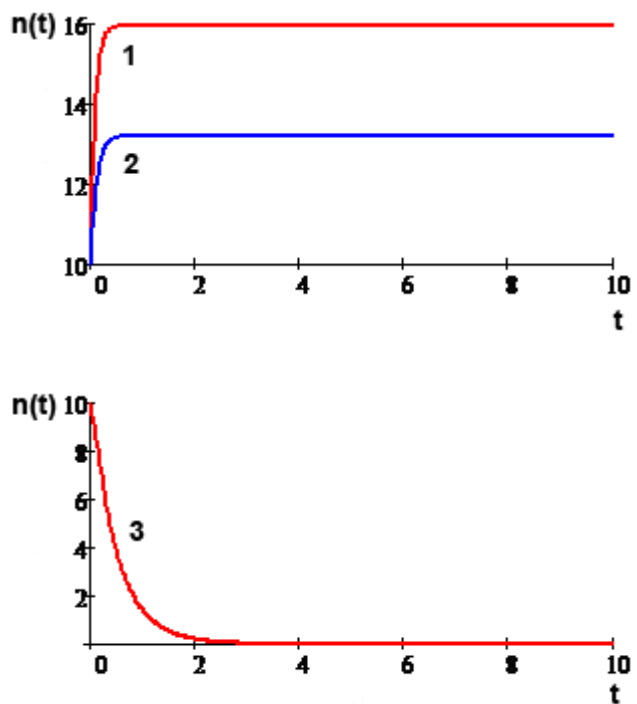


Рис. 14. Симбиоз основных сил при воздействии их на третью (случай негативного воздействия)

Отметим также, что ЭП, между которыми действуют отношения симбиоза, быстро и практически одновременно достигают своих равновесных состояний. Таким образом, этот вид взаимодействия ЭП можно назвать наиболее статичным.

Выше проанализировано ряд частных случаев динамики трех электоральных популяций. Разумеется, это далеко не все возможные варианты. Однако проиллюстрирован тот факт, что развитие электоральных процессов может полностью соответствовать интуитивным представлениям и оценкам исследователей, а при других условиях принимать неожиданные и, скажем, непривычные формы. При желании (или необходимости) каждый из рассмотренных случаев можно развивать, а также изучать другие случаи, обладающие другими характерными особенностями.

Кроме того, были показаны общие возможности математического моделирования социальных явлений. Даже простейшая модель позволила увидеть интересные и важные для понимания электоральных процессов аспекты их динамики. Полученные результаты дают основания утверждать, что такого рода модели способны правильно если не объяснять, то, по крайней мере, описывать сложные общественные процессы.

Особенностью математического моделирования электоральных процессов следует считать сравнительную простоту интерпретации получаемых результатов. Такие понятия как «численность электоральной популяции», «политический вес» и т. д. воспринимаются на интуитивном уровне даже без знакомства с точными (насколько они тут возможны) определениями. А это позволяет делать подобный анализ актуальных ситуаций предметом широкого обсуждения.

Мы не пытались получать точные, достоверные зависимости, описывающие те или иные конкретные. Для этого требуются численные значения всех параметров, а это, как уже говорилось, самостоятельная и отнюдь не простая задача. К тому же, в силу того, что некоторые решения являются неустойчивыми по отношению к своим параметрам, значения таких

параметров необходимо определять с высокой точностью. Для этого понадобился бы комплекс методик, основанных не только на обработке больших объемов статистических данных, но и на разносторонних социологических исследованиях. Разработка его потребует немало времени и усилий широкого круга специалистов.

В настоящее время более реалистичной выглядит несколько иная постановка задачи. Она состоит в использовании математических моделей электоральных процессов для прогнозирования возможных сценариев динамики ЭП на качественном уровне. В такой формулировке моделирование динамики ЭП занимает как бы промежуточный уровень между тем, что изложено здесь, и точным прогнозированием. И все же потребуются выбор значений параметров, которые бы в некотором разумном приближении соответствовали изучаемой ситуации, причем в большинстве случаев продуктивным оказывается использование относительных величин (скажем, ожидается, что значение параметра p для электоральной популяции А в полтора раза превысит его значение для электоральной популяции В; что произойдет?). Так, конечно, не получить достоверных данных о будущем развитии событий, но, скорее всего, можно составить более или менее адекватную картину того, что и как может произойти. А это уже не мало.

В любом случае не вызывает сомнений, что в той или иной форме математическое моделирование в недалеком будущем станет одним из стандартных средств аналитиков.

2.2. Индивидуум-ориентированные модели

В последнее время в связи с бурным развитием компьютерных технологий важным и перспективным с точки зрения применения на практике математического моделирования является класс т. н. имитационных моделей. Такая модель представляет собой алгоритм, при помощи которого компьютер генерирует наборы данных, описывающие заданные характеристики реальной

системы, представляющей интерес. При этом выполняемые машиной операции не имеют никакого отношения к природе и свойствам изучаемой системы. Отметим, что сам по себе факт выяснения возможности имитационного моделирования является немалым достижением современной науки. Действительно, оказывается, что структура реального процесса в известной мере не зависит от его природы и, так сказать, материальной основы. Числа, получаемые в результате манипулирования другими числами по определенным абстрактным правилам, могут в точности соответствовать числам, описывающим конкретные процессы, происходящие в нашем мире.

Разумеется, при разработке имитационной модели принимаются в расчет свойства исследуемого явления, но на уровне не внутренних механизмов, которые либо не известны, либо слишком сложны для явного использования, а общих характеристик протекания соответствующих процессов.

В плане практического применения имитационные модели хороши тем, что позволяют проводить т. н. машинные эксперименты, целью которых является изучение изменения поведения объекта исследования в зависимости от изменений внутренних параметров или (и) внешних условий. Такие методики дают возможность определять ход развития событий, которые по тем или иным причинам невозможно реализовать в реальной жизни. Например, как будут распределяться потоки беженцев в случае прорыва Киевской ГЭС. Ясно, что для выяснения этого не возможно осуществить подобную катастрофу в действительности. Имитационное же моделирование (при наличии удовлетворительных моделей) позволяет получить данные на вполне приемлемом уровне точности.

Построение имитационных моделей представляет собой достаточно сложную задачу, требующую помимо знания предметной области, еще и высокого профессионализма в сфере программирования. Однако в случае успеха результаты окупают издержки.

Индивидуум-ориентированный подход в моделировании предполагает создание имитационных моделей, учитывающих некоторые свойства

отдельных индивидуумов и их локального взаимодействия для построения интегральных моделей целых популяций, сформированных из множества индивидуумов.

Индивидуум-ориентированное моделирование (*individual-based modeling*) представляет собой отдельное направление в теории сложных систем, начало которому положено в работах Дж. Форрестера [45]. Развитию этого вида моделирования способствовало появление первых компьютеров и их использование в популяционных исследованиях.

Эти модели выделяют характерные черты поведения реальных индивидуумов (биологических организмов или участников социума) и используют их в качестве правил их локального взаимодействия.

Индивид в рамках этих моделей рассматривается как уникальная, дискретная единица, у которой есть некоторый набор характеристик, изменяющихся в течение жизненного цикла. Модели, основанные на данном подходе, строят снизу вверх, начиная с “частей” системы (индивидов), описывая в итоге всю популяцию. Целью исследования часто становится понимание того, каким образом свойства системы возникают из взаимодействия между частями [67].

Индивидуум-ориентированное моделирование снабжает исследователей инструментами, для решения задач, не поддающихся рассмотрению традиционными методами.

Первые популяционные исследования касались исключительно конкретных задач из биологической сферы и не носили системного характера. Однако, как будет показано ниже, они вполне логично переносятся на область изучения электоральных полей.

Безусловно, построение интегральной модели популяции на основе приблизительного описания правил поведения отдельного индивидуума может оказаться весьма далеким от реальности, однако в данном случае многое зависит от уровня описания этих правил, свойств отдельных индивидуумов и предполагаемой динамики популяции. Вместе с тем

индивидуум-ориентированное моделирование предоставляет ряд таких преимуществ, как простота описания отдельных особей и их локального взаимодействия, возможность детализации этих описаний в ходе моделирования, а также прозрачность обратной связи «правила – модель – реальность».

Индивидуум-ориентированное моделирование дает возможность описывать следующие свойства моделируемого объекта (особенно важные для моделирования электоральных полей):

- учет пространственных аспектов;
- учет воздействия материальных средств и влияния СМИ;
- учет социальных аспектов и индивидуальных особенностей;
- учет индивидуальных особенностей.

Учет пространственных аспектов. Индивидуум-ориентированное моделирование охватывает пространственно-распределенные модели (spatially-explicit mobile), в которых каждый индивидуум ассоциирован с определенным положением в пространстве.

Для полноценного моделирования электоральных полей, как правило, желательна пространственная распределенность, учет перемещений в пространстве. Например, при моделировании возможных фальсификаций выборов по открепительным талонам.

Учет воздействия материальных средств и влияния СМИ. Как известно, электоральные процессы существенно зависят от материальных средств и влияния средств массовой информации, используемых во время избирательных кампаний.

Модели, учитывающие информационные потоки называются информационными. В этих моделях описание изменения состояния индивидуумов базируется на правилах, основанных на анализе информационных потоков (в том числе, например, и состояние соседних индивидуумов).

Учет социальных аспектов и индивидуальных особенностей. Для моделирования электоральных процессов особенно важно писание на уровне отдельных особей.

Для электоральных процессов социальные механизмы играют очень важную роль, поэтому электоральные поля – это типичные социальные сети. Если модель направлена на исследование социальных механизмов, требует учета индивидуальных различий и обучения особей, необходимо выбирать данный подкласс моделей.

Как известно, индивидуальная изменчивость - основополагающий принцип эволюции. Вместе с тем учет этого фактора зачастую существенно усложняет модели, поэтому, не учитываются, в частности, в рассматриваемом ниже методе клеточных автоматов.

Уровень детализации. Свойства модели существенно зависят от ее пространственно-временного масштаба. Модели также различаются по количеству рассматриваемых индивидуумов. От масштаба задачи напрямую зависит объем вычислений. Этот факт приходится учитывать при выборе масштаба модели и ее реализации.

Следует отметить, что индивидуум-ориентированные модели требуют большего объема вычислений, чем аналитические модели. Вместе с тем для многих областей, в том числе и исследования электоральных полей, разработка индивидуум-ориентированной модели оправдана, в связи с тем, что:

- данных реальных наблюдений исследуемых параметров зачастую не хватает для идентификации аналитической модели;
- необходим учет пространственных аспектов;
- необходим учет социальных механизмов популяции, индивидуальных различий особи, обучения.

Дальнейшее описание посвящено модели распределения электоральных полей, построенной на основе концепции клеточных автоматов, впервые предложенной более полувека тому назад Дж. фон Нейманом [32] и развитой С. Вольфрамом в фундаментальной монографии [86].

Клеточные автоматы являются полезными дискретными моделями для разделов теории динамических систем, которые изучают характерные коллективные явления.

Дискретность модели, точнее, возможность представить модель в дискретной форме, в настоящее время относится к существенным преимуществам, поскольку открывает широкие возможности использования компьютерных технологий. Однако клеточные автоматы в этом смысле занимают особое место, поскольку в них дискретность сочетается с другим важным свойством, которое следует обсудить отдельно.

Назовем процесс пространственно распределенным, если он охватывает некоторую конечную область пространства, и каждая точка этой области обладает собственным характером изменения значений заданного набора параметров. Примером может служить заселение больших территорий осваиваемых земель: в одном районе возникает несколько быстро растущих поселков, в другом – один, не обнаруживающий тенденции к росту, а третий - остается нетронутым.

Здесь возможны два варианта. В первом - характер и динамика локальных изменений значений параметров определяется локальными свойствами соответствующего пространства (например, эффективность заселения зависит от того, насколько тот или иной район пригоден для жизни). Во втором - эти изменения порождаются неоднородностями в структуре самого процесса (в одном отряде переселенцев преобладают исполненные энтузиазма энергичные оптимисты, а в другом – исполненные сомнений пессимисты, причем соотношение тех и других меняется во времени в результате влияния друг на друга). Конечно, в чистом виде эти варианты встречаются крайне редко. Обычно наблюдается сочетание того и другого, но в большинстве случаев присутствует более или менее выраженное преобладание одного из них, что дает основание для такого рода классификации.

Особый интерес представляет частный случай второго варианта, в котором динамика дискретной системы носит итерационный характер. Иными

словами, ее эволюция может быть представлена как дискретная последовательность шагов, каждый из которых меняет состояние некоего подмножества элементов системы, причем изменения на каждом шаге определяются изменениями на предыдущем. Простейшим примером такого процесса может служить партия в шашки. После каждого хода две клетки на доске меняют свое состояние (взятие нескольких шашек рассматривается как последовательность ходов одного игрока). Состояние клетки здесь определяется одним параметром, который может принимать значения, соответствующие следующим ситуациям: клетка пустая, на клетке находится черная или белая шашка, которая может или не может быть взята и т. д.. Если на доске нет ни одной дамки, то наш процесс приобретет еще одну очень важную особенность. При каждом ходе изменение состояния фиксированной клетки определяется двумя факторами: состоянием соседних клеток и правилами игры. Под принятым нами углом зрения суть этих правил состоит в том, что они указывают, как именно состояние клетки изменяется в зависимости от состояний соседних клеток.

Этот простой пример достаточно точно иллюстрирует отнюдь не простые идеи, лежащие в основе клеточных автоматов. Они как нельзя лучше подходят для описания широкого класса процессов, допускающих представление, соответствующее приведенным выше принципам. И далеко не всегда речь идет об истинно дискретных системах. Часто непрерывные по своей природе процессы вполне приемлемо аппроксимируются подходящими дискретными конструкциями. Однако в этом случае модель, как правило, должна содержать большое (действительно большое) число дискретных элементов.

Главным достоинством клеточных автоматов является их абсолютная совместимость с алгоритмическими методами решения задач. Конечный набор формальных правил, заданный на конечном счетном множестве элементов (клеток), допускает точную реализацию в виде алгоритма. Однако отсюда вытекает и главный недостаток клеточных автоматов: вычислительные трудности, возникающие при соответствующих расчетах. Ведь на каждой

итерации необходимо сканировать весь набор клеток и для каждой из них выполнять требуемые операции. Когда и клеток, и итераций действительно много, это требует значительных ресурсов, в том числе и временных.

Поэтому долгое время клеточные автоматы воспринимались в основном как забавная, хотя и поучительная игра, не имеющая практической ценности. Но в последние годы, в связи с бурным развитием компьютерных технологий, они начинают быстро входить в арсенал инструментальных средств, используемых на практике в самых разнообразных областях науки и техники.

С помощью приведенного примера мы проиллюстрировали идею двумерных клеточных автоматов, хотя они могут быть многомерными и обладать нетривиальной топологией.

Клеточные автоматы, по своей сути, являются пространственно-немобильными дискретными индивидуально-ориентированными моделями. В традиционной системе клеточных автоматов все ячейки равноправны (пространство однородно), в то время как в индивидуум-ориентированной – помимо описания ячеек существует понятие индивидуума, который может занимать разные ячейки (и несколько различных индивидуумов могут занимать одну ячейку). Таким образом, в клеточном автомате ячейки меняют своё состояние синхронно, и цикл моделирования представляет собой перебор ячеек. В индивидуальных моделях цикл может состоять из перебора индивидуумов. Т.е. в клеточном автомате моделирование основано на разбиении пространства на однородные участки, в индивидуум-ориентированных моделях описываются сущности, которые меняют положение в пространстве. Конечно, ячейки в клеточном автомате могут находиться в различных состояниях, и с помощью определения сложных состояний можно моделировать наличие особей в ячейках и их перемещение между ячейками. Но это возможно лишь при существенных ограничениях.

Клеточный автомат представляет собой дискретную динамическую систему, совокупность одинаковых клеток, определенным образом соединенных между собой. Все клетки образуют сеть (решетку) клеточных

автоматов. Состояние каждой клетки определяется состоянием клеток, входящих в ее локальную окрестность и называемых ближайшими соседями [85]. Окрестностью конечного автомата с номером j называется множество его ближайших соседей. Состояние j -го клеточного автомата в момент времени $t + 1$, таким образом, определяется следующим образом:

$$y_j(t+1) = F(y_j(t), O(j), t),$$

где F – некоторое правило, которое можно выразить, например, языком булевой алгебры. Во многих задачах, считается, что сам элемент относится к своим ближайшим соседям, т.е. $y_j \in O(j)$, в этом случае формула упрощается: $y_j(t+1) = F(O(j), t)$. Клеточные автоматы в традиционном понимании удовлетворяют таким правилам:

- изменение значений всех клеток происходит одновременно (единица измерения – такт);
- сеть клеточных автоматов является однородной, т.е. правила изменения состояний для всех клеток одинаковы;
- на клетку могут повлиять лишь клетки из ее локальной окрестности;
- множество состояний клетки конечно.

Теоретически клеточные автоматы могут иметь любую размерность, однако чаще всего рассматривают одномерные и двумерные системы клеточных автоматов.

Предлагаемая модель является двумерной, поэтому дальнейший формализм будет относиться к этому случаю. В двумерном клеточном автомате решетка реализуется двумерным массивом. Поэтому в этом случае удобно перейти к двум индексам, что вполне корректно для конечных решеток.

В случае двумерной решетки, элементами которых являются квадраты, ближайшими соседями, входящими в окрестность элемента y_{ij} можно считать или только элементы, расположенные вверх-вниз и влево-вправо от него (т.н. окрестность фон Неймана: $y_{i-1,j}, y_{i,j-1}, y_{ij}, y_{i,j+1}, y_{i+1,j}$), либо добавленные к ним еще и диагональные элементы (окрестность Мура: $y_{i-1,j-1}, y_{i-1,j}, y_{i-1,j+1}, y_{i,j-1}, y_{ij},$

$y_{i,j+1}, y_{i+1,j-1}, y_{i+1,j}, y_{i+1,j+1}$). В модели Мура каждая клетка имеет восемь соседей. Для устранения краевых эффектов решетка топологически «сворачивается в тор» (рис. 15), т.е. первая строка считается продолжением последней, а последняя – предшествующей первой. То же самое относится и к столбцам.

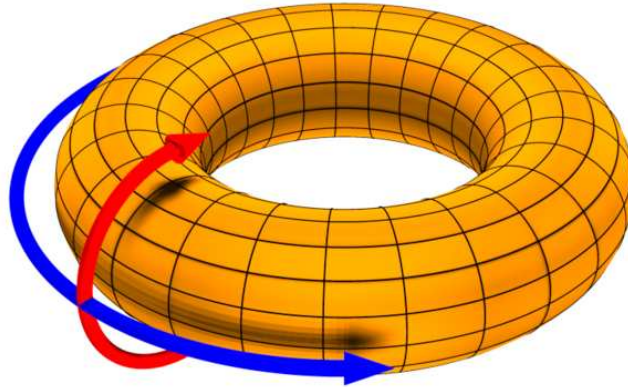


Рис. 15. Сворачивание плоскости в тор. Источник: [wikimedia.org](https://commons.wikimedia.org/wiki/File:Torus.jpg)

Это позволяет определять общее соотношение значения клетки на шаге $t + 1$ по сравнению с шагом t [29, 34, 35]:

$$y_{i,j}(t) = F(y_{i-1,j-1}(t), y_{i-1,j}(t), y_{i-1,j+1}(t), y_{i,j-1}(t), y_{i,j}(t), y_{i,j+1}(t), y_{i+1,j-1}(t), y_{i+1,j}(t), y_{i+1,j+1}(t)).$$

С. Вольфрам, классифицируя различные клеточные автоматы [79], выделил те, динамика которых существенно зависит от начального состояния. Подбирая различные начальные состояния, можно получать самые разнообразные конфигурации и типы поведения. Именно к таким системам относится классический пример - игра "Жизнь", изобретенная Дж. Конвеем и известная широкому кругу читателей благодаря публикации в книге М. Гарднера [9].

Некоторые примеры клеточных автоматов, применяемых в задачах социологии, приведены в [36, 37]. В частности, описывается модель процесса расовой сегрегации при выборе места жительства [62]. В рассматриваемом примере предполагается, что каждая расовая группа предпочитает иметь определенный процент соседей с тем же цветом кожи. Если это условие не выполняется, то семья перебирается в ближайший дом, где процентный состав соседей является приемлемым. В [62] использовалась модель конечных автоматов с простыми правилами и окрестностью

Мура. Построенная модель вполне реалистично описала процесс разделения региона на несколько расово-однородных областей.

Клеточные автоматы с успехом применяются и при моделировании процессов распространения новостей, инноваций [58]. В статье Т. Брауна [60] рассматривается модель электорального процесса. Он считает (с чем вполне солидарны авторы), что избирательные предпочтения индивида определяются установками его ближайшего окружения. В одной из моделей предполагается, что индивид принимает решение голосовать в момент $t + 1$ за республиканцев или демократов в соответствии с правилом простого большинства. Учитываются взгляды индивида и четырех его ближайших соседей в момент t (окрестность фон Неймана). Модель исследовалась на большом временном горизонте - до 20 000 тактов. Оказалось, что партийная борьба приводит к очень сложным конфигурациям, существенно зависящим от исходного распределения.

2.2.1. Модель, основанная на влиянии окружения

Авторами рассматривалось обобщение модели Брауна на случай, когда учитываются взгляды индивида и восьми его ближайших соседей (окрестность Мура). При этом электорат делится не на 2, как у Брауна, а на 4 части, распределенных в разных пропорциях (рис. 16), например, нейтральный (40% - белые клетки) и симпатизирующий трем партиям с заданными распределениями (например, 25% - черные клетки, 20% - серые клетки и 15% - светло-серые клетки), т.е. клетки в рассматриваемой модели могут принимать 4 значения (что, очевидно, не ограничивает общности). Именно поведение нейтральной части электората принципиально отличает эту модель от других и позволяет приблизиться к реалиям избирательной кампании в условиях многопартийности. Вместе с тем данная модель, описывающая сложные социально-психологические явления, безусловно, является упрощенной, однако достаточно точно описывает динамику электоральных полей и позволяет делать вполне реалистичные прогнозы на качественном уровне. Используемый авторами механизм

моделирования социальной самоорганизации может рассматриваться как дополнение к традиционным моделям динамики сложных нелинейных систем.

Онлайн-вариант модели, разработанной авторами и размещенной по адресу <http://edu.infostream.ua/vyb1.html>, позволяет наблюдать за решеткой 40 x 40 клеток. На начальном этапе клетки случайным образом распределяются по решетке (рис. 17). На каждом следующем такте модели клетки перекрашиваются в цвет, соответствующий цвету большей части клеток из окрестности (включая ее саму), кроме одного случая - исключения. Если клетка цветная, то она не может перекрашиваться в белый цвет, а перекрашивается в цвет, соответствующий цвету большинства «окрашенных» соседей. Это исключение соответствует тому факту, что в реальной жизни безразличные к политическим процессам люди редко переубеждают симпатизирующих той или иной партии. Формально эти правила можно записать следующим образом:

$$y_{i,j}(t+1) = \begin{cases} \arg \max_{k=1,\dots,3} C(k, O(i, j), t), & y_{i,j} \neq 0; \\ \arg \max_{k=0,\dots,3} C(k, O(i, j), t), & y_{i,j} = 0. \end{cases}$$

Здесь $O(i, j)$ – окрестность клетки с индексами i, j ,

$C(k, O(i, j), t)$ – количество элементов со значением k в окрестности $O(i, j)$ в момент времени t .

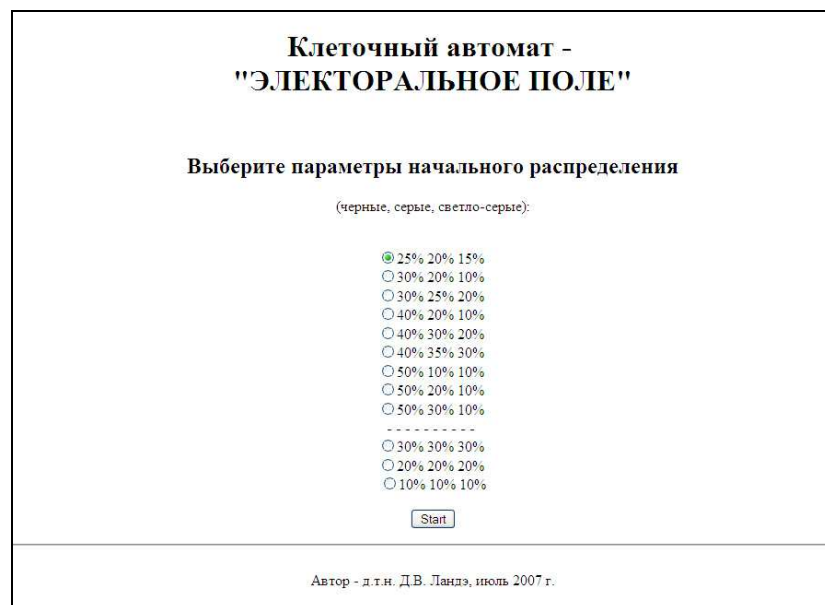


Рис. 16. Интерфейс модели «Электоральное поле»

Авторами были выполнены исследования модели, которые вполне может повторить читатель, которые свидетельствуют о том, что процесс достаточно быстро стабилизируется (10-40 тактов), принимая разнообразные конечные состояния (рис. 18).

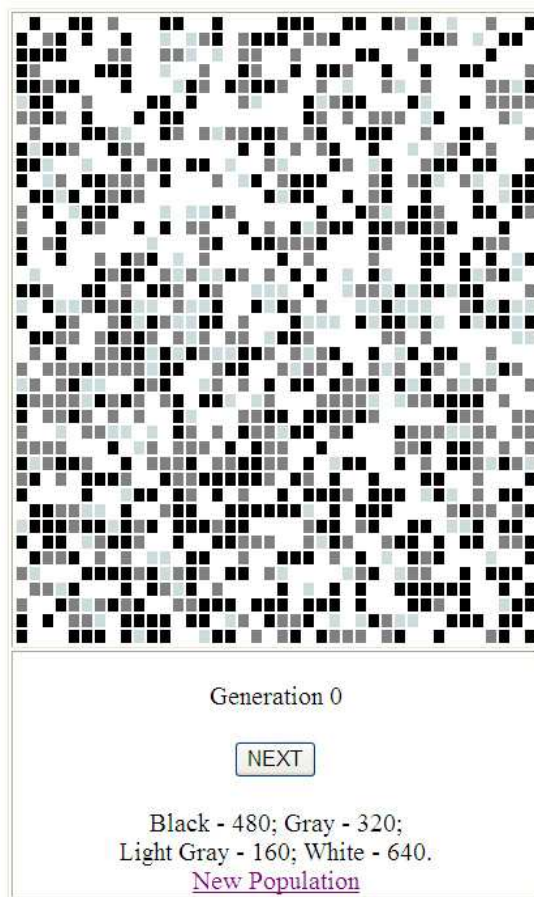


Рис.17. Вариант исходного состояния таблицы клеточных автоматов

На рис. 19 приведена динамика эволюционных предпочтений электората в рамках предложенной модели, которая позволила сделать некоторые выводы, оказавшиеся вполне реалистичными.

Островки электората, относящегося к малым партиям, чаще всего гибнут, оставаясь существовать лишь в двух случаях: когда их конфигурация стабильна (в нашем случае, образует, например квадрат со срезанными углами), либо когда они находятся в непосредственной близости к электорату других партий, которые взаимно компенсируют свое влияние.

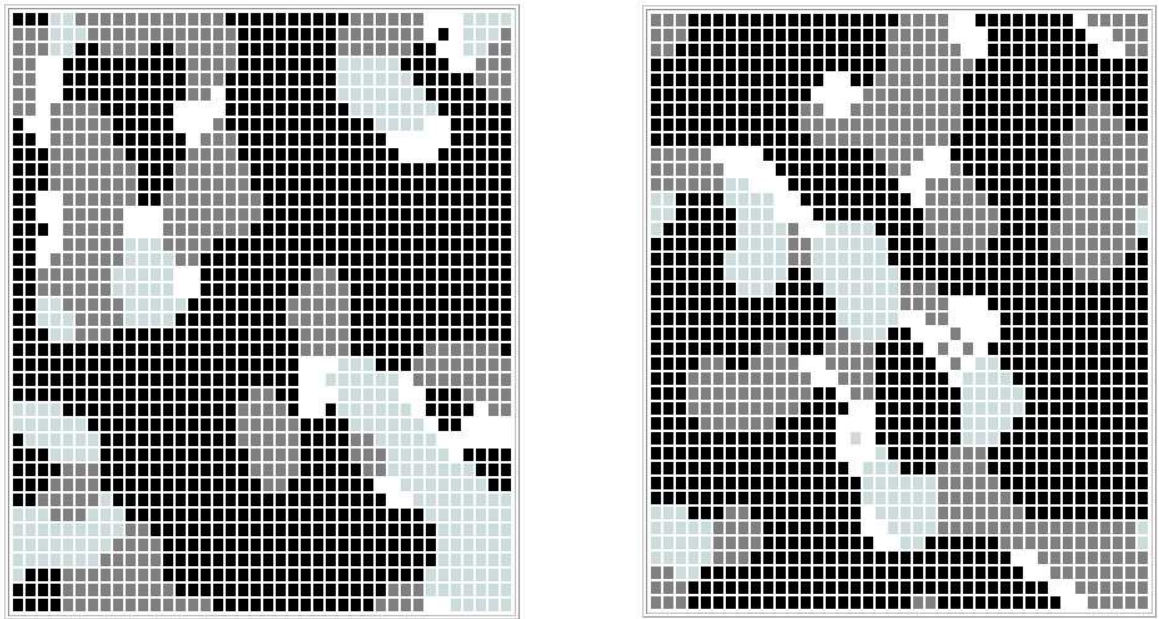


Рис. 18. Варианты конечного стабильного состояния клеточных автоматов

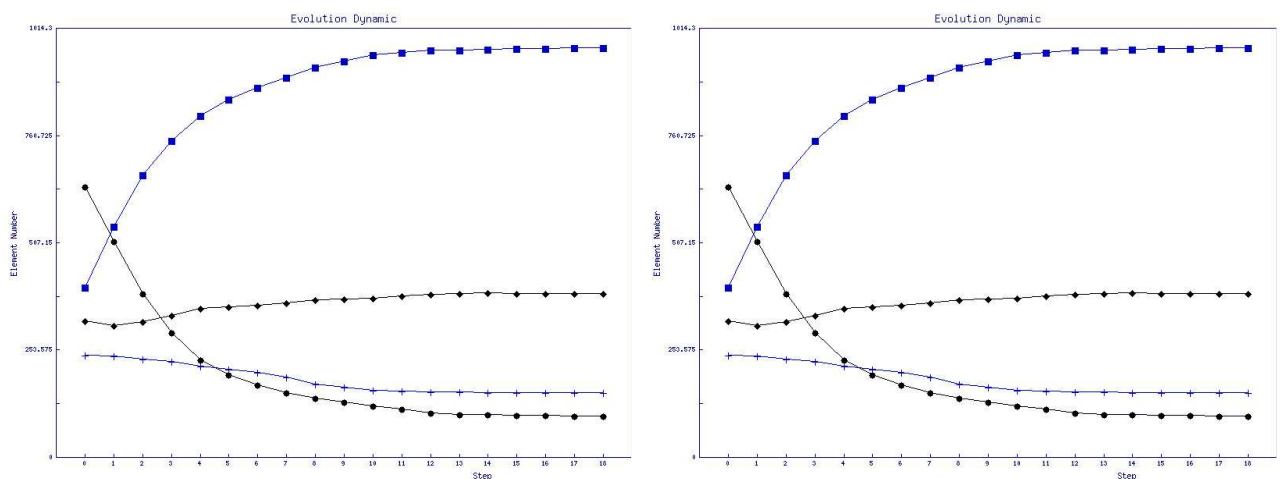


Рис. 19. Примеры динамики эволюции предпочтений электората (ось X – такты модели, ось Y – количество клеток, соответствующих электоратам):

■ - черные, ● - белые, ◆ - серые, + – светло-серые клетки

Рассмотренная модель позволила выявить некоторые общие свойства, которые вполне могут применяться во время оценки возможных результатов реальных избирательных кампаний:

- высокая сходимость - полная стабилизация происходит за 10 – 40 тактов;

- при стабилизации процент электората лидирующей партии возрастает с 25% до 55-65%;
- доля людей, симпатизирующих партии с минимальным электоратом, незначительно снижается до 5-8%;
- доля второй по числу электората партии остается стабильной;
- основной прирост сторонников лидирующей партии происходит за счет нейтральной части электората.

Покажем, как может быть осуществлен переход от приведенной базовой модели системы клеточных автоматов к системе обыкновенных дифференциальных уравнений.

Обозначим:

x – количество черных клеток в рассматриваемой системе;

y – количество серых клеток;

z – количество светло-серых клеток;

w – количество белых клеток;

t – время;

C_x, C_y, C_z, C_w – нормирующие константы;

N – количество клеток в системе клеточных автоматов.

Анализируя логику рассмотренной выше системы клеточных автоматов, можно предположить, что на скорость прироста числа черных клеток позитивно влияет количество черных и белых клеток. Количество серых и светло-серых клеток, по-видимому, отрицательно влияет на скорость прироста количества черных клеток. Высказанное утверждение можно записать в виде дифференциального уравнения:

$$\frac{dx}{dt} = c_x x - c_y y - c_z z + c_w w.$$

Аналогичные рассуждения можно привести для серых и светло-серых клеток. На скорость прироста количества серых клеток положительно влияет количество серых и белых клеток. Количество черных и светло-серых клеток, по-видимому, отрицательно влияет на скорость роста серых клеток.

Соответственно, на скорость прироста количества светло-серых клеток положительно влияет количество светло-серых и белых клеток и отрицательно – количество черных и серых клеток. Запишем высказанные предположения в виде еще двух дифференциальных уравнений системы:

$$\frac{dy}{dt} = c_y y - c_x x - c_z z + c_w w;$$

$$\frac{dz}{dt} = c_z z - c_x x - c_y y + c_w w.$$

Кроме того, справедливо условие баланса:

$$x + y + z + w = N.$$

Приведенные три дифференциальных уравнения и условия нормировки можно дискретизировать, сведя к системе из трех итерационных уравнений:

$$\begin{cases} x_{t+1} - x_t = c_x x_t - c_y y_t - c_z z_t + c_w (N - x_t - y_t - z_t); \\ y_{t+1} - y_t = c_y y_t - c_x x_t - c_z z_t + c_w (N - x_t - y_t - z_t); \\ z_{t+1} - z_t = c_z z_t - c_x x_t - c_y y_t + c_w (N - x_t - y_t - z_t). \end{cases}$$

Соответственно:

$$\begin{cases} x_{t+1} = (c_x + 1)x_t - c_y y_t - c_z z_t + c_w (N - x_t - y_t - z_t); \\ y_{t+1} = (c_y + 1)y_t - c_x x_t - c_z z_t + c_w (N - x_t - y_t - z_t); \\ z_{t+1} = (c_z + 1)z_t - c_x x_t - c_y y_t + c_w (N - x_t - y_t - z_t). \end{cases}$$

На рис. 20 представлены кривые, соответствующие динамике изменения значений x, y, z, w от времени t , полученные путем численного решения соответствующей системы уравнений итерационных уравнений с выбранными нормирующими константами $c_x = c_y = c_z = c_w = 0.15$ и $N = 1600$ (см. выше):

$$\begin{cases} x_{t+1} = 1.15x_t - 0.15(y_t + z_t) + 0.15(1600 - x_t - y_t - z_t); \\ y_{t+1} = 1.15y_t - 0.15(x_t + z_t) + 0.15(1600 - x_t - y_t - z_t); \\ z_{t+1} = 1.15z_t - 0.15(x_t + y_t) + 0.15(1600 - x_t - y_t - z_t). \end{cases}$$

Представленные зависимости вполне соответствуют кривым, представленным на рис. 19. Следует отметить, что количество «свободных» нормирующих констант, а также некоторая вольность допущений при формулировании дифференциальных уравнения значительно снижают доверие к модели, по сравнению даже с такими подходами, как индивидуум-ориентированное моделирование или его частный случай, представленный клеточными автоматами.

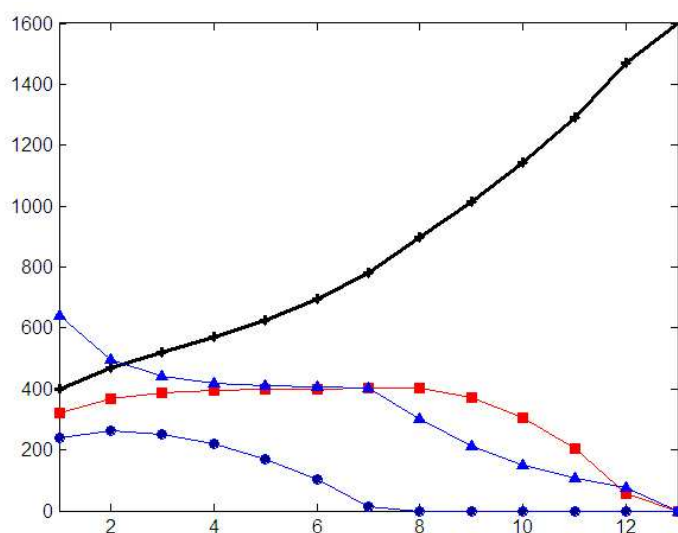


Рис. 20. Зависимости, полученные из динамической модели, соответствующей системе клеточных автоматов: $\times$ - черные клетки; $\blacksquare$ - серые клетки; $\bullet$ - светло-серые клетки; $\blacktriangle$ - белые клетки

2.2.2. Модель, учитывающая концепцию «малых миров»

Вероятно, нет нужды доказывать, что одним из важнейших факторов общественных процессов являются коммуникации между членами социума.

Часто при обсуждении вопросов информационного обмена в общественной системе вспоминают легенду о том, как изобретатель шахмат попросил себе в награду за первую клетку одно зерно, за вторую – два, за третью – четыре и т. д. Именно так происходит в реальности, если вы сообщите некую новость двум своим друзьям, каждый из них передаст ее

двум своим друзьям и т. д. Сколько потребуется итераций для того, чтобы все население страны оказалось проинформированным? Вопрос сам по себе не лишен интереса, хотя и имеет простое решение. Однако в реальном мире человеческих коммуникаций все не так просто. Для начала следует учесть, что те двое, кому вы передаете сообщение, могут его уже знать из альтернативного источника, а двое других граждан, соответственно, останутся неинформированными. В этом случае ваш канал выпадает из процесса.

Существуют и другие факторы, не столь очевидные, а в ряде случаев и просто нам неизвестные. Как бы там ни было, реальные пространственно распределенные процессы коммуникации в общественных системах обладают крайне сложным поведением и часто приводят к неожиданным конфигурациям, например образованию не связанных между собой устойчивых «островков».

Сказанное не исчерпывается одним только общением. Все это справедливо для любых сетевых межчеловеческих отношений, содержащих в себе пространственный аспект (в том или ином смысле). И, как нетрудно понять, имеет самое прямое отношение к электоральным процессам. В частности, в этой связи имеет смысл говорить о топологических свойствах электорального пространства, что может быть предметом самостоятельных исследований.

Следует отметить, что сеть отношений между людьми, которые составляют электорат, допускает аналогии на содержательном уровне с сетями, образуемыми, например, гиперссылками в Интернет или цитирования в науке. Поэтому сети, образуемые отношениями электората, с полным правом являются социальной сетью, исследование которой можно проводить, базируясь на существующем подходе анализа таких сетей - SNA (Social Network Analysis). Понятие «Социальная сеть» обозначает скопление социальных объектов, которые можно рассматривать как сеть (или граф), узлы которой - объекты, а связи - социальные отношения. В зависимости от

рода связей (ребер графа), они могут быть ненаправленными или направленными. Термин «Социальная сеть» был введен в 1954 году социологом из «Манчестерской школы» Дж. Барнсом в работе «Классы и собрания в норвежском островном приходе». Во второй половине XX века понятие «социальная сеть» стало очень популярным у западных исследователей, которые в качестве узлов социальных сетей стали рассматривать не только представителей социума, но и другие объекты, которым присущи социальные связи [38].

В ходе развития аппарата анализа социальных сетей развились такие направления, как теория социальных сетей и анализ социальных сетей (Social Network Analysis, SNA). Большинству видов социальных связей можно присвоить количественные характеристики, определяющиеся с помощью математического аппарата теории графов. Определим некоторые понятия SNA, которые понадобятся в дальнейшем изложении.

Путь между узлами - это последовательность вершин и ребер, соединяющих две вершины. Расстояние между узлами - количество шагов, которые нужно сделать, чтобы добраться от одной вершины к другой. Вершины в графе могут быть связаны напрямую или косвенно.

При анализе социальных сетей на базе теории графов выделяют:

- расчет параметров отдельных узлов;
- расчет параметров сети в целом;
- выделение сетевых подструктур.

В теории социальных сетей можно выделить три основных направления [19]:

- исследование статистических свойств, которые характеризуют поведение сетей;
- создание модели сетей;
- предсказание поведения сетей при изменении структурных свойств.

В прикладных исследованиях обычно применяют такие типичные для сетевого анализа характеристики, как размер сети, сетевая плотность, степень и плотность центральности и эквивалентность.

Для отдельных узлов выделяют следующие параметры:

- входная степень узла — количество ребер графа, входящих в узел;
- выходная степень узла — количество ребер графа, исходящих из узла;
- расстояние от данного узла до каждого из остальных;
- среднее расстояние от данного узла остальных;
- эксцентричность (eccentricity) — наибольшее из геодезических расстояний от данного узла до остальных;
- промежуточность — число путей, включающих данный узел;
- центральность (общее количество связей данного узла по отношению к остальным).

Общие параметры сети. Для расчета индексов для социальной сети в целом используют такие параметры:

- число узлов;
- число ребер;
- геодезическое расстояние между узлами (минимальное расстояние между узлами);
- среднее расстояние от одного узла до остальных;
- плотность (вычисляется как отношение количества ребер в сети к возможному максимальному количеству ребер с данным количеством узлов);
- количество симметричных, транзитивных и циклических триад;
- диаметр социальной сети (наибольшее геодезическое расстояние в социальной сети).

Задачи выявления сетевых подструктур. Существует несколько актуальных задач исследования социальных сетей [38]:

- определение клик в социальной сети. Клики - это подгруппы или кластеры, в которых узлы связаны между собой сильнее, чем с членами других клик;
- выделение компонентов (частей сети), которые связаны внутри и не связаны между собой;
- нахождение блоков и перемычек. Узел называется перемычкой (cutpoint), если при его удалении сеть распадается на несвязанные части;
- выделение группировок - групп эквивалентных узлов (которые имеют максимально похожие профили связей).

Коэффициент кластеризации. Д. Уаттс и С. Строгатц [84] в 1998 году определили такой параметр социальных сетей, как коэффициент кластеризации, который соответствует уровню связности узлов в сети.

Коэффициент кластеризации для отдельного узла сети определяется следующим образом. Пусть из узла исходит k ребер, которые, очевидно, соединяют его с k другими узлами, ближайшими соседями. Если предположить, что все ближайшие соседи соединены непосредственно друг с другом, то количество ребер между ними составляло бы $k(k-1)/2$. Т.е. это число, которое соответствует максимально возможному количеству ребер, которыми могли бы соединяться ближайшие соседи выбранного узла. Отношение реального количества ребер, соединяющих ближайших соседей к максимально возможному, и называется коэффициентом кластеризации узла. Естественно, это величина не превышает единицу.

Коэффициент кластеризации определяется как для каждого узла, так и для всей сети. Соответственно, уровень кластеризации сети определяется как нормированная по количеству узлов сумма соответствующих коэффициентов отдельных узлов. Рассматриваемый ниже феномен «малых миров» непосредственно связан с уровнем кластеризации.

Эластичности сети. Свойство эластичности сетей относится к распределению расстояний между узлами при удалении отдельных узлов.

Большинство сетей основано на их связности, то есть существовании путей между парами узлов. Если узел удален из сети, типичная длина этих путей увеличивается, и в результате пары узлов станут разъединенными [55].

Структура сообщества. О структуре сообщества [79] можно говорить тогда, когда существуют группы узлов, имеющие высокую плотность ребер между собой при том, что плотность ребер между отдельными группами – низкая. Традиционный метод для извлечения структуры сообществ – кластерный анализ. Существуют десятки приемлемых для этого методов, базирующихся на различных мерах расстояний между узлами, взвешенных путевых индексах между узлами и т.п. Для больших социальных сетей наличие структуры сообществ оказалось неотъемлемым свойством.

Модель слабых связей. Некоторые свойства социальных сетей не укладываются в рамки традиционной иерархической связи. К таким свойствам относятся и так называемые слабые связи. Аналогом слабых связей являются, например, отношения с дальними знакомыми и коллегами. В частности, слабые связи значительно расширяют масштабы привлекаемой информации. В некоторых случаях эти связи оказываются более эффективными, чем связи «сильные». Так в области мобильной связи группой учёных из Великобритании, США и Венгрии, был получен концептуальный вывод, что «слабые» социальные связи между индивидуумами оказываются самыми важными для существования социальной сети [75].

Во время исследования были проанализированы звонки 4,6 млн. абонентов мобильной связи, что составляло около 20% населения в одной европейской стране. Судя по всему, это первый случай в мировой практике, когда учёным удалось получить и проанализировать такую большую выборку данных о межличностных коммуникациях (в какой именно стране и какой оператор выдал приватную информацию — не сообщалось). Предыдущие аналогичные исследования на наземных телефонных линиях не

могли быть объективными, потому что домашний телефон используется разными людьми.

В социальной сети из 4,6 млн. было выявлено 7 млн. социальных связей, то есть взаимных звонков от одного абонента другому и обратно, если обратные звонки были сделаны в течение 18 недель. Частота и длительность разговоров использовалась для того, чтобы определить силу каждой социальной связи.

Именно слабые социальные связи (один-два обратных звонка в течение 18 недель) связывают воедино большую социальную сеть. Если эти связи убрать, то сеть распадется на отдельные фрагменты. Если же убрать сильные связи, то ничего страшного с сетью не произойдет — она останется единой (рис. 21).

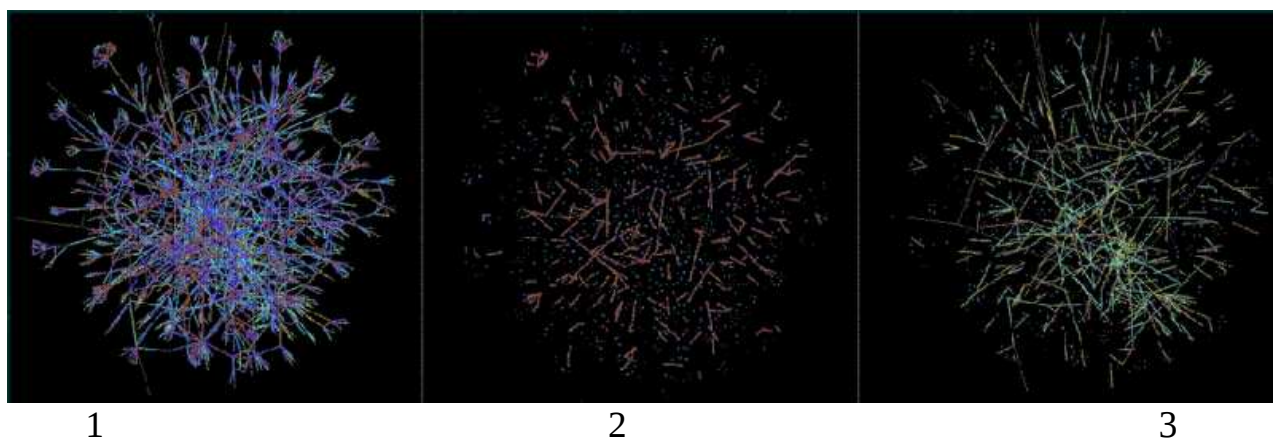


Рис. 21. Структура сети: 1) полная карта сети социальных коммуникаций; 2) социальная сеть, из которой удалены слабые связи (разбивается на множество изолированных участков); 3) карта, из которой удалены сильные связи: структура сохраняет сквозную проводимость

На основании проведенных исследований ученые сделали вывод, что именно слабые связи являются тем феноменом, который связывает большое общество в единое целое. Надо полагать, что данный вывод справедлив и для электоральных процессов, хотя академических исследований в этой области до сих пор не проводилось.

Малые миры. Несмотря на огромные размеры некоторых социальных сетей, в большинстве из них (и в WWW, в частности) существует сравнительно короткий путь между двумя любыми узлами – геодезическое расстояние. Еще в 1967 г. психолог Стенли Милгран [73] пришел к выводу, что существует цепочка знакомств, со средней длиной, равной шести, практически между двумя любыми гражданами США. Венгерскими математиками П. Эрдосом и А. Реньи [64] показано, что среднее расстояние между двумя вершинами в случайном графе растет как логарифм от числа вершин.

Эффект малых миров наглядно демонстрируется процедурой, представленной Д. Уаттсом и С. Строгатцом [84]. На рис. 22 представлены 3 состояния сети: регулярная сеть, каждый узел которой соединен с четырьмя соседними, та же сеть, у которой некоторые «ближние» связи случайным образом заменены «дальними» (именно в этом случае возникает феномен «малых миров») и случайная сеть, когда количество подобных замен превысило некий порог.

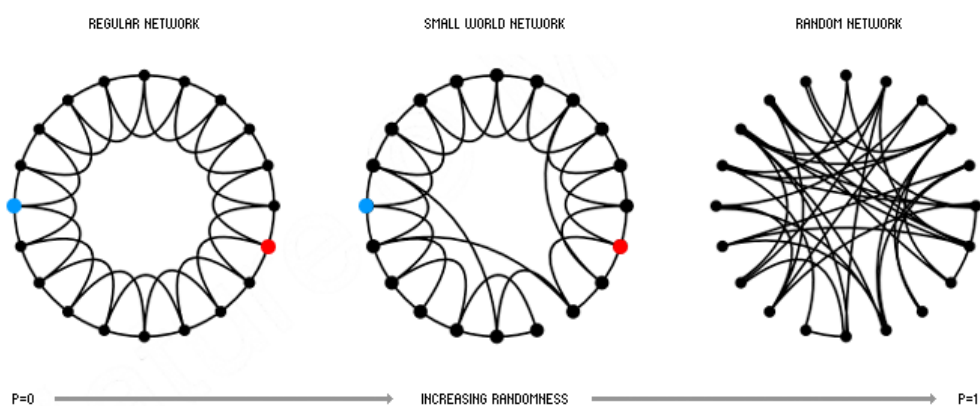


Рис. 22. Процедура моделирования «малых миров»

Как оказалось именно те сети, узлы которых имеют одновременно некоторое количество локальных и случайных «дальних» связей, демонстрируют и эффект малого мира, и большой уровень кластеризации [87]. Именно к таким сетям относится сеть человеческих отношений, для которой подтвержден феномен малых миров. В частности, с концепцией «малых миров» связан практический подход, именуемый «сетевой

мобилизацией», реализуемой над структурой «малых миров». В частности, скорость распространения информации благодаря эффекту «малых миров» возрастает на порядки, по сравнению с теоретически возможной, ведь большинство пар узлов сети соединены коротким путем.

Экспертами по безопасности эффект «малых миров» в последнее время все чаще связывается с сетями террористических организаций.

Теория перколяции. При изучении «малых миров» оказался интересен подход, логически связанный с понятием перколяции (протекания), популярном в современной физике [82, 43]. Оказывается, что многие вопросы, возникающие при анализе безопасности избирательных процессов, напрямую относятся к теории перколяции.

Перед теорией перколяции стоит множество вопросов, которые выходят за рамки дискретной математики и теории вероятностей. Самая простая, очищенная от всех физических и математических наслоений, формулировка задачи теории перколяции имеет следующий вид: «Дана решетка из связей, случайная часть которой проводит сигнал (воздух, ток, информацию ...), а остальная часть его не проводит. Основной вопрос – чему равна минимальная концентрация проводящих связей, при которой еще существует путь через всю решетку?».

В настоящее время известно много важных обобщений перколяционной задачи, например, рассматриваются случаи, когда «непроводящие» связи проводят, но много хуже проводящих; можно говорить о различных значения проводимостей для разных связей; можно рассматривать однонаправленные «диодные» связи (типичный пример – воздействие на избирателей средств массовой информации), и т.п.

К задачам, решаемым в рамках теории перколяции и анализа сетей [79] относятся такие, как определение порогового уровня проводимости (пропускной способности), изменения длины пути, и его траектории (извилистости, запараллеленности) при приближении к пороговому уровню

проводимости, количества узлов, которое необходимо вывести из строя, чтобы нарушить связность сети.

Одним из направлений анализа социальных сетей является визуализация соответствующих графов. Визуализация имеет важное значение, поскольку зачастую позволяет делать важные выводы о характере взаимодействия узлов, не прибегая к точным методам анализа. При отображении модели социальной сети целесообразным может быть:

- размещение узлов сети в двух измерениях;
- пространственное упорядочение объектов в одном измерении соответственно некоторому их количественному свойству, такому как, например, положение в организационной иерархии или иная мера социального статуса.
- использование общих для всех сетевых диаграмм методов для отображения количественных и качественных свойств объектов и отношений.

В качестве примеров визуализации сетей социальных рассмотрим некоторые разработки компании TouchGraph. Так, например, TouchGraph Amazon отображает сеть, порожденную книгами и связями между ними (тематиками, авторами, издательствами). Одним из самых динамичных новостных ресурсов Интернет сегодня можно считать и живые журналы (блоги). Компания TouchGraph, в частности, реализовала интерфейс для построения социосетей на основе Livejournal - TouchGraph LiveJournal Browser.

В случае визуализации WWW средствами TouchGraph Google Browser (<http://www.touchgraph.com/TGGoogleBrowser.html>) ребрами выступают не гиперссылки, а отношения подобия, которые в данной реализации не являются ненаправленными. Google Browser, представляющая собой Java-апплет, позволяющий визуализировать связи подобия между веб-сайтами, рассчитывающиеся в поисковой системе Google. В этом интерфейсе (рис. 23) можно увидеть все сайты, связанные отношением подобия с исходным

заданным, при этом пользователь может задавать глубину связей и отображать взаимосвязи различных сайтов. TouchGraph Google Browser весьма полезный инструмент также при поиске сайтов, связанных с исходной общей тематикой.

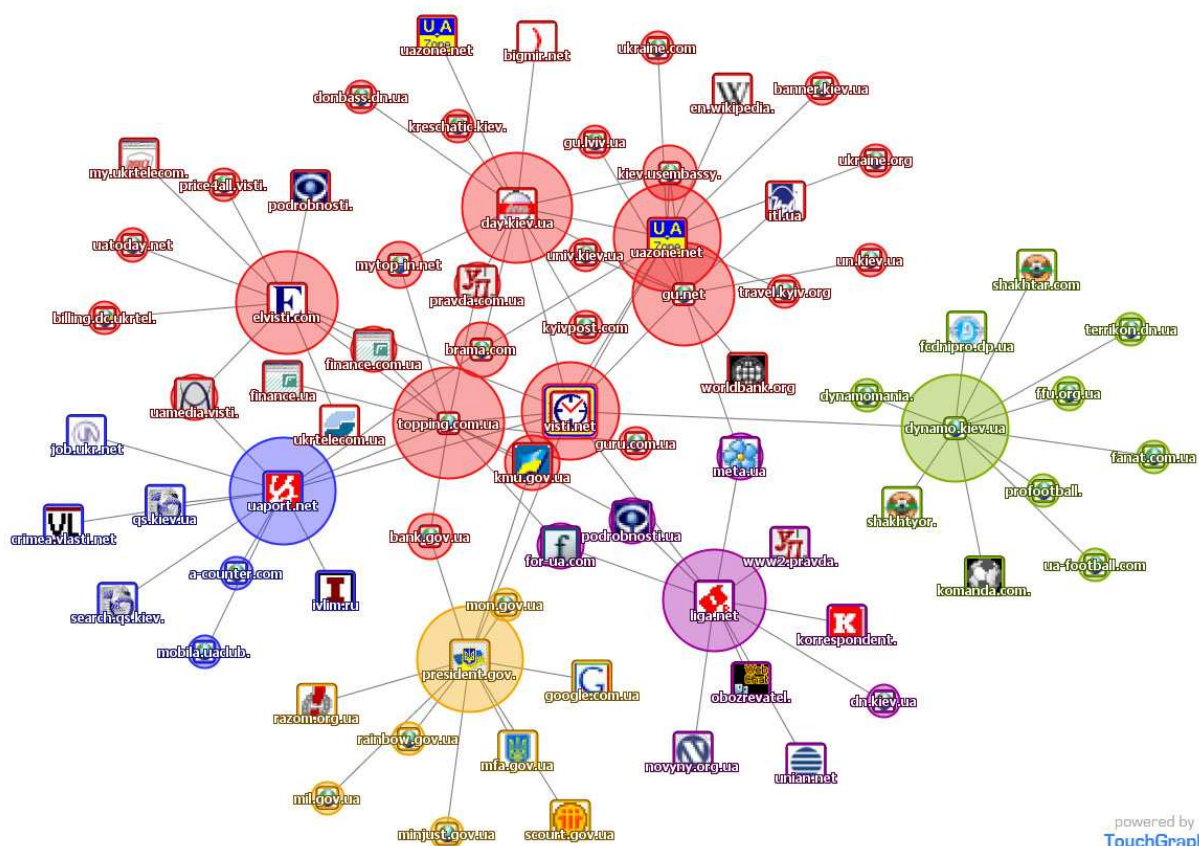


Рис. 23. Граф связи сайтов в Google Browser

В качестве еще одного инструмента для анализа и визуализации социальных сетей можно привести программу NetVis (<http://www.netvis.org>), использующую online-данные и импортированные csv файлы. Также широко известны программы визуализации и анализа социальных/организационных сетей InFlow (текущая версия 3.1 доступна по адресу <http://www.orgnet.com/inflow3.html>) и система анализа социальных сетей UCINET (<http://www.analytictech.com/ucinet/ucinet.htm>) с интегрированной в нее свободно распространяемой программой визуализации NetDraw.

В настоящее время ученые все еще далеки от понимания структуры сложных сетей и процессов, происходящих в них.

Пожалуй, лучшее отражение существующего положения с изучением топологии сложных социальных сетей, это модель «мятого Web», предложенная датчанином Л. Бйорнеборном [59]. «Мятый Web» в этой модели ассоциируется с мятой бумагой (рис. 24). При этом путь между выбранными точками на мятой бумаге зачастую короче, так как противоположные части листа бумаги соединены вместе. В соответствии с этой моделью каждая новая гиперссылка изменяет все существующие связи, создавая новые деформации «мятой» сети. То есть каждая новая гиперсвязь – «крючок», растягивающий или деформирующий форму существующей сети WWW.

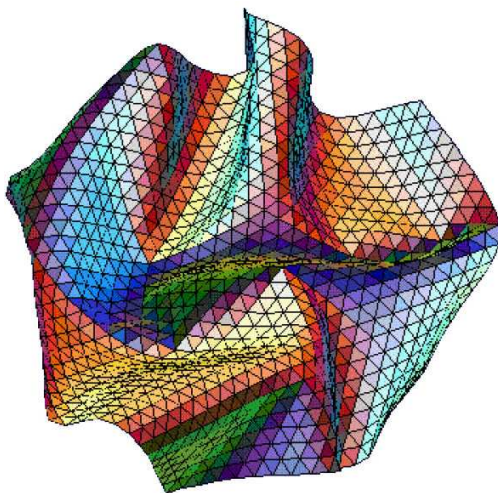


Рис.24. «Мятый Web»

Вместе с тем сегодня весьма успешно изучаются масштабируемые, статические, иерархические, самосинтезирующиеся, "малые миры" и др. сети, исследуются их фундаментальные свойства, такие, как устойчивость к деформациям и перколяции. В частности, недавно физики из университета Гранады [63] доказали, что наибольшую информационную проводимость имеет особый класс сетей, названных запутанными (entangled networks). Они характеризуются максимальной однородностью, минимальным расстоянием между любыми двумя узлами и очень узким спектром основных статистических параметров. Ученые-теоретики полагают, что запутанные сети могут найти широкое применение в области информационных и социальных технологий.

С учетом концепции «малых миров» и «дальних связей» авторами была усовершенствована клеточная модель, приведенная в предыдущем разделе.

Как и в описанном случае, электоральное поле вначале было разделено на 4 части, в пропорции 40% - нейтральные, 25% - черные; 20% - серые и 15% - светло-серые клетки. Как и ранее учитывались взгляды индивида и восьми его ближайших соседей (окрестность Мура). Кроме того, для каждой клетки учитывалось состояние одной или нескольких «дальних связей» (их количество выбиралось заранее при запуске модели, см. рис. 25).

**Клеточный автомат -
"ЭЛЕКТОРАЛЬНОЕ ПОЛЕ"**

Параметры начального распределения:
черные, серые, светло-серые: 25% - 20% - 15%

Выберите количество ДАЛЬНИХ СВЯЗЕЙ на клетку:

☒ 0
☐ 1
☐ 2

Автор - д.т.н. Д.В. Ландэ, август 2007 г.

Рис. 25. Выбор параметров усовершенствованной модели

Модифицированный указанным образом алгоритм описывает более реалистичную картину. Действительно, на отдельного человека оказывают влияние не только его близкие, друзья или сослуживцы. Очень часто существуют авторитетные для данного человека люди, удаленные, предположим, территориально.

Онлайн-вариант модернизированной модели, размещенный по адресу <http://edu.infostream.ua/vyb3.html>, также позволяет наблюдать за клеточной решеткой 40 x 40 клеток. Формально модифицированные правила можно записать следующим образом:

$$y_{i,j}(t+1) = \begin{cases} \arg \max_{k=1,\dots,3} C(k, O(i,j) \cup P(i,j), t), & y_{i,j} \neq 0; \\ \arg \max_{k=0,\dots,3} C(k, O(i,j) \cup P(i,j), t), & y_{i,j} = 0. \end{cases}$$

Здесь $O(i, j)$ – окрестность клетки с индексами i, j ,

$P(i, j)$ – клетки, соответствующие дальним связям для клетки i, j ,

$C(k, O(i, j) \cup P(i, j), t)$ – количество элементов со значением k в окрестности $O(i, j)$ в момент времени t .

Проведенные исследования модели показали, что в отличие от первой модели, скорость сходимости (приходу к стационарным состояниям), оказалась очень высокой (10-30 тактов), при этом конечные состояния свидетельствуют о значительном численном превосходстве электората, имеющего в начале незначительный перевес (рис. 26).

На рис. 27 приведена динамика предпочтений электората в рамках модифицированной модели.

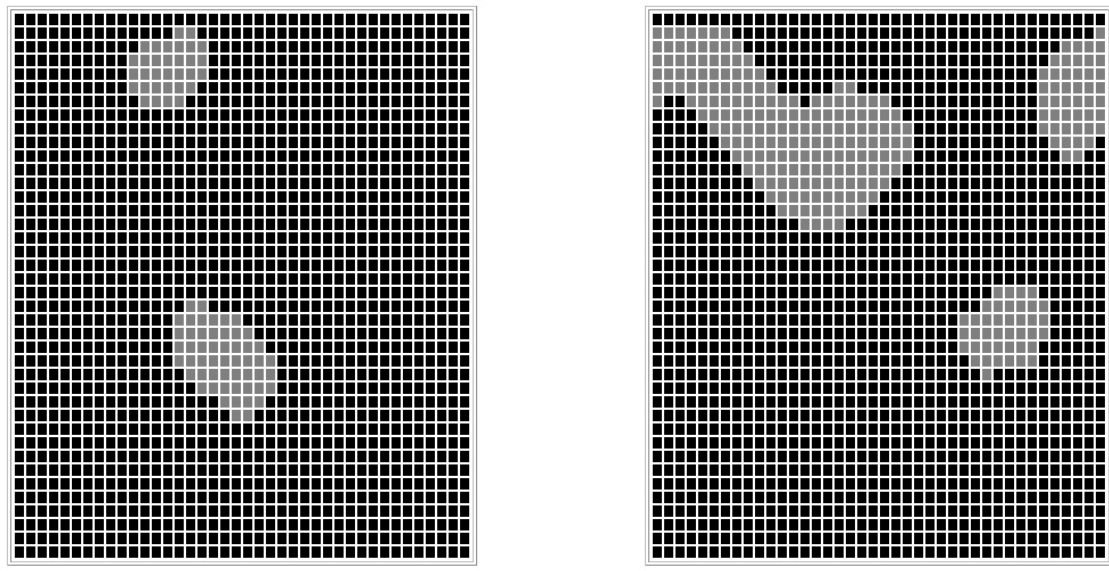


Рис. 26. Варианты конечного стабильного состояния модифицированной системы клеточных автоматов с «дальними связями»

Рассмотренная модель показывает, что электорат, относящийся к малым партиям, в конечном счете меняет предпочтения, причем значительно чаще, чем в базовой модели, представленной в предыдущем разделе. В данном случае основной прирост сторонников лидирующей партий происходит за счет электората партий-аутсайдеров.

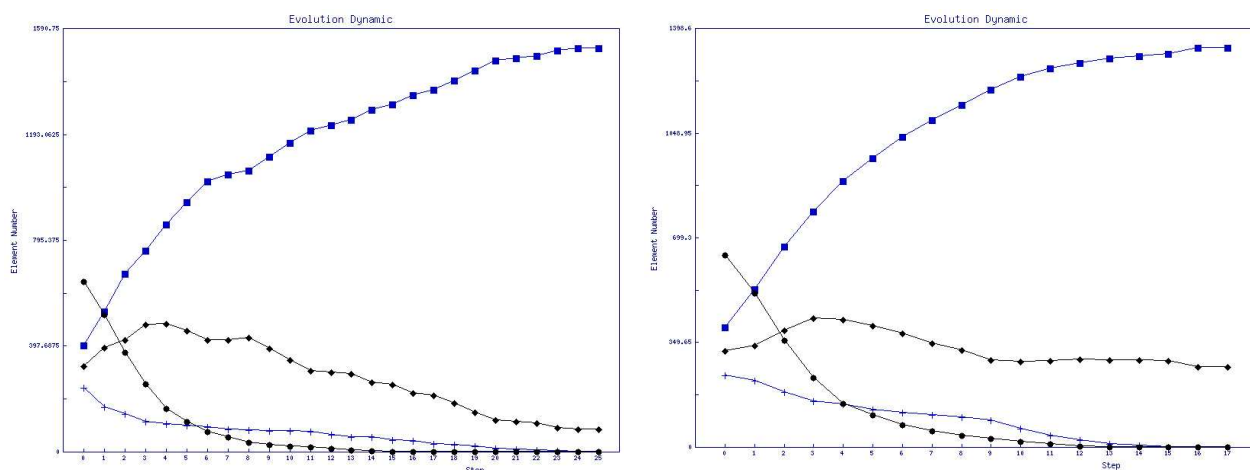


Рис.27. Примеры динамики эволюции электорального поля (ось X – такты модели, ось Y – количество клеток, соответствующих электоратам):
 ■ - черные, ● - белые, ◆ - серые, + – светло серые клетки

Вместе с тем жизненная практика показывает, что партии-аутсайдеры зачастую с успехом сохраняют и преумножают свой электорат. Возможно, эти процессы скорее всего можно связать с влиянием некоторых внешних факторов.

2.2.3. Модель, учитывающая влияние внешних факторов

Фактор внешнего влияния, который накладывается на рассматриваемую выше базовую клеточную модель, заключается в некотором изменении алгоритма потактового перекрашивания клеток. Т.е. предполагается, что под действием внешних факторов, например, в результате влияния политической рекламы в СМИ, мнение избирателя будет зависеть не только от количества соседей определенной электоральной ориентации, но и от «коэффициента усиления» их влияния. Формально процесс эволюции клетки при этом можно записать следующим образом:

$$y_{i,j}(t+1) = \begin{cases} \arg \max_{k=1, \dots, 3} C(k(1+v_k), O(i, j), t), & y_{i,j} \neq 0; \\ \arg \max_{k=0, \dots, 3} C(k(1+v_k), O(i, j), t), & y_{i,j} = 0. \end{cases}$$

Здесь $O(i, j)$ – окрестность клетки с индексами i, j ;

v_k - коэффициент внешнего влияния фактора k ;

$C(k(1+v_k), O(i, j), t)$ – количество элементов со значением k в окрестности $O(i, j)$ в момент времени t .

Авторами рассматривалась модифицированная базовая модель с внешнего влияния параметрами, выбираемыми при запуске процесса эволюции (онлайн-вариант размещен по адресу <http://edu.infostream.ua/vyb2.html>, рис. 28).

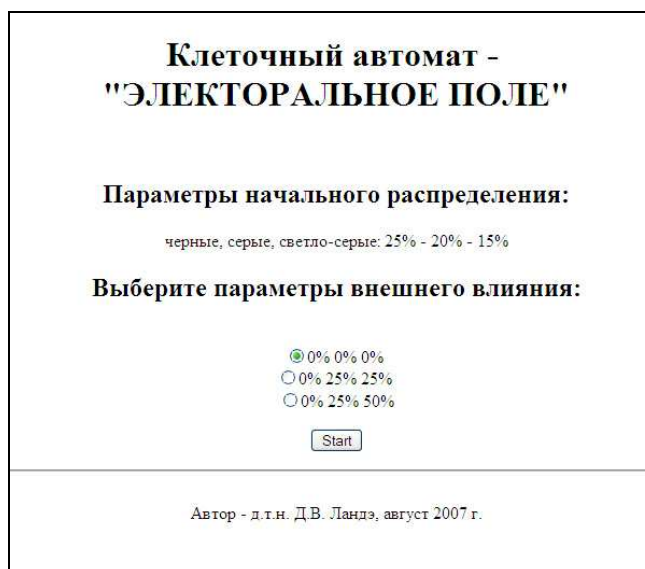
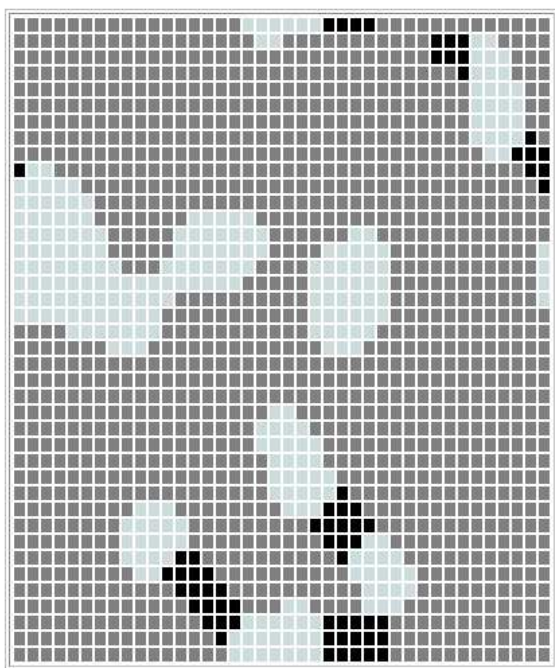


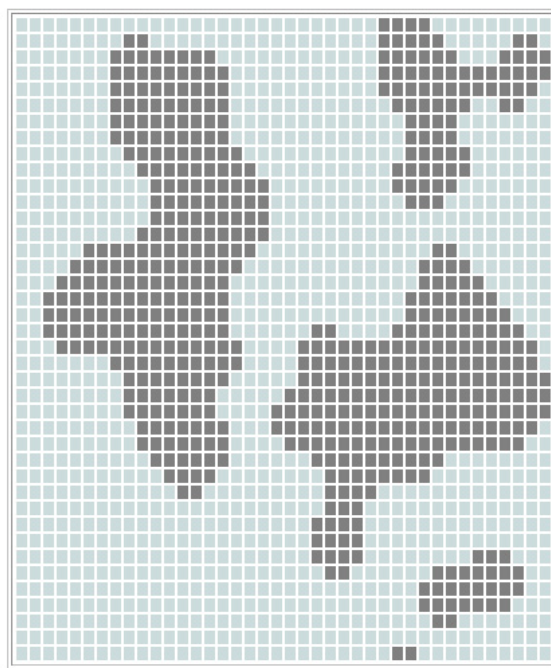
Рис. 28. Интерфейс системы клеточного автомата, учитывающего параметры внешнего влияния

Результаты моделирования, приведенные на рис. 29, свидетельствуют о решающем вкладе в динамику электоральных полей внешнего влияния. В этом случае блоки с наименьшей начальной численностью, но обладающие минимальным преимуществом во внешнем влиянии практически вытесняют лидирующие в начале электоральные силы. На рис. 30 показана соответствующая динамика изменения численности электората.

Приведенные выше модели систем клеточных автоматов обладают общим свойством, заключающемся в простоте правил «жизни» каждого отдельного автомата. Вместе с тем, множество клеток в совокупности показывают в результате эволюции совершенно нетривиальные результаты, которые на качественном уровне зачастую оказываются более реалистичными, чем традиционные аналитические модели.

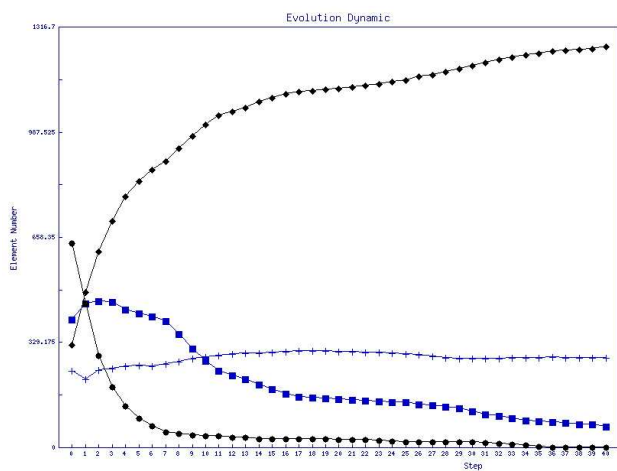


а)

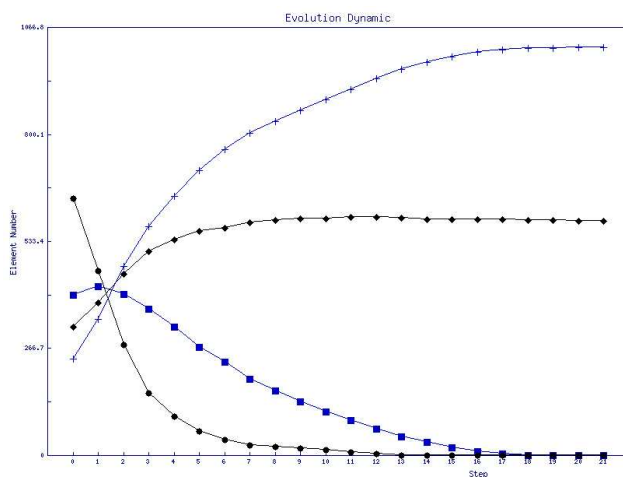


б)

Рис. 29. Итоговые «картинки» эволюции электоральных полей: а) – соотношение коэффициентов усиления: $0 - 0.25 - 0.25$; б) $0 - 0.25 - 0.50$



а)



б)

Рис. 30. Динамика эволюции электоральных полей: а) – соотношение коэффициентов усиления: $0 - 0.25 - 0.25$; б) $0 - 0.25 - 0.50$;

■ - черные, ● - белые, ◆ - серые, ⊕ – светло серые клетки

2.3. Другие подходы

2.3.1. Метод анализа иерархий

На наш взгляд, для анализа электоральных процессов необходимо снабдить аналитиков различными методами проведения экспертных оценок, нашедшими применение в других областях деятельности и подтвердивших свою эффективность. При этом надо отдавать себе отчет в том, что рассматриваемая задача усложняется значительными трудностями ее формализации и отсутствием общепризнанных методик обработки информационных потоков, что вносит существенные неточности в прогнозы и оценки ситуаций, возникающих в ходе избирательных кампаний.

Для решения задач подобного рода при прогнозировании и аналитическом планировании широко применяется метод анализа иерархий (МАИ), разработанный Т. Саати [40, 41]. Этот метод считается в наибольшей мере свободным от приведенных недостатков. МАИ включает метод парных сравнений, который, по мнению авторов, является основой для получения аналитических выводов в рассматриваемой области на основе экспертных оценок.

По определению Т. Саати, МАИ является систематической процедурой для иерархического представления элементов, определяющих суть любой проблемы. Метод состоит в ее декомпозиции на простые составляющие и дальнейшей обработке отдельных факторов по парным сравнениям. В результате может быть численно выражена относительная интенсивность взаимодействия элементов в иерархии.

В МАИ любая проблема предварительно структурируется и представляется в виде иерархии (на практике чаще всего сетевой). Таким образом, основная цель и все факторы, влияющие на достижение цели, распределяются по уровням иерархии в зависимости от интенсивности влияния.

МАИ широко применяется при анализе и прогнозировании рынков в экономике, а также при решении задач безопасности. В частности, известны работы [4, 5], в которых представлена методика оценки уровня военной безопасности для государства, разработаны методические рекомендации по прогнозированию сценариев развития военно-политической обстановки и установлению режимов функционирования элементов системы обеспечения военной безопасности государства.

Итак, ключевым моментом в МАИ является построение многоуровневой иерархии. То есть, если представить потенциал избирательного блока i в виде некоторой k -уровневой иерархии, то для оценки ее избирательного потенциала можно использовать метод анализа иерархий, включающих такие этапы [3, 41]:

Номер этапа	Содержание этапа	Примечание
1	Формулирование цели оценивания	Формируется цель оценивания и определяются сферы - оценки программ избирательного блока
2	Построение иерархии сфер и показателей оценивания	На первом уровне – цель оценивания (оценка избирательного потенциала); на втором - она связывается со сферами (политической, экономической, культурологической и др.); на третьем – отдельные показатели, относящиеся к выбранным сферам, например, отношение к двойному гражданству, нескольким государственным языкам, евроинтеграционным процессам и т.п.
3	Построение матрицы парных сравнений сфер	Разрабатываются бланки матриц парных сравнений.
4	Построение матрицы парных сравнений показателей	
5	Определение приоритетов сфер	При определении приоритетов основное внимание

Номер этапа	Содержание этапа	Примечание
6	Определение приоритетов показателей	концентрируется на сферах и показателях, имеющих наибольший вес.
7	Расчет текущего уровня избирательного потенциала	Рассчитывается итоговое нормированное значение уровня избирательного потенциала.
8	Сравнение текущего уровня избирательного потенциала с предполагаемым порогом реагирования	Полученный итоговый уровень избирательного потенциала блока i сравнивается с пороговым с целью проведения необходимых мероприятий при его превышении (в случае, если аналитик работает в интересах конкурирующего блока)

После определения цели оценивания, вторым этапом применения МАИ является структурирование проблемы в виде иерархии. В простейшем виде иерархия строится от цели, через промежуточные уровни-сферы к самому нижнему уровню, который в общем случае является набором показателей-альтернатив. На втором этапе выполняется процесс построения иерархии, который продолжается до тех пор, пока в неё не включены все основные факторы.

Допустимо практически любое количество уровней, однако в дальнейшем изложении, не ограничивая общности, остановимся на трехуровневой модели, первый уровень которой – цель – оценка программы избирательного блока, второй уровень – сферы – политические, экономические, социальные и др., третий уровень – отдельные показатели, например, отношение к двойному гражданству, нескольким государственным языкам, евроинтеграционным процессам и т.п.

После иерархического воспроизведения проблемы устанавливаются приоритеты критериев и оцениваются каждая сфера и показатель по критериям. В МАИ элементы задачи сравниваются попарно по отношению к их воздействию на общую для них характеристику. Система парных

сведений приводит к результату, который может быть представлен в виде обратносимметричной матрицы.

Выбор шкалы критериев определялся следующими требованиями:

- шкала должна давать возможность улавливать разницу в субъективных оценках, которые имеют люди;
- эксперт должен быть максимально уверенным во всех градациях своих суждений.

Для проведения парных сравнений в МАИ вводится соответствующая матрица $A = ||a_{ij}||$, элементом a_{ij} которой является интенсивность проявления элемента иерархии i относительно элемента иерархии j , оцениваемая по шкале от 1 до 9, предложенной Т. Саати для проведения субъективных парных сравнений. При этом критерии имеют следующий смысл:

Интенсивность важности	Определение	Пояснение
1	Равная важность	Равный вклад двух элементов
3	Незначительное превосходство одного над другим	Опыт и суждения свидетельствуют о легком превосходстве одного элемента над другим
5	Существенное превосходство	Опыт и суждения свидетельствуют о сильном превосходстве одного элемента над другим
7	Значительное превосходство	Одному из элементов дается настолько сильное превосходство, что он становится значимым
9	Очень сильное превосходство	Очевидность превосходства одного элемента над другим подтверждается сильно
2,4,6,8	Промежуточные решения	Применяются в компромиссных случаях

Интенсивность важности	Определение	Пояснение
Обратные величины приведенных выше чисел	Если при сравнении одного элемента с другим получено одно из вышеуказанных чисел, то при сравнении второго вида деятельности с первым записывается обратная величина	

Как показывают работы различных исследователей [3], шкала Т. Саати и ее незначительные модификации оказываются более эффективной по сравнению с другими рассмотренными шкалами. По окончании построения иерархии (этапов 3 и 4) для каждой опорной вершины соответствующего графа иерархии проводится оценка весовых коэффициентов, определяющих степень ее зависимости от влияющих на нее вершин более низкого уровня. При этом используется метод парных сравнений. Существует несколько вариантов метода парных сравнений. В рассматриваемой ниже модификации этого метода факторы сравниваются попарно по отношению к их воздействию (интенсивности или весу) на общую для них сферу или показатель.

Пусть в конкретной задаче необходимо определить состав некоторой избирательной программы. Причем пусть $A_1, A_2, \dots, A_n$ основные факторы, программные положения, относящиеся к разным сферам избирательной программы блока i . Тогда для определения структуры избирательной программы заполняется матрица парных сравнений:

	A_1	A_2	...	A_n
A_1	1	a_{12}		a_{1n}
A_2	a_{21}	1		a_{2n}
...			...	
A_n	a_{n1}	a_{n2}		1

Здесь a_{ij} – элементы матрицы парных сравнений, n – количество строк и столбцов матрицы. Если при сравнении одного фактора (сферы в случае 2-го уровня иерархии) i с другим j получено $a_{ij} = b$, то при сравнении второго фактора с первым получаем $a_{ji} = 1/b$.

При проведении парных сравнений при сравнении различных элементов ставятся следующие вопросы:

- какой из них важнее или имеет большее воздействие?
- какой из них более вероятен?
- какой из них предпочтительнее?

Если обозначить долю фактора A_i через w_i , то элемент матрицы $a_{ij} = w_i/w_j$.

Таким образом, в рассматриваемом варианте применения метода парных сравнений, определяются величины отношений значений факторов. При этом очевидно, что матрица является положительно определенной, обратносимметричной, имеющей ранг 1. Фактически основная задача МАИ состоит в отыскании вектора $(w_1, w_2, \dots, w_n)$. Применяя МАИ, эксперты, производя парное сравнение факторов $A_1, A_2, \dots, A_n$ заполняют таблицу парных сравнений. Очевидно, что когда $w_1, w_2, \dots, w_n$ неизвестны заранее, то парные сравнения производятся с использованием субъективных суждений, численно оцениваемых по шкале, аналогичной приведенной выше.

Эксперт, сравнивая n факторов, реально проводит $n(n-1)/2$ сравнений. Кроме того, ввиду справедливости соотношения $a_{ij} = a_{ik} a_{kj}$, справедливого для всех значений k , производится опосредованное сравнение факторов A_i и A_j через соответствующие сравнения этих факторов с фактором A_k . Учет этих дополнительных сравнений позволяет значительно повысить надежность получаемых результатов.

Относительная сила или вес отдельного объекта в иерархии определяется оценкой соответствующего ему элемента собственного вектора матрицы приоритетов, нормализованного к единице. Один из основных методов вычисления вектора w основывается на том утверждении линейной

алгебры, что искомый вектор является собственным вектором матрицы парных сравнений, соответствующим максимальному собственному числу (λ_{max}).

Процедура определения собственных векторов матриц поддается приближению с помощью вычисления геометрического среднего: компонента собственного вектора x_i в этом случае определяется следующим образом:

$$x_i = \sqrt[n]{\prod_{j=1}^n \frac{w_i}{w_j}}.$$

Из линейной алгебры известно, что у положительно определенной, обратносимметричной матрицы, имеющей ранг равный 1, максимальное собственное число равно размерности этой матрицы (т.е., n в рассматриваемом случае). Чаще всего вычисленное максимальное собственное число λ_{max} для матрицы составленной экспертом будет отличаться от соответствующего собственного числа для идеальной матрицы. Это различие характеризует так называемую рассогласованность реальной матрицы. И, соответственно, характеризует уровень доверия к полученным результатам. Чем больше это отличие, тем меньше доверие. Таким образом, эта модификация метода парных сравнений содержит внутренние инструменты позволяющие определить качество обрабатываемых данных и степень доверия к ним. При этом весьма полезным побочным продуктом теории является так называемый индекс согласованности (I), который дает информацию о степени нарушения согласованности и вычисляется по формуле:

$$I = (\lambda_{max} - n)/(n - 1).$$

Если такие отклонения превышают установленные пределы, то тому, кто проводит суждения, следует перепроверить их.

Если разделить I на число, соответствующее случайной согласованности матрицы того же порядка (в [41] приведены значения случайных согласованностей для матриц разных порядков), получим

отношение согласованности, величина которого должна быть порядка 0.1 или менее, чтобы быть приемлемой. В некоторых случаях допускается значение этого отношения до 0.2, но не более, иначе надо проверить свои суждения.

Как уже отмечалось, рассматриваемая версия метода парных сравнений, позволяет определить качество исходных данных. Причем Т. Саати рекомендует при плохо согласованной матрице либо сменить экспертов, найти дополнительные данные, либо решать проблему другим методом.

Еще один, более простой подход к определению вектора $(w_1, w_2, \dots, w_n)$ состоит в том, что суммируются по строкам элементы матрицы парных сравнений (для каждого значения i вычисляется сумма $a_i = a_{i1} + a_{i2} + \dots + a_{in}$). Затем все a_i нормируются так, чтобы их сумма была равна 1. В результате получаем искомый вектор $(w_1, w_2, \dots, w_n)$.

Таким образом, $w_i = a_i / (a_1 + a_2 + \dots + a_n)$.

Этот способ, значительно проще в реализации, но он не позволяет определять качество исходных данных.

Для синтеза приоритетов Саати рекомендует из группы матриц парных сравнений формировать набор локальных приоритетов, которые выражают относительное влияние множества элементов на элемент примыкающего сверху уровня. Затем определяется относительный вес каждого отдельного объекта путем решения систем уравнений, выраженных матрицами парных сравнений, каждая из которых обладает обратносимметричными свойствами. Для этого нужно вычислить множество собственных векторов для каждой матрицы, нормировать их к единице, получая тем самым вектор приоритетов.

В рассматриваемом случае вычисление глобальных приоритетов L_k , ($k = 1, \dots, n$) для сфер – 2-го уровня иерархии момент t_0 в выполняется в соответствии с известными формулами матричного исчисления:

$$L_k(t_0) = r_k / \sum_{j=1}^n r_j,$$

где

$$r_k = w_k / w_1 * w_k / w_2 * \dots * w_k / w_n.$$

После расчета приоритетов сфер L_k по аналогичному алгоритму выполняется процедура определения приоритетов показателей P_j , $j = 1, \dots, K$, где K – сумма размерностей подматриц парных сравнений, которые были построены для 3-го уровня иерархии.

Окончательное нормированное значение избирательного потенциала для отдельного избирательного блока в момент t_0 рассчитывается по формуле [41]:

$$A(t_0) = \sum_{j=1}^K (L_j P_j) / \sum_{j=1}^K \max(L_j P_j).$$

Данную методику можно применять и для других объектов, имеющих отношение к избирательным процессам. Например, в качестве факторов можно рассматривать избирательные блоки и строить матрицу их парных сравнений для различных регионов, что с учетом особенностей региональных электоральных полей, может оказаться хорошей основой для прогноза результатов избирательной кампании. При этом следует отметить, что полученная с помощью данной методики информация ввиду природы ее получения имеет лишь рекомендательный характер.

2.3.2. Теоретико-игровой подход

Характерной особенностью электоральных процессов является, в числе прочего, то обстоятельство, что их участники находятся в состоянии конфликта интересов, и при этом действуют в условиях отсутствия полной информации о намерениях друг друга. По крайней мере, это справедливо для блоков политических сил, внутри которых ситуация может быть иной. При изучении электоральных процессов практически всегда приходится анализировать конфликтные ситуации, в которых сталкиваются интересы двух или более конкурирующих сторон, преследующих различные цели. Математической теорией, которая посвящена изучению конфликтных

ситуаций, является теория игр. В ней под понятие игры подходит любая ситуация с рациональными, то есть целеполагающими, оптимизирующими субъектами (участниками), а также некоторые ситуации с неполной рациональностью [7, 23]. Таким образом, нам представляется вполне естественным попытаться применить к изучению электоральных процессов теорию игр. В обобщенной игре (в качестве игры, в нашем случае может рассматриваться ход избирательной кампании) могут сталкиваться интересы двух или нескольких противников. При этом игроки могут образовывать коалиции, в этом случае игра становится коалиционной.

Структура любой игры описывается тремя блоками:

- 1) допустимые множества ходов или стратегий участников;
- 2) цели участников;
- 3) тип поведения и информированности участников.

Анализ игры заключается в умении прогнозировать решение игры - множество возможных ходов и их результатов. Важными понятиями в теории игр являются также оптимальная стратегия, цена игры, средний выигрыш. В частности, стратегии P^* первого игрока и Q^* второго игрока называются оптимальными, а число V – ценой игры, если для любых стратегий P первого игрока и Q второго игрока выполняются неравенства [23]:

$$M(P, Q^*) \leq V \leq M(P^*, Q),$$

где $M(P, Q)$ – математическое ожидание выигрыша первого игрока, выбравшего стратегию P , при условии, что вторым выбрана стратегия Q .

Во многих задачах из теории игр неопределенность вызвана не противодействием противника, а недостаточной осведомленностью игрока об условиях, в которых действуют стороны, например, внешних воздействиях. Такие игры принято называть «играми с природой», при решении которых используют, так называемые, «матрицы рисков». Анализ матриц рисков осуществляется методами, близкими по идеологии к рассмотренным в предыдущем пункте.

Решая задачи в условиях неопределенности, когда вероятности отдельных частных исходов неизвестны, возникают трудности при математическом моделировании. В таких случаях теория принятия решений, в частности, рекомендует применять подход, базирующийся на известной теореме Байеса. Стратегия оптимизации в таких случаях строится на основе байесовской теории принятия решений. При этом принятая в теории игр функция потерь рассматривается как обобщение вероятности ошибки. Соответственно, предполагается выбирать решение, минимизирующее функцию потерь.

Байесовский подход к оценке вероятностных связей играют решающую роль в теории принятия решений в условиях неопределенности последствий этих решений или в условиях противодействия со стороны природы, или конкуренции. В этих условиях ключевой является стратегия управления, основанная на апостериорной (послеопытной) вероятности события. Обязательное условие корректности такого подхода является постоянное обучение системы. Стратегия управления в начале строится на базе определенных представлений о вероятности событий, а по мере функционирования системы реализуется коррекция управления - использования накапливаемого опыта путем перерасчета вариантов стратегий с учетом изменившихся значений вероятностей.

Нормальную форму игры часто соотносят со случаем статической или одновременной игры (однократные одновременные ходы участников), а развернутую форму - с динамическими играми (последовательные ходы), хотя мы увидим, что возможны и другие трактовки. Нормальная форма задает исходную физическую и целевую структура игры как объект

$$G := \langle I, X, u(\cdot) \rangle = \langle I, \{X_i\}_{i \in I}, \{u_i(\cdot)\}_{i \in I} \rangle,$$

где

$I := \{1, \dots, m\}$ - множество участников i ;

$X := (X_i)_{i \in I} := \prod_i X_i = (X_1 \times X_2 \times \dots \times X_m)$ - набор (профиль) допустимых множеств стратегий $(x_i)_{i \in I}$ участников;

$u := (u_i)_{i \in I}$ - набор (профиль) целевых функций участников (заметим: каждая целевая функция $u_i : X_i \rightarrow R$ зависит, вообще говоря, от всех $(x_i)_{i \in I}$).

Будем обозначать через $x_{-i} := (x_j)_{j \in I \setminus \{i\}}$ набор стратегий всех игроков кроме i , и аналогично индексировать множества и функции. Если игроки не обладают информацией ни о целях, ни о намеченных стратегиях партнеров и ведут себя "очень осторожно", то подходит концепция «максиминного» решения. Дадим формальное определение этой ситуации. Множество X_{MMi} осторожных или максиминных стратегий игрока i задается как аргументы, максимизирующие гарантированный выигрыш:

$$X_{MMi} := \{x_i \in X_i \mid \forall x_{-i} \Rightarrow u_i(x_i, x_{-i}) \geq \sup_{y_i \in X_i} (\inf_{z_{-i} \in X_{-i}} u_i(y_i, z_{-i}))\},$$

при этом $MM := \prod_{i \in I} X_{MMi}$ - множество максиминных решений игры.

В этом случае каждый максимизирует выигрыш при пессимистических ожиданиях, т.е. максимизирует гарантированный выигрыш. В антагонистической игре концепция максимина очень естественна. Но не все максиминные решения вызывают доверие как возможный результат повторяющейся игры.

Следует отметить, что применение теории игр имеет два различных аспекта: во-первых, она может использоваться в целях оптимизации механизмов принятия решений участниками выборов и, во-вторых, для выработки принципов их организации. В частности, во втором случае крайне актуальным становится вопрос об устойчивости игры (которой мы описываем электоральный процесс) в смысле Нэша. Напомним, что игра называется устойчивой в смысле Нэша, если ни один из игроков не может увеличить свой выигрыш только за счет своих собственных действий. Формально равновесие по Нэшу определяется следующим образом [22, 23]. В случаях, когда свои ожидания о поведении партнера каждый игрок строит по прошлому опыту подобных игр, устойчивое в каком-то смысле решение игры естественно называть равновесием этой популяции. В этих случаях особое значение приобретает равновесие по Нэшу - профиль стратегий, от

которого никому не выгодно отклоняться, если партнеры не отклоняются. Соответственно, множество нэшевских равновесий есть:

$$NE := \{\bar{x} \in X \mid y_i \in X_i \Rightarrow u_i(\bar{x}_i, \bar{x}_{-i}) \geq u_i(y_i, \bar{x}_{-i}) \forall i \in I\},$$

(здесь u – целевая функция) если же все неравенства строгие, то говорят о строгих равновесиях по Нэшу (SNE). Иными словами, Нэшевское равновесие – точка из которой ни одному игроку нет пользы уходить при текущих ходах партнеров, а строгое Нэшевское равновесие – точка, из которой невыгодно уходить. Если обозначить отображение рационального отклика i -го участника на ожидаемые действия x_{-i} его партнеров:

$$\chi_i^*(\cdot): X_{-i} \rightarrow X_i$$

то в этих терминах равновесие по Нэшу – это профиль рациональных откликов всех игроков на рациональные отклики партнеров:

$$\bar{x} \in NE \Leftrightarrow \prod_i \chi_i^*(\bar{x}_{-i}).$$

где $\chi_i^*(x_{-i})$ состоит из аргументов, максимизирующих его целевую функцию:

$$\chi_i^*(x_{-i}) = \arg \max_{x_i \in X_i} u_i(x_i, x_{-i}) = \{x_i \in X_i \mid u_i(x_i, x_{-i}) \geq u_i(y_i, x_{-i}) \forall y_i \in X_i\}.$$

При анализе электоральных процессов часто рассматриваются ситуации нессиметричных условий для различных игроков. В таких случаях имеет смысл рассмотреть равновесие Штакельберга [80], которое в отличие от симметричных условий, предполагает различные принципы формирования ожиданий разных игроков. Первый игрок (лидер) ориентируется на оптимальные ответы партнеров, зная их предпочтения, а остальные играют, как в NE , лишь реагируя на его ход и на ходы друг друга. Равновесие Штакельберга может возникать, например, когда один из игроков делает свой выбор раньше других и знает их цели. Или когда он один, а однотипных ведомых игроков достаточно много, чтобы каждый не пытался просчитывать общие последствия своего хода.

Равновесие Штакельберга с лидером есть такой профиль стратегий всех, что лидер с учетом целей партнеров адекватно прогнозирует

равновесия Нэша, складывающиеся после его хода, и соответственно оптимизирует свою стратегию, а остальные поступают согласно его прогнозу. Считая 1-го игрока лидером, обозначим решение Нэша среди последователей при фиксированной стратегии $\bar{x}_1$ лидера - через $NE_{-1}(\bar{x}_1)$.

Равновесие Штакельберга с лидером ($StEP_1$) есть такой набор $\bar{x}$ что

$$\bar{x}_{-1} \in NE_{-1}(\bar{x}_1),$$

$$\exists \tilde{x}_1 \in X_1 : u_1(\tilde{x}_1, \tilde{x}_{-1}) > u_1(\bar{x}_1, \bar{x}_{-1}) \forall (\tilde{x}_{-1} \in NE_{-1}(\tilde{x}), \bar{x}_{-1} \in NE_{-1}(\bar{x})).$$

В частности, пессимистическое равновесие Штакельберга с лидером есть такой набор $\bar{x} \in StEP_1$, что

$$\bar{x}_1 \in \arg \max_{x_1 \in X_1} \min_{x_{-1} \in NE_{-1}(x_1)} u_1(x_1, x_{-1}),$$

$$\bar{x}_{-1} \in \arg \min_{x_{-1} \in NE_{-1}(\bar{x}_1)} u_1(\bar{x}_1, x_{-1}).$$

Оптимистическое равновесие Штакельберга с лидером $\bar{x} \in StEO_1$ определяется так же, но с заменой $\min$ на $\max$.

Концепция $StEO_1$ предполагает доброжелательность партнеров к лидеру при выборе из эквивалентных для себя вариантов (из χ^*), а $StEP_1$ - недоброжелательность; если же выбор игроков однозначен, то разницы между $StEO$ и $StEP$ нет. Если не различать оптимистические и пессимистические решения, то можно определить $StE = \{StEO, StEP\}$.

Еще одним применением теории игр к электоральным процессам может быть оптимизация определения состава и локализации избирательных округов в плане выравнивания контролируемых игроками ресурсов.

В работе украинских математиков [33] рассматривается теоретико-игровая модель борьбы двух (модель естественно расширяется на любое число) крупных партий за электорат во время избирательной кампании. Предполагается, что в выборах участвует n регионов с электоратом $A_i, i = 1...n$. Каждая партия принимает решение об объеме средств, вкладываемых в предвыборную кампанию и по распределению их по регионам. Выигрыш партии – это максимальное количество голосов,

превышающее количество голосов, полученное другой партией. Предполагается, что одна партия делает свой выбор о величине вкладываемых средств, когда выбор другой партии известен лишь с некоторой вероятностью.

В работе [33] вводится функция выигрыша

$$f_i^j(x_i^j) = \frac{k_i^j x_i^j}{1 + k_i^j x_i^j},$$

который получит партия j по i -му региону, при условии, что в этот регион вложено x_i^j средств. Коэффициенты k_i^j отражают приоритет i -го региона для j -й партии.

Показано, что в условиях, когда обе партии одновременно «выбирают» электорат из каждого региона, общий выигрыш для j -х партий $F^j(x_i^j)$ по всем регионам имеет вид:

$$F^1(x^1, x^2) = \sum_{i=1}^n \frac{f_i^1(x_i^1)(1 - f_i^2(x_i^2))}{1 - f_i^1(x_i^1)f_i^2(x_i^2)} A_i,$$

$$F^2(x^2, x^1) = \sum_{i=1}^n \frac{f_i^2(x_i^2)(1 - f_i^1(x_i^1))}{1 - f_i^2(x_i^2)f_i^1(x_i^1)} A_i.$$

Задача оптимального распределения средств по регионам с целью максимизации выигрыша партий при существующих ограничениях на вкладываемые средства в результате представляются в виде задачи решения нелинейной системы из $2n+2$ уравнений с $2n+2$ неизвестными.

Следует иметь в виду, что модели теории игр в меньшей степени, чем рассмотренные выше, могут использоваться для более или менее точных расчетов и прогнозов. Скорее здесь можно говорить о хорошо обоснованной методологии, которая может существенно повысить эффективность действий участников электоральных процессов. Эти модели, по сути, представляют собой наборы рекомендаций, которые дают заметные преимущества тем, кто их использует.

3. МОДЕЛИРОВАНИЕ ИНФОРМАЦИОННЫХ ПОТОКОВ

Появление в современной литературе термина «Информационное общество» должно было, вероятно, отражать тот факт, что информация в наше время становится ключевым фактором динамики общественного развития. Вопрос о том, какой мере это соответствует действительности, выходит далеко за рамки нашего обсуждения, так же, как и вопрос о том, что такое на самом деле информация. Но то, что в нашем обществе информационные технологии действительно играют исключительно важную роль, и что их вклад в протекание общественных процессов постоянно растет, сомнений не вызывает.

Более того, бурное развитие информационных технологий в последние годы привело к появлению качественно новых ситуаций и проблем, о которых ранее человечество даже не подозревало.

Для иллюстрации этих, несомненно, важных и достаточно сложных, вопросов обратимся к одному из наиболее простых и наглядных аспектов понимания места и роли информации в современном обществе.

В прошлые (и даже не столь отдаленные) времена считалось, что информация всего лишь обеспечивает нашу осведомленность о событиях и фактах в окружающем нас мире. Мы либо знаем о чем-то, что для нас представляет интерес, либо не знаем. Если знаем – хорошо. Всегда лучше знать, чем не знать, а для того, чтобы знать, необходимо получить соответствующую информацию. Таким образом, информация воспринималась как полезный инструмент, предназначенный для расширения наших возможностей. Правда, уже давно было хорошо известно, что помимо информации существует также и дезинформация, но это обстоятельство рассматривалось как досадная, но незначительная неприятность. Ни у кого не

возникало сомнений в том, что та и другая качественно отличаются друг от друга, и что всегда может быть найден способ их дифференциации.

Однако со временем стало ясно, что в технологически развитых общественных системах информация, как бы мы ее ни определяли и какие свойства ее ни приписывали, представляет собой нечто самодостаточное, существующее и функционирующее по своим собственным законам, и прочно занявшее свою нишу в общей структуре общественной организации. В этом плане угол зрения, под которым информация четко отделена от дезинформации и является ее противоположностью, оказывается непродуктивным. Возьмем, например, роман Герберта Уэллса «Война миров», необычайно реалистично повествующий о вторжении на Землю марсиан. Что это: информация или дезинформация? На самом деле, это художественная литература. Отлично, но это решение проблемы только в том случае, когда вы держите в руках книгу с обложкой, на которой написано, что перед вами фантастический роман, сочиненный Гербертом Уэллсом для вашего развлечения. А если вы получите фрагмент этого романа по электронной почте безо всяких указаний на его происхождение и назначение? Вы скажете, что дезинформация здесь имеет место в действиях того, кто послал такое сообщение, и с этим трудно не согласиться. Но тогда получается, что понятие «дезинформация» (а значит, и понятие «информация») относятся не к сообщению как таковому, а к мотивации того, кто его посылает. Но ведь определенное количество бит, в которых измеряется информация, содержит сообщение, а не мотивацию. И компьютерные системы манипулируют именно с сообщением, а вовсе не с чьей-то мотивацией, о которой мы, вообще говоря, можем только догадываться, ибо, как известно, чужой мир – потемки.

Конечно, можно сказать, что подлинным критерием здесь должно быть соответствие содержания сообщения действительности. Но, во-первых, надежную проверку далеко не всегда можно осуществить технически, а во-вторых - в случае с высадкой марсиан разобраться просто, но когда речь идет

о не столь экзотических вещах, возникает проблема интерпретации (что, скажем, можно считать подарком, а что следует квалифицировать как взятку?).

В реальной жизни очень редко встречаются безупречно точные и строгие изложения фактической стороны дела, а когда встречаются, то обычно большинством потребителей воспринимаются как совершенно «несъедобные». Среднестатистического потребителя не воодушевляют колонки цифр малопонятные специальные термины. Потребитель хочет ярких, интуитивно понятных образов, желательно воздействующих на эмоциональную сферу. А спрос, как известно, рождает предложение, и в этом отношении информационные технологии отнюдь не являются исключением. Бессмысленно спорить о том, плохо это или хорошо. Так устроен современный мир, и с этим приходится считаться.

Как следствие, в современном обществе качественно меняется характер производства и потребления информации, которая, становится неотличимой (по крайней мере, на структурном уровне) от дезинформации.

Получая с помощью информационных технологий сообщение, мы уже не пытаемся решать вопрос о его достоверности так, как это делалось, скажем, в XVIII-м веке. Мы употребляем его в соответствии с неким набором правил, который сформировался на уровне коллективного сознания и прочно внедрился в сознание индивидуальное. Причем мы уже давно перестали задумываться о таких вещах, как перестали интересоваться чего больше в бифштексе – мяса или сои.

Ослабление контроля, особенно если оно не осознается, порождает химер, к числу коих относится малоизученный (несмотря на множество публикаций) феномен, известный в литературе как «виртуальная реальность». Уже сам термин должен был бы заставить нас задуматься, поскольку он явно предполагает, что могут существовать разные реальности. Как бы мы ни относились к такой возможности, в любом случае она требует глубокого переосмысления фундаментальных принципов общей картины мира.

Вероятно, нет нужды доказывать, что информационные аспекты социальных явлений исключительно важны для понимания электоральных процессов. Действительно, трудно себе представить избирателей, которые голосуют вне информационного контекста избирательной компании. Поэтому нам также придется серьезно заняться такого рода вопросами.

Приведенные выше рассуждения демонстрируют сложность и многогранность всего того, что определяет место и роль информации в современном человеческом обществе. Нас, однако, будут интересовать более прозаические, но никак не менее важные вопросы. Один из них связан с понятием «информационные потоки», получившем в последнее время широкое распространение, прежде всего в результате развития сетевых технологий.

По мере разработки и внедрения в практику компьютерных сетей стало ясно, что одним из главных свойств информации, чем бы она ни была на самом деле, является ее способность к движению. В этом отношении она во многом подобна жидкости, текущей по трубам или, скажем, электрическому току, текущему по проводам. Аналогия настолько полная, что для описания многих процессов, в которых обмен информацией играет значительную роль, оказывается продуктивным использовать представление о потоках информации (информационных потоках). В большинстве случаев для этого нет нужды строить корректные определения – интуитивного понимания того, о чем идет речь, бывает достаточно. Тем не менее, многими авторами предпринимаются усилия по созданию полноценной теории, и есть все основания полагать, что данная программа будет успешно выполнена.

Ниже мы займемся вопросами моделирования динамики информационных потоков в сети Интернет прежде всего потому, что именно при анализе процессов, происходящих в глобальной сети, наиболее рельефно выступают их главные особенности.

Для изучения и моделирования динамических свойств информационных потоков в рамках данной работы примем некоторые допущения.

Предположим, что существует система, сканирующая новостную информацию с веб-сайтов сети Интернет (либо любой другой информационной среды, сети) по мере публикации этой информации¹. Т.е. на входе такой системы – веб-пространство, а на выходе – поток сообщений, следующих одно за другим по мере публикации. В узком смысле в рамках численного моделирования в пределах данной работы под информационным потоком будем понимать дискретный числовой ряд, члены которого соответствуют количеству публикаций в единицу времени (например, за час или за сутки).

В рамках такого подхода фактически анализируются элементарные единицы содержательного наполнения информации. В качестве такой единицы будем использовать документ. В узких рамках данной работы не различаются понятия «документ», «сообщение» или «публикация». В дальнейшем будет преимущественно использоваться термин «документ», поскольку он более привычен в областях исследований, связанных с поиском и анализом информации.

3.1. Тематические информационные потоки

Под тематическим информационным потоком в широком смысле будем понимать последовательность сообщений, соответствующих определенному тематическому запросу.

Под тематическим информационным потоком в узком смысле в рамках данной работы будем понимать количество документов, относящихся к заданной теме, сканируемых из сети и фильтруемых некоторой идеальной

¹ Конечно, идеальной системы такого типа возможно не существует, но, например, в распоряжении авторов находилась система контент-мониторинга InfoStream® (торговая марка ИЦ «ЭЛВИСТИ», Украина), сканирующая около 50 тыс. новостных сообщений в сутки с открытых веб-сайтов RUNeta.

системой контент-мониторинга в единицу времени. Рассмотрим общую картину динамики тематических информационных потоков, ограничившись механизмами, типичными для новостного сегмента Интернет.

Подобное понимание сетевых документальных потоков, по-видимому, позволяет более или менее адекватно описывать общие закономерности в динамике информационных потоков.

В практическом плане часто оказывается вполне удовлетворительным упрощенное понимание информационного потока как некой зависящей от времени величины $X(t)$, описываемой уравнением:

$$\frac{dX(t)}{dt} = F(X(t), t). \quad (3.1)$$

Многочисленные факты свидетельствуют о том, что в действительности динамика тематических информационных потоков определяется комплексом внутренних нелинейных механизмов, лишь частично коррелирующих с ее объективным окружением. Очевидно, что она в принципе не может быть сведена к какому-либо одному фактору, полностью ответственному за все разнообразие наблюдаемых эффектов. Именно это обстоятельство и придает особую актуальность проблеме моделирования динамики сетевых тематических потоков.

Далее, каждая также обладает рядом характерных свойств, допускающих некоторую классификацию, например, на основе особенностей ее генерации и воспроизведения во времени:

- публикации на «разовую» тему, временная зависимость числа которых резко возрастает, выходит на насыщение, а затем убывает и асимптотически стремится к нулю;
- публикации по теме, периодически появляющиеся в информационном потоке по истечении ограниченного промежутка времени, практически исчезающие из него;

- публикации по теме, временная зависимость числа которых колеблется около некоторого значения, никогда не исчезая полностью.

Соответственно этому и сообщения могут подразделяться на аналогичные категории, причем каждая из них имеет собственную специфику развития во времени.

3.2. Корреляционный анализ информационных потоков

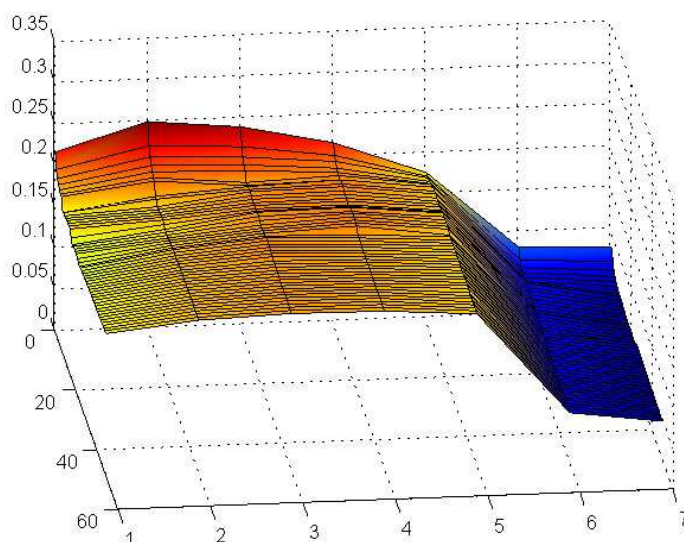
При изучении и моделировании тематических информационных потоков необходимо знание их «трендов», особенностей реального поведения, без учета периодических составляющих, образуемых, например, праздничными датами при рассмотрении информационных потоков за несколько лет или различными объемами публикаций в различные дни недели.

На рис. 31 приведено соотношение количества новостных сообщений, сканируемых системой контент-мониторинга InfoStream [6, 14] в 2006 году, по дням недели (количество сканируемых новостных сообщений превысило 10 млн. документов).

Нетрудно заметить, что она носит достаточно сложный колебательный характер. Это свидетельствует о том, что в динамике информационных потоков присутствуют более или менее устойчивые циклы, имеющие различную природу. Это обстоятельство позволяет нам говорить о наличии в динамике потоков отдельных составляющих, каждая из которых может составить предмет самостоятельного исследования.

Известно, что для решения поставленной проблемы могут применяться разнообразные подходы – от простой «компенсации» приведенного процентного соотношения публикаций по дням недели, до методов, базирующихся на преобразованиях Фурье и вейвлет-анализе. Вместе с тем

для обнаружения и обоснования последующего устранения периодических составляющих в динамике тематических информационных подходов в рамках данной работы применены подходы, обоснованные в рамках корреляционной теории [70, 20] с последующим применением метода «скользящей средней», которые представляются наиболее обоснованными и реализуемыми с инженерной точки зрения.



*Рис. 31. Недельные колебания объемов информации (ось Z), сканируемых на протяжении 2006 г. (номер недели в году - ось Y) в процентах:
 Пн – 17.23; Вт – 18.44; Ср – 18.80; Чт – 18.73; Пт – 17.86;
 Сб – 5.57; Вс – 4.16 (день недели – ось X)*

Если обозначить через X_t член ряда количества публикаций (количества электронных сообщений, поступивших, например, в день t , $t = 1, \dots, N$), то функция автокорреляции для этого ряда X определяется как:

$$F(k) = \frac{1}{N-k} \sum_{t=k}^{N-k} (X_{k+t} - m)(X_t - m), \quad (3.2)$$

где m – среднее значение ряда X , которое в дальнейшем, не ограничивая общности, будем считать равным 0 (это достигается присвоением значению X_t значения $X_t - m$). Предполагается, что ряд X может содержать скрытую периодическую составляющую.

Известно, что функция автокорреляции обладает тем свойством, что если скрытая периодическая составляющая существует, то ее значение асимптотически приближается к квадрату среднего значения исходного ряда.

Когда рассматриваемый ряд периодический, т.е. может быть представлен как:

$$X_t = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\omega t + \theta_n), \quad (3.3)$$

то его функция автокорреляции будет равна:

$$F(k) = \frac{a_0^2}{4} + \frac{1}{2} \sum_{n=1}^{\infty} a_n^2 \cos n\omega k. \quad (3.4)$$

Этот результат показывает, что функция автокорреляции периодического ряда также является периодической, содержит основную частоту и гармоники, но без фазовых углов θ_n .

Рассмотрим числовой ряд X , являющийся суммой некоторой содержательной составляющей N и синусоидальной сигнала S :

$$X_t = N_t + S_t. \quad (3.5)$$

Найдем функцию автокорреляции для этого ряда (значения приведены к среднему $m = 0$):

$$\begin{aligned} F(k) &= \frac{1}{N-k} \sum_{t=1}^{N-k} X_{k+t} X_t = \\ &= \frac{1}{N-k} \sum_{t=1}^{N-k} (N_{k+t} + S_{k+t})(N_t + S_t) = \\ &= \frac{1}{N-k} \sum_{t=1}^{N-k} N_{k+t} N_t + \frac{1}{N-k} \sum_{t=1}^{N-k} S_{k+t} S_t + \frac{1}{N-k} \sum_{t=1}^{N-k} N_{k+t} S_t + \frac{1}{N-k} \sum_{t=1}^{N-k} S_{k+t} N_t. \end{aligned} \quad (3.6)$$

Очевидно, первое слагаемое есть функция непериодическая, асимптотически стремящаяся к нулю. Так как N и S не когерентны, то взаимная корреляция между ними отсутствует, поэтому третье и четвертое слагаемое также стремятся к нулю. Таким образом, самый значительный

ненулевой вклад составляет второе слагаемое – автокорреляция сигнала S . Т.е. функция автокорреляции ряда X остается периодической.

Для экспериментального подтверждения рассмотренной гипотезы была сгенерирована последовательность, по своей природе напоминающая реальный информационный поток. Предполагалось, что ежедневное количество сообщений в сети растет по экспоненциальному закону (с очень небольшим значением экспоненциальной степени), и на это количество накладываются колебания, связанные с недельной цикличностью в работе информационных источников. Также принимается во внимание некоторый элемент случайности, выраженный соответствующими отклонениями.

Для получения соответствующего временного ряда была написана процедура на языке Perl:

```
#!/usr/bin/perl
for ($i=1; $i<1000; $i++) {
    $a=rand(1);
    $b=$a*exp($i*0.001)+sin($i*3.1415/7+$a);
    if ($b<0) {$b=0;}
    printf "%.4f\n",$b;
}
```

Итак, процедура реализует простейшую модель информационного потока – экспонента отвечает за рост количества публикаций во времени (общая тенденция), синус – за недельную периодичность, переменная a – за случайные отклонения. Количество публикаций b не может быть отрицательным числом. На рис. 32. представлен график модели (ось X – переменная i – день, ось Y – переменная b – число публикаций).

Исходный ряд был обработан: приведен к нулевому среднему и нормирован (каждый член разделен на среднее). После этого были рассчитаны коэффициенты корреляции $R(k)$. На рис. 33. представлена график

значений коэффициентов корреляций (ось X – переменная k , ось Y – коэффициент корреляции $R(k)$).

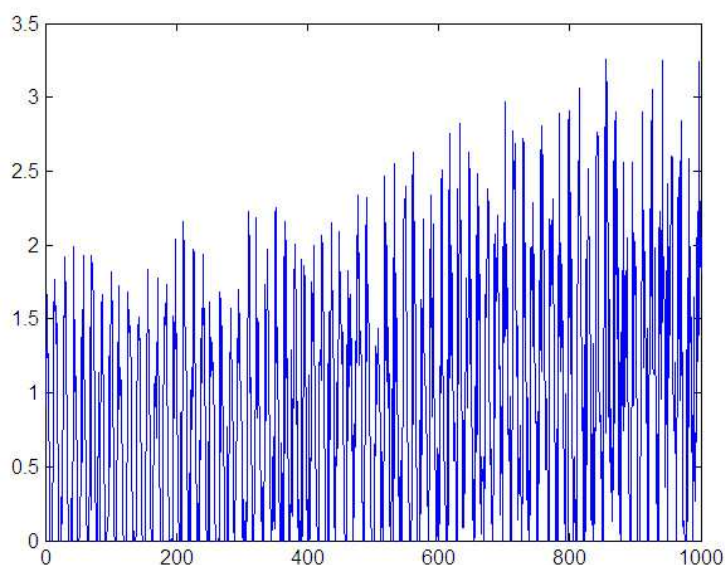


Рис. 32. Модель потока с экспоненциальным ростом

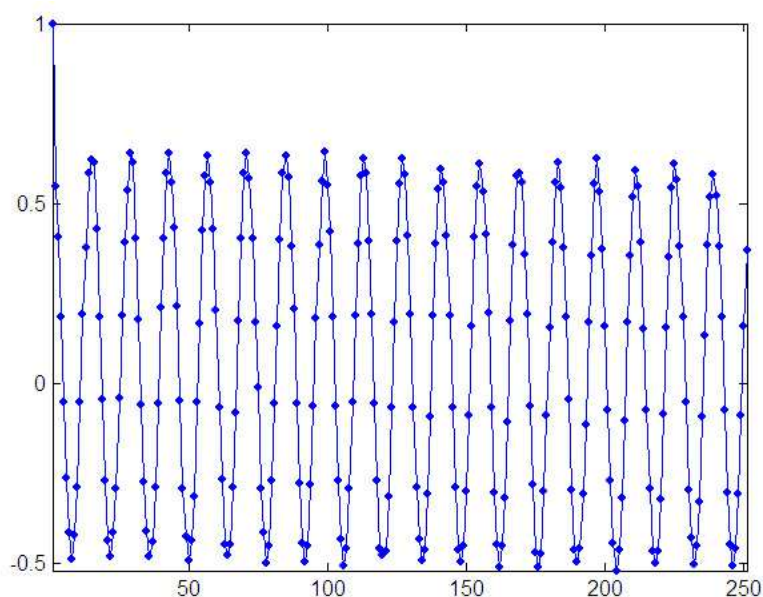


Рис. 33. Значения коэффициентов корреляции модели.

На рис. 34. представлен график реального информационного потока сообщений сетевых СМИ по теме «Коррупция в Украине», сформированный системой InfoStream, с помощью которой сканировалось свыше 2500 новостных веб-сайтов. По запросу «(корупц|коррупц) & (Украин|Україн)» было отобрано свыше 83 тыс. публикаций за 456 дней (с 1 января 2007 г. по 31 марта 2007 года).

Как известно, коэффициенты корреляции для дискретного ряда измерений рассчитываются следующим образом:

$$R(k) = \frac{F(k)}{\sigma^2}, \quad (3.7)$$

где k – ширина «окна наблюдений»; $F(k)$ — функция автокорреляции; σ^2 — дисперсия X .

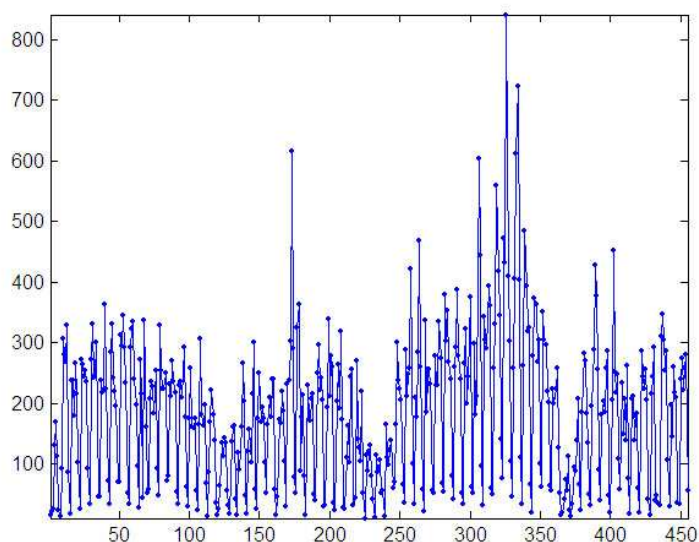


Рис. 34. Количество публикаций по заданной теме (ось Y) по дням (ось X)

*Графическое представление коэффициента корреляции для исследуемого ряда наблюдений свидетельствует о разделении корреляционных свойств по дням недели (рис. 35).

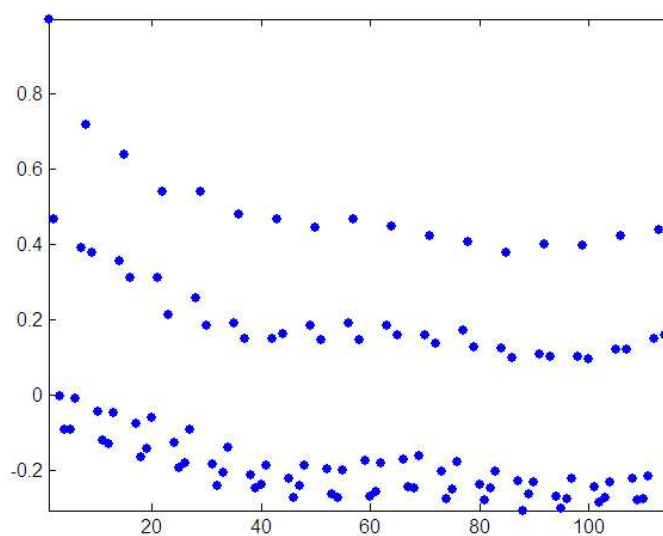


Рис. 35. Коэффициенты корреляции ряда количества публикаций в сутки (ось Y) в зависимости от окна наблюдений k (ось X)

Семь уровней значений, представленных на графике соответствуют 7 дням недели.

Вместе с тем очевидно, что приведенные выше коэффициенты корреляции представляют собой выбранные значения некоторой непрерывной функции, имеющей явно выраженную гармоническую составляющую (рис. 36).

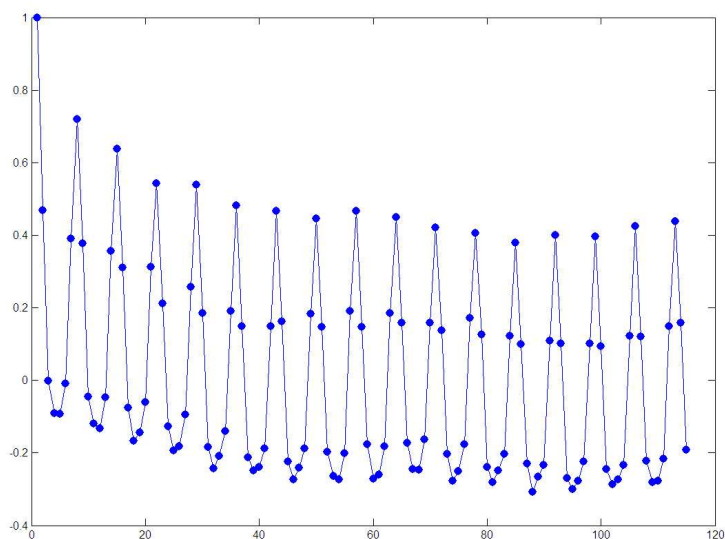


Рис. 36. Коэффициенты корреляции ряда как дискретные значения функции с гармонической составляющей

Для удаления периодической составляющей в исследуемом числовом ряду использовался метод взвешенной «скользящей средней». В основе этого метода сглаживания лежит принцип, заключающийся в том, что разброс среднего из N членов временного ряда будет характеризоваться величиной дисперсии, равной σ^2/N , где σ^2 – дисперсия исходного анализируемого ряда.

Предлагается рассматривать поведение значений данного числового ряда на промежутках, определяемых небольшими окнами наблюдений (в выбранном случае окно наблюдения равно 7 – числу дней недели), и вычислять значение нового ряда «сглаженных» величин, определяемых следующим образом:

$$S_t = \frac{1}{7} \sum_{i=t-3}^{t+3} X_i. \quad (3.8)$$

Очевидно, при изменении i от $t - 3$ до $t + 3$, происходит своеобразное «скольжение» по оси времени, соответственно, примененный метод называется методом «скользящей средней».

Следует заметить, что деление на ширину окна наблюдения позволяет при необходимости переходить к переменным окнам наблюдения, что особенно актуально в граничных точках. На рис. 37 приведен график «сглаженного» ряда, соответствующего рассматриваемому исходному.

Корреляционная функция «сглаженного» ряда (рис. 38) не содержит явно выраженных гармоник и подтверждает предположение, что основная периодическая составляющая рассматриваемого ряда соответствует 7-дневному (недельному) циклу. Одновременно, коэффициенты корреляции ряда наблюдений, усредненных по неделям, аппроксимируются гиперболической функцией, что свидетельствует о долгосрочной зависимости исходного ряда. Взаимная зависимость членов «сглаженного» ряда без учета циклической составляющей также подтверждается сравнением с «перемешанным» рядом.

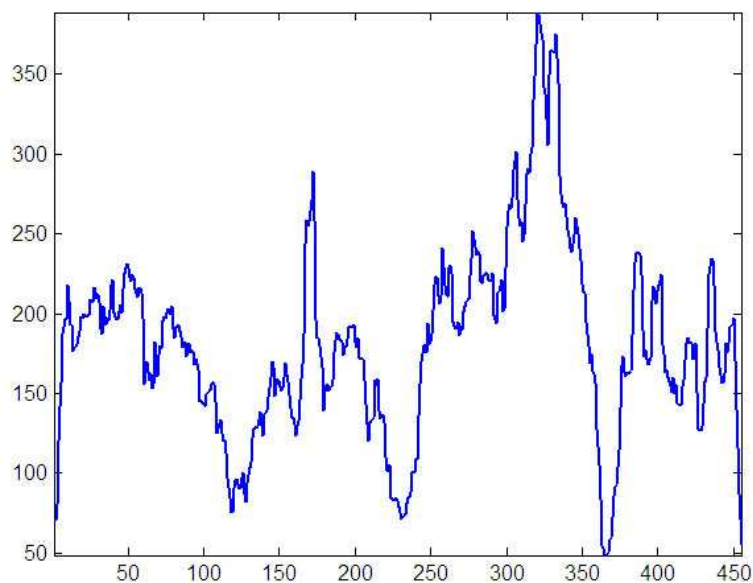


Рис. 37. «Сглаженный» ряд, соответствующий исходному

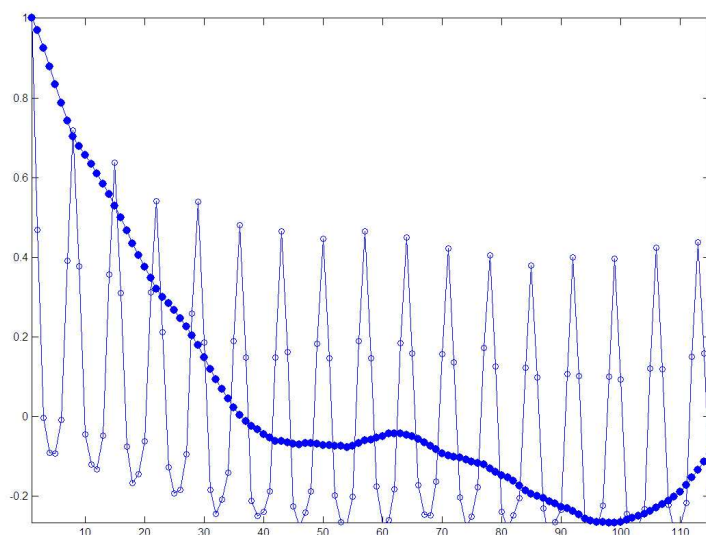


Рис. 38. Коэффициенты корреляции исходного (○) и сглаженного (●) ряда наблюдений

Предполагается, что «сглаживание» гармонической составляющей позволит выявить особенности реального поведения тематического информационного потока, его динамику, соответствующую событиям реального мира, без учета редакционной политики и периодики отдельных электронных СМИ. Таким образом, в дальнейших исследованиях информационных потоков, объемы которых представляются числовыми рядами с явно выраженной гармонической составляющей, предполагается использование «сглаженных» рядов с окном наблюдения в 7 дней, что позволит избавляться от явно выявленных с помощью корреляционного анализа недельных циклов в объемах публикаций.

В поведении современных информационных потоков [28, 26], наблюдаются две характерные особенности: отчетливая тенденция к постоянному росту их объемов и усложнение динамической структуры. В связи с этим актуальной становится проблема моделирования динамических процессов генерации и распространения информации, т.е. динамики информационных потоков.

Разнообразие поведения потоков публикаций по различным тематикам и сложность их взаимного влияния заставляют искать новые, ранее неизвестные в этой области методы. Сегодня, при моделировании информационных

потоков, особый интерес представляют модели, не претендующие на детальное описание особенностей каждого конкретного сообщения (публикации), а позволяющие обобщать и вместе с тем учитывать некоторую конкретику.

3.3. Моделирование динамики информационных потоков

Изучение динамики, построение моделей информационных потоков [28, 24, 6, 19, 20] является, бесспорно, важным и интересным, особенно учитывая то, что этот вопрос остается почти не исследованным [66].

На протяжении последних десятилетий были достигнуты определенные успехи в решении проблемы старения информации в рамках модели Бартона-Кеблера [61], которая возникла в свое время из необходимости оценки реальных сроков использования научных работ. Со временем оказалось, что полученные результаты (а также подходы, которые лежали в их основе) могут быть полезными в более широком контексте проблем информационных технологий. Однако понимание процессов динамики информационных потоков требует более глубокого анализа и более совершенной техники.

В данной работе, в частности, предлагается рассмотрение динамики тематических потоков новостной информации в рамках логистической модели [1, 2, 35, 39, 83]. Наряду с этим, выявлена ограниченность рассматриваемой модели, что открывает путь для дальнейших исследований.

Все Интернет-пространство можно с достаточной долей условности разделить на две составляющие - стабильную и динамическую [6], которые имеют очень разные характеристики с точки зрения интеграции информационных потоков. Стабильная составляющая Интернет содержит информацию долгосрочного характера, в то время как динамическая - постоянно обновляемые ресурсы. Некоторая часть этой составляющей со

временем вливается в стабильную. Однако большая часть исчезает из Интернет или попадает в сегмент "скрытого" Web-пространства, не доступного пользователям с помощью публичных информационно-поисковых систем.

Наиболее выраженным в плане динамики является, бесспорно, сегмент новостной информации. С одной стороны, он имеет высочайший уровень обновляемости, а с другой - в нем генерируются и распространяются на самом деле большие объемы данных. Поэтому именно он выглядит наиболее подходящим для исследований. В частности, процессы старения информации, потери ее актуальности в известной модели Бартона-Кеблера [54] описываются уравнением, которое состоит из двух компонент:

$$m(t) = 1 - ae^{-T} - be^{-2T},$$

где $m(t)$ – доля полезной информации в общем потоке через время T , первое вычитаемое соответствует стабильным ресурсам, а второе – динамическим – новостным.

Учет старения информации (потери части актуальности) имеет большое значение при аналитических исследованиях, создании информационных продуктов типа информационных портретов, основных сюжетов событий, ранжировании результатов работы информационно-поисковых систем. Даже приблизительная оценка скорости старения информации и отдельных документов имеет огромную практическую ценность, так как помогает держать в поле зрения только наиболее актуальную информацию.

С философской точки зрения понятие старения документов можно рассматривать как закономерный постоянный процесс уменьшения со временем их использования для получения необходимой пользователям информации, которая содержится в них. Процесс старения информации можно рассматривать как потерю информацией практической полезности для потребителя. Старение информации проявляется в том, что постоянно возникают новые документы, новые источники, которые содержат более

полную, точную, достоверную информацию. Поэтому с целью экономии времени и ресурсов оправданно первоочередное обращение именно к этим документам и источникам. При этом сложность использования закономерностей старения информационных сообщений состоит из разности характеристик уменьшения их использования во времени в разных предметных областях и для разных временных периодов. Степень старения информации неодинакова для документов разных видов и тематик. На скорость старения влияют в разной степени очень много факторов. Особенности старения информации органически связаны с тенденциями развития каждого тематического направления.

Для того чтобы количественно оценить скорость старения информации, Р. Бартон и Р. Кеблер по аналогии с периодом полураспада радиоактивных веществ также ввели понятие «полупериода жизни» научных статей. Полупериод жизни в их понимании - это время, на протяжении которого была опубликована половина всех используемых в настоящее время документов относительно выбранного события или явления. Бартон и Кеблер определили периоды полураспада публикаций по физике (4,6 года), математике (10,5), геологии (11,8) [54].

Информационная динамика в сети обусловлена многими факторами, большинство которых вообще не поддаются точному анализу. Однако в рамках задачи моделирования как разумное допущение можно предположить, что общий характер временной зависимости числа тематических публикаций в Сети определяется довольно простыми закономерностями, которые целиком допускают построение математических моделей.

В известных работах, посвященных изучению старения информации, используется модель Мальтуса [71] (возможно, с некоторыми модификациями, например, в виде суперпозиции двух кривых с разными параметрами в рамках приведенной выше модели Бартона-Кеблера). Преимущество этой модели в том, что уравнение Мальтуса имеет точное

решение в виде очень простой и удобной функции - экспоненты, но с точки зрения интерпретации результатов она выглядит довольно сомнительной. Главной проблемой является то, что экспонентой является монотонно возрастающая функция, а, следовательно, принципиально не может описывать процессы, которые по своей природе должны иметь локальные экстремумы.

То, что новости со временем теряют актуальность, и соответствующее количество публикаций уменьшается, не нуждается в доказательствах. Поэтому для получения более адекватной зависимости следует обратиться к более сложным моделям.

Одной из самых перспективных выглядит логистическая модель, которая была предложена П. Ферхлюстом [83] для описания динамики населения и Р. Перлом [77] для биологических сообществ, а со временем хорошо зарекомендовала себя в ряде направлений научных исследований. Преимуществом этой модели является, в первую очередь, то, что она объединяет относительную простоту формулирования задачи с возможностью варьировать решения с помощью набора параметров, которые могут иметь более или менее прозрачное физическое содержание.

Анализ информационных потоков, их моделирование сегодня становится одним из наиболее информативных методов количественного изучения динамики отдельных тематических направлений. По изменению величин информационных потоков судят о скорости развития, как отдельных тематических направлений, так и всего информационного пространства.

Устойчивые статистические связи между отдельными сообщениями позволяют говорить о корреляции отдельных тематик, об эффективности ссылок на публикации предшественников, более ранние работы, цитирование, републикации и т.п.

Механизмы, которые базируются на обобщенных методах кластерного анализа, разрешают обнаруживать сообщения в информационных потоках, которые формируют вокруг себя новые тематические направления.

Кластерный анализ, теория фракталов и автомодельных процессов при их корректном применении разрешают количественно оценивать степень связи в тематических информационных потоках.

Из классической пространственно-векторной модели информационного пространства принято использовать модель $TF * IDF$, где TF – это локальная частота термина (Term Frequency), а IDF – величина, обратная частоте появления сообщений во всем информационном потоке, которые содержат этот терм (Inverse Document Frequency). В то время, как локальная частота термина в документе говорит о значимости термина в пределах документа, то обратная частота появления свидетельствует об уникальности термина во всем потоке документов. Поэтому произведение этих величин – достаточно удачный критерий определения значимости термина - веса. Предполагается, что новостные сообщения стареют, теряя свою актуальность с интенсивностью, которая определяется некоторым эмпирическим законом. Для иллюстрации предположим, что это экспоненциальный закон (в дальнейшем будет показана корректность такого предположения для большого количества примеров). Один из предложенных подходов к такой части обобщения, как ранжирование сообщений, состоит в использовании параметрических множителей, которые зависят от времени. Можно определить вес сообщения как произведение элементов типа $TF * IDF * e^{-\alpha t}$, где α – некоторая константа, t – интервал времени, которое прошло с момента появления сообщения в информационном потоке. Значение α может трактоваться как коэффициент полураспада актуальности сообщения (экспертно определяемый интервал времени, на протяжении которого сообщение в результате старения теряет свою актуальность наполовину). Например, если предположить, что через сутки документ теряет половину своей актуальности, то имеем: $e^{-\alpha \cdot 24} = 1/2$, и, соответственно, $\alpha = 0,025$.

3.3.1. Линейная модель

В некоторых случаях динамика тематических информационных потоков (повышения актуальности или старения информации) происходит линейно, то есть количество сообщений в момент времени t можно, соответственно, представить формулой:

$$y(t) = y(t_0) + v(t - t_0),$$

где $y(t)$ – количество сообщений на время t , v – средняя скорость увеличения (уменьшения) интенсивности тематического информационного потока во времени (например, в результате старения). Ниже на рис. 39 а) и б) приведены примеры линейного роста количества сообщений из информационного потока в системе контент-мониторинга InfoStream, в которых встречаются слова «многопартийность» и «электорат», соответственно. Из графиков, которые отображают динамику изменений понятий на протяжении года, сглаженную в соответствии с описанной выше процедурой, можно видеть, что уровень роста связан, прежде всего, с ростом количества источников, которые сканируются на протяжении года.

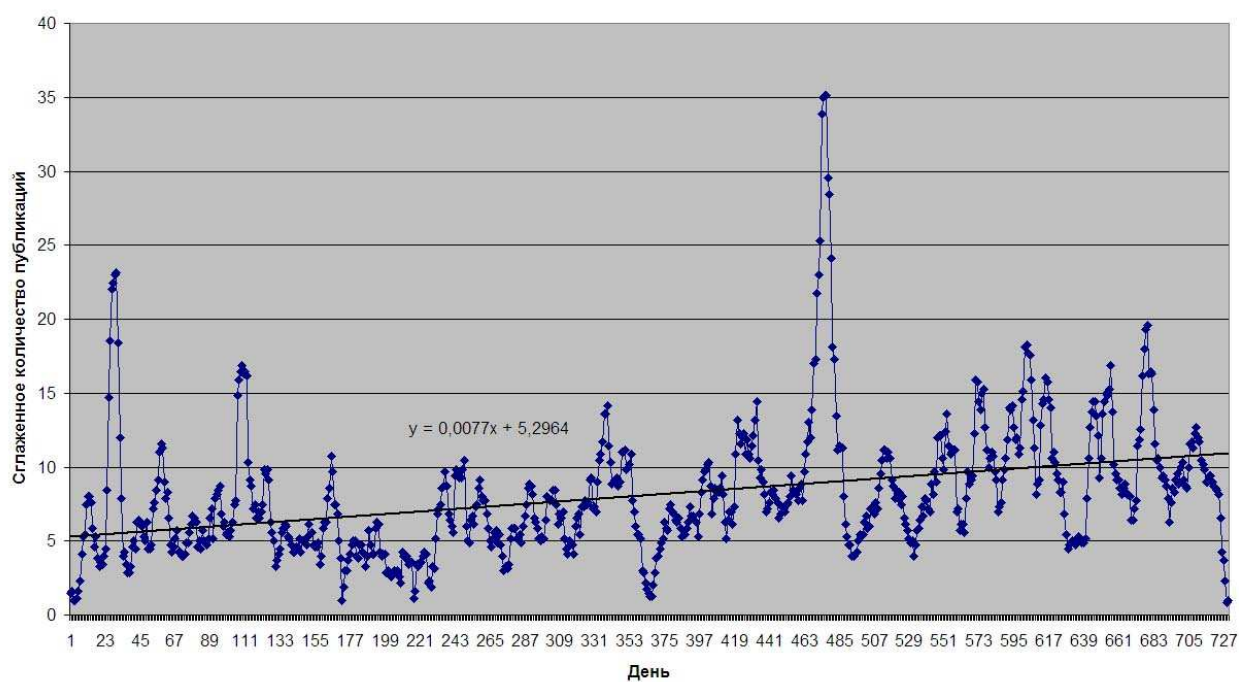
Содержательная составляющая информационного потока может быть количественно оценена как флюктуация информационного потока – изменение стандартного отклонения $\sigma(t)$, которое вычисляется по формуле:

$$\sigma(t_i) = \sqrt{\frac{1}{i} \sum_{k=0}^i [y(t_k) - (y(t_0) + v(t_i - t_0))]^2}.$$

Как показано в работе [21], если эти величины изменяются как корень квадратный из времени, то процесс изменения публикаций по теме можно считать процессом с независимыми приращениями. При этом связями с предыдущими публикациями можно пренебречь. В случае поведения стандартного отклонения по времени как $\sigma(t) \propto t^\mu$, чем большее значение μ , тем выше корреляция между текущими и предыдущими публикациями.

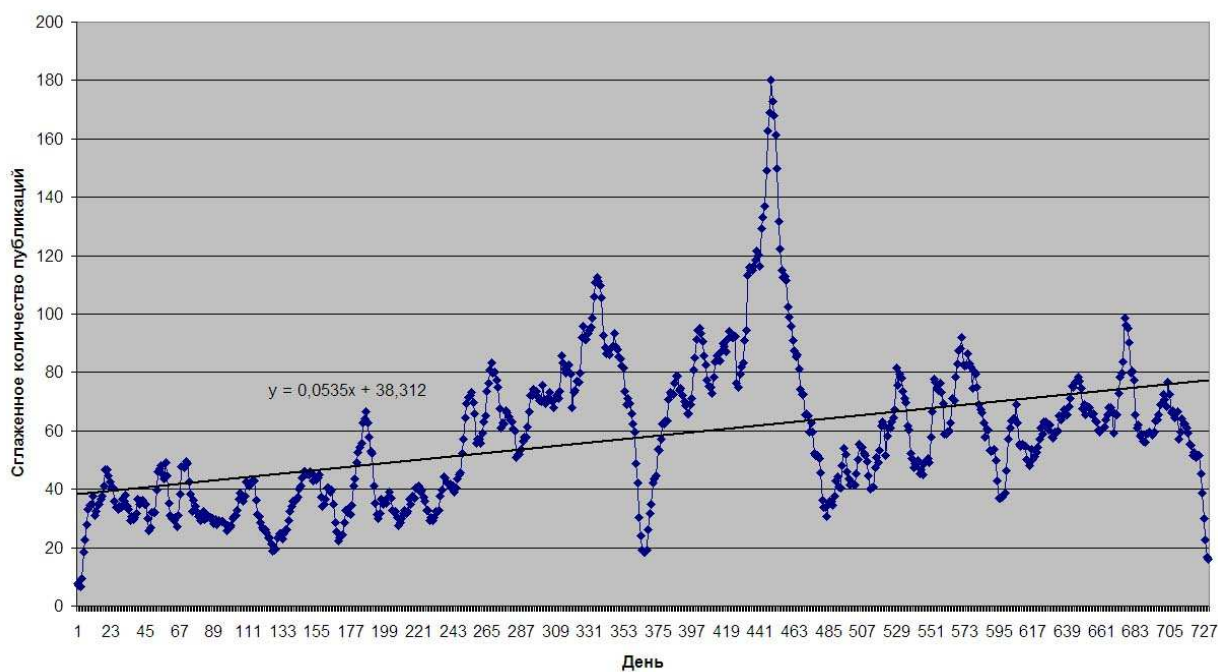
В этих случаях μ характеризует степень связи между случайными событиями и принимает значения от 1/2 до 1.

Интенсивность потока публикаций со словом "многопартийность"



а)

Интенсивность потока публикаций со словом "электорат"



б)

Рис. 39. Линейные тренды интенсивности публикаций со словами:
а) – многопартийность, б) - электорат

3.3.2. Экспоненциальная модель

В некоторых случаях процесс увеличения (роста) актуальности или старения информации описывается экспоненциальной зависимостью, которую можно аппроксимировать такой формулой:

$$N(t) = N(t_0)e^{\lambda(t-t_0)},$$

где μ - среднее относительное изменение интенсивности информационного потока.

Ниже (рис. 40) приведен пример экспоненциального роста количества сообщений из информационного потока системы контент-мониторинга InfoStream, в которых встретилось слово «блог». Экспоненциальный рост на протяжении 26 месяцев обусловлен ростом популярности нового средства общения в Интернет - «живых журналов».

Относительное изменение интенсивности в определенный момент времени вычисляется по формуле:

$$\lambda(t_i) = \frac{N(t_i) - N(t_{i-1})}{N(t_{i-1})}.$$

Изменение флуктуаций величины $\lambda(t_i)$ относительно среднего значения может быть оценено формулой:

$$\sigma(t_i) = \sqrt{\frac{1}{i} \sum_{k=0}^i [\lambda(t_k) - \lambda]^2}.$$

В этом случае также, если $\sigma(t)$ изменяется как корень квадратный от времени, то можно говорить о процессе с независимыми приращениями [21], корреляция между отдельными сообщениями несущественна. В случае наличия значительного количества зависимых сообщений справедливо: $\sigma(t) \propto t^\mu$, причем μ превышает 1/2, но ограничено 1. Значение μ , которое превышает 1/2, говорит о наличии долгосрочной памяти системы. Такие системы порождают класс процессов, который получил название

автомодельных, для которых предполагается корреляция между количеством сообщений информационных потоков в разные моменты времени.

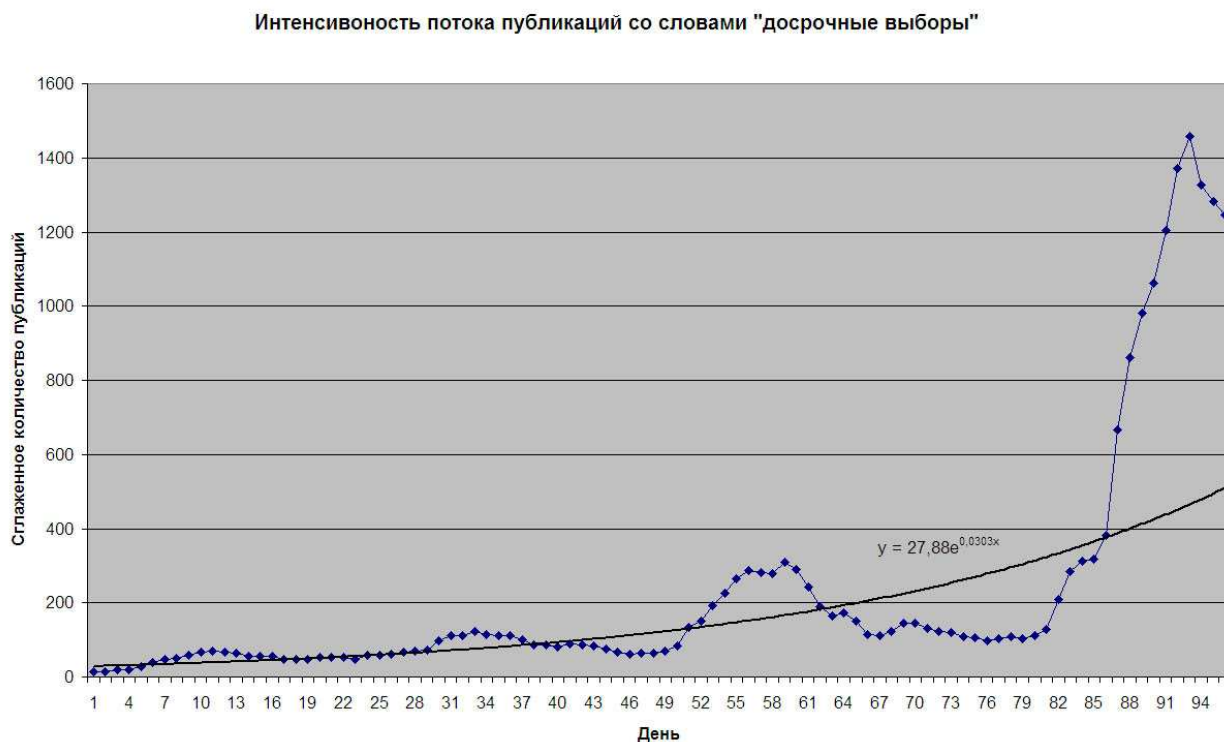


Рис. 40. Экспоненциальный тренд интенсивности публикаций

Изучение флуктуаций информационных потоков показывает наличие статистической корреляции как на коротких, так и на продолжительных временных интервалах.

Новые надежды дает применение теории фракталов, которая позволяет говорить о проявлении свойств подобия для коммуникационных процессов на разных уровнях. Такой подход разрешил расширить представление об основных закономерностях коммуникационных процессов (в том числе и процессов роста актуальности или старения информации).

3.3.3. Логистическая модель

Рассмотрим общую картину динамики тематических информационных потоков, ограничившись механизмами, типичными для новостного сегмента Интернет.

Мы исходим из того, что организации-генераторы новостной информации в абсолютном большинстве работают в стационарном режиме, который может характеризоваться максимальной емкостью информационного пространства N (укажем, что вопрос о размерности параметров, а также об их измерении мы в данной работе не рассматриваем). Это означает, что каждая организация-генератор производит поток информации, в среднем постоянный по количеству, как знаков, так и сообщений. Изменяются во времени лишь объемы сообщений, которые соответствуют той или другой теме. Другими словами, рост количества публикаций по одной теме сопровождается уменьшением публикаций на другие темы, так что для каждого промежутка времени T имеем:

$$\int_0^T \sum_{i=1}^M n_i(t) dt = NT,$$

где $n_i(t)$ – количество публикаций в единицу времени, а M – общее количество всех возможных тем. Конечно, предполагается, что часть $n_i(t)$ всегда равняется нулю.

Основной интерес в такой формулировке представляет изучение динамики отдельного тематического потока, который описывается плотностью $n_i(t)$.

Теоретически можно предположить, что множества публикаций, ассоциированных с определенным набором тем, пересекаются, то есть существуют публикации, которые могут быть отнесены одновременно к нескольким темам. Такая “политематичность” действительно является эффектом, с которым нельзя не считаться, но в первом приближении будем считать, что его вклад не искажает общую картину.

Дальше, будем считать, что на протяжении времени своего существования (актуальности) тема фиксирует комплекс механизмов, которые приводят к росту количества публикаций, имеющих определенные общие черты. Разные темпы могут порождать разные по объему потоки публикаций, ведь в этом плане они не являются равнозначными. Поэтому на

формальном уровне сопоставим с темой как абстрактным понятием два параметра: продолжительность (характерное “время жизни”) λ и интенсивность D . В рамках данной работы будем считать интенсивность величиной постоянной. Это, конечно, упрощенный взгляд, но вполне достаточный для выяснения общих тенденций.

Продолжительность, как вытекает из сказанного выше, не обязательно должна совпадать с началом и окончанием какого-то события в реальном мире (или ряда событий). Она характеризует лишь характерный промежуток времени, на протяжении которого тема имеет оконченную актуальность. Интенсивность можно определить как величину, которая характеризует порожденное соответствующей темой количество публикаций, усредненное по промежутку λ .

Реакция медийных средств, описываемая величиной D , никогда не бывает мгновенной: всегда существует определенная задержка во времени. Чтобы учесть этот аспект, введем фактор опоздания τ .

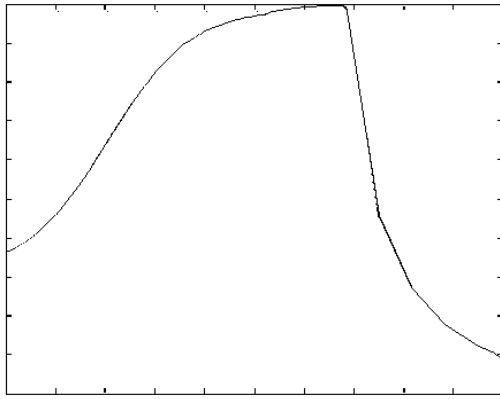
Теперь можно предложить следующую качественную картину динамики тематических информационных потоков. Генерация информационных потоков имеет две основные составляющие: фоновую и собственно тематическую. Фоновая составляющая определяется наложением многих слабо связанных между собой факторов и при определенных условиях может приближаться (с точки зрения тематических распределений) к шуму. Но она обеспечивает публикацию более или менее стабильного количества материалов, отобранных по принципу “Надо же что-то опубликовать!”

Возникновение новой темы вызывает процесс (точнее говоря, комплекс процессов) перераспределения сетевых ресурсов в связи с появлением актуальных сюжетов. Объем фоновых публикаций снижается, а тематических – возрастает. Если продолжительности двух или более тем пересекаются, то соответствующие тематические публикации также начинают перераспределяться между ними, причем характер

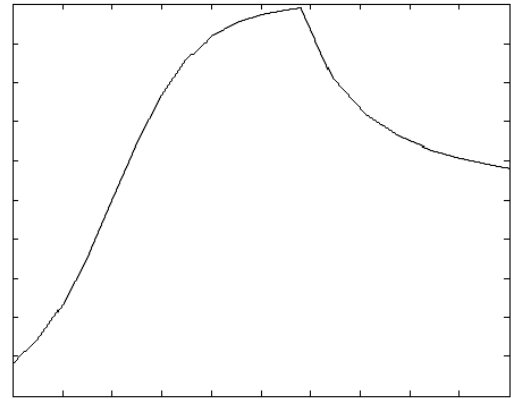
перераспределения определяется значениями параметров λ и D каждой темы. Когда же тема теряет актуальность, ассоциированные с ней ресурсы начинают переходить или в фоновые потоки, или в другие тематические.

Рассмотрим именно тематическую составляющую, причем сосредоточим внимание на динамике потоков, порожденных одной темой. Изучение взаимодействия нескольких тем представляет отдельное исследование, которое выходит за пределы поставленной нами задачи.

Приведем примеры лишь двух реальных информационных потоков, поведение которых нужно будет учесть в модели, которую опишем ниже. В первом случае (рис. 41 а) рассматривались публикации, которые сканировались системой мониторинга новостей из Интернет по тематике болезни и отхода от деятельности известного политического деятеля. К моменту обострения болезни объемы публикаций относительно его деятельности были на довольно высоком уровне. Болезнь значительно повысила количество публикаций, которое достигло верхнего уровня насыщения. Сведения относительно отхода от деятельности снизили количество публикаций до нижней планки, на этом уровне и произошла окончательная стабилизация. Второй пример - сенсационное избрание мэра большого города (рис. 41 б). До избирательной кампании об этом лице было не слишком много публикаций в Интернет, что соответствовало нижней стабильной планке. Выборы и утверждение мэра сопровождались значительным количеством публикаций как положительного, так и отрицательного характера (верхняя планка). Процесс дальнейшей деятельности мэра сопровождается объемами публикаций, которые соответствуют среднему стабилизационному уровню.



а)



б)

Рис. 41. Примеры информационных потоков

При желании логистическую модель можно рассматривать как обобщение модели Мальтуса, которая, как известно, предусматривает пропорциональность скорости роста функции ее значению в каждый момент времени:

$$\frac{dn(t)}{dt} = kn(t), \quad (3.9)$$

где k – некоторый коэффициент пропорциональности. Поскольку рассматривается динамика отдельного тематического потока, то далее не будем писать для величин $n_i(t)$ индексы, которые определяют тему.

Идея заключается в том, чтобы сделать коэффициент в уравнении Мальтуса функцией времени, причем так, чтобы решение не превышало заданного порогового значения. Существуют разные способы сделать это, но наиболее распространенным является использование константы, которая в явном виде ограничивает рост решения. В нашем случае с этой целью используем емкость N . Тогда правую часть выражения (3.9) можно представить в виде:

$$k(N - rn(t)), \quad (3.10)$$

где k – коэффициент Мальтуса, а r – фактор, который описывает отрицательные для данной системы процессы, связанные с внутренними факторами.

Теперь нам надо учесть в явном виде параметры, которые характеризуют влияние темы на динамику публикаций.

Поскольку интенсивность D определена нами как константа, ее взнос представим следующим образом:

$$y(t) = \begin{cases} D, 0 < t \leq \lambda \\ 0, t < 0, t > \lambda \end{cases} \quad (3.11)$$

Соответственно, будем рассматривать отдельно две временные области: $0 < t \leq \lambda$ с $D > 0$ и $t > \lambda$ с $D = 0$, для которых решениями являются функции $u(t)$ и $v(t)$. Полное решение получим путем “сшивки” на границе в точке λ :

$$n(t) = \begin{cases} u(t), 0 < t \leq \lambda \\ v(t), t > \lambda \end{cases}, \quad (3.12)$$

$$u(\lambda) = v(\lambda).$$

Первой области соответствует процесс роста числа публикаций на данную тему в условиях ее ненулевой актуальности ($D > 0$) и, возможно, переход к состоянию насыщения, а второй – процесс сокращения числа публикаций, обусловленный потерей актуальности ($D = 0$).

Отнормировав параметры к пороговой величине N , представим уравнение для первой области в таком виде:

$$\frac{du(t - \tau)}{dt} = pu(t - \tau)(1 - qu(t - \tau)) + Du(t - \tau), \quad (3.13)$$

$$u(0) = n_0.$$

Величина p определяет нормированную вероятность в единицу времени появления публикации независимо от актуальности данной темы. Такой фактор отображает фоновые механизмы генерации информации (типичным примером может быть механическое перепечатывание материалов престижных информационных ресурсов). Величина D характеризует непосредственное влияние актуальности данной темы. Параметр q характеризует уменьшение скорости роста количества

публикаций и является величиной, обратной асимптотическому значению зависимости $u(t)$ при $D = 0$.

Начальное условие в (3.13) отражает два аспекта информационной динамики: во-первых, наличие фоновой составляющей информационных потоков, а во-вторых, неопределенность точного момента, когда определенная тема начинает вносить свой вклад в общий процесс генерации публикаций. Ввиду этого, в момент времени $t = 0$ существует некоторое количество публикаций, которые могут быть ассоциированы с данной темой.

Для второй области, соответственно, имеем:

$$\frac{dv(t - \lambda)}{dt} = pv(t - \lambda)(1 - qv(t - \lambda)), \quad (3.14)$$

$$v(\lambda) = u(\lambda).$$

Так как во второй области тема уже не оказывает влияния на динамику публикаций (она описывает инерционные по отношению к теме процессы), в уравнении (3.13) не включается фактор запаздывания τ . Предельное условие в (3.14) обеспечивает “сшивку” функций $u(t)$ и $v(t)$.

Решение (3.13) имеет такой вид:

$$u(t) = \frac{u_s}{1 + \left(\frac{u_s}{n_0} - 1\right) \exp[-(p + D)(t - \tau)]}, \quad (3.15)$$

где u_s – асимптотическое значение u , величина которого определяет область насыщения (если, конечно, данная зависимость успевает ее достичь):

$$u_s = \frac{p + D}{pq}. \quad (3.16)$$

Заметим, что выражение (3.16) не зависит от значения n_0 , что свидетельствует о несущественности для состояния насыщения информационной динамики начальных условий. Каким бы ни было начальное количество публикаций, насыщение будет определяться исключительно параметрами, которые характеризуют фоновую скорость роста числа публикаций, количественную меру актуальности и отрицательные для процесса факторы. Поэтому с практической точки зрения

можем пренебречь фоновыми факторами, которые плохо подвергаются изучению.

Кривая (3.15) имеет точку перегиба:

$$t_{\text{inf}} = \frac{1}{p+D} \ln\left(\frac{u_s}{n_0} - 1\right) + \tau. \quad (3.17)$$

Таким образом, для первой области имеем так называемую S-подобную зависимость, а при $t \sim t_{\text{inf}}$ зависимость (3.15) приближается к линейной и соответствует линейной модели.

Представим теперь для удобства (3.15) в другом виде:

$$\frac{u_s \exp[(p+D)(t-\tau)]}{\exp[(p+D)(t-\tau)] + \left(\frac{u_s}{n_0} - 1\right)} = \frac{u_s \exp[(p+D)t]}{\exp[(p+D)t] + \left(\frac{u_s}{n_0} - 1\right) \exp[(p+D)\tau]}. \quad (3.18)$$

Из этой записи видно, что при условии

$$t < \frac{1}{p+D} \ln\left(\frac{u_s}{n_0} - 1\right) + \tau = t_{\text{inf}}, \quad (3.19)$$

зависимость $u(t)$ имеет экспоненциальный характер, причем ее выразительность определяется величиной запаздывания τ . Итак, для значений t , значительно меньше t_{inf} , наша модель совпадает с экспоненциальной моделью.

Типичная зависимость представлена на рис. 42.

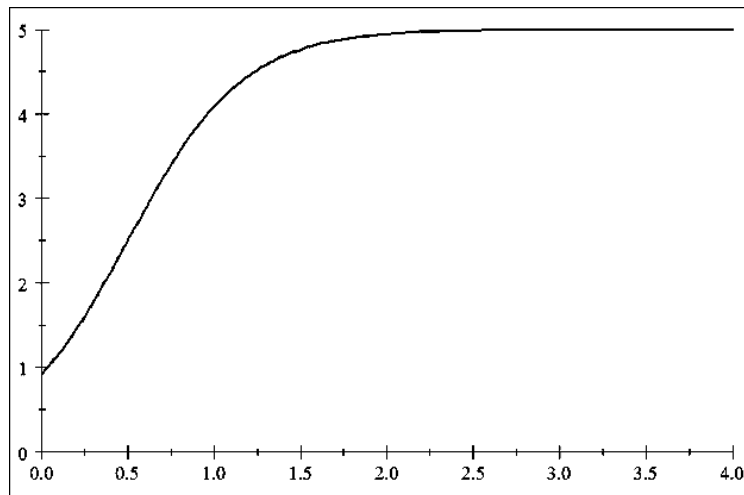


Рис. 42. Область роста

Перейдем ко второй области. Для нее решение имеет такой вид:

$$v(t) = \frac{u(\lambda)}{qu(\lambda) + (1 - qu(\lambda)) \exp[-p(t - \lambda)]} . \quad (3.20)$$

Если зависимость $u(t)$ успевает достичь насыщения за промежуток времени $t < \lambda$, можем упростить решение (3.20), представив его следующим образом:

$$v(t) = \frac{v_s (p + D)}{p + D(1 - \exp[-p(t - \lambda)])} , \quad (3.21)$$

где $v_s = 1/q$ асимптотическое значение зависимости $v(t)$.

Как и следовало ожидать, величина v_s также не зависит ни от начального условия, ни от условия “сшивки” на границе областей.

Во второй области динамика публикаций в первом приближении имеет экспоненциальный характер, что совпадает с результатами исследований.

Типичная зависимость для второй области представлена на рис. 43.

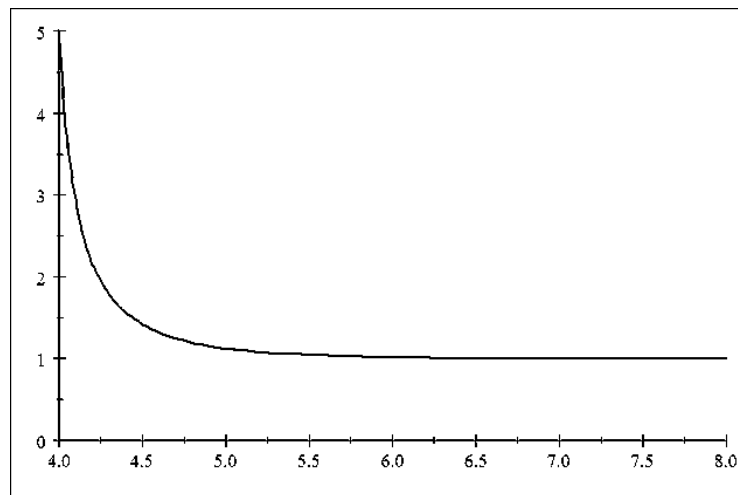


Рис. 43. Область спада

Итак, мы видим, что рассматриваемая зависимость имеет область насыщения u_s (при $t \leq \lambda$) и асимптотику v_s , которая описывает постепенное уменьшение числа публикаций к фоновому уровню. А это означает, что она,

по крайней мере качественно, согласована с общими представлениями о характере информационной динамики, полученными на основе экспериментальных данных. Кроме того, она неплохо совпадает с линейной и экспоненциальной моделями на определенных участках t .

Типичная полная зависимость $n(t)$ приведена на рис. 44.

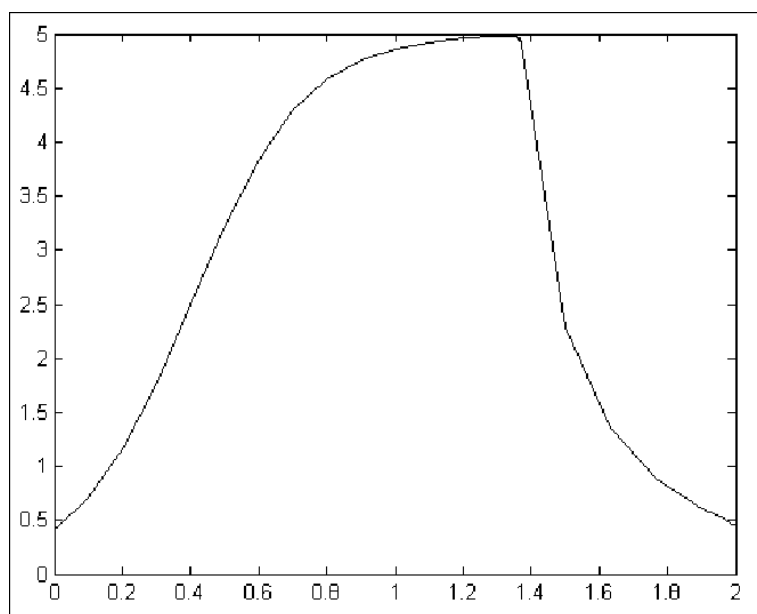


Рис. 44. Обобщенный график динамики тематического потока

Итак, предлагаемая модель правильно описывает (по крайней мере на уровне качественных свойств) временную зависимость плотности публикаций, порожденных отдельной темой. В частности, она содержит область насыщения, которую невозможно объяснить в рамках экспоненциальной модели.

Мы также видим, что полученная зависимость не является симметричной и имеет характерный “гребешок” на границе выделенных двух областей. Решения нашего уравнения для второй области, в отличие от первой, не имеет состояния насыщения: оно описывает близкий к экспоненциальному спад, который асимптотически приближается к нулю.

Такая интересная особенность поведения кривой на самом деле наблюдается на практике в определенной части случаев, но не во всех.

Экспериментальные данные свидетельствуют о наличии еще двух типов зависимостей, которые не обсуждаются в данной работе. Укажем лишь, что была рассмотрена простейшая реализация модели. Не исключено, что более сложные ее модификации дадут возможность описать все основные разновидности реальной динамики.

Отдельную проблему информационной динамики представляют циклические процессы роста и снижения активности информационных ресурсов, не связанные с собственно информационными факторами (например, периодическое снижение количества публикаций в выходные дни).

Открытой остается проблема определения соотношения решений приведенных логистических уравнений с условием баланса тем. Однако имеются веские основания для утверждения, что логистическая модель в самом деле описывает динамику определенной категории тематических информационных потоков.

3.4. Модели диффузии информации

Механизмы, связанные с распространением информации, играют настолько важную роль в электоральных процессах, что изучение их представляет собой одну из ключевых задач. Поэтому мы продолжим эту тему и проанализируем одну из наиболее интересных сторон процессов информационного обмена.

Выше уже говорилось, что одним из главных свойств информации с точки зрения современных взглядов является ее способность к движению. Здесь обратимся к еще одному направлению в изучении процессов, связанных с информацией. Конкретно, речь пойдет о крайне интересном явлении, которое часто называют диффузией информации.

Напомним, что в естественных науках под диффузией понимают взаимное проникновение друг в друга соприкасающихся веществ, вызванное тепловым движением их частиц. Важно отметить, что диффузия происходит

в направлении падения концентрации вещества и ведёт к его равномерному распределению по всему занимаемому объёму.

Для понимания существа дела следует, прежде всего, учесть, что информация также в определенном смысле состоит из «частиц» – документов (об этом уже говорилось применительно к информационным потокам). И эти «частицы» в определенных условиях могут взаимно перемещаться из одной области информационного пространства в другую.

Такая постановка вопроса предполагает, что можно придать смысл понятию границы областей информационного пространства. В настоящее время общее решение этой проблемы отсутствует, однако в большинстве конкретных случаев граница области информационного пространства может быть определена применительно к поставленной задаче. Так, например, хорошим приближением может служить выделение территорий, население которых обладает существенно большим уровнем информированности, чем их окружение (города и сельская местность и т. п.). Тогда граница областей приобретает вполне физический характер. Более сложная ситуация возникает при контактах различных социальных групп, по-разному информированных относительно определенных вопросов. Здесь граница областей возникает в условной зоне коммуникаций их представителей и потому не имеет территориального аспекта. Таким образом, она должна быть определена в некотором абстрактном пространстве, но, тем не менее, представить себе ее в смысле проявлений в механизме коммуникаций не составляет особого труда. Отметим также, что четкая граница областей имеет место только в начальный момент процесса диффузии, так как он приводит к ее размыванию и формированию градиента концентрации, и диффузия продолжается до тех пор, пока его величина остается отличной от нуля.

Главная особенность явления диффузии, позволяющая успешно использовать этот термин в расширенном варианте, применяя его к самым разнообразным, в том числе общественным явлениям, состоит в том, что здесь каждая частица «действует» вне какой бы то ни было связи с другими

частицами. Напротив, потоки возникают при наличии некой организующей силы, приводящей много частиц в согласованное движение. Под этим углом зрения и следует понимать диффузию информации. Именно, предполагается, что здесь имеет место передача отдельных сообщений от одного субъекта процесса к другому, причем носящая локальный характер. Такими субъектами могут быть отдельные личности, небольшие коллективы, обладающие локальным действием, например, информационные агентства и т. п.

Когда речь идет об электоральных процессах, понятие сообщения следует уточнить. Имеются в виду не только обычные информационные материалы, как новости, аналитические обзоры и т. д., касающиеся избирательной кампании, но и специфические формы воздействия на сознание избирателей. К ним относятся, в первую очередь, разные виды частной пропаганды и агитации на индивидуальном уровне, например, по принципу «А вот люди говорят ...». В определенных условиях такие механизмы могут играть роль, сопоставимую со средствами массовой агитации, целенаправленно используемыми в предвыборной борьбе политическими силами. Поэтому диффузные процессы должны быть одним из главных центров внимания.

Так же, как и в случае информационных потоков, аналогия с диффузией достаточно полная, что дает возможность использовать хорошо разработанный инструментарий.

Многие процессы, близкие к динамике информационных потоков, можно моделировать достаточно точно, если четко параметризовать и установить их граничные параметры. Однако, по-видимому, на данном этапе в области моделирования более сложных информационных процессов успех может быть достигнут только путем синтеза достаточно простых алгоритмов и концепций.

Следует признать перспективными в этой области и методы дискретной математики, к которым можно отнести и теорию клеточных автоматов.

В частности, данная теория успешно применялась при анализе диффузии инноваций, структурно близком интересующим нас процессам [58]. Действительно, если рассматривать инновации как элементы некоторого дискретного множества, обладающего определенным набором свойств, то окажется, что эти свойства в первом приближении совпадают со свойствами сообщений в том смысле, в котором о них договорились выше. Во всяком случае, процессы диффузии в обоих случаях выглядят одинаково. Поэтому для описания диффузии информации воспользуемся именно основами модели диффузии инноваций.

Подобная модель функционирует по следующим правилам: каждый индивид, способный принять инновацию, соответствует одной квадратной клетке, на двумерной плоскости. Каждая клетка может находиться в двух состояниях: 1 - новинка принята; 0 - новинка не принята. Предполагается, что автомат, восприняв инновацию один раз, запоминает ее навсегда (состояние 1 – не может быть изменено). Автомат принимает решение о принятии новинки, ориентируясь на мнение восьми ближайших соседей, т.е. если в окрестности данной клетки (используется окрестность Мура) имеется m сторонников новинки и, p - вероятность принятия новинки (генерируется по ходу работы модели), то при

$$pm > R, \quad (3.22)$$

где R - фиксированное пороговое значение, то клетка принимает инновацию (принимает значение 1). По мнению авторов этой модели, клеточное моделирование позволяет строить значительно более реалистические модели рынка инноваций, чем традиционные подходы.

Вместе с тем, динамике распространения информации присущи некоторые дополнительные свойства, которые были учтены в представленном ниже исследовании. В рамках описываемого исследования рассматривалась расширенная модель диффузии инноваций, которая относится к распространению новостей в информационном пространстве. При тех же условиях, касающихся клеточного пространства, окрестности Мура и вероятного

правила принятия новости, дополнительно предполагается, что клетка может быть в одном из трех состояний: 1 – «свежая новость» (клетка окрашивается в черный цвет); 2 – новость, устаревшая, но сохраненная в виде сведений (серая клетка); 3 – клетка не имеет информации, переданной новостным сообщением (клетка белая, информация не дошла или уже забыта). Правила распространения новостей следующие:

- изначально все поле состоит из белых клеток за исключением одной, черной, которая первой «приняла» новость (рис. 45 а);
- белая клетка может перекрашиваться только в черный цвет или оставаться белой (она может получать новость или оставаться «в неведении»);
- белая клетка перекрашивается, если выполняется условие, аналогичное (2), в модели диффузии инноваций: $pt > 1$;
- если клетка черная, а вокруг нее исключительно черные и серые, то она перекрашивается в серый цвет (новость устаревает, но сохраняется как сведения);
- если клетка серая, а вокруг нее исключительно серые и черные, то она перекрашивается в белый цвет (происходит забывание сведений при их общеизвестности).

Описанная система клеточных автомат вполне реалистично отражает процесс распространения новостей среди отдельных информационных источников и их публикаций. На поле размером 40 x 40 (размеры были выбраны авторами исключительно с целью наглядности) состояния системы клеточных автоматов полностью стабилизируются за ограниченное количество ходов, т.е. процесс эволюции - сходящийся. Пример работы модели приведен на рис. 45.



Рис.45. Процесс эволюции системы клеточных автоматов «диффузии новостей»: а) – исходное состояние; б-д) – промежуточные состояния; е) – конечное состояние

Многочисленные эксперименты с данным клеточным автоматом, доступным в настоящее время в сети Интернет по адресу <http://edu.infostream.ua/newsk.pl> показывают, что период его сходимости составляет от 80 до 150 шагов. Типичные зависимости количества клеток (последовательности количества однотипных клеток), пребывающих в

различных состояниях в зависимости от шага итерации приведены на рис. 46. При анализе приведенных графиков следует обратить внимание на такие особенности: 1 – суммарное количество клеток, пребывающих во всех трех состояниях на каждом шагу итерации постоянно и равно размеру поля, 2 – при стабилизации клеточных автоматов соотношения серых, белых и черных клеток приблизительно составляет: 0.75 : 0.25 : 0; существует точка пересечения кривых, определяемых всеми тремя последовательностями на уровне 33 % каждая.

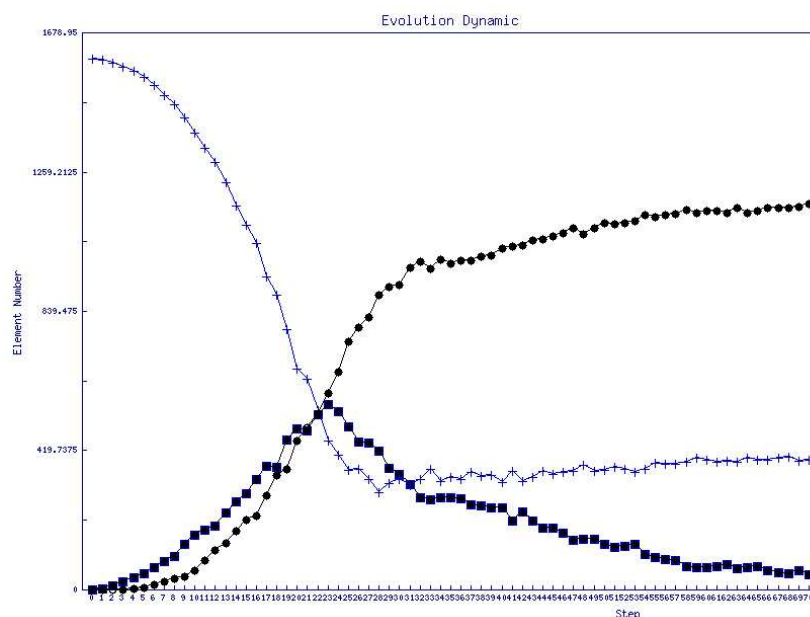


Рис. 46. Количество клеток каждого цвета в зависимости от шага эволюции: белые клетки - (+); серые клетки - (•); черные клетки – (■)

Особое внимание на графике следует обратить на зависимость, образованную черными клетками. Вид данной кривой вполне согласуется с «жизненной» динамикой новости, сначала она динамично распространяется, захватывая все новые уголки информационного пространства, затем происходит своеобразное насыщение, и информация для большинства реципиентов перестает быть новостью, переходя в разряд сведений или попросту забываясь.

Детальный анализ полученных зависимостей позволил провести аналогии данной модели «диффузии информации» со следующими аналитическими рассуждениями. Результаты моделирования позволяют

предположить, что эволюция серых клеток описывается некоторой непрерывной функцией:

$$x_g = f(t, \tau_g, \gamma_g), \quad (3.23)$$

где t - время (шаг эволюции), τ_g - сдвиг по времени, обеспечивающие получение требуемого фрагмента аналитической функции, γ_g - параметр крутизны данной функции.

Соответственно, динамика белых клеток x_w (количество клеток в момент t) может моделироваться «перевернутой» функцией x_g со своими аналогичными параметрами:

$$x_w = 1 - f(t, \tau_w, \gamma_w). \quad (3.24)$$

Поскольку, как было сказано выше, всегда выполняется условие баланса, т.е. общее число клеток в конкретный момент времени всегда постоянно, то условие нормировки можно записать:

$$x_g + x_w + x_b = 1, \quad (3.25)$$

где x_w - количество черных клеток в момент времени t .

Таким образом, в соответствии с (5):

$$x_b = 1 - x_g - x_w = f(t, \tau_w, \gamma_w) - f(t, \tau_g, \gamma_g). \quad (3.26)$$

Вид зависимости, представленной на рис. 2. позволяет предположить, что в качестве функции $f(t, \tau, \gamma)$ может быть выбрано следующее выражение:

$$f(t, \tau, \gamma) = \frac{C}{1 + e^{\gamma(t-\tau)}}, \quad (3.27)$$

где C - некоторая нормирующая константа.

На рис. 47 приведены графики зависимостей x_g, x_w, x_b от шага эволюции системы клеточных автоматов, полученные в результате аналитического моделирования, выраженного формулами:

$$\begin{aligned}
x_g &= \frac{0.75}{1 + e^{-0.15(t-30)}}; \\
x_w &= 1 - \frac{0.75}{1 + e^{-0.25(t-20)}}; \\
x_b &= 0.75 \left(\frac{1}{1 + e^{-0.25(t-20)}} - \frac{1}{1 + e^{-0.15(t-30)}} \right).
\end{aligned}
\tag{3.28}$$

Для экспериментальной проверки рассматриваемой модели были проведены исследования динамики объемов публикаций по двум резонансным темам, имеющим прямое отношение к избирательным процессам в Украине (досрочным выборам в Верховный Совет в 2007 году). Исследования проводились на базе системы интеграции информационных ресурсов сети Интернет InfoStream, обеспечивающей информационный поток в 50 тыс. сообщений в сутки.

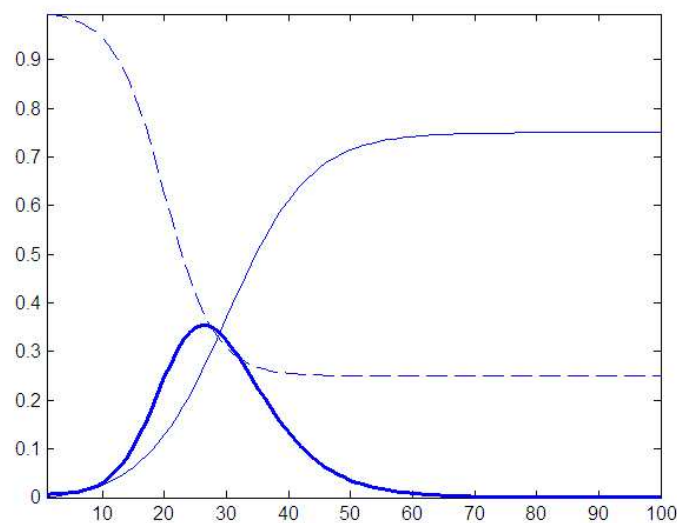
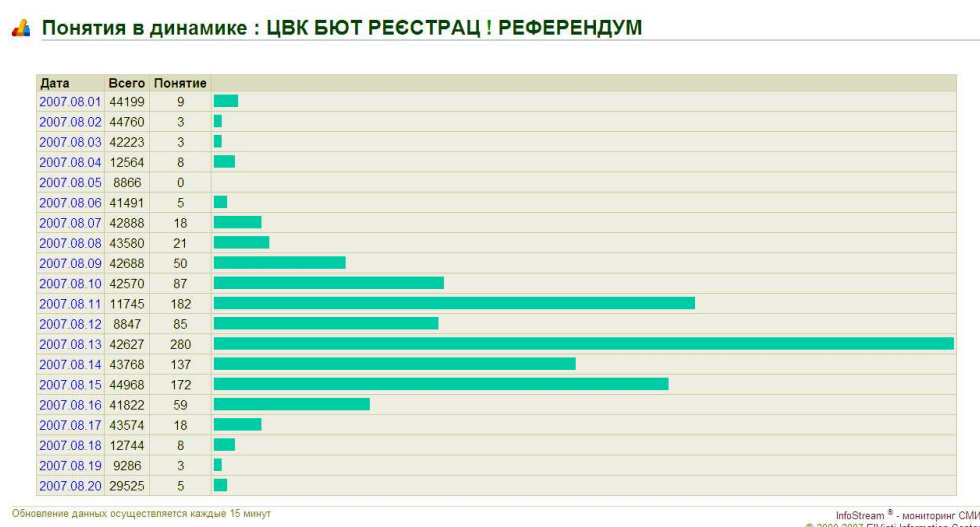


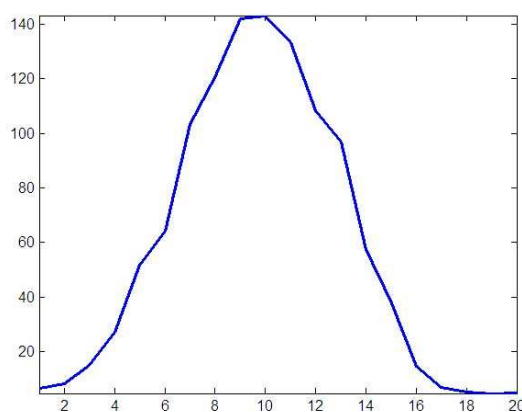
Рис.47. Непрерывные зависимости, полученные в результате аналитического моделирования, в зависимости от шага эволюции: сплошная линия – серые (x_g); пунктирная линия – белые (x_w); сплошная жирная линия – черные (x_b)

Для получения необходимых для эксперимента информационных потоков было сформировано два запроса к системе InfoStream. Первый запрос касался регистрации избирательного блока БЮТ (рис. 48), решение о которой, как известно, было принято в результате судебного рассмотрения.

Второй запрос позволил отобрать документы с информацией о четвертом указе Президента Украины, относящемся к досрочным выборам (рис. 49). На рис. 48 а) и 49 а) приведена реальная динамика объемов публикаций в Интернете по дням. На рис. 48 б) и 49 б) представлены зависимости, полученные сглаживанием соответствующих числовых рядов методом скользящих средних. Сравнение экспериментальных зависимостей с результатами аналитического моделирования позволяют говорить о высокой точности аппроксимации экспериментальных данных модельными кривыми.

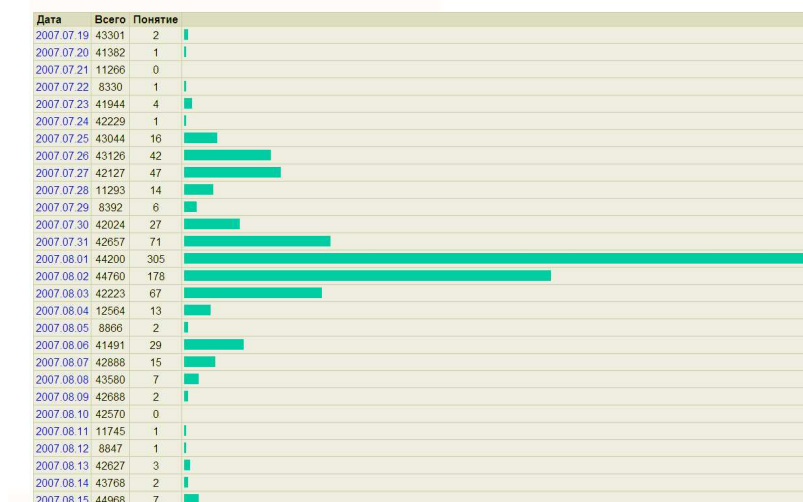


а)

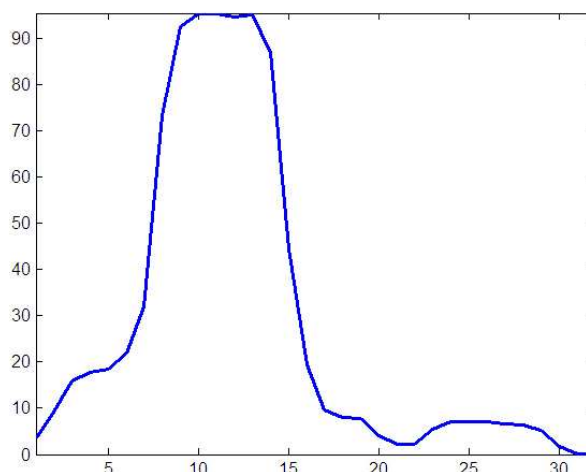


б)

Рис. 48. Динамика понятия «ЦВК, БЮТ, регистрация, но не референдум», в системе интеграции новостей InfoStream - а) сглаженная по дням недели зависимость объемов публикаций по данной теме - б)



а)



б)

Рис. 49. Динамика понятия «четвертый указ», в системе интеграции новостей InfoStream - а) сглаженная по дням недели зависимость объемов публикаций по данной теме - б)

Следует отметить, что полученная колоколообразная зависимость диффузии новостей на интернет-источниках (веб-сайтах) хорошо согласуется с «жизненным» поведением тематических информационных потоков, а на локальных временных промежутках с известными моделями, например, экспоненциальной и логистической. Вместе с тем данная модель является более реалистичной, чем названные традиционные модели.

3.5. Фрактальные свойства информационных потоков

Сегодня, наверное, уже никто не сомневается, в том, что сетевые информационные технологии начинают играть доминирующую роль в общей структуре коммуникационных процессов.

Ранее нами были рассмотрены вопросы информационной динамики электоральных процессов. Однако тот факт, что в основе сетевой информационной динамики лежат механизмы генерации, распространения и потребления сообщений (документов), придает актуальность проблеме изучения самих сетевых документальных массивов.

Ключевым моментом здесь является, несомненно, то обстоятельство, что сетевых ресурсов, образующих, так сказать, материальную базу документальных массивов, на самом деле очень много. К тому же они систематически обновляются. Общий уровень их упорядоченности намного ниже, чем традиционных СМИ. Если к этому добавить то, что значительная часть сетевых ресурсов практически не контролируется ответственными профессионалами, становится ясно, что происходящие процессы достаточно сложны и требуют для своего изучения применения развитых современных методов.

Как выяснилось в ходе исследования систем с высоким уровнем сложности, они в ряде случаев демонстрируют поведение, описание которого далеко выходит за рамки возможностей традиционной статистики. И проблема заключается даже не в том, что не удастся найти точные решения уравнений, а в том, что возникающие эффекты не могут быть описаны в статистических терминах. Для их полного понимания необходимо разработать новый понятийный аппарат, существенно расширяющий возможности описания коллективных процессов.

В результате возникли принципиально новые перспективные направления в науке, к числу которых принадлежит, например, теория детерминированного хаоса [45, 15], получившая в настоящее время широкое

распространение. Эта теория позволяет понять многие явления, казавшиеся необъяснимыми в рамках традиционных представлений.

Ниже для изучения информационных аспектов электоральных процессов воспользуемся другим, не менее перспективным направлением, известным в литературе как фрактальный анализ. Его преимущество состоит в том, что он позволяет исследовать не только и не столько процессы, связанные с динамикой сетевых документальных массивов, сколько специфические характеристики самих этих массивов. Учитывая специфику динамики сетевых информационных потоков, уместно считать эти потоки не простыми совокупностями никак не связанных между собой документов, а, напротив, системами, обладающими нетривиальным поведением.

Как уже отмечалось, система обладает свойствами, которые принципиально нельзя свести к свойствам образующих ее элементов. Поэтому документальные массивы как системы также могут демонстрировать поведение, для понимания которого уже недостаточно представлять себе характеристики отдельных документов. Так, например, заметную роль может играть распределение документов в массиве по некоему параметру. Именно для решения подобных задач и применяют фрактальный анализ. С его помощью можно абстрагироваться от природы и конкретных свойств элементов изучаемого множества и исследовать общие свойства, в том числе и те, которые принципиально не могут быть описаны в рамках традиционных представлений.

Одно из такого рода свойств, которому в последнее время принято уделять особое внимание, состоит в следующем. Выделим из некоего множества его произвольное подмножество. Очевидно, что оно в любом случае несет определенную информацию об исходном множестве, причем в той мере, в какой является его частью. Возникает вопрос: а нельзя ли выделить подмножество таким образом, чтобы оно несло информацию о множестве? Оказывается, это возможно.

Существует большой класс множеств, обладающих иерархически организованной структурой, на каждом уровне представляющей собой последовательность экземпляров некоего паттерна в соответствующем масштабе. Они в известном смысле напоминают вкладывающихся друг в друга матрешек: последовательно раскрывая их, мы на каждом шагу видим ту же самую матрешку, только меньшего размера.

Объекты, которые описываются такими множествами, называют фракталами. Оказывается, фракталы очень широко распространены в нашем мире. Настолько широко, что, кажется, проще найти фрактал, чем то, что им не является. Поэтому фрактальные свойства оказываются крайне полезны при изучении объектов реального мира.

Термин фрактал (от латинского слова *fractus* – дробный), был предложен Бенуа Мандельбротом в 1975 году для обозначения нерегулярных самоподобных математических структур. Популярная сегодня фрактальная геометрия получила свое название лишь в 1977 году благодаря его книге «The Fractal Geometry of Nature». В работах Мандельброта использованы научные результаты других ученых, работавших в этой же области (прежде всего, Пуанкаре, Кантора, Хаусдорфа). Основное определение фрактала, данное Мандельбротом, звучало так: "Фракталом называется структура, состоящая из частей, которые в каком-то смысле подобны целому".

В самом простом случае небольшая часть фрактала содержит информацию обо всем фрактале. Строгое определение самоподобных множеств было дано Дж. Хатчинсоном в 1981 году. Он назвал множество самоподобным, если оно состоит из нескольких компонент, подобных всему этому множеству, т.е. компонент, получаемых афинными преобразованиями - поворотом, сжатием и отражением исходного множества.

Однако самоподобие – это хотя и необходимое, но далеко не достаточное свойство фракталов. Ведь нельзя же, в самом деле, считать фракталом точку, или плоскость, расчерченную на клетки. Главная особенность фракталов заключается в том, что их размерность не

укладывается в привычные геометрические представления. Фракталам характерна геометрическая «изрезанность». Поэтому используется специальное понятие фрактальной размерности, введенное Ф. Хаусдорфом и А. Безиковичем. Эта размерность не соответствует привычным для нас длине, площади или объему (размерности 1, 2 или 3, соответственно). Размерность фракталов не является целым числом, характерным для привычных геометрических объектов. Вместе с тем, в большинстве случаев фракталы напоминают объекты, плотно занимающие реальное пространство, но не использующее его полностью.

В реальной жизни фрактальные объекты имеют вполне определенные границы фрактальности, в том числе и самоподобия. Тем не менее, фракталы – это очень удобная и наглядная абстракция, которая сегодня уже широко применяется при моделировании естественных процессов. При этом спектр применения фракталов постоянно расширяется, сегодня он применяется и к моделированию информационного пространства.

Один из лучших примеров проявления фракталов в природе – структура береговых линий. Действительно, на километровом отрезке побережье выглядит столь же изрезанным, как и на стокилометровом.

Опыт показывает, что длина береговой линии L зависит от масштаба l , в котором проводятся измерения, и увеличивается с уменьшением последнего по степенному закону $L = \Lambda l^{1-\alpha}$, $\Lambda = const$. Так, например, для побережья Великобритании $\alpha \approx 1.24$, то есть, так называемая фрактальная размерность береговой линии Великобритании равна 1.24.

В настоящее время информационное пространство в целом, ввиду его объемов и динамики изменения, принято рассматривать как стохастическое. Во многих моделях информационного пространства изучаются структурные связи между тематическими множествами, входящими в это пространство. При этом численные характеристики этих множеств подчиняются гиперболическому закону (с возможными степенными поправками). Сегодня в моделировании информационного пространства все чаще используется

фрактальный подход, базирующийся на свойстве самоподобия информационного пространства, т.е. сохранение внутренней структуры множеств при изменениях их размеров или масштабов их рассмотрения извне.

Самоподобие информационного пространства выражается, прежде всего, в том, что при его лавинообразном росте в последние десятилетия, частотные и ранговые распределения, получаемые в таких разрезах, как источники, авторы, тематика практически не меняют своей формы. Т.е. применение теории фракталов при анализе информационного пространства позволяет с общей позиции взглянуть на закономерности, составляющие основы информатики. Например, тематические информационные массивы сегодня представляют развивающиеся самоподобные структуры, которые по своей сути являются стохастическими фракталами, так как их самоподобие справедливо лишь на уровне математических ожиданий, например, распределения кластеров по размерам.

В информационном пространстве возникают, формируются, растут и размножаются кластеры – группы взаимосвязанных документов. Системы, основанные на кластерном анализе, самостоятельно выявляют новые признаки объектов и распределяют объекты по новым группам.

Чем же определяется природа фрактальной структуры информационного пространства, порождаемого такими кластерными структурами? С одной стороны, параметрами ранговых распределений, а с другой - механизмом развития информационных кластеров, который отражает природу информационного пространства. Появление новых публикаций увеличивает размерность уже существующих кластеров и является причиной образования новых.

Фрактальные свойства характерны для кластеров информационных Web-сайтов, на которых публикуются документы, соответствующие определенным тематикам. Эти кластеры, как наборы тематических документов, представляют собой фрактальные структуры, обладающие

рядом уникальных свойств. Например, российскими исследователями (С. Иванов и др.), определена фрактальная размерность подобных информационных массивов, изменяющаяся в пределах от 1.05 до 1.50, что свидетельствует о небольшой плотности заполнения кластеров документами по одной теме.

Как один из основных законов, отражающих самоподобие информационного пространства, можно назвать закон Ципфа. В 1949 году профессор филологии из Гарварда Дж. Ципф собрал достаточный статистический материал и экспериментально показал, что распределение слов естественного языка подчиняется закону: “Если к какому-либо достаточно большому тексту составить список всех встретившихся в нем слов, а затем ранжировать эти слова, т.е. расположить их в порядке убывания частоты встречаемости в данном тексте и пронумеровать в возрастающем порядке, то для любого слова произведения его порядкового номера (ранга) в этом списке и частоты его встречаемости в тексте будет величиной постоянной.” Ученый описал обнаруженную им закономерность распределения слов в текстах на английском языке:

- небольшое количество слов, таких как "the", "and" в английском языке, которые имеют очень высокий ранг;
- среднее количество слов имеет средний ранг;
- большое количество слов имеет очень низкий ранг.

Таким образом: $f \cdot r = c$, где f - частота встречаемости слова в тексте; r - ранг (порядковый номер) слова в списке; c - эмпирическая постоянная величина. Эту закономерность зависимости частоты от ранга называют первым законом Ципфа. То есть, зависимость количества слов с данной частотой от частоты - гипербола с постоянными параметрами для всех текстов в пределах одного языка. Значение константы в разных языках различно, но внутри одной языковой группы остается неизменным. Так, например, для английских текстов константа Ципфа равна приблизительно

0,1. Для русского и украинского языков коэффициенты Ципфа составляют приблизительно 0,06 - 0,07.

Ципф сформулировал еще одну закономерность, близкую по смыслу к своему первому закону. Он определил, что частота и количество слов, входящих в текст с этой частотой, также взаимосвязаны. Если построить диаграмму, отложив по одной оси частоту вхождения слова, а по другой - количество слов, входящих в текст с данной частотой, то получившаяся кривая будет сохранять свои параметры для всех текстов в пределах одного языка. Однако на каком бы языке текст ни был написан, форма кривой Ципфа останется неизменной – могут отличаться лишь коэффициенты. Эта закономерность получила название второго закона Ципфа - "количество - частота".

Теория фракталов тесно связана с кластерным анализом, решающим задачу выделения компактных групп объектов с близкими свойствами. Кластеризация сегодня применяется при реферировании больших документальных массивов, определении взаимосвязанных групп документов, для упрощения процесса просмотра при поиске необходимой информации, нахождения уникальных документов из коллекции, выявления дубликатов или близких по содержанию документов.

Фрактальный принцип самоподобия предполагает бесконечное дробление набора объектов с сохранением их свойств. В данном случае можно наблюдать подобие сюжетных цепочек, получаемых при уточнении запроса (конечно, в определенных рамках). Вместе с тем, сегодня многими исследователями рассматривается не дробление, а естественный рост размеров информационного пространства.

Свойства самоподобия фрагментов информационного пространства наглядно демонстрирует новый интерфейс, представленный на Web-сайте службы News Is Free (<http://newsisfree.com>). На этом сайте отображается состояние информационного пространства в виде ссылок на источники и отдельные сообщения. При этом учитываются два основных параметра

отображения – ранг популярности и «свежесть» информации. В рамках этой модели можно наблюдать «дробление» групп источников при увеличении ранга популярности и «свежести» изданий. Когда этот ранг становится достаточно высоким, дробление не позволяет без особых усилий читать названия источников и идентифицировать отдельные документы.

Пространство интернет-новостей, являясь, пожалуй, самой динамичной частью Web-пространства, характеризуется большим количеством контекстных и гипертекстовых ссылок, топология которых изложена в первой главе данной монографии.

Как и в случае модели Бредера для Web-пространства, топология и характеристики модели новостной части Интернет оказались примерно одинаковыми для различных его подмножеств, подтверждая наблюдение о том, что "информационное пространство новостей - это фрактал", т.е. свойства всей структуры этого пространства верны и для его отдельных подмножеств.

С другой стороны, информационное пространство можно рассматривать как среду, в которой возникают и развиваются кластерные структуры, которые можно изучать и моделировать, используя как методическую основу теорию фракталов.

Новостную составляющую информационного пространства Интернет можно рассматривать как мощный информационный поток [24], характеризующийся определенным набором параметров, среди которых выделяются такие, как источники информации (Web-сайт) и тематики. Именно их можно рассматривать как лежащие на поверхности основы для кластеризации [27, 42].

В то время как для традиционных средств научной коммуникации подходы к кластеризации с точки зрения теории фракталов были впервые исследованы Ван Рааном, анализирувавшим массивы статей и связи, образуемые цитированием, информационные потоки сообщений из Интернета до последнего времени не ассоциировались с фракталами, что

связано с проблемами идентификации информационных потоков как фрактальных множеств, а также с трудностью нахождения основ для построения кластеров — сообщений в политематических потоках, порождающих многократное цитирование.

По этой же причине в рамках данной работы исследуются количественные характеристики лишь тематических информационных потоков, которые характеризуются итеративностью при формировании и вполне доступны как для количественного, так и для качественного анализа.

Объемы сообщений в тематических информационных потоках образуют временные ряды. Для исследования временных рядов сегодня все шире используется теория фракталов, традиционная область применения которой — фрактальная геометрия, обработка изображений и т.п. [45]. Вместе с тем временные ряды, порождаемые тематическими информационными потоками, также обладают фрактальными свойствами [27] и могут рассматриваться как стохастические фракталы [21], [24]. Этот подход расширяет область применения теории фракталов на информационные потоки, динамика которых описывается средствами теории случайных процессов.

С другой стороны, теория фракталов рассматривается как подход к статистическому исследованию, который позволяет получать важные характеристики информационных потоков, не вдаваясь в детальный анализ их внутренней структуры и связей. Одним из основных свойств фракталов является самоподобие (скейлинг). Как показано в работах С.А. Иванова, для последовательности сообщений тематических информационных потоков в соответствии со скейлинговым принципом, количество сообщений, резонансов на события реального мира пропорционально некоторой степени количества источников информации (кластеров) и итерационно продолжается в течение определенного времени. Точно так же, как и в традиционных научных коммуникациях, растущее множество сообщений в Интернете по одной тематике во времени представляет собой динамическую

кластерную систему, возникающую в результате итерационных процессов. Этот процесс объясняется републикациями, прямой или совместной цитируемостью, различными публикациями - отражениями одних и тех же событий реального мира, прямыми ссылками и т.д. Кроме того, для большинства тематических информационных потоков наблюдается увеличение их объемов, причем на коротких временных интервалах - линейный рост, а на длительных - экспоненциальный.

Фрактальная размерность в кластерной системе, соответствующей тематическим информационным потокам, показывает степень заполнения информационного пространства сообщений в течение определенного времени:

$$N_{\text{нубл}}(\varepsilon t) = \varepsilon^{\rho} N_k(t)^{\rho}, \quad (3.29)$$

где $N_{\text{нубл}}$ - размер кластерной системы (общее число электронных публикаций в информационном потоке); N_k - размер — число кластеров (тематик или источников); ρ - фрактальная размерность информационного массива; ε - коэффициент масштабирования. В приведенном соотношении между количеством сообщений и кластеров проявляется свойство сохранения внутренней структуры множества при изменении масштабов его внешнего рассмотрения.

Изучение явлений самоподобия, применение теории фракталов при анализе информационного пространства позволяет с общей позиции взглянуть на эмпирические законы, составляющие теоретические основы информатики. Например, тематические информационные массивы сегодня представляют развивающиеся самоподобные структуры, и могут рассматриваться как стохастические фракталы [24]. Известно, что все основные законы научной коммуникации, такие как законы Парето, Лотки, Бредфорда, Ципфа, могут быть обобщены именно в рамках теории стохастических фракталов [21].

Очевидно, что резкие скачки в объемах потоков электронных публикаций по тематике выборов свидетельствуют о некоторых реальных

событиях, на которые возможна реакция соответствующих специалистов. Т.е., можно предположить, что как средство анализа электоральной ситуации само изучение параметров массивов электронных публикаций по данной теме относится к избирательным технологиям.

Как известно, возникновение детерминированного хаоса в динамике объектов тесно связано с наличием у него фрактальных свойств, важность которых в последние годы широко обсуждается в самых различных областях науки. Теория фракталов широко применяется как подход к статистическому исследованию, который позволяет получать важные характеристики информационных потоков, не вдаваясь в детальный анализ их внутренней структуры. В частности, количество тематических сообщений в Интернете, резонансов на событие реального мира, пропорционально некоторой степени количеству тематических источников (веб-сайтов). Точно так же, как и в традиционных научных коммуникациях, количество сообщений в Интернете по выбранной тематике представляет собой динамическую кластерную систему.

Как и в случае потоков энергии или вещества, проходящих через открытые системы, информационные потоки также во многих случаях обладают самоорганизацией, т.е. свойствами самоподобия, характеризующееся сильными, подчиняющимися степенному закону, корреляциями. Если рассматривать информационные потоки как ряды публикаций в течение времени, то можно воспользоваться таким определением строгого самоподобия (масштабной инвариантности, скейлинга): процесс $X(t)$ является самоподобным, если $X(t)$ и $\alpha^{-H}X(\alpha t)$ имеют одинаковые распределения вероятностей для всех $\alpha > 0$.

В предлагаемой работе исследуются временные ряды, соответствующие количеству публикаций в сети Интернет по заданной проблематике. В наблюдаемых рядах выявлено самоподобие и устойчивые взаимные корреляции. На основании обработки данных наблюдений получены

значения различных статистических показателей соответствующих рядов, а также показано, что они обладают фрактальной природой.

Исследования проводились на наборе документальных корпусов, содержащих сообщения онлайн-СМИ различных объемов, сформированные системой InfoStream [14]. Тематика исследуемого информационного потока определялась запросом к системе InfoStream: «(Киев or Київ or Україн or Украин) and (НАТО] or Євроатлант or Евроатлант)».

Ретроспективный период исследования составлял весь 2005 год, т.е. 365 дней ($N = 365$). В результате поиска по приведенному запросу было найдено 54933 релевантных документов. На основе обработки этих данных была получена полная картина экспериментальных данных - временной ряд за заданный период (рис. 50). Пиковые значения этого ряда соответствовали таким событиям: пресс-конференция Президента Украины с генеральным секретарем Североатлантического альянса в Брюсселе (февраль 2005 г.); заседание комиссии Украина-НАТО на уровне министров иностранных дел (апрель 2005 г.); выступление министра иностранных дел Украины на Форуме по безопасности СЕАП (май 2005 г.); встреча Президента Украины с Генеральным секретарем НАТО (октябрь 2005 г.); конференция «Национальная безопасность Украины в контексте евроатлантической интеграции» (ноябрь 2005 г.)

Метод DFA

Метод DFA [76, 81, 34] представляет собой вариант дисперсионного анализа, позволяющий исследовать эффекты длительных корреляций в нестационарных рядах. При этом анализируется среднеквадратическая ошибка линейной аппроксимации в зависимости от размера аппроксимируемого участка.

В рамках этого алгоритма определения DFA вначале осуществляется приведение данных к нулевому среднему (вычитание среднего значения $\langle F \rangle$ из временного ряда F_n ($n = 1, \dots, N$)) и строится случайное блуждание $y(k)$:

$$y(k) = \sum_{n=1}^N [F(n) - \langle F \rangle_N]. \quad (3.30)$$

Затем ряд значений $y(k)$, $k = 1, \dots, N$ разбивается на неперекрывающиеся отрезки (участки) длины n , в пределах каждого из которых методом наименьших квадратов определяется уравнение прямой, аппроксимирующей последовательность $y(k)$.

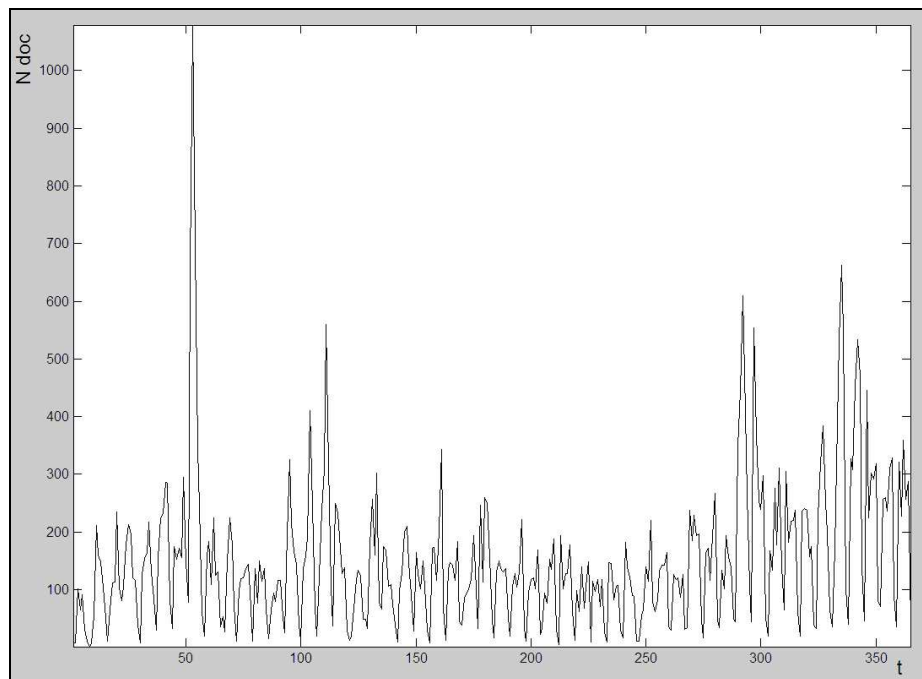


Рис. 50. Временной ряд динамики появления понятия (ось Y) за весь период (ось X)

Найденная аппроксимация $y_n(k)$ ($y_n(k) = ak + b$) рассматривается в качестве локального тренда. При этом коэффициенты a и b вычисляются следующим стандартным образом:

$$a = \frac{n \sum ky(k) - (\sum k)(\sum y(k))}{n \sum k^2 - (\sum k)^2};$$

$$b = \frac{(\sum y(k))(\sum k^2) - (\sum k)(\sum ky(k))}{n \sum k^2 - (\sum k)^2}.$$
(3.31)

Далее вычисляется среднеквадратическая ошибка линейной аппроксимации в широком диапазоне значений n . Считается, что зависимость $D(n)$ часто имеет степенной характер $D(n) \sim n^\alpha$, т.е. наличие линейного участка в двойном логарифмическом масштабе $\lg D(\lg n)$ позволяет говорить о существовании скейлинга. При использовании метода DFA для различных участков ряда наблюдений одинаковой длины n исследуемой последовательности строится линейная аппроксимация, для которой затем вычисляется среднеквадратичная ошибка $D(n)$.

Как видно по рис. 51, $D(n)$ степенным образом зависит от n , т.е. в двойном логарифмическом масштабе эта зависимость близка к линейной, что позволяет говорить о наличии скейлинга [34].

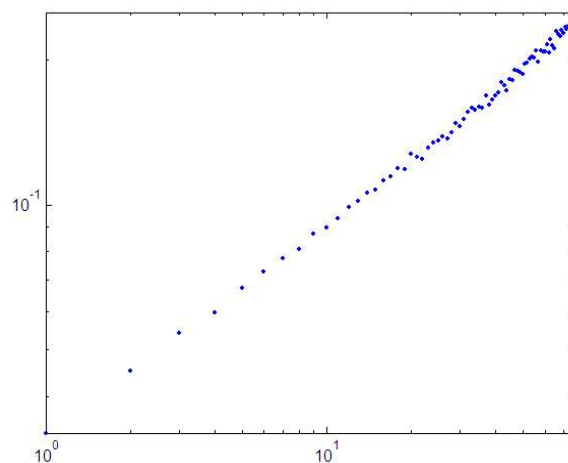


Рис. 51. Зависимость $D(n)$ ряда наблюдений (ось Y) от длины участка аппроксимации n (ось X) в логарифмической шкале

Коэффициенты корреляции

Как известно, коэффициенты корреляции для ряда измерений рассчитываются по формуле:

$$R(k) = \langle (X_{k+t} - m)(X_k - m) \rangle / \sigma^2,$$

где $R(k)$ - коэффициент корреляции; X_k - ряд измерений; m – его математическое ожидание X_k ; σ^2 – дисперсия.

Графическое представление коэффициента корреляции для исследуемого ряда наблюдений свидетельствует о разделении корреляционных свойств по дням недели (рис. 52). Вместе с тем, коэффициенты корреляции ряда наблюдений, усредненного по неделям, аппроксимируются гиперболической функцией, что свидетельствует о долгосрочной зависимости исходного ряда (рис. 53).

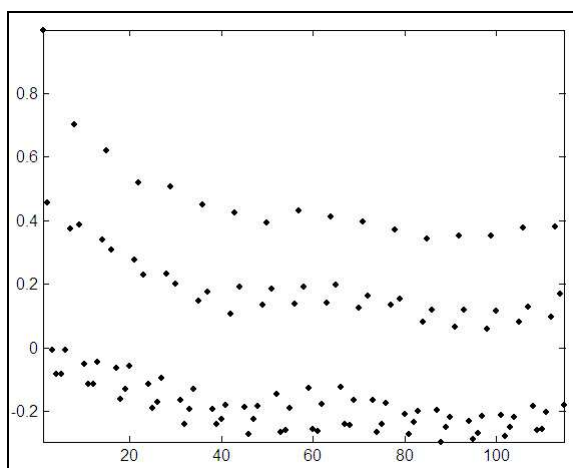


Рис. 52. Коэффициенты корреляции ряда наблюдений

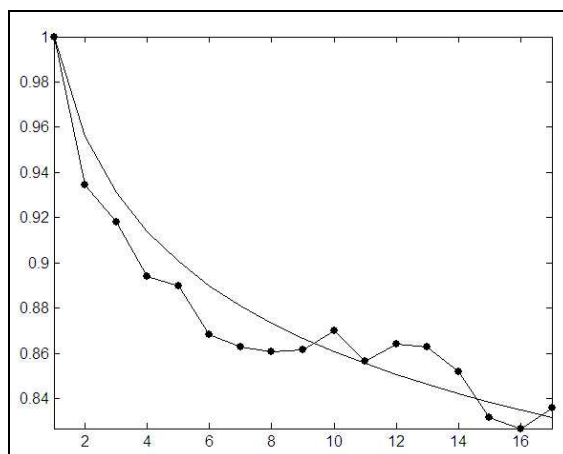


Рис. 53. Коэффициенты корреляции ряда наблюдений, усредненного по неделям

Фактор Фано

Для изучения поведения процессов и подтверждения их самоподобия принято использовать еще один показатель – индекс разброса дисперсии (IDC), так называемый, фактор Фано. Эта величина определяется как отношение дисперсии числа событий (в нашем случае – числа публикаций) временного ряда на заданном окне наблюдений k к соответствующему математическому ожиданию:

$$F(k) = \sigma^2(k)/m(k).$$

Для самоподобных процессов выполняется соотношение:

$$F(k) = 1 + Ck^{2H-1},$$

где C и H – константы. На рис. 54 приведен график значений $F(k)$ в двойном логарифмическом масштабе.

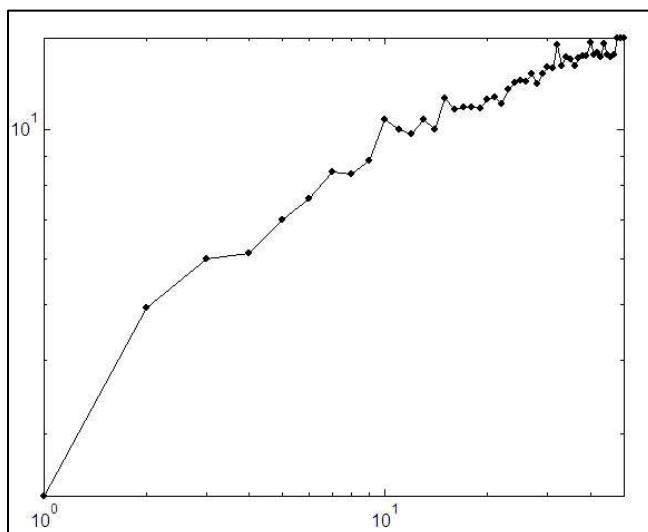


Рис. 54. Зависимость фактора Фано от ширины окна наблюдений

Показатель Херста

Основной характеристикой рядов, обладающих хаотическим поведением, является, как известно, показатель Херста [45]. Для его определения воспользуемся т. н. R/S-анализом, который успешно применялся авторами ранее в исследованиях фрактальной природы научных

коммуникаций и информационных потоков [28]. Он позволяет достаточно эффективно исследовать свойства числовых рядов на основе отношения разброса значений к среднеквадратичному отклонению.

Сегодня, в связи с развитием теории стохастических фракталов, становится популярной такая характеристика временных рядов как показатель Херста (H). В [45] показано, что он связан с традиционной «клеточной» фрактальной размерностью (Θ) простым соотношением:

$$\Theta = 2 - H. \quad (3.32)$$

Условие, при котором показатель Херста связан с фрактальной «клеточной» размерностью в соответствии с формулой (3.32), определено Е. Федером следующим образом: «... рассматривают клетки, размеры которых малы по сравнению как с длительностью процесса, так и с диапазоном изменения функции; поэтому соотношение справедливо, когда структура кривой, описывающая фрактальную функцию, исследуется с высоким разрешением, т.е. в локальном пределе». Еще одним важным условием является самоаффинность функции. Не вдаваясь в подробности, заметим, что для информационных потоков это свойство интерпретируется как самоподобие, возникающее в результате процессов их формирования. Можно отметить, что указанными свойствами обладают не все информационные потоки, а лишь те, которые характеризуются достаточной мощностью и итеративностью при формировании. При этом временные ряды, построенные на основании мощных тематических информационных потоков, вполне удовлетворяют этому условию. Поэтому при расчете показателя Херста фактически определяется и такой показатель тематического информационного потока как фрактальная размерность.

Известно, что показатель Херста представляет собой меру персистентности — склонности процесса к трендам (в отличие от обычного броуновского движения). Значение $H > 1/2$ означает, что направленная в определенную сторону динамика процесса в прошлом, скорее всего, повлечет продолжение движения в том же направлении. Если $H < 1/2$, то

прогнозируется, что процесс изменит направленность. $H = 1/2$ означает неопределенность — броуновское движение.

Для изучения фрактальных характеристик тематических информационных потоков изучались значения показателя Херста за определенный период для временных рядов, составленных из количества относящихся к ним сообщений. Согласно [45] он определяется из соотношения:

$$\frac{R}{S} = \left(\frac{N}{2} \right)^H, N \gg 1. \quad (3.33)$$

Здесь S – стандартное отклонение:

$$S = \sqrt{\frac{1}{N} \sum_{n=1}^N (F(n) - \langle F \rangle_N)^2}, \quad (3.34)$$

$$\langle F \rangle_N = \frac{1}{N} \sum_{n=1}^N F(n), \quad (3.35)$$

а R , так называемый размах:

$$R(N) = \max_{1 \leq n \leq N} X(n, N) - \min_{1 \leq n \leq N} X(n, N), \quad (3.36)$$

где

$$X(n, N) = \frac{1}{N} \sum_{n=1}^N [F(n) - \langle F \rangle_N]. \quad (3.37)$$

Были проведены исследования фрактальных свойств информационных потоков, для чего использовался документальный корпус системы мониторинга новостей из Интернет InfoStream [14]. Рассматривались ряды, соответствующие количеству публикаций в разрезе дат. На рис. 55 показан процесс вычисления показателя Херста H . При увеличении n показатель H принимает значения $0.75 \div 0.85$. Ввиду того, что значение H намного превышает $1/2$, статистика исследуемого ряда отличается от гауссовой и в этом ряду обнаруживается персистентность (существование долговременных корреляций, которые могут быть связаны с существованием детерминированного хаоса). Если считать функцию $F(n)$ самоаффинной

(этот вопрос требует отдельного исследования), то в соответствии с [45] функция $F(n)$ имеет фрактальную размерность D , равную

$$D = 2 - H \approx 1.25 \div 1.15. \quad (3.38)$$

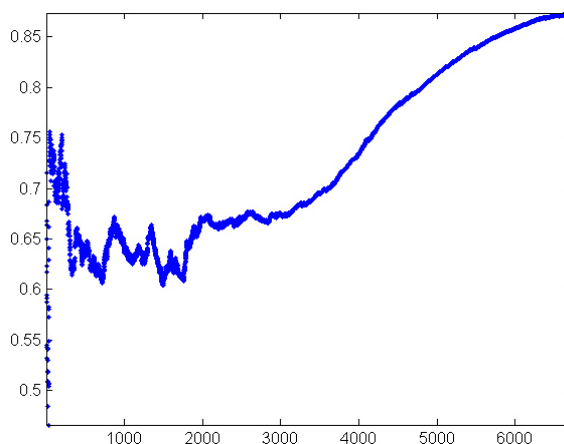


Рис. 55. Значения показателя Херста (ось Y) в зависимости от объема исследуемого массива (ось X)

Таким образом, проведенные исследования тематических информационных потоков подтвердили предположение о самоподобии и итеративности процессов в информационном пространстве. Републикации, цитирование, прямые ссылки и т.п. порождают самоподобие, проявляющееся в устойчивых статистических распределениях и известных эмпирических законах.

В результате эксперимента было подтверждено наличие высокого уровня статистической корреляции в информационных потоках на продолжительных временных интервалах. На рассмотренном примере показана высокая персистентность процесса, что, в частности, свидетельствует об общей тенденции увеличения публикации по тематике евроатлантической интеграции.

Анализ самоподобия информационных массивов, таким образом, может рассматриваться как технология, предназначенная для осуществления

аналитических исследований с элементами прогнозирования, способная к экстраполяции полученных зависимостей.

3.6. Анализ стабильности информационных источников

Один из возможных подходов к решению проблемы изучения сетевого информационного пространства основан на представлении его некоторым множеством источников, порождающих информационные потоки. Предполагается, что динамика этих потоков в определенном смысле более содержательна, чем динамика составляющих их сообщений.

При этом можно отметить разнообразный диапазон параметров этих источников как по объемам публикуемой информации, так и по содержанию - от сообщений серьезных информационных агентств - до «живых журналов» школьников.

Источники информации, очевидно, характеризуются уровнем стабильности. Примером стабильных источников могут служить крупные информационные агентства, регулярно поставляющие потребителям примерно одинаковые объемы информации на протяжении длительного времени, а примером нестабильных – «живые журналы», многие из которых активно действуют в течение нескольких дней, а затем угасают.

Нестабильные источники по-своему интересны хотя бы тем, что, видимо, именно они ответственны за хаотичность динамической части сетевого информационного пространства. Однако они не связаны с его основными тенденциями и поэтому могут не приниматься в расчет при его систематических исследованиях. Напротив, ключевую роль здесь должны играть именно стабильные источники, отражающие (и в какой-то мере порождающие) реальные закономерности сетевой динамики.

На практике среди множества проблем подбора и анализа источников контента большое значение имеет учет параметров их стабильности, в частности, тематической. При этом тематическая стабильность и

стабильность потока информации от источников зачастую играют решающую роль при проведении аналитических исследований. Например, такие важные свойства информационных источников, как их тематическая корреляция и полнота, имеет смысл учитывать только для источников, публикующих документы относительно стабильной тематической направленности.

Тематическую стабильность источника можно определить как корреляцию наборов тематических рубрик, которым соответствуют документы из этого источника в различные периоды времени. Предполагается, что конкретный набор рубрик мало влияет на предлагаемый ниже метод расчета стабильности источников (под тематической рубрикой в данном случае понимается тематика, семантика которой, в частности, находит свое отражение в виде запроса на информационно-поисковом языке). Предполагается, что документу присваивается та или иная рубрика, если он соответствует определенному запросу. Перечень рубрик и соответствующих им запросов был выбран авторами на основании опыта работы с политематическими новостными ресурсами сети Интернет. Эти рубрики и запросы установлены и апробированы в течение длительного времени в системе контент-мониторинга InfoStream. В настоящее время система включает 35 основных тематических рубрик.

При исследовании тематической направленности некоторых источников информации были обнаружены документы, отклоняющиеся от основной направленности этих источников. Такие документы, если их количество относительно невелико, не должны влиять на рассчитываемый ниже уровень стабильности источников. Конечно, автоматическая рубрикация во многом зависит от качества запросов, однако некоторыми погрешностями в рубрикации при статистическом исследовании можно пренебречь.

Для подхода к изучению стабильности источников важно знать параметры их распределения по тематическим рубрикам, т.е. количество

рубрик, соответствующих документам, входящим в эти источники. Результаты такого исследования, охватывающего 920 репрезентативных русскоязычных источников (опубликовавших за месяц более 100 сообщений), приведены на рис. 56. Об относительно невысокой тематической стабильности источников, порождающих общий информационный поток системы, свидетельствует тот факт, что около половины репрезентативных источников соответствуют более 20 рубрикам.

Для вычисления уровня разброса (нестабильности) источника информации использовалась формула, основанная на линейной метрике:

$$R = \frac{1}{N} \sum_{i=1}^N \frac{1}{M \cdot \max(r_i)} \sum_{j=1}^M |r_{ij} - \frac{1}{M} \sum_{k=1}^M r_{ik}| \quad (3.39)$$

где N – количество рубрик, M – количество дней, $\max(r_i)$ – максимальное суточное количество вхождений рубрики i в документы источника за все время, r_{ij} – количество вхождения рубрики i за день j .

Из приведенной формулы следует, что значение R , на самом деле, учитывает не только тематический разброс, но и разброс по количеству вхождений рубрики, т.е. фактическое количество документов от источника, относящихся к данной рубрике. Даже если источник соответствует одной рубрике, но его наполнение не является стабильным, значение может существенно отличаться от нулевого.

Ранговая диаграмма распределения уровня разброса для источников – Web-сайтов, ежедневно публикующих новостные сообщения в ноябре 2005 года, по уровням тематической стабильности приведена на рис. 57.

Определение стабильности документов выполнялось по такому алгоритму:

1. Проводился поиск документов в базе данных за определенный период.
2. Формировалась таблица, которая включала код источника информации, коды соответствующих ему тематических рубрик и их количество в разрезе дат.
3. Для каждого источника по приведенной выше формуле определялся

уровень разброса R .

4. Информационные источники ранжировались по рассчитанным параметрам, и строилась соответствующая диаграмма.

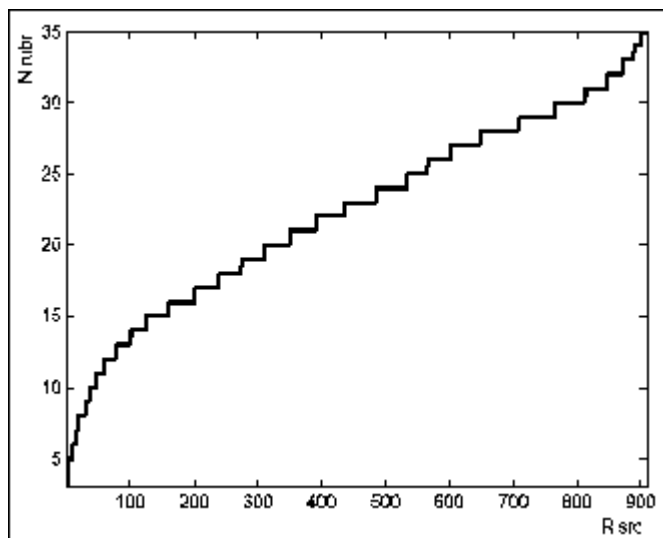


Рис. 56. Ранговая диаграмма “Источники – количество рубрик”

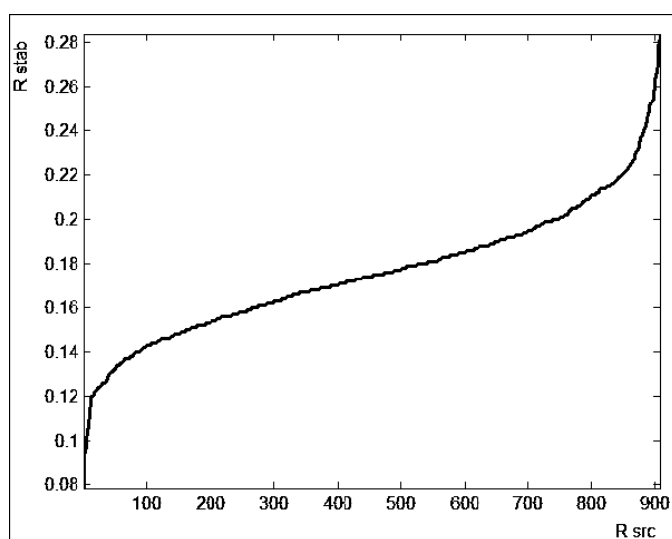


Рис. 57. Диаграмма “Ранг источника - коэффициент разброса”

Как оказалось, источники, содержащие до 5-6 рубрик, обладают исключительной стабильностью, что, в общем, достаточно очевидно. Не совсем очевидным оказался факт резкого всплеска разброса для источников, включающих документы с 25 и более рубриками.

В приведенной для вычисления уровня разброса формуле используется линейная метрика, которая позволяет достаточно точно дифференцировать источники при минимальных затратах на вычисления. Необходимо отметить, что для анализа временных рядов сегодня также широко используются вычисления, базирующиеся на евклидовой метрике. В частности, популярный в экономике подход R/S-анализа позволяет исследовать «изрезанность» кривой, образуемой временным рядом на основе отношения разброса значений к среднеквадратичному отклонению. Очевидно, что «изрезанность» кривой близка по смыслу фрактальной размерности. Заметим, что в то время как фрактальная природа научных коммуникаций исследовалась в работах Ван Раана [78] и С.А. Иванова [21], новостные информационные потоки в сети Интернет до настоящего времени не рассматривались в научных публикациях с этой точки зрения.

Аналог формулы (3.39) – мера «изрезанности» распределения временного ряда при фрактальном походе R_f выглядит следующим образом:

$$R_f = \frac{1}{N} \sum_{i=1}^N \frac{S_i}{R_i}, \quad (3.40)$$

где S_i – среднеквадратичное отклонение по рубрике i :

$$S_i = \sqrt{\frac{1}{M} \sum_{j=1}^M \left\{ r_{ij} - \frac{1}{M} \sum_{k=1}^M r_{ik} \right\}^2}; \quad (3.41)$$

R_i – размах значений по рубрике i :

$$R_i = \max_{1 \leq k \leq M} X_{ik} - \min_{1 \leq k \leq M} X_{ik};$$

X_{ik} – накопленное к моменту k отклонение по рубрике i :

$$X_{ik} = \sum_{j=1}^k \left\{ r_{ij} - \frac{1}{M} \sum_{l=1}^M r_{il} \right\}. \quad (3.42)$$

В соответствии с формулой (3.40) коэффициент «изрезанности» временного ряда количества публикаций по N темам одного источника выглядит следующим образом:

$$R_f = \frac{1}{N} \sum_{i=1}^N \left(\frac{2}{M}\right)^{H_i} = \frac{1}{N} \sum_{i=1}^N \left(\frac{2}{M}\right)^{1-\rho_i}. \quad (3.43)$$

На рис. 58 представлена кривая значений коэффициентов «изрезанности» для источников (было измерено поведение 1700 источников за 2005 год), ранжированных по этим значениям. Как видим, характер кривой вполне соответствует характеру кривой разброса источников. Более того, названные коэффициенты очень близки по своей природе.

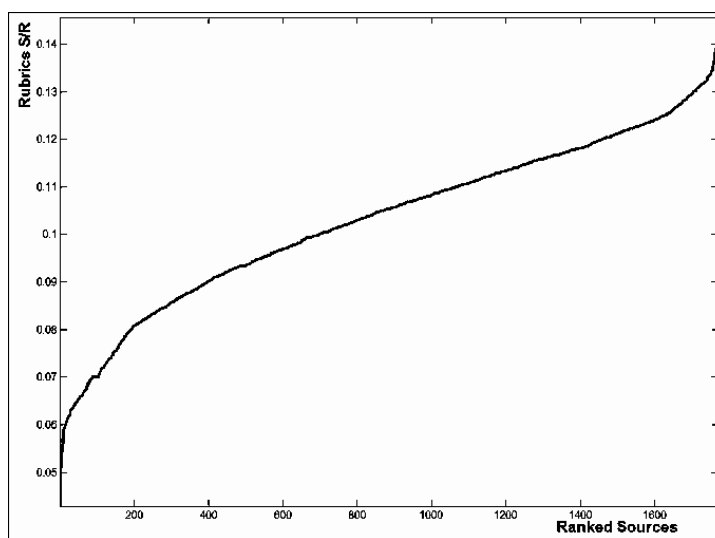


Рис. 58. Диаграмма «Ранг источника - коэффициент «изрезанности»»

Коэффициент Херста для источника i можно вычислить по формуле:

$$H_i = \log(R_i/S_i)/\log(M/2). \quad (3.44)$$

На рис. 59 приведена совместная диаграмма коэффициентов «изрезанности» и коэффициентов Херста.

Из фрактального анализа известно, что коэффициент Херста, равный $1/2$, соответствует броуновскому движению, т.е. случайному поведению временного ряда по тематикам, чем значение коэффициента Херста ближе к единице, тем процесс тематических публикаций более детерминирован, т.е. персистентен. Это означает, что если количество публикаций растет, то можно предположить, что оно будет возрастать и в дальнейшем, а если уменьшается, то и в дальнейшем будет наблюдаться спад. Значение коэффициента Херста меньшее $1/2$ свидетельствует об антиперсистентности

или эргодичности процесса. Данное свойство означает, что если количество публикаций растет, то можно предположить, что оно в дальнейшем пойдет на спад, а если уменьшается, то и в дальнейшем будет наблюдаться рост количества публикаций. Конечно, такие источники можно оценивать как крайне тематически нестабильные.

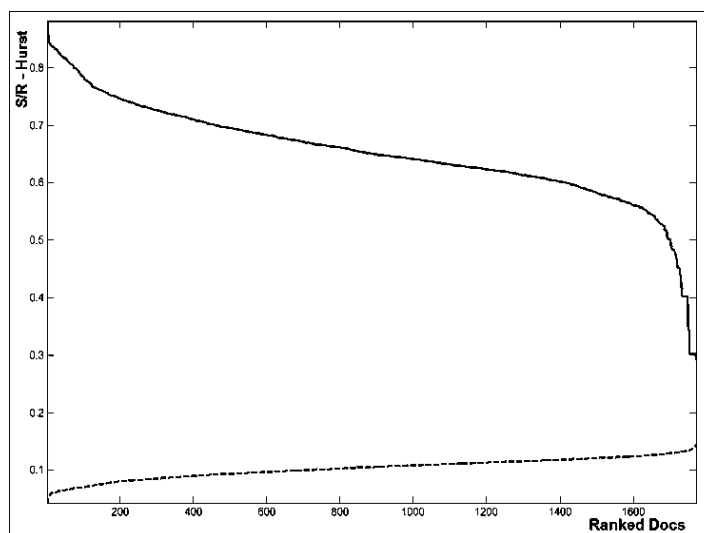


Рис. 59. Диаграмма “Ранг источника – коэффициент тематической «изрезанности» и коэффициент Херста”

При этом рассматривались «усредненные» по количеству тематических рубрик значения коэффициента Херста:

$$H = \frac{1}{N} \sum_{i=1}^N H_i, \quad (3.45)$$

Как оказалось, имеет большой практический смысл рассматривать стабильность процесса поступления информации от отдельных источников даже вне тематического разреза. Параметр «изрезанности» кривой количества поступлений от источников вне тематического разреза, конечно, более простой, однако он также может быть определяющим при выборе списка источников абонентами системы контент-мониторинга.

В этом случае:

$$R_f = \frac{S}{R} = \left(\frac{2}{M}\right)^H, \quad (3.46)$$

где S - среднеквадратичное отклонение, R – размах, H – коэффициент Херста.

На рис. 60 представлена ранговая диаграмма внетематического распределения источников по коэффициентам «изрезанности». Как видно, форма графиков отличается от представленных на рис. 56 практически отсутствием одной из точек перегиба, а значения показателя Херста в среднем существенно выше, что свидетельствует об очевидном факте: электронные издания более склонны изменять тематику своих публикаций, чем их объемы, выраженные количеством публикаций.

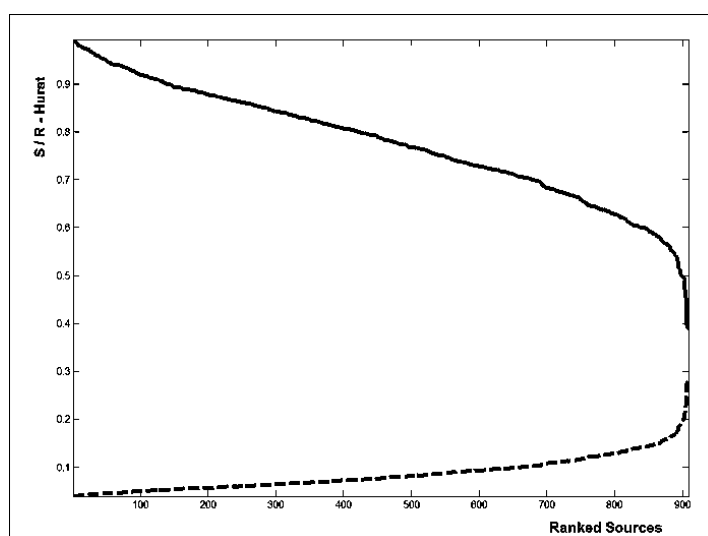


Рис. 60. Диаграмма «Ранг источника – коэффициент внетематической «изрезанности» и коэффициент «Херста»»

Результаты исследований стабильности источников могут использоваться при ранжировании выдачи информационно-поисковых систем, подсчете медиа-рейтингов, позволяют рекомендовать пользователям наиболее тематически стабильные и оригинальные источники информации, например, для включения их в список «Персональных информационных источников» в интерфейсе системы контент-мониторинга информационных ресурсов.

Сегодня становится ясно, что разработка качественно новых средств работы с сетевыми ресурсами переходит в разряд приоритетных задач. Без приемлемых средств контроля за сетевыми информационными процессами невозможно обеспечить репрезентативность выборок. В любом случае,

успешное продвижение в изучении современного информационного пространства невозможно без хотя бы общих представлений о структуре и свойствах динамики сетевых информационных процессов, что в свою очередь требует выявления их устойчивых закономерностей.

3.7. Объектно-статистический анализ информационных потоков

Динамика и постоянно увеличивающиеся объемы разноплановых публикаций в Интернет обуславливают проблему получения данных для информационно-аналитических исследований, как оперативных, так и ретроспективных по различным тематическим направлениям. Обычными, традиционными, методами поиска и экстрагирования информации, необходимой для последующей обработки, уже не обойтись. На помощь могут прийти лишь системы контент-мониторинга, охватывающие тысячи информационных ресурсов, и позволяющие выявлять тенденции, сюжеты, объекты и их связи [26]. Вместе с тем анализ процессов, которые имеют довольно значительные временные рамки, все еще ждет своего инструментария. Если вопросы визуализации результатов поиска информационного отображения подобных процессов освещаются в большом количестве работ [12, 13, 25, 68, 69], то анализу и визуализации объектного распределения отобранных информационных массивов больших объемов до сих пор не уделялось существенного внимания.

Предметной областью данного исследования является анализ и визуализация объектного распределения отобранных информационных массивов на примере анализа динамики публикаций в Интернет-пространстве о деятельности системы избирательных комиссий в Украине по выборам Президента Украины и народных депутатов Украины за 2004-2006 годы. Эта динамика отражает реальный интерес общественности, через электронные

средства информации, к избирательным процедурам, а также процессы, происходящие в ходе избирательных кампаний.

Система контент-мониторинга InfoStream позволила построить зависимость суточных объемов тематических публикаций за 3 года (1096 суток, общее количество – свыше 320 тысяч). Пики на графике (рис. 61), позволяют оценить интенсивность освещения в прессе как президентской избирательной кампании 2004 г., так выборов в Верховный Совет Украины в 2006 г.

Вместе с тем для более детального анализа процессов, общепринятыми методиками является анализ Фурье и вэйвлет-анализ [53, 16, 17]. Технология использования вэйвлетов (маленьких волн) позволяет выявлять одиночные и нерегулярные всплески, резкие изменения значений количественных показателей в различные периоды времени, в частности, объемов тематических публикаций в Интернет. При этом могут выявляться моменты возникновения циклов, а также когда за периодами регулярной динамики следуют хаотические колебания. Метод вэйвлет-анализа используется также для декомпозиции, выделения сигнала из "шума", изучения динамики различных процессов, в том числе экономических и социальных. На рис. 62 приведена спектограмма - результат вэйвлет-анализа временного ряда, соответствующего изучаемому процессу.

Прекрасно отражая спектральные характеристики сигналов, вэйвлет-анализ, однако, по своей природе, не может быть использован, когда информационный поток следует рассматривать с объектной точки зрения. В случае, рассматриваемом авторами, такими объектами выступали отдельные лица, определяемые в публикациях своими фамилиями, инициалами, должностями и т.п. В частности, с помощью средств экстрагирования информации системы InfoStream из рассматриваемого потока было выявлено упоминание о более чем 40 тыс. лицах, в той или иной мере имеющих отношение к избирательному процессу. Как промежуточный результат

исследования был выявлен тот факт, что ранговое распределение фамилий с большой точностью удовлетворяют закономерности Ципфа (рис. 63).

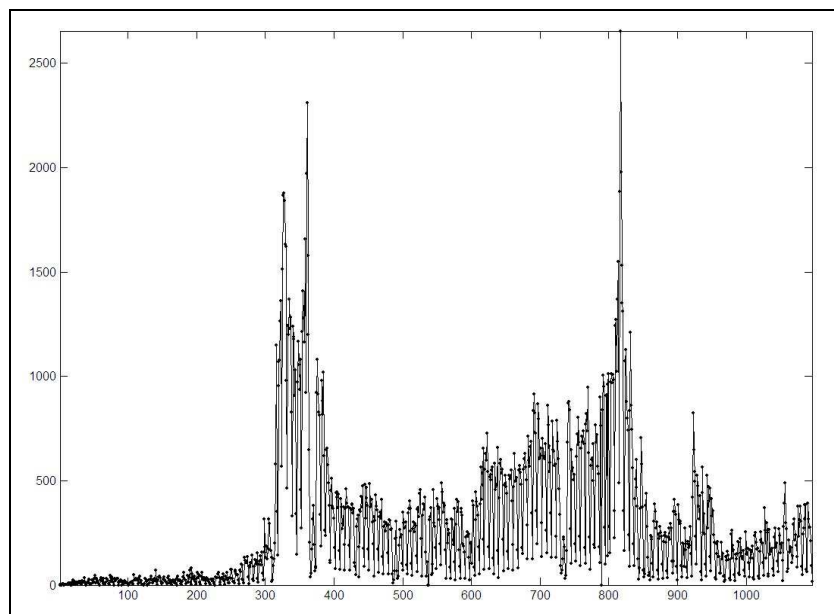


Рис. 61. Количество тематических публикаций (ось Y) по дням (ось X)

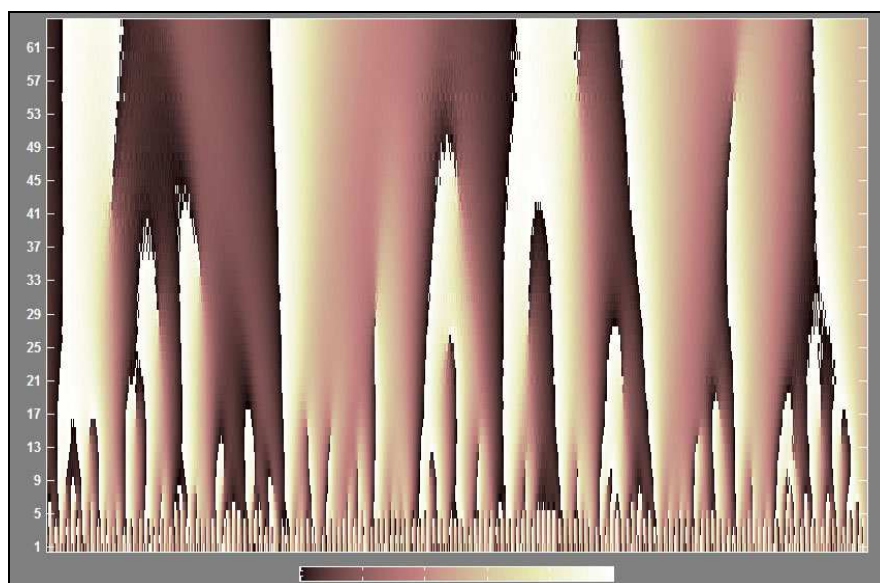


Рис. 62. Вэйвлет-спектограмма динамики тематического информационного потока (одномерное непрерывное вэйвлет-преобразование, вэйвлет Гаусса), ось X – дни, ось Y – частоты

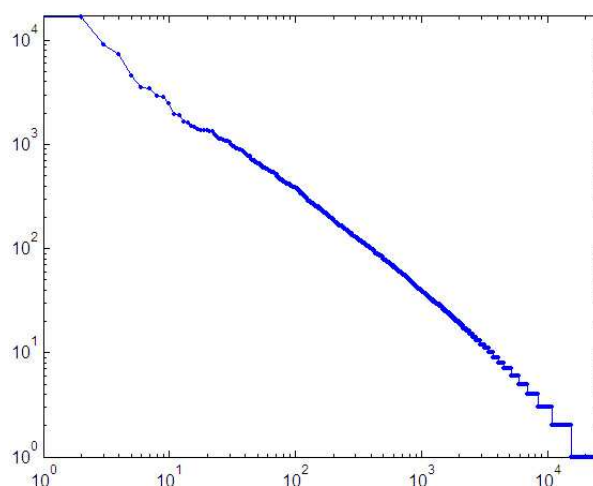


Рис. 63. Ранжированный список фамилий в текстах сообщений тематического информационного массива в двойном логарифмическом масштабе (ось X – номер фамилии, ось Y – частота встречаемости)

В экстрагированном виде каждая персона представлялась одним дескриптором. Для обеспечения учета и анализа распределения информационных потоков в разрезе интересующих персон был предложен оригинальный метод, так называемых вордлет-диаграмм. Эти диаграммы представляют собой форму визуального отображения информационного потока в разрезе объектов и дат, представляющую собой прямоугольную таблицу, ячейки которой заполнены значениями количества сообщений информационного потока за определенную дату, соответствующих определенному объекту. Столбцам этой таблицы соответствуют даты, а строкам – объекты, являющиеся своеобразными содержательными фильтрами исследуемого информационного потока. Объектам в рассматриваемом случае соответствуют определенные лица. Естественно, для визуального отображения из множества персон выбирается лишь несколько десятков интересующих исследователя.

Визуально вордлет-диаграмма представляет собой таблицу, ячейки которой закрашены оттенками серого цвета, в зависимости от значений объемов публикации по выбранному объекту в соответствующий день (большее значение соответствует более темному оттенку). Следует заметить, что многие строки вордлет-диаграммы обладают фрактальными свойствами,

которые присущи им как количественным индикаторам тематических информационных потоков. В частности, для аналогичных временных рядов было экспериментально подтверждено наличие статистической корреляции на достаточно длительных интервалах [27].

Вордлет-диаграммы для относительно небольшого количества строк (несколько десятков) позволяют визуально выявлять группы наиболее связанных по датам и интенсивностям публикаций объектов. Для большего количества объектов в процессе построения вордлет-диаграммы предлагается ее кластеризация путем перестановки строк (перегруппировки объектов) в соответствии с алгоритмом *k-means* [72]. При этом подразумевается, что, например, в случае разбиения на 2 кластера, основы кластеров i и j (центроиды), которые затем рекурсивно уточняются, выбираются для наибольших значений функции:

$$F_{ij} = \sum x_{ik} * \sum x_{jk} * R_{ij},$$

где x_{ik} – значения элементов таблицы, суммы берутся по всем датам k ; R_{ij} – «расстояние» между строками i и j , определяемое формулой:

$$R_{ij} = \sum |x_{ik} - x_{jk}|.$$

Следует отметить, что кроме названной выше тематической задачи, были получены вордлет-диаграммы, соответствующие большим информационным потокам различной тематической направленности. В качестве параметров запросов для отбора объектов выбирались такие параметры, как ключевые слова, фамилии, географические названия, названия организаций.

На рис. 64 приведена вордлет-диаграмма первого уровня (превью), позволяющая визуально выявлять аномальные корреляции. На этой диаграмме, охватывающей информацию по 49 персонам, отчетливо видны циклы праздничных дней, а также корреляции отдельных объектов. С помощью приведенной на рис. 65 уточняющей вордлет-диаграммы можно точно указать выявленные корреляции, например, персон с номерами 10, 11 и 26, 27 за последние 20 дней.

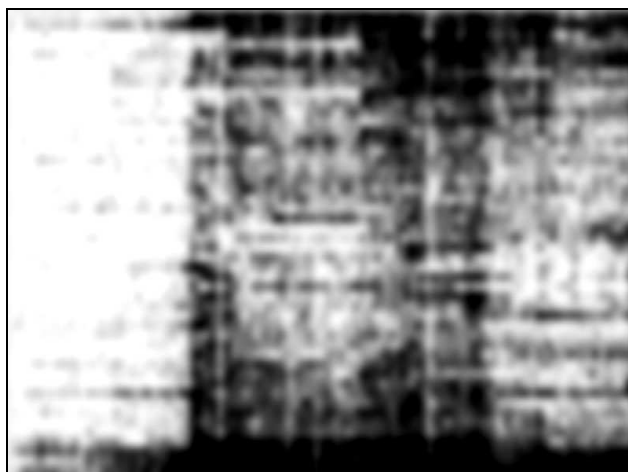


Рис. 64. Вордлет-диаграмма-превью (ось X – дни, ось Y – персоны)

В результате проведенных экспериментов есть основания предположить, что использование таких средств визуализации, как вордлет-диаграммы, позволяет «разлагать» исходные временные ряды в соответствии с объектами, обнаруживать медиа-активность по выбранным объектам, выявлять взаимосвязи объектов в разрезе дат, определять детали медиа-активности каждого объекта или группы объектов. Вордлет-диаграммы позволяют более адекватно анализировать динамику публикаций в Интернете в разрезе интересующих объектов, предоставляя в наглядном виде важную информацию о динамике реальных процессов. Использование вордлет-диаграмм представляется важным дополнением к уже признанным методам исследований, таким как анализ Фурье, корреляционный и фрактальный анализ, а также вейвлет-анализ.

Необходимо отметить, что представленный подход к решению вопроса анализа и визуализации объектного распределения отобранных информационных массивов, несмотря на то, что он продемонстрирован на примере анализа динамики публикаций в Интернет-пространстве о деятельности системы избирательных комиссий в Украине по выборам Президента Украины и народных депутатов Украины, носит общий характер.

Данный подход применим для решения вопросов анализа и визуализации объектного распределения любых отобранных информационных массивов для процессов, которые интересуют исследователя и имеют довольно значительные временные рамки.

Кроме того, необходимо отметить, что представленный подход анализа информационных потоков носит объектно-статистический характер, который, в свою очередь, представляется как существенная составляющая методологической базы прогнозно-эмпирического анализа.

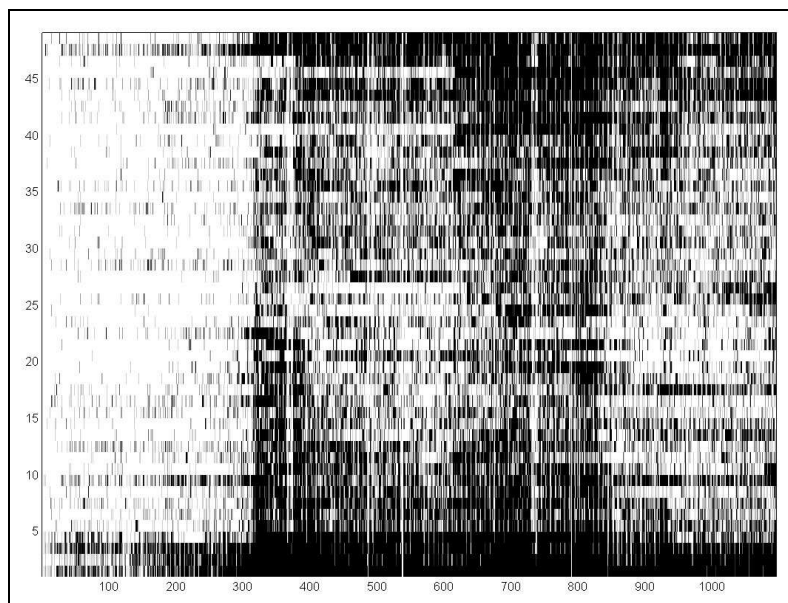


Рис. 65. Уточняющая вордлет-диаграмма (ось X – дни, ось Y – персоны)

ЗАКЛЮЧЕНИЕ

Итак, рассмотрено несколько достаточно компактных моделей, которые в некотором приближении позволяют описывать информационно-электоральные процессы.

Модели в том виде, в котором они приведены в данной работе, пригодны для описания общих тенденций в динамике информационно-электоральных процессов. Полная картина может быть получена с учетом дополнительного набора факторов, большинство которых являются случайными и потому не воспроизводятся во времени. В какой мере необходим их явный учет – зависит в первую очередь от поставленной задачи. Структура основных правил, лежащих в основе большинства из рассмотренных моделей, позволяет вносить соответствующие коррективы, например, искусственно моделируя случайные отклонения.

Вместе с тем отметим, что воспроизведение результатов во времени, вообще говоря, является в данном случае крайне серьезной проблемой. Действительно, повторяемость моделируемых явления составляет основу научной методологии. В динамике информационно-электоральных процессов точная повторяемость если и встречается, то крайне редко. Поэтому в нашем распоряжении в данный момент не оказалось надежного способа верификации результатов. По-видимому, дальнейшие исследования внесут в этот вопрос большую ясность.

ЛИТЕРАТУРА

- [1] В.И. Арнольд. Аналитика и прогнозирование: математический аспект. // Научно-техническая информация. Сер. 1. Вып. 3. - 2003. - С. 1-10.
- [2] В.И. Арнольд. Обыкновенные дифференциальные уравнения. М.: Наука, 1971, - 240 с.
- [3] О. А. Ахметов, М. Б. Мжельский. Метод анализа иерархий как составная часть методологии проведения оценки недвижимости // Актуальные вопросы оценочной деятельности. – 2001 г. (<http://www.nsk.su/~estate/articles/art001.html>)
- [4] Богданович В.Ю. Воєнна безпека України: методологія дослідження та шляхи забезпечення. - Київ, 2003.- 322 с.
- [5] Богданович В.Ю., Маначинский А.Я. Методологические основы системных исследований проблем военной безопасности государства. - Киев, 2001. - 172 с.
- [6] Брайчевский С.М., Ландэ Д.В. Современные информационные потоки: актуальная проблематика. // Научно-техническая информация. Сер. 1. Вып. 11. – 2005. - С. 21-33.
- [7] Вентцель Е. С. Элементы теории игр. Изд. 2. М., Физматгиз, 1961. 68 с.
- [8] Вольтерра В. Математическая теория борьбы за существование. М.: Наука, 1976.
- [9] Гарднер М. Математические досуги. – М.: Мир, 1972.
- [10] Гаузе Г. Ф. Борьба за существование. – М: УРСС, 2002. - 160 с.
- [11] Гаузе Г. Ф. Экология и некоторые проблемы происхождения видов. В кн.: Экология и эволюционная теория. — Л., Наука, 1984, с. 5–108.
- [12] Григорьев А.Н., Ландэ Д.В. Многоуровневый классификатор-навигатор по откликам информационно-поисковой системы // Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции Диалог'2006 – Москва: Наука, 2006. - С. 329-331.
- [13] Григорьев А.Н., Ландэ Д.В. Адаптивный интерфейс уточнения запросов

- к системе контент-мониторинга InfoStream // Труды Международного семинара «Диалог'2005». – М.: Наука, 2005. – С. 109-111.
- [14] Григорьев А.Н., Ландэ Д.В. и др. Мониторинг новостей из Интернет: технология, система, сервис: научно-методическое пособие. – К.: ООО «Старт-98», 2007. – 40 с.
- [15] Гринченко В.Т., Мацыпура В.Т., Снарский А.А. Введение в нелинейную динамику. Хаос и фракталы. Изд. 2.–М: УРСС, 2007.- 263 с.
- [16] Давыдов. А. А. Системная социология. –М: КомКнига, 2006. - 192 с.
- [17] Давыдов А.А. Вейвлет-анализ социальных процессов//Социолог.исслед. 2003, №11, С. 97-103.
- [18] Давыденко В.А., Ромашкина Г.Ф., Чуканов С.Н. Моделирование социальных сетей // Вестник Тюменского государственного университета. № 1, 2005. С. 68-79.
- [19] Додонов А.Г., Ландэ Д.В. Организация сети информационных прокси-серверов. // Реєстрація, зберігання і обробка даних, 2006, Т. 8, № 3.– С. 24 - 31.
- [20] Додонов А.Г., Ландэ Д.В. Самоподобие массивов сетевых публикаций по компьютерной вирусологии. // Реєстрація, зберігання і обробка даних, 2007, Т. 9, № 2.– С. 53 - 60.
- [21] Иванов С.А. Стохастические фракталы в Информатике // Научно-техническая информация. Сер. 2, 2002. - № 8. - С. 7-18.
- [22] Интриллигатор М. Математические методы оптимизации и экономическая теория. - М.: Прогресс. – 1975. – 607 с.
- [23] Коковин С.Г. Лекции по теории игр и политологии. Часть 1. Введение в теорию игр. 66 стр. (<http://www.math.nsc.ru/~mathecon/Kokovin/m1tigran.pdf>)
- [24] Ландэ Д.В. Основы интеграции информационных потоков - К.: Инжиниринг, 2006. - 240 с. (<http://dwl.kiev.ua/art/monogr-osnov/spusk3.pdf>)
- [25] Ландэ Д.В. Присмотритесь внимательнее или "Изюминки" поисковой визуализации // hiTech Pro - К., 2006. - декабрь, - С. 94-95.

(<http://dwl.kiev.ua/art/hitech/>)

- [26] Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа - М.: ИД Вильямс, 2005. - 271 с.
- [27] Ландэ Д.В. Фрактальные свойства тематических информационных потоков из Интернет // Реєстрація, зберігання і обробка даних, 2006, Т. 8, № 2.– С. 93 - 99.
- [28] Ландэ Д.В., Фурашев В.Н., Брайчевский С.М., Григорьев А.Н. Основы моделирования и оценки электронных информационных потоков - К.: Инжиниринг, 2006. - 176 с.
- [29] Ландэ Д.В, Фурашев В.Н. Объектно-статистический анализ информационных потоков. - Открытые информационные и компьютерные технологии. Харьков: НАКУ «ХАИ», 2007. – Вып.35 –с.133-137.
- [30] Ландэ Д.В., Фурашев В.Н. Моделирование электоральных процессов на основе концепции клеточных автоматов. - Открытые информационные и компьютерные технологии. Харьков: НАКУ «ХАИ», 2007. – Вып.36 –с.17-34.
- [31] Лоскутов А.Ю., Михайлов А.С. Введение в синергетику. М.: Наука, 1990. - 272 с.
- [32] Нейман Дж. Теория самовоспроизводящихся автоматов - М.: Мир, 1971. – 382 с.
- [33] Остапенко В.В., Остапенко О.С., Подладчикова Т.В. Оптимизация стратегии политических партий во время предвыборной кампании // Тезисы конференции «Вычислительная и прикладная математика». – Киев, 24-25 сентября 2004 г. – С. 119.
- [34] Павлов А.Н., Сосновцева О.В., Зиганшин А.Р., Мультифрактальный анализ хаотической динамики взаимодействующих систем // Изв. вузов, Прикладная нелинейная динамика, т. 11, No. 2, стр. 39-54 (2003).
- [35] Перцев Н.В. Применение одного дифференциального уравнения с последствием в моделях динамики популяций // Фундаментальная и прикладная математика / Под ред. А.К. Гуца: Сб. науч. тр. Омск, 1994. С.

- [36] Плотинский Ю.М. Модели социальных процессов. – Изд. 2-е. –М.: Логос, 2001. – 296 с.
- [37] Плотинский Ю.М. Математическое моделирование динамики социальных процессов. М., Изд-во МГУ. 1992.
- [38] А. Прохоров. Социальные сети и Интернет // КомпьютерПресс. - № 10. – 2007.
- [39] Ф.С. Робертс. Дискретные математические модели с приложениями к социальным и экологическим задачам. М. 1986.
- [40] Саати Т. Принятие решений. Метод анализа иерархий. М.: Радио и связь, 1993. - 278 стр.
- [41] Саати Т., Крис К. Аналитическое планирование. Организация систем: Пер. С англ. – М. Радио и связь, 1991. – 224 с.
- [42] Снарский А.А., Ландэ Д.В. и др. Ранжирование сайтов «по Хиршу». // Доклады международной конференции «MegaLing'2006 Горизонты прикладной лингвистики и лингвистических технологий». - С. 248-249.
- [43] Снарский А.А., Безсуднов И.В., Севрюков В.А. Процессы переноса в макроскопических неупорядоченных средах. –М.: УРСС, 2007. – 264 с.
- [44] Сорокин П.А. Классификация методов индивидуум-ориентированного моделирования. // Электронный журнал «ИССЛЕДОВАНО В РОССИИ». - 2003, - С. 574-588 (<http://zhurnal.ape.relarn.ru/articles/2003/050.pdf>)
- [45] Федер Е. Фракталы -М.: Мир, 1991. - 254 с.
- [46] Экономико-математические методы и прикладные модели / Под ред. В.В. Федосеева. Издание 2. –М.: Юнити, 2005. - 304 с.
- [47] Форрестер Дж. Мировая динамика. - М.: Наука , 1978, 268 с.
- [48] Фурашев В.М., Коваль М.І., Маглюй С.А. Системна інформатизація виборчого і референдумного процесів в Україні. - Монографія. Київ: Парламентське видавництво, 2004. – с.607.
- [49] Фурашев В.М. Деякі особливості застосування системи автоматизованого інформаційно-аналітичного забезпечення процесу проведення виборів

Президента України у 2004 році. - Открытые информационные и компьютерные технологии. Харьков: НАКУ «ХАИ», 2005. – Вып.27 – с.125-132.

- [50] Фурашев В.М. Перспективи подальшого розвитку системної інформатизації виборчих і референдумних процесів в Україні.- Наук. журнал “Правова інформатика”. Київ:АПрН, НДЦПІ, 2005 – 4(8)/2005 – с.13-17
- [51] Фурашев В.Н, Ландэ Д.В., Брайчевский С.М. Системная информатизация избирательных и референдумных процессов: методологические основы статистических исследований электронных информационных ресурсов в период избирательной кампании. - Открытые информационные и компьютерные технологии. Харьков: НАКУ «ХАИ», 2005. – Вып.29 – с.11-15.
- [52] Фурашев В.М. Нормативно-правові засади системної інформатизації інформаційно-аналітичного забезпечення здійснення процедур виборчих і референдумних процесів. - Монографія. Київ: Парламентське видавництво, 2006. – 144 с.
- [53] Чуи К. Введение в вэйвлеты.М.: Мир, 2001.
- [54] Чурсин Н.Н. Популярная информатика. -К.: Техника, 1982. -158 с.
- [55] Albert R., Jeong H., Barabási A. Attack and error tolerance of complex networks // Nature. - 2000. - Vol. 406, pp. 378–382.
- [56] Axelrod R. The Evolution of Cooperation. New York: Basic Books, 1984.
- [57] Bertalanffy L. von, An outline of general system theory, «British Journal for the Philosophy of Science», 1950, v. I, № 2.
- [58] S.C. Bhargava, A. Kumar, A. Mukherjee. A stochastic cellular automata model of innovation diffusion // Technological forecasting and social change. – 1993. Vol. 44. № 1. - P. 87-97.
- [59] Bjorneborn, L., Ingwersen, P. Toward a basic framework for webometrics. Journal of the American Society for Information Science and Technology, 55(14): 1216-1227. - 2004.

- [60] Brown T.A. Nonlinear Politics // Chaos Theory in the Social Sciences / Eds. L.D. Kiel, E. Elliot. Ann Arbor.: The Univ. Of Michigan Press. 1996. P. 119-137.
- [61] Burton R.E. and Kebler R.W. The "half-life" of some scientific and technical literatures. American Documentation 1960;1:98—109.
- [62] Casti J.L. Searching for Certainty. N.Y.: W.Morrow, 1990.
- [63] Luca Donetti, Pablo I. Hurtado, Miguel A. Munoz. Entangled Networks, Synchronization, and Optimal Network Topology // "Physical Review Letters" - Vol. 95, No. 18, Art. 188701, 28.10.2005.
- [64] P. Erdős, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hungar. Acad. Sci. 5 (1960) 17-61.
- [65] Gibbons, R. A Primer in Game Theory. Harvester Wheatsheaf, 1992.
- [66] Gianna M. Del Corso, Antonio Gullí, Francesco Romani. Ranking a stream of news. Proceedings of the 14th international conference on World Wide Web. Chiba, Japan. – 2005. - P. 97 - 106.
- [67] Grimm, V., Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? Ecological Modelling, Vol. 115 (2-3) (2002) pp. 129-148, 1999.
- [68] Z. Junliang, Javed M., Himansu T. Information Retrieval by Semantic Analysis and Visualization of the Concept Space of D-Lib[®] Magazine // D-Lib Magazine October 2002 Volume 8 Number 10.
- [69] M.M. Knepper, R. Killam, K.L. Fox O. Frieder. Information Retrieval and Visualization using SENTINEL / TREC 1998: 336-340.
- [70] Y.W. Lee, T.P. Cheatham, J.B. Wiesner, Application of correlation analysis to the detection of periodic signals in noise, PJRE 38, 1165, 1950.
- [71] Malthus T.R. An essay on the principal of Population, as it affects the future improvement of society. - 1798 (<http://etext.lib.virginia.edu/toc/modeng/public/MalPopu.html>).
- [72] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on

- Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297.
- [73] Milgram S. The small world problem, *Psychology Today*, 1967, Vol. 2, pp. 60–67.
- [74] Newman M.E.J. The structure and function of complex networks. // *SIAM Review*. - 2003. - Vol. 45. pp. 167–256.
- [75] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. "Structure and tie strengths in mobile communication networks." *Proceedings of the National Academy of Sciences*. May 1, 2007, vol. 104. no. 18, 7332-7336.
- [76] C.-K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series // *CHAOS*. 1995. Vol. 5, P. 82.
- [77] Pearl R. *The Introduction to Medical Biometry and Statistics*. Philadelphia, 1930; *Ibid*. *The Natural History of Population*. L., 1939.
- [78] Van Raan A. F. J. Fractal geometry of information space as represented by cocitation clustering // *Scientometrics*. -1991. – Vol. 20, № 3. - P. 439-449.
- [79] Nima Sarshar, P.Oscar Boykin, Vwani P. Roychowdhury. Scalable Percolation Search in Power Law Networks. Preprint. – 2004. (<http://arxiv.org/abs/cond-mat/0406152>)
- [80] Stackelberg H. von. *Marktform und Gleichgewicht*. Faksimile der 1934 in Wien und Berlin erschienenen Erstausgabe. Reihe: Klassiker der Nationalökonomie VI, 142 S., Gebunden.
- [81] H.E. Stanley, L.A.N. Amaral, A.L. Goldberger, S. Havlin, P.Ch. Ivanov, C.-K. Peng, Statistical physics and physiology: monofractal and multifractal approaches.// *Physica A*. 1999. Vol. 270, P. 309.
- [82] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, Taylor and Francis, London 1994.
- [83] Verhulst P.F. Notice sur la loi que la population suit dans son accroissement *Corr. Math. Et Phys.* 10, 113-121, 18.

- [84] Watts D.J., Strogatz S.H. Collective dynamics of "small-world" networks. // Nature. - 1998. - Vol. 393. pp. 440–442.
- [85] S.Wolfram ed. Theory and Applications of Cellular Automats. Singapore: World Scientific. 1986.
- [86] S. Wolfram. A New Kind of Science. Champaign, IL: Wolfram Media Inc., 2002. – 1197 pp.
- [87] Shi Zhou and Raul J. Mondragon. Topological Discrepancies Among Internet Measurements Using Different Sampling Methodologies, Lecture Notes in Computer Science (LNCS), Springer-Verlag, no. 3391, pp. 207-217, Feb. 2005.