БОЛЬШИЕ ДАННЫЕ
И
БОЛЬШОЕ УПРАВЛЕНИЕ

Д.А. Новиков

(Институт проблем управления РАН)

dan@ipu.ru

БОЛЬШИЕ ДАННЫЕ: ОПРЕДЕЛЕНИЕ И ПРИМЕРЫ

Большие данные (Big Data*) в информационных технологиях – направление в науке и практике, связанное с разработкой и применением методов и средств оперирования большими объемами неструктурированных данных.

*Nature. – 2008. September 3 (Special Issue).

20 терабайт в час телеметрии с одного авиадвигателя

ИНТЕРНЕТ ВЕЩЕЙ - По оценке Oracle, общее количество «умных машин» к 2020 году достигнет 50 миллиардов

По прогнозам IDC, один из основных факторов роста - увеличение доли автоматически генерируемых данных: с 11% их общего объема в 2005 году до более чем 40% в 2020-м.

Walmart осуществляет более 1 млн. транзакций каждый час. База данных по объему в сотни раз больше объема библиотеки Конгресса США

Google

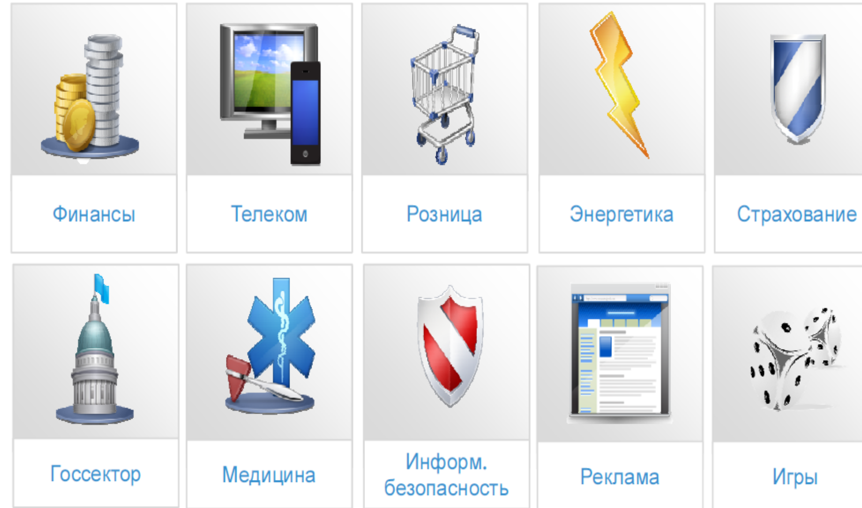
1,000,000,000
Запросов в день

Телескоп, используемый Sloan Digital Sky Survey, в первые несколько недель работы собрал больше данных, чем было собрано за всю историю астрономии

facebook

250,000,000
Новых фотографий в день

ОСНОВНЫЕ ПОТРЕБИТЕЛИ BIG DATA



– **наука** – астрономия и астрофизика, метеорология, ядерная физика, физика высоких энергий, геоинформационные и навигационные системы, дистанционное зондирование Земли, геология и геофизика, аэродинамика и гидродинамика, генетика, биохимия и биология и др.

– Интернет (в широком смысле) и другие **телекоммуникационные системы**;

– **бизнес, торговля и финансы**, а также **маркетинг и реклама** (включая трейдинг, таргетирование и рекомендательные системы, CRM-системы, RFID – радиочастотные идентификаторы, все чаще используемые в торговле, транспорте и логистике и др.);

– **мониторинг** (гео-, био-, эко-; космический, авиа- и др.);

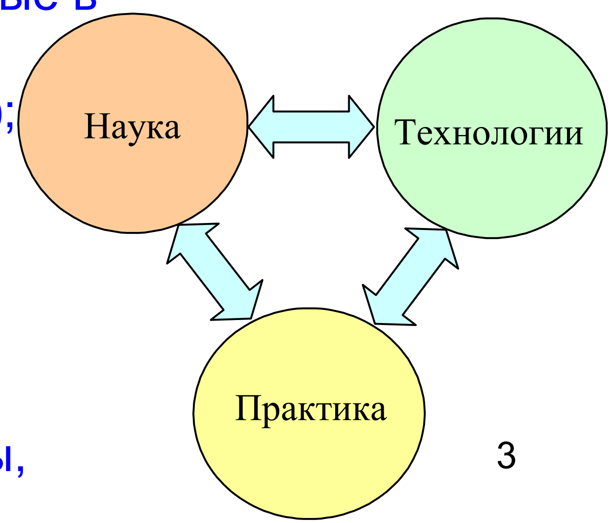
– **безопасность** (системы военного назначения, антитеррористическая деятельность и др.);

– **электроэнергетика** (включая атомную), SmartGrid;

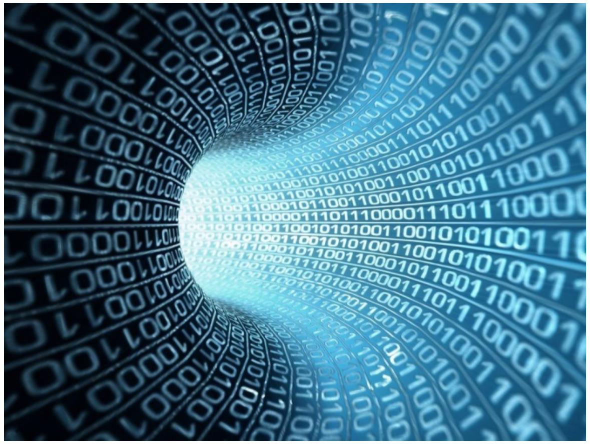
– **медицина**;

– **госуслуги** и **госуправление**;

– **производство** и **транспорт** (объекты, узлы и агрегаты, системы управления и др.).



«СТРАШИЛКИ»

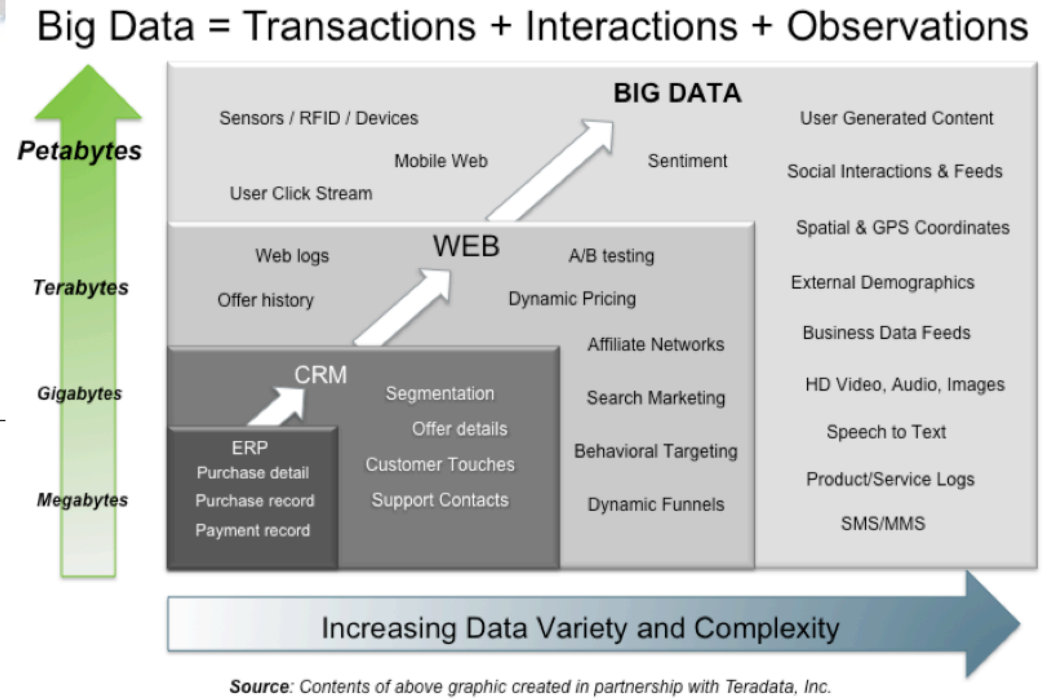


Office of Science and Technology Policy
Executive Office of the President
New Executive Office Building
Washington, DC 20502

FOR IMMEDIATE RELEASE
March 29, 2012

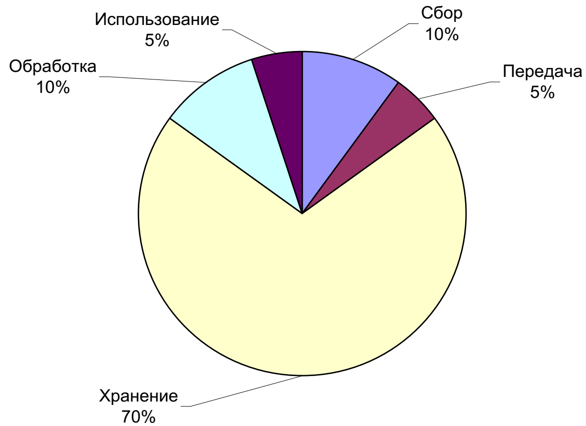
Contact: Rick Weiss 202 456-6037 rweiss@ostp.eop.gov
Lisa-Joy Zgorski 703 292-8311 lisajoy@nsf.gov

OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE:
ANNOUNCES $200 MILLION IN NEW R&D INVESTMENTS

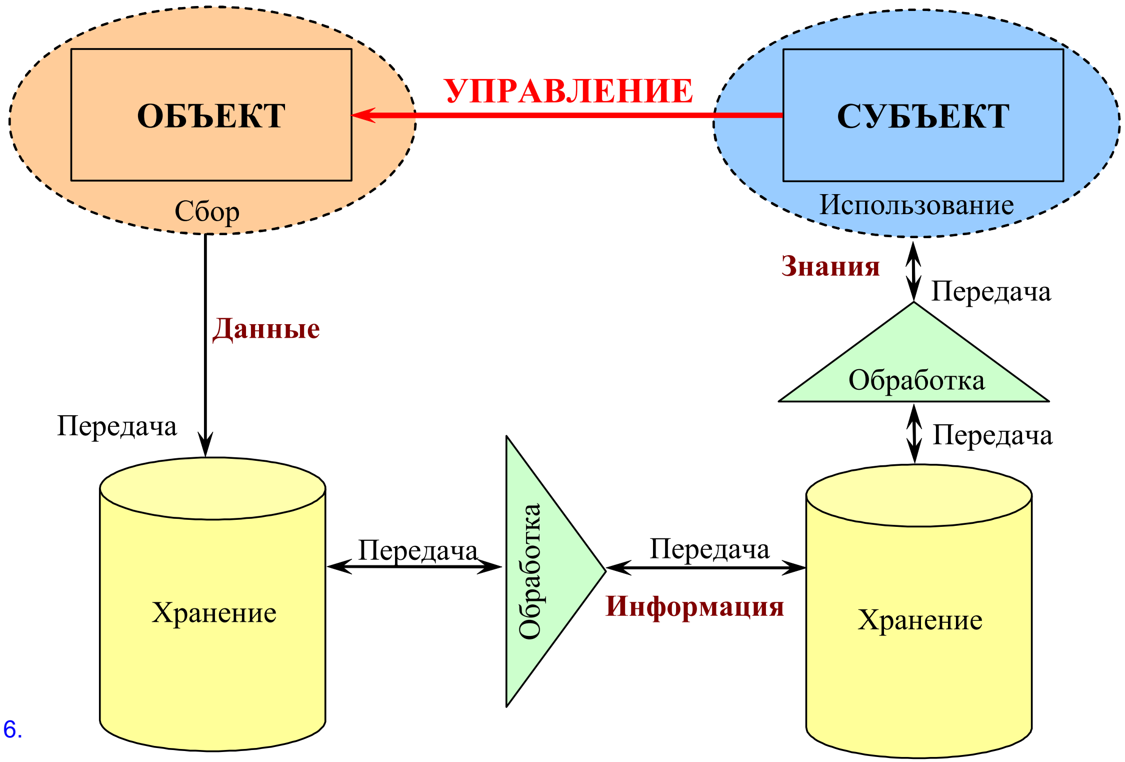


ОПЕРИРОВАНИЕ БОЛЬШИМИ ДАННЫМИ*

Оперирование Big Data включает их:
– сбор (получение);
– передачу;
– хранение (включая запись и извлечение);
– обработку (преобразование, моделирование, вычисления и анализ данных);
– использование (включая визуализацию) в практической, научной, образовательной и других видах человеческой деятельности.



Текущее распределение внимания исследователей и разработчиков к проблемам оперирования Big Data

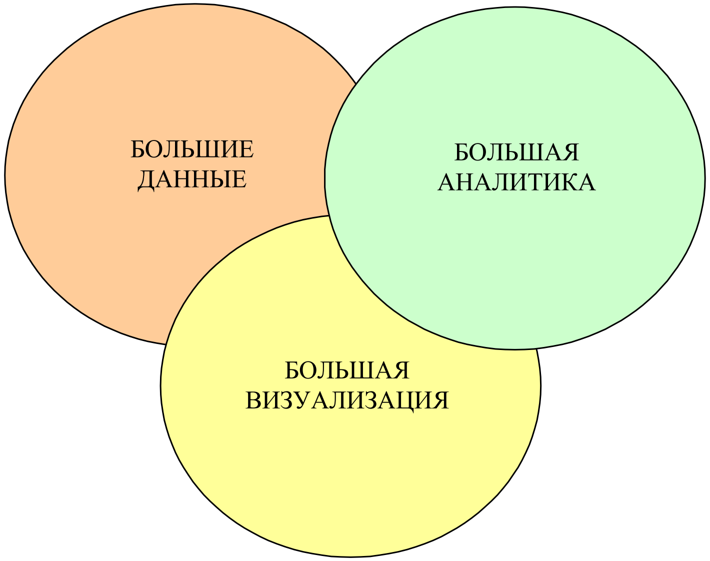


*Новиков Д.А. // Проблемы управления. 2013. № 6.

БОЛЬШИЕ: ДАННЫЕ, АНАЛИТИКА, ВИЗУАЛИЗАЦИЯ И **УПРАВЛЕНИЕ**

«Большие триады»:
– большие данные, большая аналитика, большая визуализация
– большие данные, высокопроизводительные вычисления, облачные технологии

«Большое управление»*:
– управление на основе больших данных, большой аналитики и большой визуализации**
– управление процессами оперирования Big Data



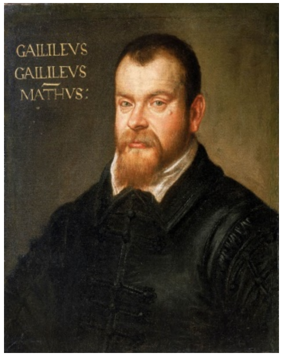
Следует иметь в виду, что специалисты по теории управления в последние полтора десятилетия все чаще говорят о совместном решении задач управления, вычислений и связи – так называемая **проблема C^3 (Control, Computation, Communication) – решения задач синтеза управляющих воздействий в реальном времени с учетом задержек в каналах связи и временных затрат на обработку информации (включая вычисления). Кроме того, существует устойчивое словосочетание «**управление большими системами**» (Large-scale Systems Control), однако большие данные могут порождаться и «маленькими» в этом смысле системами.*

**Novikov D. // Advances in Systems Studies and Applications. 2014.

ПРИМЕР: ЗАКОН ВСЕМИРНОГО ТЯГОТЕНИЯ

Тихо Браге (1546 – 1601) в течение двух десятилетий регулярно наблюдал за движением планет Солнечной системы. Его записи представляют собой Большие (по тем временам) данные.

Иоганн Кеплер (1571 – 1630), на основе данных Браге, сформулировал свои эмпирические (!) законы движения планет:
– каждая планета Солнечной системы обращается по эллипсу, в одном из фокусов которого находится Солнце;
– каждая планета движется в плоскости, проходящей через центр Солнца, причём за равные промежутки времени радиус-вектор, соединяющий Солнце и планету, описывает равные площади;
– квадраты периодов обращения планет вокруг Солнца относятся, как кубы больших полуосей орбит планет.



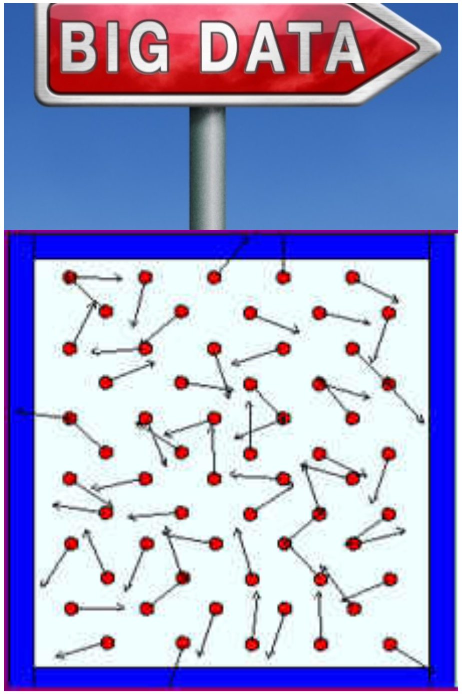
Три закона Кеплера *агрегировали* информацию Браге, и движение каждой конкретной планеты могло быть рассчитано по ним (а не по многотомным записям Браге) с высокой точностью. Другими словами, Браге научился *описывать* движение планет; Кеплер – *описывать* и *предсказывать* это движение. Но законы Кеплера ничего не говорят о том, *почему* планеты движутся в соответствии с этими законами.

Ответ на этот вопрос (т. е. объяснение) дал закон всемирного тяготения **Исаака Ньютона** (1643 – 1727).

ПРИМЕР: МОЛЕКУЛЯРНО-КИНЕТИЧЕСКАЯ ТЕОРИЯ

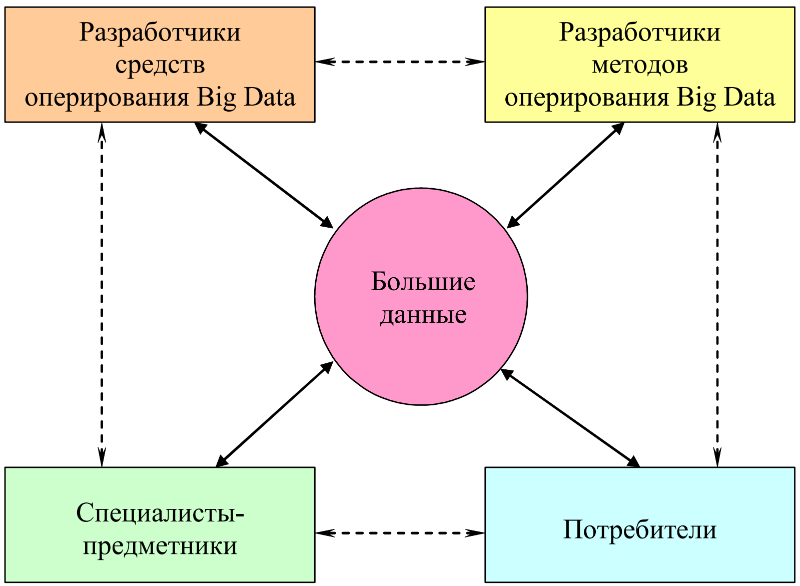
Рассмотрим следующий мысленный эксперимент – задачу детального описания поведения идеального газа. Предположим, что при нормальных условиях находится один кубический метр воздуха. В нем содержится примерно **10^25 молекул**, движение каждой из которых и их соударения исчерпывающе (в рамках модели идеального газа) описываются кинематикой и динамикой, т. е. с точки зрения физики никаких принципиальных проблем описания их движения и взаимодействия не существует. Каждая молекула испытывает порядка **10^9 столкновений/сек** с другими молекулами. Описание поведения такой системы (координаты и скорости всех молекул) в реальном времени породит поток данных не менее **10^35 байт/с**. Такой поток данных превосходит технологические возможности человечества даже сегодня!

Породить большой поток данных – не проблема, вопрос в том, что мы хотим с этими данными делать и на какие вопросы отвечать!



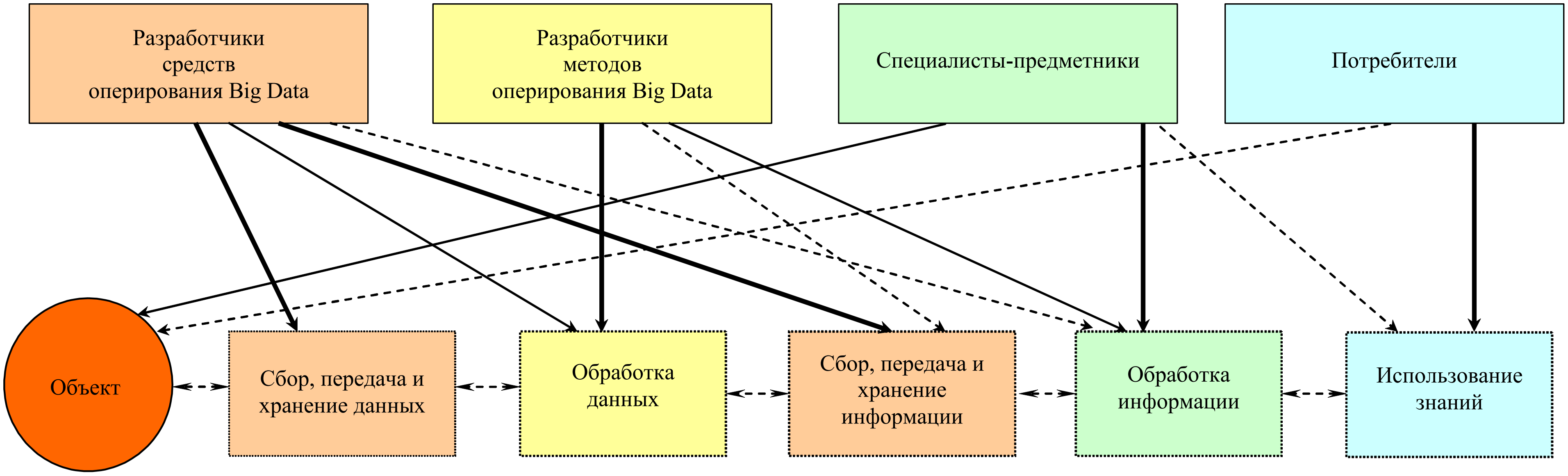
$$pV = \frac{m}{M} RT$$

СУБЪЕКТЫ И РАЗДЕЛЕНИЕ ЗОН ОТВЕТСТВЕННОСТИ



Природа устроена просто. Надо лишь уметь находить надёжные средства раскрытия этой осложнённой подробностями простоты.

Э. Резерфорд



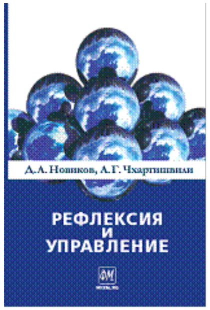
НЕКОТОРЫЕ ОПАСНОСТИ

✓ ***Информационная безопасность Big Data***. Здесь потребуется и адаптация известных, и разработка принципиально новых методов и средств.

✓ ***Энергетическая эффективность Big Data***. Уже сейчас центры обработки данных представляют собой существенный класс потребителей электроэнергии. Чем больше данных мы хотим обрабатывать, тем больше потребуется энергии.

✓ ***Принцип дополнительности*** давно известен в физике – измерение изменяет состояние системы. А как обстоит дело в социальных системах, элементы которых (люди) активны – обладают своими интересами и предпочтениями, способны самостоятельно выбирать свои действия и пр.? Одно из проявлений заключается в так называемом ***манипулировании информацией***. Снимаются или усугубляются эти и подобные (примеры – краудсорсинг, конформное поведение и др.) проблемы в области Big Data?

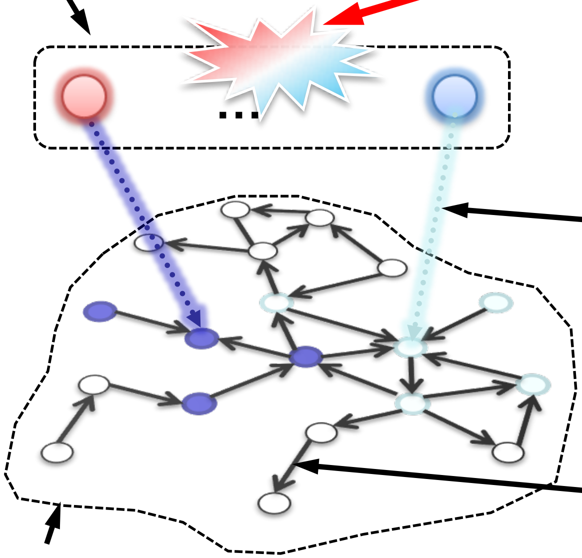
✓ «***Принцип неопределенности***» в следующем (гносеологическом) варианте: текущий уровень развития науки характеризуется определенными совместными ограничениями на «обоснованность» результатов и их области применимости. Применительно к Big Data это означает, что существует рациональный баланс между степенью детальности описания состояния интересующей нас системы и обоснованностью тех результатов и выводов, которые мы хотим сделать на основании этого описания.



✓ Традиционно при построении и эксплуатации информационных систем (будь то корпоративные системы или системы поддержки госуслуг, межведомственного документооборота и т. п.) считается, что содержащаяся в них информация должна быть максимально полной, унифицированной и общедоступной (с учетом разделения прав доступа). Но ведь возможно показывать каждому пользователю реальность, искаженную в своем «кривом зеркале» – создавать для каждого свою индивидуальную информационную картину, осуществляя тем самым ***информационное управление***. Стремиться к этому или бороться с этим в области Big Data?

МАТЕМАТИЧЕСКИЕ МОДЕЛИ АКТИВНЫХ СЕТЕВЫХ СТРУКТУР

Множество управляющих субъектов *M*



Множество управляемых субъектов *N*

5. Информационное противоборство

Каждый игрок из множества M имеет возможность влиять на начальные мнения агентов u_{ij} и заинтересован в формировании итоговых мнений X_M.

Задача – найти равновесные действия игроков в игре

$$\Gamma = (M, \{U_j\}_{j \in M}, \{G_j(\cdot)\}_{j \in M}).$$

4. Информационное управление (оптимальное)

Задача – найти такой вектор управлений u, что:

$$\Phi(X, u) = H(X) - c(u) \to \max_{u \in U},$$

где $H(\cdot)$ – выигрыш, $c(\cdot)$ – затраты на управление.

3. Информационное взаимодействие (динамика)

Агенты из N образуют социальную сеть $G = (N, E)$.
Вектор начальных мнений x, конечных X.

$$i \xrightarrow{a_{ij}} j \xrightarrow{a_{jk}} k$$

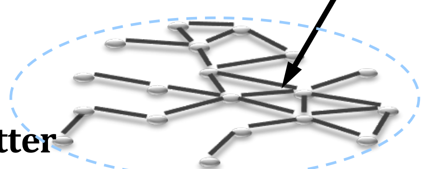
$a_{ij} \geq 0$ – степень доверия i-го агента j-му, k-й агент косвенно влияет на i-го.

$$x^{k+1} = A\,[x^k + B\,u^k].$$

Задача – найти результирующее влияние одних агентов на других; найти агентов, формирующих итоговое мнение в сети.

2. Структурный анализ

Facebook



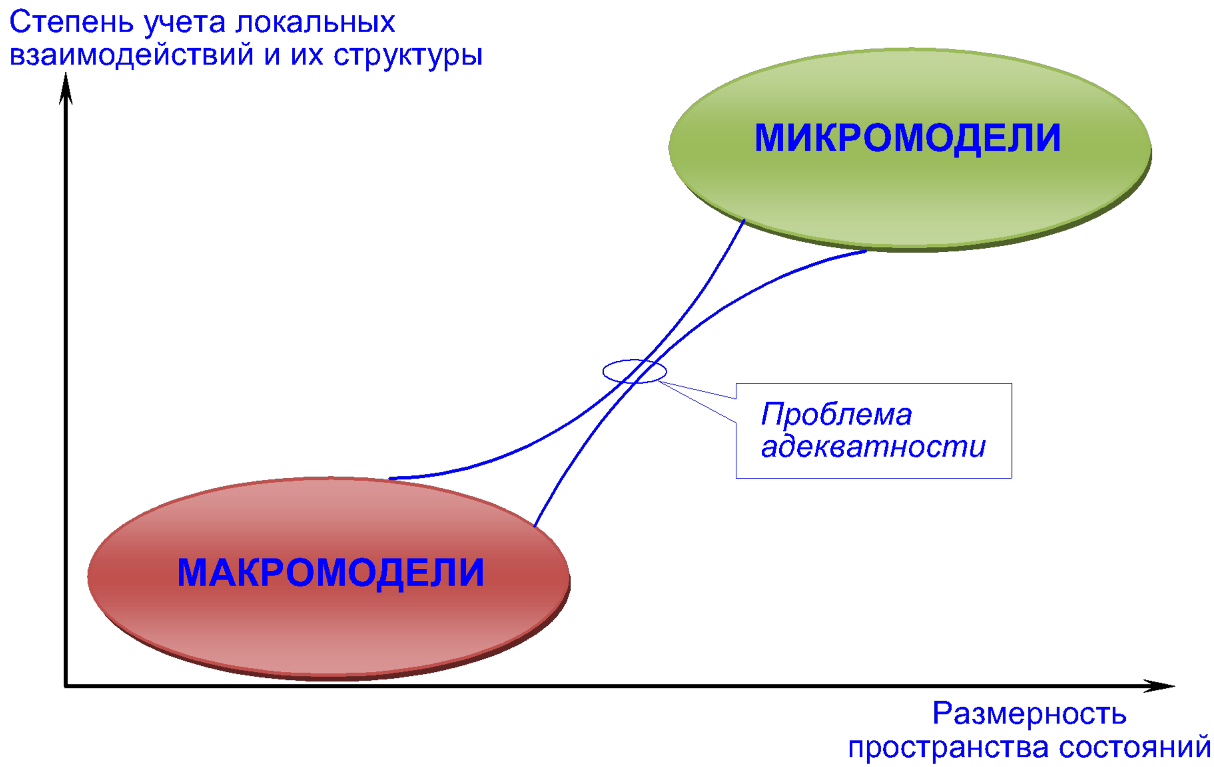
Twitter

Newsvine

Habr

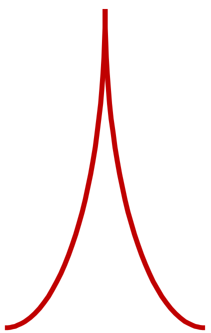
1. «Статистический» анализ

МИКРО- И МАКРОМОДЕЛИ



ВОЗМОЖНОСТЬ ОПЕРИРОВАНИЯ БОЛЬШИМИ ДАННЫМИ

5. Информационное противоборство
4. Информационное управление
3. Информационное взаимодействие
2. Структурный анализ
1. «Статистический» анализ



ПРИМЕРЫ МОДЕЛЕЙ АКТИВНЫХ СЕТЕВЫХ СТРУКТУР:

Модель ДеГроота: $x^k = A\,x^{k-1}$

Дискретная и непрерывная модели Грановеттера: $x^k = F(x^{k-1})$, $\dot{x} = F(x) - x$

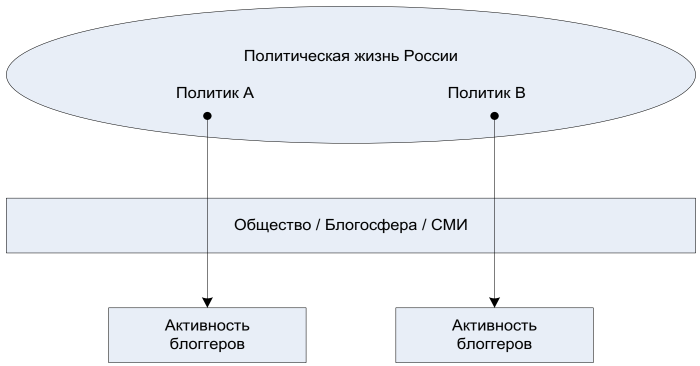
Вероятностная модель Грановеттера: $\frac{\partial}{\partial t} p(x,t) + \frac{\partial}{\partial x}\left(\left[F(x) - x\right] p(x,t)\right) = 0$

Общая модель возбуждения сети: $x^* = G(x^0)$

АНАЛИЗ И МОДЕЛИРОВАНИЕ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ В СОЦИАЛЬНЫХ МЕДИА

5. Информационное противоборство
4. Информационное управление
3. Информационное взаимодействие
2. Структурный анализ
1. «Статистический» анализ

«ПАССИВНАЯ СОЦИОЛОГИЯ»

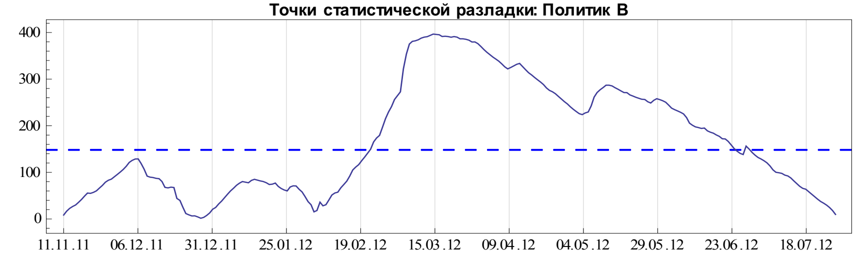


Суточное количество сообщений : Политики А и В

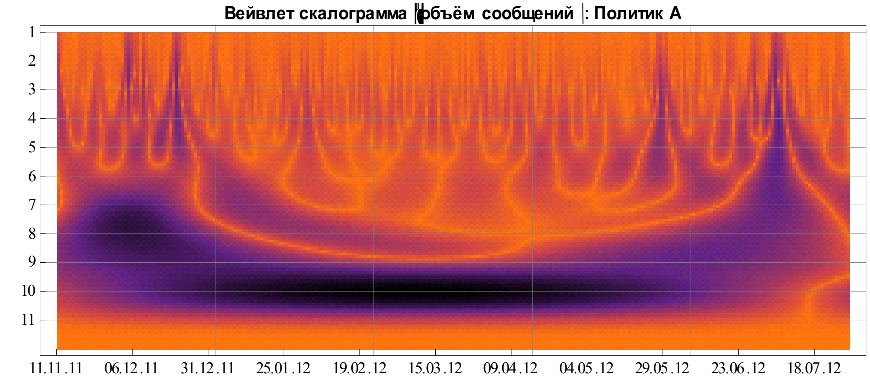
Автокорреляционная функция по сообщениям : Политики А и В

Парная корреляционная функция по сообщениям

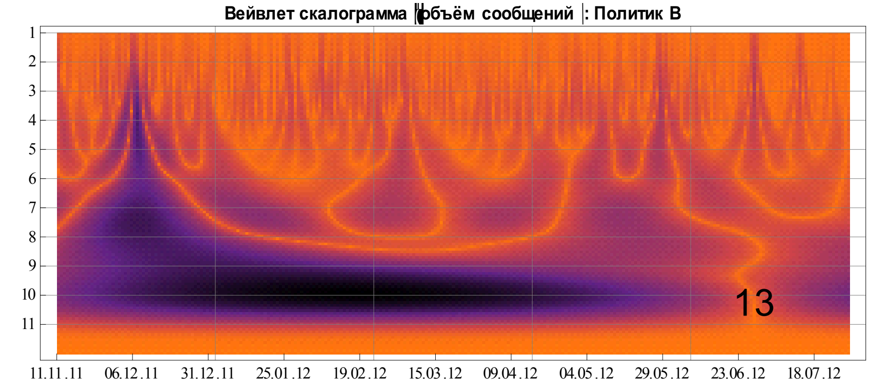
Точки статистической разладки: Политик В



Вейвлет скалограмма (объём сообщений) : Политик А



Вейвлет скалограмма (объём сообщений) : Политик В



СТРУКТУРА СЕТИ КОММЕНТИРОВАНИЯ FACEBOOK НА ИЮНЬ 2013 г.

Рассматриваются максимальные подсети из узлов с количеством комментаторов Н и выше.

Н. Число групп. Размер групп

38. 1 группа (547).
37. 1 группа (610).
36. 1 группа (660).
35. 1 группа (1 163)
…
31. 1 группа (3 013).

20. 1 группа (19 508)

10. 1 группа (114 229)

5. 2 группы (349К и 5)

3. 33 группы (694К, 21..)

1. 272К групп (1.7М., 26..)

«Остров» является «монолитным», на самой его вершине – **пользователи из Украины, оппозиционно настроенные к правительству РФ и к правительству Украины.**

ЗАДАЧА О КОНСЕНСУСЕ* В СОЦИАЛЬНОЙ СЕТИ

Информационное влияние

$$x^{k+1} = A\,[x^k + B\,u^k]$$

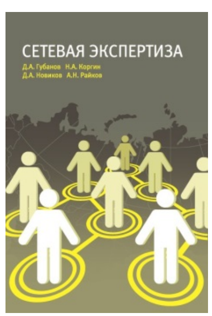
<u>Утверждение 1.</u> Пусть все элементы стохастической матрицы прямого влияния A строго положительны, а управления не ограничены. Тогда при наличии, как минимум, одного (произвольного) агента влияния может быть реализовано любое единогласное значение итоговых мнений членов социальной сети.

<u>Утверждение 2.</u> Пусть центр оказал воздействия $u^0, \ldots, u^l$, $l < +\infty$. Вектор итоговых (при $t = +\infty$) мнений агентов не изменится, если те же (по величине) воздействия были оказаны в любые другие конечные моменты времени.

<u>Утверждение 3.</u> Пусть управления не ограничены и

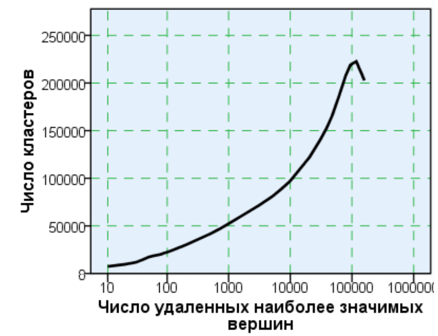
$$\mathrm{span}\,(\Phi) \subseteq \mathrm{span}\,(A^{l+1}\,B).$$

Тогда для любой конечной последовательности векторов управляющих воздействий $u^0, \ldots, u^l$, $l < +\infty$, и реализовавшегося в результате этих воздействий в момент времени l состояния x^{l+1} социальной сети, существует такой вектор управлений $\hat{v}$ в начальный (нулевой) момент времени, который приводит к тому же состоянию x^{l+1} социальной сети в момент времени $l + 1$.

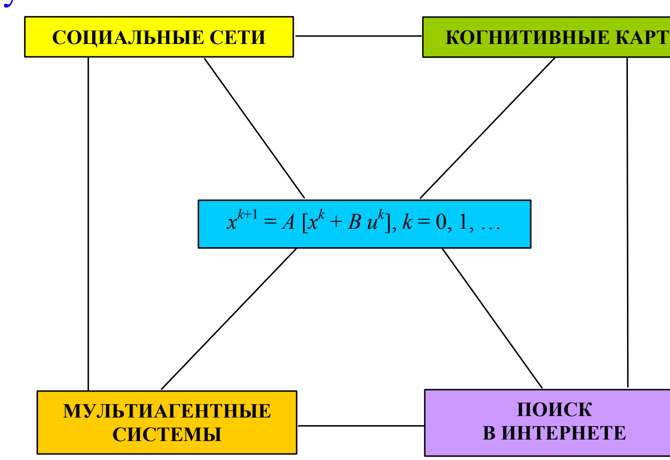


*French(1956)-DeGroot(1974)-Harary(1959)

Управление структурой



<u>Пример.</u> Удаление небольшого числа наиболее значимых узлов приводит к образованию большого числа несвязанных групп (А). Однако эти группы по своему размеру малы, наибольшая группа остается связанной (В)



ПОРОГОВОЕ ПОВЕДЕНИЕ: МАКРО- И МИКРО-МОДЕЛИ

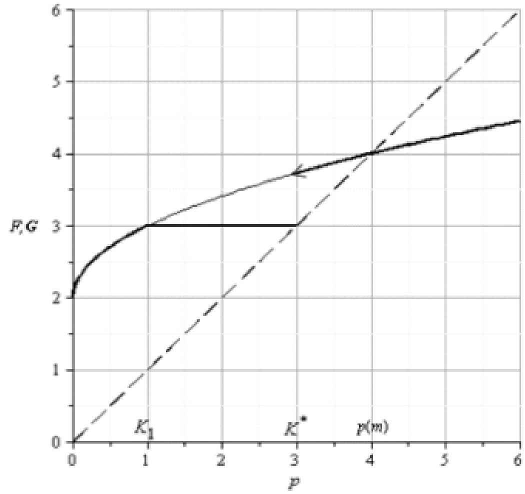
Пороговое поведение*:

$$x_i = BR_i(x_{-i}) = \begin{cases} 1, \text{ если } \sum_{j \neq i} t_{ij} x_j \geq \theta_i, \\ 0, \text{ если } \sum_{j \neq i} t_{ij} x_j < \theta_i. \end{cases} \quad (1)$$

Модель динамики коллективного поведения: в начальный момент времени все агенты бездействуют, далее в каждый из последующих моментов времени агенты одновременно и независимо действуют в соответствии с пороговой процедурой. Обозначим $Q_0 = \{i \in N \mid \theta_i = 0\}$, $Q_k = Q_{k-1} \cup \{i \in N \mid \sum_{j \in Q_{k-1}, j \neq i} t_{ij} \geq \theta_i\}$, $k = 1, 2, \ldots, n-1$,

$q(T, \theta) = \min \{k = \overline{0, n-1} \mid Q_{k+1} = Q_k\}$. **Равновесие коллективного поведения** (РКП):

$$x_i^*(T, \theta) = \begin{cases} 1, \text{ если } i \in Q_{q(T,\theta)} \\ 0, \text{ если } i \in N \setminus Q_{q(T,\theta)} \end{cases}, \ i \in N.$$



Агрегированным показателем состояния толпы будем считать число действующих агентов:

$$K(T, \theta) = |Q_{q(T, \theta)}|,$$

Утверждение.** Для любых матриц влияния T и порогов агентов θ РКП существует, единственно и является одним из равновесий Нэша для игры с наилучшим ответом (1).

Задача управления: $H(K(T, \theta)) - C(T, \theta, T^0, \theta^0) \to \max_{T \in \mathrm{T}, \theta \in \Theta}$.

Утверждение.** Затраты

$$c(K^*) = g(K^*)\left(F(0+) - K^*\right)^+ + \int_{F^{-1}(K^*)}^{K^*} g(K^* - t)\, dF(t)$$

минимальны для реализации центром РКП K^*.

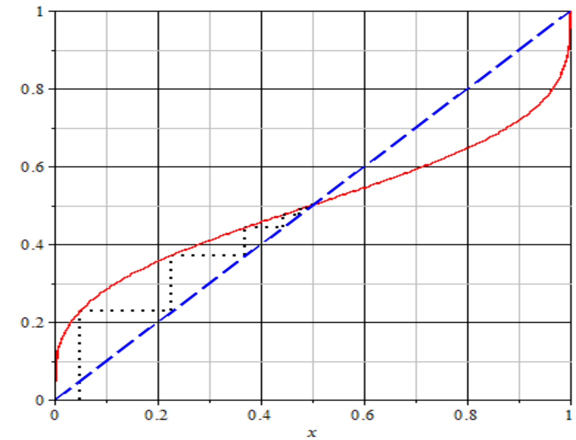
Утверждение.** Не существует стабильного информационного равновесия, при котором действует строго меньшее число агентов, чем в РКП.

*Granovetter M. // The American Journal of Sociology. 1978.
**Бреер В.В., Новиков Д.А. // Проблемы управления. 2012.

МАКРО- И МИКРО-МОДЕЛИ: ТЕОРИЯ*

В *микромодели* поведение агентов (здесь $x_i^{(k)} \in \{0,1\}$ - действует или бездействует агент i в периоде времени k) описывается с помощью его наилучшего ответа (*BR-Best Response*):

$$(1)\quad x_i^{(k)} = BR_i\left(x_{-i}^{(k-1)}\right) = \begin{cases} 1, & \sum\limits_{j \in D_i} x_j^{(k-1)} > \theta d_i, \\ 0, & \sum\limits_{j \in D_i} x_j^{(k-1)} \le \theta d_i. \end{cases}, \quad i \in N = \{1, 2, \dots, n\},$$



где D_i - множество *соседей*, $d_i = |D_i|$ - количество соседей,

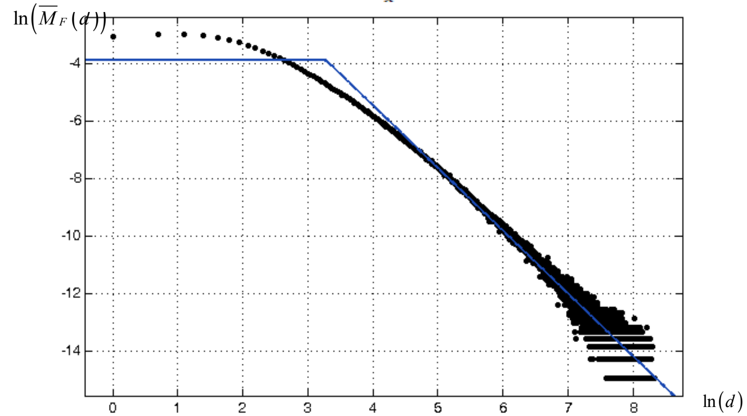
$x_{-i}^{(k-1)} = \left\{x_1^{(k-1)}, x_2^{(k-1)}, \dots, x_{i-1}^{(k-1)}, x_{i+1}^{(k-1)}, \dots, x_n^{(k-1)}\right\}$ - обстановка для агента i.

Макро-модель с единым относительным порогом описывает динамику доли действующих агентов $p_k = \frac{1}{n}\sum\limits_i x_i^{(k)} \in [0,1]$:

$$(2)\quad p_{k+1} = F_n(p_k, \theta),$$

$$(3)\quad F_n(p, \theta) = \sum_{d=1}^{n-1} B(p, d, \theta) M(d),$$

$B(p,d,\theta) = 1 - \sum\limits_{k=0}^{[\theta d]} C_d^k p^k (1-p)^{d-k}$ - функция от биномиального распределения, $M(d)$ - плотность распределения числа соседей d в графе социальной сети.



Макро-показатели СС

СС	MaxFrnds	Users	Nonzero users	Links	AvgFriends
Facebook	4 199	3 250 580	3 084 017	77 639 757	50.35
Live Journal	2 499	5 758 706	3 586 959	124 729 288	34.77
Twitter	759 313	~41 700 000	35 427 738	1 418 363 662	40.00

*Бреер В.В., Новиков Д.А., Рогаткин А.Д. // Проблемы управления. 2014. № 5.

МАКРО- И МИКРО-МОДЕЛИ: ЗАДАЧИ*

Задача 1 заключается в идентификации функций распределения $M(\cdot)$ в указанных трех СС. В рамках этой задачи строятся эмпирические функции плотности распределения $M_F(\cdot)$, $M_L(\cdot)$ и $M_T(\cdot)$, и ищутся аппроксимирующие их в аналитическом виде функции $\overline{M}_F(\cdot)$, $\overline{M}_L(\cdot)$ и $\overline{M}_T(\cdot)$.

Задача 2 заключается в построении и исследовании имитационных моделей порогового поведения, которое задается наилучшим ответом (1). Так, действующими в начальный момент времени считались случайно выбранные агенты, и вычислялось, согласно выражению (1), число действующих агентов на следующем шаге. Затем результат усреднялся по случайным множествам первоначально выбранных агентов. В результате было построено семейство функций (зависящее от параметра θ), которые сравнивались с другими функциями распределения, полученными в результате решения других задач настоящего исследования.

Задача 3 состоит в аппроксимации сигмовидными функциями зависимостей, полученных при имитационном моделировании в рамках решения задачи 2.

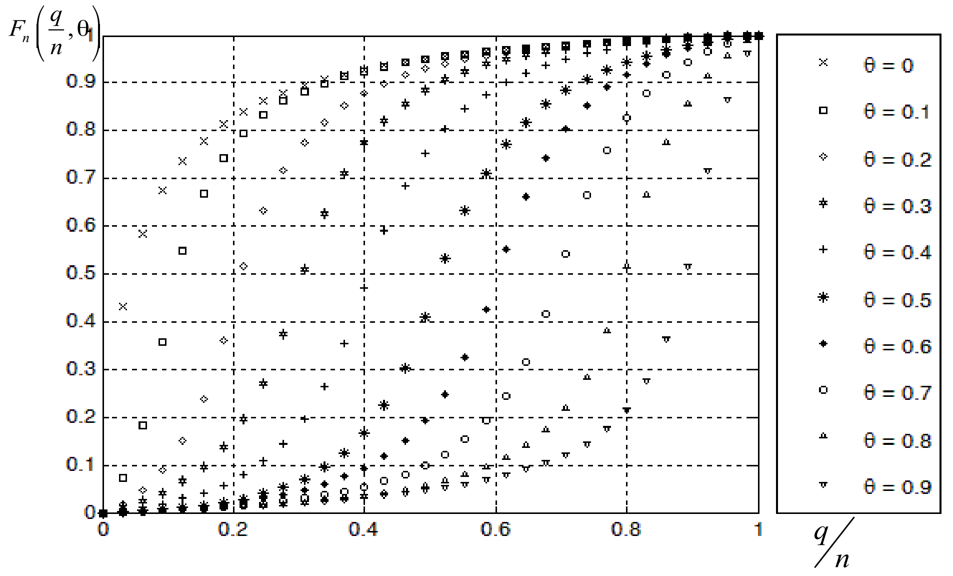
Задача 4 состоит в нахождении семейства функций распределения (3) (зависящего от параметра θ), в которые вместо $M(\cdot)$ подставляются эмпирические функции распределения степеней узлов графов связей для трех рассматриваемых реальных СС - $M_F(\cdot)$, $M_L(\cdot)$ и $M_T(\cdot)$.

Задача 5 аналогична четвертой, но вместо эмпирических функций распределения СС используются их аппроксимации $\overline{M}_F(\cdot)$, $\overline{M}_L(\cdot)$ и $\overline{M}_T(\cdot)$, найденные в результате решения задачи 1.

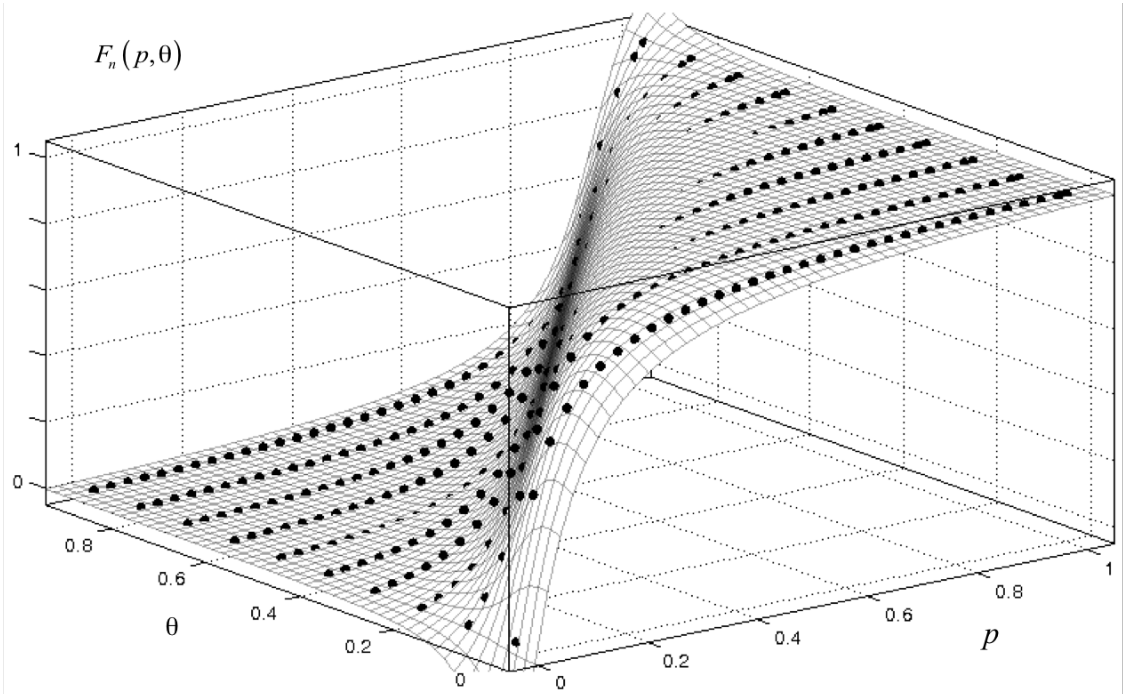
Для решения задач 2-5 использовались два метода: анализ эмпирических данных и их аналитическая аппроксимация.

*Батов А.В., Бреер В.В., Новиков Д.А., Рогаткин А.Д. // Проблемы управления. 2014. № 6.

МАКРО- И МИКРО-МОДЕЛИ: ИДЕНТИФИКАЦИЯ*



$$F(x,\theta,\gamma)=\frac{arctg\ \gamma(x-\theta)+arctg(\gamma\theta)}{arctg\ \gamma(1-\theta)+arctg(\gamma\theta)}$$



*Батов А.В., Бреер В.В., Новиков Д.А., Рогаткин А.Д. // Проблемы управления. 2014. № 6.

«ЭКВИВАЛЕНТНОСТЬ» МАКРО- И МИКРО-МОДЕЛЕЙ. ПОЛОЖЕНИЯ РАВНОВЕСИЯ*

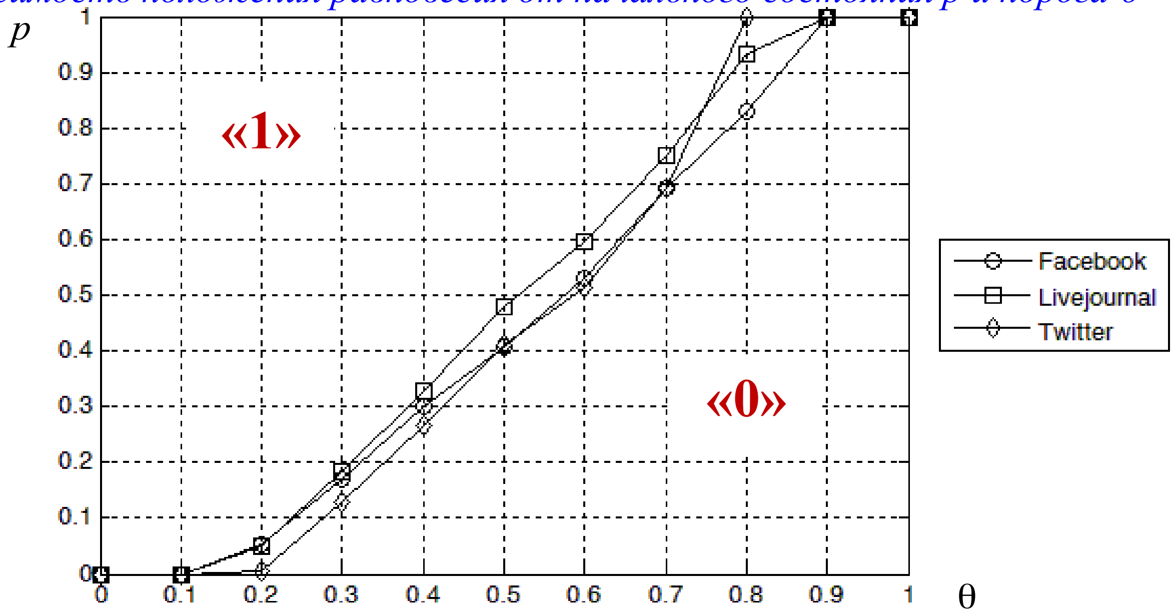
Используемые модели и методы

Метод / *Модель*	**Эмпирические данные**	**Аппроксимация**
Микро-модель СС	Задача 2	Задача 3
Макро-модель СС	Задача 4	Задача 5

Сравнение результатов (значение R^2)

СС	Задачи 2 и 4	Задачи 2 и 5	Задачи 4 и 5	Задачи 2 и 3	Задачи 3 и 4	Задачи 3 и 5
Facebook	0,9976	0,9932	0,9911	0,9973	0,9973	0,9907
LiveJournal	0,9999	0,9872	0,9872	0,9960	0,9960	0,9855
Twitter	0,9998	0,9631	0,9642	0,9949	0,9950	0,9599

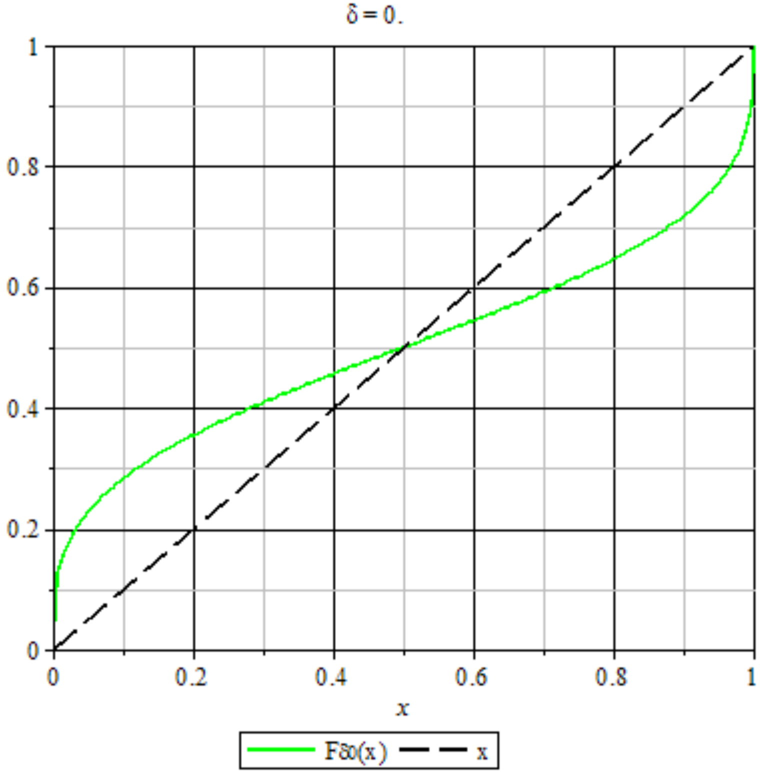
Зависимость положения равновесия от начального состояния p и порога θ

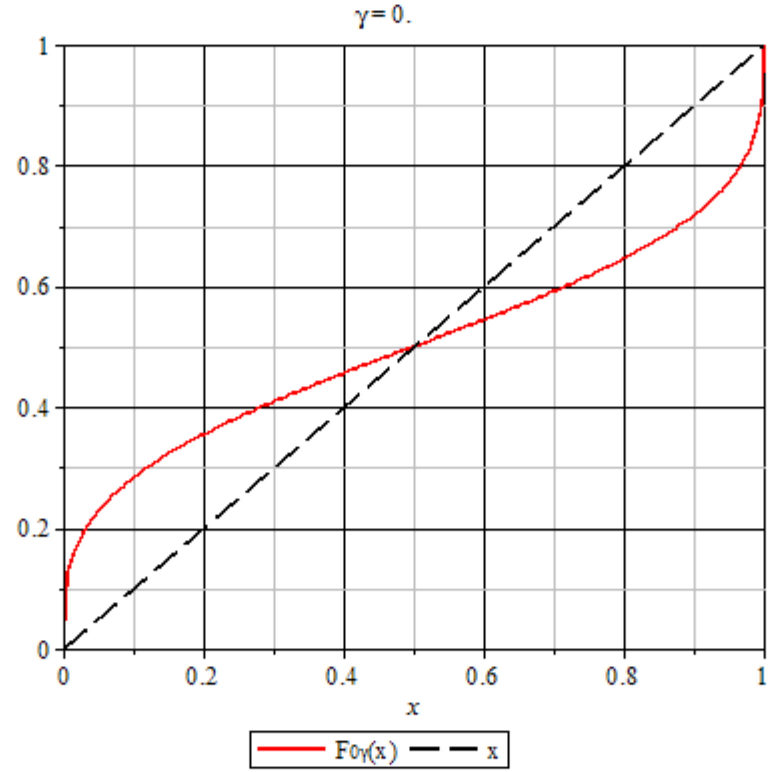


*Батов А.В., Бреер В.В., Новиков Д.А., Рогаткин А.Д. // Проблемы управления. 2014. № 6.

СТОХАСТИЧЕСКИЕ МОДЕЛИ УПРАВЛЕНИЯ ТОЛПОЙ*

$$F_{\delta\gamma}(x) = \begin{cases} \dfrac{\delta + F(x)}{1+\delta+\gamma}, x \in [0,1) \\ 1, x = 1 \end{cases}$$





*Бреер В.В., Новиков Д.А., Рогаткин А.Д. // УБС. 2014. № 52.

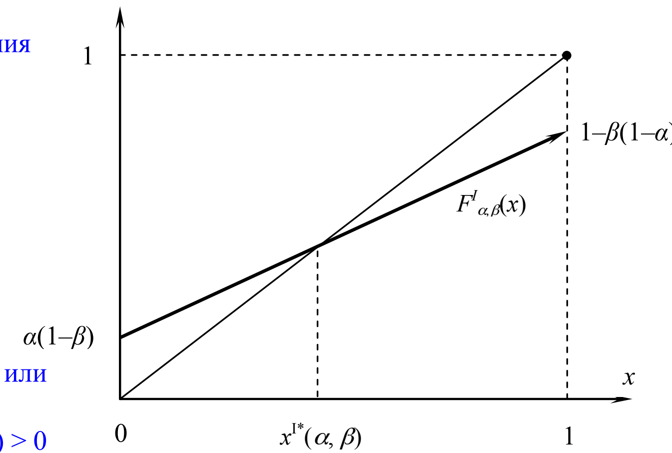
СТОХАСТИЧЕСКИЕ МОДЕЛИ УПРАВЛЕНИЯ ТОЛПОЙ*

<u>Утверждение.</u> В модели I для любого $\beta \in [0, 1]$ $x^*(\alpha, \beta)$ монотонно неубывает по α. При этом, если $F(0) > 0$ и $F(1-0) < 1$, то для любого $\beta \in [0, 1]$ $x^*(\alpha, \beta)$ монотонно возрастает по α.

<u>Утверждение.</u> В модели I для любого $\alpha \in [0,1]$ $x^*(\alpha, \beta)$ монотонно невозрастает по β. При этом, если $F(0) > 0$ и $F(1-0) < 1$, то для любого $\alpha \in [0, 1]$ $x^*(\alpha, \beta)$ монотонно убывает по β.

Обозначим через $W_{\alpha,\beta} = \bigcup\limits_{(\alpha,\beta)\in[0;1]^2} x^*(\alpha,\beta)$ *множество достижимости.*

Пример. $F(x)=x$



<u>Утверждение.</u> Точка $x \in [0, 1]$ принадлежит множеству достижимого равновесия функции распределения $F(\cdot) \in C[0, 1]$ тогда и только тогда, когда выполнено либо $F(x) = 0$, либо

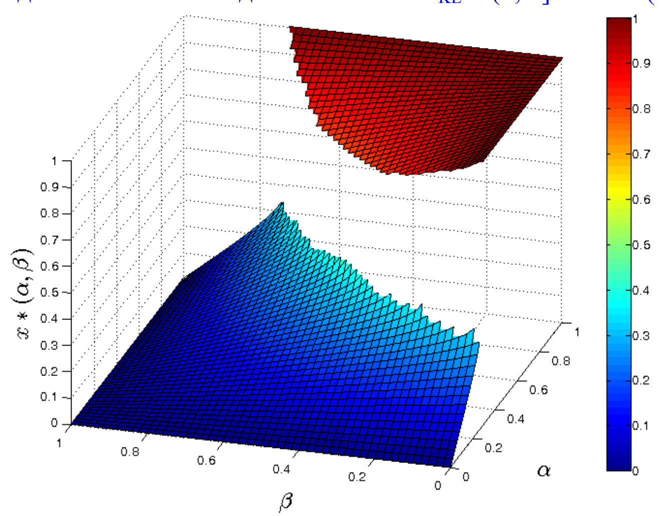
$$\frac{F'(x)}{F(x)}\cdot\left(x-\frac{\left(\sqrt{1+\frac{x}{F(x)}\left(\frac{1}{F(x)}-2\right)}-1\right)^2}{\left(\frac{1}{F(x)}-2\right)^2}\right)<1.$$

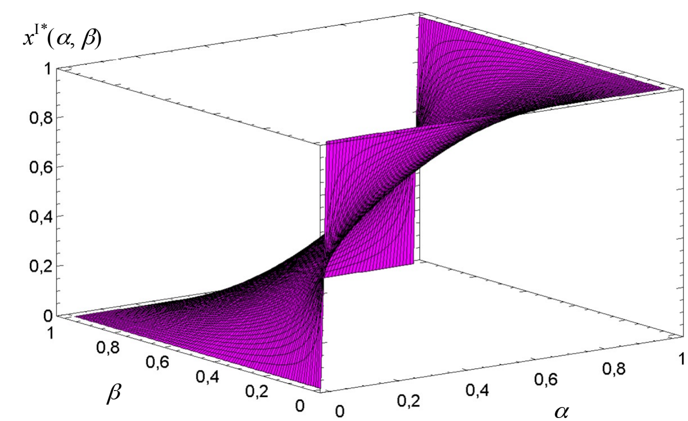
<u>Утверждение.</u> В модели II $x^*(\delta, \gamma)$:

1) монотонно неубывает по δ, а для строгой монотонности достаточно выполнения условия: $F(1\text{-}0) < 1$ или $\gamma > 0$;

2) монотонно невозрастает по γ, а для строгой монотонности достаточно выполнения условия: $F(0) > 0$ или $\delta > 0$.

<u>Утверждение</u>. В модели II множество достижимости $W_{KL} = (0, 1]$. Если $F(0) = 0$, то $W_{KL} = [0, 1]$.

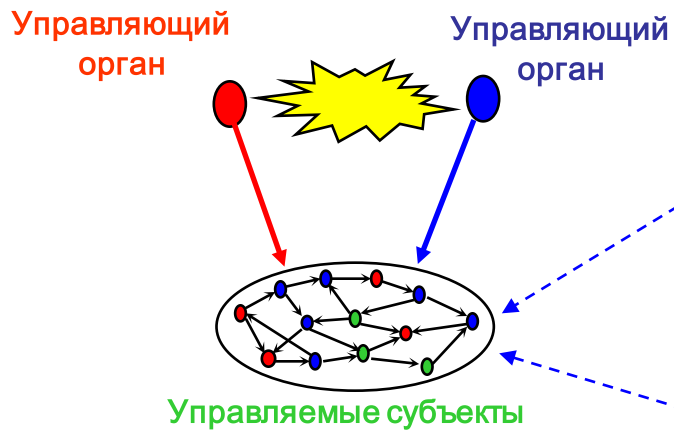




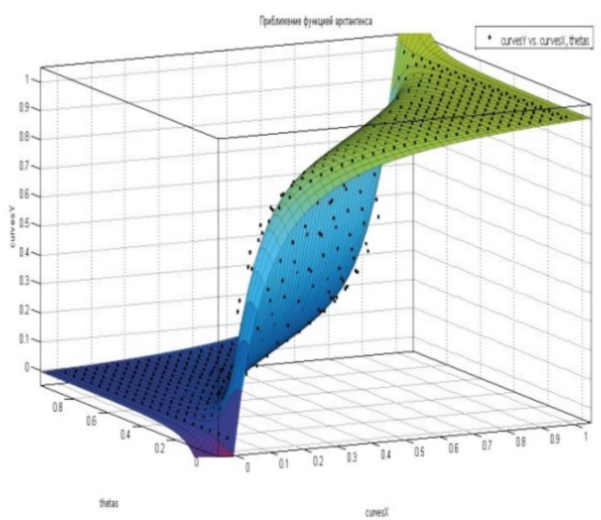
Результат численного вычисления $x^*(\alpha, \beta)$ в модели I для СС Facebook ($\theta = 0.5$, $\lambda = 13$)

*Бреер В.В., Новиков Д.А., Рогаткин А.Д. // УБС. 2014. № 52.

ИНФОРМАЦИОННОЕ ПРОТИВОБОРСТВО



«МАРКОВСКАЯ» МОДЕЛЬ
Губанов Д.А., Калашников А.О., Новиков Д.А. Теоретико-игровые модели информационного противоборства в социальных сетях // Управление большими системами. 2010. № 31. С. 192 - 204.



Идентификация порогового поведения (СС Facebook)

$$F(x,\theta,\gamma)=\frac{arctg\,\gamma(x-\theta)+arctg(\gamma\theta)}{arctg\,\gamma(1-\theta)+arctg(\gamma\theta)}$$

МОДЕЛЬ I*

«Управляемая» функция распределения порогов:

$$F_{\alpha,\beta}(x)=\begin{cases}\alpha(1-\beta)+(1-\alpha-\beta+2\alpha\beta)\,F(x), & x\in[0;1),\\ 1, & x=1,\end{cases}$$

$\alpha,\beta\in[0;\,1]$

МОДЕЛЬ II*

«Управляемая» функция распределения порогов:

$$F_{\delta,\gamma}(x)=\begin{cases}\dfrac{\delta+F(x)}{1+\delta+\gamma}, & x\in[0;1),\quad \delta,\gamma\geq 0\\ 1, & x=1,\end{cases}$$

«Равновесие»: $x^*=\begin{cases}y, \text{ если } \forall z\in[0,y]\ F_{\alpha,\beta}(z)\geq z,\\ 0, \text{ иначе}.\end{cases}$

*Бреер В.В., Новиков Д.А., Рогаткин А.Д. // УБС. 2014. № 52.

РДС И РАВНОВЕСИЕ НЭША*

Пусть целевые функции первого и второго центров имеют соответственно вид:

$$f_\alpha(\alpha, \beta) = H_\alpha(x^*(\alpha, \beta)) - c_\alpha(\alpha),\ f_\beta(\alpha, \beta) = H_\beta(x^*(\alpha, \beta)) - c_\beta(\beta),$$

причем выигрыш первого центра $H_\alpha(\cdot)$ – возрастающая функция (он заинтересован в максимизации числа возбужденных агентов), а выигрыш второго центра $H_\beta(\cdot)$ – убывающая функция (он заинтересован в минимизации числа возбужденных агентов), а обе функции затрат $c_\alpha(\cdot)$ и $c_\beta(\cdot)$ – неубывающие и $c_\alpha(0) = c_\beta(0) = 0$.

Утверждение. В модели I в игре с нулевой суммой без учета затрат центров на управление существует РДС их игры: $\alpha^{РДС} = 1$, $\beta^{РДС} = 1$.

Утверждение. В модели II в игре с нулевой суммой без учета затрат центров на управление не существует конечного РДС или РН их игры. Если в модели II множества допустимых стратегий центров ограничены: $\delta \leq \delta_{\max}$, $\gamma \leq \gamma_{\max}$, то в игре с нулевой суммой без учета затрат центров на управление существует РДС их игры: $\delta^{РДС} = \delta_{\max}$, $\gamma^{РДС} = \gamma_{\max}$.

Утверждение. Если в модели I $x^*(\alpha, \beta)$ - непрерывная монотонная функция, $W_{\alpha,\beta} = [0; 1]$, функции выигрыша центров - ограниченные и слабо вогнутые, а их функции затрат - выпуклые и асимптотически стремящиеся к бесконечности при стремлении аргумента к единице, то существует внутреннее равновесие Нэша игры центров.

Пример. Пусть $F(x) = x$, $H_\alpha(x) = x$, $H_\beta(x) = 1 - x$, $c_\alpha(\alpha) = -\ln(1 - \alpha)$, $c_\beta(\beta) = -\lambda \ln(1 - \beta)$. Из условий первого порядка получаем: $\beta = (1/\lambda)\,\alpha$. При $\lambda = 1$ находим: $\alpha^* = 1/4$, $\beta^* = 1/4$. При этом:

$$x^{I*}(\alpha^*, \beta^*) = 1/2,\ f_\alpha(\alpha^*, \beta^*) = f_\beta(\alpha^*, \beta^*) \approx -0{,}2.$$

Утверждение. Если в модели II $x^*(\delta, \gamma)$ - непрерывная монотонная функция, функции выигрыша центров ограниченные и слабо вогнутые, а их функции затрат - выпуклые и асимптотически стремящиеся к бесконечности при стремлении аргумента к бесконечности, то существует конечное равновесие Нэша игры центров.

Пример. Пусть $F(x) = x$, $H_\delta(x) = x$, $H_\gamma(x) = 1 - x$, $c_\delta(\delta) = \delta^2$, $c_\gamma(\gamma) = \lambda^2 \gamma^2$. РКП: $x^*(\delta, \gamma) = \delta / (\delta + \gamma)$. Получаем следующие выражения для целевых функций центров: $f_\delta(\delta, \gamma) = \delta / (\delta + \gamma) - \delta^2$, $f_\gamma(\delta, \gamma) = 1 - \delta / (\delta + \gamma) - \lambda^2 \gamma^2$.

Находим РН: $\delta^* = \sqrt{\dfrac{\lambda}{2}}\dfrac{1}{1+\lambda}$, $\gamma^* = \dfrac{1}{\sqrt{2\lambda}}\dfrac{1}{1+\lambda}$.

*Новиков Д.А. // ПУ. 2015. № 3.

ПОРОГОВЫЕ ФУНКЦИИ ВЫИГРЫША ЦЕНТРОВ: ОПИСАНИЕ МОДЕЛИ

$$H_{\alpha(\beta)}(x) = \begin{cases} H^+_{\alpha(\beta)}, \text{ если } x \geq (\leq)\ \theta_\alpha(\theta_\beta), \\ \qquad H^-_{\alpha(\beta)}, \text{ иначе,} \end{cases}$$

где $H^+_{\alpha(\beta)} > H^-_{\alpha(\beta)}$, то есть первый центр получает больший выигрыш тогда, когда доля действующих агентов не меньше порога $\theta_\alpha \in [0; 1]$, а второй центр – при условии, что доля действующих агентов не превышает порога $\theta_\beta \in [0; 1]$. Обозначим через x^0 РКП в отсутствии воздействий центров, т.е. $x^0 = x^*(0, 0)$. Введем следующие предположения.

А.1. Множество достижимости составляет весь единичный отрезок, $x^*(\alpha, \beta)$ – строго монотонная непрерывная функция своих переменных (соответствующие достаточные условия приведены в разделе 3 и/или могут быть проверены в каждом конкретном случае), а функции затрат центров строго монотонны.

А.2. Первый центр при нулевых действиях второго может реализовать самостоятельно любое РКП из $[x^0; 1]$; а второй центр при нулевых действиях первого может реализовать самостоятельно любое РКП из $[0; x^0]$

А.3. Первый центр при нулевых действиях второго может реализовать самостоятельно РКП θ_α; а второй центр при нулевых действиях первого может реализовать самостоятельно РКП θ_β.

Обозначим через $\alpha(\theta) = \min \{\alpha \in [0; 1] \mid x^*(\alpha, 0) = \theta\}$, $\beta(\theta) = \min \{\beta \in [0; 1] \mid x^*(0, \beta) = \theta\}$.

Определим множество

$$\Omega_{\alpha,\beta}(\theta) = \{(\alpha, \beta) \in [0; 1]^2 \mid x^*(\alpha, \beta) = \theta,\ c_\alpha(\alpha)\ ?\ H^+_\alpha - H^-_\alpha,\ c_\beta(\beta)\ ?\ H^+_\beta - H^-_\beta\},\ \ \theta_\beta < x^0 < \theta_\alpha.$$

Определим следующие функции (если множество, по которому вычисляется минимум, пусто, то будем считать, что значение функции равно $+\infty$):

$$C_\alpha(x, \beta) = \min_{\{\alpha \in [0;1] \mid x^*(\alpha,\beta)=x\}} c_\alpha(\alpha),\ C_\beta(x, \alpha) = \min_{\{\beta \in [0;1] \mid x^*(\alpha,\beta)=x\}} c_\beta(\beta).$$

*Новиков Д.А. // ПУ. 2015. № 3.

ПОРОГОВЫЕ ФУНКЦИИ ВЫИГРЫША ЦЕНТРОВ: РН и РБС

5. Информационное противоборство
4. Информационное управление
3. Информационное взаимодействие
2. Структурный анализ
1. «Статистический» анализ

Утверждение. Если выполнены предположения А.1 и А.3, то РН игры центров характеризуются следующим образом:

- (0; 0) является РН, если выполнено $\begin{cases} H_\alpha^+ - c_\alpha(\alpha(\theta_\alpha)) \le H_\alpha^-, \\ H_\beta^+ - c_\beta(\beta(\theta_\beta)) \le H_\beta^-; \end{cases}$

- $(\alpha(\theta_\alpha); 0)$ является РН, если выполнено $\begin{cases} H_\alpha^+ - c_\alpha(\alpha(\theta_\alpha)) \ge H_\alpha^-, \\ H_\beta^- \ge H_\beta^+ - C_\beta(\theta_\beta, \alpha(\theta_\alpha)); \end{cases}$

- $(0; \beta(\theta_\beta))$ является РН, если выполнено $\begin{cases} H_\beta^+ - c_\beta(\beta(\theta_\beta)) \ge H_\beta^-, \\ H_\alpha^- \ge H_\alpha^+ - C_\alpha(\theta_\alpha, \beta(\theta_\beta)). \end{cases}$

Определим следующие функции: $C_\delta(x, \gamma) = \min\limits_{\{\delta \ge 0 | x^*(\delta, \gamma) = x\}} c_\delta(\delta)$, $C_\gamma(x, \delta) = \min\limits_{\{\gamma \ge 0 | x^*(\delta, \gamma) = x\}} c_\gamma(\gamma)$.

Утверждение. Пусть выполнены предположения А.1 и А.2. Тогда:

1) $(\delta_{\text{РБС}} + \varepsilon; 0)$ является РБС, если существует минимальное неотрицательное значение $\delta_{\text{РБС}}$, для которого выполнено:

$$\begin{cases} x^*(\delta_{\text{РБС}}; 0) \ge \theta_\delta; \\ H_\delta^+ - c_\delta(\delta_{\text{РБС}}) \ge H_\delta^-; \\ H_\gamma^+ - C_\gamma(\theta_\gamma, \delta_{\text{РБС}}) \le H_\gamma^-; \end{cases}$$

2) $(0; \gamma_{\text{РБС}} + \varepsilon)$ является РБС, если существует минимальное неотрицательное значение $\gamma_{\text{РБС}}$, для которого выполнено:

$$\begin{cases} x^*(0; \gamma_{\text{РБС}}) \le \theta_\gamma; \\ H_\gamma^+ - c_\gamma(\gamma_{\text{РБС}}) \ge H_\gamma^-; \\ H_\delta^+ - C_\delta(\theta_\delta, \gamma_{\text{РБС}}) \le H_\delta^-, \end{cases}$$

где ε - произвольно малая строго положительная константа.

*Новиков Д.А. // ПУ. 2015. № 3.

ТЕОРЕТИКО-ИГРОВАЯ МОДЕЛЬ ВОЗБУЖДЕНИЯ СЕТИ*

Сеть. $N = \{1, 2, ..., n\}$ – конечное множество *агентов*, входящих в *социальную сеть*, описываемую ориентированным графом $G = (N, E)$, где $E \subseteq N \times N$ – множество дуг. Пусть каждый агент может находиться в одном из двух состояний – «0» или «1» (например, бездействовать или действовать, не быть или быть в возбужденном состоянии). Обозначим через $y_i \in \{0; 1\}$ состояние i-го агента, $i \in N$, через $y = (y_1, y_2, \ldots, y_n)$ – вектор их состояний. Условно переход от бездействия к действию будем называть «возбуждением» соответствующего агента.

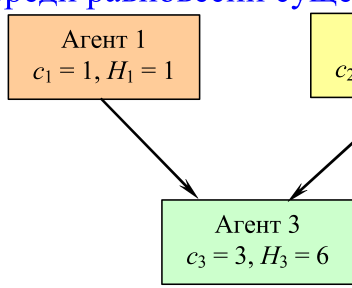
Поведение агентов. Предположим, что первоначально все агенты бездействуют, а «управляемая динамика» сети описывается отображением $\Phi: 2^N \to 2^N$, где $\Phi(S) \subseteq N$ – множество агентов, находящихся в состоянии «1» после окончания переходного процесса, вызванного изменением в начальный момент времени состояния (с бездействия на действие) агентов из множества (*коалиции*) $S \subseteq N$.

Задача централизованного управления (выбор центром такого множества первоначально возбуждаемых агентов, которое максимизировало бы его целевую функцию)**: $v(S) = H(\Phi(S)) - C(S) \to \max\limits_{S \subseteq N}$,

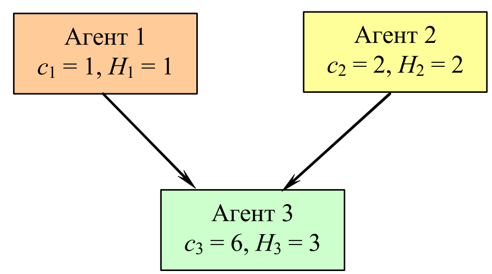
Децентрализованное управление: игра в нормальной форме. Пусть задан *механизм* $\sigma = \{\sigma_i(G(y)) \geq 0\}_{i \in N}$ распределения выигрышей между агентами, причем дополнительные (в соответствии с механизмом σ) выигрыши могут получать только те агенты, которые в начальный момент времени действовали (распределению подлежит «выигрыш» от опосредованного возбуждения других агентов). Целевую функцию i-го агента $f_i(y)$ будем считать равной разности между его выигрышем и затратами: $f_i(y) = \sigma_i(G(y)) + (H_i - c_i)\, y_i$, $i \in N$.

Гипотеза 1: одно из эффективных состояний является равновесным.

Гипотеза 2: среди равновесий существует хотя бы одно эффективное.



Пример 1. Эффективными являются три вектора состояний агентов: (0; 1; 0), (1; 0; 0) и (1; 1; 0). Равновесиями Нэша, независимо от механизма распределения выигрышей, являются четыре остальных вектора состояний, отличных от (0; 0; 0). Таким образом, ни одно из четырех равновесий Нэша не эффективно, и ни одно из эффективных состояний не является равновесным.



Пример 2. Эффективными являются те же три вектора состояний агентов: (0; 1; 0), (1; 0; 0) и (1; 1; 0). Равновесие Нэша единственно – вектор (1; 1; 0), то есть равновесие Нэша эффективно.

*Новиков Д.А. / Procedia Computer Science. 2014.

**Частный случай - Kempe D., Kleinberg J., Tardos E. / KDD. 2003.

ЦЕНТРАЛИЗОВАННОЕ УПРАВЛЕНИЕ «ВОЗБУЖДЕНИЕМ» СЕТИ*

Целевые функции агентов:

(1) $f_i(y) = s_i(y) + (H_i - c_i)\, y_i$,

где $s_i(y) \geq 0$ – выплачиваемое центром i-му агенту вознаграждение, зависящее в общем случае от вектора состояний (действий) всех агентов, $i \in N$.

Фиксируем некоторое множество агентов V. Рассмотрим следующую **задачу «институционального управления»** – найти минимальную (в смысле суммарных затрат центра на стимулирование) вектор-функцию стимулирования, которая реализует «возбуждение» в точности заданного множества V агентов как **равновесие Нэша** $y^*(s(\cdot))$ их игры:

(2) $$\begin{cases} \sum\limits_{i \in N} s_i(y^*) \to \min\limits_{s(\cdot)} \\ y_i^*(s(\cdot)) = 1,\ i \in V, \\ y_j^*(s(\cdot)) = 0,\ j \notin V. \end{cases}$$

Утверждение. При стратегии центра

(3) $$s_i^*(y) = \begin{cases} c_i - H_i + \varepsilon_i, \text{ если } i \in V \text{ и } y_i = 1 \\ 0, \text{ иначе} \end{cases},$$

где ε_i – сколь угодно малая строго положительная константа, выбор единичных действий агентами из множества V (и только ими!) является единственным *равновесием в доминантных стратегиях* игры агентов с целевыми функциями (1). Более того, стратегия (3) является ε_V-оптимальным решением задачи (2), где $\varepsilon_V = \sum\limits_{i \in V} \varepsilon_i$.

*Новиков Д.А. / Procedia Computer Science. 2014.

НЕКОТОРЫЕ ВЫВОДЫ

Модель ДеГроота ($x^k = A\, x^{k-1}$)

ВРЕМЯ / УПРАВЛЕНИЕ	Статика	Дискретное	Непрерывное
Постоянное			
Программное			
Позиционное			
Противоборство			

АиТ. 2010. № 11; УБС. 2010. № 31.

Простая (линейная) модель большой размерности

Дискретная и непрерывная модели Грановеттера ($x^k = F(x^{k-1})$, $\dot{x} = F(x) - x$)

ВРЕМЯ / УПРАВЛЕНИЕ	Статика	Дискретное	Непрерывное
Постоянное			
Программное			
Позиционное			
Противоборство			

УБС. 2014. № 52; ПУ. 2015. № 3; АиТ. 2016 (в печати).

Сложная (нелинейная) скалярная модель

Вероятностная модель Грановеттера ($\frac{\partial}{\partial t} p(x,t) + \frac{\partial}{\partial x}\big(\left[F(x) - x\right] p(x,t)\big) = 0$)

ВРЕМЯ / УПРАВЛЕНИЕ	Статика	Дискретное	Непрерывное
Постояное			
Программное			
Позиционное			
Противоборство			

УБС. 2015 (в печати)

Сложная (нелинейная) скалярная «функцио-нальная» модель

Общая модель возбуждения сети ($x^* = G(x^0)$)

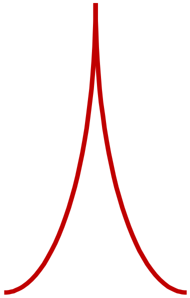
ВРЕМЯ / УПРАВЛЕНИЕ	Статика	Дискретное	Непрерывное
Постоянное			
Программное			
Позиционное			
Противоборство			

Procedia Computer Science. 2014. Vol. 31.

Сложная (нелинейная) «функциональная» модель большой размерности

ВОЗМОЖНОСТЬ ОПЕРИРОВАНИЯ БОЛЬШИМИ ДАННЫМИ

5. Информационное противоборство
4. Информационное управление
3. Информационное взаимодействие
2. Структурный анализ
1. «Статистический» анализ



Проблемы идентификации: обоснование перехода от микро- к макро-описанию

БОЛЬШИЕ ДАННЫЕ
И
БОЛЬШОЕ УПРАВЛЕНИЕ

Д.А. Новиков

(Институт проблем управления РАН)

dan@ipu.ru