

УДК 004.9+004.21

## СУПЕРКОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ КРИТИЧЕСКИХ ЯВЛЕНИЙ В СЛОЖНЫХ СОЦИАЛЬНЫХ СИСТЕМАХ

П.М.А. Слоот<sup>a</sup>, Я. Холист<sup>b</sup>, Дж. Кампис<sup>c</sup>, М.Х. Лис<sup>a</sup>, С.А. Митягин<sup>d</sup>, С.В. Иванов<sup>d</sup>,  
К.О. Боченина<sup>d</sup>, В.Ю. Гулева<sup>d</sup>, К.Д. Мухина<sup>d</sup>, Д.А. Насонов<sup>d</sup>, Н.А. Бутаков<sup>d</sup>, В.А. Леоненко<sup>d</sup>,  
А.А. Ланцева<sup>d</sup>, А.В. Бухановский<sup>d</sup>

<sup>a</sup> Университет г. Амстердам, Амстердам, 1012WX, Нидерланды

<sup>b</sup> Варшавский политехнический университет, Варшава, 00-661, Польша

<sup>c</sup> Будапештский университет имени Лоранда Этвёша, Будапешт, 1053, Венгрия

<sup>d</sup> Университет ИТМО, г. Санкт-Петербург, 197101, Российская Федерация

Автор для переписки: k.bochenina@gmail.com

### Информация о статье

Поступила в редакцию 20.09.16, принята к печати 28.10.16

doi: 10.17586/2226-1494-2016-16-6-967-995

Язык статьи – русский

**Ссылка для цитирования:** Слоот П.М.А., Холист Я., Кампис Дж., Лис М.Х., Митягин С.А., Иванов С.В., Боченина К.О., Гулева В.Ю., Мухина К.Д., Насонов Д.А., Бутаков Н.А., Леоненко В.А., Ланцева А.А., Бухановский А.В. Суперкомпьютерное моделирование критических явлений в сложных социальных системах // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 6. С. 967–995. doi: 10.17586/2226-1494-2016-16-6-967-995

### Аннотация

В статье рассмотрена проблема компьютерного моделирования критических явлений в сложных социальных системах на основе формализма комплексных сетей с использованием вычислительных систем петафлопсного диапазона производительности. Предложена система вложенных моделей комплексной сети на трех уровнях, включающая агрегированную аналитическую модель для идентификации критических состояний, детализированную модель динамики сети на индивидуальном уровне, а также модель, уточняющую особенности топологической структуры комплексной сети. Разработан масштабируемый параллельный алгоритм, в совокупности покрывающий все уровни моделирования комплексных сетей, и исследована его производительность на различных суперкомпьютерных системах. Обсуждены вопросы программно-информационной инфраструктуры моделирования комплексных сетей, включая организацию распределенных вычислений, сбор данных в социальных сетях, а также визуализацию результатов. Рассмотрены приложения разработанных методов и технологий в части моделирования распада криминальных сетей, быстрого распространения слухов в социальных сетях Интернет, эволюции финансовых сетей и распространения эпидемий.

### Ключевые слова

критические явления, комплексные сети, суперкомпьютерное моделирование, динамические процессы, параллельный алгоритм, распространение эпидемий, криминальные сети, финансовые сети

### Благодарности

Работа выполнена при поддержке проекта Российского научного фонда, соглашение № 14-21-00137.

## SUPERCOMPUTER SIMULATION OF CRITICAL PHENOMENA IN COMPLEX SOCIAL SYSTEMS

P.M.A. Sloot<sup>a</sup>, J. Holyst<sup>b</sup>, G. Kampis<sup>c</sup>, M. H. Lees<sup>a</sup>, S.A. Mityagin<sup>d</sup>, S.V. Ivanov<sup>d</sup>,  
K.O. Bochenina<sup>d</sup>, V.Yu. Guleva<sup>d</sup>, K.D. Mukhina<sup>d</sup>, D.A. Nasonov<sup>d</sup>, N.A. Butakov<sup>d</sup>, V.A. Leonenko<sup>d</sup>,  
A.A. Lantseva<sup>d</sup>, A.V. Boukhanovsky<sup>d</sup>

<sup>a</sup> University of Amsterdam, Amsterdam, 1012WX, the Netherlands

<sup>b</sup> University of Warsaw, Warsaw, 00-661, Poland

<sup>c</sup> Eötvös University, Budapest, 1053, Hungary

<sup>d</sup> ITMO University, Saint Petersburg, 197101, Russian Federation

Corresponding author: k.bochenina@gmail.com

### Article info

Received 20.09.16, accepted 28.10.16

doi: 10.17586/2226-1494-2016-16-6-967-995

Article in Russian

**For citation:** Sloot P.M.A., Holyst J., Kamps G., Lees M.H., Mityagin S.A., Ivanov S.V., Bochenina K.O., Guleva V.Yu., Mukhina K.D., Nasonov D.A., Butakov N.A., Leonenko V.A., Lantseva A.A., Boukhanovsky A.V. Supercomputer simulation of critical phenomena in complex social systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 6, pp. 967–995. doi: 10.17586/2226-1494-2016-16-6-967-995

#### **Abstract**

The paper describes a problem of computer simulation of critical phenomena in complex social systems on a petascale computing systems in frames of complex networks approach. The three-layer system of nested models of complex networks is proposed including aggregated analytical model to identify critical phenomena, detailed model of individualized network dynamics and model to adjust a topological structure of a complex network. The scalable parallel algorithm covering all layers of complex networks simulation is proposed. Performance of the algorithm is studied on different supercomputing systems. The issues of software and information infrastructure of complex networks simulation are discussed including organization of distributed calculations, crawling the data in social networks and results visualization. The applications of developed methods and technologies are considered including simulation of criminal networks disruption, fast rumors spreading in social networks, evolution of financial networks and epidemics spreading.

#### **Keywords**

critical phenomena, complex networks, supercomputer simulation, dynamical processes, parallel algorithm, epidemics spreading, criminal networks, financial networks

#### **Acknowledgements**

This paper is financially supported by the Russian Scientific Foundation, Agreement No.14–21–00137 (15.08.2014).

### **Введение**

Современный уровень развития методов и технологий сбора, анализа и интерпретации больших объемов данных, а также компьютерного моделирования на их основе делает возможным изучение глобальной изменчивости географически распределенных систем реального мира (например, энергетической, транспортной, финансовой) с учетом их социальной и социотехнической составляющих. На практике это может являться основой для выработки стратегий устойчивого развития таких систем и поддержки принятия решений в критических ситуациях с целью устранить или хотя бы минимизировать ожидаемый ущерб. Однако с вычислительной точки зрения прогресс в данной области ограничен свойством сложности (complexity) социальных систем, которое проявляется в наличии между их элементами (индивидуумами или социальными группами) множественных связей с нетривиальной топологией. Поскольку сложные социальные системы, как правило, многомасштабны, обладают неоднородностью и выраженным нестационарным поведением, а также содержат внутренние механизмы самоорганизации и самоодерации, то динамические процессы в них обычно имеют яркую нелинейную природу. Это связано с возникновением критических состояний с последующими фазовыми переходами, с кардинальным изменением свойств системы на макроуровне (аналогично физическим системам) [1]. Для описания критических явлений в социальных системах применяются модели на основе методов равновесной и неравновесной статистической физики [2], включая многокомпонентное приближение среднего поля и многоуровневое уравнение Фоккера–Планка–Колмогорова. При этом специфика работы с социальными системами учитывается в них посредством таких механизмов, как теория социальных воздействий (social impact theory) [3] и модели голосования [4], а также эвристические модели возникновения коллективных эмоций [5–7]. Для численного моделирования динамики социальных групп применяются мультиагентные методы [8] и модели на основе клеточных автоматов [9]. Однако наибольшей общностью и детализацией с точки зрения исследования самой природы критических явлений обладает математический формализм комплексных сетей (КС) [10]. Основы КС заложены в работах таких исследователей, как М.Е.Д. Newman, А.Л. Barabasi, А. Vespignani, S.H. Strogatz, J.M. Kleinberg, и др., при этом наибольшее внимание ими уделено изучению сетей модельной структуры, для которых характеристики индивидуальных узлов (на микроуровне) и интегральные характеристики состояния КС в целом (на макроуровне) связаны аналитическими соотношениями. Численный аппарат моделирования КС и процессов на них [11, 12] развит несколько беднее.

Изучение сложных социальных систем на основе КС в общем случае требует применения современных суперкомпьютерных технологий, как в силу размеров самих сетей (в реальных системах – порядка 1 млрд узлов), так и в силу специфики развития критических явлений (чувствительность к проблеме масштаба, что приводит к необходимости моделировать сети на 1–2 порядка большего размера, чем исходная система). Однако известные алгоритмы для моделирования больших сетей [13, 14] применимы лишь для КС определенной структуры; широта их использования в общем случае дискуссионна. Во многом это связано с тем, что КС представляет собой сильно связанную структуру с динамически изменяющимися связями, т.е. теоретически любые два узла сети имеют возможность взаимодействовать между собой. Как следствие, это ограничивает возможности распараллеливания вычислений. Кроме того, критические явления, по своей сути, являются редкими событиями; поэтому для того, чтобы определить вероятность их появления, требуется модельное воспроизведение всего ансамбля возможных состояний КС. Это не только увеличивает общие требования к вычислительным ресурсам, но и отражается на эффективности их использования, поскольку ресурсоемкость задач моделирования одной и той же системы

в устойчивом и в критическом состоянии может существенно различаться. При этом использование результатов моделирования для оперативной поддержки принятия решений налагает достаточно жесткие временные ограничения на получение результата – от десятков минут до нескольких часов, что на текущем уровне техники и технологий не всегда возможно.

В данной статье комплексно рассматриваются различные аспекты создания математического, алгоритмического и программного обеспечения для моделирования критических явлений в сложных социальных системах на современных суперкомпьютерах петафлопсного диапазона производительности. В их основу заложен общий подход к суперкомпьютерному моделированию КС, подразумевающий совместное использование трех вложенных моделей.

«А» – Агрегированная аналитическая модель КС на макроуровне, позволяющая качественно оценить наличие критических состояний, представляется в виде системы дифференциальных уравнений относительно интегральных характеристик КС (например, распределения узлов по состояниям и распределения степеней вершин сети). Необходима для моделирования устойчивых состояний КС, перемежаемых критическими явлениями, обеспечивает, таким образом, «разгон» процесса формирования критической ситуации в социальной системе (создание соответствующих условий).

«Б» – Прямая модель КС на индивидуальном (микро-) уровне, на основе модельной топологии с достаточно простым воспроизводящим алгоритмом, в качестве начального состояния включает в себя результаты, полученные по приближенной аналитической модели: например, КС строится на основе эволюционирующего закона распределения степеней вершин. При этом задача моделирования распадается на две части: генерация самой КС с заданными характеристиками и эволюция динамического процесса в ней (включая изменение топологической структуры).

«В» – Детализированная прямая модель КС на индивидуальном (микро-) уровне, учитывающая неоднородность и иерархичность социальной системы, необходимая для описания динамики критических состояний. Целесообразность использования модели связана с тем, что для протекания критических явлений характерны специфические свойства сети: наличие кластерной структуры, определенной иерархии, центральностей и точек переходов. Моделями типа «Б», ориентированными на большие, но однородные сети, такие эффекты в целом не воспроизводятся.

Аналитическая модель «А» позволяет оценить поведение системы в устойчивом состоянии и определить условия наступления критических явлений, когда целесообразно использовать прямое моделирование «Б». Аналогично, модель «В» КС на модельной топологии обеспечивает воспроизведение динамического процесса (распространения информации) в КС и позволяет декомпозировать сеть на отдельные области, каждая из которых создается независимо на базе детализированной модели. Таким образом, на основе результатов расчетов по более грубым моделям может быть инициирован процесс планирования и балансировки нагрузки для детализированного расчета динамики КС, что и обеспечивает эффективность использования ресурсов суперкомпьютера.

#### Аналитические модели критических явлений в комплексных сетях

Определение возможности наступления критического состояния, исходя из логики «точек перелома» (tipping point) в модели КС, возможно путем использования ее обобщенного аналога в виде агрегированной аналитической модели. Для определения критических состояний в аналитической агрегированной модели общего вида можно использовать следующие критерии.

I. Критерий формального выполнения определения точки перелома в формулировке Ламберсона и Пейджа [15]. Пусть рассматривается агрегированная модель, представленная в виде системы разностных уравнений:  $x_{t+1} = F(x_t, y_t)$ ,  $x_t \in \Omega_x$ ,  $y_t \in \Omega_y$ . Задана функция  $L_\Delta(x_t, y_t) = x_{t+\Delta}$ , где  $\Delta \in \mathbb{N} \cup \{0\}$ . Точка  $\tau \in \Omega_y$  называется точкой перелома (tipping point) для  $x_t$ , если существует путь  $q: (-1, 1) \rightarrow \Omega_y$ :  $q(0) = \tau$ , такой что функция  $L_\Delta(x_t, q(s))$  является разрывной в точке  $s=0$  как функция аргумента  $s$  для некоторого значения  $\Delta$ . В частном случае, когда в качестве  $y_t$  рассматривается сама переменная состояний системы  $x_t$ , соответствующая точка перелома называется прямой (direct). В противном случае она называется контекстуальной (contextual).

II. Критерий наступления катастрофических бифуркаций (проход пороговых значений соответствующих бифуркационных параметров).

III. Критерии, связанные с предсказанием критических состояний по изменениям режимов динамики наблюдаемых переменных (эффект «critical slowing down» и подобные ему) [16].

Проблемой использования агрегированной модели «А» вместо модели на индивидуальном уровне «Б» для выявления критических ситуаций является неизбежное падение детализации описания процессов. Природа и величина этих изменений зависят от многих факторов, связанных со структурой исходной сетевой модели. Как следствие, поиск критических состояний по критериям (I–III) должен сопровождаться валидацией найденных условий на прямой модели КС (на микроуровне) с разделением их на три категории:

1. условия наступления критических состояний, совпадающие или незначительно отличающиеся для агрегированной и прямой моделей;
2. условия наступления критических состояний для агрегированной модели, которые позволяют сузить диапазон значений параметров сетевой модели;
3. условия наступления критических состояний для агрегированной модели, не отражающие поведение прямой модели.

На КС критерии определения критических явлений делятся на две основные категории по типу:

- с упором на конкретные типы взаимодействий в статических сетях на микроуровне;
- с упором на состояние системы на макроуровне (например, для явлений синхронизации, конденсации, перколяции) [17].

Исторически развитие аналитических моделей критических явлений в социальных системах базируется на методах статистической физики, включая теорию фазовых переходов Ландау, и использует термодинамические характеристики сети, такие как энергия, температура, энтропия. Так, задаваемой обобщенным функционалом энергии  $(\Omega(\gamma, \lambda))$  поверхности Гиббса может быть сопоставлен Парето-фронт, поверхность которого определяет своей формой фазовые переходы и критические точки [18]. Парето-фронт образован множеством решений оптимизационной задачи для множества целевых функций  $(t_1, t_2)$ , определяемых требованием (1):

$$\Omega(\gamma, \lambda) = \lambda t_1(\gamma) + (1 - \lambda)t_2(\gamma), \quad (1)$$

где  $\gamma$  – некоторая сеть;  $\lambda$  – управляющий параметр. При скачке  $\lambda$  такая модель отражает фазовый переход первого рода; фазовому переходу второго рода будет соответствовать равенство нулю производной параметра порядка по  $\lambda$ . Недостатком данного подхода является строгая привязка к определениям качества функционирования системы, без учета обобщенного изменения ее состояния.

Не требует формализации качества функционирования системы термодинамическая модель, ориентированная на квантификацию изменений КС, представляемой динамическим (темпоральным) графом [19]. Энергия ( $U$ ), температура ( $T$ ) и энтропия спектра нормированного лапласиана графа ( $H$ ) связаны уравнением (2):

$$\frac{1}{T} = \frac{dH}{dU}. \quad (2)$$

Лапласиан графа отражает его топологию, причем изоспектральные по нормированному лапласиану сети имеют более схожие свойства проводимости, чем изоспектральные по матрицам смежности графы. Энергия в рассматриваемой постановке соответствует числу ребер графа, а температура отражает меру изменения энтропии системы вместе с энергией за время между эпохами  $t$  и  $t'$  при неизменном числе вершин в графе. Аппроксимируя энтропию спектра нормированного лапласиана согласно (3),

$$H = 1 - \frac{1}{N} - \frac{1}{N^2} \sum \frac{1}{d_u d_v}, \quad (3)$$

где  $N$  – количество узлов графа, а  $d_u, d_v$  – степень вершин  $u, v$  соответственно, получаем представление (4):

$$\frac{1}{T(G, G')} = \frac{1}{\Delta|E|} \cdot \sum \left( \frac{1}{d_u d_v} - \frac{1}{d_u d_v} \right), \quad (4)$$

где  $T$  – температура,  $G, G'$  – состояния графа для эпох  $t, t'$  соответственно,  $E$  – множество ребер графа. Критическими предполагаются «точки перелома» на графиках температур, которые соответствуют смене темпа изменения сети. Таким образом, вычислительная сложность снижается за счет представления энтропии через попарные произведения степеней вершин (3). Предложенная аппроксимация, с другой стороны, опирается на разложение в ряд Тейлора, в результате чего полученный инвариант не может быть интерпретирован с точки зрения изоспектральности графов. В то же время прямое вычисление спектра с целью интерпретируемости применения данной методики на сетях большого размера неприменимо (фактически – это переход к модели «Б»). Следует отметить, что данная модель не учитывает ориентацию и вес ребер сети, а также состояния отдельных ее узлов, что усложняет ее применение к значительному классу задач.

Учет направленности графа в рамках поставленной задачи решается построением матрицы передачи информации на основании представляющего систему графа, которая призвана отобразить возможные потоки информации между узлами сети, что отличным от лапласиана образом отражает взаимодействия узлов [20]. Матрица содержит условные энтропии распространения информации между парами узлов (5) для некоторого момента времени в будущем с учетом конечного ряда прошлых состояний рассматриваемых узлов:

$$\begin{aligned} T_{X \rightarrow Y} &= H(y(t + \tau) | y^{(k)}(t)) - H(y(t + \tau) | y^{(k)}(t), x^{(l)}(t)) = \\ &= \sum_{y(t+\tau), y^{(k)}(t), x^{(l)}(t)} p(y(t + \tau), y^{(k)}(t), x^{(l)}(t)) \log \frac{p(y(t+\tau) | y^{(k)}(t), x^{(l)}(t))}{p(y(t+\tau) | y^{(k)}(t))}, \end{aligned} \quad (5)$$

где  $H$  – энтропия Шеннона;  $x, y$  – узлы сети;  $x(t)$  – состояние узла  $x$  в момент времени  $t$ ;  $x^{(l)}(t)$  – временной ряд состояний узла  $\{x(t-l), x(t-l+1), \dots, x(t)\}$ . Минимум спектра матрицы соответствует критической точке. Данный подход учитывает направления ребер графа и динамику распространения

процесса, однако необходимость хранения временных рядов состояний узлов и многочисленных условных вероятностей для определения энтропии распространения информации значительно усложняет вычисления для сетей большого размера.

Потенциальный эффект от взаимодействия пар узлов  $IP$  (6), принимающий в расчет как динамику изменения их состояний, так и топологию графа, может быть вычислен ресурсоемко при условии хранения относительных данных о состояниях узлов, таких как скорость изменения их состояний и линейно зависящая от скорости удаления узла от его критического состояния [21]:

$$IP(S) = \sum_{(i,j,w,m) \in E} \frac{|v_i|+w}{s_j} \cdot \mathbb{1}_{(\min\{r_i, r_j\}; +\infty)}(m) + \sum_{(i,j,w,m) \in E} w \cdot \left(\frac{1}{s_i} + \frac{1}{s_j}\right) \cdot \mathbb{1}_{(-\infty; \min\{r_i, r_j\}]}, \quad (6)$$

где  $S$  – текущее состояние системы;  $(s_1, \dots, s_n)$  – состояния каждого узла графа;  $i, j$  – вершина-источник и вершина-приемник соответственно;  $w$  – вес ребра;  $m$  – время жизни ребра в системе. При этом  $v_k$  и  $r_k$  – скорость и удаленность узла  $k$  соответственно.

Подавляющее большинство исследований, посвященных изучению «ранних сигналов» фазовых переходов и определений самих критических точек, предлагают некую характеристику системы или инвариант графа, динамика значений которых была бы согласована с динамикой изменения системы. При этом часть из них склонна приводить критерий качества функционирования исследуемой системы, в то время как другие опираются лишь на интенсивность изменений, в связи с чем критерием наступления критического состояния (в этих случаях) является резкое изменение закономерностей развития предлагаемых инвариантов.

Демонстрацией применения аналитических моделей для быстрого обнаружения критических состояний системы, требующих более детального моделирования, может служить агрегированная модель развития эпидемии ОРВИ/гриппа. Модель описывает два режима заболеваемости острых респираторных вирусных инфекций (ОРВИ) в течение года: межэпидемический, когда заболеваемость незначительна и тяготеет к стационару, и эпидемический. При этом развитие эпидемии рассматривается как критическое явление в контактной КС, и ему предшествует точка перелома, связанная с началом эпидемии (рис. 1). Учитывая, что каждый год время наступления роста заболеваемости может меняться, такой механизм позволяет гибко отслеживать ситуацию и инициировать процедуру перехода от агрегированной модели «А» к детализированной модели «Б».

Сезонная заболеваемость ОРВИ в Санкт-Петербурге с 1 июля 1987 г. по 30 июня 1988 г.

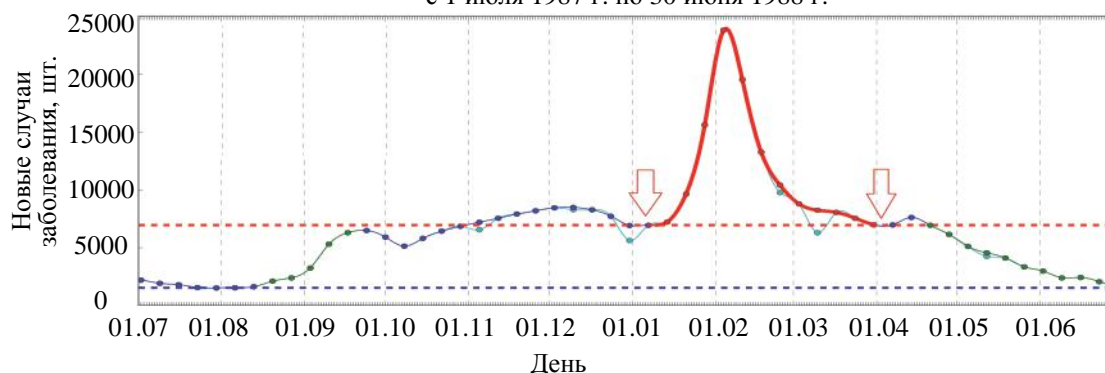


Рис. 1. График ежедневной заболеваемости гриппом в Санкт-Петербурге с 1 июля 1987 г. по 30 июня 1988 г. Голубым цветом обозначены отклонения в данных, вызванные недоучетом заболевших органами здравоохранения в праздничный период. Синие и красные горизонтальные линии отмечают средние уровни заболеваемости ОРВИ в летний и осенне-весенний периоды соответственно. Красная сплошная линия отмечает заболеваемость в период эпидемии гриппа. Моменты возникновения критических состояний (переходов от сезонного ОРВИ к эпидемии и обратно) отмечены стрелками

### Численное моделирование динамики комплексных сетей на индивидуальном уровне

Существующие методы прямого моделирования динамики процессов в КС ориентированы на итеративное воспроизведение взаимодействий вершин в соответствии с набором изначально заданных правил. Результатом таких взаимодействий может становиться как изменение переменных состояний вершин (бинарных – при моделировании случайных булевых сетей [22], категориальных – при моделировании эпидемиологических процессов [23], вещественных – при моделировании финансовых сетей [24]), так и модификация топологии сети за счет добавления/удаления вершин и связей между ними. Значительное увеличение вычислительной емкости задач моделирования процессов на сетях, требующее применения высокопроизводительных вычислений, возникает преимущественно в рамках следующих трех сценариев.

1. С1. Небольшой размер сети, простые правила изменения состояний, большое число итераций моделирования. Пример – адаптивные случайные булевы сети, где число итераций при поиске аттрактора растет экспоненциально с ростом размера сети.
2. С2. Большой размер сети, простые правила изменения состояний. Пример – распространение инфекции на сетях размером  $2^{20} - 2^{30}$  узлов.
3. С3. Наличие правил эволюции сети, требующих вычислительно интенсивных расчетов для определения результата взаимодействия вершин.

Известные методы параллельного моделирования динамики процессов для сценариев С1–С3 (см., например, [25–27]) осуществляют итеративный расчет состояний вершин для распределенно-храняемой сети, что имеет ограниченную масштабируемость для большого количества вычислительных узлов (ВУ). Во избежание этой проблемы целесообразно использовать механизм моделирования на основе стохастических графов Кронекера (СГК) [14]. Его преимуществом является возможность применения фрагмента КС (сэмпла, sample) малого размера для генерации всей сети, что:

- позволяет использовать сэмплы сетей реального мира, имеющие ограниченный размер вследствие принципиальной частичной наблюдаемости [28] таких сетей;
- естественным образом снижает вычислительные требования к ресурсоемким моделям «В». Более того, при параллельной реализации модели «В» свойства кронекеровских графов позволяют выполнять независимую генерацию подсетей одновременно на нескольких вычислительных узлах. Таким образом, в рамках модели «В» воссоздается сеть заданного размера по образцу сети реального мира или по сэмплу, сгенерированному одним из указанных способов.

Специфичными для выбранного способа реализации модели «В» являются:

- возможность генерации сети одновременно с моделированием динамики процесса;
- возможность достижения хорошей сбалансированности вычислительной и коммуникационной нагрузки при параллельном моделировании динамического процесса за счет использования свойств кронекеровских графов, что будет показано ниже.

Для реализации модели «В» на основе сэмпла заданного размера выполняется оценка кронекеровской иницирующей матрицы (ИМ). ИМ оценивается по методу максимального правдоподобия, после чего по ней с использованием операции кронекеровского произведения может быть получена матрица вероятностей ребер между вершинами сети.

Поскольку ИМ формально обеспечивает воспроизведение сети того размера, по которому она была оценена, необходимо масштабирование ИМ с целью учета характера изменения топологии с ростом сети. Для этого вводится алгоритм масштабирования ИМ, предназначенный для подстройки характеристик (числа ребер, максимальной степени вершины) полноразмерной сети при ее генерации по сэмплу меньшего размера. Алгоритм валидирован в рамках статистической теории экстремальных значений, исходя из соответствия ожидаемых и фактических распределений максимальных степеней вершин в КС (рис. 2). Согласно теории, при возрастании размера выборки (т.е. размера сети) в  $K$  раз функция распределения максимумов (Generalized Extreme Valuedistribution, GEV) должна быть возведена в степень  $K$ . На рис. 2, а, проиллюстрировано масштабирование GEV для сэмплов из сети Барабаши–Альберт, размер которых различается в два раза. Полученные результаты подтверждают совпадение эмпирического распределения с масштабированным. На рис. 2, б, приведен пример применения масштабирования GEV для сети реального мира. Из выбранной сети (Autonomous systems) было получено 250 сэмплов размера 200, по которым оценены параметры GEV, затем осуществлено масштабирование GEV для возрастающих размеров сэмплов (график fitted). Так как по факту распределение максимальных степеней является усеченным (степень не может быть больше числа узлов в сети), распределения GEV при масштабировании также усекались по размеру сети. Масштабированные распределения сравнивались с эмпирическими (empirical), полученными путем сэмплирования из исходного графа. Приведенные результаты демонстрируют воспроизводимость эмпирического распределения максимальных степеней по GEV, оцененной для набора сэмплов меньшего размера.

СГК может быть сгенерирован по ИМ практически полностью параллельно [29–30], однако схема распределения сети по вычислительным процессам (ВП), реализуемая в ходе такой процедуры, непригодна для эффективного параллельного моделирования динамики процессов на КС. Это связано с тем, что предложенная в [14] параллельная версия СГК реализует модель хранения сети, ориентированную на ребра (так называемую edge-centric, EC). EC-модель эффективна для реализации сетей заданной топологии, однако ребра, инцидентные вершинам, могут храниться разрозненно на разных ВУ, что критично с точки зрения моделирования динамических процессов. Для преодоления указанных ограничений авторами [30] были предложены пуассоновские СГК (ПСГК). ПСГК позволяют генерировать ребра отдельно для каждой вершины по ее индексу и тем самым поддерживают вершино-ориентированную (vertex-centric, VC) модель хранения сети. Поскольку в [30] показано, что ПСГК наследуют топологические свойства СГК, то для распараллеливания процесса генерации достаточно назначить имеющимся ВУ различные диапазоны индексов вершин КС, после чего сгенерировать фрагменты сети параллельно. Более

того, описанная возможность позволяет добавлять вершины в сеть по мере необходимости, т.е. генерировать КС параллельно с протеканием процесса по ней.

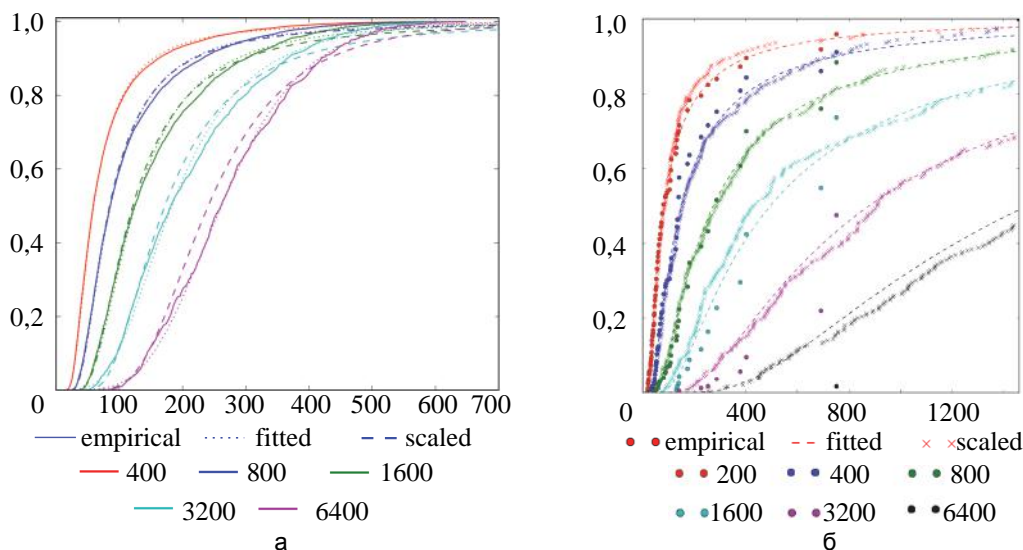


Рис. 2. Масштабирование GEV для сэмплов различных размеров, полученных из графа Барабаши–Альберт размером 25600 узлов (empirical – эмпирическое распределение, fitted – оценка GEV по эмпирическому распределению, scaled – распределение, полученное возведением в квадрат fitted для сэмпла вдвое меньшего размера). Усреднение по 1000 запускам (а); масштабирование GEV для сэмплов из графа AutonomousSystems (роутеров в сети Интернет), размер графа 6400 узлов. Размер набора сэмплов для оценки GEV – 250 (б)

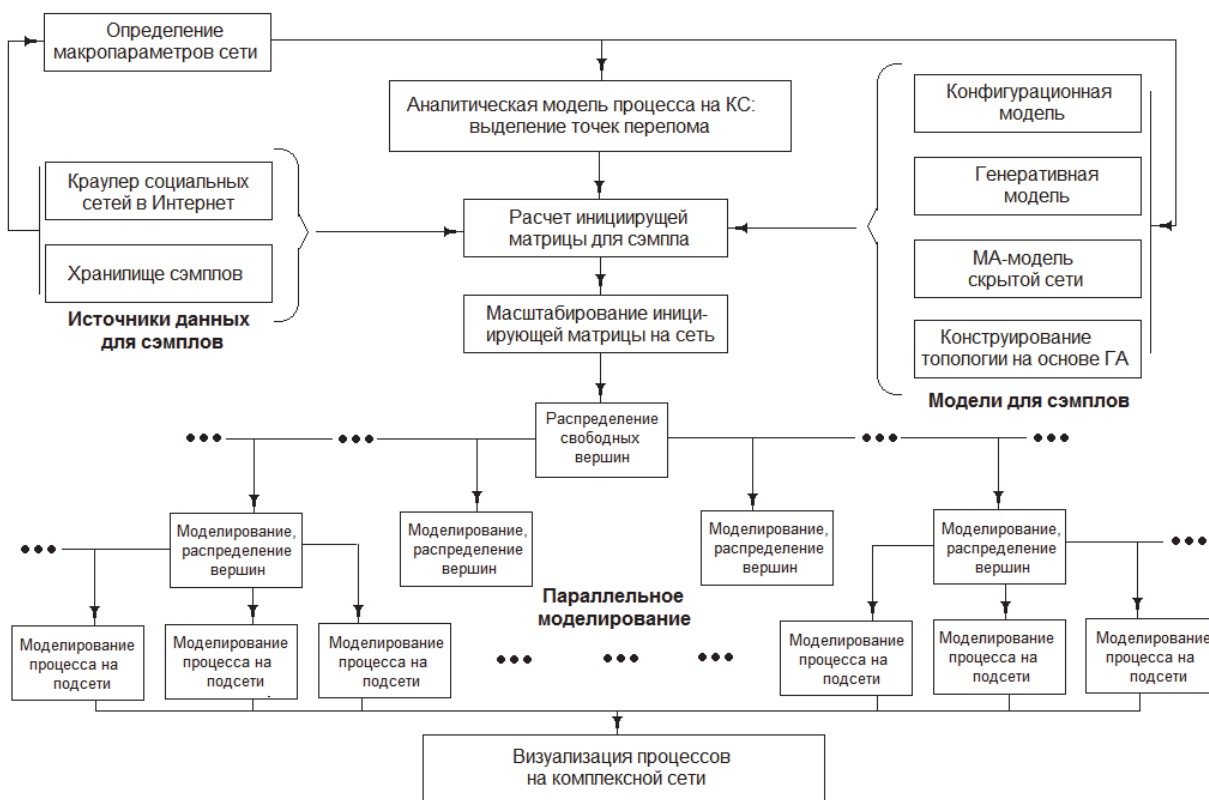


Рис. 3. Общая методика моделирования критических состояний в сложных социальных системах на основе формализма комплексных сетей с использованием моделей «А»–«В» (ГА – генетический алгоритм)

Алгоритм моделирования на основе SGK [14] целесообразно использовать для совместной реализации моделей «Б» и «В», при этом модель «В» используется только для воспроизведения относительно небольших сэмплов на каждом вычислительном узле параллельно. В целом данный подход к численному



моделированию самой КС на микроуровне позволяет органично отобразить исходную логику сочетания моделей «А»–«В» (рис. 3). Аналитическая модель «А» идентифицируется по макрохарактеристикам самого явления (построенным, в том числе, на фрагментарных данных – наблюдаемых сэмплах) и используется для выявления возможных критических состояний (в логике точек перелома). Источником сэмплов служат как сети реального мира (имеющиеся в базе или же полученные с применением средств автоматизированного сбора данных), так и сети, синтезированные по одной из моделей. По сэмплу строится ИМ, на основе которой воспроизводится пуассоновский граф Кронекера и выполняется параллельное моделирование процесса, приводящего к критической ситуации. Результаты расчетов могут быть интерпретированы и визуализированы непосредственно в динамике.

Моделирование «В» сэмплов сетей с заданными свойствами подразумевает воспроизведение специфических топологических или функциональных свойств сети, влияющих на процесс развития критических явлений. Помимо наиболее популярного свойства – распределения степеней вершин, к важным топологическим свойствам относятся малый диаметр (Small diameter), распределение числа ближних пар (Hop-plot), количество треугольников и малых сообществ и пр. В простейшем варианте сэмпл можно получить, выделив фрагмент сети реального мира. Выделение подсети выполняется путем обхода графа (сети) в ширину. Это позволяет сохранять в выделенном фрагменте естественные пути распространения критического процесса. Такой подход характерен для формирования сетевых структур на основе краулинга web-ресурсов, когда получение графа связей целиком является слишком затратным из-за технических ограничений при доступе к самим ресурсам.

Для воспроизведения сетей, имитирующих реальные процессы, популярным подходом является использование случайных графов и различных модификаций алгоритмов их генерации. В контексте математического формализма термин «случайный граф» традиционно относится к модели случайных графов Эрдеш–Реньи (Erdos&Renyi) [31]. Другими популярными моделями являются модель малого мира WS (Watts&Strogatz) [32] и модель предпочтительного добавления BA (Barabasi&Albert) [33], которые позволяют воспроизводить такие важные свойства сетей, как средние длины путей между узлами, безмасштабность, коэффициент кластеризации и другие. Общим недостатком этих подходов является необходимость выражать желаемые свойства итоговой сети через параметры алгоритма генерации, что не всегда возможно или имеет серьезные ограничения. В этом случае свойства сети, не заложенные в модель, воспроизводятся алгоритмом генерации неустойчиво и имеют значительный разброс при изменении параметров генерации. Случайные графы тесно связаны с теорией перколяции, которая описывает их связность. Математически перколяция основана на удалении части ребер сети в соответствии с определенными правилами, что позволяет получить сети с новыми, в том числе желаемыми, свойствами. Кроме того, используются методы на основе воспроизведения характерных макрохарактеристик сети, например, конфигурационный алгоритм [34], ориентированный на модельное воссоздание одномерного распределения степеней вершин.

Для более точного воспроизведения сети с заданными свойствами, в том числе с учетом протекающих процессов, используется модель на основе метаэвристических алгоритмов, когда структура сети является объектом эволюции [35]. В данном случае особь генетического алгоритма представляет собой матрицу смежности моделируемой сети. Начальные особи генерируются с помощью традиционных моделей случайного графа, а процедура мутации особи выполняется путем добавления или удаления ребра между случайной парой вершин, а также с помощью перевязки случайного ребра. Скрещивание производится путем обмена между особями подматриц матриц смежности. Главным преимуществом данного подхода является гибкость алгоритма генерации, который позволяет воспроизводить как топологические свойства сетей, так и их функциональные свойства, связанные с особенностями протекающих в сетях процессов. Это позволяет, в том числе, получать сети с оптимизированный под заданный процесс структурой. Главный недостаток подхода связан с высокой вычислительной сложностью алгоритмов, что является следствием применения эволюционных вычислений, традиционно подразумевающих вычисление большого числа альтернативных вариантов для сужения пространства поиска решений.

В качестве альтернативного пути построения сети-прототипа с заданными топологическими свойствами используют мультиагентные (МА) модели [36]. Вершинами таких сетей являются агенты, а ребра генерируются при возникновении контакта между агентами. Термин «контакт» в данном случае имеет широкую интерпретацию и может подразумевать набор различных правил взаимодействия между агентами. Простейший вариант – это попадание одного агента в зону, ограниченную радиусом другого агента, с учетом возможных препятствий, таких как, например, стены. Более сложным вариантом является, например, тесный контакт между агентами, такой как рукопожатие или общение в течение некоторого промежутка времени. Такие сети используются для получения реалистичной картины взаимодействия между людьми в реальном мире с учетом темпорального характера самих связей, которые возникают в разный момент времени, без необходимости запускать трудоемкое МА-моделирование каждый раз, когда нужно оценить характер протекания какого-либо процесса в обществе. Примерами таких процессов являются распространение слухов, контактных болезней, обмен мнениями и т.п. Наличие контактной сети незави-



симо от запуска моделирования позволяет исследовать многие процессы с учетом случайного характера самих процессов взаимодействия и неопределенности в начальных условиях.

Выбор того или иного алгоритма моделирования сэмплов сетей с заданными свойствами зависит от конкретной задачи и наличия априорных знаний о топологических или функциональных свойствах сети. Когда имеются обширные экспериментальные данные, такие как, например, граф связей социальных медиа, целесообразно получить фрагмент сети путем обхода сети в ширину или использовать алгоритмы на основе случайных графов, а для ненаблюдаемых, например, темных (dark) сетей или множества агентов, имитирующих жизнь общества, МА-моделирование является единственно возможным способом получения структуры сети. При необходимости получения большого числа сэмплов с одинаковыми свойствами основной подход связан с алгоритмами случайной генерации графов на основе макрохарактеристик сети, заданных в агрегированной форме (иницирующая матрица, плотность распределения степеней вершин). Самым точным, но и одновременно самым трудоемким является подход на основе метаэвристических алгоритмов, который, в отличие от других подходов, позволяет воспроизводить не только топологические, но и функциональные свойства сетей, оптимизируя их структуру.

### Параллельный алгоритм прямого моделирования комплексных сетей

Из рис. 3 видно, что основной потенциал для распараллеливания доступен на уровне модели «Б»; так, ресурсоемкость агрегированной модели «А» не зависит от числа вершин КС, а модель «В» применяется лишь к небольшому фрагменту (сэмплу) исходной КС. Алгоритм распространения процессов в КС в рамках модели «Б» (при наличии предварительно оцененной и масштабированной ИМ) предусматривает постепенное увеличение размеров КС за счет добавления в нее вершин, инцидентных активным на текущем шаге. Процесс на сети моделируется пошагово с использованием принудительной барьерной синхронизации после каждого шага, позволяющей гарантировать завершение приема/передачи всех сообщений на текущем шаге. Для каждого ВУ задается диапазон индексов вершин сети, для которых данный ВУ будет генерировать исходящие ребра (а также хранить информацию об истории состояний вершин).

Таким образом, общую логику функционирования одного шага алгоритма для одного ВУ можно описать следующим образом:

1. для каждой активной вершины определить списки дочерних вершин, которые подлежат активации;
2. активировать вершины, принадлежащие ВУ;
3. сформировать и отправить сообщения на активацию для других ВУ;
4. принять сообщения от других ВУ и активировать вершины.

Базовый алгоритм, предложенный в [14], генерирует СКГ по ребрам и имеет сложность  $O(E)$  (где  $E$  – число ребер). Вершины, между которыми будет добавлено ребро, определяются путем  $N_k$  рекурсивных выборов регионов матрицы с вероятностями ребер. В [30] показано, что указанная процедура может производиться в два шага, с выбором вершины-источника на первом шаге и вершины-приемника на втором шаге, причем выбор вершины-источника аппроксимируется генерацией случайной величины, имеющей мультиномиальное распределение. Основанием для перехода от СКГ к ПСГК является замена мультиномиальной случайной величины независимыми одинаково распределенными пуассоновскими случайными величинами. ПСГК позволяет определить для вершины с заданным индексом число исходящих из нее ребер, после чего найти их вершины-приемники. Благодаря этому становится возможной поддержка вершино-ориентированной модели распределенного хранения сети. Двумя главными факторами, определяющими масштабируемость параллельного алгоритма моделирования динамики процесса на сети, являются:

- способ распределения вершин между вычислительными процессами (ориентирован на снижение вычислительной и коммуникационной неоднородности);
- способ организации коммуникаций между подсетями.

Для заданного способа распределения вершин ПСГК по ВП эффективность параллельного моделирования процесса на ПСГК определяется паттерном коммуникации между подсетями. Способ обмена «все–всем» («all-to-all») плохо масштабируется вследствие квадратичного роста числа операций обмена с ростом размера коммуникатора. В [25] описано применение парадигмы «master-slave» с одним мастер-процессом для коммуникаций между подсетями при моделировании эпидемии на сети. С ростом размера коммуникатора увеличивается нагрузка на мастер-процесс, осуществляющий маршрутизацию сообщений. В данной работе мы используем схему с переменным числом мастер-процессов с целью достижения баланса между нагрузкой на мастер-процессы и сокращением числа операций коммуникации.

Схема организации ВП представлена на рис. 4, а. Процессы, осуществляющие маршрутизацию сообщений, обозначены как  $M$  (Master, мастер), а процессы, хранящие подсети и производящие активацию вершин в ходе моделирования, обозначены как  $L$  (Leaf, лист). С каждым  $M$ -процессом ассоциировано некоторое число  $L$ -процессов. В конце каждой итерации  $L$ -процесс передает информацию о вершинах, активированных в других подсетях, ассоциированному  $M$ -процессу.  $M$ -процесс принимает эти данные от своих листов и передает их как ассоциированным, так и неассоциированным листам. До начала модели-

рования  $L$ -процесс (рис. 4, б, шаги 1\*–1) принимает индексы изначально активированных вершин и активирует их. На каждой итерации каждая активная вершина пытается активировать соседей (шаг 2 на рис. 4, б). Если вершина активируется впервые, генерируются и сохраняются ее соседи. Если вершина, подлежащая активации, находится в другой подсети (шаг 2в на рис. 4, б), она помещается в пул вершин для передачи другому листу. Для того чтобы мастер-процесс мог выделить память под консолидированные пулы, по завершении формирования пулов на листах они отправляют своему  $M$ -процессу сообщение с размерами пулов, предназначенных для других подсетей (шаг 3 на рис. 4, б, и шаг 1 на рис. 4, в). После того, как память выделена,  $M$ -процесс принимает пулы от ассоциированных листов (шаг 4 на рис. 4, б, и шаг 3 на рис. 4, в) и отправляет консолидированные пулы  $L$ -процессам (шаг 5 на рис. 4, б, и шаг 4 на рис. 4, в). После того как  $L$ -процесс получает пулы от всех мастеров, он активирует вершины из этих пулов в своей подсети (шаг 7 на рис. 4, б).

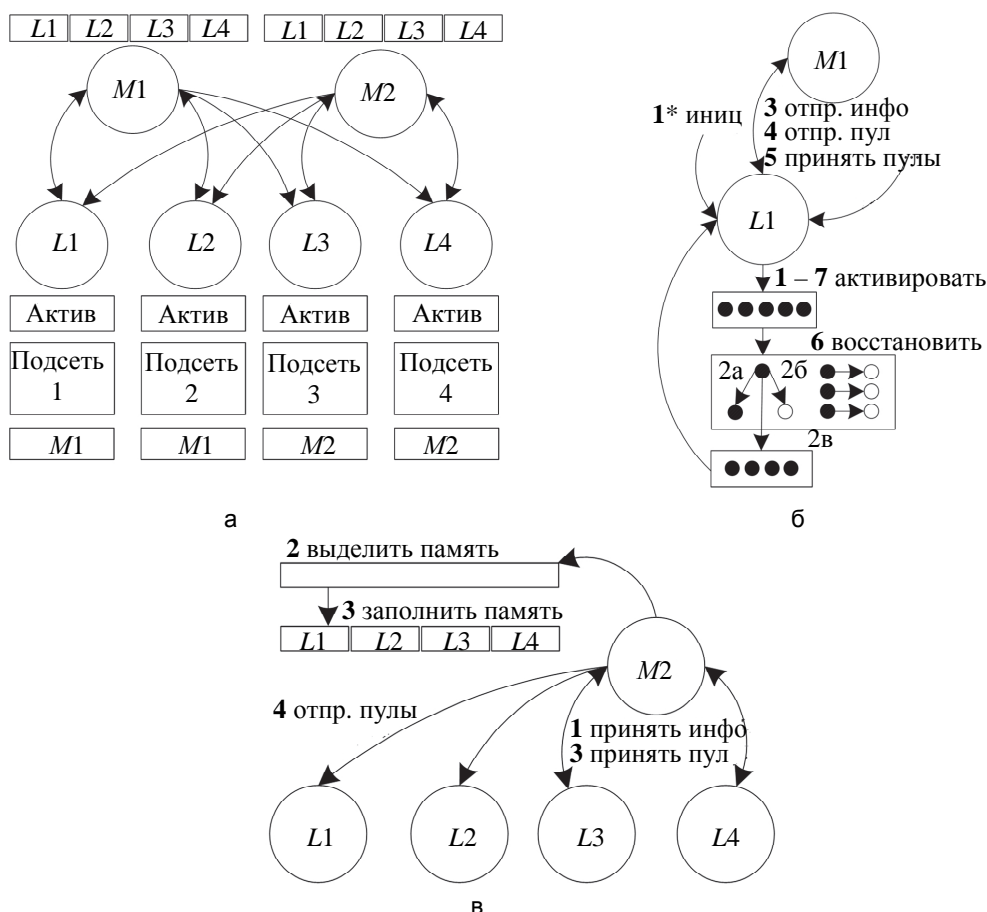


Рис. 4. Организация коммуникации между подсетями (а); алгоритм функционирования  $M$ -процесса (б); алгоритм функционирования  $L$ -процесса (в)

Формальное описание алгоритмов для  $M$ -процесса и  $L$ -процесса приведено ниже.

Алгоритм 1. Мастер-процесс.

- |  |  |
|--|--|
| 1: <b>if</b> ( $rank == 0$ )                   | 10: $Ready = 0$                                  |
| 2: <b>for each</b> $l$ in $Leafs$ :            | 11: <b>while</b> $Ready \neq Children.size()$ :  |
| 3:         Отправить $APool$ $l$               | 12:         получить ранг отправителя            |
| 4: <b>for</b> $i \in 1 \dots N$ <b>do</b>      | 13:         рассчитать смещение                  |
| 5: $Ready = 0$                                 | 14:         сообщения в $MPool$                  |
| 6: <b>while</b> $Ready \neq Children.size()$ : | 14:         получить сообщение в $MPool$         |
| 7:         принять инф. сообщение,             | 15:         со смещением, инкр. $Ready$          |
| 8:         инкрементировать $Ready$            | 16: <b>for each</b> $l$ in $Leafs$ :             |
| 9:         определить размер $MPool$           | 16:             создать тип данных $DestType$    |
|  | 17:             отправить $l$ сообщение типа     |
|  | 17: $DestType$                                   |
|  | 18:             освободить $MPool$ и $DestTypes$ |

Алгоритм 2. Листовой процесс.

```

1:   принять APool
2:   for  $i \in 1 \dots N$  do
3:      $ASize = APool.size()$ 
4:     for each  $j \in 0 \dots ASize - 1$ :
5:        $v = ActivePool[j]$ 
6:       if  $v$  добавл. в  $S$  на шаге  $i - 1$ :
7:         добавить соседей  $v$  в  $S$ 
8:       for each соседа  $nb$  верш.  $v$ :
9:         if  $nb$  должна быть
           активирована:
10:          if  $nb \notin S$ :
11:             $dest = get\_dest(nb)$ 
12:             $DPools[dest].add(nb)$ 
13:          else infect( $nb$ )
14:        for each  $DPool$  in  $DPools$ :
15:          copy( $DPool, LPool$ )
16:        отправить информационное
           сообщение родителю
17:        отправить  $LPool$  родителю
18:         $Ready = 0$ 
19:        while  $Ready \neq Masters.size()$ :
20:          получить сообщение в  $RPool$ 
21:          for each  $v$  in  $RPool$ :
22:            infect( $v$ )
23:          for each  $j \in 0 \dots ASize - 1$ :
24:            repair( $APool[j], i$ )
1:   procedure infect( $v$ ):
2:     if  $v$  может быть активирована
           и  $v \notin APool$ :
3:        $APool.add(v)$ 
1:   procedure repair( $v$ , шаг  $i$ ):
2:     if  $(rand(0,1.0) \leq p_{rec})$ 
3:        $APool.remove(v)$ 
4:       if  $(rand(0,1.0) \leq p_{res})$ :
5:         установить время
           возможной активации
           равным  $i + R_t + 1$ 

```

Используются следующие обозначения: *Leafs* – набор  $L$ -процессов, *Masters* – набор  $M$ -процессов, *Children* – набор ассоциированных  $L$ -процессов для  $M$ -процесса,  $S$  – набор вершин, добавленных в подсеть для  $L$ -процесса, *APool* – пул активированных вершин, *DPools* – набор пулов, предназначенных для разных процессов, *LPool* – консолидированный пул  $L$ -процесса, *MPool* – консолидированный пул  $M$ -процесса. Функция  $get\_dest(v)$  возвращает назначение для вершины  $v$ . Процесс с рангом 0 передает изначально активированные вершины  $L$ -процессам (шаги 1–3 Алгоритма 1). После этого на каждой итерации  $M$ -процесс получает информационные сообщения (шаги 5–7), выделяет память и принимает сообщения в *MPool* (шаги 9–14), после чего отправляет сообщения листам. Сообщения от *Children* принимаются в порядке поступления.

$L$ -процесс (Алгоритм 2) принимает изначально активированные вершины (шаг 1). На каждой итерации предпринимаются попытки активации соседей вершин из *APool* (шаги 5–13). При необходимости вершины добавляются в подсеть (шаги 6–7). Вершины, принадлежащие подсетям другим листов, помещаются в соответствующие пулы (шаги 11–12). После отправки информационного сообщения  $M$ -процессу (шаг 16) осуществляется отправка консолидированного пула *LPool*. Сообщения от мастер-процессов также обрабатываются в порядке поступления (шаги 19–22). После обработки сообщений от  $M$ -процессов, листовой процесс осуществляет деактивацию некоторых вершин из *APool* (за исключением вновь принятых), если такая логика предусмотрена рассматриваемым динамическим процессом (пример реализации процедуры деактивации `repair` приведен в Алгоритме 2, здесь  $p_{rec}$  – вероятность деактивации,  $p_{res}$  – вероятность получения иммунитета на  $R_t$  шагов).

Эффективность алгоритма напрямую зависит от сбалансированности вычислительной нагрузки между  $L$ -процессами, что обусловлено неявной синхронизацией между листами на этапе получения пулов от мастер-процессов. Одной из характеристик СГК является зависимость ожидаемой выходной степени  $E[d_u]$  вершины  $u$  от ее индекса ( $d_u$  – степень вершины  $u$ ) [37]:

$$E[d_u] = (a + b)^{N_k - w} \cdot (b + c)^w, \quad (7)$$

где  $a = \theta_{00}, b = \theta_{01} = \theta_{10}, c = \theta_{11}$ ,  $a \geq b \geq c$  и  $w$  – число единиц в двоичной записи индекса вершины  $u$ . Формула (7) представляет собой выражение для ненаправленного графа (так как  $\theta_{01} = \theta_{10}$ ). Условие  $a \geq b \geq c$  ограничивает рассмотрение наиболее изученным типом СГК. Выражение (7) показывает, что для симметричной относительно главной диагонали ИМ с  $a \geq b \geq c$  ожидаемые степени вершин монотонно убывают с возрастанием их индексов (так как  $a + b \geq b + c$ , вершина с меньшим  $w$  имеет меньшую ожидаемую степень). Вследствие этого блочное и построчное распределение вершин по ВП неэффективно: в обоих случаях сумма ожидаемых ребер (а также вычислительная нагрузка по генерации и обработке вершин) будет уменьшаться с увеличением ранга  $L$ -процесса.

Как следствие, целесообразно распределять вершины между ВП с использованием простой эвристики (Алгоритм 3). Вершины назначаются листам последовательно в цикле, однако на каждой итерации цикла направление обхода листов меняется. Интуитивно  $L$ -процессы, получившие вершины с наибольшими ожидаемыми степенями на предыдущей итерации, получают на следующей итерации вершины с

наименьшими степенями, и наоборот.

*Алгоритм 3. Поиск индекса L-процесса для вершины u*

**Входные данные:**  $u$  — индекс вершины,  $V$  — число вершин,  $LN$  — число  $L$ -процессов

- |   |   |
|---|---|
| 1: $BlockIndex = V \operatorname{div} LN$ | 4: <i>else:</i>                                     |
| 2: <b>if</b> $BlockIndex$ четный:         | 5: $LeafIndex = (LN - 1) - u \operatorname{mod} LN$ |
| 3: $LeafIndex = u \operatorname{mod} LN$  | 6: <b>return</b> $LeafIndex$                        |

### Экспериментальные исследования масштабируемости параллельного алгоритма

Для исследования масштабируемости параллельного алгоритма, описанного выше, в качестве исходных использовались ИМ, оцененные для синтетических образцов сетей Барабаши–Альберт (BA), Эрдеша–Реньи (ER) и Ваттса–Строгатца (WS) с помощью алгоритма KronFit [14]. Также была использована ИМ [0,987; 0,571; 0,049], оцененная в [26] для сети автономных систем Интернет (обозначается как *as-route*).

В качестве примера динамического процесса, моделируемого на ПСКГ, рассматривался SIRS<sup>1</sup>-подобный процесс. Предположим, что задана ИМ  $\theta = \{\theta_{ij}\}$ ,  $0 \leq i, j \leq 1$ , полученная методом максимального правдоподобия для некоторой сети-образца (оценка ИМ может быть произведена с помощью алгоритма KronFit). Для получения полноразмерного СКГ также необходимо знать  $N_k$  – число раз, которое нужно умножить ИМ саму на себя (с помощью произведения Кронекера) для получения матрицы с вероятностями ребер.

По аналогии с SIRS-моделью мы выделяем три возможных состояния вершины:

1. активная (А) – вершина присутствует в сети и может активировать соседей (кроме активных и устойчивых);
2. устойчивая (У) – вершина присутствует в сети, и через нее не может осуществляться распространение процесса;
3. потенциальная (П) – вершина, которая не является ни активной, ни устойчивой.

Вершина в состоянии А может активировать каждого из ее соседей, находящихся в состоянии П, с вероятностью  $p_{act}$ . На каждой итерации активная вершина может стать потенциальной с вероятностью  $p_{rec}$ . В случае если происходит переход А→П, вершина также может стать устойчивой на  $R_t$  итераций с вероятностью  $p_{res}$ . Также задается часть  $P_{ini}$  изначально активированных вершин. Целью предложенного алгоритма для заданных иницирующей матрицы  $\theta$  и параметров  $p_{act}, p_{rec}, p_{res}, R_t, P_{ini}$  является параллельное моделирование  $N$  итераций процесса, меняющего состояние вершин в ходе своего распространения, на динамически генерируемой сети.

Для сокращения времени исполнения последовательной версии алгоритма сумма элементов рассматриваемых ИМ нормировалась на некоторую константу  $C$ . Все результаты приводятся с усреднением по трем запускам. Эксперименты были проведены на Linux-сегменте суперкомпьютера «Лобачевский» (ННГУ им. Лобачевского, г. Нижний Новгород). Один узел данного сегмента содержит 2 процессора Intel Sandy Bridge E5-2660 2.2GHz с 8 ядрами и 64 GbRAM.

Рис. 5 демонстрирует параллельную эффективность (отношение параллельного ускорения к числу процессоров) для *as-route*, ER и BA ИМ для сетей от  $2^{25} \approx 33$  миллиона до  $2^{30} \approx 1$  миллиарда вершин на 1–24 узлах. Были протестированы топологии для 1–21 мастер-процесса, на рис. 5 представлены результаты для конфигураций, имеющих наивысшую эффективность. Значительность преимущества для случая с 8 процессами объясняется тем, что в этом случае отсутствовали коммуникации между узлами. Средняя эффективность для *as-route* варьируется от 0,33 до 0,79, для ER – от 0,2 до 0,7, для BA – от 0,23 до 0,76 и имеет тенденцию расти с ростом размера сети.

На рис. 6, а, приведены число активированных вершин и время активации для BA ИМ и 7  $L$ -процессов. Похожесть указанных характеристик для разных листов подтверждает эффективность алгоритма распределения подсетей по ВП. Рис. 6, б, иллюстрирует, что увеличение числа мастер-процессов способно значительно улучшить результирующую параллельную эффективность (например, для 384 процессов на рис. 6, б, параллельное ускорение возросло с 70 для 1 мастер-процесса до 144 для 19 мастер-процессов).

<sup>1</sup>SIRS (Susceptible–Infected–Recovered–Susceptible) – модель, представляющая распространение инфекционного процесса как переход индивидов между состояниями «восприимчивый – инфицированный – имеющий иммунитет – восприимчивый».

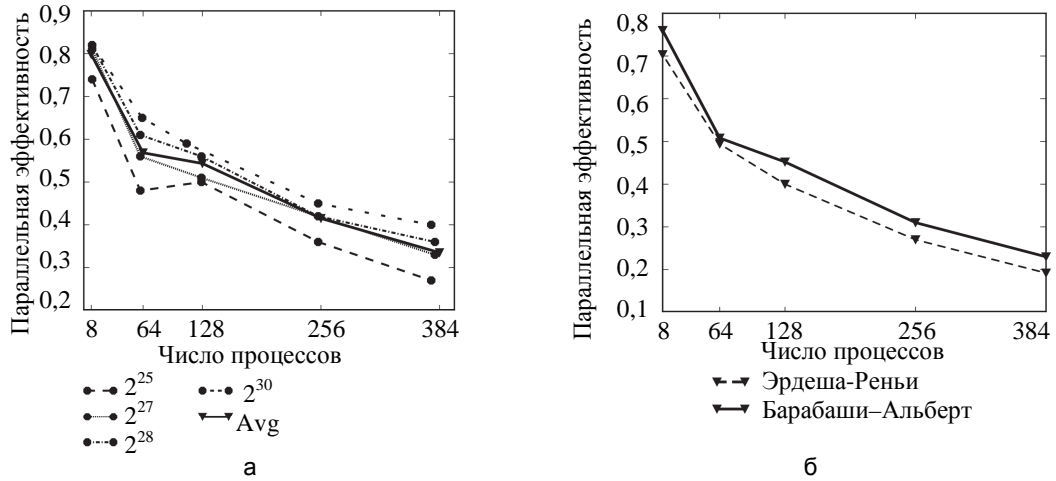


Рис. 5. Параллельная эффективность: иницирующая матрица *as-route*,  $p_{act} = 0,5$ ,  $p_{rec} = 0,2$ ,  $C = 2,05$ ,  $P_{ini} = 0,01$ ,  $N = 25$ , Avg – усреднение по различным размерам графов (а); иницирующие матрицы Эрдеша-Реньи,  $p_{act} = 0,4$ ,  $p_{rec} = 0,142$ ,  $C = 2,05$ ,  $P_{ini} = 0,1$ ,  $N = 50$  и Барабаши-Альберт (параметры см. на рис. 5, а) (б)

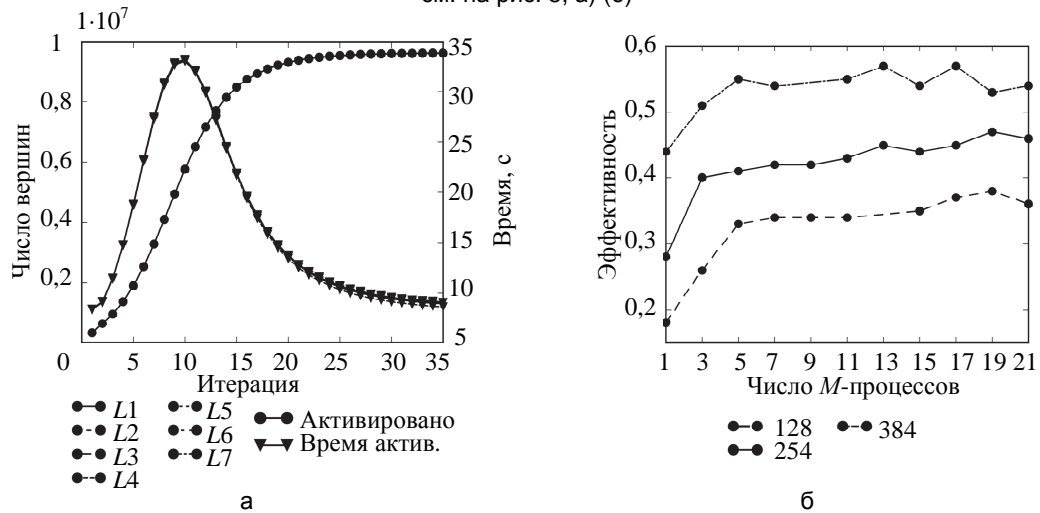


Рис. 6. Число активированных вершин и время активации для разных  $L$ -процессов для иницирующей матрицы Барабаши-Альберт,  $p_{act} = 0,3$ ,  $p_{rec} = 0,2$ ,  $C = 2,005$ ,  $P_{ini} = 0,015$ ,  $N = 35$ ,  $V = 2^{30}$  (а); эффективность для разного числа  $M$ -процессов, иницирующая матрица *as-route*,  $V = 2^{30}$  (б)

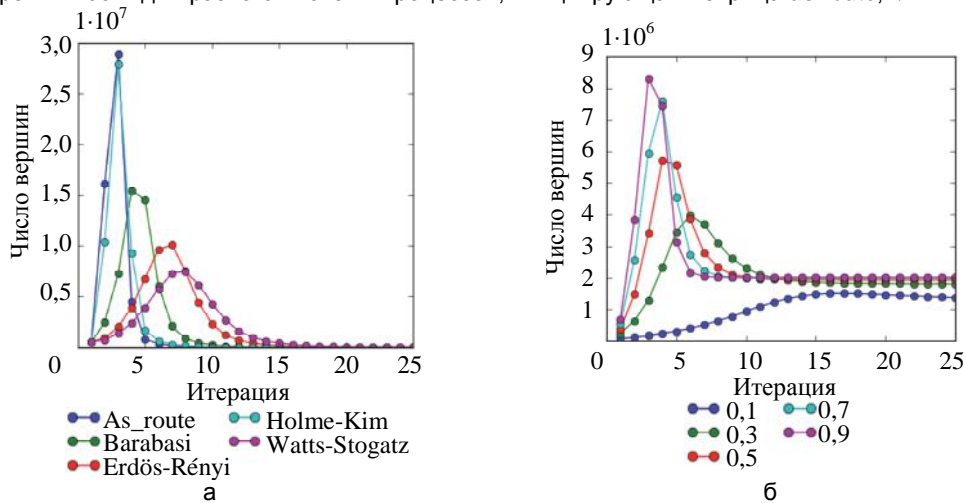


Рис. 7. Число ребер, добавляемых за итерацию, для различных типов иницирующих матриц,  $p_{act} = 0,5$ ,  $p_{rec} = 0,3$ ,  $C = 2,1$ ,  $P_{ini} = 0,015$ ,  $N = 35$ ,  $N_k = 24$  (а); число активированных вершин для различных  $p_{act}$ , иницирующая матрица Барабаши-Альберт с параметрами:  $N_k = 25$ ,  $p_{rec} = 0,1$ ,  $P_{ini} = 0,01$ ,  $N = 25$  (б)

На рис. 7, а, представлена динамика добавления ребер в сети различных типов (т.е. сети, сгенерированные по различным иницирующим матрицам). Наличие хабов (узлов с экстремально высокими степенями) для сетей со степенным (power-law) распределением степеней вершин (модели Холме–Кима, Барабаши–Альберт и граф *as-route*) приводит к более быстрому добавлению вершин и ребер в генерируемую сеть. Увеличенное значение коэффициента кластеризации (модель Холме–Кима отличается от модели Барабаши–Альберт возможностью его подстройки) также ведет к активизации распространения процесса. Рис. 7, а, б, подтверждают, что аппроксимация сэмплов сетей, сгенерированных по модели «В», графами Кронекера в рамках модели «Б» сохраняет между ними топологические различия, определяющие протекание процесса, в достаточной мере. Рис. 7, б, иллюстрирует динамику активации вершин для различных вероятностей  $P_{act}$  при прочих фиксированных параметрах. На данном рисунке можно наблюдать влияние начальных условий на значение числа вершин, при котором наблюдается стабилизация процесса, и на скорость достижения такой стабилизации.

### Вычислительная и информационная инфраструктура моделирования, обработки данных и визуализации комплексных сетей

Программная реализация схемы моделирования (рис. 3) является нетривиальной в силу необходимости объединения для решения задачи различных вычислительных моделей, которые могут быть представлены различными программными модулями, совместно работающими в распределенной среде. При этом сами сетевые структуры реального мира являются достаточно объемными; потому для хранения и обработки таких данных целесообразно использовать технологии, предполагающие перемещение вычислений к данным (Computations to Data, C2D), в отличие от классических суперкомпьютерных вычислений, основанных на перемещении данных к вычислениям (D2C). Как следствие, вычислительная инфраструктура для решения таких задач является гибридной, интегрирующей суперкомпьютерные технологии и технологии BigData.

**Организация вычислительной инфраструктуры.** Программный комплекс суперкомпьютерного моделирования критических состояний сложных социальных систем, реализующий схему на рис. 3, включает набор вычислительных модулей, которые должны взаимодействовать на основе общей интеграционной платформы в распределенной среде. В качестве основы комплекса была использована облачная платформа CLAVIRE второго поколения [38], что обеспечивает соответствие следующим требованиям:

- работа как с базовыми вычислительными приложениями (пакетами), так и со сложными (композиционными) приложениями, составленными на их основе;
- интеграция мультиплатформенной системы выполнения приложений с современными облачными и кластерными технологиями, а также возможность быстрого расширения доступного множества вычислительных ресурсов;
- поддержка средств интерактивного взаимодействия и визуализации;
- работа со сверхбольшими объемами данных с возможностью использования технологии BigData;
- поддержка ограничений на время выполнения (дедлайнов) и приоритетов для эффективного исполнения расчетных приложений на распределенных системах и суперкомпьютерах.

Общая архитектура программного комплекса представлена на рис. 8. В основе лежит ядро CLAVIRE WF Core, которое позволяет создавать и управлять отдельными проектами композитных приложений, основываясь на предметно-ориентированном языке EasyFlow, унифицировать и встраивать новые приложения (пакеты) с использованием компонента PackageBase, а также контролировать права доступа и авторизацию пользователей через компоненты UM, Facade и TokenBroker.

Подсистема META Execution базируется на компоненте MetaExecutor, позволяющем проводить интеграцию компонентов CLAVIRE с различными системами выполнения, помимо базовой системы CLAVIRE Execution, что принципиально для использования распределенных систем работы с большими данными. Так, возможна интеграция с системой Mesos Execution, которая имеет свой менеджер инфраструктуры. META Execution включает компонент MetaScheduler для глобального планирования выполняемых расчетов. Этот компонент принимает решение о том, какие части (пакеты) в составе схемы на рис. 3 отдавать на выполнение каждой из вычислительных систем, участвующих в процессе. Помимо этого, META Execution включает компонент MetaResourceBase с информацией обо всех ресурсах и их состояниях.

В состав расчетных пакетов, функционирующих под управлением META Execution, входят различные универсальные пакеты общего назначения и прикладные программы, обеспечивающие сервисные функции для задачи моделирования на рис. 3. К ним, в частности, относится пакет GVIS для визуализации динамики комплексных сетей со сложной структурой и многопараметрическими свойствами узлов. Он позволяет осуществлять динамическую визуализацию сетевых процессов с оперативным изменением графического представления свойств узлов и пространственной раскладки сети. Для сбора данных о социальных сетях в Интернет применяется пакет TwitterCrawler – приложение-краулер для интеграции ком-

плекса с внешними источниками данных. Краулер реализован на языке программирования Scala. В основе архитектуры краулера лежит акторная модель, с помощью которой реализуется распределенная обработка задач краулинга и первичная обработка собранных данных. Краулер реализует компонент scheduler для взаимодействия с системой ApacheMesos, позволяющий использовать ресурсы кластера, предоставляемые менеджером инфраструктуры Mesos. Это позволяет получить доступ к таким популярным ресурсам, как социальные сети VK, Twitter и LiveJournal. Получаемые таким образом данные могут храниться в системе Storm, ориентированной на распределенную обработку больших потоков данных в реальном масштабе времени.

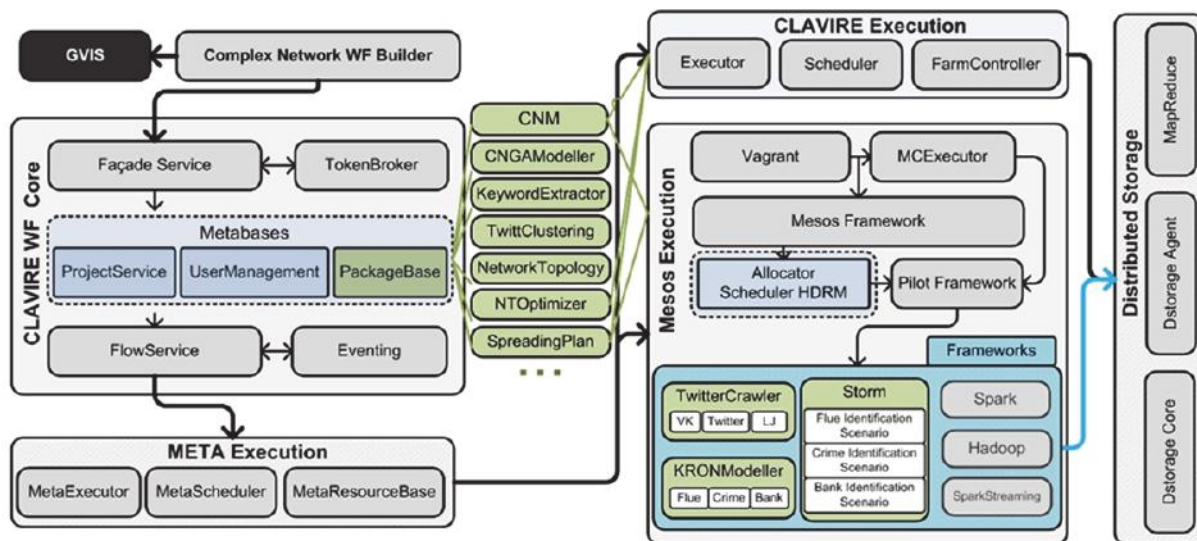


Рис. 8. Общая схема архитектуры комплекса моделирования критических явлений сложных социальных систем

Сам процесс моделирования реализуется на основе приложения KronModeller, которое позволяет осуществлять выращивание кронкерерской сети произвольного размера с одновременным моделированием изменения состояний узлов сети в ходе распространения процесса. KronModeller, в том числе, включает в себя процедуры генерации специфических сэмплов, например, для банковских и криминальных сетей. Решение прикладных задач осуществляется специальными прикладными пакетами, интегрируемыми в общее композитное приложение под управлением платформы CLAVIRE совместно с приложением KronModeller и другими универсальными пакетами.

Все указанные программы интегрируются в CLAVIRE с помощью механизмов компонента PackageManager и могут вызываться как по отдельности, так и в составе композитных приложений.

**Сбор данных для моделирования: социальные сети в Интернет.** Для получения характерных сэмплов для моделирования процессов в социальных сетях в Интернет используется специальная процедура сбора данных на основе обхода сети – краулинг. Краулинг может различаться по следующим характеристикам данных: прямые запросы на получение данных о конкретных объектах социальной сети – пользователь, пост; обход графа с целью построения точной топологии или сэмплирования; сбор потока событий, происходящих в социальной сети – создание постов, добавление пользователей, установление отношений. Технологически реализация каждого из указанных видов краулинга существенно различается; на практике часто требуется совмещение более чем одного вида краулинга [39–41]. Второй и третий типы получают развитие за счет ориентированных на них методов сбора. Помимо методов, ориентированных на полный обход (таких как поиск в глубину и поиск в ширину), для второго типа наибольшую популярность приобрели стохастические методы – Метрополиса–Гастингса и случайных блужданий [42, 43], нацеленные на сохранение основных топологических свойств сэмпла (что принципиально для задачи моделирования) при сокращенном количестве операций по получению данных. Также в рамках этого направления развиваются методы совмещенного краулинга, включая создание специальных языков, содержащих специальные операторы для задания желаемых стратегий сбора данных и правил передачи данных между ними.

Отдельное значение имеет унификация сбора данных из различных сетей [44, 45]. Существующие подходы основаны на введении общей модели, представляющей ту или иную сущность, встречающуюся в различных социальных сетях, а также набора правил отображения данных от конкретной социальной сети в такую модель. Можно выделить подходы, основанные также на построении предметно-ориентированного языка, который позволяет описать желаемые данные, обычно в декларативном стиле, а затем по полученному описанию построить набор запросов для получения данных. Работы [46, 47] предлагают в качестве такого языка SQL-подобный язык, а работа [48] предполагает использование языка



общего назначения. Следует также отметить, что развитие подходов к унификации краулинга социальных сетей и повышение потребности в рассмотрении сетей не как отдельных изолированных объектов, а как взаимодействующих, связанных между собой особыми типами связи, привело к появлению подходов, нацеленных на создание моделей глобальных социальных сетей, предполагающих переходы между социальными сетями отдельных провайдеров. В работе [48] предложен язык описания запросов, который позволяет описывать переходы между сетями, что достигается за счет унифицированной модели социальной сети и применяемых методов идентификации межсетевых связей – поиска особым образом сформированных URL-адресов. Также существуют подходы, опирающиеся на онтологические модели сущностей социальных сетей, что позволяет определять тип найденных объектов и осуществлять более эффективный сбор по определенной пользователем стратегии за счет фильтрации.

С точки зрения обеспечения вычислительной производительности важное значение имеет организация распределенного и многоканального краулинга. Под многоканальным краулингом понимается сбор данных с помощью множества сетевых (например, прокси-сервер) или информационных (например, аккаунт, использующийся для сбора) ресурсов. Наиболее распространен агентный подход к организации краулинга. Каждая из программ-агентов обладает одним или несколькими ресурсами, которые он использует для сбора данных и которые независимы от ресурсов других агентов. Все агенты разделяют одну очередь, из которой каждому из агентов выдаются задачи на сбор данных, состоящие из одного или нескольких запросов к социальным сетям. Агенты получают задачи по мере их собственной скорости обработки. Описанный выше подход к организации краулинга позволяет значительно повысить скорость сбора информации. На рис. 9 приведены экспериментальные исследования по многоканальному сбору данных, осуществляемому для множества пользователей с индивидуальными запросами, реализованные в рамках программного комплекса на рис. 8. Видно, что за счет применения агентного подхода время сбора данных можно существенно сократить – с 1429 секунд до 91 секунды (рис. 9, а), или пропорционально можно увеличить объем собираемых данных. Однако при росте числа агентов на одном ВУ наблюдается снижение общей эффективности, что говорит о необходимости применения распределенного сбора данных. Применяемый подход позволяет интегрироваться с платформой CLAVIRE для гибкого предоставления ресурсов в зависимости от нагрузки, а также расширять систему краулинга путем добавления новых агентов краулера с новой логикой сбора.

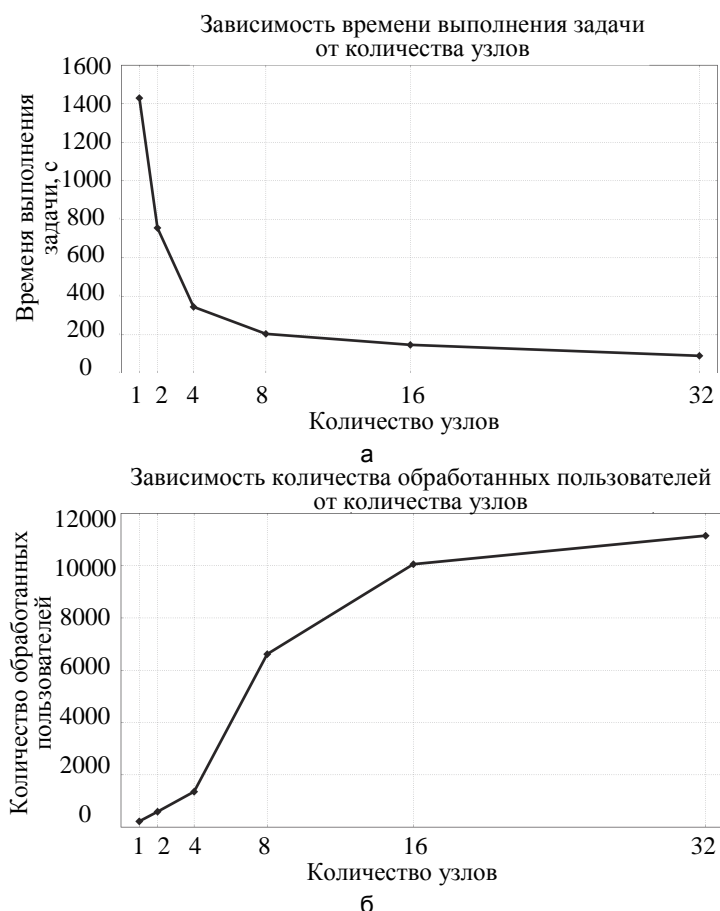


Рис. 9. Оценка производительности сбора данных в зависимости от числа агентов краулера на примере социальной сети Twitter: время (в секундах) сбора информации о 4000 пользователей (а); количество собранных пользователей за фиксированное время – 7 мин (б)

**Визуализация процессов на комплексных сетях.** При работе с результатами моделирования процессов на больших КС важно не только их корректно визуально отобразить, но и по возможности избежать перенасыщения изображения излишней информацией [49]. Значительную роль здесь играет масштабируемость графа КС [50], когда на разных уровнях детализации отображается разное число узлов с сохранением общей структуры. Визуально масштабируемость графа может быть реализована двумя способами: с помощью изменения пространственного масштаба, т.е. с помощью управления камерой или линзой [51], либо с помощью масштабируемости непосредственно графа – на различных масштабах структура графа может быть отображена с помощью объединения узлов, принадлежащих одному кластеру, в общий мета-узел для улучшения восприятия.

Динамические процессы в комплексных сетях могут быть отображены через появление и исчезновение узлов со временем [52], а также с помощью изменения взаимодействия между вершинами и управлением их визуальными характеристиками, такими как форма или размер узла [53]. Кроме того, для изучения процессов в динамических графах может использоваться цветовое кодирование, например, цветовая палитра для индикации активности узлов [54]. Для того чтобы отделить актуальные данные от устаревших, интенсивность цвета может варьироваться в зависимости от прошедшего времени: ранние события отображаются менее яркими, чем события, протекающие в данный момент [55].

Выделение наиболее важных узлов и связей, равно как кластеров и сообществ в сети позволяет не только многое сказать о структуре сети на этапе предварительного анализа, но и играет ключевую роль в исследовании различных процессов. Для визуального выделения интересующих элементов в сети используются два способа:

- управление визуальными характеристиками сети – изменение размера, цвета, формы узлов при изменении атрибутов узла или достижения какого-либо значения параметра, например, при увеличении степени узла;
- управление физикой узлов, их размещением в пространстве, т.е. влияние на взаимодействие узлов между собой.

Для больших и сложно структурированных КС изменение только цвета или размера не позволяет провести четкую идентификацию интересующего узла. В таких случаях целесообразно менять позицию узлов в зависимости от исследуемых параметров, тогда даже при беглом взгляде на сеть соответствующие узлы будут визуальным образом отделены от остальных.

На рис. 10 представлен пример динамического процесса на сети – распространение информации (слухов) в КС мобильной связи. Три группы узлов, которые демонстрируют различное поведение, расположены на окружностях разного радиуса. При инициализации связи (звонок по мобильному телефону) появляется ребро красного цвета, если во время разговора осуществляется передача информации, и белого – в обратном случае. По окончании звонка яркость ребра уменьшается.

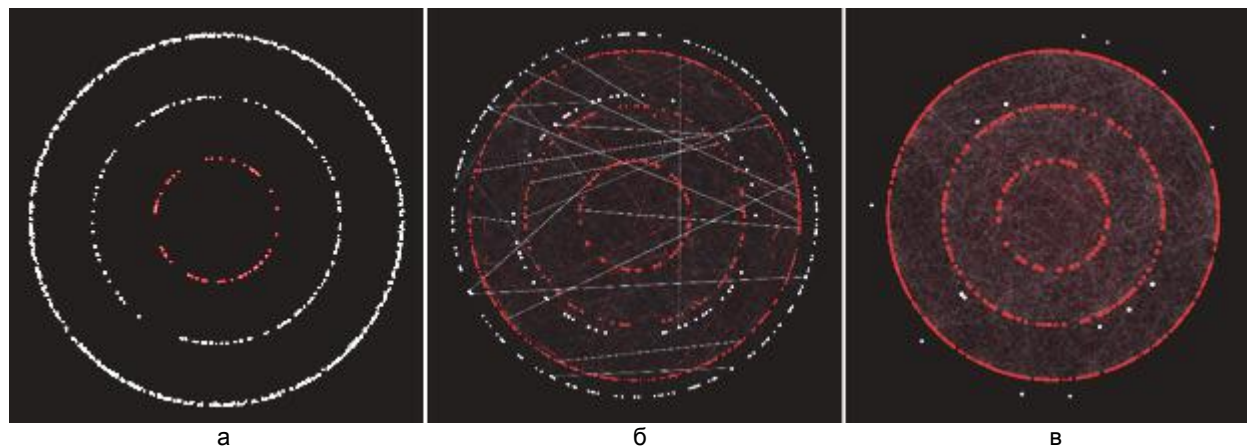


Рис. 10. Пример визуализации динамического процесса распространения слухов через сети мобильной связи: распределение информированности в начальный момент времени (а); спустя 20 часов модельного времени (б); после окончания моделирования (в)

Для наглядного представления графа размером порядка 100 тыс. узлов и выше реализован метод на основе динамической силовой модели, позволяющей позиционировать узлы графа в 2D- или 3D-пространстве с помощью связывания виртуальных сил с вершинами и ребрами графа и последующего воспроизведения движения вершин и ребер графа. В качестве основного алгоритма для раскладки графов была выбрана силовая модель с адаптивным шагом вычислений [56]. Это позволяет ассоциировать итерации алгоритма раскладки с шагами изменения динамического процесса, что обеспечивает снижение ресурсоемкости алгоритма до квадратичной. Положение узлов вычисляется исходя из их взаимодействия. Взаимодействие между узлами определяется с помощью следующих сил: сила отталкивания, в качестве

которой выступает сила, аналогичная кулоновскому взаимодействию, и сила притяжения, которая выражается через закон Гука, параметры сил зависят от свойств самих узлов, изменяющихся во времени. В случае если визуализируемая КС содержит элементы разных типов/классов или предполагается, что физическое положение узлов несет дополнительную информацию, данная сила используется для того, чтобы притянуть узлы к конкретной точке, оси или области. КС, представленные на рис. 11, визуализированы с помощью описанного выше метода.

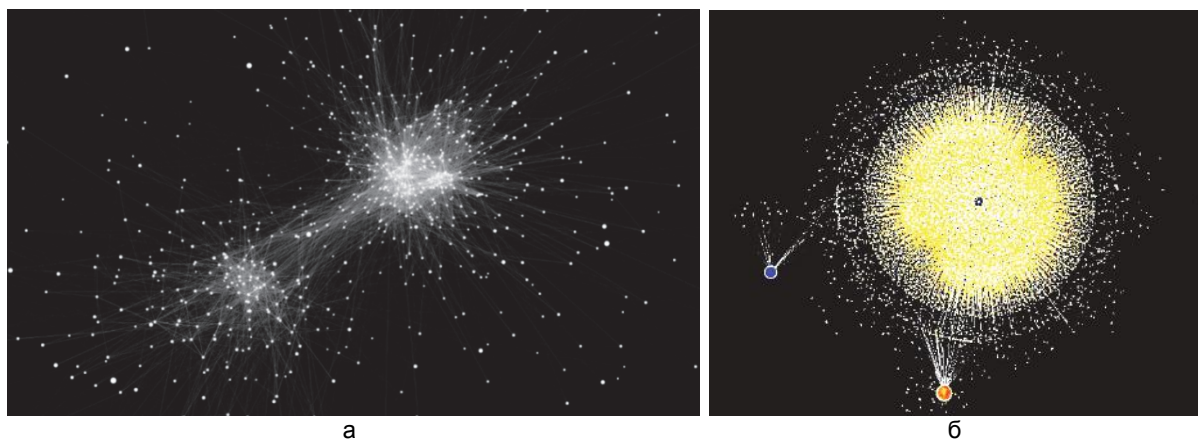


Рис. 11. Примеры работы алгоритма визуализации комплексных сетей: граф профилей лечения пациентов крупного медицинского центра (а); сеть репостов из социальной сети ВКонтакте (б)

Сам параллельный алгоритм предназначен для реализации на GPU и должен учитывать особенности платформы. Для эффективной работы алгоритма для сил, действующих на узлы, определен их радиус действия, что существенно сокращает объем вычислений. Экспериментальные исследования масштабируемости показали, что для 100 тыс. узлов алгоритм обеспечивает 25 FPS, что соответствует плавному изменению изображения на экране.

#### Приложения технологий суперкомпьютерного моделирования критических явлений в комплексных сетях

Математическое, алгоритмическое и программное обеспечение моделирования критических явлений в сложных социальных системах на основе КС позволяет решать различные прикладные задачи, связанные с предсказательным моделированием как самих процессов, приводящих к критическим ситуациям, так и методов управления.

**Моделирование развития эпидемии.** Классическим примером КС, отражающей особенности развития критических явлений в социальных системах, является эпидемиологическая сеть распространения ОРВИ и гриппа для большого города (порядка 10 млн жителей).

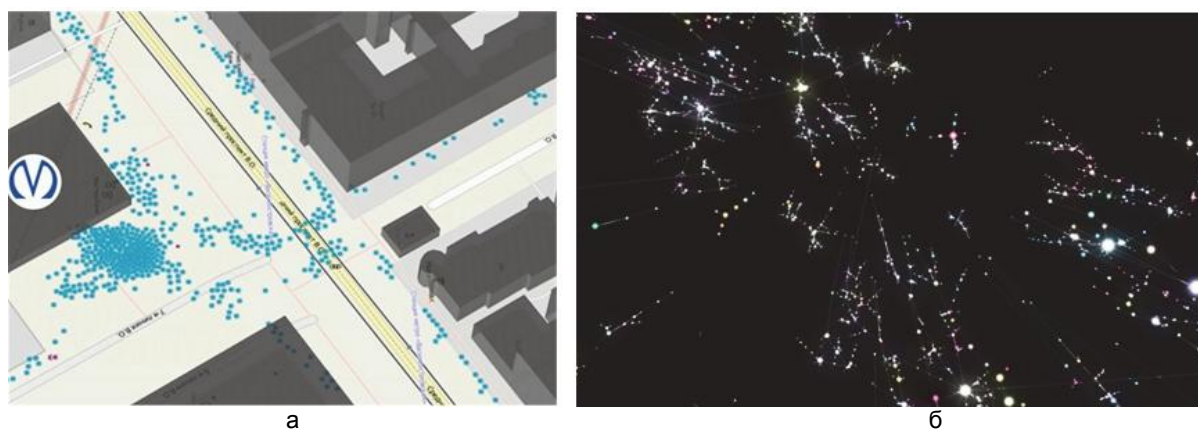


Рис. 12. Построение контактной эпидемиологической сети на основе МА-моделирования мобильности населения на микроуровне: фрагмент сцены движения агентов в городской среде (а); моделирование процессов распространения инфекции в контактной сети (б)

Для построения сети использованы данные о мобильности населения, получаемые посредством МА-моделирования [57], а также макрохарактеристики заболеваемости простудными заболеваниями для Санкт-Петербурга, полученные на массиве данных по заболеваемости за 30 лет (см. также рис. 1). МА-модель имеет одноуровневую систему навигации агентов, обеспечивающую планирование мобильности групп агентов между точками интереса в зависимости от социально-экономических характеристик

населения моделируемой территории. С целью перемещения агентов для каждого из них задается путь на заранее заданном графе мобильности. Перемещения в модели происходят между отдельными зданиями – объектами жилой, производственной, общественной инфраструктуры. При этом в модели учитывается не только перемещение агента по территории, но и их нахождение в зданиях совместно с другими агентами. Для ретроспективного выявления критической ситуации был разработан алгоритм выделения эпидемической и межэпидемической фаз динамики болезней с последующим обнаружением моментов наступления критических состояний (начало и конец эпидемической вспышки). После обнаружения факта наступления эпидемиологической ситуации посредством проведения вычислительного эксперимента с МА-моделью мобильности населения (рис. 12, а) строится контактная эпидемиологическая сеть, на основе которой моделируются процессы распространения инфекции (рис. 12, б).

**Моделирование и оптимизация путей распространения информации в социальных сетях Интернет.** Задача мониторинга и оптимизации распространения информации (в том числе слухов) в сети Интернет основывается на совместном использовании методов сбора и анализа данных, и методов моделирования процессов в КС, необходимых для создания эталонных решений. Наибольший интерес представляют ситуации, связанные с резким возрастанием скорости распространения – количества пользователей, участвующих в процессе.

Модель распространения информации в сети Интернет параметрически учитывает характеристики активных пользователей: вероятность распространения твитов, частоту генерации твитов и стоимость использования самого источника. Вероятность, частота и время жизни сообщения определяют успешность и интенсивность распространения. Данная модель позволяет оценить возможности развития ситуации с учетом различных принципов распространения, а также построить оптимальные пути распространения информации для различных категорий сообщений и групп пользователей.

Так, на рис. 13 представлены результаты моделирования распространения информации в сети Твиттер с применением двух алгоритмов, оптимизирующих использование источников информации. В первом случае рассматривается «жадный» алгоритм, ориентированный на распространение информации через лидеров мнений – узлы с наибольшим количеством подписчиков. Во втором случае моделирование процессов на КС используется для определения целевой функции оптимизационной задачи, которая направлена на выявление оптимальной структуры связей и расположения узлов, приводящей к наискорейшему оповещению максимального числа участников сети.

Из рис. 13 видно, что решения, найденные с помощью формального решения задачи оптимизации на основе моделирования КС, могут быть до 34% эффективнее, чем более наглядные «жадные» подходы. Стоимость решения, состоящего из набора вершин сети, представлена в условных единицах (у.е.). Стоимость отдельной вершины определяется как количество имеющихся у нее подписчиков (т.е. входящая степень вершины) и уровень ее активности (количество сообщений, производимых вершиной в единицу времени).

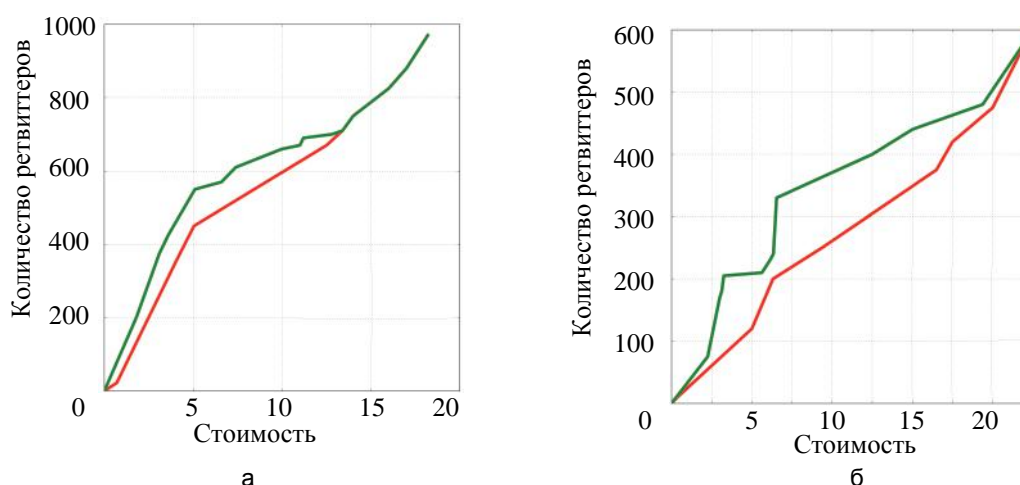


Рис. 13. Оценка эффективности способов оптимизации процесса распространения информации в сети Твиттер: с использованием решений, найденных GA (genetic algorithm – генетический алгоритм, зеленая линия) и решений с наибольшим числом подписчиков HD (highest degree – наибольшая степень вершины, красная линия): актуальность твита 24 часа (а); актуальность твита 12 часов (б)

**Моделирование эволюции финансовых сетей.** Несмотря на ощутимую связь между финансовыми и банковскими кризисами [58], существующие макроэкономические модели [59], по сути, относятся к категории «А» и не учитывают топологическую структуру рынка межбанковских кредитов. Однако в кризисной ситуации топология межбанковских взаимодействий может сыграть драматическую роль при де-



фолтах и распространении финансовых шоков. Тем не менее, учет всех значимых параметров в одной модели представляется нецелесообразным не только с точки зрения вычислительной сложности, но также в силу многогранности и изменяемости риска: в случае более детального представления модель, точно аппроксимирующая эволюционирующую систему сегодня, может потерять актуальность завтра. Таким образом, для моделирования финансового риска, обусловленного хрупкостью банковской системы, допустимо использовать систему вложенных моделей «А»–«В», которая учитывает непосредственно сетевую структуру отношений между отдельными заемщиками, отдельными банками, а также банками и заемщиками. Такая модель обеспечивает не только ситуационную осведомленность о последствиях развития кризиса, но позволяет отработать стратегии действий регулирующего органа (центробанка) для обеспечения устойчивости финансовой системы в целом (рис. 14).

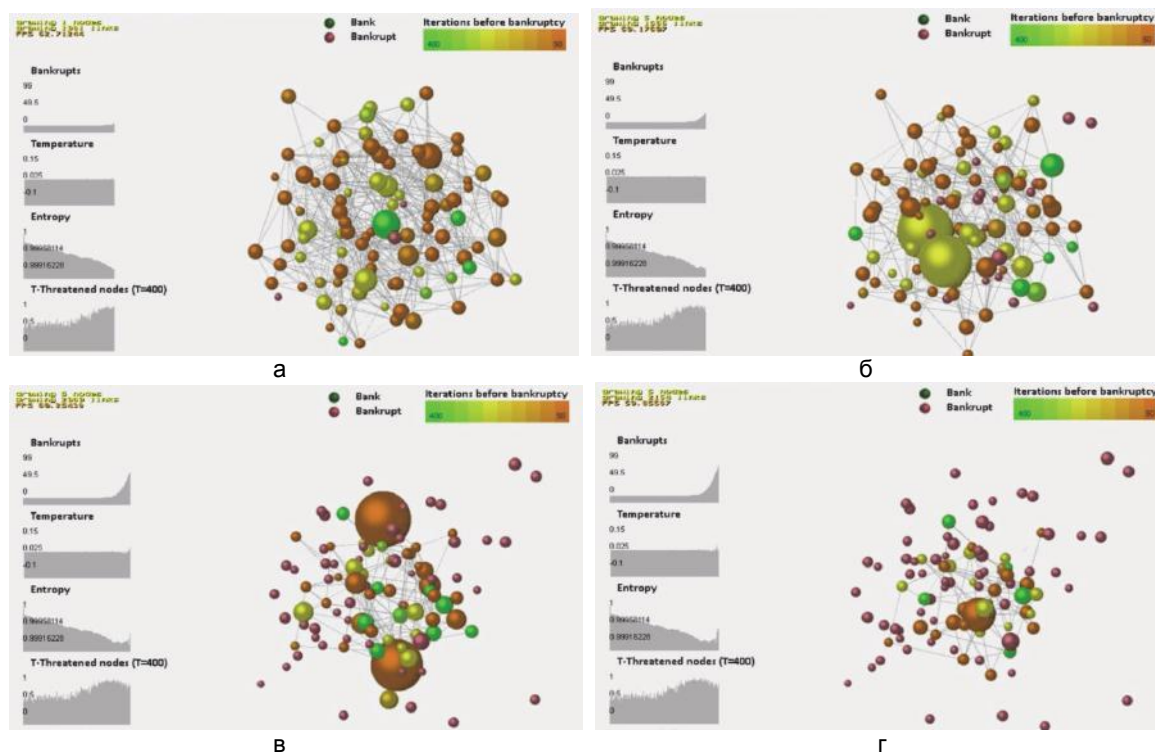


Рис. 14. Модельная динамика изменения банковской сети до полного разрушения в результате каскада контагиозных дефолтов (а–г). Размер точек соответствует значениям активов банка, цвет соответствует предполагаемому количеству итераций до наступления дефолта банка. Графики на рисунках (сверху вниз) отражают динамику изменения количества дефолтов банков в системе, температуру, энтропию, а также долю узлов с негативной динамикой

**Моделирование разрушения криминальных сетей.** Криминальная сеть представляет собой структуру, связывающую между собой представителей криминального рынка оборота наркотиков с присутствующими им функциональными ролями и индивидуальными свойствами, такими как пол, возраст, национальность, страна рождения. В [60] была отмечена аналогия между криминальными сетями и традиционными социальными сетями, широко встречающимися в реальном мире. Однако в силу своей скрытой («dark») природы криминальные сети требуют особого подхода, связанного с необходимостью учета скрытых связей и высокой устойчивостью к разрушению из-за децентрализованной структуры и процессов самовосстановления потерянных связей после нарушения целостности структуры сети в результате внешнего воздействия. В более поздних исследованиях [61–63] функционирование криминальных сетей рассматривается в контексте сетевого анализа в симбиозе с экономической составляющей криминального рынка, что позволяет детализировать структуру преступных связей, выявить скрытые процессы и установить взаимосвязь между внешними воздействиями и устойчивостью сети. В рамках такого подхода было предложено рассматривать структуру криминальной сети как бизнес-процесса в рамках концепции «Value Chain» (цепочка поставок). В основе предложенного подхода лежит представление о функциональной целостности криминальной сети, когда каждый элемент принадлежит хотя бы одной цепочке. Это естественным образом формирует основу для алгоритма восстановления скрытых связей. Под восстановлением понимается процесс, в котором каждый узел сети, не имеющий необходимых связей в рамках хотя бы одной цепочки, будет присоединен как минимум к одной уже существующей цепочке. В этом случае криминальная сеть может быть рассмотрена в рамках совокупности цепочек поставок, а количество цепочек может выступать в качестве метрики эффективности функционирования сети в целом. Однако процесс восстановления по своей природе имеет вероятностный характер, что приводит к возникно-

вению ансамбля различных конфигураций криминальной сети, каждой из которых соответствует индивидуальный набор цепочек, при сохранении макроскопических характеристик сети в целом (рис. 15).

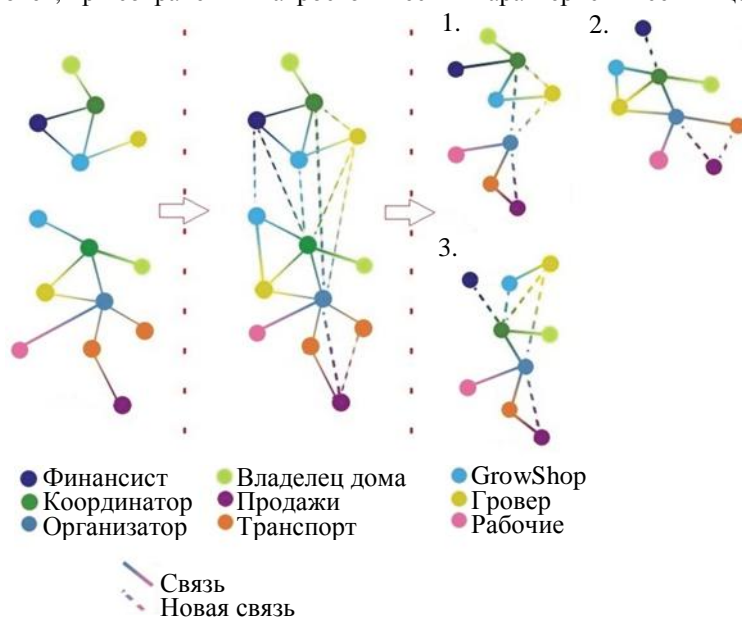


Рис. 15. Восстановление фрагмента криминальной сети с последующим выделением трех комбинаций цепочек

Основной задачей при работе с криминальными сетями является исследование и оценка их устойчивости в результате удаления части узлов из сети, что соответствует арестам членов криминального сообщества в реальном мире. Для воспроизведения реальных свойств криминальных сетей, связанных с восстановлением утраченных связей после внешнего воздействия, каждый эксперимент с разрушением восстановленной сети сопровождается ее повторным восстановлением с использованием алгоритма восстановления связей, который применялся к исходной сети.

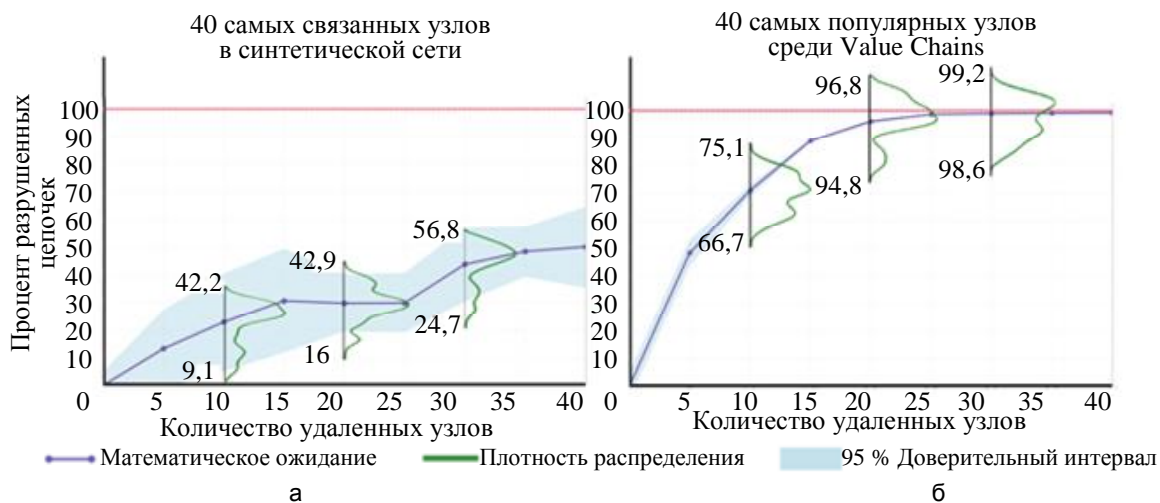


Рис. 16. Результаты работы двух стратегий разрушения: удаление вершин с наибольшей степенью (а); удаление вершин с наиболее частым появлением в различных цепочках (б)

В результате из ансамбля первично восстановленных сетей мы получаем новый ансамбль повторно восстановленных сетей для каждой из стратегий. Полученный ансамбль позволяет провести анализ неопределенности результатов применения различных стратегий разрушения сети, который помогает в сравнении и оценке эффективности различных стратегий (рис. 16).

### Заключение

В работе рассмотрены основные аспекты суперкомпьютерного моделирования критических явлений в сложных социальных системах. Поскольку критические явления обычно связаны с потерей устойчивости системы в целом, то они могут вызываться незначительными различиями в поведении индивидов и протекать достаточно быстро. Это ограничивает возможности применения для их изучения классического аппарата динамических систем, которые (в силу перехода к интегральным характеристикам

путем введения усреднения) основаны на том или ином предположении об однородности и (или) эргодичности. Как следствие, для решения данной задачи целесообразно использовать формализм комплексных сетей с учетом их многомасштабной изменчивости и неоднородности, что реализовано в форме семейства вложенных моделей «А»–«В», предназначенных для последовательного уточнения характеристик комплексных сетей в зоне наступления критической ситуации. Такой подход к моделированию обладает существенным потенциалом параллельной масштабируемости на уровне модели «Б» динамики комплексных сетей на индивидуальном уровне, что подтверждено экспериментально. Разработана информационная и вычислительная инфраструктура моделирования, обеспечивающая выполнение расчетов в распределенной среде, сбор исходных данных в социальных сетях в Интернет, а также визуализацию результатов. Предложенные методы, модели, алгоритмы и программное обеспечение могут быть применены для исследования и оптимизации различных видов процессов в социальных системах на комплексных сетях, включая распространение эпидемий, разрушение криминальных сетей, распространение информации в социальных медиа Интернет, а также управления финансовыми (банковскими) сетями.

### Литература

1. Castellano C., Fortunato S., Loreto V. Statistical physics of social dynamics // *Reviews of Modern Physics*. 2009. V. 81. N 2. P. 591–646. doi: 10.1103/RevModPhys.81.591
2. Jiang L.L., Hua D.Y., Chen T. Nonequilibrium phase transitions in a model with social influence of inflexible units // *Journal of Physics A: Mathematical and Theoretical*. 2007. V. 40. N 37. P. 11271–11276. doi: 10.1088/1751-8113/40/37/005
3. Bordogna C.M., Albano E.V. Statistical methods applied to the study of opinion formation models: a brief overview and results of a numerical study of a model based on the social impact theory // *Journal of Physics Condensed Matter*. 2007. V. 19. N 6. Art. 065144. doi: 10.1088/0953-8984/19/6/065144
4. Fortunato S., Castellano C. Scaling and universality in proportional elections // *Physical Review Letters*. 2007. V. 99. N 13. Art. 138701. doi: 10.1103/PhysRevLett.99.138701
5. Klimek P., Lambiotte R., Thurner S. Opinion formation in laggard societies // *EPL*. 2008. V. 82. N 2. Art. 28008. doi: 10.1209/0295-5075/82/28008
6. Chmiel A., Sienkiewicz J., Thelwall M., Paltoglou G., Buckley K., Kappas A., Holyst J.A. Collective emotions online and their influence on community life // *PloS One*. 2011. V. 6. N 7. Art. e22207 doi: 10.1371/journal.pone.0022207
7. Chmiel A., Holyst J.A. Transition due to preferential cluster growth of collective emotions in online communities // *Physical Review E*. 2013. V. 87. N 2. Art. 022808. doi: 10.1103/PhysRevE.87.022808
8. Gil S., Zanette D.H. Coevolution of agents and networks: opinion spreading and community disconnection // *Physics Letters A*. 2006. V. 356. N 2. P. 89–94. doi: 10.1016/j.physleta.2006.03.037
9. Kacperski K., Holyst J.A. Phase transitions and hysteresis in a cellular automata-based model of opinion formation // *Journal of Statistical Physics*. 1996. V. 84. N 1-2. P. 169–189.
10. Newman M. *Networks: An Introduction*. Oxford University Press, 2010. 784 p.
11. Snijders T.A., Van de Bunt G.G., Steglich C.E.G. Introduction to stochastic actor-based models for network dynamics // *Social Networks*. 2010. V. 32. P. 44–60. doi: 10.1016/j.socnet.2009.02.004
12. Kumar R., Novak J., Tomkins A. Structure and evolution of online social networks / In: *Link Mining: Models, Algorithms, and Applications*. NY: Springer, 2010. P. 337–357. doi: 10.1007/978-1-4419-6515-8-13
13. Leskovec J., Huttenlocher D., Kleinberg J. Predicting positive and negative links in online social networks // *Proc. 19<sup>th</sup> Int. Conf. on World Wide Web*. Raleigh, USA, 2010. P. 641–650. doi: 10.1145/1772690.1772756
14. Leskovec J., Chakrabarti D., Kleinberg J., Faloutsos C., Ghahramani Z. Kronecker graphs: an approach to modeling networks // *Journal of Machine Learning Research*. 2010. V. 11. P. 985–1042.
15. Lamberson P.J., Page S.E. Tipping points // *Quarterly Journal of Political Science*. 2012. V. 7. N 2. P. 175–208. doi: 10.1561/100.00011061
16. O'Regan S.M., Drake J.M. Theory of early warning signals of disease emergence and leading indicators of elimination //

### References

1. Castellano C., Fortunato S., Loreto V. Statistical physics of social dynamics. *Reviews of Modern Physics*, 2009, vol. 81, no. 2, pp. 591–646. doi: 10.1103/RevModPhys.81.591
2. Jiang L.L., Hua D.Y., Chen T. Nonequilibrium phase transitions in a model with social influence of inflexible units. *Journal of Physics A: Mathematical and Theoretical*, 2007, vol. 40, no. 37, pp. 11271–11276. doi: 10.1088/1751-8113/40/37/005
3. Bordogna C.M., Albano E.V. Statistical methods applied to the study of opinion formation models: a brief overview and results of a numerical study of a model based on the social impact theory. *Journal of Physics Condensed Matter*, 2007, vol. 19, no. 6, art. 065144. doi: 10.1088/0953-8984/19/6/065144
4. Fortunato S., Castellano C. Scaling and universality in proportional elections. *Physical Review Letters*, 2007, vol. 99, no. 13, art. 138701. doi: 10.1103/PhysRevLett.99.138701
5. Klimek P., Lambiotte R., Thurner S. Opinion formation in laggard societies. *EPL*, 2008, vol. 82, no. 2, art. 28008. doi: 10.1209/0295-5075/82/28008
6. Chmiel A., Sienkiewicz J., Thelwall M., Paltoglou G., Buckley K., Kappas A., Holyst J.A. Collective emotions online and their influence on community life. *PloS One*, 2011, vol. 6, no. 7, art. e22207 doi: 10.1371/journal.pone.0022207
7. Chmiel A., Holyst J.A. Transition due to preferential cluster growth of collective emotions in online communities. *Physical Review E*, 2013, vol. 87, no. 2, art. 022808. doi: 10.1103/PhysRevE.87.022808
8. Gil S., Zanette D.H. Coevolution of agents and networks: opinion spreading and community disconnection. *Physics Letters A*, 2006, vol. 356, no. 2, pp. 89–94. doi: 10.1016/j.physleta.2006.03.037
9. Kacperski K., Holyst J.A. Phase transitions and hysteresis in a cellular automata-based model of opinion formation. *Journal of Statistical Physics*, 1996, vol. 84, no. 1-2, pp. 169–189.
10. Newman M. *Networks: An Introduction*. Oxford University Press, 2010, 784 p.
11. Snijders T.A., Van de Bunt G.G., Steglich C.E.G. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 2010, vol. 32, pp. 44–60. doi: 10.1016/j.socnet.2009.02.004
12. Kumar R., Novak J., Tomkins A. Structure and evolution of online social networks. In *Link Mining: Models, Algorithms, and Applications*. NY, Springer, 2010, pp. 337–357. doi: 10.1007/978-1-4419-6515-8-13
13. Leskovec J., Huttenlocher D., Kleinberg J. Predicting positive and negative links in online social networks. *Proc. 19<sup>th</sup> Int. Conf. on World Wide Web*. Raleigh, USA, 2010, pp. 641–650. doi: 10.1145/1772690.1772756
14. Leskovec J., Chakrabarti D., Kleinberg J., Faloutsos C., Ghahramani Z. Kronecker graphs: an approach to modeling networks. *Journal of Machine Learning Research*, 2010, vol. 11, pp. 985–104.
15. Lamberson P.J., Page S.E. Tipping points. *Quarterly Journal of Political Science*, 2012, vol. 7, no. 2, pp. 175–208. doi: 10.1561/100.00011061
16. O'Regan S.M., Drake J.M. Theory of early warning signals of disease emergence and leading indicators of elimination.



- Theoretical Ecology, 2013. V. 6. N 3. P. 333–357. doi: 10.1007/s12080-013-0185-5
17. Dorogovtsev S.N., Goltsev A.V., Mendes J.F.F. Critical phenomena in complex networks // *Reviews of Modern Physics*. 2007. V. 80. N 4. P. 1275–1335. doi: 10.1103/RevModPhys.80.1275
  18. Seoane L.F., Sole R. Phase transitions in Pareto optimal complex networks // *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. 2015. V. 92. Art. 032807. doi: 10.1103/PhysRevE.92.032807
  19. Ye C., Torsello A., Wilson R.C., Hancock E.R. Thermodynamics of time evolving networks // *Lecture Notes in Computer Science*. 2015. V. 9069. P. 315–324. doi: 10.1007/978-3-319-18224-7\_31
  20. Ni K.-Y., Lu T.-C. Information dynamic spectrum characterizes system instability toward critical transitions // *EPJ Data Science*. 2014. V. 3. N 1. P. 1–25. doi: 10.1140/epjds/s13688-014-0028-7
  21. Guleva V.Y. The combination of topology and nodes' states dynamics as an early-warning signal of critical transition in a banking network model // *Procedia Computer Science*. 2016. V. 80. P. 1755–1764. doi: 10.1016/j.procs.2016.05.436
  22. Luque B., Ballesteros F.J., Muro E.M. Self-organized critical random Boolean networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2001. V. 63. N 5. Art. 51913. doi: 10.1103/PhysRevE.63.051913
  23. Han H., Ma A., Huang Z. An improved SIRS epidemic model on complex network // *Proc. Int. Conf. on Computational Intelligence and Software Engineering*. Wuhan, China, 2009. Art. 5365422. doi: 10.1109/CISE.2009.5365422
  24. Delpini D., Battiston S., Riccaboni M., Gabbi G., Pammolli F., Caldarelli G. Evolution of controllability in interbank networks // *Scientific Reports*. 2013. V. 3. Art. 1626. doi: 10.1038/srep01626
  25. Bisset K.R., Chen J., Feng X., Kumar V.S.A., Marathe M.V. EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems // *Proc. 23<sup>rd</sup> Int. Conf. on Supercomputing*. Yorktown, USA, 2009. P. 430–439. doi: 10.1145/1542275.1542336
  26. Hou B., Yao Y., Wang B., Liao D. Modeling and simulation of large-scale social networks using parallel discrete event simulation // *Simulation*. 2013. V. 89. N 10. P. 1173–1183. doi: 10.1177/0037549713495752
  27. Gonzalez J.E., Low Y., Gu H., Bickson D., Guestrin C. PowerGraph: distributed graph-parallel computation on natural graphs // *Proc. 10<sup>th</sup> USENIX Conference on Operating Systems Design and Implementation*. Hollywood, USA, 2012. P. 17–30.
  28. Liu Y.-Y., Slotine J.-J., Barabasi A.-L. Observability of complex systems // *Proc. National Academy of Science of the USA*. 2013. V. 110. N 7. P. 2460–2465. doi: 10.1073/pnas.1215508110
  29. Yoo A., Henderson K. Parallel generation of massive scale-free graphs // *Proceedings of CoRR*. 2010.
  30. Ryu E., Choi S. Stochastic Kronecker Graph on Vertex-Centric BSP // *Proceedings of CoRR*. 2012. Art. arXiv1208.2239.
  31. Paul G., Tanizawa T., Havlin S., Stanley H.E. Optimization of robustness of complex networks // *The European Physical Journal B - Condensed Matter and Complex Systems*. 2004. V. 38. N 2. P. 187–191. doi: 10.1140/epjb/e2004-00112-3
  32. Watts D.J., Strogatz S.H. Collective dynamics of 'small-world' networks // *Nature*. 1998. V. 393. N 6684. P. 440–442. doi: 10.1038/30918
  33. Albert R., Barabasi A.L. Statistical mechanics of complex networks // *Reviews of Modern Physics*. 2002. V. 74. N 1. P. 47–97. doi: 10.1103/RevModPhys.74.47
  34. Newman M.E.J. The structure and function of complex networks // *SIAM Review*. 2003. V. 45. N 2. P. 167–256.
  35. Kashirin V.V. Evolutionary simulation of complex networks structures with specific topological properties // *Procedia Computer Science*. 2014. V. 29. P. 2401–2411. doi: 10.1016/j.procs.2014.05.224
  36. Voloshin D.V., Puzyreva K.A., Karbovskii V.A. Agent-based virtual society polygon for simulation and evaluation in massive mobile services // *IERI Procedia*. 2014. V. 10. P. 231–238. doi: 10.1016/j.ieri.2014.09.082
  37. Mahdian M., Xu Y. Stochastic kronecker graphs // *Lecture Notes in Computer Science*, 2007, vol. 4863, pp. 179–186.
  17. Dorogovtsev S.N., Goltsev A.V., Mendes J.F.F. Critical phenomena in complex networks. *Reviews of Modern Physics*, 2007, vol. 80, no. 4, pp. 1275–1335. doi: 10.1103/RevModPhys.80.1275
  18. Seoane L.F., Sole R. Phase transitions in Pareto optimal complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2015, vol. 92, art. 032807. doi: 10.1103/PhysRevE.92.032807
  19. Ye C., Torsello A., Wilson R.C., Hancock E.R. Thermodynamics of time evolving networks. *Lecture Notes in Computer Science*, 2015, vol. 9069, pp. 315–324. doi: 10.1007/978-3-319-18224-7\_31
  20. Ni K.-Y., Lu T.-C. Information dynamic spectrum characterizes system instability toward critical transitions. *EPJ Data Science*, 2014, vol. 3, no. 1, pp. 1–25. doi: 10.1140/epjds/s13688-014-0028-7
  21. Guleva V.Y. The combination of topology and nodes' states dynamics as an early-warning signal of critical transition in a banking network model. *Procedia Computer Science*, 2016, vol. 80, pp. 1755–1764. doi: 10.1016/j.procs.2016.05.436
  22. Luque B., Ballesteros F.J., Muro E.M. Self-organized critical random Boolean networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2001, vol. 63, no. 5, art. 51913. doi: 10.1103/PhysRevE.63.051913
  23. Han H., Ma A., Huang Z. An improved SIRS epidemic model on complex network. *Proc. Int. Conf. on Computational Intelligence and Software Engineering*. Wuhan, China, 2009, art. 5365422. doi: 10.1109/CISE.2009.5365422
  24. Delpini D., Battiston S., Riccaboni M., Gabbi G., Pammolli F., Caldarelli G. Evolution of controllability in interbank networks. *Scientific Reports*, 2013, vol. 3, art. 1626. doi: 10.1038/srep01626
  25. Bisset K.R., Chen J., Feng X., Kumar V.S.A., Marathe M.V. EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. *Proc. 23<sup>rd</sup> Int. Conf. on Supercomputing*. Yorktown, USA, 2009, pp. 430–439. doi: 10.1145/1542275.1542336
  26. Hou B., Yao Y., Wang B., Liao D. Modeling and simulation of large-scale social networks using parallel discrete event simulation. *Simulation*, 2013, vol. 89, no. 10, pp. 1173–1183. doi: 10.1177/0037549713495752
  27. Gonzalez J.E., Low Y., Gu H., Bickson D., Guestrin C. PowerGraph: distributed graph-parallel computation on natural graphs. *Proc. 10<sup>th</sup> USENIX Conference on Operating Systems Design and Implementation*. Hollywood, USA, 2012, pp. 17–30.
  28. Liu Y.-Y., Slotine J.-J., Barabasi A.-L. Observability of complex systems. *Proc. National Academy of Science of the USA*, 2013, vol. 110, no. 7, pp. 2460–2465. doi: 10.1073/pnas.1215508110
  29. Yoo A., Henderson K. Parallel generation of massive scale-free graphs. *Proceedings of CoRR*, 2010.
  30. Ryu E., Choi S. Stochastic Kronecker Graph on Vertex-Centric BSP. *Proceedings of CoRR*, 2012, art. arXiv1208.2239.
  31. Paul G., Tanizawa T., Havlin S., Stanley H.E. Optimization of robustness of complex networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 2004, vol. 38, no. 2, pp. 187–191. doi: 10.1140/epjb/e2004-00112-3
  32. Watts D.J., Strogatz S.H. Collective dynamics of 'small-world' networks. *Nature*, 1998, vol. 393, no. 6684, pp. 440–442. doi: 10.1038/30918
  33. Albert R., Barabasi A.L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 2002, vol. 74, no. 1, pp. 47–97. doi: 10.1103/RevModPhys.74.47
  34. Newman M.E.J. The structure and function of complex networks. *SIAM Review*, 2003, vol. 45, no. 2, pp. 167–256.
  35. Kashirin V.V. Evolutionary simulation of complex networks structures with specific topological properties. *Procedia Computer Science*, 2014, vol. 29, pp. 2401–2411. doi: 10.1016/j.procs.2014.05.224
  36. Voloshin D.V., Puzyreva K.A., Karbovskii V.A. Agent-based virtual society polygon for simulation and evaluation in massive mobile services. *IERI Procedia*, 2014, vol. 10, pp. 231–238. doi: 10.1016/j.ieri.2014.09.082
  37. Mahdian M., Xu Y. Stochastic kronecker graphs. *Lecture Notes in Computer Science*, 2007, vol. 4863, pp. 179–186.

- Notes in Computer Science. 2007. V. 4863. P. 179–186.
38. Knyazkov K.V., Kovalchuk S.V., Tchurov T.N., Maryin S.V., Boukhanovsky A.V. CLAVIRE: e-Science infrastructure for data-driven computing // *Journal of Computational Science*. 2012. V. 3. N 6. P. 504–510. doi: 10.1016/j.jocs.2012.08.006
  39. Xiong F., Liu Y., Zhang Z.-J., Zhu J., Zhang Y. An information diffusion model based on retweeting mechanism for online social media // *Physics Letters A*. 2012. V. 376. N 30. P. 2103–2108. doi: 10.1016/j.physleta.2012.05.021
  40. Gjoka M., Kurant M., Butts C.T., Markoulou A. Walking in facebook: a case study of unbiased sampling of OSNs // *Proceedings IEEE INFOCOM*. San Diego, USA, 2010. Art. 5462078. doi: 10.1109/INFCOM.2010.5462078
  41. Arnaboldi V., Conti M., Passarella A., Pezzoni F. Ego networks in twitter: an experimental analysis // *Proceedings IEEE INFOCOM*. Turin, Italy, 2013. P. 3459–3464. doi: 10.1109/INFCOM.2013.6567181
  42. Lee C.H., Xu X., Eun D.Y. Beyond random walk and metropolis-hastings samplers: why you should not backtrack for unbiased graph sampling // *Performance Evaluation Review*. 2012. V. 40. N 1. P. 319–330. doi: 10.1145/2254756.2254795
  43. Wang F., Landau D.P. Efficient, multiple-range random walk algorithm to calculate the density of states // *Physical Review Letters*. 2001. V. 86. N 10. P. 2050–2053. doi: 10.1103/PhysRevLett.86.2050
  44. Buccafurri F., Lax G., Nocera A., Ursino D. Crawling social internetworking systems // *Proc. Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM)*. Istanbul, Turkey, 2012. P. 506–510. doi: 10.1109/ASONAM.2012.87
  45. Buccafurri F., Lax G., Nocera A., Ursino D. A system for extracting structural information from Social Network accounts // *Software – Practice and Experience*. 2015. V. 45. N 9. P. 1251–1275. doi: 10.1002/spe.2280
  46. Serrano D., Stroulia E., Barbosa D., Guana V. SociQL: a query language for the socialweb // *Advances in Network Analysis and its Applications*. 2012. V. 18. P. 381–406. doi: 10.1007/978-3-642-30904-5\_17
  47. Marcus A., Bernstein M.S., Badar O., Karger D.R., Madden S., Miller R.C. Processing and visualizing the data in tweets // *SIGMOD Record*. 2012. V. 40. N 4. P. 21–27. doi: 10.1145/2094114.2094120
  48. Psallidas F., Ntoulas A., Delis A. SocWeb: efficient monitoring of social network activities // *Lecture Notes in Computer Science*. 2013. V. 8181. P. 118–136. doi: 10.1007/978-3-642-41154-0\_9
  49. Hu Y., Shi L. Visualizing large graphs // *Wiley Interdisciplinary Reviews: Computational Statistics*. 2015. V. 7. N 2. P. 115–136. doi: 10.1002/wics.1343
  50. Auber D., Chiricote Y., Jourdan F., Melancon G. Multiscale visualization of small world networks // *Proc. IEEE Symposium on Information Visualization*. Seattle, USA, 2003. P. 75–84.
  51. Tominski C., Abello J., Van Ham F., Schumann H. Fisheye tree views and lenses for graph visualization // *Proc. Int. Conf. on Information Visualization*. London, 2006. P. 17–24. doi: 10.1109/IV.2006.54
  52. Bender-deMoll S., McFarland D.A. The art and science of dynamic network visualization // *Journal of Social Structure*. 2006. V. 7. N 2. P. 1–38.
  53. Moody J., McFarland D., Bender-deMoll S. Dynamic network visualization // *American Journal of Sociology*. 2005. V. 110. P. 1206–1241. doi: 10.1086/421509
  54. Collberg C., Kobourov S., Nagra J., Pitts J., Wampler K. A system for graph-based visualization of the evolution of software // *Proc. ACM Symposium on Software Visualization*. San Diego, USA, 2003. P. 77–86.
  55. Ahn J.-W., Taieb-Maimon M., Sopan A., Plaisant C., Shneiderman B. Temporal visualization of social network dynamics: prototypes for nation of neighbors // *Lecture Notes in Computer Science*. 2011. V. 6589. P. 309–316. doi: 10.1007/978-3-642-19656-0\_43
  56. Egorov D., Bezgodov A. Improved force-directed method of graph layout generation with adaptive step length // *Procedia Computer Science*. 2015. V. 66. P. 689–696. doi: 10.1016/j.procs.2015.11.078
  57. Guschin A., Shmelev V., Karbovskii V. Use of modern information
  38. Knyazkov K.V., Kovalchuk S.V., Tchurov T.N., Maryin S.V., Boukhanovsky A.V. CLAVIRE: e-Science infrastructure for data-driven computing. *Journal of Computational Science*, 2012, vol. 3, no. 6, pp. 504–510. doi: 10.1016/j.jocs.2012.08.006
  39. Xiong F., Liu Y., Zhang Z.-J., Zhu J., Zhang Y. An information diffusion model based on retweeting mechanism for online social media. *Physics Letters A*, 2012, vol. 376, no. 30, pp. 2103–2108. doi: 10.1016/j.physleta.2012.05.021
  40. Gjoka M., Kurant M., Butts C.T., Markoulou A. Walking in facebook: a case study of unbiased sampling of OSNs. *Proceedings IEEE INFOCOM*. San Diego, USA, 2010, art. 5462078. doi: 10.1109/INFCOM.2010.5462078
  41. Arnaboldi V., Conti M., Passarella A., Pezzoni F. Ego networks in twitter: an experimental analysis. *Proceedings IEEE INFOCOM*. Turin, Italy, 2013, pp. 3459–3464. doi: 10.1109/INFCOM.2013.6567181
  42. Lee C.H., Xu X., Eun D.Y. Beyond random walk and metropolis-hastings samplers: why you should not backtrack for unbiased graph sampling. *Performance Evaluation Review*, 2012, vol. 40, no. 1, pp. 319–330. doi: 10.1145/2254756.2254795
  43. Wang F., Landau D.P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters*, 2001, vol. 86, no. 10, pp. 2050–2053. doi: 10.1103/PhysRevLett.86.2050
  44. Buccafurri F., Lax G., Nocera A., Ursino D. Crawling social internetworking systems. *Proc. Int. Conf. on Advances in Social Networks Analysis and Mining, ASONAM*. Istanbul, Turkey, 2012, pp. 506–510. doi: 10.1109/ASONAM.2012.87
  45. Buccafurri F., Lax G., Nocera A., Ursino D. A system for extracting structural information from Social Network accounts. *Software – Practice and Experience*, 2015, vol. 45, no. 9, pp. 1251–1275. doi: 10.1002/spe.2280
  46. Serrano D., Stroulia E., Barbosa D., Guana V. SociQL: a query language for the socialweb. *Advances in Network Analysis and its Applications*, 2012, vol. 18, pp. 381–406. doi: 10.1007/978-3-642-30904-5\_17
  47. Marcus A., Bernstein M.S., Badar O., Karger D.R., Madden S., Miller R.C. Processing and visualizing the data in tweets. *SIGMOD Record*, 2012, vol. 40, no. 4, pp. 21–27. doi: 10.1145/2094114.2094120
  48. Psallidas F., Ntoulas A., Delis A. SocWeb: efficient monitoring of social network activities. *Lecture Notes in Computer Science*, 2013, vol. 8181, pp. 118–136. doi: 10.1007/978-3-642-41154-0\_9
  49. Hu Y., Shi L. Visualizing large graphs. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2015, vol. 7, no. 2, pp. 115–136. doi: 10.1002/wics.1343
  50. Auber D., Chiricote Y., Jourdan F., Melancon G. Multiscale visualization of small world networks. *Proc. IEEE Symposium on Information Visualization*. Seattle, USA, 2003, pp. 75–84.
  51. Tominski C., Abello J., Van Ham F., Schumann H. Fisheye tree views and lenses for graph visualization. *Proc. Int. Conf. on Information Visualization*. London, 2006, pp. 17–24. doi: 10.1109/IV.2006.54
  52. Bender-deMoll S., McFarland D.A. The art and science of dynamic network visualization. *Journal of Social Structure*, 2006, vol. 7, no. 2, pp. 1–38.
  53. Moody J., McFarland D., Bender-deMoll S. Dynamic network visualization. *American Journal of Sociology*, 2005, vol. 110, pp. 1206–1241. doi: 10.1086/421509
  54. Collberg C., Kobourov S., Nagra J., Pitts J., Wampler K. A system for graph-based visualization of the evolution of software. *Proc. ACM Symposium on Software Visualization*. San Diego, USA, 2003, pp. 77–86.
  55. Ahn J.-W., Taieb-Maimon M., Sopan A., Plaisant C., Shneiderman B. Temporal visualization of social network dynamics: prototypes for nation of neighbors. *Lecture Notes in Computer Science*, 2011, vol. 6589, pp. 309–316. doi: 10.1007/978-3-642-19656-0\_43
  56. Egorov D., Bezgodov A. Improved force-directed method of graph layout generation with adaptive step length. *Procedia Computer Science*, 2015, vol. 66, pp. 689–696. doi: 10.1016/j.procs.2015.11.078
  57. Guschin A., Shmelev V., Karbovskii V. Use of modern information and computer technologies in historical and urban

- mation and computer technologies in historical and urban studies with the example of epidemiological situation in the city of Simbirsk during the First World War // Proc. 9<sup>th</sup> Int. Conf. on Application of Information and Communication Technologies. Rostov-on-Don, Russia, 2015. P. 527–530. doi: 10.1109/ICAICT.2015.7338616
58. Goodhart C., Hofmann B., Segoviano M. Bank regulation and macroeconomic fluctuations // *Oxford Review of Economic Policy*. 2004. V. 20. N 4. P. 591–615. doi: 10.1093/oxrep/grh034
  59. Rudebusch G.D. et al. Macroeconomic models for monetary policy // FRBSF Economic Letter. 2002.
  60. Sparrow M.K. The application of network analysis to criminal intelligence: an assessment of the prospects // *Social Networks*. 1991. V. 13. N 3. P. 251–274. doi: 10.1016/0378-8733(91)90008-H
  61. Duijn P.A.C., Klerks P. Social network analysis applied to criminal networks: recent developments in Dutch law enforcement / In: *Networks and Network Analysis for Defense and Security*. Springer, 2014. P. 121–159. doi: 10.1007/978-3-319-04147-6\_6
  62. Milward H.B., Raab J. Dark networks as organizational problems: elements of a theory // *International Public Management Journal*. 2006. V. 9. N 3. P. 333–360. doi: 10.1080/10967490600899747
  63. Duijn P.A.C., Kashirin V., Sloot P.M.A. The relative ineffectiveness of criminal network disruption // *Scientific Reports*. 2013. V. 4. Art. 4238. doi: 10.1038/srep04238
  - studies with the example of epidemiological situation in the city of Simbirsk during the First World War. *Proc. 9<sup>th</sup> Int. Conf. on Application of Information and Communication Technologies*. Rostov-on-Don, Russia, 2015, pp. 527–530. doi: 10.1109/ICAICT.2015.7338616
  58. Goodhart C., Hofmann B., Segoviano M. Bank regulation and macroeconomic fluctuations. *Oxford Review of Economic Policy*, 2004, vol. 20, no. 4, pp. 591–615. doi: 10.1093/oxrep/grh034
  59. Rudebusch G.D. et al. Macroeconomic models for monetary policy. *FRBSF Economic Letter*, 2002.
  60. Sparrow M.K. The application of network analysis to criminal intelligence: an assessment of the prospects. *Social Networks*, 1991, vol. 13, no. 3, pp. 251–274. doi: 10.1016/0378-8733(91)90008-H
  61. Duijn P.A.C., Klerks P. Social network analysis applied to criminal networks: recent developments in Dutch law enforcement. In: *Networks and Network Analysis for Defense and Security*. Springer, 2014, pp. 121–159. doi: 10.1007/978-3-319-04147-6\_6
  62. Milward H.B., Raab J. Dark networks as organizational problems: elements of a theory. *International Public Management Journal*, 2006, vol. 9, no. 3, pp. 333–360. doi: 10.1080/10967490600899747
  63. Duijn P.A.C., Kashirin V., Sloot P.M.A. The relative ineffectiveness of criminal network disruption. *Scientific Reports*, 2013, vol. 4, art. 4238. doi: 10.1038/srep04238

#### Авторы

**Слоот Петрус Мария Арнольдус** – доктор физико-математических наук, профессор, заведующий кафедрой, Университет г. Амстердам, Амстердам, 1012WX, Нидерланды, p.m.a.sloot@uva.nl

**Холист Януш** – доктор физико-математических наук, профессор, директор научно-инновационного центра «Исследование сложных систем», Варшавский политехнический университет, Варшава, 00-661, Польша, jholyst@if.pw.edu.pl

**Кампис Джордж** – доктор наук, профессор, профессор, Будапештский университет имени Лоранда Этвёша, Будапешт, 1053, Венгрия, gk@hps.elte.hu

**Лис Майкл Харольд** – PhD, доцент, Университет г. Амстердам, Амстердам, 19268, Нидерланды, m.h.lees@uva.nl

**Митягин Сергей Александрович** – кандидат технических наук, директор Института дизайна и урбанистики, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, mityagin@corp.ifmo.ru

**Иванов Сергей Владимирович** – кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sergei.v.ivanov@gmail.com

**Боченко Клавдия Олеговна** – кандидат технических наук, младший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, k.bochenina@gmail.com

**Гулева Валентина Юрьевна** – инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Valentina.gul.va@gmail.com

**Мухина Ксения Дмитриевна** – инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, mukhinaks@gmail.com

**Насонов Денис Александрович** – преподаватель, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, denis.nasonov@gmail.com

**Бутаков Николай Алексеевич** – младший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, alipoov.nb@gmail.com

**Леоненко Василий Николаевич** – кандидат физико-математических наук, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, vnleonenko@yandex.ru

**Ланцева Анастасия Андреевна** – инженер, Университет

#### Authors

**Petrus M.A. Sloot** – D.Sc., Professor, Head of Chair, University of Amsterdam (UvA), Amsterdam, 1012WX, the Netherlands, p.m.a.sloot@uva.nl

**Janusz Holyst** – D.Sc., Professor, Director of the Center of Excellence “Complex Systems Research”, University of Warsaw, Warsaw, 00-661, Poland, jholyst@if.pw.edu.pl

**George Kampis** – D.S., Full Professor, Eötvös University, Budapest, 1053, Hungary, gk@hps.elte.hu

**Michael Harold Lees** – PhD, Associate professor, University of Amsterdam (UvA), Amsterdam, 19268, the Netherlands, m.h.lees@uva.nl

**Sergey A. Mityagin** – PhD, Head of Institute of Design and Urban Studies, ITMO University, Saint Petersburg, 197101, Russian Federation, mityagin@corp.ifmo.ru

**Sergey V. Ivanov** – PhD, Associate professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sergei.v.ivanov@gmail.com

**Klavdiya O. Bochenina** – PhD, junior researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, k.bochenina@gmail.com

**Valentina Yu. Guleva** – engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, Valentina.gul.va@gmail.com

**Ksenia D. Mukhina** – engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, mukhinaks@gmail.com

**Denis A. Nasonov** – lecturer, ITMO University, Saint Petersburg, 197101, Russian Federation, denis.nasonov@gmail.com

**Nikolay A. Butakov** – junior researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, alipoov.nb@gmail.com

**Vasily N. Leonenko** – PhD, engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, vnleonenko@yandex.ru

**Anastasia A. Lantseva** – engineer, ITMO University, Saint

ИТМО, Санкт-Петербург, 197101, Российская Федерация, stasja16@inbox.ru  
 Petersburg, 197101, Russian Federation, stasja16@inbox.ru

**Бухановский Александр Валерьевич** – доктор технических наук, директор мегафакультета, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, avb\_mail@mail.ru

**Alexander V. Boukhanovsky** – D.Sc., Head of School of Translational Information Technologies, ITMO University, Saint Petersburg, 197101, Russian Federation, avb\_mail@mail.ru



**Петрус Мария Арнолдус Слоот** – заслуженный профессор-исследователь Института информатики Университета г. Амстердам (UvA), приглашенный профессор в Наньянском технологическом университете (НТУ), Сингапур и профессор кафедры высокопроизводительных вычислений Университета ИТМО. Получил ученую степень PhD в 1989 г. в Университете г. Амстердам. Специалист в области информационных технологий поддержки принятия решений в слабоформализованных областях знания (включая биомедицину и социальные науки) на основе методов математического моделирования сложных систем. Автор около 500 научных публикаций (более 4 тыс. цитирований) и 4 объектов интеллектуальной собственности. Подготовил 15 PhD. Руководитель и участник международных исследовательских проектов, включая европроекты FP-6, FP-7 и HORIZON'2020. Главный редактор рецензируемых международных журналов: Journal of Computational Science и Future Generation of Computing Systems издательства Elsevier Science. Соорганизатор Института исследования сложных систем (Complexity Institute) в

Наньянском технологическом университете.

**P.M.A. Sloot** is a distinguished research professor at UvA, The Netherlands, professor of Complex Systems, NTU, Singapore, and endowed professor of advanced Computing, ITMO, St. Petersburg State University, Russia. His research focuses on understanding how nature processes information. He studies this 'natural information processing' in complex systems by computational modeling and simulation as well as through formal methods. His work is applied to a large variety of disciplines with a focus on -but not limited to- Biomedicine and Socio-Economic systems. Recent work is on modeling the virology and epidemiology of infectious diseases, notably HIV, through Complex Networks, Cellular Automata and Multi-Agents. Recently he tries to build bridges to socio-dynamics. He has published some 500 research papers, edited many books. He is editor in chief of both the Journal of Computational Science and Future Generation Computing Systems, and is one of the founding fathers and long-term chairman of the International Conference on Computational Science. He has been involved in many EU research projects, and was coordinator of EU-funded VIROLAB and DYNANETS projects.



**Януш Холист** – профессор факультета физики Варшавского политехнического университета, руководитель научно-инновационного центра «Исследование сложных систем» и лаборатории «Физика в экономике и социальных науках». Научные интересы включают моделирование эволюционирующих сетей, обработку информации в иерархических системах, модели коллективного формирования эмоций и мнений, эконофизику, равновесную и неравновесную статистическую физику. Является одним из исследователей, заложивших основы применения физических методов к экономическим и социальным системам, и работает в тесном сотрудничестве с многими научными учреждениями Германии, Италии, Великобритании, Швейцарии, Японии и США. Редактор European Physical Journal B и Journal of Computational Science. Сооснователь и действительный член Национального совета координаторов научных проектов, сооснователь и бывший член отделения «Физика в экономике и социальных науках» Польского физического общества. Был координатором проекта CYBEREMOTIONS.

**Janusz Holyst** is full professor at the Faculty of Physics, Warsaw University of Technology, where he leads the Center of Excellence for Complex Systems Research and the lab 'Physics in Economy and Social Sciences'. His current research includes simulations of evolving networks, information processing in hierarchical systems, models of collective opinion and emotion formation, econophysics, and equilibrium and nonequilibrium statistical physics. He is one of the pioneers in applications of physical methods to economic and social systems and maintains a close collaboration with many institutes in Germany, Italy, UK, Switzerland, Japan and USA. He is editor of the European Physical Journal B and Journal of Computational Science. He is a co-founder and acting chairman of KRAB (National Council for Research Projects Coordinators) as well as a co-founder and the former chairman of FENS (Physics in Economy and Social Sciences, Division of the Polish Physical Society). Janusz Holyst was coordinator of the CYBEREMOTIONS project.



**Джордж Кампис** – основатель и профессор (с 1994 года) кафедры истории и философии науки в Будапештском университете имени Лоранда Этвёша. Имеет степени PhD и доктора биологических наук, доктора в области философии науки. Основные научные интересы – область искусственной жизни, когнитивистика, сложные системы и эволюционное моделирование. Занимал должность приглашенного профессора в Университете Хоккайдо, Японском университете науки и технологий (JAIST), Университете Восточного Теннесси, Университета Индианы. Доктор Кампис имеет около 150 научных публикаций и является автором и издателем нескольких книг издательств Elsevier, Kluwer и Springer.

**George Kampis** has been acting as founding chairman and Professor (since 1994) of the department of History and Philosophy of Science at Eotvos University in Budapest. Holds a Ph.D. and a Habilitation in Biology and a D.Sc. in Philosophy of Science. Main research

interests in Artificial Life, cognitive science, complex systems and evolutionary modeling. He was guest professor at Hokkaido University, Fujitsu Chair of Complex Systems at JAIST (Japan Advanced Institute for Science and Technology), Wayne G. Basler Chair of Excellence at East Tennessee State University, and Fulbright fellow at Indiana University. Dr Kampis has about 150 scientific publications, and he is the author or editor of several books with international publishers Elsevier, Kluwer, and Springer.



**Майкл Харолд Лис** – доцент кафедры вычислительной науки на факультете естественных наук Университета г. Амстердам, Нидерланды. Защитил свою докторскую диссертацию в области адаптивного моделирования многоагентных систем и окончил Эдинбургский университет в июле 2001 года по двум направлениям – компьютерные науки и искусственный интеллект. В настоящее время доктор Лис также проводит совместные исследования с Наньянским технологическим университетом (НТУ), Сингапур.

**Michael Lees** is an Assistant Professor at the Universiteit van Amsterdam in the Computational Science Lab (Informatics Institute). He currently conducts regular research visits to Nanyang Technological University (NTU), Singapore. At the University of Amsterdam he is the Director for the Master Computational Science. He has completed his PhD in Adaptive Optimistic Simulation of Multi-Agent Systems at and graduated from Edinburgh University in July of 2001 with joint honours in Computer Science/Artificial Intelligence degree.



**Сергей Александрович Митягин** – директор Института дизайна и урбанистики Университета ИТМО. В 2012 году защитил кандидатскую диссертацию на тему «Моделирование процессов наркотизации населения на основе комплексных сетей». Сфера научных интересов – комплексные сети, моделирование критических явлений в социальных системах, модели формирования мнений, анализ данных социальных медиа.

**Sergey Mityagin** is a Head of Institute of Design and Urban Studies. In 2012 he defended PhD theses on the topic “Simulation of narcotization processes on the basis of complex networks”. His research interests include complex networks, simulation of critical phenomena in social systems, models of opinions formation, social media analysis.



**Сергей Владимирович Иванов** – старший научный сотрудник Университета ИТМО. Его научные интересы заключаются в моделировании процессов и систем с широким спектром приложений – от наук о Земле до вычислительной социологии. Он имеет длительный опыт применения технологий распределенной обработки данных и высокопроизводительных вычислений, а также моделирования сложных систем. Его последние исследования были направлены на разработку средств поддержки принятия решений для системы раннего предупреждения Комплекса защитных сооружений Санкт-Петербурга от наводнений. Является автором более 50 публикаций, в том числе в реферируемых в международных журналах. Также является победителем конкурса грантов Президента Российской Федерации для поддержки молодых кандидатов наук.

**Sergey Ivanov** is a Senior Researcher at ITMO University, Saint Petersburg, Russia. His research focuses on simulating real-world processes and systems, mainly in social and environmental science. He has a long-standing expertise in high performance/distributed computing and

Complex Systems simulations. His recent work concentrates on decision-making processes for the Early Warning System of the St Petersburg Flood Protection Barrier. He has published over 50 research papers. He has extensive experience participating in Russian and international projects. He is the winner of a grant competition of the President of the Russian Federation for support of young Ph.D scientists.



**Клавдия Олеговна Боченина** – младший научный сотрудник НИИ наукоемких компьютерных технологий Университета ИТМО, кандидат технических наук. Защитила диссертацию по алгоритмам планирования научных композитных приложений в 2014 году. Сферы научных интересов: анализ и разработка алгоритмов, моделирование динамических процессов в комплексных сетях, суперкомпьютерное моделирование критических явлений в сложных системах, параллельные и распределенные вычисления, машинное обучение.

**Klavdiya Bochenina** is a Junior Researcher at eScience Research Institute of ITMO University (Saint-Petersburg). She obtained a PhD in computer science from the same institution with a work on scientific workflow scheduling. Her research activities are focused on several fields including algorithms analysis and design, modeling of dynamical processes in complex networks, supercomputer simulation of critical phenomena in complex systems, parallel and distributed computing and machine learning.





**Валентина Юрьевна Гулева** окончила математико-механический факультет Санкт-Петербургского государственного университета в 2014 году. С 2014 года является инженером Института наукоемких компьютерных технологий, сотрудником международной лаборатории «Городская информатика» и аспирантом кафедры высокопроизводительных вычислений в Университете ИТМО. Работает в научной группе «Алгоритмика сложных систем», где занимается моделированием банковских систем, исследованием критических явлений на комплексных сетях. Ее основные научные интересы лежат в области имитационного моделирования сложных систем, в частности, сетей рынка межбанковских кредитов, экономических моделей и моделирования критических явлений.

**Valentina Guleva** graduated from the Mathematics and Mechanics Faculty of Saint-Petersburg State University with specialist degree in 2014. Since 2014 she is an engineer of eScience Research Institute, an employee of The International Laboratory “Urban Informatics” and post-graduate student at High-Performance Computing Department of ITMO University. She is currently working in “Algorithmics of Complex Systems” research team, where she models interbank market and studies critical phenomena in complex networks. The main research interests include: modeling and simulation of complex systems, particularly, interbank networks, economic models and critical phenomena modeling.



**Ксения Дмитриевна Мухина** – инженер Института наукоемких компьютерных технологий и ассистент кафедры высокопроизводительных вычислений Университета ИТМО. С 2014 года работает в научной группе «Визуализация и компьютерная графика», где занимается научной визуализацией многомасштабных комплексных сетей и участвует в создании интерактивных приложений в различных научных областях (городские науки, социальные сети, медицина и здравоохранение и т.д.). В область ее научных интересов входят визуальный анализ динамических комплексных сетей, визуализация карт и географические информационные системы.

**Ksenia Mukhina** is a Lecturer at High Performance Computing Department and a Research Assistant at eScience Research Institute in ITMO University. Ksenia works in Visualization & Computer Graphics research team where she focuses on scientific visualization of multiscale complex networks and participates in developing interactive applications for variety of applied research fields (healthcare, medicine, urban sciences, etc.). Her research interests are in the areas of complex network visualization, map visualization, and geographical information systems.



**Денис Александрович Насонов** – старший разработчик проекта CLAVIRE. С 2010 года он активно участвует в разработке предметно-ориентированных языков Easy Flow и EasyPackage для объединенного воркфлоу и управляемых данными композитных приложений облачной платформы CLAVIRE. Денис разрабатывает решения для СППР в критических ситуациях с серьезными социальными последствиями (например, акты терроризма, преступная деятельность) и реакцию общества на внешние раздражители (например, паника при пожарах, наводнениях или эпидемиях).

**Denis Nasonov** is a Senior Developer of CLAVIRE Project. Since 2010 he is actively involved in the development of coupled domain-specific languages EasyFlow and EasyPackage for workflow scheduling design and data-driven composite applications of CLAVIRE platform. Denis develops solutions for decision-making process in critical situations with significant social impact (e.g. acts of terrorism, criminal activity) and social response to external stimuli (e.g. panic at fires or flooding, epidemics, etc.).



**Николай Алексеевич Бутаков** – окончил Ульяновский государственный университет и в 2015 году аспирантуру Университета ИТМО. Старший разработчик проекта CLAVIRE. С 2013 года он участвует в разработке управляемых данными композитных приложений облачной платформы CLAVIRE. Его научные интересы – распределенные вычисления, планирование композитных приложений, анализ социальных сетей, машинное обучение.

**Nikolay Butakov** has graduated from Ulyanovsk State University and in 2015 received his PhD degree at ITMO University. Starting 2013 he is an integral part of CLAVIRE project and engaged in the design of data-driven composite applications. His research interests are distributed computations, workflow scheduling, social network analysis, machine learning.



**Василий Николаевич Леоненко** закончил в 2008 году математический факультет Омского государственного университета. В 2012 году он защитил кандидатскую диссертацию по специальности 05.13.18 «Математическое моделирование, численные методы и комплексы программ» в ученом совете Института математики им. С.Л. Соболева СО РАН. К его научным интересам относятся математическое моделирование в биологии и эпидемиологии, в частности, детерминистские и имитационные модели распространения заболеваний в человеческих популяциях.

**Vasily Leonenko** graduated from the Mathematical faculty of Omsk State University in 2008. In 2012 he defended a Ph. D. thesis on Mathematical Modeling in Sobolev Institute of Mathematics of Russian Academy of Sciences. His main area of interest is mathematical modeling in biology and epidemiology, particularly the deterministic and imitational models of disease propagation in human populations.



**Анастасия Андреевна Ланцева** – инженер Института наукоемких компьютерных технологий в Университете ИТМО, ассистент кафедры высокопроизводительных вычислений Университета ИТМО. С 2014 года является сотрудником международной лаборатории «Городская информатика». Научные интересы – моделирование сложных систем, в частности, криминальных сетей.

**Anastasia Lantseva** is a Research Assistant at eScience Research Institute and a Lecturer at High Performance Computing Department at ITMO University. Since 2014 she has been working in an international laboratory “Urban Informatics”. Her research interests include complex system modeling, particularly “Dark networks”.



**Александр Валерьевич Бухановский** – доктор технических наук. В 1997 году защитил кандидатскую диссертацию (физико-математические науки), а в 2005 году – докторскую диссертацию (технические науки) в области обработки больших данных и статистического моделирования. С 2006 года работал профессором кафедры информационных систем Университета ИТМО. В 2007 году организовал НИИ наукоемких компьютерных технологий Университета ИТМО, а в 2011 году – кафедру высокопроизводительных вычислений Университета ИТМО, которыми руководит по сей день. Специалист в области высокопроизводительных вычислений и компьютерного моделирования сложных систем. Автор 230 публикаций (более 1 тыс. цитирований), 28 объектов интеллектуальной собственности. Подготовил 23 кандидата наук. Идеолог и организатор первых в России международных магистерских программ двойного диплома в области суперкомпьютерных технологий и больших данных. Редактор рецензируемого международного журнала Future Generation of Computing Systems издательства Elsevier Science.

**Alexander Boukhanovsky** is the Chair of High Performance Computing (HPC) Department in ITMO University. In 2005, he defended his dissertation on Concurrent Software Statistical Measurements of Spatial-Temporal Fields. Since 2006 he has been working as a Professor of Information Systems and Head of the Parallel Software lab in ITMO University. In 2007 he created the eScience Research Institute where his team has created CLAVIRE (CLOUD Applications VIRTUAL Environment). In recent years he attracted several grants including mega grants of the Russian Federation Government, e.g. decree #220 “on measures to attract Leading Scientists in the Russian educational institution” and decree #218 “cooperation of Russian higher education institutions and organizations implementing complex projects of high-tech industry”. His research interests are high-performance computing, computer modelling of complex systems, intelligent computational technologies, statistical analysis and synthesis of spatial-temporal fields, parallel and distributed computing, distributed environments for multidisciplinary researches, decision support systems and technologies, statistical analysis and simulation in marine sciences. He is the author of 230 publications (cited over 1000 times) and has successfully advised 23 PhD candidates.