Petrovich V. N., Ph.D. in Engineering Science
Trebina N. N., leading engineer-mathematician,
Dvirnichuk K. V., researcher,
Taras Shevchenko National University of Kyiv

ABOUT ONE APPROACH TO PROBLEM SOLVING OF ONE DECISION OF TASK OF THE MATHEMATICAL PROGRAMMING OF UNIDIMENSIONAL DYNAMIC SYSTEM WITH INCOMPLETE CERTAIN BORDER STATE

The task of construction of integral after Laplace image of function of the state of arbitrary is set forth and untied incompletely on the regional terms of the looked after unidimensional distributed dynamic system at the terms of complete information about her initial state. Executed estimation of exactness of the got decision and laid down condition his unambiguity.

Key words: *linear systems, Grine functional, dynamical process.*

УДК 519.1 + 681.518.

В. Г. Скобелев, д-р фіз.-мат. наук, д-р техн. наук, проф.
Институт кибернетики имени В. М. Глушкова НАН Украины, Киев

ПРОБЛЕМЫ АНАЛИЗА И СИНТЕЗА КРУПНОМАСШТАБНЫХ СЕТЕЙ (ОБЗОР)

В настоящей работе содержится обзор состояния исследований некоторых актуальных проблем анализа и синтеза крупномасштабных информационных сетей. Подробно рассмотрены наиболее часто используемые методы выделения как непересекающихся, так и пересекающихся сообществ, основанные на анализе только топологии исследуемой сети. Охарактеризованы основные подходы, модели и методы, используемые в процессе анализа социальных сетей. Выделены некоторые актуальные проблемы, возникающие в процессе проектирования крупномасштабных информационных сетей, и кратко рассмотрены существующие подходы к их решению. Охарактеризованы основные модели и методы, применяемые для обеспечения безопасности крупномасштабных информационных сетей.

Ключевые слова: крупномасштабные информационные сети, анализ, синтез, безопасность, сообщества, онлайновые социальные сети.

Введение. В настоящее время проникновение информационных технологий во все сферы жизнедеятельности человечества достигло "критической массы", давшей старт фундаментальным изменениям в общении людей, экономике, технике, военной деятельности, науке, медицине, образовании и т. д. Эти изменения напрямую связаны с развитием сетевых информационных технологий, где одно из центральных мест принадлежит разработке крупномасштабных информационных сетей (КИС), т. е. масштабируемых информационных сетей (ИС), покрывающих большие географические регионы, и включающих в себя ИС различных типов. Сказанное подтверждается значительными усилиями и средствами, направляемыми ведущими странами для формирования глобальной информационной инфраструктуры − GII, реализуемой на основе концепции открытых систем.

По уровню архитектуры КИС можно выделить коммуникационные ИС (т. е. системы физических каналов связи, реализующие протокол передачи данных между территориально разделенными пользователями и абонентскими системами) и информационно-вычислительные ИС, предоставляющие по запросам пользователей и систем информационные и вычислительные ресурсы и услуги. К КИС второго относятся корпоративные и военные телекоммуникационные сети, ИС управления энергетикой, авиа- и ж.д.-перевозками, транспортировкой нефти и газа, квантовыми физическими экспериментами и т. д. Свои особенности в решение проблем анализа и синтеза КИС вносит также тот фактор, что объекты исследования являются сложными взаимодействующими между собой аппаратно-программными комплексами, распределенными на значительных расстояниях.

Цель настоящего обзора состоит в том, чтобы рассмотреть существующие подходы к решению некоторых проблем, связанных с анализом и синтезом КИС.

1. Выделение сообществ в КИС. Известно, что до сих пор отсутствует общепринятое определение понятия "сообщество в сети" (см., напр., [1–4]), а для любого из существующих определений несложно построить контрпример. Тем не менее, общепризнано, что для большинства КИС характерны общинные структуры. На их исследование направлены значительные усилия разработчиков КИС, государственных и коммерческих структур. Чтобы охарактеризовать методы выделения сообществ в КИС, достаточно считать, что для каждой КИС S можно построить (динамическую) сеть U_S, отражающую коммуникации ее пользователей, а под сообществом в сети S понимать ту или иную подсеть сети U_S с относительно редкими подключениями и уходом частей.

Выделение сообществ в КИС сводится к решению одной из следующих двух задач: 1) выделение сообществ только на основе топологии сети; 2) выделение и анализ сообществ пользователей (*в этом случае обычно КИС называют социальной сетью, а сообщества − социальными группами*), объединенных в группы либо явно за счет встроенных в КИС средств образования групп, либо неявно (т. е. за счет установления связей на основе общих интересов, деятельности, кругов общения и т. д.). Вторая задача намного сложнее первой, так как социальные группы, как правило, являются сильно перекрывающимися частями сообществ, выделенных только на основе топологии сети [5, 6]. Рассмотрим методы решения этих задач.

1.1. Выделение сообществ на основе топологии КИС. Из-за большой сложности алгоритмического решения этой задачи (*известно, что даже разбиение сети на данное число подсетей примерно равных размеров при условии минимизации числа ребер между ними является NP-трудной задачей* [7, 8]) единственным приемлемым на практике способом ее решения являются эвристические методы. Охарактеризуем эти методы (*сравнительный анализ эвристических методов построения сообществ в сетях и их реализаций содержится в* [9–11]).

1.1.1. Выделение непересекающихся сообществ. При решении этой задачи часто используется схема, основанная на следующих двух предположениях: 1) сеть U_S − это обычный граф; 2) число ребер внутри любого сооб-

щества существенно больше, чем число случайным образом построенных ребер между этим сообществом и любым другим сообществом, при условии, что сохраняются степени вершин. Вычисления состоят в последовательном измельчении разбиения (*иными словами, схема организована по принципу "сверху вниз"*) множества вершин сети U_S [12–14], и осуществляются следующим образом.

Разбиение множества вершин $V=\{1,\ldots,n\}$ сети U_S на два непересекающихся сообщества сводится к вычислению наибольшего значения квадратичной формы (*модульность (modularity)*) $Q=(4m)^{-1}\sum_{i,j=1}^{n}(a_{ij}-(2m)^{-1}k_ik_j)s_is_j$, где m − число ребер сети, a_{ij} $(i,j=1,\ldots,n)$ − элементы матрицы смежности A сети, k_i $(i=1,\ldots,n)$ − степень вершины i, а $s_i\in\{-1,1\}$ $(i=1,\ldots,n)$ − переменная (если i принадлежит 1-му сообществу, то $s_i=1$, а если 2-му сообществу, то $s_i=-1$) (*содержательно, значение квадратичной формы Q равно разности между долей ребер, связывающих вершины, принадлежащие сообществу и ожидаемой такой же долей для случая, когда ребра сети распределены случайным образом, при условии, что сохраняются степени вершин*). Доказано, что все значения Q принадлежат отрезку $[-0.5,1]$. При этом $Q>0$ на тех и только на тех наборах значений переменных, которые разбивают сеть U_S на два непустых сообщества.

Вычисление наибольшего значения квадратичной формы Q осуществляется следующим образом. Матрицей модульности сети U_S называется $n\times n$-матрица B, элементы которой имеют вид $b_{ij}=a_{ij}-(2m)^{-1}k_ik_j$ $(i,j=1,\ldots,n)$ (*отметим, что в терминах этой матрицы $Q=(4m)^{-1}\mathbf{s}^TB\mathbf{s}$, где $\mathbf{s}=(s_1,\ldots.s_n)^T$*). Пусть β − наибольшее собственное число матрицы B, а $\mathbf{u}=(u_1,\ldots,u_n)^T$ − собственный вектор, отвечающий этому собственному значению. Если $\beta\le 0$, то разбиения сети U_S на два сообщества не существует. В случае, когда $\beta>0$, полагаем $s_i=1$, если $u_i>0$, и $s_i=-1$ иначе (*если $\mathbf{u}=(u_1,\ldots,u_n)^T$ − нормализованный собственный вектор, то число $|u_i|$ может интерпретироваться как "вес" вершины i в 1-м сообществе, если $s_i=1$, и как "вес" вершины i во 2-м сообществе, если $s_i=-1$*).

Если произошло разбиение сети U_S на два сообщества, то описанные выше вычисления применяются к подсетям, определяемым каждым из двух блоков полученного разбиения и т. д., до тех пор, пока дальнейшее измельчение разбиения невозможно.

В результате экспериментов были выявлены следующие два недостатка базовой схемы. Во-первых, не всегда обнаруживались относительно небольшие сообщества (*по-видимому, это связано с расплывчатостью предположения о том, что "число ребер внутри любого сообщества существенно больше, чем случайным образом построенное количество ребер между этим сообществом и любым другим сообществом"*). Во-вторых, некоторые сообщества нельзя было выделить в принципе, так как в базовой схеме заложено предположение о том, что каждая вершина сети принадлежит единственному сообществу. В терминах теории графов это означает, что структура сети U_S представляет собой иерархию вложенных друг в друга двудольных графов. Однако эксперименты показали, что значительное число реальных КИС содержат многодольные графы, в которых длина цикла между некоторыми долями − нечетное число (*именно в таких сетях встречались сообщества, которые в принципе не могла выявить базовая схема*). Для выявления сообществ в таких КИС в [15] была предложена следующая мера оценки сцепления (cohesion) подмножества вершин X сети U_S, основанная на использовании треугольников:

$$c(X)=\frac{\Delta_i(X)}{\binom{|X|}{3}}\cdot\frac{\Delta_i(X)}{\Delta_i(X)+\Delta_o(X)},$$

где $\binom{n}{m}$ − число сочетаний из n по m, $\Delta_i(X)$ − число треугольников в подсети, определяемой множеством вершин X, а $\Delta_o(X)$ − число треугольников в сети U_S, у которых в точности две вершины принадлежат множеству X. В [16] понятие "сцепление подмножества вершин" было обобщено для взвешенных и направленных графов (*отметим, что понятие "мультиграф" является специальным случаем понятия "взвешенный граф"*).

Высокая трудоемкость базовой схемы стимулировала разработку эвристических методов, построенных как с учетом, так и без учета значений модульности. Рассмотрим некоторые из таких методов, применяемых на практике.

1. Жадные методы, использующие значения модульности [17, 18]. В них осуществляется иерархическая кластеризация КИС U_S (*иными словами, эти методы основаны на принципе "снизу вверх"*) за счет ее преобразования в последовательность мультиграфов с петлями.

Вершинами очередного мультиграфа являются сообщества, построенные на данный момент. Кратные ребра, связывающие две вершины мультиграфа, соответствуют ребрам исходной сети, связывающим сообщества, соответствующие этим вершинам. Кратная петля в вершине мультиграфа соответствует ребрам исходной сети, находящимся внутри сообщества, представленного этой вершиной. Преобразование мультиграфа осуществляется за счет объединения его i-й и j-й вершин, т. е. в замене в матрице смежности i-й и j-й строк, а также i-го и j-го столбцов их суммами. Приращение значения модульности за счет такого объединения вершин вычисляется по формуле $\Delta Q=2(e_{ij}-a_ia_j)$, где e_{ij} − доля ребер мультиграфа, соединяющих его i-ю вершину с j-й, а $a_i=\sum_r e_{ir}$. На каждом шаге объединяются две вершины (*склеиваются в одну вершину*), для которых происходит максимальный прирост значения модульности.

Эффективность программных реализаций жадных методов достигается за счет выбора структур данных, позволяющих устранять "ненужные" операции при построении последовательности мультиграфов (*например, представле-*

ние матрицы смежности массивом списков смежности вершин дает возможность при объединении 2-х вершин складывать только ненулевые значения).

2. Методы, основанные на случайных блужданиях [19, 20]. В них используется то обстоятельство, что случайные блуждания по графу имеют тенденцию оказываться в "ловушке" в плотно соединенных его частях. Иерархическая кластеризация графа осуществляется применением математического аппарата конечных цепей Маркова. С этой целью определяется матрица переходов $P = D^{-1}A$ случайного блуждания, где D − $n \times n$-матрица, элементы которой имеют вид: $d_{ii} = k_{ii}$ $(i = 1,\ldots n)$ и $d_{ij} = 0$ $(i \neq j)$. В терминах матрицы P определяется расстояние между вершинами, используемое для построения возрастающей последовательности разбиений $\pi_1, \pi_2, \ldots, \pi_n$ сети на сообщества, где $\pi_1 = \{\{i\} \mid i \in V\}$, $\pi_n = \{V\}$, а π_{h+1} $(h = 1,\ldots,n-1)$ получается из π_h за счет объединения двух сообществ. Вычисления сводятся к поиску числа h $(1 \le h \le n)$, что π_h − результат выделения непересекающихся сообществ в сети. В [19] такой подход реализован следующим образом. Фиксируется достаточно большое натуральное число t. Расстояние между вершинами i и j определяется формулой $\rho_{ij} = \sqrt{\sum_{r=1}^{n} k_r^{-1}(p_{ir}^{(t)} - p_{jr}^{(t)})^2}$ (*это обычное расстояние между вероятностными распределениями* $p_{i*}^{(t)}$ *и* $p_{j*}^{(t)}$), где $p_{ij}^{(t)}$ $(i, j = 1,\ldots n)$ − элемент матрицы P^t (*таким образом,* $p_{ij}^{(t)}$ $(i, j = 1,\ldots n)$ *− это вероятность перехода из вершины* i *в вершину* j *за* t *шагов*). Так как вероятность перехода из сообщества C в вершину j равна $p_{Cj}^{(t)} = |C|^{-1} \sum_{i \in C} p_{ij}^{(t)}$, то расстояние между сообществами C_1 и C_2 определено формулой $\rho_{C_1C_2} = \sqrt{\sum_{i \in V} k_i^{-1}(p_{C_1 i}^{(t)} - p_{C_2 i}^{(t)})^2}$. На каждом шаге кластеризации объединяются два сообщества C_1 и C_2, для которых достигается минимум величины

$$\Delta(C_1, C_2) = \frac{1}{n} \cdot \frac{|C_1| \cdot |C_2|}{|C_1| + |C_2|} \cdot \rho_{C_1C_2}^2 .$$

После объединения сообществ C_1 и C_2 пересчитываются расстояния между сообществами, принадлежащими полученному разбиению множества вершин сети. В качестве такого числа h, что разбиение π_h − результат выделения непересекающихся сообществ в КИС U_S, выбирается число $r \in \{2,\ldots,n-1\}$, которому соответствует наибольшее значение величины $\eta_r = \frac{\sigma_{r+1} - \sigma_r}{\sigma_r - \sigma_{r-1}}$, где $\sigma_r = n^{-1} \sum_{C \in \pi_r} \sum_{i \in C} \rho_{iC}^2$. Эффективность программной реализации достигается за счет того, что пересчитывать необходимо только расстояния между смежными сообществами.

3. Методы, основанные на спектральной кластеризации [21, 22]. Осуществляют поиск таких попарно ортогональных собственных векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_h$ (число h заранее неизвестно), соответствующих наименьшим собственным значениям матрицы Лапласа $L = D - A$ сети U_S, что матрица X, столбцами которой являются эти векторы, минимизирует значение целевой функции $Tr(X^T D^{-0.5} L D^{-0.5} X)$. Каждый собственный вектор определяет двухблочное разбиение множества вершин сети. Пересечение этих разбиений − результат разбиения КИС U_S на непересекающиеся сообщества. Сложность применения таких методов обусловлена вычислением собственных векторов для достаточно больших сетей. Выходом из такой ситуации является многократное порождение "супер-вершин", соединенных с обычными вершинами, для преобразования сети в разряженный двудольный граф.

4. Методы, основанные на имитации отжига [23–25]. Являются вероятностными процедурами, направленными на максимизацию значения выбранной целевой функции $f(s)$ за счет последовательного преобразования состояния s сети (*например,* f *− модульность, а* s *− разбиение множества вершин сети*). Для этого выбирается начальная "температура" T_0 (эвристический параметр) и начальное состояние s_0. Переход из одного состояния в другое осуществляется в соответствии со следующей схемой:

Шаг 1. Вычисляется текущее значение "энергии" $E_{current} = 1 - f(s)$.

Шаг 2. В результате воздействия на состояние s строится новое состояние s' (*в стандартной реализации воздействие строится из двух видов преобразований: локальных, при которых одна вершина перемещается из одного кластера в другой, взятый наугад, кластер и глобальных, состоящих из слияний и разбиений сообществ. На практике, как правило, воздействие представляет собой комбинацию, состоящую из* l^2 *локальных и* l *глобальных преобразований, где* l *− фиксированное число*).

Шаг 3. Вычисляется новое значение "энергии" $E_{new} = 1 - f(s')$.

Шаг 4. Принять решение, полученное на шаге 2, с вероятностью $\mathrm{P} = exp(-T^{-1}(t) \cdot \max\{0, E_{new} - E_{current}\})$ (*эксперименты показали целесообразность использования схемы охлаждения* $T(t) = T_0 \cdot (1 + r \cdot t)^{-1}$, *где* r *− скорость охлаждения (эвристический параметр), а* t *− номер итерации*).

Шаг 5. Если энергия системы больше заданного значения c (критерий сходимости) или температура $T(t)$ не опустилась до предельно возможной, то увеличить номер итерации t и перейти на шаг 1, иначе – конец.

5. Методы, основанные на муравьиных колониях [26–31]. Предназначены для построения на множестве V ($|V| = n$) вершин сети U_S двухблочного разбиения $\pi = \{V_1, V_2\}$ ($|V_1| = n_1, |V_2| = n_2$) (n_1 *и* n_2 *− такие заданные натуральные числа, что* $n_1 + n_2 = n$) с как можно меньшим числом ребер между блоками. В стандартной реализации задействовано n муравьев. Каждый из них в течение $n_1 - 1$ итераций формирует свою строго возрастающую по включению последовательность множеств $V_{1i}(1), \ldots, V_{1i}(n_1)$, где i − номер муравья, $|V_{1i}(1)| = 1$ и $|V_{1i}(n_1)| = n_1$, при-

чем $V_{1i}(1)\cap V_{1j}(1)=\varnothing$ $(i\neq j)$ (*известны реализации, в которых задействовано* nl *муравьев, при этом каждая группа из* l *муравьев использует одно и то же множество* $V_{1i}(1)$). На каждой итерации моделирование поведения муравьёв осуществляется в терминах распределения ими феромона на ребрах полного графа $K_n=(V,E)$. Вначале на каждом ребре графа $K_n(1)$ откладывается количество феромона, равное $Q|E|^{-1}$, где Q − фиксированное число. В стандартной реализации метода t-я итерация $(t=1,\ldots,n_1-1)$ состоит из трех этапов.

Этап 1 (каждый муравей строит следующий элемент своей последовательности множеств вершин). Действия i-го муравья $(i=1,\ldots,n)$ определяются следующей вероятностной процедурой. Для каждой вершины $v\in\overline{V_{1i}(t)}$ $(\overline{V_{1i}(t)}=V\setminus V_{1i}(t))$ вычисляется число $a_i(t,v)$ ребер сети U_S, соединяющих вершину v с множеством $V_{1i}(t)$, суммарный уровень феромона $b_i(t,v)$ на ребрах графа $K_n(t)$, соединяющих вершину v с множеством $V_{1i}(t)$, стоимость $f_i(t,v)$ связей вершины v с множеством $V_{1i}(t)$ (*стоимость* $f_i(t,v)$ *представляется либо в мультипликативной форме* $f_i(t,v)=(b_i(t,v))^{\alpha}(a_i(t,v)+1)^{\beta}$, *либо в аддитивной форме* $f_i(t,v)=(b_i(t,v))^{\alpha}+(a_i(t,v))^{\beta}$, *где неотрицательные параметры* α *и* β *подбираются экспериментально. Если* $\alpha=0$, *то каждый муравей реализует жадный метод. Если* $\beta=0$, *то каждый муравей действует только на основании феромона, что, как правило, приводит к решениям, которые не являются оптимальными*) и вероятность $P_i(t,v)=f_i(t,v)\cdot(\sum\limits_{v\in\overline{V_{1i}(t)}}f_i(t,v))^{-1}$ включения вершины v в множество $V_{1i}(t+1)$. Случайным образом выбирается одна вершина $v\in\overline{V_{1i}(t)}$, для которой эта вероятность − наибольшая, и полагается $V_{1i}(t+1):=V_{1i}(t)\cup\{v\}$.

Этап 2 (муравьи откладывают феромон на ребрах графа $K_n(t)$). Количество феромона, откладываемого i-м муравьем $(i=1,\ldots,n)$ на каждом ребре подграфа графа $K_n(t)$, определенного множеством ребер $V_{1i}(t+1)$, равно $\Delta_i(t)=Q\cdot(c_i(t))^{-1}$, где $c_i(t)$ − число ребер сети U_S, соединяющих множества вершин $V_{1i}(t+1)$ и $\overline{V_{1i}(t+1)}$.

Этап 3 (испарение феромона). На каждом ребре графа $K_n(t)$, построенного на этапе 2, количество ферамона пересчитывается в соответствии с формулой $d:=d\cdot(1-\rho)$, где ρ − подбираемый экспериментально коэффициент обновления. В результате построен граф $K_n(t+1)$.

По окончанию описанной выше итеративной процедуры среди всех разбиений $\{V_{1i}(n_1),\overline{V_{1i}(n_1)}\}$ $(i=1,\ldots,n)$ выбирается разбиение с наименьшим числом ребер между блоками.

1.1.2. Выделение пересекающихся сообществ. Часто применяются следующие методы.

1. Методы, основанные на перколяции клик (CPM) [32–35]. В них осуществляется поиск всех максимальных клик в КИС U_S. После этого объединяются клики, имеющие общие вершины и удовлетворяющие заданному метрическому или топологическому критерию. Недостатком этих методов является экспоненциальная сложность от параметров КИС U_S.

2. *Методы, основанные на распространении меток* (SLPA) [36–38]. Первоначально каждой вершине КИС U_S присваивается уникальная метка. Далее начинается итерационный процесс (*количество итераций определяется пользователем*), в котором каждая вершина принимает метку (или некоторое множество меток) по согласованию с большинством ее соседей. По окончанию этого процесса множества вершин с одной и той же меткой, определяющие связные подграфы, объявляются сообществами. Достоинство SLPA − простота реализации и быстрое исполнение (*по этой причине SLPA часто используют в качестве составляющей части при построении различных методов выделения структур сообществ в реальных КИС* [39–44]). Основной недостаток SLPA − различные их реализации приводят к различным решениям, некоторые из которых имеют очень плохое качество (*это вызвано тем, что методы SLPA ориентированы на достижение некоторого локального минимума*). В [45] показано, что SLPA эквивалентны применению упрощенной модели Поттса (*модель Поттса используется в статистической механике для описания взаимодействия спинов в кристаллической решетке*) для выделения структур сообществ в сети [46].

3. Методы, основанные на хешировании, чувствительном к местоположению (LSH) [47–51]. Основаны на уменьшении размерности данных за счет такого подбора хэш-функций, чтобы похожие объекты с высокой вероятностью попадали в одну и ту же корзину. Вычисления состоят в следующем. Заменим КИС U_S взвешенным направленным графом, где $w(v_j,v_l)\in[0,1]$ $(v_j,v_l\in V;v_j\neq v_l)$ − вес того, что дуга идет из вершины v_j в вершину v_l ($\sum\limits_{v_l\in V\setminus\{v_j\}}w(v_j,v_l)=1$ для всех $v_j\in V$). Далее многократно применяется следующая схема.

Генерируем n случайных перестановок $\pi_1,\ldots,\pi_n$ множества V. Для каждой вершины $v_j\in V$ вычислим n значений $mh_1,\ldots,mh_n$, где mh_i $(i=1,\ldots,n)$ − это элемент множества $N_j=\{v_j\}\cup\{v_l\,|\,w(v_j,v_l)>0\}$, с наименьшим индексом в перестановке π_i. Выберем случайным образом r $(r<n)$ попарно различных чисел $l_1,\ldots,l_r\in\mathbf{N}_n$. Сопоставим с каждой вершиной сигнатуру, являющуюся конкатенацией значений $mh_{l_1},\ldots,mh_{l_r}$. Хешируем вершины на основе их сигнатур. Вершины с одним и тем же хэш-значением помещаются в одно сообщество (*вероятность того, что вершины* v_j *и* v_l *согласованы на этой сигнатуре, равна* $\left(\frac{|N_j\cap N_l|}{|N_j\cup N_l|}\right)^r$, *так как вероятность того, что* v_j *и* v_l *согласованы по одному значению (т. е. индекс Жаккарда) равна* $\frac{|N_j\cap N_l|}{|N_j\cup N_l|}$).

4. Методы, использующие функцию локальной плотности (*функция локальной плотности представляет собой один из вариантов функции полезности*) [52–55]. Их суть состоит в следующем. В соответствии с выбранным критерием каждой вершине сети приписывается ранг. Последовательным удалением вершин (в направлении от максимального ранга к минимальному рангу) формируются попарно непересекающиеся ядра предполагаемых сообществ. Каждое ядро последовательно преобразуется за счет добавления или удаления вершин, обеспечивающего не убывание значения функции локальной плотности. Эта процедура выполняется до тех пор, пока невозможно осуществить дальнейшее увеличение значения этой функции. В [52] использовалась функция локальной плотности $f(C)=\frac{w_{in}^{(C)}}{w_{in}^{(C)}+w_{out}^{(C)}}$, где $w_{in}^{(C)}$ и $w_{out}^{(C)}$, соответственно, внутренний и внешний вес сообщества C (*в частности, в качестве внутреннего веса* $w_{in}^{(C)}$ *и внешнего веса* $w_{out}^{(C)}$ *могут быть выбраны суммы, соответственно, внутренних и внешних степеней вершин, принадлежащих сообществу* C). Ее недостаток в том, что могут порождаться сообщества, являющиеся несвязными подграфами (*так как в процессе построения сообщества допускается удаление вершин*). Для его устранения в [53] было предложено использовать функцию $f(C)=\frac{w_{in}^{(C)}}{w_{in}^{(C)}+w_{out}^{(C)}}+\lambda e_p$, где $e_p=\frac{2w_{in}^{(C)}}{|C|(|C|-1)}$ − реберная вероятность (*реберная вероятность отражает уровень связанности в сообществе* C), а λ − параметр, и при построении сообщества сохранять только одну компоненту, имеющую максимальную плотность. В [54] использовалась функция $f(C)=\frac{k_{in}^{(C)}}{(k_{in}^{(C)}+k_{out}^{(C)})^{\alpha}}$, где α − параметр, управляющий размером сообществ. При этом после нахождения очередного сообщества случайным образом выбиралась вершина, не принадлежащая ни одному из построенных сообществ, которая использовалась для построения нового сообщества. Предложенная в [55] функция $f(C)=\frac{k_{in}^{(C)}+1}{(k_{in}^{(C)}+k_{out}^{(C)})^{\alpha}}$ дает возможность рассматривать одновершинные сообщества, а также ускорить процесс объединения сообществ при их построении.

1.1.3. Выделение сообществ на основе статистического вывода. В этих методах используют стохастические блочные модели (SBM) (*эти модели являются Байесовскими моделями для графов* [56, 57]) в предположении, что известно число l сообществ КИС U_S (*основная предпосылка этой модели состоит в том, что вершины, принадлежащие одному и тому же сообществу стохастически эквивалентны, т. е. для всех вершин в сообществе одинаковы вероятности их связей со всеми другими вершинами*). Основная задача − максимизировать значение функции правдоподобия. Рассмотрим основные SBM.

SBM Бернулли применяется для выделения непересекающихся сообществ. В этом случае осуществляется вывод вектора $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_n)$ ($\alpha_i\in\{1,\ldots,l\}$ $(i\in V)$ − номер сообщества, содержащего вершину i) на основе $l\times l$-матрицы P, где p_{ij} $(i,j\in\{1,\ldots,l\})$ − вероятность связи между сообществами i и j. Функция правдоподобия имеет вид $L(U_S\,|\,\boldsymbol{\alpha},P)=\prod_{i<j}p_{\alpha_i\alpha_j}^{a_{ij}}(1-p_{\alpha_i\alpha_j})^{1-a_{ij}}$ (*таким образом, предполагается, что каждое ребро генерируется независимо, а величины* a_{ij} *распределены по закону Бернулли, где* $\mathrm{P}(a_{ij}=1)=p_{\alpha_i\alpha_j}$).

SBM Пуассона (*эту модель часто рассматривают как базовую SBM*) применяется для выделения непересекающихся сообществ. Предполагается, что величины a_{ij} распределены по закону Пуассона, т. е. $a_{ij}\sim\mathrm{Poi}(\omega_{\alpha_i\alpha_j})$ (*таким образом,* $l\times l$*-матрица* P *заменяется* $l\times l$*-матрицей* Ω [58]). Для любых сообществ r и s оценка максимального правдоподобия величины ω_{rs} имеет вид $\hat{\omega}_{rs}=\frac{m_{rs}}{n_r n_s}$, где m_{rs} − это число ребер между сообществами r и s, если $r\neq s$, и удвоенное число ребер внутри сообщества r, если $r=s$, а n_r − число вершин в сообществе r. Функция log-правдоподобия имеет вид $\log L(U_S\,|\,\boldsymbol{\alpha})=0.5\sum_{r,s}m_{rs}\log\frac{m_{rs}}{n_r n_s}$ [59].

В SBM Пуассона вершины сети, степени которых существенно отличаются друг от друга, имеют мало шансов попасть в одно и то же сообщество. Поэтому эта модель не дает возможность адекватно выделять сообщества в сетях, для которых характерна большая вариация распределения степеней вершин в сообществах. Был разработан ряд SBM, допускающих неоднородности степеней вершин в пределах сообщества. Наибольшее распространение получила DC SBM [58]. Для нее функция log-правдоподобия имеет вид $\log L(U_S\,|\,\boldsymbol{\alpha})=0.5\sum_{r,s}m_{rs}\log\frac{m_{rs}}{k_r k_s}$, где k_r − сумма степеней вершин в сообществе r.

Существуют SBM предназначенные для выделения пересекающихся сообществ в сетях. Наиболее известной из них является MMSB [56]. В этой модели вычисления осуществляются следующим образом. Вначале с использованием распределения Дирихле каждой вершине v ставится в соответствие вектор длины l, определяющий распределение этой вершины по сообществам. Далее инициируется итерационный процесс стохастической оптимизации (число итераций определяется пользователем), преобразующий структуру сообществ в сети на основе вариационного вывода методом самосогласованного поля для приближенного вычисления апостериорных распределений принадлежности вершин сообществам.

1.1.4. Нечеткие методы выделения сообществ. В них первоначально для каждой вершины рассчитывается "мягкий" вектор ее принадлежности (*этот вектор является стохастическим вектором, т. е. его компоненты неотрицательны и их сумма равна единице* [62]) сообществам [60–62]. При этом естественно возникает проблема определения

размерности l этого вектора (*это значение l может быть выбрано либо в виде параметра метода, либо определено на основе анализа данных*), что существенно ограничивает успешное применение нечетких методов.

В [63] выделение пересекающихся сообществ сети сводится к решению нелинейной оптимизационной задачи (с ограничениями). Для этого предполагается, что задана непрерывная дифференцируемая функция подобия (similarity function) вершин исследуемой сети, а для решения задачи используется градиентный метод, т. е. вычисления заведомо направлены на поиск локального минимума.

Несмотря на то, что разработке нечетких методов выделения сообществ в сетях в последнее время уделяется значительное внимание, основной вопрос, на который пока нет удовлетворительного ответа, состоит в следующем: как сравнивать между собой четкие и нечеткие методы выделения сообществ в сетях.

1.2. Анализ социальных сетей. Проблема выделения сообществ пользователей в КИС значительно сложнее и многограннее, чем выделение сообществ на основе топологии КИС. Это обусловлено следующими причинами.

Понятие социальной сети как общественного явления сформировалось в 1-й половине XX века, а элементы сетевого анализа впервые были использованы именно в классической социологии в середине этого же века (*благодаря этим исследованиям под социальной сетью стали понимать "набор социально-релевантных узлов, связанных одним или несколькими отношениями", что дало возможность основным социологическим понятиям "актор" и "взаимодействие" поставить в соответствие термин теории графов, а именно: вершину и ребро. Операционализация основных социологических понятий в терминах теории графов и использование широкого спектра математических методов привело к созданию методологии исследования социальных сетей, обладающей мощными инструментами анализа данных и визуализации результатов*).

Работы в области формальной социологии привели к пониманию того, что представление объекта исследования в сетевом виде требует структурирования описывающих его переменных. Эта проблема была исследована в [64]. Были выделены следующие четыре основных типа свойств (иными словами, характеристик) члена коллектива (актора): 1) *абсолютные*, т. е. полученные без учета свойств коллектива или взаимоотношений в нем; 2) *относительные*, т. е. установленные путем обработки информации об отношениях между данным членом коллектива и другими его членами; 3) *сравнительные*, т. е. полученные при сопоставлении характеристики члена коллектива с ее средним значением по группе; 4) *контекстуальные*, т. е. выведенные из свойств самого коллектива. Были выделены следующие типы свойств (характеристик) группы (коллектива): 1) *аналитические*, т. е. основанные на данных о каждом члене группы; 2) *структурные*, т. е. основанные на сведениях об отношениях между членами группы; 3) *глобальные*, т. е. основанные на коллективных проявлениях.

Формирование формальной социологии в значительной мере осуществлялось в рамках общей теории идентификации [65]. Именно с этих позиций были выявлены следующие основные моменты. Во-первых, люди присоединяются к группам исходя из их идентичности (т. е. на основании интереса к обсуждаемым темам) или из социальных отношений (bond attachment). Во-вторых, поддержание кругов общения в явном виде громоздко и занимает слишком много времени. В-третьих, для многих людей по разным причинам неудобно явно объявлять свое членство в тех или иных группах. В-четвертых, некоторые типы групп (географические, образовательные, профессиональные и т. д.) основаны на сходстве атрибутов и не могут быть полностью восстановлены из-за неполноты информации об участниках. В-пятых, построение структуры неявных сообществ сети в явном виде является ценной информацией для различных аналитических центров и служб (именно это обстоятельство, во многом, стимулировало разработку методов кластеризации графов). В-шестых, одной из основных характеристик отношений между узлами социальной сети является "сила связи", определяемая частотой, интенсивностью, регулярностью и устойчивостью связей (поэтому можно выделить сильные, слабые и латентные связи).

На этом фоне в 90-х годах XX века и появились первые онлайновые сервисы социальных сетей (Facebook, Twitter, YouTube и т. д.). Практика показала, что методы анализа социальных сетей (*методы анализа социальных сетей используется для исследования взаимодействий между участниками сети, прогнозирования их поведения, классификации, моделирования информационных потоков в сетях*) могут успешно применяться при анализе онлайн-взаимодействий [66–69], что обусловлено следующими обстоятельствами. Онлайн-взаимодействия всегда имеют сетевую структуру (*узлами сети могут быть люди, организации, вэб-страницы, публикации, страны и т. д., а связями − потоки информации, различные виды ресурсов, взаимодействие, социальные отношения и т. д. Таким образом, классификации участников сети, анализ их взаимодействий и прогнозирование их поведения играют важную роль при решении исследовательских и бизнес-задач, для борьбы с сетевыми мошенничествами, отмыванием денег, кибератаками, а также для создания вспомогательных сервисов и приложений для пользователей социальных сетей*), так как электронная коммуникация − сеть, состоящая из отправителей и получателей.

Представления о коммуникационных системах и отношениях были сформированы задолго до возникновения Интернета. Однако до сих пор сбор личных данных − сложная и громоздкая процедура, занимающая много времени. Кроме того, часто люди часто сами не понимают, кто находится в их личной сети или насколько сильна та или иная связь, а исследователю необходимо собрать точные данные о взаимодействиях. Эти сложности могут быть снижены за счёт онлайн-исследований, так как в них информация является цифровой и закодированной через акт отправки сообщения или добавления друга с помощью функционала Интернет-страницы. В электронных социальных сетях также несложно копировать сообщения для их дальнейшего анализа. Отметим, что источниками данных о структуре социальной сети могут служить API (application programming interface − интерфейс программирования приложений) популярных социальных сетей (Facebook, Twitter, Linkedin), такие универсальные децентрализованные семантические сети, как Giant Global Graph на RDF2 протоколах, электронные данные, имеющиеся у государства, а также закрытые электронные данные банков и частных компаний.

Под онлайновой социальной сетью понимают Интернет-сервис, позволяющий пользователям публиковать на своих страницах ту или иную информацию, и служащий для упрощения коммуникации и обмена информацией между пользователями сети Интернет. Выделяют следующие типы онлайновых социальных сетей: 1) общего назначения (ВКонтакте, Мой Мир, Одноклассники, Facebook, и т. д.), предназначенные для поддержания существующих контактов из реального мира, обсуждения событий и развлечений; 2) контентные (Twitter, YouTube, и т. д.), предназначенные для обмена контентом, распространения новостей, создания и развития сообществ по интересам и раз-

влечений; 3) тематические (LinkedIn, Academia.edu, ResearchGate и т. д.), предназначенные для поддержания существующих и установления новых профессиональных контактов, дискуссий по общим интересам и решения профессиональных вопросов; 4) остальные (FourSquare − сеть с функцией геопозицирования, Instagram − сеть с упором на выкладывание пользоватем фотографий и т. д.).

Несмотря на многообразие онлайновых социальных сетей, в них естественно выделяются следующие типы данных: 1) сетевые данные, т. е. отношения связи между пользователями, а также между пользователями и объектами; 2) профили пользователей, т. е. их CV, биографии, взгляды, интересы и т. д.; 3) текстовые данные, т. е. сообщения, комментарии и т. д.; 4) мультимедийные данные, т. е. фото-, аудио- и видео-материалы; 5) объекты уровня сети, т. е. группы, сообщества, приложения и т. д.; 6) объекты внешнего мира, т. е. ссылки на ресурсы, расположенные за пределами сети; 7) журналы активности пользователей, т. е. записи о взаимодействии пользователей между собой и с различными объектами. Таким образом, онлайновая социальная сеть представляет собой граф (*часто для онлайновых социальных сетей адекватным является представление размеченным ориентированным, либо направленным мультиграфом с петлями* [70]. *Некоторые онлайновые социальные сети естественно представлять гиперграфами (например, взаимодействие типа "назначение тега фотографии" включает пользователя, фотографию и тег)*), у которого вершины и ребра снабжены некоторыми наборами атрибутов.

Выделяют следующие четыре подхода к анализу онлайновых социальных сетей [71, 72]:

1. Структурный подход. Участники онлайновой социальной сети представляются вершинами графа, влияющими на конфигурацию ребер и других участников сети. Исследуется геометрическая форма сети и интенсивность взаимодействий (вес ребер), т. е. такие характеристики, как взаимное расположение вершин, центральность и транзитивность взаимодействий. Для интерпретации результатов применяются структурные теории и теории сетевого обмена.

2. Ресурсный подход. Исследуются возможности участников онлайновой социальной сети по привлечению индивидуальных и сетевых ресурсов (*индивидуальными ресурсами могут быть знания, престиж, богатство, раса, пол и т. д., а сетевыми ресурсами − влияние, статус, информация, капитал и т. д.*) для достижения поставленных целей. Осуществляется дифференциация участников, находящихся в идентичных структурных позициях социальной сети, в соответствии с их ресурсами.

3. Нормативный подход. Исследуется уровень доверия между участниками онлайновой социальной сети, нормы, правила и санкции, влияющие на их поведение и процессы их взаимодействий. Анализируются социальные роли, связанные с данным ребром сети (*например, отношения руководителя и подчиненного, дружеские или родственные связи и т. д.*). Комбинация индивидуальных и сетевых ресурсов участника с нормами и правилами, действующими в данной социальной сети, образует его "сетевой капитал", т. е. преимущества, которые участник может получить в произвольный момент времени для достижения поставленной цели.

4. Динамический подход. Исследуются изменения в структуре онлайновой социальной сети во времени, т. е. влияние внешних и внутренних воздействий на структуру сети, выделение тех или иных стационарных конфигураций в сети и т. д.

В [72, 73] эти подходы охарактеризованы в терминах следующих актуальных для анализа онлайновых социальных сетей задач и методов их решения.

1. Статистический анализ (*решение этой задачи для обычных социальных сетей было получено в 70-е годы XX столетия* (*см., напр.,* [74])). На основе использования статистических методов исследуется изменение во времени основных характеристик типичных онлайновых социальных сетей, т. е. структуры, поведения и т. д. Обзор аналитических моделей, используемых в процессе событийного моделирования онлайновых социальных сетей содержится в [75].

2. Выделение сообществ. Модели и методы, рассмотренные в п.1.1, адаптируются для онлайновых социальных сетей с учетом следующих обстоятельств:

а) вершина может принадлежать нескольким сообществам с разной степенью принадлежности, т. е. осуществляется поиск нечетких кластеров (fuzzy clusters) [33, 60, 76, 77];

б) структура сообществ может быть иерархической [54, 78–80];

в) в предположении истинности гипотезы случайного распределения ребер при заданных значениях степеней вершин необходимо отсекать "псевдосообщества" на основе статистической значимости конфигурации [77, 81].

3. Классификация вершин. В онлайновой социальной сети вершины, у которых выделенные атрибуты удовлетворяют заданным условиям, снабжаются метками, которые в соответствии с тем или иным отношением толерантности (и, возможно, при наличии дополнительных ограничений) распространяются по сети (*такой подход используется в маркетинговых исследованиях заинтересованности участников сети в конкретном продукте* [82, 83], *в экспертных методах решения крупных проблем для выявления экспертов в группах специалистов в конкретных узких областях, а также при анализе текстовой* [84, 85] *и мультимедийной* [86] *информации в сети*).

4. Анализ социального влияния. Исследуется распространение по онлайновой социальной сети той или иной информации, влияющей на действия ее пользователей [87–90]. Для решения этой задачи чаще всего используются методы, получившие имя "вирусный маркетинг". Их суть состоит в том, чтобы главным распространителем информации являлись ее адресаты (*это достигается путем формирования содержания, способного привлечь новых получателей информации за счет яркой, творческой, необычной идеи или с использованием естественного или доверительного послания*). В результате информация распространяется по сети по закону, близкому к геометрической прогрессии. Следует отметить, что в последнее время у государственных органов различных стран вызывает большую озабоченность усовершенствованная разновидность вирусного маркетинга, получившая имя "террористический маркетинг". Такие методы используют современные медийные средства для того, чтобы достаточно быстро и эффективно осуществить подрывную деятельность небольшой инициативной группой лиц, внедренной в онлайновую сеть для воздействия на целевую аудиторию.

5. Визуализация социальных сетей. Предназначена для наглядного представления структуры больших онлайновых социальных сетей в статическом и динамическом режимах [91]. Основная сложность состоит в разработке алгоритмов, сочетающих методы анализа сети и ее визуализации, обеспечивающих достижение поставленной цели. В настоящее время для визуализации и анализа данных в онлайновых социальных сетях создан ряд инстру-

ментальных средств. Одной из наиболее известных таких систем является SWIFT-3D [92], созданная в AT&T Labs. Эта система объединяет набор интерактивных инструментов, который включает пиксель-ориентированные 2D карты, интерактивные 3D-карты, статистические отображения, диаграммы сетевой топологии и интерактивный интерфейс запросов, организованный по принципу сверху-вниз.

Отметим, что при исследовании реальных онлайновых социальных сетей приходится решать рассмотренные выше задачи в их совокупности. Поэтому дополнительно возникает проблема согласования между собой используемых моделей, структур данных и методов решения этих задач.

2. Проектирование КИС. Наряду с решением обычных проблем, возникающих при проектировании любой IT-системы (*обзор этих проблем представлен, например, в* [93–95]), возникают следующие проблемы, вызванные именно наличием сетевой структуры в КИС [96].

2.1. Формулировка требований к проектируемой КИС. Хотя требования существенно зависят от назначения КИС (*например, для КИС, используемых в цифровой экономике, типичны следующие требования: 1) сеть должна быть в рабочем состоянии все время, даже в случае несостоявшихся соединений, отказов оборудования и при условиях ее перегрузки; 2) сеть должна надежно доставлять приложения и обеспечить приемлемые сроки реагирования между любыми хостами; 3) сеть должна быть безопасной, т. е. должны быть защищены данные, которые передаются по сети и данные, хранящиеся на устройствах, которые подключаются к сети; 4) сеть должна легко адаптироваться к ее росту и общим изменениям бизнеса; 5) так как время от времени возникают неисправности в сети, то их поиск, локализация и устранение не должны занимать слишком много времени*), они направлены на достижение следующих четырех целей: 1) масштабируемость, т. е. возможность роста сети за счет подключения новых групп пользователей и удаленных объектов, поддержка новых приложений без снижения уровня обслуживания существующих пользователей; 2) доступность, т. е. обеспечение стабильного и надежного непрерывного функционирования, причем отказ одного звена или оборудования не должен существенно влиять на производительность сети; 3) безопасность, т. е. планирование местоположения устройств безопасности, фильтров, функций межсетевого экрана, а также формулирование требований к криптографической защите информации и системе доступа к сети; 4) управляемость, т. е. имеющийся в наличии персонал должен быть в состоянии управлять сетью и поддерживать ее (*сеть, которую слишком сложно поддерживать, не сможет функционировать эффективно и результативно*).

2.2. Выбор структурной организации проектируемой КИС. Современные КИС имеют иерархическую структурную организацию. Так как увеличение количества уровней упрощает процесс построения сети, но ухудшает временные характеристики, то целесообразно уменьшать количество уровней при одновременном использовании виртуальных каналов связи.

В настоящее время при проектировании КИС базовой считается 3-х уровневая иерархическая структурная организация, в которой основной уровень обеспечивает подключение к устройствам уровня распределения, уровень распределения соединяет локальные сети, а уровень доступа обеспечивает подключение сетевых хостов и концевых устройств. Для дальнейшего разделения такой структурной в модульных областях достаточно успешно используется методология Cisco Enterprise Architecture Model (*Cisco Systems − один из мировых лидеров по разработке и поставке на рынок сетевого оборудования*) [97, 98].

2.3. Разработка сетевой топологии и архитектуры КИС. Является сложной комбинаторной многофакторной оптимизационной проблемой, решение которой существенно зависит от типов подсетей (т. е. кабельных, беспроводных, оптических), входящих в КИС. Рассмотрим кратко ее основные составляющие.

Известно, что для кабельных сетей минимизация стоимости соединений и переключателей в сети при наличии ограничений на ее производительность является NP-трудной задачей [7]. Поэтому на практике для ее решения часто используют эвристические методы: табу-поиск [99], имитацию отжига [100], генетические алгоритмы [101], методы, основанные на муравьиной колонии [102]. Для реальных кабельных КИС ситуация усложняется еще и тем, что стоимости соединений и переключателей зависят от топологии сети, т. е. не являются фиксированными ни вначале, ни в процессе оптимизации. Такая ситуация исследована в [103], где на основе локального поиска и геометрического подхода предложены пять методов решения этой задачи начиная от исчерпывающего поиска и заканчивая быстрыми эвристиками. Приведен сравнительный анализ этих методов на основе результатов проведенных экспериментов.

Для беспроводных сетей (кроме геометрического размещения объектов) возникает необходимость назначения таких определенных стандартами дефицитных параметров, как мощность и частота с целью максимизации количества охваченных приемников или доходов, связанных с их охватом. Известно, что (см., напр., [104]) эта задача является NP-трудной. В [105] предложен генетический алгоритм решения этой задачи. При этом в дополнение к эмиссии мощности и частоты в качестве параметра, значение которого устанавливается в результате решения задачи, рассматривается схема передачи (взрыв профиля) (*таким образом, для беспроводной КИС решается задача назначения мощности, частоты и модуляции*). Это обстоятельство дает возможность применить предложенный в [105] метод к таким беспроводным сетям нового поколения, как WiMAX [106].

Для оптоволоконных сетей естественно выделяются следующие две задачи: 1) проектирование физической топологии и конфигурации сети; 2) маршрутизация и присвоение длин волн. Проектирование физической топологии состоит в определение числа и связности оптических кросс-соединений (*современные оптические кросс-соединения осуществляют мультиплексирование разделения длины волны, т. е. длина волны оптического сигнала на входном оптоволокне может быть переключена на любую длину волны в выходном оптоволокне. Протоколы управления такими кросс-соединениями представлены, например, в* [107, 108]), предназначенных для обеспечения непрерывной цепи прохождения луча света при наличии большого числа отмеченных коммутирующих маршрутизаторов. Проектирование конфигурации сети состоит в определении размеров оптических кросс-соединений, количества оптоволоконных соединений и множества путей прохождения света. Маршрутизация включает в себя отображение путей прохождения света на физическую топологію (*таким образом, мы приходим к виртуальной топологии, в которой узлы соответствуют фактическим физическим сетевым узлам, а связи соответствуют путям прохождения света*). Различные формулировки и решения задач маршрутизации и присваивания длин волн вдоль маршрутов в предположении, что зафиксирована оптоволоконная топология, представлены в [109]. Даже при фиксированной физической топологии проектирование конфигурации сети является NP-трудной задачей (*эта задача сводится к задаче целочисленного программирования*), а

для ее решения, как правило, применяется метод ветвей и границ. При фиксированном множестве маршрутов задача присваивания длин волн вдоль маршрутов эквивалентна задаче раскраски графа. В этом случае для ее решения применяют эвристические методы [110]. В [111] предложен генетический алгоритм для решения задачи маршрутизации и присвоения длин волн в предположении, что зафиксирована физическая топология. В [112] предложены эвристические методы решения задачи минимизации количества оптических кросс-соединений при заданных ограничениях на количество длин волн в волоконно-оптической линии, в предположении, что физическая топология оптоволоконной сети является двусвязной.

Сложной задачей является размещение средств, позволяющих обеспечить эффективное функционирование КИС при больших объемах сетевого трафика и большом количестве устройств, подключенных к сети. Для этой цели предназначены центры обработки данных (ЦОД), которые используются для хранения и обработки огромного количества данных. В настоящее время проектирование ЦОД становится одним из наиболее востребованных направлений (*по прогнозу аналитиков, к 2018 году объем мирового рынка ЦОД вырастет до 35 млрд долларов, причем 40% всех расходов на КИС будут так или иначе связаны с ЦОД. В первую очередь ЦОД будут востребованы сервис-провайдерами, облачными коммерческими центрами обработки данных, крупными корпоративными КИС*), и требует существенных изменений в подходах к построению (*понятие ЦОД приводит к пересмотру сетевой архитектуры, а именно: к отделению управления от передачи данных, а также к автоматизации процесса администрирования сетевого оборудования*), эксплуатации и управлению КИС. При построении ЦОД основной топологией является fat-tree (*fat-tree (утолщенное дерево) − дерево, в котором связи становятся более производительными по пропускной способности с каждым уровнем по мере приближения к корню дерева (часто используют удвоение пропускной способности на каждом уровне)*) [113–115]. Однако сама по себе эта топология не направлена на синтез оптимальных ЦОД (***качество ЦОД оценивается по совокупности параметров:*** *стоимость-задержка-энергопотребление*). В [116] предложена архитектура Diamond для ЦОД, представляющая собой модификацию fat-tree. В этой модификации улучшение качества ЦОД достигается за счет того, что все агрегированные переключатели в fat-tree заменяются переключателями связей, которые непосредственно подсоединяются к базовым переключателям. Такой подход дает возможность снизить в среднем на 10% задержку между концевыми вершинами. В [117] показано, каким образом при проектировании ЦОД может быть использована dragon fly [118] топология, с тем, чтобы сократить задержки в сети за счет уменьшения количеств переходов между узлами, и общую стоимость ЦОД за счет уменьшения количества оптических кабелей, используемых для подключения подсетей.

Большие усилия исследователей посвящены разработке методологии проектирования распределенных КИС (см., напр., [119–121]). В этом направлении особое место занимают исследования пиринговых сетей (P2P nets) [122–124], представляющих собой оверлейные компьютерные сети, основанные на равноправии участников. Часто в такой сети отсутствуют выделенные серверы, а каждый узел является как клиентом, так и выполняет функции сервера. Такая организация позволяет сохранять работоспособность сети при любом количестве и любом сочетании доступных узлов. По-видимому, одним из наиболее перспективных вариантов пиринговой сети является сеть Gnutella [121, 126, 127]. Она представляет собой полностью децентрализованную файлообменную сеть в рамках Интернета, в которой отсутствует центральный сервер. Сеть Gnutella формируется, когда один ее пользователь соединяется с другим пользователем, после чего может происходить обмен информацией (*обмениваться можно файлами любых форматов, для этого предусмотрено большое количество удобных функций*) между ними. В результате полной децентрализации сеть Gnutella практически невозможно уничтожить (так как для этого необходимо вывести из строя каждый ее узел).

2.4. Эмуляция КИС. Является одной из ключевых технологий в исследованиях компьютерных сетей. Дает возможность проводить эксперименты, которые являются более реалистичными, чем моделирование, и более дешевыми, чем эксперименты на сетях, построенных на основе аппаратных средств (*кроме того, сети, реализованные аппаратно, трудно перенастраивать, что существенно ограничивает возможности крупномасштабных экспериментов, в которых необходимы несколько различных тестовых вариантов сети*). Существует ряд инструментальных средств, предназначенных для эмуляции КИС (например, EINAR [128], VNUML [129], Netkit [130, 131]). Хотя эмуляция устраняет необходимость физического размещения соединительных проводов, она требует конфигурирования таких устройств, как маршрутизаторы и коммутаторы. Известно, что конфигурирование маршрутизаторов − сложная задача [132, 133], а ее решение вручную − источник "человеческих ошибок", которые трудно выявить. Поэтому разработка средств автоконфигурации маршрутизаторов при эмулировании КИС является актуальной задачей. В [134] предложен один из вариантов решения этой задачи − инструментальное средство автоконфигурации маршрутизаторов AutoNetkit, предназначенное для эмуляции КИС с использованием инструментальных средств Netkit.

2.5. Управление КИС. Представляет собой сложную многогранную проблему. Рассмотрим кратко некоторые ее основные составляющие.

Основные принципы конструирования Интернет-трафика, его онлайн и оффлайн технические возможности, системы его поддержки и варианты его развития систематически изложены в [135, 136]. В [137] исследована задача конструирования трафика в рамках одной автономной системы (такой как фирма, университетский городок, провайдер интернет услуг и т. д.). Показано, как адаптировать конфигурацию весов связей с позиции трафика всей сети и топологии внутри домена. В [138] предложен метод быстрого вывода матриц трафика в IP-сетях на основе измерений нагрузки связей, дополненных легкодоступной информацией о сетевой конфигурации и маршрутизации. Эффективность предложенного метода проиллюстрирована его применением к вычислению матриц трафика между магистральными маршрутизаторами достаточно большой Tier-1 IP сети (*Tier-1 IP сеть − сеть, в которой оператор имеет доступ к сети Интернет исключительно через пиринговые соединения*). В [139] содержится обзор состояния инжиниринга Интернет трафика с позиции оптимизации маршрутизации. Соответствующие алгоритмы классифицированы в следующих измерениях: одноадресный / многоадресный, внутри-доменный / меж-доменный, IP / MPLS, оффлайн схема / онлайн схема. Охарактеризована надежность Интернет трафика и его совместимость с оверлейной маршрутизацией. Отмечены сложности, возникающие при применении инжиниринга Интернет трафика и некоторые проблемы, требующие дальнейшего исследования.

В п. 2.3 отмечено, что в настоящее время большое внимание уделяется разработке ЦОД. Однако известно мало результатов, связанных с анализом характеристик трафика сетевого уровня современных ЦОД. В [140] исследованы сетевые трафики десяти ЦОД различных типов организаций: университет, предприятие и облако. При этом в определение ЦОД облака кроме ЦОД крупных провайдеров, предлагающих интернет-приложения, включены также ЦОД, используемые для приложений с интенсивной обработкой данных (*иными словами, приложения, написанные с помощью MapReduce*). Была проанализирована статистика протокола сетевого управления, топология и следы на уровне пакетов. Развернутые в исследуемых ЦОД приложения рассмотрены с позиции их размещения, уровня потока данных, передачи на уровне следов пакетов, влияния на использование сети и связей в ней, загруженности сети и потери пакетов. Охарактеризовано влияние установленных закономерностей на инжиниринг внутреннего трафика, а также на архитектуру современных ЦОД. В [141] представлена система MicroTE, функционирующая поверх основных топологий ЦОД, предназначенная для смягчения последствий перегруженности сети, вызванной непредсказуемым трафиком, за счет использования частично предсказуемой матрицы краткосрочного трафика (*таким образом, система MicroTE адаптирует сеть к изменениям трафика*). Система MicroTE реализована рамках OpenFlow с незначительными изменениями в концевых хостах. Показано, что система MicroTE реализует решение, близкое к оптимальному решению, при минимальных накладных расходах на сеть, что делает его подходящим для существующих и будущих ЦОД. Обзор существующих схем передачи потоков данных в ЦОД содержится в [142]. Отмечены преимущества и недостатки этих схем. Обосновано, почему некоторые версии транспортных протоколов не могут быть эффективными в ЦОД.

Значительное число реализованных в сетях алгоритмов и сетевых протоколов включают в себя механизмы распространения информации (*например, при маршрутизации в сети для пересылки сообщения осуществляется поиск адресата и отслеживание маршрута в его направлении*). Исследования, направленные на разработку механизмов перераспределения трафика в интернете из мест, менее релевантных запросу в места с большей релевантностью по этому запросу получили имя поисковый маркетинг. Эффективность механизма распространения информации в КИС оценивается по совокупности параметров: число сообщений и время завершения. В [143] исследованы две основные составляющие механизма распространения информации в КИС − поиск и маркетинг. Предложены критерии, характеризующие правильность, быстроту и честность (fairness) применения этого механизма.

Программно-конфигурируемые сети (SDN) передачи данных [144, 145] являются одной из форм виртуализации вычислительных ресурсов. Они характеризуется тем, что в них уровень управления отделён от устройств передачи данных и реализуется программно. В [146] содержится обзор современного состояния исследований SDN. Основное внимание уделено анализу управления потоками данных, отказоустойчивости сетей, обновлению топологии и средствам анализа трафика.

Эффективность КИС в значительной мере зависит от количества и расположения сервисных центров, размещенных на различных хостах. В [147] предложен новый подход к добавлению или удалению серверов в КИС на основе использования только локальной информации о топологии сети и запросах. Показано, что в течение одной-двух итераций достигается производительность КИС, сопоставимая с применением централизованных подходов, требующих полную информацию о топологии и запросах всей КИС.

3. Безопасность КИС. Является одной из ключевых проблем современных сетевых технологий. Рассмотрим кратко некоторые ее основные составляющие.

Практика показала, что к масштабным отказам, сбоям и разрушению маршрутизации КИС могут привести как стихийные бедствия [148–150], так и преднамеренные нападения [151–153]. Большинство существующих подходов к восстановлению КИС являются, по своей сути, проактивными и используют резервные предварительно вычисленные пути для перенаправления трафиков, пострадавших от неисправностей, во время конвергенции IGP (протокол внутреннего шлюза) [154–156]. Такие проактивные подходы построены на неявном предположении о том, что в КИС возникают только спорадическиие изолированные неисправности связей. По этой причине эти подходы не применимы для крупномасштабных неисправностей, так как связи и резервные пути могут быть разрушены одновременно. Кроме того, ни один маршрутизатор не имеет всю информацию о неисправностях; он только знает, достижимы ли его соседи. Для недостижимого соседа маршрутизатор не может различить ситуации: неисправен ли сосед или соединяющая его связь. В [157] решена задача быстрого восстановления маршрутизации внутри домена при крупных авариях. Идея предложенного подхода состоит в следующем. Вначале осуществляется сбор информации об отказе с помощью разработанного протокола пересылки специальных пакетов вокруг области неисправности и записи информации о неисправности в заголовке пакета. Затем вычисляются кратчайшие исправные пути, по которым направляются пакеты. В [158] исследуется эксплуатационная практика, стандарты и текущие исследования в области безопасности междоменной маршрутизации, сходства и различия в существующих подходах к построению более безопасной инфраструктуры Интернет.

Для обеспечения безопасности КИС большое значение имеет защита от нападений, осуществляемых на базе Интернета. Эта проблема особенно остро стоит перед корпоративными КИС. Одной из составляющих для ее решения является развертывание набора продуктов мониторинга безопасности, которые генерируют результаты "ситуационной разведки" в форме различных журналов. Такие журналы, как правило, содержат большие объемы информации о деятельности в сети, и являются одними из первых источников данных, которые анализируют специалисты по информационной безопасности, когда они подозревают, что произошла атака на сеть. Продукты мониторинга безопасности часто поступают от различных поставщиков, устанавливаются и администрируются несогласованно друг с другом. Как следствие, форматы генерируемых журналов отличаются друг от друга, журналы часто являются неполными, взаимно противоречивыми и очень большими по объему. В результате, собранная информация, хотя и является полезной, часто беспорядочна и содержит много "мусора". В [159] представлена система Beehive, осуществляющая автоматическое извлечение знаний из "мусорных данных" журналов, созданных для широкого спектра продуктов мониторинга безопасности корпоративных КИС. Эта система вместо выявления подозрительного поведения узла сообщает о потенциальных инцидентах, связанных с безопасностью. Далее инциденты могут быть проанализированы группами реагирования, чтобы определить, произошло ли нарушение политики безопасности, либо атака на сеть. На основе проведенных экспериментов уста-

новлено, что система Beehive в состоянии идентифицировать злонамеренные действия и нарушения политики безопасности, которые в противном случае остались бы незамеченными.

Известно, что в онлайновых социальных сетях существует достаточно много рисков, связанных с нарушением конфиденциальности и безопасности (см., напр., [160]). Эти риски возникают из-за неявного доверия, встроенного в заявленных социальных отношениях, что дает возможность для сбора личной информации о пользователях для тех или иных целей. Отметим, что в настоящее время в онлайновых социальных сетях принято выставлять информацию о пользователе, собранную из различных социальных сфер (например, личную информацию из Facebook, профессиональную деятельность из LinkedIn и т. д.), что приводит к излишне детализированным профилям [161]. Используя такую доступность информации, фирмы выискивают на Facebook и Twitter тенденции для создания вирусного контента для акций и лайков, работодатели проверяют на Facebook, LinkedIn и Twitter профили кандидатов на должность [162], правоохранительные органы подбирают в онлайновых социальных сетях доказательства при раскрытии преступлений [163]. Кроме того деятельность в онлайн социальных платформах используется для изменения политических режимов [164] и для влияния на результаты выборов [165]. Пресечение масштабных вползаний (crawls) пользовательских профилей в такие онлайновые социальные сети, как Facebook и Renren отвечает интересам как пользователей, так операторов этих сайтов. Пользователи стремятся сохранить контроль над своей личной информацией, а операторы − защитить свои активы и бизнес-репутацию. Существующие методы ограничения скорости неэффективны против краулеров (*Краулер − программа, являющаяся составной частью поисковой системы и предназначенная для перебора страниц Интернета с целью занесения информации о них в базу данных поисковика. По принципу действия краулер напоминает обычный браузер. Он анализирует содержимое страницы, сохраняет его в некотором специальном виде на сервере поисковой машины, которой принадлежит, и отправляется по ссылкам на следующие страницы. Владельцы поисковых машин нередко ограничивают глубину проникновения краулера внутрь сайта и максимальный размер сканируемого текста, поэтому слишком большие сайты могут оказаться не полностью проиндексированными поисковой машиной*) с большим количеством аккаунтов, будь то фальшивые аккаунты (Sybils) или взломанные аккаунты реальных пользователей, полученные на черном рынке. В [166] представлена система Genie, предназначенная для защиты от краулеров в крупномасштабных онлайновых социальных сетях. Эта система использует тот фактор, что шаблоны просмотра честными пользователями и краулерами сильно отличаются друг от друга: даже краулеру с доступом ко многим аккаунтам необходимо просмотреть намного больше профилей на каждый аккаунт, чем честному пользователю. Краулер также просматривает профили пользователей, которые сильно отдалены друг от друга в социальной сети. Эксперименты с использованием реальных данных, собранных из популярных крупномасштабных онлайновых социальных сетей показали, что Genie расстраивает масштабные вползания и редко оказывает влияние на честных пользователей; пострадавших пользователей можно легко восстановить, добавив несколько ссылок о дружбе.

Критическую угрозу для сетевой безопасности представляют различные вредоносные программы (malware). Современные модели распространения вредоносных программ в КИС делятся на два класса. К первому классу относятся модели, построенные на основе теории управляющих систем [167]. На их основе разрабатываются методы обнаружения и ограничения распространения вредоносных программ. Ко второму классу относятся модели, построенные на основе эпидемиологии [168, 169]. Эти модели предназначены для исследования законов распределения количества зараженных хостов. Отметим, что на сегодняшний день нет полного понимания поведения вредоносных программ в КИС. Поэтому исследование законов их распространения в КИС является актуальной задачей. В [170] с глобальной точки зрения исследуется, каким образом вредоносные программы распространяются в КИС. Новизна предложенного подхода по сравнению с традиционным подходом, основанным на использовании эпидемиологических моделей, состоит в том, что анализ нарушений в модели осуществляется в соответствии со следующими двумя этапами. Вначале на основе восприимчивой к инфицированию модели [171] вычисляется количество сетей, скомпрометированных в течение заданного времени с момента прорыва КИС вредоносной программой. Далее для взломанной сети рассчитывается, сколько хостов было скомпрометировано с момента нарушения сети. Доказано, что распределение вредоносной программы в КИС подчиняется экспоненциальному закону на ранней стадии, степенному распределению с коротким экспоненциальным хвостом на поздней стадии, и сходится к степенному распределению. Достоверность теоретических результатов проиллюстрирована их применением к реальным данным. В [172] представлена система Nazca, предназначенная для обнаружения инфицирования КИС. Эта система выявляет явные признаки сетевых инфраструктур, которые руководят инсталяцией вредоносных программ. Процесс выявления этих признаков осуществляется на основе анализа коллективного трафика, производимого большим числом пользователей КИС (*именно такой подход обеспечивает системе Nazca большую скорость и эффективность по сравнению с системами, основанными на анализе соединений, свойств загруженных программ и репутации серверов*). На систему Nazca не влияет полнота зон покрытия пробелов в репутационных базах (т. е. черных списках). Кроме того, эта система не поддается на запутывание кода. Эффективность системы Nazca была проиллюстрирована результатами ее работы в течение семи дней с крупным Интернет провайдером. Были обнаружены ранее не выявленные вредоносные программы, причем число ложных срабатываний системы оказалось небольшим.

Заключение. В работе кратко рассмотрены некоторые актуальные проблемы в следующих направлениях, связанных с анализом и синтезом КИС: выделение сообществ на основе топологии сети, анализ социальных сетей, проектирование КИС, обеспечение безопасности КИС. Следует отметить, что существующие в настоящее время подходы к решению этих проблем весьма далеки от того, что принято называть проработанной технологией.

За рамками данного обзора осталось много актуальных проблем в области анализа и синтеза КИС, которые в настоящее время находятся только на стадии исследования, и анализ ситуации с каждой из которых является темой для соответствующего аналитического обзора. К ним, в частности, относится весь комплекс проблем анализа и синтеза сенсорных сетей, проблемы эффективной и безопасной организации облачных вычислений, проблемы эффективной криптографической защиты информации в КИС с критической областью применения и т. д.

В заключение отметим, что большинство проблем анализа и синтеза КИС настолько тесно взаимосвязаны, что попытки обеспечить оптимальное решение каждой из проблем при ее изолированном исследовании из-за возможных противоречий между этими решениями приведут к созданию КИС, которая окажется неэффективной и недостаточно защищенной.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.

1. Girvan M., Newman M. Community structure in social and biological networks // Proc. of the National Acad. of Sci. of the USA. – 2002. – Vol. 99. − P. 7821–7826.
2. Radicchi F., Castellano C., Cecconi F., at al. Defining and identifying communities in networks // Proc. of the National Acad. of Sci. of the USA. – 2004. – Vol. 101. − P. 2658–2663.
3. Hastings M. Community detection as an inference problem. − http://arxiv.org/pdf/cond-mat/0604429.pdf
4. Buzun N., Korshunov A. Innovative methods and measures in overlapping community detection // Proc. of International Workshop on Experimental Economics in Machine Learning. − 2012. − P. 20–32.
5. Yang J., Leskovec J. Defining and evaluating network communities based on ground-truth // Proc. of IEEE 12th International Conference on Data Mining. − 2012. − P. 745–754.
6. Chykhradze K., Korshunov A., Buzun N., at al. Distributed generation of billion-node social graphs with overlapping community structure // Proc. of Complex Networks V. Studies in Computational Intelligence. − 2014. − Vol. 549. − P. 199–208.
7. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. – М.: Мир, 1982. – 416 с.
8. Tomita E, Tanaka A, Takahashi H. The worst-case time complexity for generating all maximal cliques // Computing and Combinatorics. − Berlin/Heidelberg: Springer. − 2004. − P. 161–170.
9. Fortunato S., Castellano C. Community structure in graphs. − http://arxiv.org/abs/0712.2716.pdf
10. Xie J., Kelly S., Szymanski B. Overlapping community detection in networks: the state of the art and comparative study. − http://dx.doi.org/10.1145/2501654.2501657
11. Harenberg S., Bello G., Gjeltema L., at al. Community detection in large-scale networks: a survey and empirical evaluation // WIREs Comput. Statistics. – 2014. − Vol. 6. − P. 426–439.
12. Newman M. Modularity and community structure in networks. − http://www.pnas.org/content/103/23/8577.full
13. Newman M., Girvan M. Finding and evaluating community structure in networks. –http://link.aps.org/abstract/PRE/v69/e026113
14. Ziv E., Middendorf M., Wiggins C. Information-theoretic approach to network modularity. – http://journals.aps.org/pre/abstract/10.1103/PhysRevE.71.046117
15. Friggeri A., Chelius G., Fleury E. Triangles to capture social cohesion. – http://arxiv.org/pdf/1107.3231.pdf
16. Arenas A, Fernandez A., Fortunato S., Gomez S. Motif-based communities in complex networks. – http://arxiv.org/pdf/0710.0059.pdf
17. Clauset A., Newman M., Moore C. Finding community structure in very large networks. – http://arxiv.org/abs/cond-mat/0408187
18. Newman M. Fast algorithm for detecting community structure in networks. – http://arxiv.org/abs/cond-mat/0309508
19. Pons P., Latapy M. Computing communities in large networks using random walks // Proc. of the 20th International Conference on Computer and Information Sciences. − 2005. – P. 284–293.
20. De Meo P., Ferrara E., Fiumara G., Provetti A. Mixing local and global information for community detection in large networks // Journal of Computing Systems Sciences − 2014. − Vol. 80. −P:72–87.
21. White S., Smyth P. A spectral clustering approach to finding communities in graph // Proc. of SIAM Data Mining Conference. − 2005. − P. 76–84.
22. Newman M. Finding community structure in networks using the eigenvectors of matrices. – http://arxiv.org/abs/physics/0605087
23. Kirkpatrick S., Gelatt C.D., Vecchi M.P. Optimization by simulated annealing // Science. − 1983. − Vol. 220. − P. 671–680.
24. Guimera R., Sales-Pardo M., Amaral L. Modularity from fluctuations in random graphs and complex networks. − http://dx.doi.org/10.1103/PhysRevE.70.025101
25. Liu J., Liu T. Detecting community structure in complex networks using simulated annealing with k-means algorithms // Physica A.: Statistical Mechanics and its Applications. – 2010. – Vol. 389. – № 11. – P. 2300–2309.
26. Штовба С. Муравьиные алгоритмы // Exponenta Pro. Математика в приложениях. − 2003. – №4. − С. 70–75.
27. Dorigo M., Stützle T. Ant colony optimization: overview and recent advances // IRIDIA Technical Report Series. − 2009. − Technical Report № TR/IRIDIA/2009-013. − 32 p.
28. Лебедев О.Б. Разбиение на основе муравьиной колонии // Материалы XV Международной конференции по нейрокибернетике. Т. 2. − Ростов-на-Дону: Изд-во ЮФУ. − 2009. − С. 102–105.
29. Liu Y., Luo J., Yang H., Liu L. Finding closely communicating community based on ant colony clustering model // Proc. of the International Conference on Artificial Intelligence and Computational Intelligence. − 2010. − P. 127–131.
30. Sadi S., Oguducu S., Uyar A. An efficient community detection method using parallel clique-finding ants // Proc. of IEEE Congress on Computational Intelligence. − 2010. − P. 1–7.
31. Jin D., Liu D., Yang B., at al. Ant colony optimization with markov random walk for clustering in complex networks // Proc. of the 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining. − 2011. − P. 123–134.
32. Palla G., Dereny I., Farkas I., Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society // Nature. − 2005. − Vol. 435. – P. 814–818.
33. Lee C., Reid F., McDaid A., Hurley N. Detecting highly overlapping community structure by greedy clique expansion // Proc. of SNAKDD Workshop. − 2010. − P. 33–42.
34. Gregori E., Lenzini L., Orsini C. k-clique communities in the Internet AS-level topology graph // Proc. of ICDCS Workshops. − 2011. − P. 134–139.
35. Gregori E., Lenzini L., Mainardi S. Parallel k-clique community detection on large-scale networks // IEEE Trans. on Parallel and Distributed Systems. – 2012. – № 8. − P. 1651–1660.
36. Bagrow J., Bollt E. A local method for detecting communities. – arXiv:cond-mat/0412482v2
37. Raghavan U., Albert R., Kumara S. Near linear-time algorithm to detect community structures in large-scale networks. – http://arxiv.org/abs/0709.2938v1
38. Costa L. Hub-based community finding. – http://arxiv.org/abs/cond-mat/0405022v1
39. Xie J., Szymanski B. Towards linear time overlapping community detection in social networks // Proc. of PAKDD. − 2012. − P. 25–36.
40. Xie J., Szymanski B. Community detection using a neighborhood strength driven label propagation algorithm. – http://arxiv.org/pdf/1105.3264
41. Barber M. Detecting network communities by propagating labels under constraints. –http://arxiv.org/abs/0903.3138
42. Leung I., Hui P., Li P., Crowcroft J. Towards real-time community detection in large networks. – http://arxiv.org/pdf/0808.2633.pdf
43. Gregory S. Finding overlapping communities in networks by label propagation. – http://arxiv.org/pdf/0910.5516
44. Prabavathi G., Thiagarasu V. Design and development of overlapping community detection algorithm using multi-level propagation. – http://www.ijarcsms.com/February2014.htm
45. Tibely G., Kertesz J. On the equivalence of the label propagation method of community detection and a potts model approach // Physica A. − 2008. − Vol. 387. – P. 4982–4984.
46. Kumpula J., Saramaki J., Kaski K., Kertesz J. Limited resolution in complex network community detection with potts model approach // European Physical Journal B. − 2007. − Vol. 56. − P. 41–45.
47. Macropol K., Singh A. Scalable discovery of best clusters on large graphs // Proc. of VLDB Endowment. − 2010. − № 3. − P. 693–702.
48. Satuluri V., Parthasarathy S., Ruan Y. Local graph sparsification for scalable clustering // Proc. of the International Conference on Management of Data. – 2011. – P. 721–732.
49. Song H. Clustered embedding of massive social networks // ACM SIGMETRICS Performance Evaluation Review. – 2012. – №. 1. – P. 331–342.
50. Tantawi A. A scalable algorithm for placement of virtual clusters in large data centers // Proc. of MASCOTS. − 2012. − P. 3–10.
51. Sui X., Lee T., Whang J., at al. Parallel clustered low-rank approximation of graphs and its application to link prediction // Languages and Compilers for Parallel Computing. – Berlin/Heidelberg: Springer. − 2013. – P. 76–95.
52. Baumes J., Goldberg M., Krishnamoorthy M., at al. Finding communities by clustering a graph into overlapping subgraphs // Proc. of the IADIS International Conference on Applied Computing. − 2005. − P. 97–104.
53. Kelley S. The existence and discovery of overlapping communities in large-scale networks. − Ph.D. thesis, Rensselaer Polytechnic Institute, Troy, NY. − 2009.
54. Lanchinettii A., Fortunato S., Kertesz J. Detecting the overlapping and hierarchical community structure of complex networks. New J. Phys. − 2009. − Vol. 11. − № 3. − 20 p.
55. Havemann F., Heinz M., Struck A., Glaser J. Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels. − 2011. − J. Statist. Mech. − № 1. − PO1023.
56. Airoldi E., Blei D., Fienberg S., Xing E. Mixed membership stochastic blockmodels // J. Mach. Learn. Res. − 2008. − № 9. − P. 1981–2014.
57. Wasserman S., Anderson C. Stochastic a posteriori blockmodels: Construction and assessment // Social Networks. − 1987. − № 1. − P. 1–36.
58. Karrer B., Newman M.. Stochastic blockmodels and community structure in networks. − http://dx.doi.org/10.1103/PhysRevE.83.016107
59. Bickel P., Chen A. A nonparametric view of network models and Newman–Girvan and other modularities // Proc. of the National Acad. of Sci. of the USA. – 2009. – Vol. 106. − P. 21068–21073.

60. Zhang S., Wang R., Zhang X. Identification of overlapping community structure in complex networks using fuzzy c-means clustering // Physica A. − 2007. − Vol. 374: −P. 483–490.
61. Gregory S. Fuzzy overlapping communities in networks. – http://arxiv.org/abs/1010.1523
62. Liu J. Fuzzy modularity and fuzzy community structure in networks // The European Physical Journal B. – 2010. − № 4. – P. 547–557.
63. Nepusz T., Petróczi A., Négyessy L., Bazsó F. Fuzzy communities and the concept of bridgeness in complex networks. – http://arxiv.org/abs/0707.1646v3
64. Lasarsfeld P.F. On Social Research and Its Language. − University of Chicago Press. − 1993. − 342 p.
65. Prentice D. A., Miller D. T., Lightdale J. R. Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond groups // Personality and Social Psychology Bulletin. − 1994. − № 5. − P. 484–493.
66. Garton L., Haythornthwaite C., Wellman B. Studying online social networks // Journal of Computer Mediated Communication. − 1997. − Vol. 3. − № 1. − P. 75–106.
67. Yang B., Liu D., Liu J. Discovering communities from social networks: methodologies and applications // Handbook of Social Network Technologies and Applications. − Berlin/Heidelberg: Springer, 2010. − P. 331–346.
68. Tang L., Liu H. Community detection and mining in social media // Synthesis Lectures on Data Mining and Knowledge Discovery. − 2010. − № 1. − P. 1–137.
69. Martin A., Wellman B. Social Network Analysis: An Introduction // Handbook of Social Network Analysis / ed. by P. Carrington, J. Scott, 2011. − Thousand Oaks, CA: Sage. − P. 11–25.
70. Bollobas B. Modern graph theory. – NY: Springer-Verlag, 1998. – 394 p.
71. Чураков А. Н. Анализ социальных сетей // СоцИс. − 2001. − № 1. − С. 109–121.
72. Батура Т.В. Методы анализа компьютерных социальных сетей // Вестник НГУ. − 2012. − Т.12. − Вып. 4. − С. 13–28.
73. Social network data analytics / Ed. Charu C. Aggarwal. − Springer US: Science+Business Media, LLC, 2011. − eBook ISBN 978-1-4419-8462-3. − 520 p.
74. Granovetter M. S. The strength of weak ties // American Journal of Sociology. − 1973. − № 6. − P. 1360–1380.
75. Scheidegger M., Baumgartner F., BraunT. Simulating large-scale networks with analytical models // International Journal of Simulation. − 2005. − Vol.6. − № 1–2. − P. 24–31.
76. Gregory S. An algorithm to find overlapping community structure in networks. − https://www.cs.bris.ac.uk/~steve
77. Lancichinetti A., Radicchi F., Ramasco J., Fortunato S. Finding statistically significant communities in networks. − http://santo.fortunato.googlepages.com/inthepress2
78. Clauset A., Moore C., Newman M. Hierarchical structure and the prediction of missing links in networks // Nature. − 2008. − Vol. 453. − P. 98–101.
79. Ahn Y., Bagrow J., Lehmann S. Link communities reveal multi-scale complexity in networks // Nature. − 2010. − Vol. 466. − P. 761–764.
80. Rosvall M., Bergstrom C. Maps of random walks on complex networks reveal community structure // Proc. of the National Acad. of Sci. of the USA. − 2008. − Vol. 105. − P. 1118–1123.
81. McDaid A., Hurley N. Using model-based overlapping seed expansion to detect highly overlapping community structure. − http://sites.google.com/site/aaronmcdaid/moses
82. Wang X., Sun J., Chen Z., Zhai C. Latent semantic analysis for multiple-type interrelated data objects // Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06). − 2006. − NY: ACM. − P. 236–243.
83. Stern D., Herbrich R., Graepe T. Matchbox: large scale online bayesian recommendations // Proceedings of the 18th International Conference on World Wide Web (WWW '09). − 2006. − NY: ACM. − P. 111–120.
84. Машечкин Д. В. Петровский И. В., Царев М. И. Методы вычисления релевантности фрагментов текста на основе тематических моделей в задаче автоматического аннотирования // Вычислительные методы и программирование. – 2013. – Т. 14. – С. 91–102.
85. Коршунов А. В. Гомзин А. Г. Тематическое моделирование текстов на естественном языке // Труды Института системного программирования РАН. – 2012. – Вып. 23. – С. 216–243.
86. Давыдов А. А. Системная социология: анализ мультимедийной информации в Интернете. − http://www.isras.ru/index.php?page_id=988
87. Прокофьев В. Ф. Тайное оружие информационной войны: атака на подсознание. − М.: СИНТЕГ, 2003. − 408 с.
88. Freeman L. The development of social network analysis: a study in the sociology of science. − Vancouver: Empirical Press, 2004. − 208 p.
89. Губанов Д. А., Новиков Д. А., Чхартишвили А. Г. Модели информационного влияния и информационного управления в социальных сетях // Проблемы управления. − 2009. − № 5. − С. 28–35.
90. Губанов Д. А., Новиков Д. А., Чхартишвили А. Г. Социальные сети: модели информационного влияния, управления и противоборства. − М.: Изд-во физико-математической литературы, 2010. − 228 с.
91. Ware C. Information visualization: perception for design. Waltham, MA: Morgan Kaufmann. − 2013. − 512 p.
92. Koutsofio E., North S., Truscott R., Keim D. Visualizing Large-Scale Telecommunication Networks and Services // Proc. of IEEE Visualization. − 1999. − P. 457-461.
93. Скобелев В. Г. Безопасность IT систем (обзор) // Радіоелектронні і комп'ютерні системи. – 2013. – № 5. − С. 352–361.
94. Буй Д. Б., Скобелев В. Г. Безопасность программных средств: модели и методы (обзор) // Радіоелектронні і комп'ютерні системи. – 2014. – № 1. − С. 42–54.
95. Буй Д. Б., Скобелев В. Г. Модели, методы и алгоритмы оптимизации запросов в базах данных (обзор) // Радіоелектронні і комп'ютерні системи. – 2014. – № 1. − С. 43–58.
96. Ретана А., Слайс Д., Уайт Р. Принципы проектирования корпоративных IP-сетей. − М.: Вильямс, 2002. − 368 с.
97. http://www.netakademija.rs/pdf/ccna%20r&s/04.Connecting%20Networks/CN_instructorPPT_Chapter1_final.pdf
98. Hutton K., Schofield M., Teare D. Authorized self-study guide: designing Cisco network service architectures (ARCH), 2009. − Indianapolis, IN: Cisco Press. − 672 p.
99. Glover F., Laguna M. Tabu search // Modern heuristic techniques for combinatorial problems / Ed. C.R. Reeves. − 1993. − NY: John Wiley & Sons, Inc. − P. 70–150.
100. Aarts E., Lenstra J. Local search in combinatorial optimization. − NY: John Wiley & Sons, Inc., 1997. − 488 p.
101. Winter G., Periaux J., Galan M. Genetic Algorithms in Engineering and Computer Science. − NY: John Wiley & Sons, Inc., 1995. − 195 p.
102. Dorigo M., Di Caro G., Gambardella L. Ant Algorithms for Discrete Optimization // Artificial Life. − 1999. − № 5. − P. 137–172.
103. Grout V., Cunningham S., Picking R. Practical large-scale network design with variable costs for links and switches // International Journal of Computer Science and Network Security. − 2007. − № 7. − P. 113–125.
104. Mannino C., Rossi F., Smriglio S. The network packing problem in terrestrial broadcasting // Operation Research. − 2006. − № 6. − P. 611-626.
105. D'Andreagiovanni F. On improving the capacity of solving large-scale wireless network design problems by genetic algorithms // LNCS. − 2011. − Vol. 6625. P. 11–20.
106. Andrews J., Ghosh A., Muhamed R. Fundamentals of WiMAX: understanding broadband wireless networking. − Upper Saddle River: Prentice Hall, 2007. − 448 p.
107. Su D., Griffith D. Standards activities for MPLS over WDM networks // Optical Networks. − 2000. − № 3. − P. 6–69.
108. CISCO MPLS Web Page. − http://www.cisco.com/warp/public/732/Tech/mpls/
109. Karasan E., Ayanoglu E. Effects of wavelength routing and selection algorithms on wavelength conversion gain in WDM optical networks // IEEE Journal on Selected Areas of Communications. − 1998. − Vol. 16. − P. 1081–1096.
110. Zhang Z., Acampora A. A heuristic wavelength assignment algorithm for multihop WDM networks with wavelength routing and wavelength reuse // IEEE/ACM Trans. Networking. − 1995. − Vol. 3. − P. 281–288.
111. Saha D., Purkayastha M., Mukherjee B. An approach to wide area WDM optical network design using genetic algorithms // Computer Communications. − 1999. − Vol. 22. − P. 156–172.
112. Xin Y., Rouskas G., Perros H. On the physical and logical topology design of large-scale optical networks // Journal of Lightwave Technology. − 2003. − № 4 − P. 904–915.
113. InfiniBand in the Enterprise Data Center: Scaling 10Gb/s Clustering at Wire-Speed. White Paper, Mellanox Technologies, 2006. − http://www.mellanox.com/pdf/whitepapers/InfiniBand__EDS.pdf
114. Al-Fares M., Loukissas A., Vahdat A. A Scalable, commodity datacenter network architecture // ACM SIGCOMM Computer Communication Review. − 2008. − № 4. − P. 63–74.
115. Mysore R., Pamboris A., Farrington N., at al. A scalable fault-tolerant layer 2 data center network fabric // ACM SIGCOMM Computer Communication Review. − 2009. − № 4. − P. 39–50.
116. Sun Y., Chen J., Liu Q., Fang W. Diamond: an improved fat-tree architecture for large-scale data centers // Journal of Communications. − 2014. − № 1. − P. 91–98.

117. Delimitrou C., Mohammadi M., Nothaft F., Sharpless L. Datacenter network design: performance/power comparison of large-scale network configurations and a way to avoid it! − https://stanford.edu/~milad/DC_Network_v1.pdf
118. Kim J., Dally W., Scott S., Abts D. Technology-driven, highly-scalable dragonfly topology // Proc. of the International Symposium on Computer Architecture (ISCA). − 2008. − P. 77–88.
119. Menasce D., Gomaa H., Kerschberg L. A performance oriented design methodology for large-scale distributed data intensive information systems // Proc. of The First IEEE International Conference on Engineering of Complex Computer Systems. − 1995. − P. 72–79.
120. Kerschberg L., Weishar D. Conceptual Models and Architectures for Advanced Information Systems // Applied Intelligence. − 2000. − Vol. 13. − P. 149–164.
121. Risse T. Design and configuration of distributed job processing systems. − http://tuprints.ulb.tu-darmstadt.de/665/1/thomas_risse_diss.pdf
122. Saroiu S., Gummadi P., Gribble S. A measurement study of peer-to-peer file sharing systems // University of Washington Technical Report UW-CSE-01-06-02, July 2001.
123. Karagiannis T., Broido A., Brownlee N., at al. File-sharing in the Internet: A characterization of P2P traffic in the backbone − http://www.cs.ucr.edu/-tkarag
124. Karagiannis T., Broido A., Faloutsos M. Transport layer identification of P2P traffic // Proc. of the 4th ACM SIGCOMM Conference on Internet Measurement. − 2004. − P.121–134.
125. Adar E., Huberman D. Free riding on Gnutella. − http://firstmonday.org/ojs/index.php/fm/article/view/792
126. DSS Group, Gnutella: To the bandwidth barrier and beyond. − http://dss.clip2.com
127. Ripeanu M., Foster I., Iamnitchi A. Mapping the Gnutella network: properties of large-scale peer-to-peer systems and implications for system design // IEEE Internet Computing Journal. − 2002. − № 1. − P. 50–57.
128. EINAR. Einar router simulator. − http://www.isk.kth.se/proj/einar
129. Galan F., Fernadez D., Ruiz A., at al. Use of virtualization tools in computer network laboratories // Proc. International Conference on Information technology Based Higher Education and Training. − 2004. − P. 209–214.
130. University of Roma Tre. Computer Networks Research Group. Netkit.− http://www.netkit.org/
131. Rimondini M. Emulation of Computer Networks with Netkit. − Roma Tre University: Technical Report RT-DIA-113-2007.
132. Enck W., McDaniel P., Sen S., at al. Configuration management at massive scale: system design and experience // Proc. of the 2007 USENIX Annual Technical Conference. − 2007. − P. 1–14.
133. Bellovin S., Bush R. Configuration management and security // IEEE Journal on Selected Areas in Communications. − 2009. − № 3. − P. 268–274.
134. Nguyen H., Roughan M., Knight S., Falkner N., at. al. How to build complex, large-scale emulated networks. − https://nguyenxuanhung.files.wordpress.com/2014/07/autonetkit.pdf
135. Awduche D., Chiu A., Elvalid A., at al. Overview and principles of Internet traffic engineering. Internet-Draft: draft-ietf-tewg-principles-00.txt. − https://tools.ietf.org/html/draft-ietf-tewg-principles-00
136. Choi T., Yoon S., Chung H., at al. Design and implementation of traffic engineering server for a large-scale MPLS-based IP Network // LNCS. − 2002. − Vol. 2343. − P. 699–711.
137. Fortz B., Rexford J., Thorup M. Traffic engineering with traditional IP routing protocols // IEEE Communications Magazine. − 2002. − Vol. 40. − P. 118–124.
138. Zhang Y., Roughan M., Duffield N., Greenberg A. Fast accurate computation of large-scale IP traffic matrices from link loads // Proc. of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems. − 2003. − P. 206–217.
139. Wang N., Ho K., Pavlou G., at al. An overview of routing optimization for Internet traffic engineering // IEEE Communications Surveys & Tutorials. − 2008. − № 1. − P. 36–56.
140. Benson T., Akella A., Maltz D. Network traffic characteristics of data centers in the wild // Proc. of the Internet Measurement Conference. − 2010. − P. 267–280.
141. Benson T., Anand A., Akella A., Zhang M. MicroTE: Fine grained traffic engineering for data centers // Proc. of the Seventh Conference on Emerging Networking Experiments and Technologies. − 2011. − P. 1–12.
142. Rojas-Cessa R., Kaymak Y., Dong Z. Schemes for fast transmission of flows in data center networks // IEEE Communications Surveys & Tutorials. − 2015. − № 3. − P. 1391–1422.
143. Oikonomou K., Kogias D., Tzevelekas L., at al. Investigation of information dissemination design criteria in large-scale network environments // Proc. of the 13th Panhellenic Conference on Informatics. − 2009. − P. 163–167.
144. Openflow switch specification v1.0–v1.4. − https://www.opennetworking.org/sdn-resources/onf-specifications
145. Agarwal S., Kodialam M., Lakshman T. Traffic engineering in software defined networks // Proc. of the 32nd IEEE International Conference on Computer Communications. − 2013. − P. 2211–2219.
146. Akyildiz I., Lee A., Wang P., at al. A roadmap for traffic engineering in SDN-OpenFlow networks // Computer Networks. − 2014. − Vol. 71. − P. 1–30.
147. Lautarias N., Smarandakis G., Oikonomou K., at al. Distributed Placement of Service Facilities in Large-Scale Networks // Proc. of IEEE INFOCOM, 2007. − P. 2144–2152.
148. Neumayer S., Zussman G., Cohen R., Modiano E. Assessing the vulnerability of the fiber infrastructure to disasters // IEEE/ACM Transactions on Networking. − 2011. − № 6. − P. 1610–1623.
149. Neumayer S., Modiano E. Network reliability with geographically correlated failures // Proc. of the 29th Conference on Information Communications. − P. 1658–1666.
150. Ran Y. Considerations and suggestions on improvement of communication network disaster countermeasures after the wenchuan earthquake // IEEE Communications Magazine. − 2011. − № 1. − P. 44–47.
151. Ogielski A., Cowie J. Internet routing behavior on 9/11 and in the following weeks. − http://www.renesys.com/tech/presentations/pdf/renesys-030502-NRC-911.pdf
152. Bellovin S., Gansner E. Using link cuts to attack Internet routing. − http://www.cs.columbia.edu/smb/papers/reroute.pdf
153. Wilson C. High altitude electromagnetic pulse (HEMP) and high power microwave (HPM) devices: threat assessment. − http://www.fas.org/man/crs/RL32544.pdf
154. Kvalbein A., Hansen A., Cicic T., at al. Fast IP network recovery using multiple routing configurations. − http://folk.uio.no/amundk/infocom06.pdf
155. Kini S., Ramasubramanian S., Kvalbein A., Hansen A. Fast recovery from dual link failures in IP networks. − http://www2.engr.arizona.edu/~srini/papers/Srini-2009-INFOCOM.pdf
156. Wang Y., Wang H., A. Mahimkar A., at al. R3: Resilient routing reconfiguration. − http://cs-www.cs.yale.edu/homes/yry/projects/reinforce/r3-sigcomm10.pdf
157. Zheng Q., Cao G., La Porta T., Swami A. Optimal recovery from large-scale failures in IP networks // Proc. of the IEEE 32nd International Conference on Distributed Computing Systems. − 2012. − P. 295–304.
158. Butler K., Farley T., McDaniel P., Rexford J. A survey of BGP security issues and solutions // Proc. of the IEEE. − 2010. − № 1. − P. 100-122.
159. Yen T., Oprea A., Onarlioglu K., at al. Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks // Proc. of the 29th Annual Computer Security Applications Conference. − 2013. − P. 199–208.
160. Kayes I., Iamnitchi A. A Survey on Privacy and Security in Online Social Networks. − http://arxiv.org/abs/1504.03342
161. Nissenbaum H. A contextual approach to privacy online // Daedalus. − 2011. − № 4. − P. 32–48.
162. http://goo.gl/kHJFI5
163. http://www.cnn.com/2012/08/30/tech/social-media/fighting-crime-social-media/
164. Lotan G., Graeff E., Ananny M., at al. The arab spring − the revolutions were tweeted: information flows during the 2011 tunisian and egyptian revolutions // International Journal of Communication. − 2011. − № 5. − P. 1375–1405.
165. http://goo.gl/9A6FR
166. Mondal M., Viswanath B., Clement A., at al. Defending against large-scale crawls in online social networks // Proc. of the 8th International Conference on Emerging Networking Experiments and Technologies. − 2012. − P. 325–336.
167. Dantu R., Cangussu J., Patwardhan S. Fast worm containment using feedback control // IEEE Transactions on Dependable and Secure Computing. − 2007. − № 2. − P. 119–136.
168. Sellke S., Shroff N., Bagchi S. Modeling and automated containment of worms // IEEE Transactions on Dependable and Secure Computing. − 2008. − № 2. − P. 71–86.
169. Willinger W., Alderson D., Doyle J. Mathematics and the internet: a source of enormous confusion and great potential // Notices of the American Mathematical Society. 2009. − № 5. − P. 586–599.
170. Yu S., Gu G., Barnawi A., at al. Malware propagation in large-scale networks // IEEE Transactions on Knowledge & Data Engineering. − 2015. − № 1. − P. 170–179.

171. Zou C., Gong W., Towsley D., Gao L. The monitoring and early detection of internet worms // IEEE/ACM Transactions on Networking. − 2005. − № 5. − P. 961–974.
172. Invernizzi L., Miskovic S., Torres R., at al. Nazca: Detecting malware distribution in large-scale networks // Proc. of the IEEE Network and Distributed System Security Symposium. − 2014. − 16 p.

Стаття надійшла до редколегії 25.09.15

Скобелєв В. Г., д-р фіз.-мат. наук, д-р техн. наук, проф.
Інститут кібернетики імені В. М. Глушкова НАН України, Київ

ПРОБЛЕМИ АНАЛІЗУ ТА СИНТЕЗУ ВЕЛИКОМАСШТАБНИХ МЕРЕЖ (ОГЛЯД)

У даній роботі міститься огляд стану досліджень деяких актуальних проблем аналізу та синтезу великомасштабних інформаційних мереж. Детально розглянуті найбільш часто використовувані методи, що засновані на аналізі лише топології досліджуваної мережі та призначені для виділення в мережі або спільнот які не перетинаються, або спільнот які можуть перетинатися. Охарактеризовано основні підходи, моделі і методи, використовувані в процесі аналізу соціальних мереж. Виділено деякі актуальні проблеми, що виникають у процесі проектування великомасштабних інформаційних мереж, і коротко розглянуті існуючі підходи до їх вирішення. Охарактеризовано основні моделі і методи, застосовувані для забезпечення безпеки великомасштабних інформаційних мереж.

Ключові слова: *великомасштабні інформаційні мережі, аналіз, синтез, безпека, спільноти, онлайнові соціальні мережі.*

Skobelev V. G., Dr. Phys. Math. Sci., Dr. Tech. Sci., Prof.
V. M. Glushkov Institute of Cybernetics NAS of Ukraine, Kyiv

PROBLEMS OF ANALYSIS AND SYNTHESIS OF LARGE-SCALE NETWORKS (SURVEI)

Given paper consists some survey of the state of the art of research for some actual problems of analysis and synthesis of large-scale information networks. There are considered in detail the most frequently used methods based only on the analysis of the topology of the network and intended for extracting non-overlapping communities, as well as of overlapping communities. Main approaches, models and methods used in the analysis of social networks are described. There are extracted some actual problems arising in the design of large-scale information networks, and existing approaches to solving them are briefly discussed. Basic models and methods used to ensure the security and safety of large-scale information networks are characterized.

Key words: *large-scale information networks, analysis, synthesis, security, safety, communities, online social networks.*

УДК 512.7+512.9, 688. 321

Р. В. Скуратовський, асп.
інституту математики НАН України, Київ

МОДЕРНІЗОВАНИЙ АЛГОРИТМ ПОЛІГА-ХЕЛМАНА, ШЕНКСА

Не викликає сумніву, те що більшість з методів криптоаналізу можуть бути перевтілені завдяки застосуванню паралельних алгоритмів та алгебраїчного апарата, зокрема теорії груп. Одним з таких методів є метод Шенкса розв'язання ПДЛ. Ціллю данної роботи є побудова алгоритма, що параллельно знаходить всі значення з таблиць малого кроку і великого кроку, також зробити цей пошук більш спрямованим і впорядкованим для всіх значень елементів таблиць, що дозволить застосувати метод блокового пошука і дасть можливість розбиття на впорядковані підблоки, прискорить застосування метода індексації значень (чи хеш від значень). Методом є паралельна оптимізація і блочна параллельне поразрядне сортування, яка стала можливою завдяки швидким пересилкам в дуплексном режиме і математичні моделі алгоритму. В даній роботі запропоновано метод паралельного обчислення векторів координатами яких є значення таблички BS. Також знайдена оптимальна довжина малого кроку і як наслідок і великого кроку для методу. В роботі запропоновано метод покращення алгоритма Шенкса шляхом його композиції с методом Поліга-Хелмана.

Ключові слова: Алгоритм, метод Шенкса, криптоаналіз.

Вступ. За часів виникнення метода Шенкса (1973р.) для розв'язання проблеми дискретного логарифма його ефективність [1] була незначною з причини невисокої направленості перебору, який він використовує, і наявності великої кількості чисел для пошуку рівності. Це був один з перших методів, більш швидкий, ніж метод прямого перебору. Ним займались такі відомі криптографами, як Л. Адлеман і А. Стеін [2, 3]. Завдяки розвитку комп'ютерної техніки з'явилась можливість вдосконалення методу. В наш час, коли можлива паралельна обробка великих масивів інформації, його ефективність може зрости в стільки разів, скільки комп'ютерів ми застосовуємо. Це можливо саме завдяки придатності методу до розпаралелювання обчислень, які в ньому проводяться, але цьому не приділялось достатньої уваги відносно цього методу.

За основу для подальшої розробки взято класичний метод Шенкса [1]. Також порівняти затрати часу на пошук у невпорядкованому наборі масивів – табличок з методу Шенкса і у впорядкованих блоках, які відповідають тим же табличкам хіба дещо іншого розміру.

1. Постановка задачі. Мета роботи – створити алгоритм розв'язання проблеми дискретного логарифму (ПДЛ) із застосуванням паралельних обчислень і знайти його оптимальні параметри та оцінки складності обчислень.

Основні поняття. В роботі ми прагнемо зробити обчислення таблиць паралельним. Зменшення часу на пошук рівності досягаємо впорядкування елементів таблиць, що досягається швидкими пересилками; потім використовуємо пошук по лінійно впорядкованим множинам і по частково впорядкованим множинам. Також необхідно порівняти затрати часу на пошук у невпорядкованому наборі масивів – табличок з методу Шенкса і у впорядкованих блоках, які відповідають тим же табличкам, хіба дещо іншого розміру.

Розв'язання ПДЛ за Даніелем Шенксом [1, 2] спирається на пошук рівностей в наборах чисел. В циклічній мультиплікативній групі $\mathbb{C}_n$ обчислюються два рядка чисел: $(a, ga, g^2a, g^3a, ..., g^{m-1}a) \bmod n$ цей набір (таблицю) назвемо BS, $(g^m, g^{2m}, g^{3m}, g^{4m}, ..., g^{(m-1)m}) \bmod n$ цей набір (таблицю) назвемо GS.

Знаходять такі i та j, для яких $g^i a = g^{jm}$. Тоді в $g^x = a$ маємо $x = jm - i$.