

А.Г. КОЧАРЯН

ПРОГРАММНАЯ СИСТЕМА XRANDNET ДЛЯ ИЗУЧЕНИЯ
ТОПОЛОГИЧЕСКИХ ХАРАКТЕРИСТИК СЛУЧАЙНЫХ СЕТЕЙ

Активное изучение случайных сетей делает актуальным вопрос о разработке системы, которая способна имитировать поведение случайных сетей различных моделей (*Erdős-Rényi*, *Watts-Strogatz*, *Baraba'si-Albert*, *Regular Block-Hierarchical* и *Non Regular Block-Hierarchical*) и проводить анализ их топологических свойств. Разработанная система *xRandNet* направлена на эффективную имитацию именно блочно-иерархических сетей, которые являются достаточно новыми объектами в данной области.

Ключевые слова: случайные сети, блочно-иерархические сети, топологические характеристики.

Введение. В различных областях науки все чаще можно встретить так называемые сложные, комплексные системы, которые, как правило, нетривиальны и по своей структуре, и по характеру их изменений. Такими являются биологические системы (в частности, биополимеры – ДНК, РНК, белки), технические системы, социальные, экономические и т.д. [1]. На протяжении последних десятилетий сложные системы моделируются с помощью сетей, узлами (вершинами) которых являются элементы системы, а связями – взаимодействие этих элементов. В начале 1960-ых Пал Эрдеш и Альфред Реньи обратили внимание на то, что в изучении сложных систем можно использовать вероятностные методы, и основали теорию случайных сетей. Сеть называется случайной, если возникновение любой связи в ней подчиняется определенному вероятностному правилу. Далее будем рассматривать только ненаправленные (*non directed*) сети, в которых не допускается существование петель (*self-reference*) и кратных связей. Эрдеш и Реньи предложили модель случайной сети (классическая модель), для которой задаются число узлов N и вероятность p , определяющая наличие связи между любыми двумя узлами сети.

Изучение случайных сетей предполагает статистическое исследование различных топологических характеристик, таких как средняя длина пути, диаметр, распределение степени узлов, коэффициента кластеризации, циклов различной длины, расстояний между узлами и т.д.

Помимо изучения классических случайных сетей *Erdős-Rényi*, растет интерес и к новым классам сетей – “малым мирам” (*small-world*), для которых значение среднего расстояния между узлами по сравнению с размером сети очень малое; *безмасштабным* (*scale-free*), для которых характерно степенное поведение распределения степеней узлов и т.д. А в последнем десятилетии было предложено изучение принципиально иного класса случайных сетей - *блочно-иерархических*, с помощью которых, как оказалось, можно моделировать различные биологические структуры, начиная от белков до нейронных сетей [2,3].

Существующие программные системы (*GraphCrunch*, *mfinder*, *FANMOD*, *pajek*) позволяют проводить генерацию и анализ случайных сетей для стандартных моделей *Erdős-Rényi*, *Watts-Strogatz*, *Baraba'si-Albert* с числом узлов порядка 10^4 . В рассматриваемой системе *xRandNet*, помимо стандартных моделей, проводятся генерация и анализ сетей блочно-иерархического класса с числом вершин, превышающим 10^6 . Для моделей *Regular Block-Hierarchical* и *Non Regular Block-Hierarchical* разработаны новые алгоритмы, позволяющие довольно быстро проводить различные исследования. Оценка и описание этих алгоритмов даны в [4-6], а в данной статье приведены их времена выполнения. Помимо анализа по основным топологическим свойствам, включены конкретные виды исследований, которые описаны ниже. Особенностью системы является простое подключение новых моделей, для которых выполняются все имеющиеся исследования.

Необходимость исследования случайных сетей вообще и сетей блочно-иерархического класса, в частности, делает актуальными вопросы разработки инструмента, предназначенного имитировать генерацию случайных сетей различных моделей, вычисления топологических характеристик, выполнение различных исследований и проведения статистического анализа полученных результатов. Детальное описание генерации ансамбля случайных сетей разных моделей приведено в [7].

В данной статье представлены функциональность системы *xRandNet*, ее архитектура, детали реализации и иллюстрация ее работы.

Функциональность системы *xRandNet*. Данная система дает возможность генерировать ансамбли случайных сетей моделей *Erdős-Rényi*, *Watts-Strogatz*, *Baraba'si-Albert*, *Regular Block-Hierarchical* (далее *RBH*) и *Non Regular Block-Hierarchical* (далее *NRBH*), вычислять топологические характеристики для каждой сети в ансамбле, сохранять полученные результаты в хранилище данных и проводить статистический анализ этих данных.

Рассмотрим виды исследований, проведение которых поддерживается системой *xRandNet*.

Базовое исследование (Basic Research). Под базовым исследованием понимается генерация ансамбля случайных сетей конкретной модели и вычисление выбранного набора топологических характеристик этих сетей. Это могут быть как глобальные свойства, такие как средняя длина пути, диаметр, средний коэффициент кластеризации, так и локальные свойства - распределение степеней узлов, распределение коэффициентов кластеризации узлов и т.д. В рамках базового исследования можно также анализировать реальные сети, представленные в виде матрицы связей или списка смежностей в текстовом файле (рис. 1).

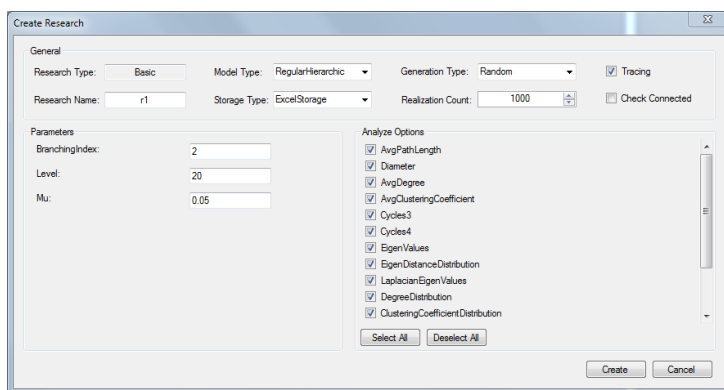


Рис. 1. Окно для создания базового исследования (*Basic Research*) в системе *xRandNet*

Исследование эволюции (Evolution Research). В данном исследовании рассматривается направленная эволюция случайных сетей по выбранному глобальному свойству при "замораживании" одной из степеней свободы. В системе *xRandNet* реализован процесс развития с увеличением числа циклов длиной 3 при сохранении числа связей или степеней вершин. Для семплирования использован алгоритм Метрополиса–Гастингса [8].

Исследование пороговой величины (Threshold Research). Для некоторой модели случайной сети вероятностным назовем параметр, с помощью которого определяется вероятность появления связей между узлами сети. Например, для модели *Erdős-Rényi* таким является параметр p , а для модели *RBH* – параметр плотности сети μ [4] и т.д. В теории случайных сетей одним из актуальных вопросов является нахождение порогового значения вероятностного параметра, при котором в сети появляется выбранное глобальное свойство Q . В системе *xRandNet* с этой целью реализовано исследование *Threshold Research*, для которого в качестве Q выбрано свойство связности сети. В процессе

исследования генерируются ансамбли сетей с изменением вероятностного параметра в данном диапазоне с шагом дельта. Для каждого ансамбля вычисляются размер максимальной связной компоненты, размер второй связной компоненты и средний размер всех остальных связных компонент, что позволяет изучать их зависимость от вероятностного параметра.

Для всех исследований результаты вычислений сохраняются в выбранном хранилище данных – XML файлы (*.xml), MS Excel файлы (*.xlsx) или MS SQL таблицы.

Статистический анализ. В системе *xRandNet* также предусмотрен инструмент для предварительного статистического анализа полученных результатов – *xRandNetStat (Extended Random Networks Statistics)*. Он имеет функциональность загрузки результатов из хранилища данных, группирования их по входным параметрам и получения соответствующих таблиц усредненных значений или графиков усредненных распределений. Для последних можно задавать аппроксимацию – степенную, экспоненциальную, гауссовскую, параметр сгущения для сглаживания графика и т.п. (рис. 2).

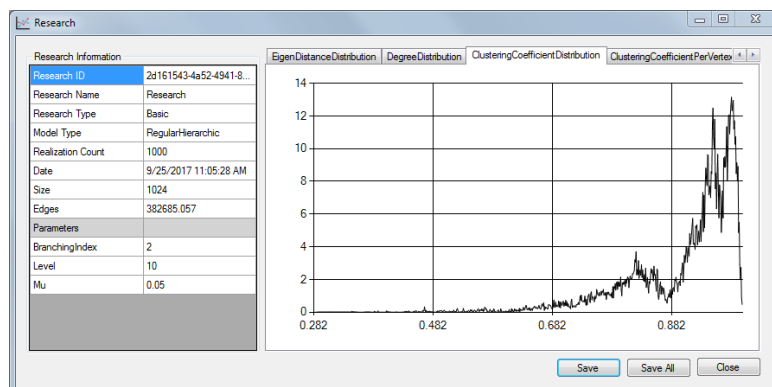


Рис. 2. Окно статистического анализа в системе *xRandNet*

Наличие такого инструмента позволяет предварительно фильтровать полученные результаты, выявляя особенности поведения для последующего детального анализа полученных данных с помощью более специализированных инструментов (таких как *Origin Pro*).

Архитектура системы *xRandNet*. Перечислим технические требования к разработке системы *xRandNet*:

1. *Эффективное использование памяти.* Это требование вытекает из необходимости изучения сетей с 10^6 узлами в рамках ансамблей из 1000 реализаций.

2. *Скорость вычисления топологических характеристик случайной сети.* Изучение случайных сетей подразумевает также возможность вычисления широкого набора топологических характеристик для каждой сети сгенерированного ансамбля. Оценки сложностей соответствующих классических алгоритмов вычисления, как правило, варьируются от $O(N)$ до $O(N^k)$ (k зависит от алгоритма) [9], что значительно затрудняет проведение исследований для больших сетей.

Данные проблемы в системе *xRandNet* решены для моделей RBH и NRBH путем разработки и реализации новой структуры хранения и новых оптимальных алгоритмов, оценки сложностей которых не превышают $O(N)$ [4-6].

Кроме чисто технических требований, перед системой *xRandNet* были поставлены ряд функциональных задач:

1. *Масштабируемость.* В систему *xRandNet* минимальными средствами можно включать как принципиально новые исследования, так и новые модели случайных сетей, а также реализовать новые типы хранилищ данных.

2. *Способность к взаимодействию.* В системе *xRandNet* в целях возможности передачи данных были учтены распространённые форматы вход/выходов других программных средств, например, *Tulip*, *Origin Pro* и т.п.

Разработанная система *xRandNet* спроектирована соответственно требованиям, сформулированным выше. Система реализована на языке *C#* платформы *.Net Framework 4.5*. Архитектуру системы можно разделить на три независимых слоя (*layers*): графический интерфейс пользователя (*GUI*), управление сеансом системы (*Session Manager*) и ядро (*Core*) (рис. 3).

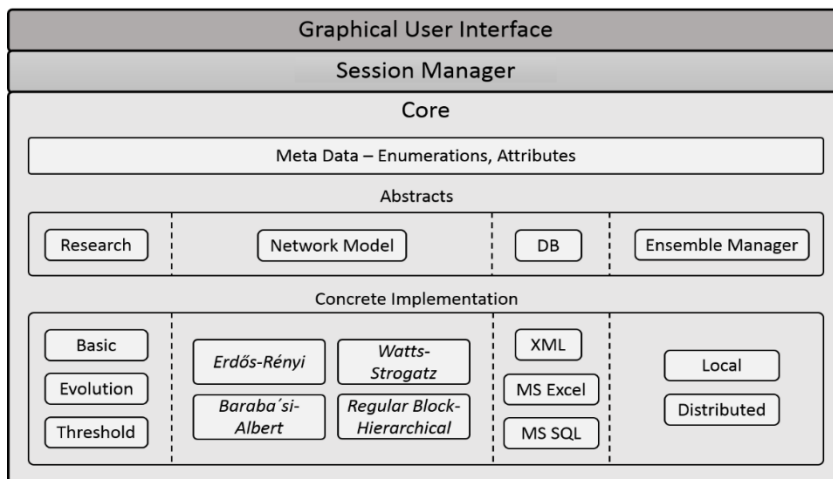


Рис. 3. Архитектурная структура *xRandNet*

Слой *GUI* использует элементы управления *Windows Forms* и технологию визуализации графики *GDI+*. Для взаимодействия *GUI* с более низкими слоями разработан модуль управления сеансом системы – *Session Manager*, который предоставляет единый интерфейс для запуска доступных исследований, процессов генерации и анализа случайных сетей, сохранения результатов и дальнейших запросов к полученным данным.

Слой *Core* состоит из трех основных компонент - организация работы с мета-данными типов (*Meta Data*) по технике позднего связывания с использованием атрибутов и отражений типов (*Enumerations, Attributes*), абстрактные интерфейсы, взаимодействие между ними (*Abstracts*) и имплементация (*Concrete Implementation*). В компоненте *Abstracts* определяются базовые классы для последующей имплементации видов исследований (*Research*), моделей сетей (*Network Model*), хранилищ данных (*DB*) и управления ансамблем (*Ensemble Manager*).

Важнейшей абстракцией в системе является *Network Model* (рис. 4).

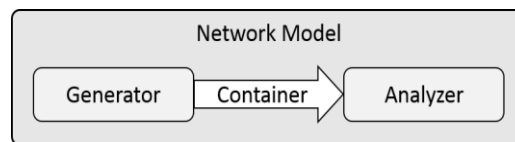


Рис. 4. Абстракция *Network Model*

Модель основана на том, что случайная сеть определяется тремя компонентами: генератор (*Generator*), контейнер (*Container*) и анализатор (*Analyzer*). Логика взаимодействия этих трех абстрактных компонентов такова: генератор на основе входных параметров генерирует случайную сеть в контейнере, затем на сгенерированном контейнере анализатор запускает алгоритмы вычислений топологических характеристик.

Так как процесс генерации и вычислений характеристик выполняется независимо, в системе реализован параллелизм, управляемый в модуле *Ensemble Manager*. Параллельная обработка каждой сети в ансамбле выполняется или с помощью потоков (с учётом числа ядер на машине), или распределением работы в локальной сети.

В компоненте *Concrete Implementation* представлены конкретные реализации исследований, моделей случайных сетей, взаимодействие с разными типами хранилищ данных, многопоточная или распределенная организация работы над ансамблями и т.д. Именно на этом этапе для каждой модели сети разработаны соответствующие алгоритмы генерации, структуры хранения в памяти (контейнеры) и алгоритмы вычислений топологических характеристик,

оптимизированные на основе особенностей данной модели. В *Concrete Implementation* использовались средства *ADO.NET* для доступа к базам данных, объекты синхронизации пространства имен *Threading* и *WCF* технология для распределения в локальной сети.

Структура хранения блочно-иерархических сетей. Учитывая, что система *xRandNet* сориентирована на проведение исследований над случайными сетями моделей *RBH* и *NRBH*, можно сказать, что важную роль в разработке имеет структура хранения блочно-иерархических сетей.

Приведем определение блочно-иерархических сетей. Пусть b и Γ – натуральные числа, $b > 1$. Для заданных b и Γ определяется класс **регулярно-ветвящихся блочно-иерархических сетей** $\mathcal{R}_{b,\Gamma}$. Число узлов сети $G_{b,\Gamma} \in \mathcal{R}_{b,\Gamma}$ равно b^Γ . Сеть строится по уровням. На каждом новом уровне γ , $0 \leq \gamma \leq \Gamma$, новые кластеры (подсети) формируются посредством объединения кластеров, построенных на предыдущем уровне, и введения новых связей между ними посредством соединения некоторых из них [4] (рис. 5).

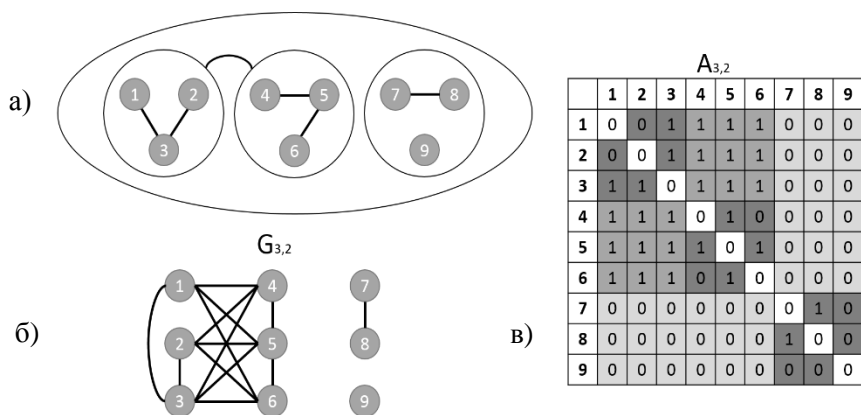


Рис. 5. Блочно-иерархическая сеть $G_{3,2}$: а- вид с выделенными кластерами, б- вид с выделенными связями, в- вид ее матрицы смежности $A_{3,2}$

Рассмотрен также нерегулярный случай, когда число кластеров, которые объединяются в один кластер, определяется случайным образом от единицы до значения b [6].

Такие сети представляются в виде **дерева связей**, листьями которого являются узлы сети, а поддеревья соответствуют кластерам (группам узлов) и отмечены последовательностью нулей и единиц, которые представляют связи между соответствующими узлами кластера (рис. 6). Все алгоритмы вычисления топологических характеристик сети блочно-иерархической модели используют дерево связи.

В системе *xRandNet* дерево связей реализовано в виде двумерного массива (*BitArray*[[[] на языке C#). Для каждого уровня конструируется битовая последовательность, являющаяся совокупностью меток всех узлов данного уровня. Очевидно, что на уровне $\gamma, 0 \leq \gamma \leq \Gamma$ дерево имеет b^γ узлов, каждый из которых “отмечен” последовательностью 0 и 1 длиной $b * (b - 1)/2$. Следовательно, размер использованной памяти равен

$$\Gamma * \frac{b*(b-1)}{2} * \sum_{\gamma=0}^{\Gamma} b^\gamma. \quad (1)$$

Как известно, платформа *.Net Framework* имеет ограничение на размер *BitArray*-а (4ГБ для 64-разрядной машины), что, как видно из (1), может привести к переполнению при достаточно больших значениях b или Γ . Во избежание этого битовая последовательность для каждого уровня разбивается на нужное количество *BitArray*-ов, размер которых не превышает некоторое фиксированное значение (в системе *xRandNet* это значение равно $2 * 10^9$) (рис. 6). Так как изучение блочно-иерархических сетей, как правило, предполагает значения $b < 10$ и $\Gamma < 20$, то можно сказать, что с помощью данной структуры можно представлять сети практически любого размера.

Для выбора структуры хранения также важно оценить сложность алгоритма определения наличия связи между любыми двумя вершинами сети. Очевидно, что для дерева связи этот алгоритм потребует не больше Γ шагов (высота дерева).

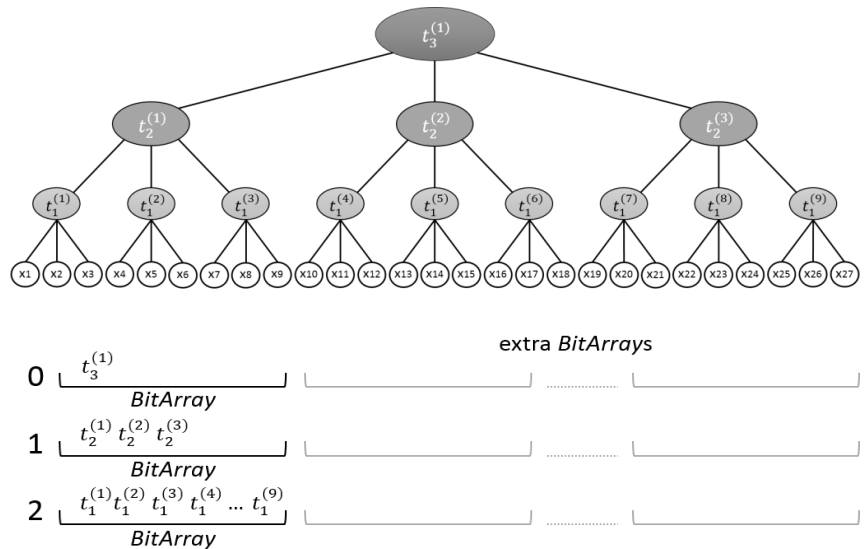


Рис. 6. Дерево связей для сети $G_{3,3}$ и его представление в памяти

Как известно, для представления сетей в памяти традиционно используются два подхода: матрица смежности и список смежности [9]. Матрица смежности, реализованная с помощью битовой последовательности, для сети размером N использует $(N^2 - N)/2$ бит. А размер использованной памяти для списка смежности равен $N * (32 + N/2 * 64) = 32 * N * (1 + N)$ с допущением, что машина 32-разрядная, размер целочисленного значения равен 32 бита, и в среднем каждая вершина может иметь порядка $N/2$ соседей.

Основываясь на вышесказанном, можно сделать вывод, что структура хранения блочно-иерархических сетей в виде дерева связей является наиболее оптимальной, что также видно из данных табл. 1. Заметим, что для нерегулярного случая дополнительно хранится информация о ветвлении [6].

Таблица 1

Размеры использованной памяти в битах для разных структур хранения. Размеры вычислены для сети порядка $N = 3^{12} = 531\,441$ на 32-разрядной машине, где Γ - максимальный уровень блочно-иерархической сети (высота дерева связи)

Реализация	Размер памяти	Сложность алгоритма определения наличия связи
Матрица смежности	$\approx 10^{12}$	$O(1)$
Списки смежности	$\approx 9 * 10^{12}$	$O(N)$
Структура хранения для блочно-иерархических сетей	$\approx 3 * 10^7$	$O(\Gamma)$

Работа системы xRandNet. Для иллюстрации работы системы xRandNet проведены базовые исследования для моделей сетей Erdős-Rényi и RBH. В табл. 2 приведены средние значения времен выполнения основных алгоритмов для модели Erdős-Rényi и модели RBH (специализированных алгоритмов). По этим данным можно сделать вывод, что система имеет высокую производительность для сетей блочно-иерархического класса.

Сравнение времен выполнения проведено на машине с процессором Intel(R) Core(TM) i7-3540M, 3.00 GHz и ОЗУ 4.00 GB.

Таблица 2

Среднее время выполнения алгоритмов вычисления средней степени, среднего коэффициента кластеризации, числа циклов длиной 3, 4 и распределения степеней вершин для сетей модели Erdős–Rényi размером $N = 512$ с вероятностью связи 0,3 и для сетей модели RBH размером $N = 2^9 = 512$, $N = 2^{20} = 1\,048\,576$, $N = 3^{13} = 1\,594\,323$, $N = 5^{10} = 9\,765\,625$ и $N = 7^8 = 5\,764\,801$ с вероятностью связи 0,3

Топологическая характеристика	ER	RBH				
		512	2^{20}	3^{13}	5^{10}	7^8
Размер сети N	512	512	2^{20}	3^{13}	5^{10}	7^8
Средняя степень	200 с	$\ll 1$ с	1 с	1 с	7 с	5 с
Средний коэффициент кластеризации	200 с	$\ll 1$ с	60 с	60 с	180 с	180 с
Число циклов длиной 3	200 с	$\ll 1$ с	2 с	4 с	24 с	24 с
Число циклов длиной 4	360 с	$\ll 1$ с	5 с	5 с	60 с	60 с
Распределение степеней	360 с	$\ll 1$ с	3 с	3 с	25 с	25 с

Заключение. Спроектирована и реализована программная система *xRandNet*, которая используется для исследования новых моделей RBH и NRBH при изучении поведения широкого набора статистических характеристик и позволяет проводить сравнение со стандартными моделями.

Предполагается развитие системы с добавлением новых моделей и новых видов исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Albert R., Barabási A.L. Statistical mechanics of complex networks // Rev. Mod. Phys.-2002.-74.-P. 47–97.
2. Some Physical Applications of Random Hierarchical Matrices / V.A. Avetisov, A.Kh. Bikulov, O.A. Vasilyev, et al // JETP.-2009.-109(3).-P. 485-504.
3. Mirny L.A. The fractal globule as a model of chromatin architecture in the cell // Chromosome Res.-2011, Jan.-19(1).-P. 37-51.
4. Алгоритмы вычислений статистических свойств регулярных блочно-иерархических сетей / С. Аветисян, А. Арутюнян, Д. Асланян и др. // Шестая годовичная научная конференция (5–9 декабря 2011г.): Сборник научных статей.-Ер.: Изд-во РАУ, 2012.-С. 108-121.
5. Аветисян С., Кочарян А. О циклах длины четыре в регулярных блочно-иерархических сетях // Ученые записки АрГУ.-2013.-2.-С. 26-35.
6. Аветисян С., Самвелян М., Карапетян М. Случайная нерегулярная блочно-иерархическая сеть: алгоритмы вычислений основных свойств // Девятая годовичная научная конференция (3–8 декабря 2014г.): Сборник научных статей.– Ер.: Изд-во РАУ, 2015.-С. 48-60.

7. **Avetisyan C., Kocharyan A.** The System of Generation of Random Networks and Computation of Their Topological Properties // Proceedings of the Conference CSIT-2013, Publishing House of NAS of RA.-Yerevan, 2013.-P. 381-384.
8. **Hovhannisyan M., Avetisyan S.** Clustering of random networks under topological constraints // Proceedings of the Conference CSIT-2015, Publishing House of NAS of RA.-Yerevan, 2015.-P. 155-159.
9. **Уилсон Р.** Введение в теорию графов.-М.: Мир, 1977.- 208 с.

Ереванский государственный университет. Материал поступил в редакцию 31.01.2017.

Ա.Գ. ԳՈՉԱՐՅԱՆ

XRANDNET ԾՐԱԳՐԱՅԻՆ ՀԱՄԱԿԱՐԳ՝ ՊԱՏԱՀԱԿԱՆ ՑԱՆՑԵՐԻ ՏՈՊՈԼՈԳԻԱԿԱՆ ԲՆՈՒԹԱԳՐԻՉՆԵՐԻ ՀԵՏԱԶՈՏՄԱՆ ՀԱՄԱՐ

Պատահական ցանցերի ակտիվ հետազոտությունն արդիական է դարձնում այնպիսի համակարգի մշակումը, որն ունակ է իմիտացնելու տարբեր մոդելների (*Erdős-Rényi*, *Watts-Strogatz*, *Baraba'si-Albert*, *Regular Block-Hierarchical* և *Non Regular Block-Hierarchical*) պատահական ցանցերի վարքը և իրականացնելու դրանց տոպոլոգիական հատկանիշների վերլուծությունը: Ներկայացված *xRandNet* համակարգը նպատակաուղղված է հատկապես բլոկ-հիերարխիկական ցանցերի արդյունավետ իմիտացիային, որոնք բավականին նոր օբյեկտներ են տվյալ ոլորտում:

Առանցքային բառեր. պատահական ցանցեր, բլոկ-հիերարխիկական ցանցեր, տոպոլոգիական բնութագրիչներ:

A.G. KOCHARYAN

APPLICATION XRANDNET FOR STUDYING THE TOPOLOGICAL CHARACTERISTICS OF RANDOM NETWORKS

The active research of random networks makes the development of an application capable of imitating the random networks of different models (*Erdős-Rényi*, *Watts-Strogatz*, *Baraba'si-Albert*, *Regular Block-Hierarchical* and *Non Regular Block-Hierarchical*), relevant and analyzing their topological properties. The developed application *xRandNet* is aimed at the efficient imitation of block-hierarchical networks which are quite new objects in this field.

Keywords: random networks, block-hierarchical networks, topological properties.