УДК 004.942

Н. Н. Николаев, **П. Ю. Шамин**, **М. Ю. Звягин**, **В. А. Скляренко**

Владимирский государственный университет
ул. Горького, 87, Владимир, 600000, Россия

E-mail: nikolaevn.home@gmail.com

РАЗРАБОТКА ПРОГРАММНОГО СРЕДСТВА МОДЕЛИРОВАНИЯ ЗАДАЧ ПОИСКА ИНФОРМАЦИИ В СОЦИАЛЬНЫХ ГРАФАХ *

Рассматривается инструмент моделирования социальных сетей. Предметом изучения являются модели, описывающие взаимодействия агентов внутри сообществ сети и между сообществами для решения общей задачи.

Ключевые слова: моделирование социальных сетей, граф, распределенное моделирование, перколяция.

Введение

Социальная сеть представляет собой граф, вершинами которого являются агенты, а ребрами – связи между ними. В качестве примера такой сети можно привести сети неформальных знакомств, партнерских связей, контактов в неком социуме и т. д. Будем называть их соседними узлами, или соседями вершины графа с общим ребром, или, иначе, агентов, имеющих связь между собой. В структуре большинства социальных сетей проявляются сообщества – группы агентов с большей плотностью связей между собой, нежели с агентами, не входящими в состав этой группы. Таким образом, формально можно рассмотреть еще один граф – граф связей между агентами и сообществами, ребро в котором означает принадлежность агента к сообществу. В процессе исследования глобальных сообществ в сети Интернет выяснилось, что крупным социальным сетям присущи несколько свойств. Как правило, для таких сетей справедливо так называемое «Правило шести рукопожатий». Средняя длина пути от одного случайного агента до другого невелика. Вершины, имеющие общего соседа, как правило, являются связанными между собой, т. е. в такой сети много треугольников или кластеров.

Существующие модели сетей рассматриваемого типа делятся на два больших класса: вероятностные и теоретико-игровые. Вероятностные изучают свойства типичных графов, построенных согласно некоторому распределению. Теоретико-игровые вводят некоторые правила игры и моделируют сети, возникающие в равновесных состояниях [1].

В данной работе рассматривается еще один вид теоретико-игровых моделей. Исследуется перколяционная задача поиска параметров сети, влияющих на возможность решения сетью некоей головоломки, части которой распределены между агентами сети. Решение головоломки возникает при обмене информацией между узлами и нахождении подходящих частей

* Работа выполнена в рамках НИР по госзаданию «Наука» (№ 8.3303.2011 и 8.3534.2011 от 23.11.2011 г.).

Николаев Н. Н., Шамин П. Ю., Звягин М. Ю., Скляренко В. А. Разработка программного средства моделирования задач поиска информации в социальных графах // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11, вып. 3. С. 47–52.

© Н. Н. Николаев, П. Ю. Шамин, М. Ю. Звягин, В. А. Скляренко, 2013

агентами сети [2]. В качестве примера можно привести задачу подбора команды специалистов для работы над некоторым проектом.

Способы распространения информации о рабочих местах играют большую роль с точки зрения предложения труда. Согласно исследованиям более чем в половине случаев трудоустройства используется информация из неформальных источников [3; 4]. В терминах социальной сети – информация распространяется агентами сети. Предполагается, что сеть состоит из некоторого числа сообществ, объединенных общей профессиональной направленностью. Основываясь на том, что в реальном профессиональном сообществе группы специалистов, занятых в одной профессии, с большой долей вероятности знакомы между собой, нежели с представителями других профессий, шанс распространения информации внутри одного сообщества больше, чем между сообществами.

В данной работе исследуется возможность решения задачи подбора группы агентов заданных профессий и специализаций в зависимости от конфигурации сети. Конкретно ставится задача оценки показателей успешности или результативности решения сетью этой задачи при заданных параметрах.

Опишем постановку задачи более формально. Пусть имеется множество агентов $V = \{1, ..., n\}$. Каждый агент может принадлежать к некоторому подмножеству $A = \{1, ..., m\}$, являющемуся в условии задачи поиска специалистов множеством профессий. В свою очередь, каждый элемент множества A принадлежит множеству $B = \{1, ..., k\}$, определяющему специализацию элемента в подгруппе A. Таким образом, каждый агент может быть обозначен $v_n^{m,k}$, где n – индекс вершины графа V; m – индекс подгруппы из множества A (профессия); k – индекс из множества B (специализация). Задача состоит в поиске набора агентов сети с параметрами, соответствующих запросу в виде $\left\{ x_i^{j,t} : i \in \mathrm{N}, j \in 1, ... m, t \in 1, ... k \right\}$, где j – требование принадлежности искомого агента к соответствующей профессии; t – принадлежность агента к специализации внутри профессии. Размер запроса может варьироваться. Запрос инициализируется одним из агентов и передается соседним агентам. Те, в свою очередь, проверяют свои параметры на соответствие одному из элементов запроса и передают его своим соседям. При обнаружении соответствия агент связывается напрямую с инициатором запроса.

Каждый узел v_i имеет некоторое количество соседних узлов как среди представителей «своей» подгруппы, так и среди представителей других подгрупп. Представитель, получив и проанализировав запрос, ретранслирует его своим соседям с некоторой долей вероятности, равной p; причем для соседа, находящегося внутри той же подгруппы, вероятность ретрансляции запроса будет больше, чем для соседей вне ее.

Разработка средства моделирования

В последнее десятилетие фокус исследований сетей смещается от изучения небольших графов и анализа свойств их структуры и связей к рассмотрению статистических свойств крупномасштабных сетей, состоящих из миллионов и миллиардов узлов. Если раньше хорошим подспорьем для исследований было визуальное представление сети и связей в ней, то с увеличением масштабов такой подход становится бесполезным. В связи с этим возникает необходимость создания инструмента, позволяющего дать ответ на вопрос, что представляет собой сеть, если на нее нельзя взглянуть. В настоящее время существует достаточное количество как коммерческих, так и свободно распространяемых проектов – симуляторов сетей различной направленности, использующих различные подходы к моделированию сетей и анализу результатов [5]. Из недостатков большинства подобных систем следует отметить сложность их использования, в том числе требования к знанию языков программирования для задания параметров исследуемой сети. Многие из систем налагают ограничения на используемое программное обеспечение и оборудование. Значительным затруднением является также автоматизация проведения экспериментов и сбор результатов.

В качестве основы для реализации инструмента моделирования для рассматриваемой задачи мы выбрали программный комплекс «Параллельный сетевой симулятор» (ПСС), разработанный на кафедре физики и прикладной математики ВлГУ [6]. Симулятор представляет собой распределенное приложение, что позволяет задействовать в ходе моделирования большие объемы оперативной памяти и обеспечивает масштабируемость процесса моделирования в плане размерности сети. Универсальность применения симулятора достигается модульной архитектурой приложения. Выбор модулей расширения дает возможность определять способ конструирования топологии сети, способы взаимодействия и обмена информацией между узлами сети и сбора необходимой информации о функционировании модели в ходе процесса моделирования.

В процессе адаптации ПСС для решения поставленной задачи был реализован набор модулей, включающих генератор топологии сети, модуль логики взаимодействия узлов, модуль ретрансляции информационных запросов в сети, модуль сбора статистики. Рассмотрим кратко назначение каждого модуля.

Модуль генерации топологии определяет количество узлов сети согласно параметрам, передаваемым при запуске ПСС. В поставленной задаче предполагается, что связи между узлами двунаправленные. Входными параметрами для модуля генерации топологии являются минимальное и максимальное число соседей узла и значения вероятностей ретрансляции, зависящие от принадлежности либо непринадлежности соседа той же подгруппе, что и текущий узел. С учетом этих параметров модуль строит матрицу связности для графа сети. Затем набор непосредственных соседей передается каждому узлу. Число подгрупп и число специализаций внутри подгруппы также определяется входными параметрами модуля.

Модуль логики взаимодействия проверяет соответствие параметров принадлежности текущего узла набору параметров пришедшего запроса. Если соответствие найдено, устанавливается соответствующий признак в запросе.

Модуль ретрансляции запроса осуществляет инициализацию запросов для передачи их соседним узлам. В зависимости от принадлежности узла к той же подгруппе, что и текущий узел и вероятности ретрансляции, присвоенной на этапе генерации сети, принимается решение о посылке запроса соседнему узлу.

Модуль сбора статистики осуществляет подсчет количества ретрансляций, потребовавшихся для подбора узлов из запроса, подсчитывает общее число запросов, созданных в узлах сети при моделировании, фиксирует факт успешного нахождения набора.

Таким образом, в качестве входных данных для работы алгоритма ПСС при моделировании данной задачи были определены следующие параметры: количество подгрупп и специализаций в них; диапазон возможного количества связей внутри подгрупп и между подгруппами; вероятность ретрансляции внутри и между подгруппами; объем запроса.

Результаты моделирования

В ходе моделирования были получены зависимости вероятности успешного решения сетью задачи поиска группы специалистов от параметров сети.

Как отмечается в некоторых работах [3; 4], большое влияние на распространение информации в социальных сетях оказывают так называемые слабые связи, т. е. связи вне профессиональных сообществ. В качестве примера, подтверждающего эту зависимость, была проведена серия экспериментов, в которых изменялось среднее количество связей с агентами сети вне сообществ и вероятность ретрансляции запросов этим агентам. Полученные результаты действительно показывают увеличение числа успешных результатов решений задачи в реализованной модели сети при увеличении количества слабых связей и вероятности ретрансляции (рис. 1).

Была также рассмотрена зависимость числа успешных решений от плотности связей сети. Отмечается рост числа успешных решений при увеличении среднего числа связей одного агента даже при малом количестве связей с агентами вне сообщества (рис. 2).

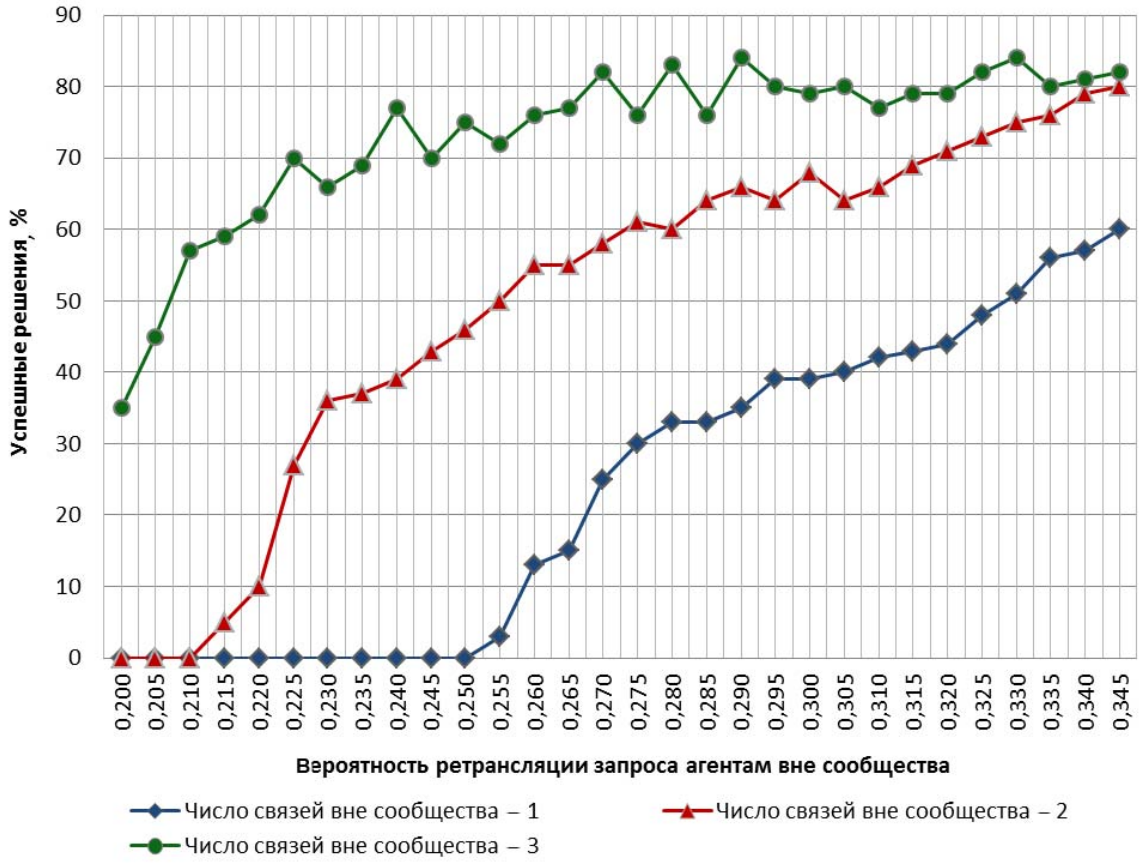

Рис. 1. Графики зависимостей числа успешных решений задачи сетью при изменении вероятности ретрансляции запроса агентам вне сообществ для трех вариантов среднего количества связей с агентами вне сообщества

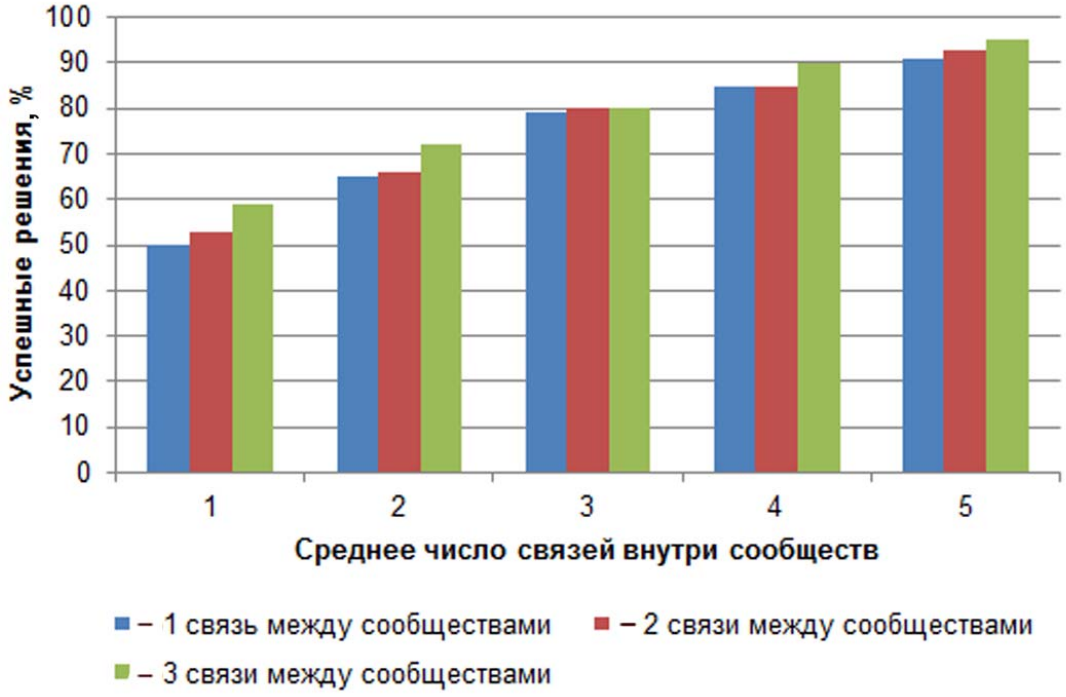

Рис. 2. Зависимость числа успешных решений от среднего количества связей агентов

Направления дальнейшей работы

Как было проиллюстрировано, реализованный симулятор с соответствующим набором модулей расширения позволяет проводить исследования модели при изменении таких параметров, как количество агентов в сети, количество связей агента, количество связей внутри и между сообществами, значения вероятности ретрансляции запроса между агентами сообщества и между сообществами. При этом численные эксперименты могут проводиться последовательно с изменением одного или нескольких параметров с целью получения числовых рядов для дальнейшего анализа. По завершении сеанса моделирования система регистрирует начальные параметры, количество успешных решений задачи сетью, потребовавшееся количество шагов, количество созданных в системе запросов между агентами.

Все доработки, планируемые нами, направлены в первую очередь на расширение областей применения симулятора и улучшение адекватности моделей в плане приближения их к реально существующим социальным сетям.

Во-первых, рассматривается переход к сетям с динамическим изменением активности связей. Для этого необходимо реализовать механизм, который с вероятностью $0 \le f \le 1$ будет изменять статус активности связи с определенной периодичностью. При этом агент при ретрансляции запроса будет определять активность соседнего узла и учитывать не только вероятность передачи p, описанную в постановке задачи, но и активность связи на момент инициализации запроса. Это позволит имитировать, например, такие ситуации, когда человек переезжает в другой город, меняет место работы и т. п., в результате чего он, как правило, теряет часть имевшихся связей, но при этом приобретает новые.

Еще одно направление нашей дальнейшей работы ориентировано на совершенствование управления модельным временем. На данный момент в системе ПСС реализовано два варианта симуляции работы сети, отличающихся методами изменения и учета модельного времени. В первом варианте отсчет модельного времени ПСС производится циклично фиксированными шагами по времени. Второй вариант подразумевает изменение модельного времени с переменным шагом, но также синхронно в пределах всей модели, недостатки такого подхода в определенных ситуациях могут негативно сказаться на производительности симулятора. В качестве дальнейшего развития предполагается реализация модели в концепции непрерывного модельного времени с асинхронной работой сегментов модели, что позволит обеспечить большую адекватность моделирования процессов во времени [7]. Эта доработка даст возможность отказаться от учета различных вероятностей ретрансляции запроса внутри сообществ агентов и между ними и применять одно значение вероятности ретрансляции в пределах всей сети. Взамен предлагается увеличивать интенсивность рассылки запросов внутри сообществ и уменьшать при ретрансляции запроса за пределы сообщества. Переход к непрерывному времени позволит также более гибко определять алгоритм изменения активности связей, что облегчит работу и по первому из описанных направлений развития инструмента моделирования, и самой модели.

Список литературы

1. *Borgs C., Chayes J. T., Ding J., Lucier B.* The Hitchhiker's Guide to Affiliation Networks: A game-theoretic Approach // Proceedings of the 2nd Symposium on Innovations in Computer Science. Tsinghua University, Beijing, China, 2011. P. 389–400.

2. *Brummit Ch. D., Shirshendu Chatterjee, Dey P. S., Sivakoff D.* Jigsaw Percolation: What Social Networks Can Collaboratively Solve a Puzzle? University of California, Davis, Courant Institute of Mathematical Sciences and Duke University, 2012. P. 1–18.

3. *Давыденко В. А., Ромашкина Г. Ф., Чуканов С. Н.* Моделирование социальных сетей // Вестн. Тюмен. гос. ун-та. 2005. № 1. С. 68–79.

4. *Granovetter M.* The Sociological Approaches to Labor Market Analysis: A Social Structural View // Granovetter I., Swedberg R. (Eds.) The Sociology of Economic Life. Boulder, Westview Press, 1992. P. 244–245.

5. *Di Caro G. A.* Analysis of Simulation Environments for Mobile ad Hoc Networks // Technical Report IDSIA-24-03 December 2003. IDSIA, Lugano, Switzerland, 2003. URL: http://www.idsia.ch/idsiareport/IDSIA-24-03.pdf (дата обращения: 18.07.2013)

6. *Шамин П. Ю., Алексанян А. С., Прокошев В. В.* Параллельный сетевой симулятор: концепция и перспективы развития // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. 2009. № 3. С. 18–24.

7. *Николаев Н. Н., Шамин П. Ю.* Модернизация параллельного сетевого симулятора для реализации «мягкого» варианта непрерывного модельного времени // Актуальные проблемы естественных и математических наук: Материалы Междунар. заоч. науч.-практ. конф. / 04 марта 2013 г. Новосибирск: СибАК, 2013. С. 48–52.

Материал поступил в редколлегию 12.06.2013

N. N. Nikolaev, P. Yu. Shamin, M. Yu. Zvyagin, V. A. Sklyarenko

IMPLEMENTING SOFTWARE SYSTEM FOR MODELLING INFORMATION SEARCH IN SOCIAL GRAPHS

This article is focused on development of the instrument for social networks modeling. Subject of studying are models describing interactions of agents in communities of a network and between these communities for the solution of the general task.

Keywords: social networks modeling, graph models, distributed modeling, percolation.

References

1. *Borgs C., Chayes J. T., Ding J., Lucier B.* The Hitchhiker's Guide to Affiliation Networks: A game-theoretic Approach // Proceedings of the 2nd Symposium on Innovations in Computer Science. Tsinghua University, Beijing, China, 2011. P. 389–400.

2. *Brummit Ch. D., Shirshendu Chatterjee, Dey P. S., Sivakoff D.* Jigsaw Percolation: What Social Networks Can Collaboratively Solve a Puzzle? University of California, Davis, Courant Institute of Mathematical Sciences and Duke University, 2012. P. 1–18.

3. *Davydenko V. A., Romashkina G. F., Chukanov S. N.* Modelirovanie socialnyh setey // Vestn. Tumen. gos. un-ta. 2005. № 1. S. 68–79.

4. *Granovetter M.* The Sociological Approaches to Labor Market Analysis: A Social Structural View // Granovetter I., Swedberg R. (Eds.) The Sociology of Economic Life. Boulder, Westview Press, 1992. P. 244–245.

5. *Di Caro G. A.* Analysis of Simulation Environments for Mobile ad Hoc Networks // Technical Report IDSIA-24-03 December 2003. IDSIA, Lugano, Switzerland, 2003. URL: http://www.idsia.ch/idsiareport/IDSIA-24-03.pdf (дата обращения: 18.07.2013)

6. *Shamin P. U., Aleksanyan A. S., Prokoshev V. V.* Parallelnyy setevoy simulyator: koncepciya i perspektivy razvitiya // Nauchno-tehnicheskie vedomosti Sankt-Peterburgskogo gosudarstvennogo politehnicheskogo universiteta. 2009. № 3. S. 18–24.

7. *Nikolaev N. N., Shamin P. U.* Modernizaciya parallelnogo setevogo simulyatora dlya realizacii «myagkogo» varianta nepreryvnogo modelnogo vremeni // Aktualnye problemy estestvennyh i matematicheskih nauk: Materialy Mezhdunar. zaoch. nauch.-prakt. konf. / 04 marta 2013 g. Novosibirsk: SibAK, 2013. S. 48–52.