

«УТВЕРЖДАЮ»

Директор ФИЦ ИУ РАН

И.А. Соколов

И.А. Соколов

«11» октября 2019 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального исследовательского центра «Информатика и управление»

Российской академии наук (ФИЦ ИУ РАН)

на диссертацию Дробышевского Михаила Дмитриевича

«Методы и программные средства моделирования и генерации сложных сетей с сохранением графовых свойств»,

представленной на соискание учёной степени кандидата физико-математических наук по специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Актуальность темы

В реальном мире многие данные имеют графовую структуру: это социальные сети (Фейсбук, Твиттер и др.), интернет, графы цитирований, метаболические сети, нейронные сети, автономные системы и т.д., обладающие сложной топологической структурой. В рамках науки о сложных сетях последние десятилетия активно разрабатываются модели случайных графов, используемые в различных прикладных задачах, связанных с анализом сетей. Порождение случайных графов позволяет решить задачу анонимизации реальных сетевых сообществ, а также даёт возможность создать масштабные выборки для машинного обучения, состоящие из сгенерированных графов, которые в определённом смысле аналогичны исходным данным. Основной трудностью, возникающей при таком подходе, является адекватное воспроизведение необходимых топологических свойств реальных графов из различных областей.

Целью диссертации Дробышевского М.Д. является создание метода генерации сложных сетей, способного воспроизводить заранее неизвестные топологические свойства заданного графа. При этом рассматриваются направленные взвешенные графы со структурой сообществ и требуется возможность генерации графов произвольного размера. Суть работы заключается в применении методов вложения графа для создания моделей случайных графов, сохраняющих топологические свойства.

Общая характеристика работы

Диссертация состоит из введения, трёх глав, заключения и двух приложений. Полный объем диссертации составляет 163 страницы с 42 рисунками и 8 таблицами. Список литературы содержит 183 наименований.

Во введении обоснована актуальность работы, которая обусловлена наличием двух проблем, часто возникающих на практике: нехватка графовых данных для проведения тестирования алгоритмов интеллектуального анализа графов и проблема анонимизации графов для возможности их публикации.

Первая глава посвящена обзору существующих моделей случайных графов. Проведен тщательный анализ литературы, в котором представлен новый взгляд на предметную область. В процессе обзора составлена иерархическая таксономия подходов моделирования графов. По результатам обзора подготовлена статья к публикации в журнале «ACM Computing Surveys», прошедшая первое рецензирование.

Во второй главе представлен новый подход — Embedding based Random Graph (ERGG), основанный на вложении графа (graph embedding), при котором признаки графа кодируются в наборе векторов его вершин в пространстве малой размерности. В рамках подхода разработан метод, позволяющий автоматически извлекать признаки ориентированного графа из любого домена и генерировать похожие графы, масштабируя размер исходного графа на вещественный коэффициент.

В третьей главе описана программная реализация метода, с помощью которой проведено экспериментальное исследование метода и его сравнение с ближайшими аналогами. Показана способность метода генерировать графы, близкие к заданному по ряду известных признаков, и в то же время имеющие достаточный разброс по ним для возможности имитации соответствующего разброса в реальных графах.

В заключении сформулированы основные результаты диссертационной работы:

Основные результаты диссертационной работы

В ходе исследований автором получены следующие результаты, выносимые на защиту:

1. Проведен анализ существующих моделей случайных графов, в результате которого предложена таксономия подходов к решению задачи исследования.
2. Предложен новый подход ERGG к генерации случайных направленных графов, похожих на данный, основанный на вложении графа в низкоразмерное пространство. Подход позволяет совмещать автоматическое извлечение признаков заданного графа, контролируемый размер генерируемых графов и поддержку ориентированных взвешенных ребер и структуры сообществ.
3. В рамках подхода ERGG предложен метод ERGG-dwc (от англ. directed, weighted, communities), решающий задачу генерации графов, похожих на данный, и удовлетворяющих условиям: автоматическое извлечение признаков, контролируемый размер генерируемых графов,

поддержка направленных взвешенных ребер и структуры сообществ. Состоятельность метода ERGG-dwc подтверждена экспериментально.

4. Создана программная система, реализующая метод ERGG-dwc и проведено его экспериментальное сравнение с ближайшими аналогами. Показано что ERGG-dwc способен генерировать графы, близкие к исходному по ряду известных графовых признаков, и при этом имеющие разброс по признакам, аналогичный таковому для графов из одного домена. Это означает, что ERGG-dwc можно применять для создания искусственных выборок для тестирования статистической значимости и масштабируемости алгоритмов «майнинга» графов, а также в целях анонимизации графовых данных.

Все результаты, выносимые на защиту, являются новыми.

Достоверность результатов исследования

Состоятельность предложенного подхода и работоспособность разработанного метода подтверждаются экспериментами.

Основные результаты работы докладывались на двух российских и одной зарубежной конференциях, а также обсуждалась на научных семинарах. На конференции ECML PKDD 2017 получена награда «лучшая студенческая работа».

Результаты исследования опубликованы в журналах из списка ВАК (три научные работы), еще одна статья подготовлена к публикации и прошла первый этап рецензирования в журнале «ACM Computing Surveys».

Теоретическая и практическая значимость полученных результатов.

Теоретическая значимость диссертации состоит в том, что автором разработан новый метод генерации случайных графов на основе вложения в условиях поставленных требований, а также изучены свойства графов, порождаемых по этому методу: доказаны соответствующие теоремы.

С практической точки зрения показана практическая применимость разработанного метода для создания искусственных графовых данных, а также для решения задачи анонимизации графов. Создание выборки графов одинакового размера позволяет тестировать статистическую значимость, а варьирование размера графа делает возможным оценку масштабируемости алгоритмов на практических задачах.

Замечания

В работе Дробышевского М.Д. отмечены следующие недостатки:

1. Из изложения работы не совсем ясен объем произведенных модификаций существующих алгоритмов вложения при разработке нового метода вложения графа.
2. Веб-демонстрация разработанного алгоритма упомянута, но не описана в тексте работы.
3. Формализация постановки задачи могла бы быть представлена автором в более полном и развернутом виде.

Заключение

Указанные недостатки не снижают однозначно положительную оценку работы и не умаляют высокой значимости полученных результатов. Диссертационная работа Дробышевского М. Д. обладает научной новизной и практической значимостью, представляя собой законченное научное исследование по актуальной теме. Диссертационная работа соответствует требованиям, предъявляемым ВАК РФ к работам на соискание степени кандидата физико-математических наук по специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Результаты диссертации Дробышевского М. Д. были представлены и рассмотрены на семинаре Отделения № 7 "Искусственный интеллект и принятие решений" (Институт проблем искусственного интеллекта) ФИЦ ИУ РАН 11 октября 2019 г. (Протокол № 7).

Заведующий Отделом 73 "Интеллектуальные технологии и системы" ФИЦ ИУ РАН
к.ф.-м.н.

И. В. Соченков
«11» октября 2019 г.
e-mail: sochenkov@isa.ru
тел.: +7 (499) 135-04-63