

ОТЗЫВ

официального оппонента на диссертационную работу Самосвата Е.А.
“Моделирование интернета с помощью случайных графов”,
представленную на соискание ученой степени кандидата
физико-математических наук по специальности 05.13.18 –
“математическое моделирование, численные методы и комплексы
программ”

В последнее время наблюдается существенный интерес к исследованию сложных сетей. Такие сети естественным образом возникают в самых разных областях. Например, в электроэнергетике, биологии, телекоммуникациях, экономике, социологии и т.д. Одной из важных сетей, возникающих в реальной жизни, является граф Интернета. Вершинами этого графа являются веб-страницы, а ребра показывают, как страницы связаны между собой посредством гиперссылок. Стремительное и широкое развитие интернета сделало актуальными исследования свойств графа Интернета. Важным инструментом при исследовании сложных сетей является их моделирование. Хорошие модели помогают понять и предсказать ключевые свойства растущих сетей, а также могут быть использованы при разработке различных графовых алгоритмов, приспособленных к определенным типам сетей. Диссертационная работа Самосвата Егора Александровича посвящена разработке, анализу и применению моделей графа Интернета. Актуальность темы диссертации не вызывает сомнений.

Диссертация состоит из введения, трех глав, заключения и списка литературы. Во введении изложена история и актуальность моделирования сложных сетевых структур, а также основные результаты работы.

Первая глава посвящена моделям предпочтительного присоединения. В 1999 году два физика Барабаши и Альберт заметили, что степени вершин ссылочного графа Интернета распределены в соответствии со степенным законом. В поисках естественного обоснования этой закономер-

ности они предположили, что новые страницы ссылаются на уже существующие по принципу предпочтительного присоединения. Другими словами, вероятность того, что новая страница процитирует другую страницу тем выше, чем больше ссылок последняя страница уже имеет. Позже рядом математиков были предложены различные формальные модели сетей, основанные на идее предпочтительного присоединения. Помимо распределения степеней вершин создание таких моделей было нацелено на описание и других свойств, присущих реальным сетям. Например, две из таких характеристик - это глобальный и средний локальный кластерные коэффициенты. Математический анализ, сопровождающий предлагаемые сетевые модели, направлен на доказательство того, что модель обладает теми или иными свойствами. Зачастую для разных моделей такой анализ схож и громоздок.

В разделе 1.1 описывается идея предпочтительного присоединения и приводятся конкретные модели, реализующие эту идею. В частности, дается формальное определение известной LCD-модели, формулируются свойства данной модели. В разделе 1.1.3 определяется модификация LCD-модели, свойства которой исследуются в разделе 1.2. В рамках этой модели доказаны два общих результата, названных теоремами о длинном и коротком спуске. С помощью этих сильных средств получена асимптотически точная оценка математического ожидания количества подграфов, изоморфных произвольному фиксированному графу.

В разделе 1.3 предлагается общий подход к моделям предпочтительного присоединения. А именно, в разделе 1.3.1 вводится параметризованный PA-класс моделей, который, как показано в работе, с одной стороны содержит многие известные модели предпочтительного присоединения, а с другой стороны обладает достаточно широкими свойствами. Так в разделе 1.3.2 доказано, что любая модель из PA-класса имеет степенной закон распределения вершин с параметром, зависящим от параметра модели. В разделе 1.3.3 получены оценки кластерных коэффициентов для моделей из PA-класса. Эти результаты позволяют легко получать новые модели, обладающие разнообразными свойствами. При этом свойства конкретной модели получаются как следствия из общих результатов о свойствах моделей из PA-класса и не требуют дополнительного громоздкого анализа. В качестве примера в разделе 1.3.4 приводится конкретная модель из PA-класса, позволяющая настраивать параметр распределения степеней и кластерный коэффициент. В разделах 1.3.5 и 1.3.6 проводится экспериментальная проверка свойств графов из приве-

денной модели.

Вполне естественно предположить, что разные части Интернет графа могут эволюционировать по разным законам. Действительно, личный опыт подсказывает, что популярность страницы с какими-нибудь качественными учебными материалами обычно монотонно возрастает со временем, тогда как страницы с новостями набирают популярность в короткие сроки, но при этом столь же быстро ее теряют. Другими примерами страниц со свойствами, похожими на свойства новостных страниц, могут служить записи в блогах и на форумах. Часть интернета, соответствующая таким страницам, называется медиа-вебом. Изучению законов развития медиа-веба и построению адекватных моделей соответствующего подграфа графа Интернета посвящена глава 2 работы Самосвата Е.А.

В разделе 2.1 обсуждается необходимость моделирования медиа-веба, а также применимость для этого существующих моделей сети интернета.

В разделе 2.2 эмпирически анализируется влияние возраста страницы на шансы получения ею новых ссылок. Обнаружено, что возрастом страница медиа-веба очень быстро теряет свою привлекательность с точки зрения получения новых ссылок. Таким образом замечено, что возраст страницы играет существенную роль в эволюции медиа-веба и, очевидно, должен учитываться при его моделировании. Вместе с тем, модели предпочтительного присоединения не учитывают этот фактор и, вероятно, в своем исходном виде малопригодны для моделирования медиа-веба.

Основываясь на результатах раздела 2.2, в разделе 2.3 предложен новый класс сетевых моделей, учитывающих такие факторы, как популярность, качество и возраст страницы. Стоит отметить, что предложенный класс, помимо новых моделей, содержит также известные до этого модели предпочтительного присоединения и модели приспособления.

В разделе 2.4 проводится теоретический анализ предложенного класса моделей. Используя полученные результаты, в классе выделяются модели, наиболее пригодные для моделирования медиа-веба. Для выделенных моделей в разделе 2.5 проводится статистический анализ их адекватности реальным данным, а также их сравнение с существующими моделями. Результаты показывают, что предложенные модели отражают свойства медиа-веба значительно лучше ранее известных моделей.

В главе 3 предложенная модель медиа-веба успешно применяется для разработки алгоритма эффективной индексации эфемерных страниц, то есть страниц, к которым быстро пропадает пользовательский интерес. Перед тем, как поисковая система сможет показывать страницы в ре-

зультатах поиска, эти страницы должны быть проиндексированы, то есть обнаружены и скачаны поисковым роботом. В случае эфемерных страниц, актуальность которых быстро теряется, особую важность приобретает задача уменьшения задержки между моментом создания страницы и временем ее скачивания. Оптимальность решения этой задачи сказывается на удовлетворенности пользовательского интереса, которая в свою очередь влияет на экономические показатели поисковой системы.

Для формализации решаемой задачи в разделе 3.1 вводится метрика, отражающая удовлетворенность пользователя в зависимости от задержек между созданием и скачиванием страниц. В разделе 3.2 выдвигается и проверяется гипотеза о том, что большинство новых эфемерных страниц можно обнаружить на небольшом наборе источников контента. Предложен простой метод формирования множества таких источников.

Результаты последних двух разделов используются в разделе 3.3 при разработке алгоритма переобхода источников контента. В этом разделе формулируется и решается задача оптимального расписания обхода источников и скачивания новых эфемерных страниц. Основываясь на решении этой задачи, предлагается алгоритм вычисления оптимального расписания. Предложенный алгоритм динамически оценивает качество и скорость появления новых ссылок на каждом из источников контента, что позволяет ему автоматически подстраиваться к изменениям интернета. В этом же разделе обсуждаются возможные реализации алгоритма. В разделе 3.4 экспериментально сравнивается эффективность расписаний, построенных как различными реализациями предложенного алгоритма, так и другими методами, основанными на естественных идеях. Результаты показывают, что предложенный алгоритм экономно расходует ресурсы и при этом достигает результатов, близких к оптимальным. Одной из сильных сторон алгоритма является то, что он формирует единое расписание для операций обхода источников и операций скачивания новых страниц, что позволяет ему более эффективно распределять ресурсы.

Замеченные недостатки работы относятся не к ее содержанию, а к стилю изложения:

1. название алгоритма 2, а также поясняющий текст ко всем графикам приведены на английском языке, что немного нарушает однородность оформления;
2. не приводится определение понятия масштабно-инвариантного графа, используемого на стр. 41;

3. на стр. 38 в третьей строчке с конца вместо слова “данному” должно быть слово “данном”.

Все математические утверждения диссертации снабжены строгими доказательствами. Сильной стороной работы является то, что разработанный теоретический аппарат был успешно применен для решения реальной прикладной задачи – предложенный в работе алгоритм составления расписания для поискового робота внедрен в компании Яндекс. Полученные в работе результаты могут найти применение в исследованиях по моделированию сложных сетей, а также при решении практических задач информационного поиска.

Все сказанное выше свидетельствует о том, что диссертационная работа Самосвата Е.А. является существенным вкладом в области моделирования сложных сетей.

Результаты диссертации представлены в 5 работах, все они опубликованы в журналах, рекомендованных ВАК. Тема диссертации соответствует специальности 05.13.18 – “математическое моделирование, численные методы и комплексы программ”, а сама диссертация удовлетворяет всем требованиям, предъявляемым к кандидатским диссертациям.

Считаю, что Самосват Е.А. несомненно заслуживает присвоения ученой степени кандидата физико-математических наук.

Официальный оппонент
кандидат физ.-мат. наук, доцент

ВВ

28.05.2014

Замараев В.А.



В.А. Замараев

НАЧАЛЬНИКОМ

ОТДЕЛА КАДРОВ

И.А. ЕРМОЛИНОЙ

[Signature]