МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(государственный университет)

ФАКУЛЬТЕТ ИННОВАЦИЙ И ВЫСОКИХ ТЕХНОЛОГИЙ
КАФЕДРА АНАЛИЗ ДАННЫХ

(Специализация «Прикладные информационные технологии
в управлении и бизнесе»)

ОТЫСКАНИЕ ОПТИМАЛЬНЫХ ПАРАМЕТРОВ В МОДЕЛЯХ СЛУЧАЙНЫХ ВЕБ-ГРАФОВ

Магистерская диссертация
студента 793 группы
Жернова Павла Владимировича

Научный руководитель
Райгородский А.М., д.ф.-м.н.

г. Москва
2013

Оглавление

Глава 1

Введение

Теория графов играет огромную роль, как в фундаментальной, так и в прикладной математике. Нас будет интересовать лишь одно направление, которое становится все более актуальным с каждым годом. Это графы, которые изучаются с вероятностной точки зрения. Случайные графы, которые описывают рост различных сетей, – социальных, биологических, транспортных – наиболее современны в этом направлении. В первую очередь, это связано, конечно же, с Интернетом.

Глава 2

Модели случайных веб-графов

Модели случайных веб-графов позволяют генерировать WWW-подобные графы, которые значительно меньше и проще, чем реальные WWW-графы, однако сохраняют определенные ключевые свойства структуры ребер веба. Такие искусственные графы можно рассматривать как экспериментальную платформу для получения новых подходов к поиску, индексации и т.д. Вершины веб-графа соответствуют веб-страницам, а ребра – гиперссылкам между ними. Веб-графы довольно активно изучались на предмет различных числовых характеристик таких, как распределение, диаметр, число связных компонент, макроскопическая структура. Ниже приведем различные модели, призванные описывать реальные веб-графы.

Один из возможных теоретических подходов к модели веб-графа – это математическая концепция случайного графа. Суть этого подхода заключается в том, что веб-граф развивается стохастически. Было предпринято множество попыток смоделировать граф гиперссылок интернета как случайный граф. Наиболее простой и исторически первой является модель

Эрдеша-Реньи.

2.1 Модель Эрдеша-Реньи

Пусть $V_n = \{1, \dots, n\}$ – множество вершин графа. Именно на них мы и будем строить наш случайный граф. Соединим любые две вершины a и b ребром с вероятностью $p \in [0, 1]$ независимо от всех остальных пар вершин. Другими словами, ребра в графе будут появляться в соответствии со схемой Бернулли, в которой вероятность успеха p и C_n^2 испытаний (нас не интересуют кратные ребра, петли; граф неориентирован). Пусть E – случайное множество ребер, полученное в результате реализации такой схемы. Граф $G = (V_n, E)$ и есть случайный граф в модели Эрдеша-Реньи.

2.2 Модели предпочтительного присоединения

В 90-е годы XX века в своих работах Барабаши и Альберт описали некоторые статистики интернета – веб-графа, вершинами которого являются страницы в интернете, а ребрами – гиперссылки между ними. На самом деле, похожую структуру имеют также большинство других реальных сетей – социальные, биологические, транспортные.

Основные результаты исследования Барабаши и Альберта состоят в следующем.

1. Веб-граф – это «разреженный» граф. У него на n вершинах всего mn ребер, где $m \in \mathbb{Z}$ – некоторая константа. Для сравнения, у полного графа на n вершинах $C_n^2 \sim n^2$ ребер.

2. Диаметр веб-графа очень мал (5-7, результат 1999 года). Это хорошо

известное свойство любой социальной сети, которое принято называть «мир тесен». Например, говорят, что любые 2 человека в мире «знакомы через 5-6 рукопожатий». В интернете это свойство заключается в том, что кликая 5-7 раз по ссылкам можно перейти между любыми двумя страницами. (Если говорить более точно, то в интернете есть только что появившиеся сайты, которые могут быть не связаны с остальными сайтами. Поэтому правильнее сказать, что в интернете есть огромная компонента, диаметр которой мал). Итак, веб-граф обладает интересным свойством – он разрежен, но при этом «тесен».

3. Для веб-графа характерен степенной закон распределения степеней вершин. То есть, вероятность того, что вершина веб-графа имеет степень d равна $cd^{-\gamma}$, где $\gamma = 2.1$. Интересно, что этот закон характерен для всех реальных сетей, но у каждой из них своя γ.

Таким образом, описанная выше модель случайного графа Эрдеша-Реньи плохо описывает реальные веб-графы, поскольку графы, полученные в этой модели, не имеют степенного закона распределения степени вершины.

Барабаши и Альберт предложили концепцию предпочтительного присоединения: граф строится с помощью случайного процесса, на каждом шаге которого добавляется новая вершина и фиксированное число ребер из новой вершины в уже существующие. При этом, вершины с большей степенью приобретают ребра с большей вероятностью, которая линейно зависит от их степени.

2.3 Модель Боллобаша-Риордана

Общая идея предпочтительного присоединения строго математически формулируется в модели Боллобаша-Риордана. Конструируется набор графов (марковская цепь) G_m^n, $n = 1, 2, \ldots$, с n вершинами и mn ребрами, где $m \in \mathbb{Z}$ – целое число. Сначала рассмотрим случай $m = 1$. Пусть граф G_1^1 – граф, состоящий из одной вершины и одного ребра (петля). Граф G_1^t получается из графа G_1^{t-1} добавлением вершины t и ребра из вершины t в вершину i, где i выбирается из существующих в графе вершин случайно, согласно следующему распределению вероятностей:

$$P(i = s) = \begin{cases} d_{G_1^{t-1}(s)}/(2t-1), & \text{если } 1 \leq s \leq t-1, \\ 1/(2t-1), & \text{если } s = t, \end{cases}$$

где $d_{G_1^{t-1}(s)}$ – степень вершины s в графе G_1^t.

Заметим, что распределение вероятностей задано корректно, поскольку:

$$\sum_{i=1}^{t-1} \frac{d(i)}{2t-1} + \frac{1}{2t-1} = \frac{2t-2}{2t-1} + \frac{1}{2t-1} = 1$$

Случайный граф G_1^n построен и он удовлетворяет принципу предпочтительного присоединения. Далее, граф G_m^n строится из графа G_1^{mn} объединением вершин $1, \ldots, m$ в вершину 1 нового графа, объединением вершин $m+1, \ldots, 2m$ в вершину 2 нового графа и так далее. Замети, что можно аналогичным образом строить ориентированные графы: ребро между вершинами i и j идет из i в j, если $i > j$.

Модель Боллобаша-Риордана хорошо отражает некоторые ключевые свойства различных реальных графов. Боллобаш и Риордан доказали, что диаметр графа G_m^n равен $\log n / \log\log n$ при больших n. Они также показали, что распределение степеней вершин графа G_m^n подчиняется степенному закону: число вершин со степенью d в модели хорошо аппроксимируется функцией $d^{-\gamma}$, где $\gamma = 3$. Однако, наблюдалось также расхождение с реальными графами, для которых $\gamma_{www} = 2.1$. Это означает, что хоть модель Боллобаша-Риордана и отражает некоторые свойства интернета, она должна быть видоизменена, чтобы лучше соответствовать реальности.

2.4 Модель Бакли-Остхуса

Возможный подход к такому видоизменению – это модель, независимо предложенная двумя группами исследователей. Они предложили расширить модель с помощью параметра, называемого *начальная аттрактивность вершины*. Это положительная константа, которая не зависит от степени. Позже Бакли и Остхус предложили явную конструкцию данной модели. Распределение степеней вершин в модели Бакли-Остхуса также подчиняется степенному закону, однако теперь варьируя значение параметра a в определении модели можно изменять значение γ результирующего графа.

Более строго, модель генерирует набор графов $H_{a,m}^n, n = 1, 2, \ldots$, с n вершинами и mn ребрами, где $m \in \mathbb{Z}$ – фиксированное число. Определение $H_{a,1}^n$ повторяет определение G_1^n с одним отличием, заключающимся в том, что вероятность нового ребра, добавляемого в $H_{a,1}^n$ равна

$$P(i = s) = \begin{cases} \frac{d_{H_{a,1}^{t-1}}(s)+a-1}{(a+1)t-1}, & \text{если } 1 \le s \le t-1, \\ \frac{a}{(a+1)t-1}, & \text{если } s = t, \end{cases}$$

Граф $H_{a,m}^n$ получается из графа $H_{a,1}^{mn}$ так же, как и G_m^n получается из G_1^{mn}. Заметим, что при $a = 1$ мы получаем изначальную модель Боллобаша-Риордана G_m^n. Для целых a Бакли и Остхус доказали, что распределение степеней вершин случайного графа в модели соответствует степенному закону с $\gamma = 2 + a$.

Глава 3

Previous work

A much longer LaTeX 2_ε example was written by Gil [?].

Глава 4

Results

In this section we describe the results.

4.1 Conclusions

We worked hard, and achieved very little.

Литература

[1] Степанов В. Е. О вероятности связности случайного графа gm(t) // Теория вероятностей и ее применения. 1970. Т. 15. № 1. С. 55–67.

[2] Степанов В. Е. Фазовый переход в случайных графах // Теория вероятностей и ее применения. 1970. Т. 15. № 2. С. 187–203.

[3] Степанов В. Е. Структура случайных графов gn(x|h) // Теория вероятностей и ее применения. 1972. Т. 17. № 3. С. 227–242.

[4] Колчин В. Ф. Случайные графы. М.: Физматлит, 2004.

[5] Bollobas B. Random Graphs. Cambridge: Cambridge Univ. Press, 2001.

[6] Алон Н., Спенсер Дж. Вероятностный метод. М: Бином. Лабо- ратория знаний, 2007.

[7] Janson S., Luczak T., Rucinski A. Random graphs. N.Y.: Wiley, 2000.

[8] Маргулис Г. А. Вероятностные характеристики графов с боль- шой связностью // Проблемы передачи информации. 1974. Т. 10. С. 101–108.

[9] Karp R. The transitive closure of a random digraph // Random structures and algorithms. 1990. V. 1. P. 73–94.

[10] Карлин С. Основы теории случайных процессов. М: Мир, 1971.

[11] Barabasi L.-A., Albert R. Emergence of scaling in random networks // Science. 1999. V. 286. P. 509–512.

[12] Barabasi L.-A., Albert R., Jeong H. Scale-free characteristics of random networks: the topology of the world-wide web // Physica A. 2000. V. 281. P. 69–77.

[13] Albert R., Jeong H., Barabasi L. A. Diameter of the world-wide web // Nature. 1999. V. 401. P. 130–131.

[14] Bollobas B., Riordan O. Mathematical results on scale-free random graphs. Handbook of graphs and networks. Weinheim: Wiley-VCH. 2003. P. 1–34.

[15] Райгородский А. М. Экстремальные задачи теории графов и ана- лиз данных. М.–Ижевск: НИЦ «РХД», 2009.

[16] Stoimenow A. Enumeration of chord diagrams and an upper bound for Vassiliev invariants // J. Knot Theory Ramifications. 1998. V. 7. N. 1. P. 93–114.

[17] Bollobas B., Riordan O. The diameter of a scale-free random graph // Combinatorica. 2004. V. 24. N. 1. P. 5–34.

[18] Bollobas B., Riordan O., Spencer J., Tusnady G. The degree sequence of a scale-free random graph process // Random Structures Algorithms. 2001. V. 18. N. 3. P. 279–290.

[19] Kumar R., Raghavan P., Rajagopalan S., Sivakumar D., Tomkins A., Upfal E. Stochastic models for the web graph // Proc. 41st Symposium on Foundations of Computer Science. 2000.