



Рис. 6. Фрагмент дорожной сети 3

В данном случае, учитывая аналогию прецедента c_3 , существует несколько вариантов планирования: либо снижение скорости транспортного средства на указанном участке, либо построение новой траектории маршрута (пунктирная линия на рис. 5). Новый маршрут оказался на 50 км длиннее предыдущего, однако позволяет сократить риск возникновения аналогичного прецедента.

В данной работе применен логический подход к отбору прецедентов при планировании перевозки. Прецеденты должны нести полезную информацию. Совместное применение ГИС с механизмом прецедентов позволяют повысить эффективность планирования и управления логистическим процессом. Подход на основе отбора прецедентов по методу аналогий помогает учитывать риски на этапе планирования транспортировки.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Логистика автомобильного транспорта: Учеб. пособие / В.С. Лукинский, В.И. Бережной, Е.В. Бережная и др. – М.: Финансы и статистика, 2004.
2. Транспортная логистика: Учебник для транспортных вузов / Под общей редакцией Л.Б. Миротина. – М.: Изд-во «Экзамен», 2002.
3. Курганов В.М. Логистика. Транспорт и склад в цепи поставок товаров. Учебно-практическое пособие. – М.: Книжный мир. – 2006.
4. Шелобаев. Экономико-математические методы и модели: Учеб. пособие для вузов. 2-е изд. перераб. и доп. – М.: ЮНИТИ-ДАНА, 2005.
5. Просветов Г.И. Математические методы в логистике: Учебно-методическое пособие. – М.: Изд-во РДЛ, 2006.
6. Сток Джеймс Р., Ламберт Дуглас М. Стратегическое управление логистикой. – М.: ИНФРА-М, 2005.
7. Логистика: учеб. пособие / Б.А. Аникин [и др.]; под ред. Б.А. Аникина, Т.А. Родкиной. – М.: ТК Велби, Изд-во Проспект, 2006. – 408 с.
8. Губин П. «Размышления на развилке у трех дорог». – «Логистик & система». 7/05.
9. Геоинформатика: Учеб. для студ. вузов / Е.Г. Капралов, А.В. Кошкарев, В.С. Тикунов и др.; Под ред. В.С. Тикунова. – М.: Издательский центр «Академия», 2005. – 480 с.
10. Берштейн Л.С., Боженьюк А.В. Нечеткие модели принятия решений: дедукция, индукция, аналогия. Монография. – Таганрог. Изд-во ТРТУ, 2001.

А.А. Целых

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ИНТЕРНЕТ-СИСТЕМ

Введение. Распределенные системы, в том числе сети сотовой связи, компьютерные сети и интернет, обладают сложной топологией и имеют в своей осно-

ве социальные самоорганизующиеся процессы, управляемые общими законами. Такие системы, состоящие из большого числа взаимодействующих элементов, могут демонстрировать интересные и весьма неожиданные временные и пространственные эффекты: устойчивость к ошибкам и внешним воздействиям, эффект массовой мобилизации и др.

Разработка адекватного математического аппарата для объяснения этих явлений ведется в рамках современной теории социальных сетей. Одним из подходов к анализу такого рода систем является теория «малого мира», обитателей которого отделяют друг от друга несколько «рукопожатий».

Вывод о том, что большинство членов сообщества малого мира имеют примерно одинаковое количество контактов, был впервые сделан в работах ученых Эрдоса и Реньи [1].

Даже с учетом растущей численности населения планеты, структура социальных сетей такова, что люди очень тесно связаны друг с другом. Два случайно выбранных человека, как правило, связаны весьма короткой цепочкой промежуточных знакомств.

По результатам «почтового» исследования ученого Милграма в 1967 г., средняя длина такой цепочки составила 5,5 звеньев. Из материалов, опубликованных журналом *Science* в 2003 г., следует, что гипотеза о «шести степенях» справедлива и для коммуникаций по электронной почте. Предметом другого исследования в 2000 г. стали свойства сети гиперссылок между документами в Интернете. Усредненная длина такой цепочки составила 19 промежуточных переходов для $N \sim 8 \cdot 10^8$ документов. В дальнейшем действует логарифмическая зависимость – при увеличении числа документов на порядок число переходов увеличится на два.

Модели случайных сетей. Современная теория случайных сетей постулирует случайность связей между вершинами сети.

Под *случайным графом* [2] будем понимать граф с заданным числом вершин, смежность которых определяется некоторым случайным образом.

Теория случайных графов использует вероятностные методы для исследования некоторых видов графов и определения ряда свойств «почти всех» графов. Таким образом, предметом исследования этой теории являются ожидаемые свойства графа.

Предположим, что имеется сеть из N документов. В среднем, каждый из них содержит z ссылок на другие документы в сети. Таким образом, во всей сети имеется $Nz/2$ связей между документами. Модель случайных сетей предполагает, что связи устанавливаются между случайно выбранными парами документов, вершины в которых соединены с вероятностью p .

При этом модель демонстрирует эффект малого мира. Усредненное расстояние между двумя вершинами в случайном графе растет как логарифм от числа вершин. Таким образом, средняя длина D цепочки, соединяющей два случайных документа, равна $D = \ln N / \ln z$. Поскольку $\ln N$ медленно увеличивается с ростом N , то D также мало даже для очень больших систем.

Согласно модели Эрдоса-Реньи, случайный граф состоит из N помеченных вершин, соединенных p ребрами, которые выбираются случайным образом из $N(N-1)/2$ возможных. Существует всего $C_{N(N-1)/2}^p$ графов с N вершинами и p ребрами, которые образуют вероятностное пространство с одинаковой вероятностью для каждой реализации.

Теория случайных графов изучает вероятностное пространство графов с N вершинами при $N \rightarrow \infty$, используя методы случайного анализа.

В процессе эволюции случайного графа, начиная с N изолированных вершин, последовательно добавляются случайные ребра. Графы, полученные на каждой последующей стадии эволюции, соответствуют все большим вероятностям p . В итоге эволюции при $p = 1$ приходим к полному графу с числом ребер $n = N(N-1)/2$.

Характерно, что многие важные свойства случайных графов проявляются неожиданно при достижении некоторой критической вероятности $p_c(N)$ [3]. До этого либо почти каждый граф обладает некоторым свойством Q , либо почти ни один граф им не обладает. Если $p(N)$ возрастает медленнее, чем $p_c(N)$ при $N \rightarrow \infty$, то почти ни один граф не будет обладать свойством Q . Если $p(N)$ возрастает быстрее, чем $p_c(N)$, почти любой граф будет обладать свойством Q .

Вероятность того, что граф с N вершинами и функцией распределения ребер $p = p(N)$ обладает свойством Q , определяется выражением

$$\lim_{N \rightarrow \infty} P_{N,p}(Q) = \begin{cases} 0, & \text{если } \frac{p(N)}{p_c(N)} \rightarrow 0 \\ 1, & \text{если } \frac{p(N)}{p_c(N)} \rightarrow \infty \end{cases}.$$

Проявление некоторых свойств связано с количеством вершин в графе. Так, свойство появления циклов при неизменном p скорее проявится в графах большего размера. Критическая вероятность наличия цикла порядка k составляет $p_c(N) = cN^{-1}$. С другой стороны, средняя степень графа в принципе имеет критическое значение, не зависящее от размера системы.

Диаметр графа. Определим диаметр графа как наибольшее расстояние между любыми двумя его вершинами. При этом диаметр несвязного графа с изолированными кластерами либо бесконечен, либо определяется как максимальный из диаметров его кластеров.

При малом p случайные графы имеют малый диаметр, поскольку количество вершин с расстоянием l от выбранной вершины зависит логарифмически только от количества вершин, то есть пропорционально $\ln(N)/\ln(\langle k \rangle)$.

Для большинства значений p практически все графы имеют сходный диаметр [3], который определяется выражением:

$$d = \frac{\ln(N)}{\ln(pN)} = \frac{\ln(N)}{\ln(\langle k \rangle)}.$$

Кластерность. Однако модель случайного графа не отражает в полной мере всех свойств реальных сетей. Во многих социальных сетях прослеживается тенденция к кластерности – существованию групп тесно связанных между собой элементов.

Можно ввести коэффициент кластерности C [4], представляющий собой среднюю долю таких пар соседних вершин узла, которые также являются соседями друг друга.

Выберем некоторую вершину i , имеющую K_i ребер, соединяющих ее с K_i вершинами. В случае, когда первые ближние соседи вершины являются частью клики, между ними существует $K_i(K_i-1)/2$ ребер.

Коэффициент кластерности вершины определяется как $C_i = 2E_i/(K_i(K_i-1))$, где E_i – число ребер между K_i вершинами. Коэффициент кластерности сети находится как сумма коэффициентов отдельных вершин.

В полностью связанной сети $C = 1$. В случайном графе коэффициент кластерности равен вероятности $C = p$. В реальных сетях коэффициент кластерности обыч-

но значительно больше, чем в случайных сетях с тем же количеством ребер и вершин – значение C хотя и много меньше единицы, однако существенно превосходит $1/N$.

Противоположностью случайным графам являются регулярные решетки с полностью упорядоченными связями каждого элемента с некоторым числом соседей. Легко видеть, что часть непосредственных соседей каждого элемента всегда связана друг с другом, благодаря чему обеспечиваются кластерные свойства сети. Однако из-за отсутствия «дальних» связей такая сеть не обладает эффектом малого мира.

Модель, узлы которой имеют одновременно некоторое количество локальных и случайных «дальних» связей [5], демонстрирует как эффект малого мира, так и кластеризацию.

Распределение степеней. Легко заметить, что не все вершины сети имеют одинаковое количество ребер. Степень вершины характеризуется функцией $P(k)$, которая определяет вероятность того, что случайно выбранная вершина будет иметь ровно k ребер. В случайном графе большая часть вершин имеет приблизительно равную степень, близкую к средней степени $\langle k \rangle$ сети.

Считается, что распределение степеней вершин случайного графа является распределением Пуассона с пиком в $P(\langle k \rangle)$. Однако, согласно эмпирическим результатам [6], для большинства реальных сетей распределение степеней значительно отличается от распределения Пуассона и является степенным:

$$P(k) \approx k^{-\alpha}.$$

Для модели «голливудовского графа» характерна разреженность, высокий коэффициент кластеризации и малый диаметр. Вероятность того, что страница имеет гиперссылку к k страницам, составляет $k^{-2.45}$, а вероятность того, что m страниц имеют ссылку к данной странице, равняется $m^{-2.1}$.

Теория «галстука-бабочки» предполагает, что Интернет – сильно направленная сеть, разделенная на крупные области, приблизительно равные по числу страниц.

С любой страницы сильносвязанного ядра, узла галстука-бабочки, можно попасть на любую другую страницу ядра. Таким образом, посредством гиперссылок пользователи могут легко перемещаться между такими сайтами в любом направлении. Одна сторона «бабочки» содержит «стартовые» страницы, с которых доступны все страницы ядра, но не наоборот – это база графа. Другая сторона содержит «конечные» страницы, которые могут быть достигнуты из страниц ядра, но не имеют обратных ссылок на него – антибаза графа.

Заключение. Таким образом, на сегодня существуют три основных подхода к математическому моделированию социальных сетей. Случайные графы широко применяются для моделирования и эмпирических исследований. Класс моделей малого мира опирается на базовое понятие кластерности. Модели без масштабирования объясняют происхождение степенного распределения степеней вершин и прочих отклонений от распределения Пуассона высокой интерактивностью и динамикой интернет-систем и позволяют весьма достоверно описать почти все особенности структуры Интернета.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Erdos, P., Renyi, A.* On Random Graphs // *Publ. Math. Debrecen*, Vol. 6, 1959. Pp. 290–297.
2. *Gross, J., Yellen, J.* Graph Theory and Its Applications. Chapman & Hall/CRC, 2005.
3. *Колчин В.Ф.* Случайные графы. – М.: Физматлит, 2004.
4. *Watts, D., Strogatz, S.* Collective Dynamics of Small-World Networks // *Nature*, Vol. 393, 1998. Pp. 440–442.

5. Kleinberg, J. The small-world phenomenon: An algorithmic perspective // Proc 32nd Symposium on Theory of Computing, 2000.
6. Reka, A., Jeong, H., and Barabasi, A. Diameter of the World Wide Web // *Nature*, Vol. 401, 1999. Pp. 130–131.

Е.Е. Краснощеков

**ПОИСК РЕЛЕВАНТНЫХ ДОКУМЕНТОВ
В ЭКОНОМИЧЕСКИХ СИСТЕМАХ ПО ЗАПРОСУ,
РАСШИРЕННОМУ ПАРАДИГМАТИЧЕСКИМИ ОТНОШЕНИЯМИ**

В статье рассмотрены технология адаптивного распознавания образов APRP и общие подходы к организации оптимального поиска с использованием нечеткого сравнения. Проанализированы существующие подходы к решению задач поиска и предложено их улучшение. Рассмотрено применение при поиске автоматически построенных ассоциативных отношений вместо парадигматических отношений, построенных вручную. Проведена экспериментальная оценка метода поиска по расширенному запросу.

Введение

Объем электронных документальных баз данных в экономических системах за последние годы значительно возрос. В настоящее время значительная часть экономических документов доступна в глобальной сети Интернет, которая предоставляет не только почти неограниченные возможности для размещения данных, но и возможности для доступа к документам из любой точки земного шара при условии наличия компьютера, подключенного к «всемирной паутине».

Количество электронных документов в сети Интернет настолько велико, что обнаружить необходимый ресурс без «путеводителя» почти невозможно. Сегодня роль путеводителя и справочника в Интернете играют поисковые системы.

В подобной ситуации резко возросла потребность в системах поиска и анализа данных. Именно поисковая система определяет, превратятся ли в знания многочисленные разрозненные данные, поступающие по различным каналам связи и накапливаемые в разнообразных государственных, ведомственных, частных и прочих электронных архивах.

Поиск документов можно отнести к наиболее важным задачам содержательной обработки текстовой информации, что, в частности, обусловлено потребностью поиска информации в сети Интернет.

Разработка методов текстового поиска имеет давнюю историю, насчитывающую более сорока лет [1]. За это время информационно-поисковые системы эволюционировали от систем формально-логического типа [2] к системам нечеткого поиска, основными особенностями которых являются следующие [3]:

- запрос задается на естественном языке, а не в виде формального выражения булево-контекстного типа;
- некоторые или даже все найденные документы содержат только часть информативных слов запроса;
- найденные документы выдаются в ранжированном виде, т. е. в порядке убывания их соответствия запросу.