

УДК 004.942

СЛУЧАЙНЫЕ ГРАФЫ КАК МОДЕЛЬ СРЕДЫ РАСПРОСТРАНЕНИЯ И ВЗАИМОДЕЙСТВИЯ САМОРАЗМНОЖАЮЩИХСЯ ОБЪЕКТОВ

© 2015 К.Е. Климентьев

Самарский государственный аэрокосмический университет им. академика С.П. Королева
(национальный исследовательский университет)

Рассматриваются аспекты построения и применения некоторых разновидностей случайных графов в качестве модели среды распространения и взаимодействия саморазмножающихся объектов, таких как компьютерные вирусы и черви, болезнетворные микроорганизмы и т.п.

Ключевые слова: самовоспроизводящийся объект, компьютерный вирус, сетевой червь, распространение, взаимодействие, имитационная модель, случайная сеть, случайный граф, коэффициент кластеризации, вероятностное распределение.

Климентьев Константин Евгеньевич, кандидат технических наук, доцент кафедры информационных систем и технологий. E-mail: climentieff@ro.ru

RANDOM GRAPHS AS MODEL OF PROPAGATION AND INTERACTION AREA OF SELFREPRODUCTIVE OBJECTS

© 2015 K.E. Klimentiev

Samara State Aerospace University named after academician S.P. Korolyov
(national research university)

Some types of random graphs are discussed. Also, some examples of graph building are presented.

Key words: selfreproductive object, computer virus, network, spreading, interaction, simulation, homogeneous graph, random graph, random geometrical graph, ad hoc network, cluster coefficient, probability distribution.

Klimentiev K.E., Candidate of technics, an Associate professor of Information systems and technologies Department. E-mail: climentieff@ro.ru

Введение. Силами преподавателей и студентов факультета информатики СГАУ ведется работа над реализацией системы имитационного моделирования распространения и взаимодействия саморазмножающихся объектов. Под саморазмножающимися объектами в данной работе понимаются сущности реального мира, способные к обладанию свойством саморепликации, причем это свойство передается от объекта к объекту в результате их взаимодействия. Примерами являются:

- инфицирование живых существ болезнетворными микроорганизмами [1];
- размножение компьютерных вирусов и сетевых червей [3];
- передача информационных сообщений между людьми и устройствами;
- и даже распространение зомби [5].

В качестве моделей среды распространения и взаимодействия объектов традиционно принимаются графы различных видов (см. табл. 1).

Таблица 1 - Графовые модели среды обитания саморазмножающихся объектов

Объект	Вид графа
Сетевые черви	Полный граф
Файловые черви и вирусы, почтовые черви, сообщения в социальных сетях	Безмасштабные (scale free) случайные графы с экспоненциальным распределением степеней вершин [4]
Блезнетворные микроорганизмы, мобильные вирусы, сообщения в сетях сотовой связи	Случайные графы с неэкспоненциальным распределением степеней вершин [4]

В данной работе речь пойдет о графах, используемых для моделирования эпидемий «мобильных червей» [3] и, возможно, воздушно-капельных инфекций среди высших животных [1]. Эти эпидемии характеризуется следующими особенностями:

- возможностью передачи свойства саморепликации лишь в пределах ограниченного «радиуса действия» (например, на расстояниях не более 10-15 м. по протоколу Bluetooth);
- изменчивостью топологии среды в результате перемещения объектов в пространстве;
- для случая «мобильных червей» - детерминированным алгоритмом поведения, содержащим, возможно, элементы стохастизма [3].

1. Постановка задачи. Класс сетей, служащий для моделирования среды существования таких объектов, носит наименование «специального» (ad hoc) и представляет собой множество случайных графов с различной топологией [4]. Задачи изучения процессов распространения и взаимодействия саморазмножающихся объектов близки к множеству задач, рассматриваемых в теории перколяции.

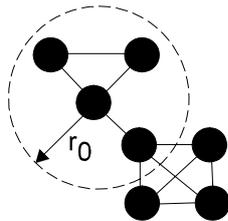
Общими характеристиками таких графов, влияющими на динамику распространения и взаимодействия объектов, являются:

- вероятностное распределение степеней k_i вершин;
- вероятностное распределение локальных коэффициентов кластеризации c_i вершин.

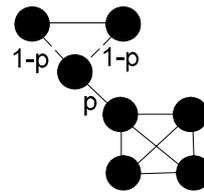
Под локальным коэффициентом кластеризации вершины обычно понимают долю «треугольников» данной вершины, образованных из «соседей» данной вершины и являющихся «соседями» для друг друга, в общем количестве потенциально возможных «треугольников» $c_{\max} = k_i(k_i - 1) / 2$, где k_i - степень i -ой вершины графа.

Важными частными характеристиками таких графов являются средняя степень вершин $\bar{k} = \frac{1}{N} \sum_{i=1}^N k_i$ и средний коэффициент кластеризации $\bar{c} = \frac{1}{N} \sum_{i=1}^N c_i$, где i – номер вершины; N – количество вершин в графе.

Наиболее близкими абстракциями для сетей, моделирующих среду обитания «близкодействующих» саморазмножающихся объектов, являются случайный граф Радо (он же граф Эрдеша-Реньи) и «геометрический» случайный граф. Они различаются методами построения.



а) «Геометрический» граф



б) Граф Радо (Эрдеша-Реньи)

Рис. 1. Случайные графы

Целью исследования, положенного в основу данной статьи, является изучение возможностей построения заданных типов случайных графов с заданными средней степенью вершины \bar{k} и средним коэффициентом кластеризации \bar{c} .

2. Случайный граф Радо (RRG). Он получается из «полного» графа, каждое ребро которого остается с заранее выбранной вероятностью p и удаляется с вероятностью $1-p$ (см. рис. 1,б). Степени k_i всех вершин в таком графе примерно одинаковы, более точно - их количество подчиняется биномиальному распределению $P(k_i = k) = NC_{N-1}^k p^k (1-p)^{N-1-k}$ со средней степенью вершины $\bar{k} = p \times N$, где N – количество вершин [4]. Очевидно, для таких графов средний коэффициент кластеризации $\bar{c} = p$.

3. Случайный «геометрический» граф (RGG). В построении графа со случайной геометрией (RGG) участвуют «координаты» узлов. Например, если считать, что граф размещен в единичном квадрате, тогда каждому узлу приписываются координаты, представляющие собой равномерно распределенные на интервале $[0..1]$ случайные величины. Затем для каждой пары узлов с индексами i и j рассчитывается «расстояние» r_{ij} , и ребро между узлами проводится в том случае, если расстояние между ними не превышает заранее назначенного «радиуса» r_0 (см. рис. 1,а).

3.1. Получение желаемой средней степени вершин. Для того, чтобы при построении случайного «геометрического» графа получать желаемое значение \bar{k} , необходимо изучить распределение величины $r = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, где (x_i, y_i) и (x_j, y_j) – координаты любых двух случайных точек в «единичном квадрате». Более общая задача решена, например, в работе [7], но там отсутствует необходимое выражение, которым можно было бы воспользоваться при генерировании графов. Поэтому пользуясь известными соотношениями теории вероятности [2], самостоятельно получены следующие результаты.

Расстояние r между двумя случайными точками в «единичном квадрате» подчиняется закону с плотностью:

- $f(r) = 2r^3 - 8r^2 + 2\pi r$, если обе точки внутри «единичного круга»;
- $f(r) = -2r^3 + 2r^2\sqrt{r^2 - 1} + r(4 - 2\pi - 8 \arctan(\sqrt{r^2 - 1}))$, иначе;

и функцией распределения:

- $F(r) = 0.5r^4 - \frac{2}{3}r^3 + \pi r^2$, если обе точки внутри «единичного круга»;
- $F(r) = 1 - \frac{1}{3}(2r^2 + 1)\sqrt{r^2 - 1} + (\pi - 2)r^2 + \frac{1}{3}$, иначе.

Математическое ожидание расстояния $m(r) = \frac{1}{15}(2 + \sqrt{2} + 5 \arcsin(1)) \approx 0.52$.

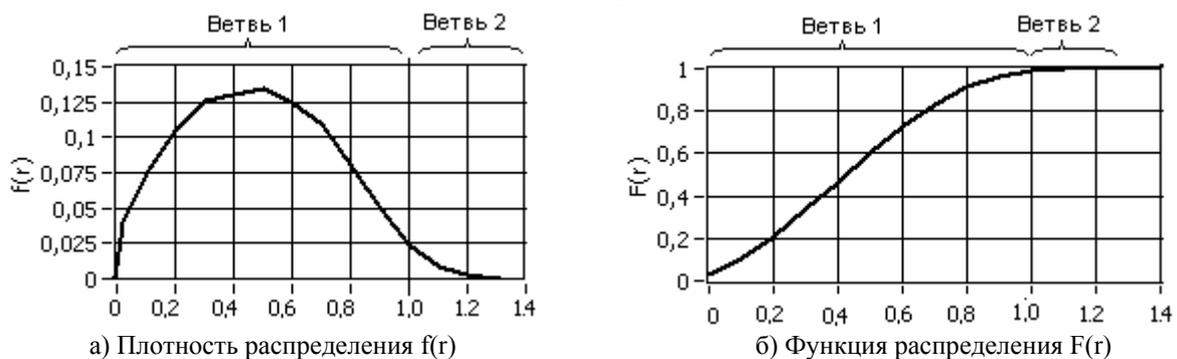


Рис 2. Распределение расстояния между двумя случайными точками в единичном квадрате

Таблица 2. – Числовые значения $F(r)$

r	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0	1.1	1.2	1.3	1.4
$F(r)$	0	.028	.105	.214	.344	.483	.621	.746	.852	.932	.980	.982	.984	.990	.997

Очевидно, средняя степень вершин соответствующего графа зависит от r_0 :
 $\bar{k} = (N-1) \times P(r < r_0) = (N-1) \times F(r_0)$.

3.2. Получение желаемого среднего коэффициента кластеризации. Для того, чтобы при построении случайного «геометрического» графа получать желаемое значение \bar{c} , необходимо изучить распределение величины

$$r = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (1)$$

где (x_i, y_i) и (x_j, y_j) – координаты любых двух случайных точек в круге радиусом r_0 . В случае $r \leq r_0$ вершины, являющиеся «соседями» заданной вершины, являются так же «соседями» друг для друга.

Выражение для плотности распределения можно найти в [6]:

$$g(r) = \frac{4r}{\pi r_0^2} \left(\arccos\left(\frac{r}{2r_0}\right) - \frac{r}{2r_0} \sqrt{1 - \frac{r^2}{4r_0^2}} \right).$$

Выражение для функции распределения получено самостоятельно:

$$G(r) = \frac{1}{\pi r_0^3} \left(2r^2 r_0 \arccos(r/2r_0) - \frac{1}{4} (r^3 + 2rr_0^2) \sqrt{4 - r^2/r_0^2} + 2r_0^3 \arcsin(r/2r_0) \right).$$

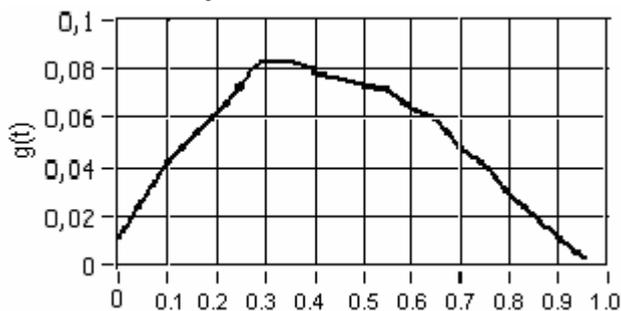
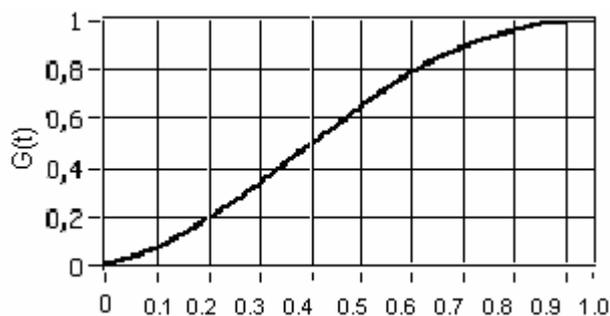
а) Плотность распределения $f(r)$ б) Функция распределения $F(r)$

Рис 3. Распределение расстояния между двумя случайными точками в единичном круге

Вероятность образования «треугольника» из заданной точки и двух случайных, находящихся на расстоянии меньше r_0 от заданной, составляет

$$G(r_0) = \frac{1}{\pi} \left(\arccos\left(\frac{1}{2}\right) - \frac{\sqrt{3}}{4} + \arcsin\left(\frac{1}{2}\right) \right) = 1 - \frac{3\sqrt{3}}{4\pi} \approx 0.59. \quad \text{Таким образом, как только что}$$

показано, в «геометрических» случайных графах коэффициент кластеризации есть константа, и это обстоятельство совпадает с выводом, полученным в [8].

Однако данный вывод справедлив лишь при следующих ограничениях:

- 1) $0 < r_0 \leq 0.5$, поскольку при больших r_0 любые вписанные круги начинают превышать размеры единичного квадрата;
- 2) $N \rightarrow \infty$.

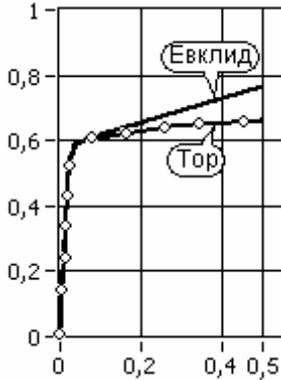
Несоблюдение второго ограничения приводит к существенным отклонениям $G(r_0)$ от теоретического значения, более того, при малых N поведение коэффициента кластеризации описывается монотонно возрастающей функцией (см. рис. 4). Разумеется, при увеличении N наблюдается приближение коэффициента к теоретическим значениям (см. рис. 4, б). Кроме того, в некоторой степени избавиться от «эффекта малого N » позволяет изменение алгоритма построения «геометрического» случайного графа, заключающееся в отказе от Евклидовой метрики (1) в пользу «тороидальной»:

$$r = \min\left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \sqrt{\Delta x^2 + \Delta y^2}\right), \text{ где} \quad (2)$$

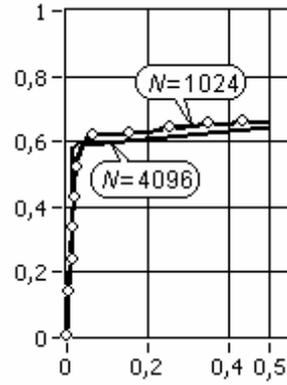
$\Delta x = x_0 + 1 - x_1$ при $x_0 \leq x_1$, $\Delta x = x_1 + 1 - x_0$ иначе;

$\Delta y = y_0 + 1 - y_1$ при $y_0 \leq y_1$, $\Delta y = y_1 + 1 - y_0$ иначе.

Сравнительные результаты применения различных метрик, полученные при помощи численных экспериментов, приведены на рис. 4,а.



а) Зависимость \bar{c} от вида метрики при $N=1024$

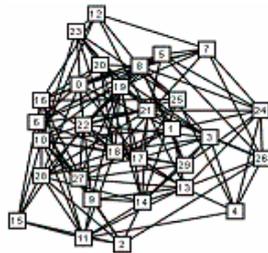


б) Зависимость \bar{c} от размера графа N (метрика «торидальная»)

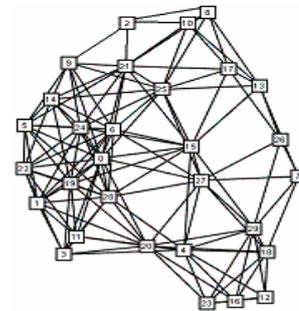
Рис. 4. Коэффициент кластеризации RGG при малых N

Полученные в результате экспериментов числовые значения зависимостей при разных r_0 и N позволяют строить графы с любым желаемым \bar{c} .

4. Примеры использования полученных соотношений и результатов. В качестве первого примера применения полученных соотношений приведем результаты генерации случайных графов типов RRG и RGG с $N=30$ вершинами и ожидаемой средней степенью вершин $\bar{k} = 10$.



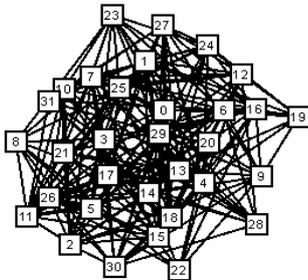
а) Граф типа RRG, $p=0.33$, $\bar{k} \approx 9.13$



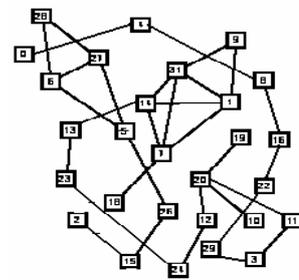
а) Граф типа RGG, $r_0=0.39$, $\bar{k} \approx 9.46$

Рис. 5. Графы разных типов с одинаковой средней степенью вершин

Второй пример – результаты генерации случайных графов типа RRG и RGG (метрика «торидальная») с $N=32$ вершинами и ожидаемым средним коэффициентом кластеризации $\bar{c} = 0.44$.



а) Граф типа RRG, $p=0.45$, $\bar{c} \approx 0.45$



а) Граф типа RGG, $r_0=0.012$, $\bar{c} \approx 0.44$

Рис. 6. Графы разных типов с одинаковым средним коэффициентом кластеризации

Интересно, что во втором случае желаемое значение среднего коэффициента кластеризации \bar{c} достигнуто при на порядок меньшем значении средней степени вершин \bar{k} .

Изучение влияния \bar{c} и \bar{k} на динамику распространения и взаимодействия саморазмножающихся объектов послужит целью дальнейших исследований.

Литература

1. Братусь А.С., Новожилов А.С., Платонов А.П. Динамические системы и модели биологии. - М.: Физматлит, 2010. – 400 с.
2. Вентцель Е.С., Овчаров Л.А. Прикладные задачи теории вероятностей. – М.: Радио и связь, 1983. – 416 с.
3. Климентьев К.Е. Компьютерные вирусы и антивирусы: взгляд программиста. – М.: ДМК-Пресс, 2013. – 656 с.
4. Райгородский А.М. Модели случайных графов и их применения // М.: Труды МФТИ. – 2010. – Том 2, №4. – С. 130-140.
5. Smith R., Munz P., Hudea I, Imad J. When Zombies Attack: Mathematical Modelling of an Outbreak of Zombie Infection // Infectious Disease Modelling Research Progress. - Nova Science Publishers, Inc. – pp. 133-157.
6. Mathai A.M. An introduction to geometrical probability: distributional aspects with applications. - CRC Press, 1999. - 576 pp.
7. Philip J. The probability distribution of the distance between two random points in a box // TRITA MAT 7(10), 2007. – 13 pp.
8. Dall J., Christensen M. Random geometric graphs / Phys. Rev., vol.66, no 016121, 2002.