

КОЛИЧЕСТВЕННЫЕ КРИТЕРИИ КЛАССИФИКАЦИИ СЕТЕЙ МАЛОГО МИРА ДЛЯ ПРОМЫШЛЕННЫХ СИСТЕМ¹

Семенков К.В.

Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия
semenkovk@ipu.ru

Промыслов Г.В.

*Национальный исследовательский университет “Высшая школа экономики”,
Москва, Россия*
georgypr@yahoo.com

Аннотация. В работе рассмотрена возможность распознавания систем имеющих топологию сетей малого мира без детального анализа их топологических свойств. Для этого используются методы распознавания образов и машинного обучения. Проведенное исследование позволяет сделать вывод, что распознавания с обучением дает более высокие результаты правильной классификации сетей.

Ключевые слова: сети малого мира, критерии классификации, графы.

Введение

Сети малого мира известны давно. Впервые они стали широко известны благодаря эксперименту американского психолога Милгрема [1], где случайно выбранные добровольцы получали задание переправить письмо случайно выбранным незнакомым людям, привлекая для пересылки только своих друзей и знакомых. Оказалось, что письма в большинстве своем дошли до адресатов и что по пути следования письма обычно проходили через 5-6 посредников, поэтому результаты эксперимента Милгрема часто называют «правилом шести рукопожатий».

С формальной точки зрения, сети малого мира — это такие сети, в которых среднее геодезическое расстояние между узлами l есть величина порядка логарифма от количества n узлов в сети: $l = O(\log n)$. Дополнительным критерием наличия у сети свойств малого мира является достаточно большое значение глобального коэффициента.

Новый всплеск интереса к сетям малого мира случился в конце 90-х годов XX века и начале XXI века и связан с бурным развитием Интернета, быстрым ростом глобальной сети и появлением социальных сетей. Была создана модель малого мира Уотса-Строгаца [2]; Барабаши и Альберт [3] создали модель роста масштабно-инвариантной сети, а Боллобаш и Риордан [4] доказали её ключевые свойства, дав таким образом математические методы исследования социальных сетей. В работе [5] сформулирована проблема навигации в сетях малого мира сформулирован ряд условий возможности децентрализованной навигации для сетей определенного вида.

С другой стороны, развитие технологий телекоммуникаций сделало обыденностью создание беспроводных сетей произвольной топологии на базе небольших устройств (датчиков, устройств интернета вещей), а также устойчивых сетей между подвижными объектами. Давно и широко применяются стандарты IEEE 802.11 и 802.15.4 [6, 7] для различных беспроводных сетей, в том числе беспроводных сетей ячеистой топологии (mesh-сетей) [8]. Появившиеся технологические возможности позволяют рассматривать практические задачи о создании технических, промышленных систем, систем мобильных устройств, обладающих свойствами малого мира [9, 10].

Однако в имеющихся работах вопрос о критериях отнесения системы к сетям малого мира не рассматривается, а принадлежность системы к искомому классу просто неявно постулируется. Мы же полагаем, что при создании технической, промышленной системы необходимо определить измеримые критерии проверки тех или иных свойств данной системы. Настоящая работа посвящена вопросу задания количественных критериев отнесения промышленной системы информационно связанных устройств к сетям малого мира и их распознавания без детального анализа их топологических свойств с помощью методов распознавания образов и машинного обучения.

¹Исследование выполнено за счет гранта Российского научного фонда № 23-19-00338, <https://rscf.ru/project/23-19-00338/>

1. Основные определения и ограничения системы

Стандартной формой представления сетей является граф. В зависимости от задачи, он может быть ориентированным, неориентированным, раскрашенным, с весами вершин и ребер и т.д. В данной работе мы будем рассматривать только «обычные» неориентированные графы без петель и кратных ребер.

Напомним, что в графе расстоянием, или геодезическим расстоянием между двумя вершинами называется количество ребер на кратчайшем пути, соединяющем эти вершины. Если путь между вершинами отсутствует, то расстояние между вершинами считается бесконечным. Соответственно, средним расстоянием между вершинами связного графа называется величина

$$l = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i>j} d_{ij}$$

где d_{ij} — расстояние между вершинами i и j . Если граф содержит несколько компонент связности, то будем брать средневзвешенное значение.

Глобальный коэффициент кластеризации характеризует транзитивность графа — свойство, когда две вершины, смежные данной, также смежны между собой. Следуя [11], будем пользоваться следующим определением глобального коэффициента кластеризации:

$$C = \frac{6N_{\Delta}}{N_2} ,$$

где N_{Δ} — количество треугольников в графе, а N_2 — количество путей длины 2. Из определения, в частности, следует, что в полном графе $C = 1$, а в графе, состоящем из квадратных ячеек, $C = 0$.

Очевидно, что эффект малого мира на малом количестве узлов не проявляется. С другой стороны, в технических системах количество объектов не может быть слишком большим, в отличие от социальных сетей и веб-графов, поэтому нас будут интересовать сети с числом узлов порядка $10^3 - 10^4$.

В работе мы поставили две задачи: во-первых, можно ли сказать, что сеть принадлежит к малому миру, не исследуя детально свойства графа, и, во-вторых, каковы критерии отнесения сети к малому миру для промышленных, инженерных систем.

Для проведения распознавания использовались следующие методы:

Первый метод, представляющий задачу кластеризации без обучения [12], состоит из трех подзадач. Сначала набор данных представлялся в виде матрицы координат приведенного к единому масштабу множества графов на плоскости. Для каждой пары графов проводилось выполнение прокрустового преобразования и построение матрицы расстояний между полученными кривыми, к которой затем применяется метод главных компонент, после чего полученные вектора проекций разделялись на два кластера методом k-средних.

Второй метод относится к методам классификации с обучением и реализован на основе метода опорных векторов [12].

2. Результаты распознавания

Результаты распознавания для каждого из методов представлены на рисунках 1-4. Каждый рисунок содержит три графика, каждый из которых соответствует заданному коэффициенту перераспределения. По оси x указано количество точек, заданное в эксперименте, по оси y — процент правильно классифицированных графов малого или не малого мира.

Рисунки 1-2 содержат результаты распознавания методом кластерного анализа. Из рисунка 1 видно, что метод плохо отделил графы малого мира от остальных. Для всех исследованных сочетаний коэффициента перераспределения и количества вершин, степень разделения графов, относящихся к малому миру, в среднем составила 60%, где хуже всего распознали графы с коэффициентом 0.5, за исключением единичного случая, тогда как для остальных коэффициентов правильность распознавания была на 20% лучше.

Из рисунка 2 можно заключить, что чем ниже коэффициент перераспределения, тем с большей уверенностью можно отличить граф, не принадлежащий к малому миру. Доля графов, не принадлежащих к малому миру во втором кластере, варьируется от 65% до 95% в зависимости от коэффициента и количества вершин в графе.

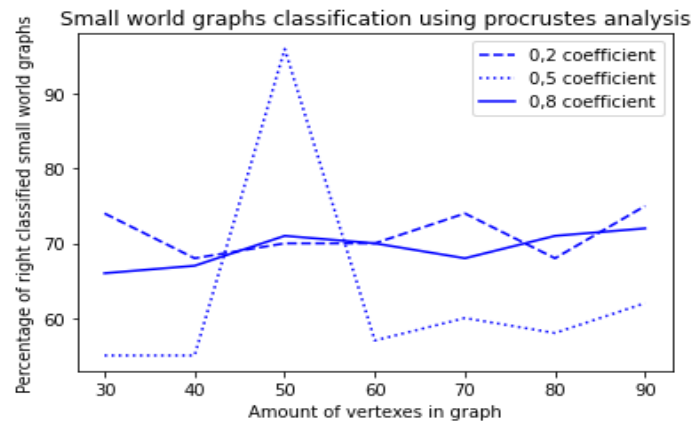


Рис. 1. Результаты классификации методом кластерного анализа графов малого мира

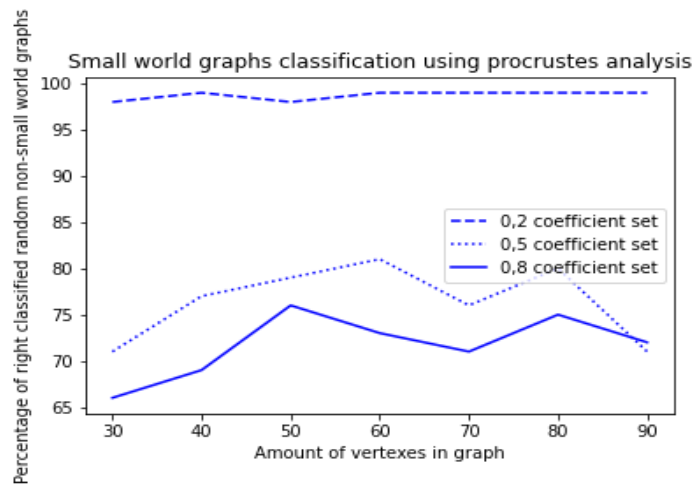


Рис. 2. Результаты классификации методом кластерного анализа графов не малого мира

Рисунок 3 содержит результаты распознавания графов малого мира методом с обучением. Средний результат для небольших коэффициентов составил примерно 90% правильно распознанных графов, тогда как для высокого коэффициента наблюдалось падение точности с ростом количества вершин в графе, на 25% по результатам экспериментов.

Из рисунка 4 можно заключить, что как очень низкий, так и очень высокий коэффициент перераспределения ведут к усилению различий с обычными графами: процент правильно классифицированных графов составил примерно 95% во всех случаях. В свою очередь средний коэффициент привел только к 65% точности распознавания графов.

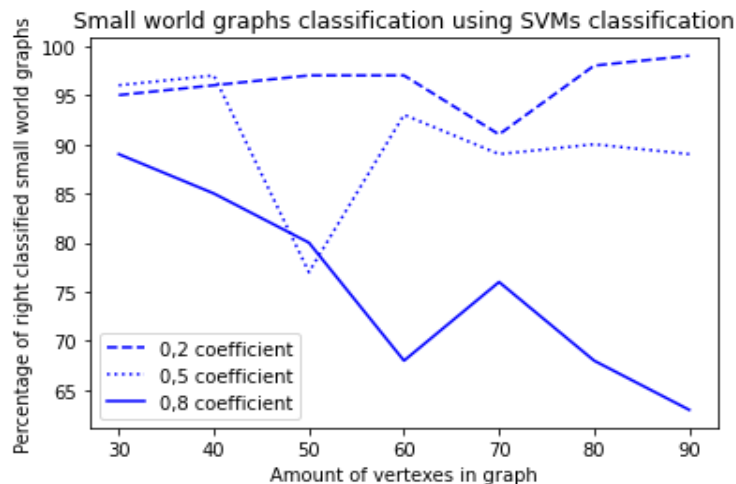


Рис. 3. Результаты классификации методом опорных векторов графов малого мира

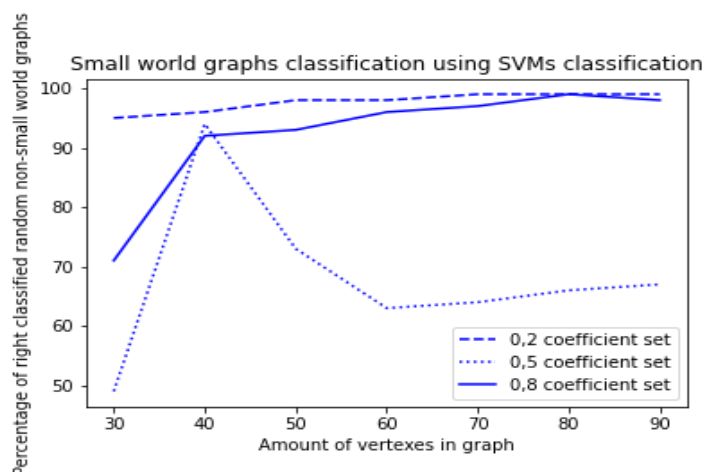


Рис. 4. Результаты классификации методом опорных векторов графов не малого мира

3. Результаты исследования количественных критериев графов Уоттса-Строгаца

На рисунках 5 и 6 представлены усредненные значения основных свойств графов, определяющих принадлежность к малому миру: средней длины пути и глобального коэффициента кластеризации. Значения не нормированы и являются средневзвешенными по всем компонентам связности (рисунок 7) случайно сгенерированных графов (пример на рисунке 8)

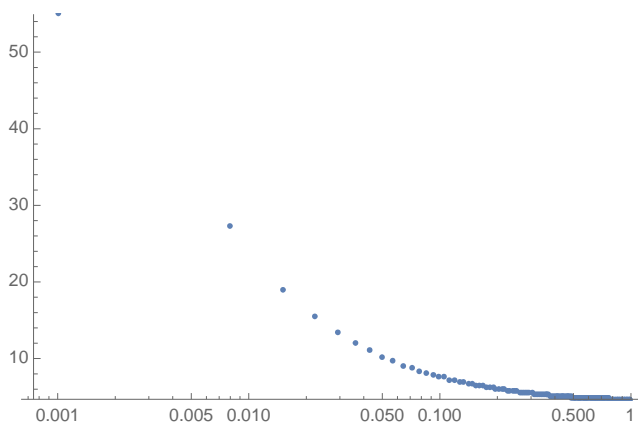


Рис. 5. Зависимость $L(p)$ от p . $L(p)$ – среднее расстояние между вершинами при данном коэффициенте p

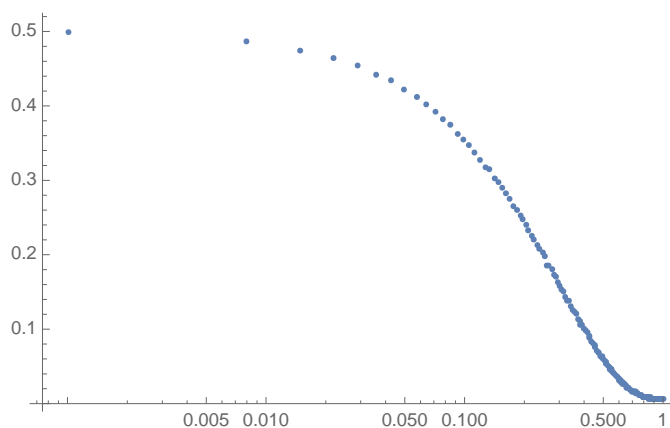


Рис. 6. Зависимость $C(p)$ от p . $C(p)$ – глобальный коэффициент кластеризации при данном коэффициенте p

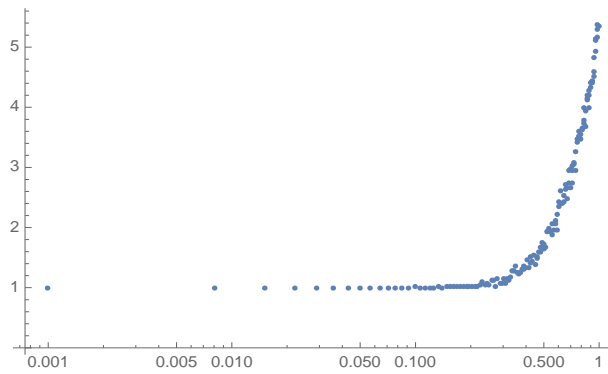


Рис. 7. Количество компонент связности в графе в зависимости от коэффициента p

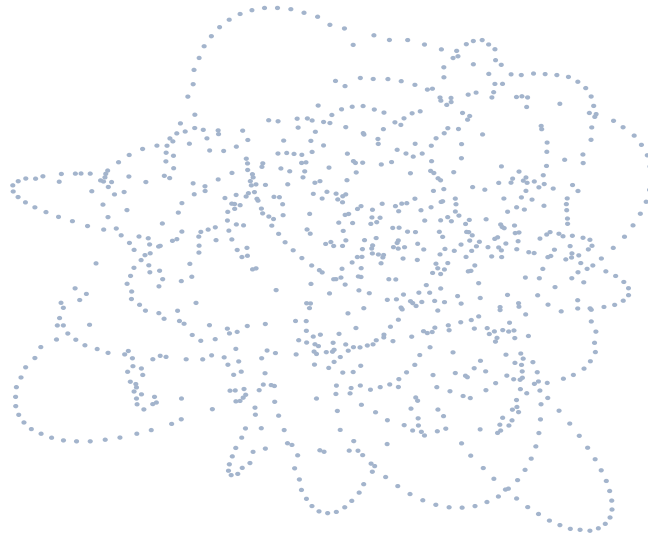


Рис. 8. Граф малого мира, сгенерированный методом Уоттса-Строгаца. 900 вершин, коэффициент перепутывания 0.05

4. Результаты исследования количественных критериев ячеистых сетей

На рисунках 9 и 10 представлены усредненные значения основных свойств ячеистых графов, определяющих принадлежность к малому миру: среднего расстояния между вершинами и глобального коэффициента кластеризации. Значения не нормированы и являются средневзвешенными по всем компонентам связи (рисунок 11) случайно сгенерированных графов (пример на рисунке 12)

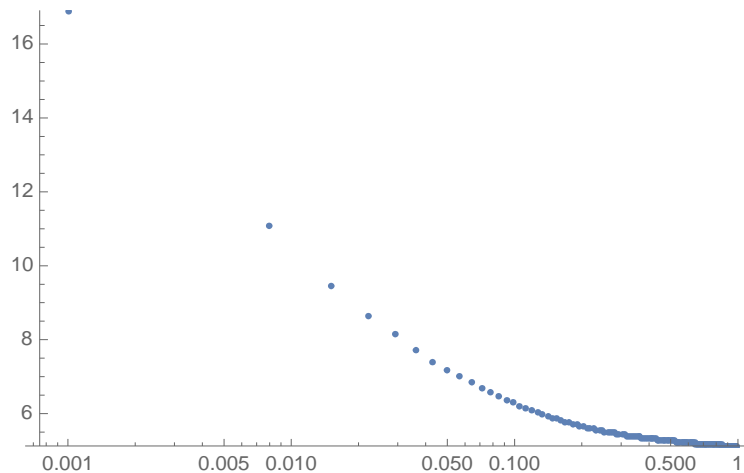


Рис. 9. Зависимость $L(p)$ от p . $L(p)$ – среднее расстояние между вершинами при данном коэффициенте p

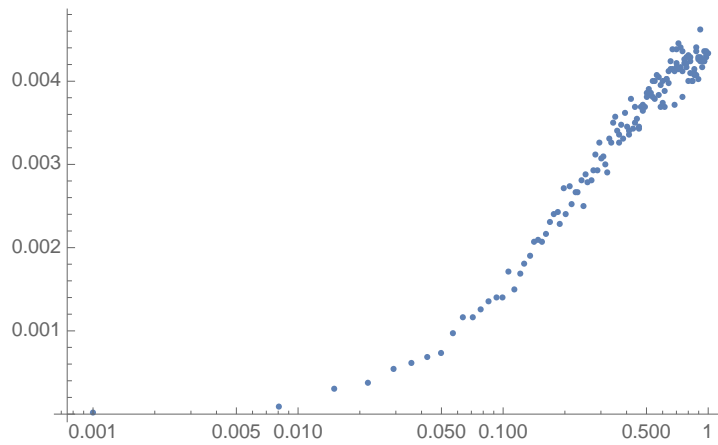


Рис. 10. Зависимость $C(p)$ от p . $C(p)$ – глобальный коэффициент кластеризации при данном коэффициенте p

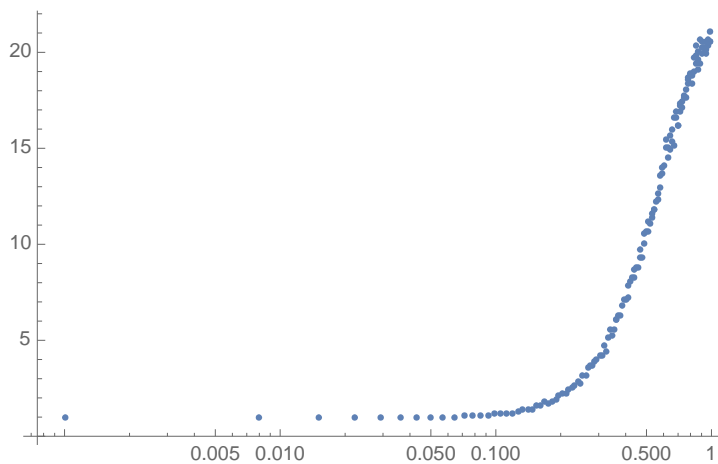


Рис. 11. Количество компонент связности в графе в зависимости от коэффициента p



Рис. 12. Ячеистая сеть, 900 вершин, коэффициент перепутывания 0.01

Можно видеть, что по сравнению с “классическим” графом малого мира в модели Уотса-Строгаца ячеистые графы, являясь по основному свойству (уравнение 1) и рисунок 9 графами малого мира, по дополнительному свойству глобального коэффициента кластеризации имеют характеристики фактически обратные к свойствам графов в модели Уотса-Строгаца (рисунки 6 и 10 соответственно). Это объясняется особенностью определения кластера как треугольника, используемого в расчете глобального коэффициента кластеризации. Поэтому данное отличие стоит учитывать при анализе сетей, описывающих реальные технические системы.

5. Заключение

Проведенное исследование с применением методов распознавания без обучения и с обучением позволяет сделать вывод, что, хотя существуют отдельные случаи, показывающие хороший результат, среднее качество работы метода составляет около 70%.

Метод распознавания с обучением также не дает очень высоких результатов, тем не менее средний результат во всех случаях был на 10-20% лучше аналогичного в кластерном анализе. Как показало проведенное моделирование, при классификации сетей, описывающих реальные технические системы, следует с осторожностью использовать глобальный коэффициент кластеризации в качестве признака сети малого мира, так как он зависит от особенности определения кластера.

Литература

1. *Milgram S.* The Small World Problem // *Psychology Today*. – 1967. – Vol. 1, N. 1. – P. 61-67.
2. *Watts D. J., Strogatz S. H.* Collective dynamics of ‘small-world’ networks // *Nature*. – 1998. – Vol. 393, N 6684. – P. 440–442.
3. *Barabási A.-L., Albert R.* Emergence of Scaling in Random Networks // *Science*. – 1999. – Vol. 286, Iss. 5439. – P. 509-512.
4. *Bollobás B., Riordan O.* The Diameter of a Scale-Free Graph // *Combinatorica*. – 2004. – Vol. 24, Iss. 1. – P. 5-34.
5. *Kleinberg J.* The small-world phenomenon: an algorithmic perspective // *Proc. of the 32nd Annual ACM Symposium on Theory of Computing (STOC00)*. – Portland: ACM, 2000. – P. 163-170.
6. *IEEE.* IEEE Standard 802.11-2020: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. – IEEE, 2021. 4379 с.
7. *IEEE.* IEEE Standard 802.15.4 for Low-Rate Wireless Networks. – IEEE, 2020. – 800 с.
8. *Вишневский В., Лаконцев Д., Сафонов А., Шилев С.* Mesh-сети стандарта IEEE 802.11s - технологии и реализация // *Первая миля*. – 2008. – Т. 8, №2-3. – С. 26–31.
9. *Roveri N., Carcaterra A., Molinari L., Pepe G.* Safe and Secure Control of Swarms of Vehicles by Small-World Theory // *Energies*. – 2020. – Vol. 13, N 5. – P. 1043–1071.
10. *Ji X., Zhang W., Chen S., Luo J., Lu L., Yuan W., Hu Z., Chen J.* Speeding Up Velocity Consensus Control with Small World Communication Topology for Unmanned Aerial Vehicle Swarms // *Electronics*. – 2021. – V. 10, № 20. – P. 2547–2563.
11. *Newman, M. E. J.* The Structure and Function of Complex Networks // *SIAM Rev.* – 2003. – V. 45, № 2. – P. 167-256.
12. *Ванник В. Н. и др.* Теория распознавания образов. — М.: Наука, 1974.