

АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ МОДЕЛИРОВАНИЯ ПЕРКОЛЯЦИОННЫХ ПРОЦЕССОВ В СЛУЧАЙНЫХ СЕТЕВЫХ СТРУКТУРАХ

Зальцман А.Д.

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технологический университет», 119454, Москва, Проспект Вернадского, д. 78, Россия. e-mail: ad.zaltsman@gmail.com

Исследования процессов распространения информации и кластеризации узлов в сетях социальных связей, имеющих случайную топологию, является очень актуальной задачей для экономики, рекламы, маркетинга, социологии, политологии и т.д. В работе рассматриваются вопросы разработки алгоритмов для построения сетевых структур, имеющих случайное строение, а так же алгоритмы исследования в них перколяционных процессов, что необходимо для ответа на вопрос, как влияет среднее число связей на величину порога перколяции, а также на ряд других важных процессов.

Ключевые слова: кластеризация узлов, случайная топология, алгоритмы построения случайных сетей, перколяционные процессы, порог перколяции

ALGORITHMIC MEANS OF PERCOLATIVE PROCESSES MODELLING IN RANDOM NETWORK STRUCTURES

¹Zaltsman A.

MIREA (Federal State Budget Educational Institution of Higher Education "Moscow Technological University"), 119454, 78 Prospekt Vernadskovo st., Moscow, Russia, ad.zaltsman@gmail.com

Researching distribution and clustering of information in randomly topologised social network knots is very relevant task for economy, advertizing, marketing, sociology, political science, etc. This work considers either randomly composed social structures development algorithms or its percolative processes development algorithms in order to define the influence of average communications number on percolation threshold size and also on other important processes.

Key words: knots clustering, random topology, random networks organisation algorithms, percolative processes, percolation threshold

Введение

Исследования процессов распространения информации и кластеризации узлов (кластер – группа связанных между собой узлов, выделенных по определенным свойствам или типам) в сетях социальных связей, имеющих случайную топологию, является очень важной и актуальной задачей для экономики, рекламы, маркетинга, социологии, политологии и т.д.

Узлами социальной сети являются отдельные люди, а ребрами - коммуникативные связи, число которых может иметь произвольное значение.

В социальной сети, распространение информации может одновременно происходить множеством путей через различные узлы сети. Отдельный узел сети может получать от другого узла некоторую информацию (рекламные предложения, идеи, политические взгляды, профессиональные сведения, мнений и т.д.) и передавать её другим узлам, если имеет согласную с поступившей информацией позицию (является активным проводником) или блокирует её в противном случае.

Моделирование и анализ информационных процессов, протекающих в социальных сетях со случайной структурой возможен с применением методов теории перколяции, которая может позволить, например, ответить на следующие вопросы:

Как происходит кластеризация общества на группы связанных между собой негативными взглядами индивидуумов, в зависимости от среднего числа связей на узел.

При какой доле негативно настроенных индивидуумов (узлов сети) может создаваться условие для беспрепятственного распространения негативных влияний между двумя любыми, произвольно выбранными узлами (индивидуумами) социальной сети (будет пройден порог перколяции), что можно рассматривать как

готовность общества к революционным изменениям. Доля негативно (оппозиционно, криминально и т.д.) настроенных граждан может быть выявлена путем социологических опросов, что предоставляет возможность оценить насколько социальная сеть близка к порогу перколяции (а состояние государства к социальному взрыву).

Кластеризация социальной сети. Если переход любого узла социальной сети (индивидуума) из одного состояния в другое рассматривать как случайный процесс (с некоторой вероятностью перехода, определяемой множеством случайных факторов, в том числе зависящих от влияния средств массовой информации), то вероятность перехода будет влиять на средний размер кластера пользователей сети (группа напрямую связанных между собой узлов).

Перколяция в социальной сети. В теории перколяции, доля проводящих (не заблокированных) узлов при которой возникает проводимость между двумя различными, произвольно выбранными узлами сети называется порогом перколяции (протекания).

Обзор существующих исследований

Наиболее часто для проведения исследований операций и процессов в сетевых социальных структурах сегодня очень часто используются готовые средства анализа, например инструменты социального сетевого анализа (SNA - social network analysis), позволяющие получить количественные характеристики параметров графа сети, таких как "центральность", "промежуточность", "плотность" (среднее число связей приходящиеся на один узел сети). Центральность характеризует степень влияния данного узла на всю сеть. Промежуточность характеризует степень включенности объекта в маршруты связей между другими участниками сети. Промежуточность показывает, насколько часто данный узел встречается на кратчайших путях между другими узлами.

Использование готовых инструментов имеет как свои преимущества, так и ряд недостатков. К преимуществам можно отнести то, что готовые инструменты социального сетевого анализа позволяют сравнивать между собой по количественным характеристикам однотипные сети. Основные недостатки заключаются в том, что они не позволяют создавать новые, более информативные модели и время анализа больших структур может быть очень большим. Например, для реальных сетей, состоящих из миллионов узлов (пользователей) время поиска промежуточных и центральных узлов может достигать нескольких часов, что является неприемлемым для оперативного анализа. Поэтому в ряде работ уделяется внимание разработке новых алгоритмов ускоряющих поиск центральных узлов. Например, в работе [1] для выявления влияния в очень больших сетях социальных медиа был рассмотрен оптимизированный алгоритм коллективного влияния (CI), использованный для поиска минимального набора узлов влияния с помощью оптимальной перколяции. Было показано, что в отличие от оригинального алгоритма [Morone, Makse, Nature 524, 65 (2015)] сложность может быть уменьшена до $O(N \log N)$, где N общее число узлов в сети) при удалении один за одним узлов из сети. Это становится возможным благодаря использованию соответствующей структуры данных и выбором радиуса сферы коллективного влияния CI (центральный узел влияния является центром окружности, длина которой определяется числом связей). Также возможно расширение данного алгоритма для случаев, когда сфера CI не ограничена. Применение данного алгоритма позволяет существенно сократить затраты времени на проведение расчетов.

В работе [2] для выделения групп сообществ (кластеризации) и выделения коммуникационных связей в больших и малых сетях был предложен эвристический метод, основанный на блочной оптимизации. Использованный авторами [2] метод позволяет быстро находить блоки в больших сетях (кластеры узлов) и определять их полную иерархическую структуру, предоставляя доступ к различным слоям выделения сообществ (кластеризации сети на подгруппы пользователей по интересам или заданным параметрам). Этот метод был использован при отождествлении языкового сообщества в бельгийской сети мобильной телефонной связи (2,6 млн. абонентов) по анализу веб-графа из 118 миллионов узлов и более чем одному миллиарду ссылок. При реализации предлагаемого метода в работе [2] были использованы несколько типов алгоритмов обнаружения сообществ (выделенных кластеров): алгоритм обнаружения межобщинных ссылок и удаление их из сети [3 - 5]; рекурсивный алгоритм слияния аналогичных узлов сообществ (выделенных кластеров) [6].

Приведенный литературный обзор зарубежных работ является далеко не полным. Однако он позволяет представить, насколько тематика представленной работы является актуальной, а вопрос исследования перколяционных свойств социальных сетевых структур является важным.

Поэтому в представленной работе рассматриваются вопросы разработки алгоритмов и программного обеспечения для построения сетевых структур, имеющих случайное строение и исследования в них

перколяционных процессов, что например, необходимо для ответа на вопрос, как влияет среднее число связей (плотность сети) на величину порога перколяции, а также ряд других важных процессов [16-18].

Алгоритм построения случайной сети с множеством путей между узлами

1. Суть предлагаемого алгоритма построения сети с множеством путей между узлами заключается в следующем:

2. Задаем максимальное число M узлов в сети.

3. Берем в качестве текущего узла (для которого генерируются связи) первый узел сети.

4. Задаем случайным образом для текущего узла число связей n_i ($i=1$), n_i – целое число от 1 до какого-либо числа n_{max} (n_{max} – заданное максимальное значение).

5. Проверяем, имеет ли узел некоторые связи, если у узла уже есть связи, то проверяем, не превышает ли их количество то число связей, которое сгенерировано на шаге 3. Если на шаге 3 сгенерировано больше связей, чем в настоящий момент есть у узла, то переходим к шагу 5, иначе считаем, что узел насыщен связями, и мы исключаем его из рассмотрения и переходим к шагу 9.

6. Выбираем из всех оставшихся узлов (исключая рассматриваемый текущий узел) случайным образом количество узлов равное значению свободных связей у узла (число связей сгенерированное на шаге 3, минус число связей уже имеющееся для рассматриваемого узла).

7. Для каждого выбранного узла проверяем количество связей на нем (ограничение по максимуму связей на узле n_{max}). Если на узле достигнут максимум возможных связей n_{max} , считаем его связи сгенерированными и исключаем его из дальнейшего рассмотрения, при этом выбирая произвольным образом другой узел из всех оставшихся к текущему моменту узлов сети (кроме рассматриваемого текущего узла) и повторяя для выбранного узла шаг 6.

8. Если после шага 6 в сети не осталось узлов для образования связей, то считаем все связи образованными и заканчиваем дальнейшее выполнение алгоритма.

9. Соединяем текущий узел (первый) и узлы, отобранные на шагах 5 и 6.

10. Считаем все связи на текущем узле сгенерированными, убираем его из рассмотрения и переходим к следующему узлу сети, повторяя для него шаги 3 – 8, пока не останется узлов в сети или не выполнится условие шага 7.

Алгоритм моделирования перколяции, расчета размеров и числа кластеров

Задача узлов

После того, как вся сеть построена, проводится процедура исследования образования кластеров. Основная идея этого процесса заключается в том, что:

1. Нумеруются все узлы полученной сети (от 1 до M).

2. Исходя из заданной величины Q_i – вероятности того, что узел сети находится в заблокированном состоянии, определяется число заблокированных узлов N ($N=Q_i \cdot M$).

3. Из M чисел случайным образом выбираем N узлов (блокированные узлы).

4. Определяется, какие из этих N – чисел (узлов) имеют связи и связаны между собой в пары, тройки и далее группы узлов (кластеры).

Таким образом, можно проанализировать как вероятность того, что узел находится в заблокированном состоянии, влияет на образование кластеров заблокированных узлов и как, например, среднее число связей приходящиеся на один узел сети (плотность сети) может сказываться на этом процессе, так как при построении сети среднее число связей, приходящееся на один узел, будет зависеть от их допустимого максимального числа n_{max} .

Для расчета числа кластеров и анализа их размеров используются список всех узлов сети, которые находятся в состоянии «исключен» (блокирован). Это позволяет уменьшить число обрабатываемых узлов. Вычисление размера кластера производится по следующему алгоритму:

1. Проверяется, есть ли в массиве заблокированных узлов необработанные узлы. Если нет – то работа алгоритма прекращается.

2. Если массив с исключенными узлами не пуст, то устанавливается текущий размер кластера равным 1.

3. Выбираем первый узел из массива исключенных узлов и добавляем все его связи в список связей для анализа. После этого исключаем выбранный узел из дальнейшего рассмотрения (он уже обработан).

4. Если список связей для анализа не пуст, то для каждой связи проверяем, находится ли на её другом конце исключенный узел. При наличии исключенного узла переходим к шагу 5, иначе считаем связь рассмотренной и переходим к следующей.

5. Увеличиваем размер кластера на 1. Находим обнаруженный на шаге 4 исключенный узел в массиве исключенных узлов, после чего добавляем связи этого узла к текущему списку связей для анализа и исключаем узел из дальнейшего рассмотрения.

6. По окончании рассмотрения списка связей исключенных узлов запоминаем текущие значения размеров кластера (увеличиваем количество кластеров данного размера), после чего переходим к шагу 2.

По окончании работы данного алгоритма будет получен список кластеров в виде: размер кластера и количество таких кластеров.

Задача связей

1. Нумеруются все узлы полученной сети (от 1 до M).
2. Исходя из заданной величины Q_i – вероятности того, что связь сети находится в заблокированном (разорванном) состоянии, определяется число заблокированных связей и узлов.
3. Из M чисел случайным образом выбираем N узлов (блокированные узлы).
4. Определяется, какие из этих N – чисел (узлов) имеют прямые связи и объединены между собой в пары, тройки и далее группы узлов (кластеры).
5. Далее работа алгоритма совпадает с аналогичным алгоритмом для задачи связей.

Алгоритм моделирования перколяции и расчета величины её порога

В общем случае алгоритм моделирования перколяции информации в сети может иметь следующий вид:

1. Задаем число узлов M (не менее 10^6) и строим сеть. Для случайных сетей задаем число n_{min} и n_{max} возможного числа связей на один узел. Индексируем узлы сети.

2. Случайным образом выбираем два узла сети A и B. Можно условно разбить множество узлов пополам, и выбрать узел A случайным образом из первой половины, а узел B случайным образом из второй. Кроме того, наложим на их выбор ещё одно ограничение: между ними должен быть хотя бы один промежуточный узел. Если данное условие не выполняется перебираем один из узлов A, или B до выполнения наложенного ограничения.

3. Первый шаг. Задаем величину вероятности блокирования узлов (исключения из сети в задачи узлов) или разрыва связи (для задачи связей) и случайным образом блокируем число узлов сети (или связей) в общей доли, равной данной вероятности.

4. Проверяем есть ли во всем нашем графе (сети) хотя бы один “открытый” путь (путь из неблокированных узлов или неразорванных связей) от A до B. Если хотя бы один из узлов A или B заблокирован значит открытого пути в любом случае нет (число “открытых” путей равно 0). Если есть хотя бы один “открытый” путь записываем 1 (и запоминаем, что “открытый” путь есть).

5. Второй шаг. Увеличиваем величину вероятности блокирования узлов (исключения из сети в задачи узлов) или разрыва связи (для задачи связей) и случайным образом блокируем число узлов сети (или связей) в общей доли, равной данной вероятности. Определяем заблокированные узлы сети.

6. Проверяем есть ли во всем нашем графе (сети) хотя бы один “открытый” путь (путь из неблокированных узлов) от A до B. Если хотя бы один из узлов A или B заблокирован значит открытого пути в любом случае нет (число “открытых” путей равно 0). Если есть хотя бы один “открытый” путь записываем 1 (и запоминаем, что “открытый” путь есть).

7. Третий и последующие шаги. Возвращаемся к выполнению пунктов 6 и 7 до тех пор, пока не будут заблокированы все узлы сети.

8. Возвращаемся к пункту 3, и заново проводим выполнение пунктов 4 - 7 - Q раз (пусть $Q=100$). Для каждого из шагов от первого и до последнего (когда вся сеть оказывается заблокированной), по всем N экспериментам, находим сумму числа раз, когда был найден хотя бы один “открытый” путь (назовем это число ξ). Сказанное можно пояснить следующим примером. Пусть, например, на $h=5$ шаге в 1, 2, 10, 34, 58, 89 и 96 экспериментах из N хотя бы один “открытый” путь был найден, тогда число $\xi(5)=7$ (7 суммарное число

“открытых” путей). Находим для каждого шага величину $\bar{\rho}(h) = \xi(h)/Q$, где h номер шага. Кроме того, рассчитываем средние значения размеров кластеров заблокированных узлов, количества кластеров, количества узлов в кластерах и размер каждого кластера (параметры кластеризации сети) для каждого из шагов по всем N экспериментам. Сказанное можно пояснить следующим примером. Пусть, например, на $h=5$ шаге процесса блокирования в 1 эксперименте получено 3 кластера размера 100 узлов, во 2 – ом 94 кластера, в 3 – им 102 и т.д. в 100 – 90. Тогда среднее число кластеров размера 100 будет равно: $(100+94+102+ \dots +90)/100$. Средний размер

кластера считается как отношение суммы средних значений, полученных на данном шаге кластеризации по всем Q экспериментам к числу экспериментов Q.

9. Возвращаемся к пункту 2. Заново проводим выбор узлов сети A и B и повторяем выполнение пунктов 3-8 ещё W раз (пусть W=100). Для каждого из W испытаний находим зависимость $\bar{\rho}(h)$, но в данном случае мы будем обозначать их $\bar{\rho}_w(h)$. Наличие индекса w говорит о котором из всех W испытаний идет речь.

10. После окончания всех W испытаний рассчитываем для каждого из h шагов развития эпидемии

величину $\langle \bar{\rho}(h) \rangle = \sum_{w=1}^{W=100} \bar{\rho}_w(h) / W$. Величина $\langle \bar{\rho}(h) \rangle$ – среднее значение вероятности

прохождения данных через сеть по неблокированным узлам для каждого шага, по различным конфигурациям путей в сети. Кроме того, рассчитываем средние значения размеров кластеров блокированных узлов, количества кластеров, количества узлов в кластерах и размеры каждого кластера (параметры кластеризации сети) для каждого из шагов по всем W конфигурациям путей в сети (с учетом усреднения выполненного по всем Q экспериментам).

11. Строим зависимости номера шага, на котором образуется кластер заданного размера от среднего числа связей, входящих на один узел сети.

12. Строим зависимость среднего значения вероятности прохождения данных через сеть $\langle \bar{\rho}(h) \rangle$ от номера шага (доли блокированных узлов сети). Определяем для данного типа сети величину f_c – номер шага потери способности к передачи данных (при какой доле блокированных узлов сеть теряет работоспособность).

13. Возвращаемся к выполнению пункта 1, и повторяем все процедуры алгоритма, описанные в пунктах 1-13 для разных n_{min} и n_{max} .

14. Для каждого n_{min} и n_{max} вычисляем среднее число связей на один узел сети. Назовем это число S. Строим зависимость $f_c(\lambda)$ от S для сети. При необходимости экспортируем полученные данные в другие инструментальные средства для дальнейшей обработки (например, excel) и анализа.

Выводы и рекомендации

Имеющиеся средства сетевого анализа и проводимые на их основе исследования не позволяют целенаправленно изучать такие вопросы, как влияние плотности (среднего числа связей в расчете на один узел) сети на её кластеризацию и величину порога перколяции (как в задаче узлов, так и в задаче связей), как в масштабируемых, так и случайных сетях. И учитывать в целом всю совокупность свойств сети, которые определяют порог перколяции и кластеризацию.

Описанные в представленной работе алгоритмы и программное обеспечение позволяет провести ряд важных исследований [7-9] и сделать, например следующие выводы:

1. В сетях имеющих случайную структуру пороги перколяции как в задаче узлов, так и в задаче связей при большой плотности сети (среднее число связей на один узел) практически достигают величины насыщения (0,24 для задачи связей и 0,10 для задачи узлов) и затем слабо зависят от неё. Величина насыщения порога перколяции в задаче связей почти в 2,5 раза больше чем в задаче узлов.

2. С точки зрения создания проводимости случайной сети в целом, образование проводящих связей в задаче связей менее эффективно, чем образование проводящих узлов в задаче узлов (например, при плотности сети равной 5 для возникновения проводимости необходимо иметь долю проводящих узлов равную 0,22, в то время как доля проводящих связей должна быть равной 0,78).

3. С ростом среднего числа связей (плотности сети), при фиксированной вероятности воздействия, размер кластера увеличивается, а скорость роста кластеризации узлов, находящихся в данном состоянии наиболее сильно увеличивается в области значений вероятности перехода единичных узлов от 0,4 до 0,6; а при малых и высоких значениях возрастает не так сильно.

4. С точки зрения информационного влияния сети, имеющей случайную структуру, увеличение плотности связей оказывает большее влияние, чем наличие отдельных “центральных” узлов имеющих множество связей. Иными словами 100 пользователей социальной сети, имеющие по 20 друзей (связей с другими узлами) будут более эффективно распространять свое мнение, чем один блогер с 1000 подписчиками.

5. Полученные результаты могут быть применены для определения доли числа лиц любой национальности, культурных или религиозных традиций, которые могут жить в иной социальной среде без возникновения социальных конфликтов с представителями других этносов. Необходимо чтобы доля данных

индивидуумов не достигала порога перколяции для случайной сети с типичным для представителей данной национальности средним числом социальных связей.

Работа выполнена при финансовой поддержке РФФИ, грант № 16-29-09458 офи_м «Разработка перколяционных топологических моделей описания виртуальных социальных систем, процессов кластеризации их участников по группам настроений, стохастической динамики распространения влияния и управления переходами».

Список литературы

1. Flaviano Morone, Byungjoon Min, Lin Bo, Romain Mari, and Hern´an A. Collective Influence Algorithm to find influencers via optimal percolation in massively large social media. 2016 // arXiv: 1603.08273v1 [physics.soc-ph] 28 Mar 2016
2. Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre. Fast unfolding of communities in large networks. 2008 // arXiv: 0803.0476v2 [physics.soc-ph] 25 Jul 2008.
3. Girvan M and Newman M E J, 2002 Proc. Natl. Acad. Sci. USA 99 7821.
4. Newman M E J and Girvan M, 2004 Phys. Rev. E 69 026113.
5. Radicchi F, Castellano C, Cecconi F, Loreto V and Parisi D, 2004 Proc. Natl. Acad. Sci. USA 101 2658.
6. Pons P and Latapy M, 2006 Journal of Graph Algorithms and Applications 10191.
7. Lesko, S.A., Zhukov, D.O. Percolation models of information dissemination in social networks. Proceedings - 2015 IEEE International Conference on on Social Computing and Networking, SocialCom 2015, **DOI:** 10.1109/SmartCity.2015.73.
8. T. Khvatova, M. Block, D. Zhukov, S. Lesko. Studying the Structural Topology of the Knowledge Sharing Network. / 11th European Conference on Management Leadership and Governance, Military Academy, Lisbon, Portugal, 12-13 November 2015., ECMLG 2015., E Book ISBN: 978-1-91081-077-4, Book version ISBN: 978-1-91081-076-7, p. 20-27.
9. T. Khvatova, M. Block, D. Zhukov, S. Lesko. How to measure trust: the percolation model applied to intra-organizational knowledge sharing network. Journal of Knowledge Management. 2016, Vol. 20, Issue 5, pp.918 – 935.

References

1. Flaviano Morone, Byungjoon Min, Lin Bo, Romain Mari, and Hern´an A. Collective Influence Algorithm to find influencers via optimal percolation in massively large social media. 2016 // arXiv: 1603.08273v1 [physics.soc-ph] 28 Mar 2016
2. Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre. Fast unfolding of communities in large networks. 2008 // arXiv: 0803.0476v2 [physics.soc-ph] 25 Jul 2008.
3. Girvan M and Newman M E J, 2002 Proc. Natl. Acad. Sci. USA 99 7821.
4. Newman M E J and Girvan M, 2004 Phys. Rev. E 69 026113.
5. Radicchi F, Castellano C, Cecconi F, Loreto V and Parisi D, 2004 Proc. Natl. Acad. Sci. USA 101 2658.
6. Pons P and Latapy M, 2006 Journal of Graph Algorithms and Applications 10191.
7. Lesko, S.A., Zhukov, D.O. Percolation models of information dissemination in social networks. Proceedings - 2015 IEEE International Conference on on Social Computing and Networking, SocialCom 2015, **DOI:** 10.1109/SmartCity.2015.73.
8. T. Khvatova, M. Block, D. Zhukov, S. Lesko. Studying the Structural Topology of the Knowledge Sharing Network. / 11th European Conference on Management Leadership and Governance, Military Academy, Lisbon, Portugal, 12-13 November 2015., ECMLG 2015., E Book ISBN: 978-1-91081-077-4, Book version ISBN: 978-1-91081-076-7, p. 20-27.
9. T. Khvatova, M. Block, D. Zhukov, S. Lesko. How to measure trust: the percolation model applied to intra-organizational knowledge sharing network. Journal of Knowledge Management. 2016, Vol. 20, Issue 5, pp.918 – 935.