



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 27 за 2020 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Кислицын А.А., Орлов Ю.Н.

Структура сильно связной
компоненты сетевого графа

Рекомендуемая форма библиографической ссылки: Кислицын А.А., Орлов Ю.Н. Структура сильно связной компоненты сетевого графа // Препринты ИПМ им. М.В.Келдыша. 2020. № 27. 16 с. <http://doi.org/10.20948/prepr-2020-27>
URL: <http://library.keldysh.ru/preprint.asp?id=2020-27>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

А.А. Кислицын, Ю.Н. Орлов

**Структура сильно связной
компоненты сетевого графа**

Москва — 2020

Кислицын А.А., Орлов Ю.Н.

Структура сильно связной компоненты сетевого графа

В работе представлено исследование структуры сетевого графа, образованного дружескими связями социальной сети «ВКонтакте» между городами России. Оказалось, что распределение вершин графа по степеням близко к равномерному. Следствием этого является наличие высокоразмерной полносвязной области и тонкой периферии. Для плотной и разреженной областей графа оценена вероятность полносвязных сообществ.

Ключевые слова: сетевой граф, сильно связная компонента, распределение рангов, структурная устойчивость

Kislitsyn A.A., Orlov Yu.N.

Structure of the network graph connections

In this paper the investigation of the structure of network graph is presented. The social network between the Russian towns is considered. It is shown, that the distribution of vertex powers is uniform. As a consequence there is a high dimension region with whole connection. The probability of special sub-graphs is estimated.

Key words: explanatory dictionary, strongly connected component, clusterization, semantic analysis

Работа выполнена при поддержке гранта РФФИ № 19-01-00602.

Содержание

1. Введение и постановка задачи	3
2. Алгоритм выделения связных компонент	4
3. Статистика степеней вершин	6
4. Центры силы, ядро и периферия.....	10
5. Связные подграфы в модели равномерного распределения степеней вершин.....	12
6. Заключение.....	15
Литература	16

1. Введение и постановка задачи

Анализ графов социальных сетей представляет собой в настоящее время весьма популярную прикладную математическую задачу, имеющую также и важный теоретический аспект, связанный с разработкой алгоритмов быстрого поиска определенных конфигураций. Теоретическая сторона задачи интересна в первую очередь тем, что в настоящее время отсутствуют доказательные утверждения относительно статистических свойств больших связных графов, позволяющие строить эффективные алгоритмы поиска. Большая часть таких алгоритмов связана с модельными графами (типа «колесо», «звезда», «пропеллер», «дерево» и т.п.) либо основана на эвристиках, проверяемых лишь постфактум. В качестве примеров теоретических работ, в которых обсуждаются основные характеристики графов социальных сетей и алгоритмы, их вычисляющие, следует указать работу [1], в которой обсуждаются некоторые практические методы анализа, работу [2] об алгоритмах визуализации и анализа графов больших размеров, а также работу [3] с подробным анализом методов исследования сетевых структур.

В настоящей работе исследуется структура связей неориентированного графа, представляющего социальную сеть «дружбы» между жителями различных территориальных единиц (условно «городов») России. Данные для анализа взяты с ресурса «ВКонтакте» за 2015 г. в виде, обработанном в [4] и описанном в [5, 6]. Прежде чем выявлять и анализировать особенности такого общения в зависимости от индивидуальных параметров отдельных «подписчиков» (пол, возраст, образование и др.), необходимо определить, имеются ли чисто территориальные особенности графа, позволяющие выделить в нем некоторые априорные структурные блоки.

Вершину графа, представляющего отдельный «город», будем обозначать Γ_i , где $i = 1, 2, \dots, N$, N есть полное число вершин в графе. Сам граф обозначаем символом G . «Населением» города, т.е. числом подписчиков, проживающих в нем (связанных с ним адресом компьютера), равно как и фактическим населением, зарегистрированным на данной территории, а также и его параметрами (пол, возраст, социальный статус, образование и т.п.), на данном этапе анализа не интересуемся.

Структура графа характеризуется структурной матрицей с элементами A_{ij} , которые равны единице, если между вершинами Γ_i и Γ_j есть связь, и нулю, если связи нет. Так как граф неориентированный, а внутригородская сеть не рассматривается, то матрица A_{ij} симметрична, а на ее главной диагонали стоят нули. Количество ребер, выходящих из данной вершины Γ_i , называется степенью n_i этой вершины. Степень вершины вычисляется по формуле:

$$n_i = \sum_{j=1}^N A_{ij}. \quad (1)$$

Граф G называется полным, если каждая его вершина связана с любой другой вершиной. В этом случае степени всех вершин одинаковы и равны $n_{tot} = N - 1$, чем исчерпывается информация о структуре связей графа.

Чаще, однако, наблюдается другая ситуация: матрица A_{ij} разреженная, так что число ребер хотя и больше, чем число вершин, но существенно меньше, чем половина квадрата числа вершин. В таком случае возникает нетривиальный вопрос о связности данной структуры. В первую очередь следует выяснить, содержит ли G так называемую сильно связную компоненту, т.е. подграф, в котором для любых двух его вершин существует путь, соединяющий их через некоторые другие вершины. Поскольку мы рассматриваем неориентированный граф, то для него сильно связная компонента является просто связной компонентой, а сам он представляется в виде некоторой совокупности таких подграфов. Следовательно, дальнейший анализ относится к связным графам.

Для связного графа нас будет интересовать вопрос о том, насколько прочно он связан. С этой целью рассматриваются различные его модификации, порожденные удалением либо ребра, либо вершины. Мы будем рассматривать «вершинную» устойчивость связности графа. Индексом устойчивости будем называть минимальное число вершин, которые достаточно удалить из G , чтобы он перестал быть связным. Это – очевидная задача управления информационными потоками применительно к социальным сетям. Практический интерес представляет анализ реальной структуры [4].

2. Алгоритм выделения связных компонент

В соответствии с данными [4] матрица A_{ij} известна, поэтому естественным алгоритмом поиска связных подграфов является так называемый «поиск в глубину». Он состоит в следующем.

На шаге 1 выбирается произвольная (допустим, первая) строка матрицы A_{ij} и запоминаются ненулевые номера столбцов $\{j_1^1, \dots, j_{p_1}^1\}$ в этой строке.

На шаге 2 для каждого номера j_p^1 , где $1 \leq p \leq p_1$, берется строка матрицы, имеющая этот номер, после чего для нее определяются номера столбцов $\{j_{1,p}^{1,p}, \dots, j_{s_{2,p}}^{1,p}\}$ с ненулевыми (т.е. единичными) значениями. Если при этом какие-либо из этих номеров совпали с номерами, отмеченными на шаге 1 или на предыдущих этапах шага 2, они исключаются из дальнейшего рассмотрения. Каждый номер строки учитывается только один раз. Пусть таких новых номеров на шаге 2 всего p_2 .

На шаге 3 повторяется процедура с новыми номерами, отмеченными на шаге 2. Поиск останавливается на таком шаге h_1 , для которого $p_{h_1} = 0$, т.е. новых номеров нет.

Минимальная по всем вершинам величина

$$H = \min_{\{\Gamma_i\}} (h_i - 1) \quad (2)$$

называется глубиной графа G .

Сложность алгоритма – линейная по числу вершин и ребер графа, т.е. она не выше, чем N^2 .

Применительно к данному конкретному графу выяснилось следующее.

1. Граф связный.

2. Число вершин равно $N = 2441$.

3. Число ребер равно $K = 1\,876\,564$, средняя плотность $\rho = \frac{2K}{N(N-1)} \approx 0,63$.

Эта плотность равна вероятности того, что произвольные две вершины графа являются связными.

4. Глубина графа равна $H = 1$: существует несколько вершин (а именно, 7), которые напрямую – в одну связь – связаны со всеми остальными вершинами.

На рис. 1 представлена схема связей анализируемой сети.

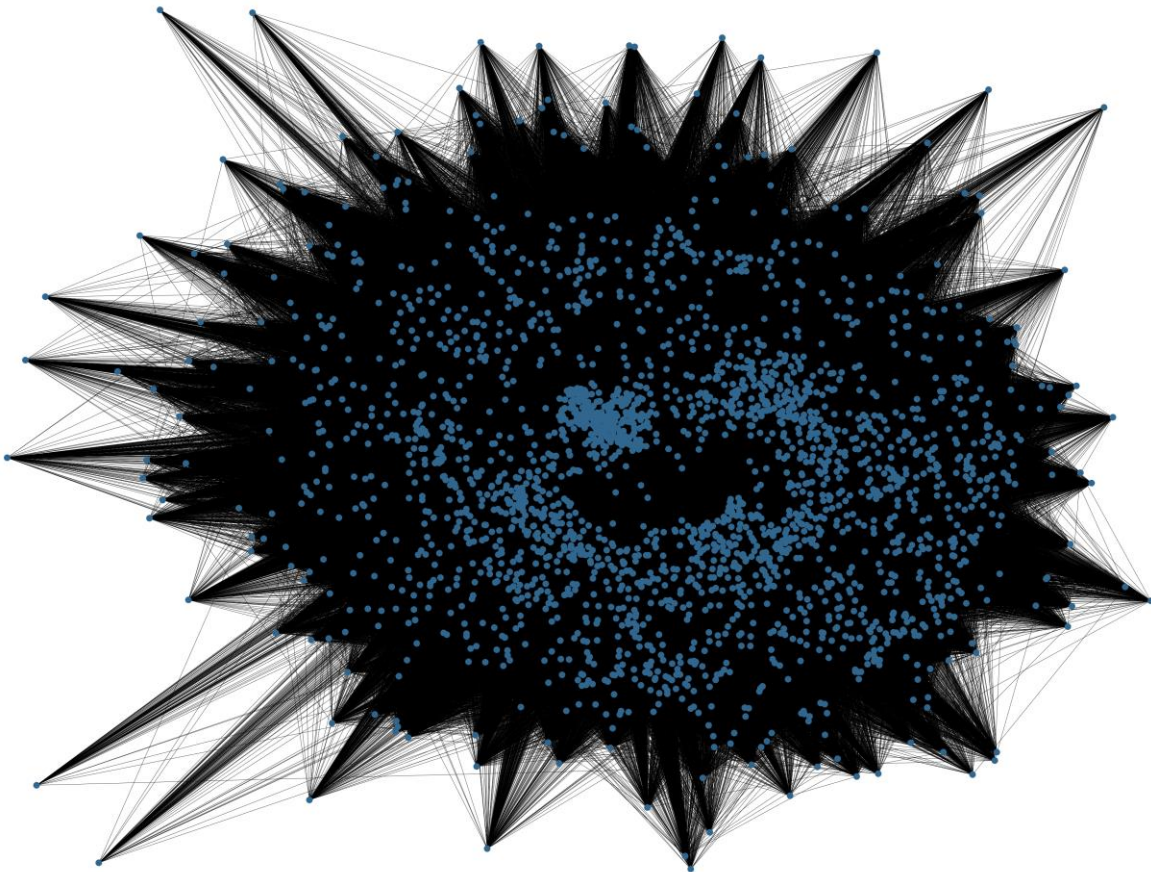


Рис. 1 – Схематичное представление связей графа G

На рис. 1 видны скопления точек в центральной области, имеющие большое количество связей, а также наличие разреженной периферии с относительно небольшим числом связей. Периферийные точки имеют, тем не менее, довольно большое число связей с центральными вершинами, хотя между

собой они почти не связаны, а в центре наблюдается картина, характерная для полной связности.

Поскольку $\sqrt{2K}/N \approx 0,8$, то граф хотя и не полный, но весьма плотный, что затрудняет поиск в нем сообществ, т.е. таких подграфов, плотность которых была бы выше, чем средняя плотность графа, а плотность оставшейся части была бы ниже средней плотности. Если в графе есть полный подграф, то любое подмножество вершин этого последнего образует сообщество. Интерес в таком случае представляет выделение в исходном графе максимально полных независимых подграфов.

3. Статистика степеней вершин

Исследуем статистику степеней вершин графа. Представляют интерес два графика: зависимость степени вершины от ее ранга при упорядочении вершин по убыванию степеней (рис. 2) и собственно распределение вершин по рангам (рис. 3).

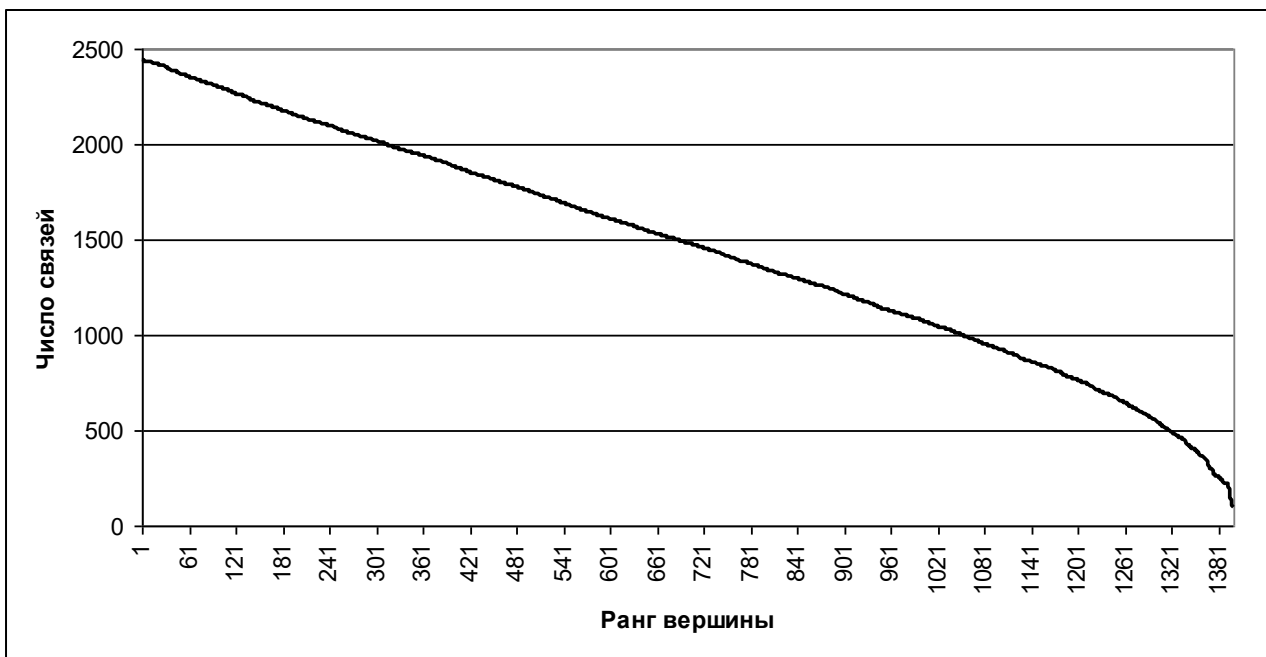


Рис. 2 – Зависимость степени вершины графа от ранга

График на рис. 2 слабо нелинейный, отличие от прямой линии заметно лишь в семи первых и примерно в полутора сотнях последних рангов. Суммарное количество городов в этих фрагментах равно соответственно 55 и 180, количество связей (включая возможные повторы из-за двойного учета вершин) составляет 130 тыс. в городах первых рангов и 65 тыс. в городах последних рангов.

Зависимость степени n вершины от ее ранга r приближенно выражается формулой

$$n = (N - 1) \left(1 - \frac{r - 1}{\alpha^2 (N - 1)} \right)^\alpha, \quad (3)$$

где N есть число вершин графа, а численный параметр α близок к $3/4$ (его оценка в рассматриваемом случае равна $\alpha \approx 0,7628$). Среднеквадратичное отклонение аппроксимации (3) от квазилинейной части кривой на рис. 2 составляет 0,08.

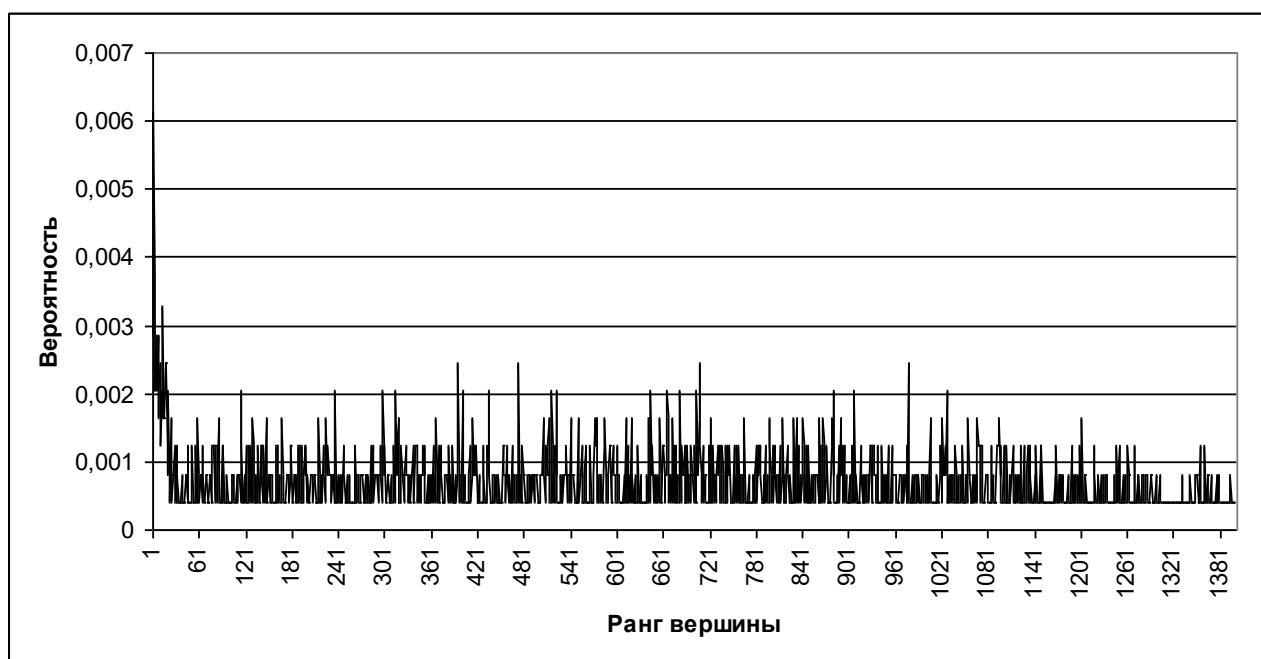


Рис. 3 – Распределение вершин графа по рангу

Статистические характеристики этого распределения следующие.

1. Наиболее вероятное значение числа связей у города 2439, таких городов в системе 15, соответствующий ранг этих городов равен 2.
2. Медиана распределения составляет 1550 связей, ей отвечает ранг 645.
3. Средняя степень вершины равна 1537, чему отвечает ранг 655.
4. Стандартное отклонение степени вершины равно 732.
5. Порядковый номер города по возрастанию ранга, совпадающий с количеством связей этого города, равен 1423, соответствующий ранг равен 745. Этот номер города можно назвать аналогом «индекса Хирша» системы городов.
6. Стационарная точка распределения степени вершины равна 1518, что близко к медиане и характерно для равномерного распределения. Этой точке отвечает ранг города, равный 670. Отметим, что функция распределения вершин по рангу действительно близка к линейной (рис. 4).

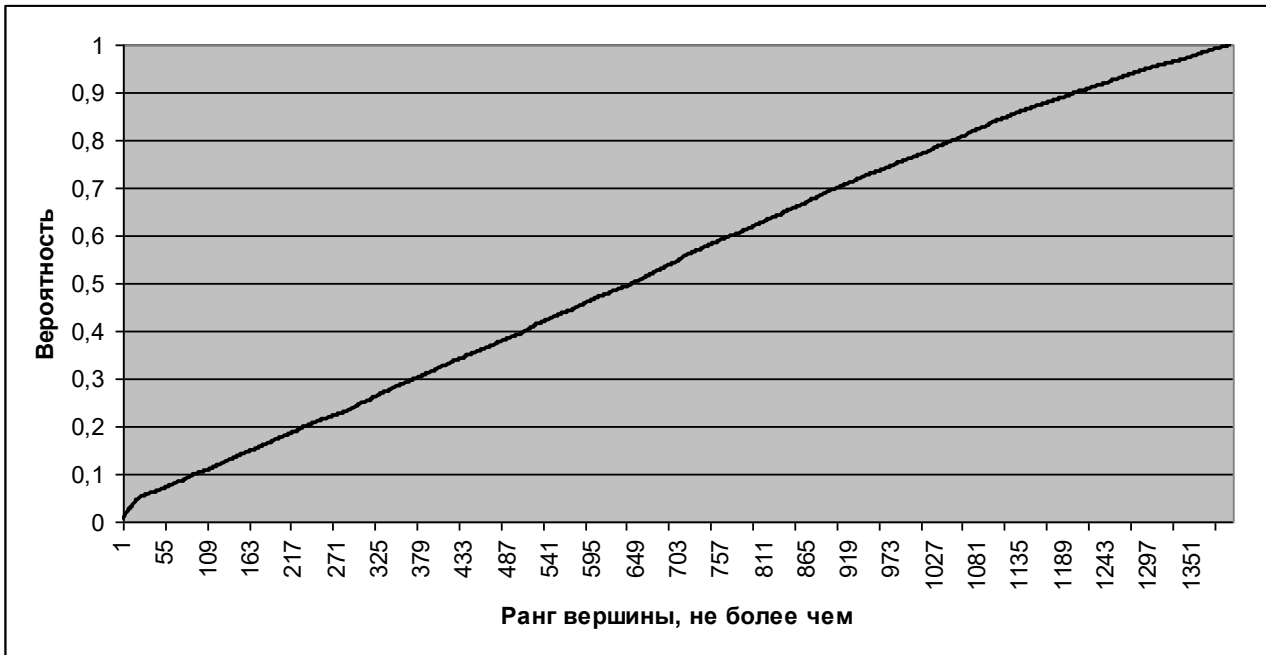


Рис. 4 – Функция распределения вершин графа по рангу

Стационарной точкой распределения мы называем квантиль этого распределения, численно равный доле дополнения до максимума носителя этого распределения, т.е. это такое число ребер k^* , которое определяется из уравнения

$$\sum_{k=k_{\min}}^{k^*} f(k) = 1 - \frac{k^*}{N-1}, \quad (4)$$

где $f(k)$ есть эмпирическая частота вершины со степенью k . Как уже говорилось выше, в данном случае $k^* = 1518$, $\bar{k} = 1537$, $k_{med} = 1550$. Соответствующие значения рангов: $r^* = 670$, $\bar{r} = 655$, $r_{med} = 645$.

7. Минимальное число ребер у вершины равно 97, ранг этого города равен 1400, такой город один, все его связи – с вершинами первых рангов.

8. Максимальное число ребер у вершины равно 2440 (из 2441 городов), т.е. город первого ранга связан со всеми прочими городами в системе. Таких городов первого ранга имеется 7.

9. Городов второго ранга имеется 15, каждый из них не связан с одним каким-то городом в системе. Пять из этих второранговых городов не связаны с одним – одним и тем же – городом последнего ранга 1400, другие пять не связаны с городом ранга 1399, еще два города не связаны с городом ранга 1398 и по одному городу не связаны с городами рангов 1392, 1393 и 1394.

10. Имеется 9 городов третьего ранга, не связанных с некоторыми двумя городами. Ранги этих пар городов, связи с которыми нет, следующие: четыре раза встречается пара (1398, 1400), два раза – пара (1378, 1394) и по одному разу встречаются пары (1394, 1400), (1382, 1400), (1387, 1399).

11. Имеется 5 городов четвертого ранга, не связанных с некоторыми тремя городами в системе. Ранги этих троек городов следующие: (1387, 1396, 1399), (1378, 1397, 1400), (1392, 1397, 1399), (1396, 1397, 1398), (1380, 1387, 1392).

На рис. 5 показано распределение количества городов, имеющих один и тот же ранг. Это распределение нужно для понимания того, как в среднем зависит порядковый номер города от ранга, поскольку распределение степеней вершин в целом линейно зависит от ранга.

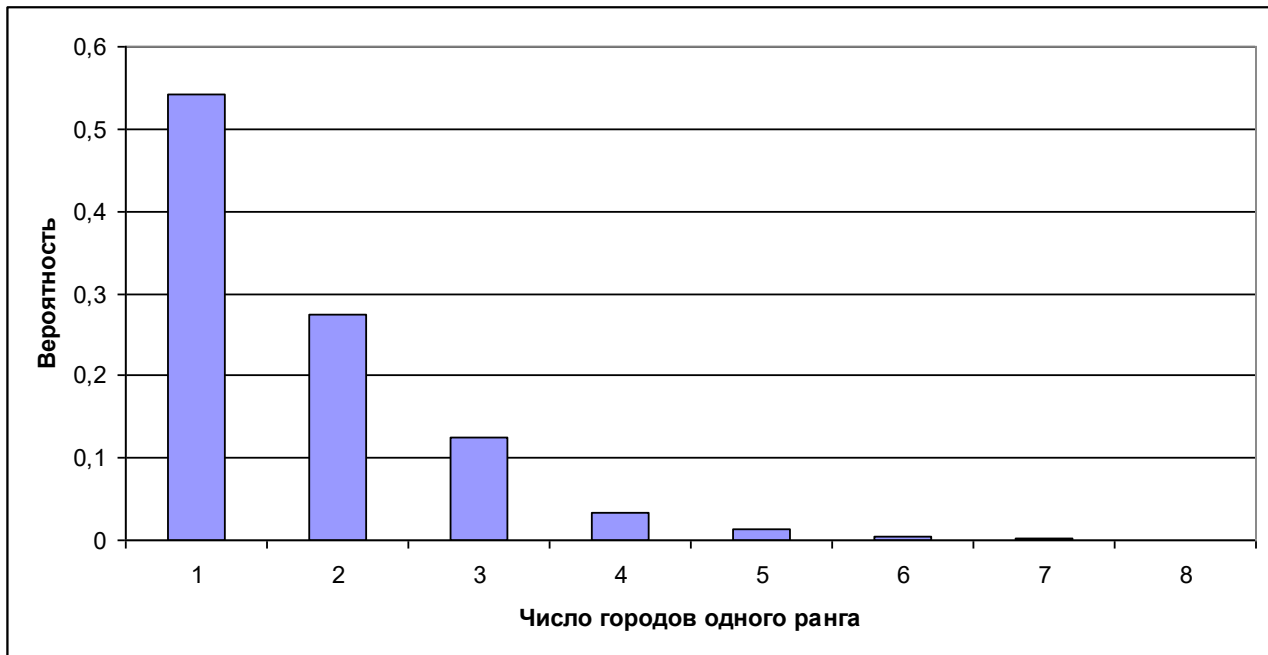


Рис. 5 – Распределение городов по одинаковым рангам

Вероятность на рис. 5 имеет экспоненциальный профиль и с детерминацией 0,99 моделируется зависимостью

$$P(n) = 1,8e^{-n}. \quad (5)$$

При этом количество городов с одинаковым рангом, большее трех, относится к группе рангов менее 1100, большее четырех – к группе с рангами менее 1030, больше пяти – к группе с рангами менее 980. Это означает, что чем выше степень вершины, тем больше вероятность, что вершин с одинаковыми степенями будет больше. В среднем же на 1 ранг приходится 1,6 вершин или на 2 ранга примерно 3 вершины.

Из сравнения рангов городов, с которыми не связаны лидеры, следует, что граф, вообще говоря, не имеет строго вложенной структуры, когда все связи вершин меньшего ранга содержатся в связях вершин большего ранга. Указанным свойством обладают только города первого ранга, связанные со всеми городами системы. Например, среди некоторых городов второго ранга отсутствует связь с городом ранга 1392, тогда как все города третьего ранга с ним связаны. Также и среди некоторых городов третьего ранга отсутствует связь с городом ранга 1394, с которым связаны все города четвертого ранга. Однако указанные исключения немногочисленны, так что приближенно можно

считать, что если город с относительно небольшим рангом r имеет связи с n городами, то с ними имеют связи и все города рангов $r' < r$.

Найденные статистические свойства данного графа существенно отличаются от свойств разреженной сильно связанной компоненты ориентированного графа, отвечающего, например, структуре ссылок внутри информационного иерархического ресурса [7]. Это отличие важно, поскольку позволяет обнаружить, что различные «самоорганизующиеся» структуры в зависимости от плотности могут иметь совершенно несходные свойства, что требует разработки специфических алгоритмов поиска в каждом конкретном случае.

4. Центры силы, ядро и периферия

Центром силы будем называть совокупность вершин, которая содержит связи (необязательно в виде связанного подграфа) со всеми вершинами графа G , причем при удалении из этой совокупности любой вершины оставшиеся вершины не образуют центр силы.

Центр силы связный (полный), если входящие в него вершины образуют связный (полный) подграф.

Центров силы может быть несколько. В частности, в рассматриваемом случае центром силы является любая из вершин первого ранга, как и любая их совокупность. Также центр силы образуют некоторые пары городов второго ранга, у которых разрыв связи происходит по разным городам системы. Аналогичный вывод можно сделать и про вершины третьего ранга.

Центры силы будем упорядочивать по возрастанию количества входящих в них вершин. Первый центр – с минимальным количеством вершин – будем называть ядром графа G . Если есть несколько центров с одинаковым числом вершин, то главным среди них будем считать тот, у которого больше вершин с наибольшей степенью в этой группе. Если таких вершин у рассматриваемых центров одинаковое количество, то центры ранжируются по величине следующих по убыванию степеней вершин. Если степени вершин, входящих в центры, совпадают, ситуацию будем называть вырожденной с соответствующей кратностью вырождения (по числу таких центров силы).

Если, например, максимальное число ребер у некоторой вершины равно $n_{\max} = N - 1$, то соответствующая вершина связана со всеми остальными вершинами графа, глубина графа равна $H = 1$, а сама эта вершина является главным центром силы или ядром графа.

В данном случае ядро графа вырождено, кратность вырождения равна 7, входящие в вырожденное ядро вершины образуют полный подграф. Наибольший ранг вершины, входящей в ядро, будем обозначать R_{ker} . Здесь $R_{\text{ker}} = 1$.

Периферию естественно определить как центр силы с наибольшим количеством вершин. Выяснилось, что ее образуют города с рангами от последнего (1400-го) до 1252-го, куда входят 182 города со связями от 97 до

666. Совокупно эти города составляют 0,075 от общего числа городов системы, так что периферия содержится в последнем дециле распределения городов по степеням вершин, точнее – в третьем квартиле последнего дециля. Обращаясь к Рис. 2, видим, что примерно в этой точке линейная зависимость переходит в корневую.

На рис. 6 показан рост доли достижимых вершин исходного графа из области периферии с увеличением количества включаемых в нее рангов. Выяснилось, что города последних 149 рангов, которым отвечают 182 города, совокупно имеют связи со всеми вершинами графа G .

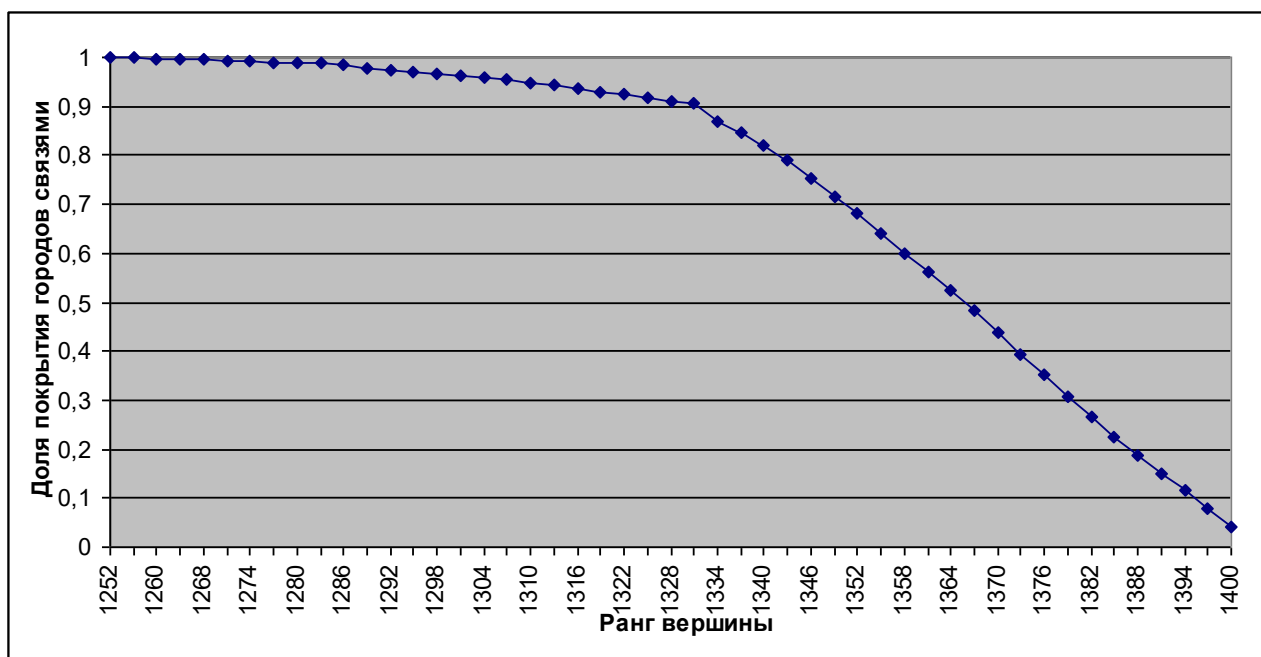


Рис. 6 – Совокупная доля покрытия городов G связями периферии

Итак, любой набор из 182 городов или большего их количества образует граф (необязательно связный), охватывающий своими связями все вершины исходного графа. При этом, как видно из Рис. 6, примерно 90 % вершин исходного графа покрываются 70 последними рангами периферии или 86 городами, а оставшиеся 10 % покрываются 79 рангами и 96 городами. То есть точка излома графика отвечает примерно середине периферии (как по городам, так и по рангам). Количество же связей, которые образуют 86 последних городов, равно 22532 (покрытие 2017 вершин), а следующие 96 городов покрывают своими связями оставшиеся 224 вершины, образуя еще 44973 связи.

Из рис. 6 также видно, что нелинейный фрагмент распределения вершин периферии по числу связей можно аппроксимировать двумя линейными участками с разными углами наклона: правая часть графика имеет наклон $-0,04$ (каждый больший ранг увеличивает покрытие графа на 4 %), а левая – в 10 раз более пологая, наклон $-0,004$.

5. Связные подграфы в модели равномерного распределения степеней вершин

Из проведенного анализа структуры графа G следует, что граф весьма тесно связан. Важно подчеркнуть, что узкой части, по которой его можно было бы разрезать на два несвязных подграфа примерно одинаковой величины, нет. Чтобы отделить периферию (а это всего лишь порядка 10 % вершин графа) от ядра, надо разорвать примерно 75 тыс. ребер.

Кроме того, поскольку имеется полный подграф, состоящий примерно из половины вершин исходного графа, то разорвать эту плотную часть невозможно: удаление любой вершины из этого подграфа оставляет его полным. Более того, если удалить весь этот полный подграф, то оставшаяся половина городов все равно останется связной, хотя и не полным образом.

Заметим также, что города первых рангов связаны между собой, т.е. они образуют полный подграф. Количество городов, с которыми нет связи у городов первых рангов, линейно возрастает с ростом ранга, что позволяет оценить максимальный полный подграф по максимальному количеству связей последнего по рангу входящего в него города. Это количество приближенно оценивается из линейной связи степени вершины с порядковым номером города. При возрастании ранга степень вершин, как было показано выше, уменьшается почти линейно. В этой связи удобно считать идеальную модель рассматриваемой системы дружеских связей в виде линейной зависимости числа связей от ранга (номера) вершины.

Рассмотрим модель равномерно распределенных степеней вершин. Отличаясь на концах от рангового распределения (3), в основной части графа оно примерно адекватно описывает его статистические свойства.

Итак, положим, что вершины занумерованы в порядке убывания числа имеющихся связей, линейно зависящих от номера. Первая вершина имеет максимальное число связей $N - 1$, а последняя (с номером N) – минимальное, равное N_{\min} . Тогда число связей у вершины с номером k равно

$$n(k) = N - 1 - [\beta(k - 1)], \quad \beta = 1 - N_{\min} / (N - 1). \quad (6)$$

Здесь квадратные скобки означают целую часть числа, а коэффициент β в нашем случае приблизительно равен $\beta = 0,96$.

Формула (6) не полностью определяет структуру графа: требуется еще указать типичные правила соединения вершин. Мы выделяем три области графа в соответствии со структурой матрицы инцидентности, схематично представленной на рис. 7. Первая область отвечает вершинам с большим числом связей; убывание количества связей происходит из-за исключения из орбиты дружбы городов третьей области с малым числом связей. Эта область простирается условно до границы максимально полного подграфа. Третья область городов с малым количеством дружеских связей характеризуется нулевой диагональной подматрицей, указывающей на то, что города с малым числом связей дружат не между собой, а с городами первых рангов. Между этими двумя областями расположена средняя – вторая зона, где происходит переход от центра к периферии. В идеале этой прослойки нет.

1	2	3	4	*	*	*	*	*	*	N-1	N
2	0	1	1	*	*	*	1	*	*	*	1
3	1	0	1	*	*	*	1	*	*	*	1
*	1	1	0	*	*	*	1	*	*	*	1
*	*	*	*	0	*	*	1	*	*	1	*
*	*	*	*	*	0	*	1	*	1	*	*
*	*	*	*	*	*	0	1	1	*	*	*
*	1	1	1	1	1	1	0		1		
*						1		0			
*					1	*	1		0	0	0
N-1				1		*			0	0	0
N	1	1	1			*			0	0	0

Рис. 7 – Схематичная структура матрицы инцидентности

Оценим размерность максимального полного подграфа, предполагая вторую зону отсутствующей. Пусть искомая размерность равна k_t . Тогда количество ребер такого упрощенного графа равно, как легко подсчитать, величине

$$K = \frac{k_t(k_t - 1)}{2} + (N - k_t)(2k_t - N) + \frac{(N - k_t)(N - k_t + 1)}{2}.$$

Пренебрегая для простоты единицей по сравнению с много большими N и k_t и вводя величину $x = k_t / N$, получаем отсюда уравнение относительно x :

$$x^2 - 2x + (1 + \rho)/2 = 0,$$

где $\rho = 2K/N^2$ (плотность графа). Следовательно, максимальная размерность полного подграфа равна

$$k_t = N \left(1 - \sqrt{\frac{1 - \rho}{2}} \right). \quad (7)$$

Для рассматриваемой системы городов получаем $x = 0,57$ и $k_t = 1390$. На долю этого подграфа приходится примерно 970 тыс. ребер или 52 % всех связей графа. В идеальной модели других непересекающихся с этим подграфом полных подграфов нет.

В реальной системе существует непустая прослойка в виде второй зоны, где спорадически рассеяны полные подграфы малых размерностей (3, 4, 5, ...), образованные вершинами с промежуточным числом связей. Эти промежуточные вершины связаны не только с центром графа, но еще и между собой. Это уменьшает фактическую размерность нулевой подматрицы в последних рангах. Чтобы оценить ее, подсчитаем непосредственно полные подграфы малых размерностей для вершин последних рангов.

Рассмотрим вероятность связности для трех вершин. Произвольная пара связна между собой с вероятностью $p_2 = \rho$. Что касается конкретной вершины

с числом связей $n(k)$ согласно (6), то если все ее связи находятся в первой зоне, вершины которой заведомо связаны между собой, то вероятность образовать треугольник с вершинами, номера которых больше k_t , равна нулю, а в среднем по графу эта вероятность для данной вершины равна $n(n-1)/(N(N-1))$, т.е. приблизительно равна квадрату вероятности пары. На рис. 8 показана зависимость вероятности образовать треугольник из вершин, номера которых не превосходят указанного по оси абсцисс.

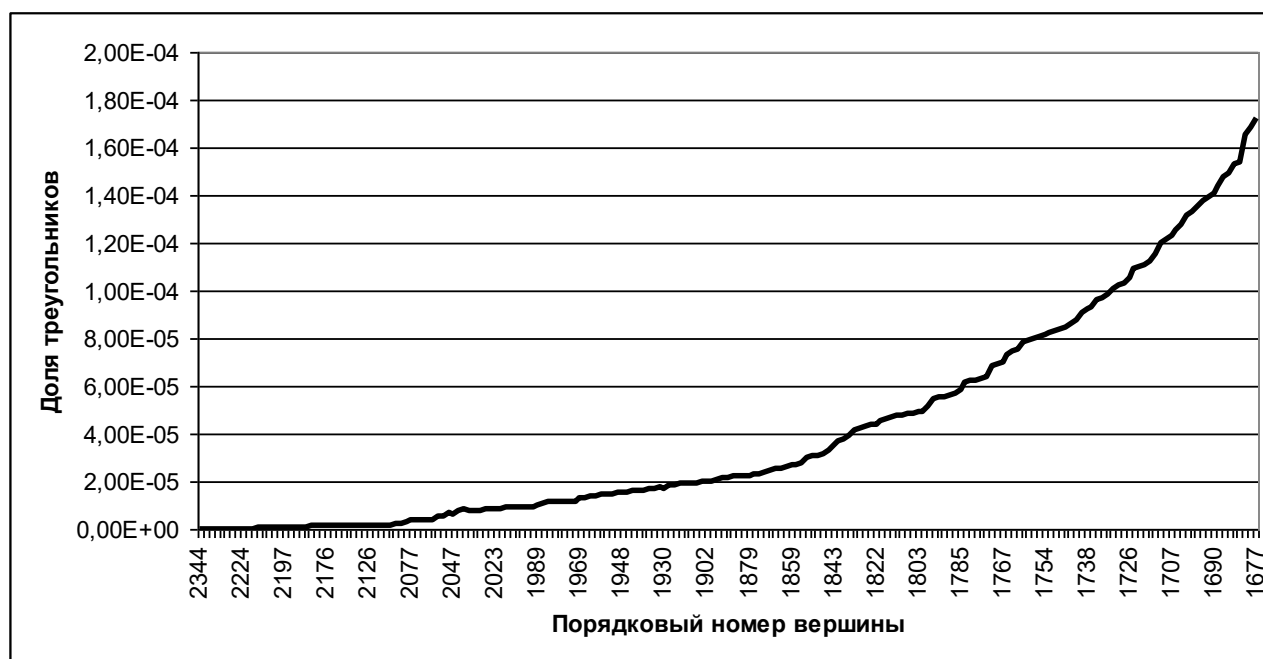


Рис. 8 – Вероятность тройственных дружеских связей в зависимости от номера вершины, открывающей «переходную» вторую зону

Из рис. 8 видно, что последние двести вершин (фактически периферия, выделенная в п. 4) не связаны между собой дружескими связями, заметные же отличия начинаются после четвертой сотни. Вероятность растет экспоненциально, пропорционально $e^{0,01 \cdot (2040-n)}$, достигая при $n = k_t$ величины $2 \cdot 10^{-3}$. В среднем для вершины из периферийной области вероятность связи с какой-нибудь вершиной графа равна 0,45. Если все связи периферии находятся в полном подграфе, то вероятность треугольника, одна из вершин которого лежит в периферии, а две – в плотной области, равна квадрату указанной вероятности, т.е. 0,222. Как показал расчет рис. 8, отличие от этой величины за счет треугольников в области периферии составляет 1 %, т.е. пренебрежимо мало. Тем самым применяемая модель графа G в виде двух зон – полного подграфа и связанной только с ним периферии – достаточно адекватна.

Отметим здесь, что хотя изучаются дружеские связи между городами, но данный граф сильно отличается от так называемого «графа дружеских отношений» [8], в котором любые две вершины имеют в точности одного общего соседа. Это показывает, насколько далеки между собой могут быть теоретический и практический объекты, имеющие одинаковую семантику.

На рис. 9 приведен аналогичный график для вероятности полносвязного четырехугольника.



Рис. 9 – Вероятность четверных дружеских связей в зависимости от номера вершины, открывающей «переходную» вторую зону

Аналогично анализу треугольных связей здесь также можно сделать вывод о том, что связи периферии находятся в основном в области центрального полносвязного подграфа.

6. Заключение

В результате проведенного анализа сети «ВКонтакте» дружеских связей между городами РФ можно сделать следующие выводы.

1. Граф связей плотный, не содержащий узких мест типа «бутылочное горло». Он состоит из полносвязного ядра, состоящего примерно из половины городов, и связанных с ним городов периферии, которые почти не связаны между собой.

2. Распределение вершин по количеству связей почти равномерное, что позволяет аналитически оценить размерность полного подграфа и периферии.

3. Совокупно периферия связана со всеми вершинами графа, так что минимальный по количеству удаленных вершин вариант с разрезанием графа абсолютно неэффективен, поскольку после удаления примерно 200 городов из периферии остается переходная зона и полносвязный подграф.

Литература

1. Батура Т.В. Методы анализа компьютерных социальных сетей // Вестник НГУ, 2012. Т.10, вып. 4. С. 13-28.
2. Коломейченко М.И., Чеповский А.Н. Визуализация и анализ графов больших размеров // Бизнес-информатика, 2014. №4. С. 7-16.
3. Гусарова Н.Ф. Анализ социальных сетей. Основные понятия и метрики. – СПб, ИТМО, 2016. – 67 с.
4. Виртуальное население России // <http://webcensus.ru/>
5. Чекмышев О.А., Яшунский А.Д. Извлечение и использование данных из электронных социальных сетей // Препринты ИПМ им. М.В. Келдыша. 2014. № 62. 16 с. URL: <http://library.keldysh.ru/preprint.asp?id=2014-62>
6. Замятина Н.Ю., Яшунский А.Д. Виртуальная география виртуального населения // Мониторинг общественного мнения: Экономические и социальные перемены, 2018. № 1. С. 117-137
7. Кислицын А.А., Орлов Ю.Н. Структура семантических связей слов русского толкового словаря // Препринты ИПМ им. М.В. Келдыша. 2018. № 252. 28 с. URL: <http://library.keldysh.ru/preprint.asp?id=2018-252>
8. Erdosh P., Renyi A., Sos V.T. On a problem of graph theory // Studia Sci. Mat. Hungar, 1966. V.1. P. 215-235.