

Анализ графа социальных связей в МФТИ

Андрей Сандлер

13 января 2018 г.

Недавно меня посетила одна мысль - у меня почти совсем нет друзей с ФФКЭ и ФПФЭ. Я так понимаю, что я не один такой с ФИВТа :) Развивая эту мысль, я подумал, как можно было бы проверить, сколько друзей с разных факультетов есть у каждого физтеха? Ответ прямо под рукой — ВКонтакте!

Итак, я решил поэзучать граф дружеских связей на Физтехе, используя данные сайта vk.com. Тут уместно сделать некоторые пояснения.

Во-первых, подсчет друзей каждого человека даст много информации, а в итоге из-за её огромного количества не будет видно общей картины. Поэтому лучше будет собрать некоторую статистику с ребер "графа дружбы". Например, почему бы не подсчитать, сколько ребер проведено между факультетами? Это куда как глобальнее, и, в принципе, отражает среднее количество друзей с разных факультетов.

Но количества ребер мне мало, поскольку факультеты по численности бывают разные, и число ребер без учета количества человек ничего не скажет. Как учесть размер факультета? Подумав, я решил считать плотность графа, то есть на ребре между ФИВТ и ФФКЭ писать плотность двудольного графа, где вершины одной доли — это фивты, а вершины другой доли — кванты. Это число получалось маленьким, поэтому я его умножал на 10000.

Возникает один технический вопрос: как по данным со страницы понять, является ли человек студентом Физтеха? Многие (но не все) указывают МФТИ(ГУ) или МФТИ(ГУ) (см. в Москве) в качестве вуза на своей странице. Факультет чаще всего указывается вместе с вузом. Используя эти данные и год выпуска, можно предположительно понять, учится ли человек в МФТИ.

И, наконец, как эти самые данные получить? ВКонтакте уже давно решили вопрос с парсингом страниц и придумали VK API. Оно умеет отдавать список друзей человека, в котором есть поля «вуз», «факультет», «год выпуска», и вообще всё, что можно указать на странице. Чтобы воспользоваться этим API, нужно завести на своей страничке пустое приложение, а данные этого приложения отдать библиотеке, которая умеет работать с VK API. Я нашел на Github библиотеку, написанную на моем любимом Ruby, и дело оставалось за малым — запустить поиск в ширину и ждать результатов.

Результаты заставили себя ждать 15 минут. Оказывается, мой BFS обошел две тысячи человек, найдя между ними чуть больше 60000 ребер. Неплохо! Но мне показалось, что я получил не полную информацию о графе дружбы, поэтому я запустил поиск с трех других аккаунтов (ФИВТ, ФУПМ, ФАКИ), и нашел ещё 10000 ребер. Как это получилось, я до сих пор не понимаю. Вроде бы все четверо (включая меня) состоят в друзьях друг у друга в контакте, списки друзей принудительно открыты несколько месяцев назад, но все же поиск куда-то не доходит. Очень странное явление (возможно, это следствие того, что часть друзей можно скрыть).

Тем не менее, результаты получены, таблицы заполнены, давайте на них смотреть.

	FRTK	FOPF	FAKI	FMBF	FFKE	FALT	FUPM	FPFE	FIVT	NBIK
FRTK	696	49	59	52	62	10	65	47	47	25
FOPF	49	822	63	74	42	7	69	68	62	45
FAKI	59	63	705	50	67	13	61	52	76	27
FMBF	52	74	50	811	43	13	68	69	58	36
FFKE	62	42	67	43	803	4	44	51	36	25
FALT	10	7	13	13	4	1246	8	15	17	17
FUPM	65	69	61	68	44	8	836	49	71	30
FPFE	47	68	52	69	51	15	49	757	60	25
FIVT	47	62	76	58	36	17	71	60	797	34
NBIK	25	45	27	36	25	17	30	25	34	2047

Рис. 1: Плотность графа дружеских связей между факультетами

Что сразу бросается в глаза? Числа на диагонали значительно больше, чем остальные, что и понятно — друзей со своего факультета обычно больше. Второе, что очень заметно — это ФАЛТ. Внутренняя плотность там (и на НБИКе) огромная, а внешняя — очень маленькая. География сказывается.

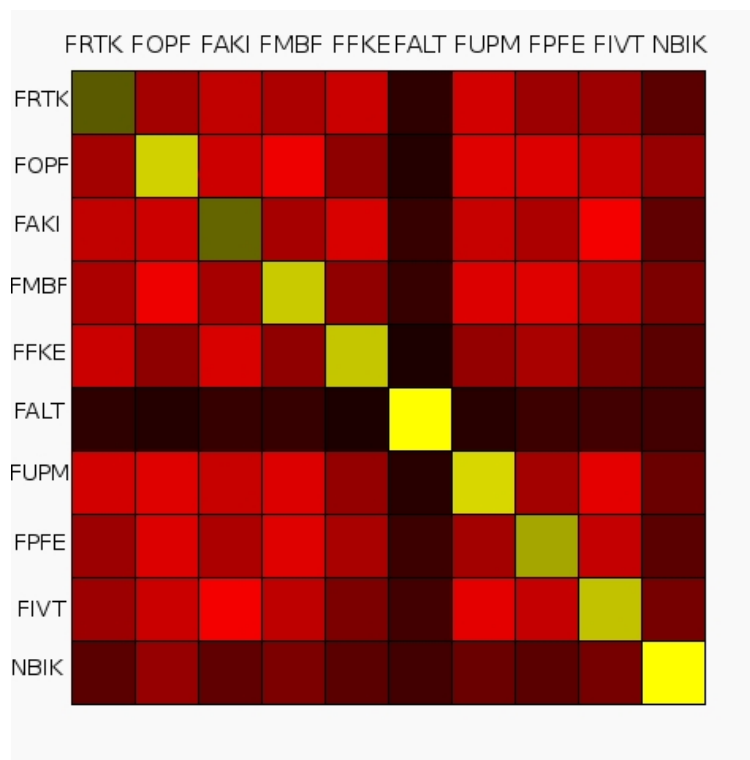


Рис. 2: Плотность графа дружеских связей между факультетами, цветная версия

Глянем на цветную картинку. Желтым здесь обозначена внутренняя плотность, красным — внешняя, градация цвета от темного к светлому означает величину плотности.

Получилось, что ФРТК и ФАКИ не очень дружны внутри. Интересно, как это на самом деле? ФОПФ дружит с ФМБФ, ФИВТ — с ФАКИ (туда на ПМФ переводятся) и с ФУПМом (исторически получилось). НБИК и ФОПФ довольно хорошо общаются.

Так, а сколько же человек с каждого факультета сидит в контакте?

Таблица 1: Количество студентов МФТИ по факультетам

FUPM	362
FOPF	308
FAKI	302
FRTK	296
FIVT	283
FMBF	247
FPFE	194
FFKE	177
FALT	110
NBIK	74
TOTAL	2353

ФУПМ - очевидный лидер!

Что можно сказать по итогам исследования? Результаты показательные, но это отнюдь не значит, что на самом деле картина такая же - это всего лишь статистика по данным ВКонтакте.

Надеюсь, было не очень скучно:)

Бонус для тех, кто дочитал: 10 самых "дружеобильных" людей с Физтеха (по данным vk.com):

- Николай Синюхин, НБИК
- Леночка Обрайен, ФОПФ
- Георгий Речинский, ФУПМ
- Константин Виноградов, ФИВТ
- Егор Ефименко, ФОПФ
- Галка Болдырева, ФРТК
- Даша Милиженко, ФУПМ
- Михаил Ерохин, ФОПФ
- Андрей Криворучко, ФУПМ
- Анна Андреева, ФИВТ