

ПОДХОД К ФОРМИРОВАНИЮ ПРОЕКТНЫХ КОМАНД ИЗ СТУДЕНЧЕСКОЙ ГРУППЫ

Л.Э. Булыгин

Научный руководитель: доцент, к.ф.-м.н. М.Е. Семенов

Национальный исследовательский Томский политехнический университет,

Россия, г.Томск, пр. Ленина, 30, 634050

E-mail: leb1@tpu.ru

APPROACH TO SPLITTING AN ACADEMIC GROUP INTO PROJECT TEAMS

L.E. Bulygin

Scientific Supervisor: Associate Prof., PhD M.E. Semenov

Tomsk Polytechnic University, Russia, Tomsk, Lenin str., 30, 634050

E-mail: leb1@tpu.ru

Abstract. *A detection of communities in a group of people allows researchers to study the modular organization of the network and use this information for various applications. The purpose of the study is to develop an approach of uniform split the academic group on project teams. The project team is a group of people who are able to act in concert and collectively to achieve a common goal. In our experiments, we formed a social network based on reciprocal nomination. Using methods of social network analysis in the designed social network, project teams have been identified. Programming language R and library igraph were used to simulation.*

Обнаружение сообществ внутри коллектива позволяет исследовать модульную организацию сети и использовать полученную информацию для решения различных задач [1, 2]. Цель исследования – разработать социальную технологию равномерного разбиения коллектива на проектные команды. Под проектной командой будем понимать группу людей, которые готовы действовать согласованно и коллективно для достижения общей цели [3]. В наших экспериментах мы сформировали социальную сеть на основе взаимного номинирования [4] коллектива из 20 человек. Используя методы анализа социальных сетей [5] в построенной социальной сети были выявлены проектные команды.

В группе студентов-первокурсников Томского политехнического университета был проведено анкетирование (осенний семестр 2015/2016 учебного года). Анкета включала следующие вопросы:

1. С кем из группы вы общаетесь больше всего?
2. К кому из группы вы обычно обращаетесь за информацией, связанной с учебой?
3. Вы бы учились лучше, если бы смогли больше общаться с
4. С кем из группы Вы бы не хотели сотрудничать в рамках творческого проекта?

Студентам было разрешено номинировать (выбирать) от одного до четырех одноклассников. На основе ответов анкеты были сформированы четыре квадратные матрицы A_1, A_2, A_3 и A_4 размера $n=20$. В каждой матрице смежности A_1, A_2, A_3 элемент (i, j) равен 1, если i -ый студент выбрал j -го студента, в противном случае элемент (i, j) равен 0. В матрице A_4 элемент (i, j) равен -1 , если i -ый студент выбрал j -го студента, в противном случае элемент (i, j) равен 0. Затем матрицы A_1, A_2, A_3 были просуммированы с матрицей A_4 , к полученному результату была применена процедура бинаризации: если элемент (i, j) меньше или равен

0, то элементу (i, j) присваивается значение 0, в противном случае – значение 1. Далее на основании матриц A_1, A_2, A_3 были сформированы три связанных, ориентированных графа G_1, G_2, G_3 . Множество вершин каждого графа соответствует $n=20$ участникам опроса, а множество ребер представляет отношение «взаимная номинация». Для проведения моделирования был использован язык R и библиотека *igraph* [6]. На рис. 1а показан граф G_3 , отражающий структуру взаимных номинаций в группе по третьему вопросу анкеты. Граф G_3 включает $m = 71$ ребро и $n = 20$ вершин (студенты обозначены под номерами $A01, A02, \dots, A20$). Для всех графов вычислены основные характеристики (табл. 1).

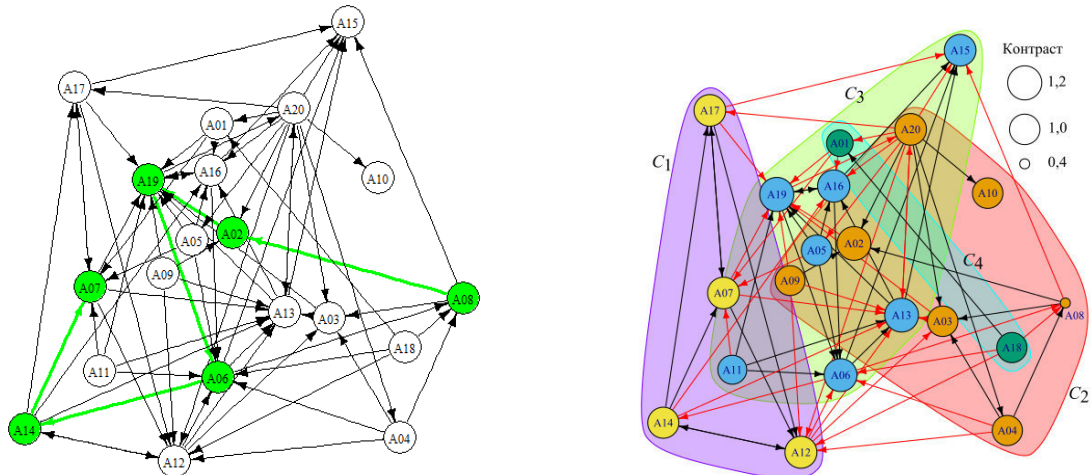


Рис. 1. а) социальная сеть G_3 , диаметр обозначен зеленым цветом,

б) кластеризация сети G_3 , красным обозначены междукластерные ребра, черным – внутрикластерные

Таблица 1

Структурные характеристики

Характеристика	Описание	G_1	G_2	G_3	$\langle G_3 \rangle$
Средний кратчайший путь	среднее минимальное расстояние между парой вершин	2,166	2,236	2,19	2,249
Транзитивность	доля вершин в графе, для которых справедливо: если $i \rightarrow j$ и $j \rightarrow k$, то $j \rightarrow k$	0,399	0,409	0,400	0,3973
Коэффициент кластеризации	показывает тенденцию графа быть разделенным на кластеры	0,3763	0,3767	0,248	0,2488
Взаимность	доля ребер в графе, которые ориентированы в обоих направлениях	0,666	0,472	0,11	0,163
Плотность	отношение числа ребер к числу возможных ребер	0,213	0,189	0,186	0,186
Диаметр	максимальный кратчайший путь между парой вершин	5	6	5	5

В наших экспериментах мы выбрали три алгоритма выявления сообществ: алгоритм промежуточной центральности (edge betweenness) [7], алгоритм случайного блуждания (walktrap) [8] и оптимальный (optimal) алгоритм [9]. Будем называть $C = \{C_1, C_2, \dots, C_p\}$ кластеризацией вершин графа G , где $C_i \neq \emptyset, i=1, 2, \dots, p$. В качестве иллюстрации приведем результат работы оптимального алгоритма кластеризации графа G_3 (рис. 1б). С применением этого алгоритма выделено четыре кластера

$$C_1 = \{A02, A03, A04, A08, A20, A09, A10\}, \quad C_2 = \{A05, A06, A11, A13, A15, A16, A19\},$$

$$C_3 = \{A01, A18\}, \quad C_4 = \{A07, A12, A14, A17\}.$$

Как видно число участников в кластерах разное: $V(C) = \{7, 7, 2, 4\}$. Для оценки качества проведенной кластеризации использован коэффициент модулярности [1], который равен 0,183, также было вычислено

отношение количества внутрикластерных ребер $|m(C)|$ и количества междукластерных ребер $|\bar{m}(C)|$, которое называется *проводимостью* (conductance) графа. В нашем случае это отношение близко к единице: $m(C)/\bar{m}(C)=34/37$.

Для оценки значимости структурных характеристик (табл. 1) проведено моделирование случайных графов. Топологические характеристики графа ($n=20$, $m=71$), а также распределение входящих и исходящих степеней вершин графа G_3 были использованы для генерации 100 случайных графов [6]. Средние значения вычисленных структурных характеристик приведены в столбце $\langle G_3 \rangle$ табл. 1. Для 100 случайных графов среднее значение и среднеквадратическое отклонение модулярности составили соответственно $M=0,2081$, $SD=0,019$. Для контроля успеваемости в проектных группах была использована «Сводная ведомость оценки текущей успеваемости», согласно которой каждый из студентов $A01, A02, \dots, A20$ в осеннем семестре 2015/2016 учебного года мог максимально набрать 480 баллов, средний балл в группе составил $M=323$ балла, $SD=55,9$, размах успеваемости $R=[107; 367]$. На основе текущей успеваемости для каждого студента был вычислен *контраст*, как отношение успеваемости к среднему баллу, размах контраста $R=[0,3; 1,2]$. На рис. 1б вершины графа G_3 отмасштабированы на величину контраста. Средний балл и среднеквадратическое отклонение успеваемости в полученных кластерах составили $M(C)=\{296,33; 343,28; 304,50; 338,50\}$, $SD(C)=\{94,00; 20,54; 26,16; 3,69\}$ соответственно. Для получения равномерного разбиения группы на проектные команды было принято решение объединить кластеры C_3 и C_4 . Как и следовало ожидать, после объединения коэффициент модулярности уменьшился с 0,183 до 0,161, а проводимость $m(C)/\bar{m}(C)=35/36$ возросла, $M(C_3 \cup C_4)=327,16$, $SD(C_3 \cup C_4)=21,29$. В итоге были сформированы три команды $T_1=\{A02, A03, A04, A08, A20, A09, A10\}$, $T_2=\{A05, A06, A11, A13, A15, A16, A19\}$, $T_3=\{A01, A18, A07, A17, A12, A14\}$.

СПИСОК ЛИТЕРАТУРЫ

1. Бузун, Н., Коршунов, А. Выявление пересекающихся сообществ в социальных сетях. Доклады всероссийской научно-практической конференции «Анализ Изображений, Сетей и Текстов» (АИСТ 2012), 16–18 марта, 2012. – М.: Национальный Открытый Университет «ИНТУИТ» 2012. – с. 87-96.
2. Пронин, А.С., Веретенник, Е.В., Семенов, А.В. Формирование учебных групп в университете с помощью анализа социальных сетей // Вопросы образования. – 2014. – № 3. – с. 54-73.
3. Zaccaro, S.J., Rittman, A.L., Marks, M.A. (2001) Team leadership. The Leadership Quarterly (12), 451-483.
4. Pijl, S.J., Koster, M., Hannink, A., Stratingh, A. (2011) Friends in the classroom: a comparison between two methods for the assessment of students' friendship networks. Soc Psychol Educ (14), 475–488.
5. Fortunato, S. (2010) Community detection in graphs. Physics Reports 486 (3-5), 75–174.
6. Csardi, G., Nepusz, T. (2006) The igraph software package for complex network research. Inter Journal Complex Systems, 1695.
7. Girvan, M., Newman, M.E.J. (2002) Community structure in social and biological networks. PNAS 99 (12), 7821–7826.
8. Pons, P., Latapy, M. (2005) Computing communities in large networks using random walks, <http://arxiv.org/abs/physics/0512106>.
9. Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., Wagner, D. (2008) On modularity clustering. IEEE Transactions on Knowledge and Data Engineering 20 (2), 172-187.