

УДК 519.8

ББК 22.18

СВЯЗИ ДРУЖБЫ И КОММЕНТИРОВАНИЯ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНОЙ СЕТИ FACEBOOK¹

Губанов Д. А.², Чхартишвили А. Г.³
(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Статья посвящена анализу связей дружбы пользователей сети Facebook и изучению их взаимосвязи со связями комментирования: определены общие характеристики сети дружбы, дано определение сильным связям дружбы и проанализирована сила связей дружбы пользователей Facebook, исследована связность друзей пользователя, выявлена зависимость между количеством друзей у пользователя и у его друзей, рассмотрена взаимосвязь связей дружбы и комментирования.

Ключевые слова: социальная сеть, связи дружбы и комментирования, сильные и слабые связи.

1. Введение

Термин «социальная сеть» ввел социолог Джеймс Барнс в 1954 году [4], хотя отдельные социально-сетевые исследования проводились еще в 1930-е годы [9]. Сегодня социальная сеть трактуется как, во-первых, социальная структура, состоящая из множества узлов-индивидов и определенного на этом множестве множества связей-отношений (отношений дружбы, обще-

¹ Статья написана при поддержке гранта РФФИ № 13-07-00876 А

² Дмитрий Алексеевич Губанов, кандидат технических наук, старший научный сотрудник (dmitry.a.g@gmail.com).

³ Александр Гедеванович Чхартишвили, доктор физико-математических наук, главный научный сотрудник (sandro_ch@mail.ru).

ния и т.п.), и, во-вторых, как интернет-реализация этой социальной структуры (например, Facebook).

Связи между узлами-пользователями можно определять по-разному, т.е. фактически можно говорить о разных социальных сетях с одним и тем же множеством узлов. В частности, связи бывают сильными (например, регулярная еженедельная переписка) и слабыми (например, обмен сообщениями раз в год) [5, 6]. В данной статье рассматриваются связи дружбы пользователей и связи комментирования – то и другое в слабом и сильном вариантах.

На основе анализа эмпирических данных ищется ответ на следующий вопрос: обусловлены ли связи комментирования связями дружбы либо формальные связи дружбы существуют сами по себе, а общение пользователей происходит по другим каналам? Ответ на этот вопрос представляется важным для решения самых разных задач, например, для прогноза существования связей между пользователями социальных сетей или для изучения распространения информации по связям социальной сети. В частности, положительный ответ означает, что для имитационного моделирования распространения информации (а также для моделирования управляющих воздействий) в социальных сетях использование только данных о связях дружбы является недостаточным.

Структура статьи следующая. В первом разделе исследуется сеть дружбы: рассматриваются ее общие характеристики, дается определение сильным связям дружбы и анализируется сила связей дружбы пользователей Facebook, затем исследуется связность друзей пользователя и дается ответ на вопрос о том, как связаны между собой чем-то похожие пользователи. Во втором разделе рассматриваются вопросы обусловленности связей дружбы связями комментирования и наоборот.

Открытые данные были предоставлены для исследований компанией Лаборатория цифрового общества (www.digsolab.com); исследования проводились для русскоязычного сегмента сети Facebook (связи комментирования рассматривались за период с 1 июня 2012 года по 1 июня 2013 года, связи дружбы рассматривались на сентябрь

2013 года). Расчеты на данных были произведены при помощи пакета GNU Octave (<http://www.gnu.org/software/octave/>).

2. Сеть дружбы

2.1. БАЗОВЫЕ ХАРАКТЕРИСТИКИ СЕТИ ДРУЖБЫ

Приведем базовые характеристики сети дружбы. Количество пользователей русскоязычного сегмента Facebook составляет 3,3 млн. (3 279 156), количество связей дружбы между ними – 77,6 млн. (77 639 757), среднее число друзей у пользователя – 47.

Распределение пользователей по количеству друзей приведено на рис. 1.

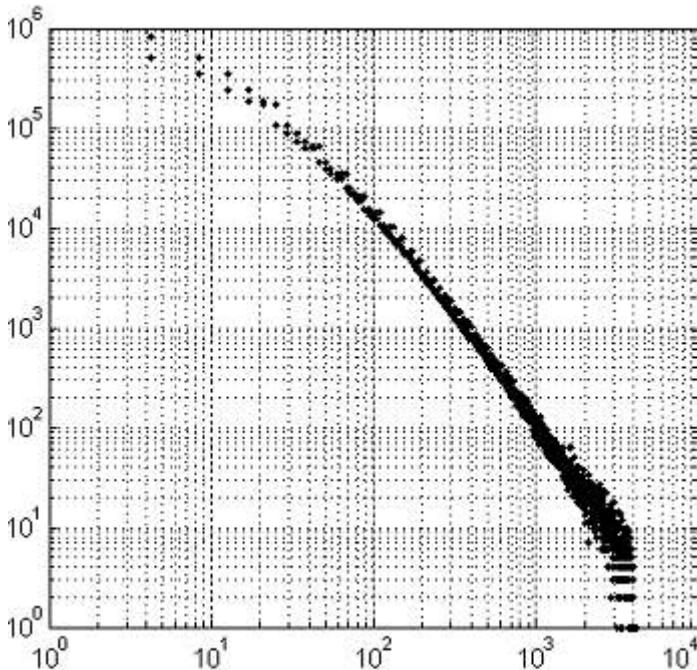


Рис. 1. Распределение пользователей по числу друзей (по оси X указано число друзей, по оси Y – число пользователей)

Распределение количества друзей похоже на степенное ($\alpha = 2,24$): двадцать процентов пользователей имеют не более трех друзей, восемьдесят процентов – не более 45 друзей. Таким образом, рассматриваемый сегмент сети Facebook является безмасштабной сетью¹.

В распределении компонент связности сети дружбы преобладает одна наибольшая компонента (около 3,1 миллионов, или 94% пользователей); одиночные компоненты наиболее распространены (около 197 тыс. пользователей), остальные компоненты по размеру не превосходят двух десятков пользователей и встречаются значительно реже.

2.2. СЛАБЫЕ И СИЛЬНЫЕ СВЯЗИ ДРУЖБЫ

Дадим определение сильной связи дружбы. Связь дружбы между пользователями назовем *сильной* (неслучайной), если для них найдется общий друг. Тогда *силу связей дружбы* пользователя u будем определять как долю друзей, имеющих как минимум одного совместного друга с u ($|\cdot|$ обозначает мощность множества):

$$w^{sf}(u) = \frac{\sum_{v \in \text{Друзья}(u)} I^{sf}(u, v)}{|\text{Друзья}(u)|},$$

где $I^{sf}(u, v) = \begin{cases} 1, & \text{если между } u \text{ и } v \text{ сильная связь дружбы} \\ 0, & \text{иначе.} \end{cases}$

В целом можно отметить (см. рис. 2), что чем больше друзей у пользователя, тем более сильными (менее случайными) являются его связи дружбы (они подтверждаются совместными друзьями).

¹ Сеть, в которой степени вершин распределены по степенному закону [3]. Примером безмасштабной сети является сеть Web, считается также (и подтверждается на практике), что онлайн-социальные сети являются безмасштабными.

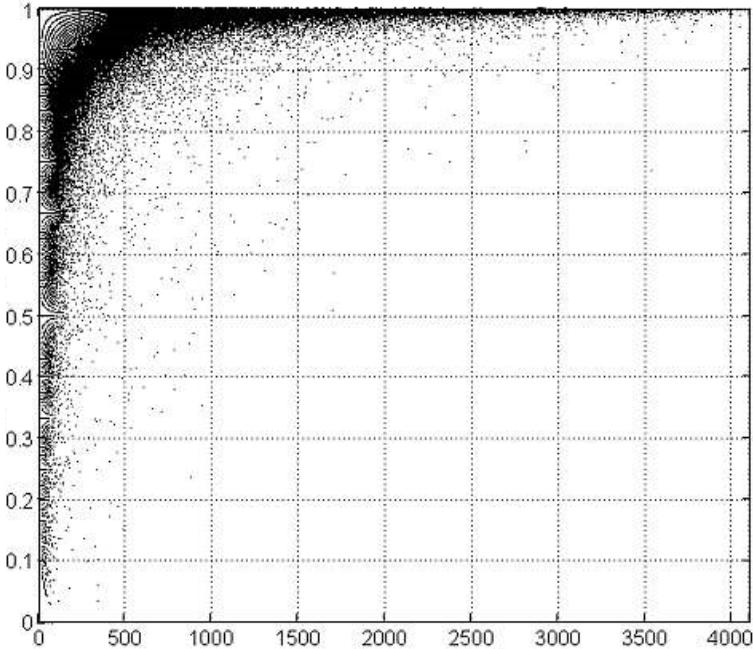


Рис. 2. Диаграмма рассеяния пользователей
(одна точка – один пользователь)

2.3. СВЯЗЬ МЕЖДУ ДРУЗЬЯМИ ПОЛЬЗОВАТЕЛЯ

Дружат ли между собой друзья пользователя? Для ответа на этот вопрос мы используем два показателя:

- связность друзей пользователя;
- количество компонент связности в множестве друзей пользователя.

Связность друзей пользователя. Связность друзей пользователя u (или коэффициент кластеризации в теории анализа сложных сетей [3]) рассчитывается следующим образом:

$$c^d(u) = \frac{2 \cdot ef}{d(d-1)},$$

Где ef – фактическое количество связей между друзьями пользователя u , а d – количество друзей пользователя u . Показатель

$c^d(u)$ принимает значение 0, если друзья никак не связаны между собой, и 1, если каждый друг связан с каждым.

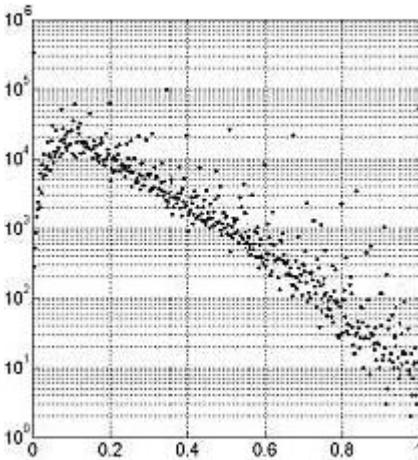


Рис. 3а. Диаграмма рассеяния пользователей (по оси X – связность друзей, по оси Y – число пользователей с данным значением связности)

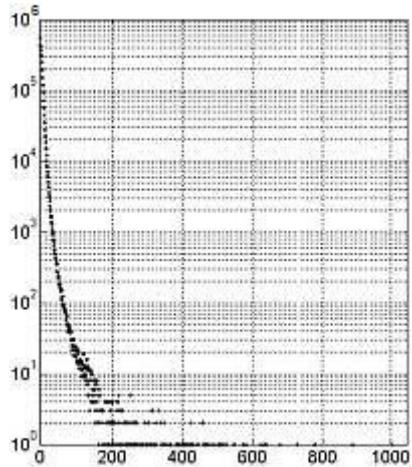


Рис. 3б. Диаграмма рассеяния пользователей (по оси X – число компонент, по оси Y – число пользователей с данным числом компонент)

На рис. 3 видно, что максимум достигается на значении 0,1, т.е. типично, что для пользователя имеется 10% связей дружбы между его друзьями от максимально возможного их числа.

В целом для сети дружбы связность друзей пользователей имеет значение 0,2.

Количество компонент связности в множестве друзей пользователя. С другой стороны, связь друзей пользователя u , $c^{wc}(u)$, можно определить, выделив компоненты связности в сети друзей пользователя (см. рис. 3б).

Из рис. 4 видно, что с ростом числа друзей в окружении пользователя доминируют одиночные компоненты и наибольшие компоненты. Одиночных компонент, впрочем, больше в процентном отношении. Но будут ли такие соотношения сохра-

няться, если мы будем учитывать размер компонент (число пользователей в компоненте)?

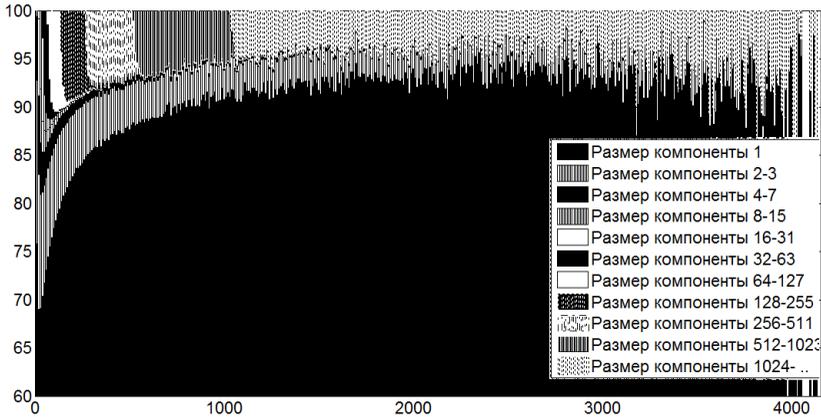


Рис. 4. Процент компонент связности определенного размера пользователей с определенным числом друзей (размер компоненты – количество пользователей в компоненте) (оси X – число друзей, по оси Y – процент компонент)

Рис. 5 показывает, что совокупный объем пользователей в наибольших компонентах значительно превосходит совокупный объем пользователей в одиночных компонентах.

Возникает вопрос, а сколько компонент того или иного типа приходится на пользователя? В сети наблюдается следующая динамика количества наибольших компонент связности: с ростом числа друзей сначала количество наибольших компонент возрастает, достигая пика (1 наибольшая компонента на пользователя в среднем), а затем снижается вплоть до нуля – ее замещает следующая по размеру наибольшая компонента.

Следовательно, мы можем говорить о том, что среди друзей пользователя существует одна наибольшая компонента связности, в которую входит большая часть его друзей (рис. 5). Процентные соотношения между компонентами разного размера можно увидеть на рис. 4.

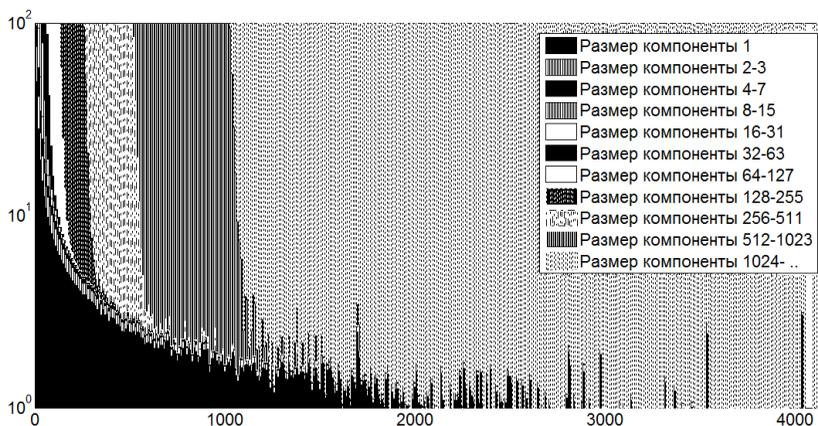


Рис. 5. Процент пользователей в компонентах связности определенной размерности с определенным числом друзей (оси X – число друзей, по оси Y – процент пользователей в компонентах)

2.4. ДРУЖБА ПОХОЖИХ ПОЛЬЗОВАТЕЛЕЙ

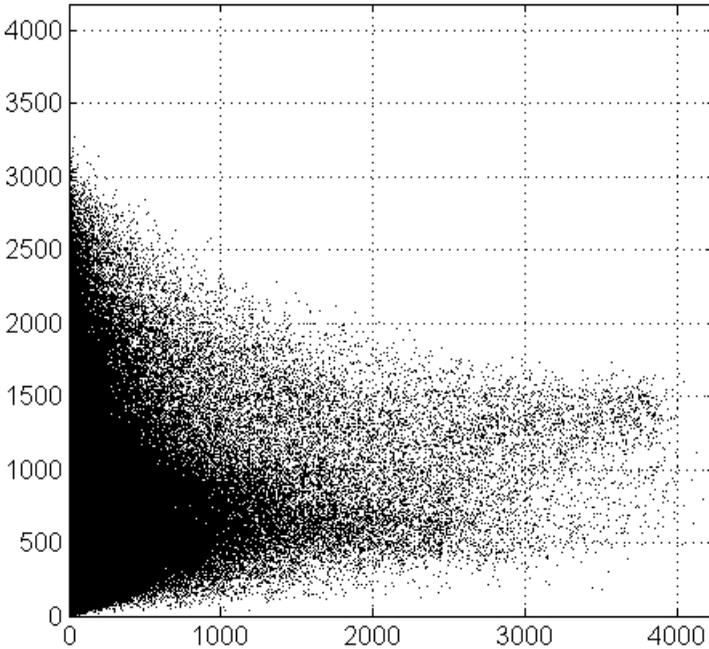
Понятие *гомофилии* (homophily) было введено в социологию Полем Лазарфельдом и Робертом Мертоном [8]. Оно означает, что если два индивида похожи (у них близкие характеристики, например, возраст, пол, профессия или ценности), то они с большей вероятностью устанавливают связь (например, связь дружбы), чем в случае различия характеристик. Верно и обратное: если между двумя индивидами имеется связь, то они скорее всего обладают похожими характеристиками.

Общая постановка вопроса о гомофилии такова: «Какие характеристики определяют гомофилию в той или иной ситуации»? В случае социальных сетей: «Какие характеристики пользователей будут объяснять существование связей между ними (пользователями)»?

Для ответа на вопрос «Дружат ли между собой похожие пользователи?» будем рассматривать в качестве характеристики похожести пользователей количество у них друзей.

Сравним количество друзей заданного пользователя сети со средним количеством друзей у его друзей. Результаты этого сравнения изображены на графике (см. рис. 6), по оси X которо-

го – число друзей, по оси Y – среднее число друзей у друзей. Каждый пользователь обозначается отдельной точкой.



*Рис. 6. По оси X – число друзей,
по оси Y – среднее число друзей у друзей*

Если бы пользователи с одинаковым числом друзей предпочитали дружить между собой, то точки на рис. 6 группировались бы вблизи диагонали. Видно, что это не так, хотя наблюдается зависимость: чем больше друзей у пользователя, тем больше среднее количество друзей у его друзей.

На рис. 6 не видно количество пользователей, приходящихся на одну «точку» графика. Поэтому дополнительно построим гистограмму распределения пользователей по разнице между числом друзей и средним числом друзей у друзей (рис. 7).

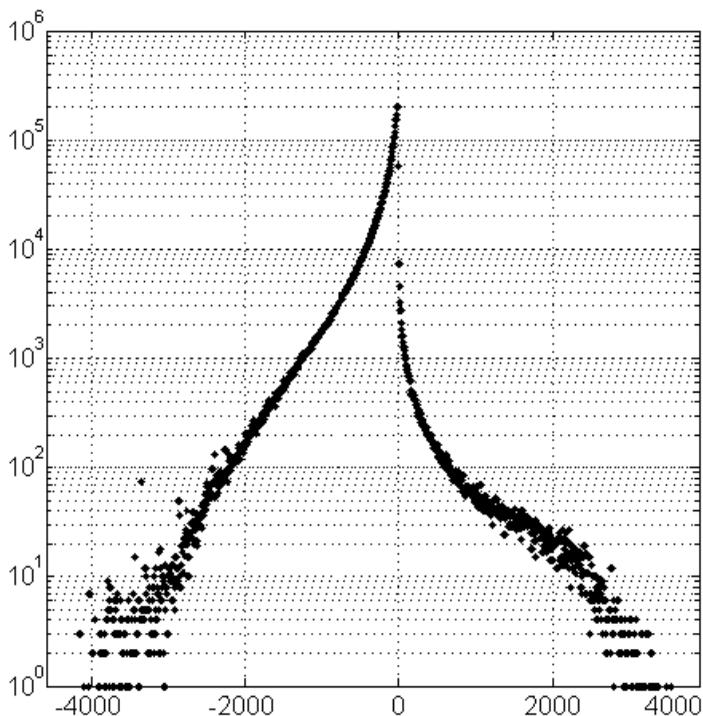


Рис. 7. Гистограмма распределения пользователей по разнице между числом друзей и средним числом друзей у друзей (по оси X – разность числа друзей и среднего числа друзей у друзей, по оси Y – число пользователей)

Видно, что существенна доля пользователей, у которых число друзей одного порядка со средним числом друзей друзей.

Для более строгой оценки того, что пользователи дружат с похожими на них пользователями (в смысле количества друзей), можно использовать показатель ассортативности (assortativity coefficient) [10], который принимает значения из отрезка $[-1, 1]$ (1 – дружат с похожими, 0 – нет никакой связи, -1 – дружат с непохожими):

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk (e_{jk} - q_j q_k),$$

где e_{jk} – вероятность того, что случайно выбранное в графе ребро будет инцидентно вершинам со степенями $j + 1$ и $k + 1$, q_j (q_k) – вероятность того, что для случайно выбранного ребра степень инцидентной ему вершины равна $j + 1$ ($k + 1$), $\sigma_q^2 = \sum_k k^2 q_k - (\sum_k k q_k)^2$.

Для сети дружбы русскоязычного сегмента Facebook показатель ассортативности $r = 0,267$, т.е. гомофилия наблюдается, но в относительно небольшой степени. Для сравнения [10]: коэффициент ассортативности для сети сотрудничества между учеными-физиками составляет 0,363, для сети директоров компаний – 0,276, для интернета составляет - 0,189. Более подробно о значении коэффициента ассортативности в различных сетевых структурах см. [10].

3. Взаимосвязь связей дружбы и связей комментирования

В данном разделе мы рассмотрим взаимосвязь дружбы и комментирования в Facebook. При этом связь комментирования будем считать *сильной*, если она подтверждается 9 месяцев в году (т.е. хотя бы один комментарий в каждом из как минимум девяти месяцев).

Введем следующие обозначения событий: F – два случайно выбранных пользователя являются друзьями; C – между двумя случайно выбранными пользователями существует связь комментирования; SF – два случайно выбранных пользователя являются «сильными» друзьями; SC – между двумя случайно выбранными пользователями существует «сильная» связь комментирования.

Для определения того, в какой мере связи комментирования обусловлены связями дружбы, рассчитаем следующие вероятности:

- $$P(C) = \frac{|E_c|}{|V| * (|V| - 1) / 2} = \frac{7113600}{3279156 * (3279156 - 1) / 2} = 1,3 * 10^{-6}$$
 и

$$P(SC) = \frac{|E_{sc}|}{|V| * (|V| - 1) / 2} = \frac{35713}{3279156 * (3279156 - 1) / 2} = 6,6 * 10^{-9}$$

– вероятности того, что между двумя случайно выбранными пользователями существует связь комментирования;

$$\bullet P(C | F) = \frac{P(C, F)}{P(F)} = \frac{|E_f \cap E_c|}{|E_f|} = \frac{3194403}{77639757} = 0,041 \text{ и}$$

$$P(SC | F) = \frac{P(SC, F)}{P(F)} = \frac{|E_f \cap E_{sc}|}{|E_f|} = \frac{24386}{77639757} = 0,0003$$

– вероятности того, что два пользователя связаны отношением комментирования, если между ними существует связь дружбы;

$$\bullet P(C | SF) = \frac{P(C, SF)}{P(SF)} = \frac{|E_{sf} \cap E_c|}{|E_{sf}|} = \frac{2982964}{18041237} = 0,165 \text{ и}$$

$$P(SC | SF) = \frac{P(SC, SF)}{P(SF)} = \frac{|E_{sf} \cap E_{sc}|}{|E_{sf}|} = \frac{23797}{18041237} = 0,0013$$

– вероятности того, что два пользователя связаны отношением комментирования, если между ними существует неслучайная (сильная) связь дружбы.

Вероятности возрастают, следовательно, связи дружбы значимы для комментирования (в том числе сильного комментирования). Сильная связь дружбы повышает вероятность комментирования (в том числе сильного комментирования) более чем в 4 раза по сравнению с «обычной» связью дружбы (что, в частности, свидетельствует о важности рассмотрения сильных связей дружбы).

Для определения того, в какой мере связи дружбы обусловлены связями комментирования, рассчитаем следующие вероятности:

$$\bullet P(F) = \frac{|E_f|}{|V| * (|V| - 1) / 2} = \frac{77639757}{3279156 * (3279156 - 1) / 2} = 1,4 * 10^{-5} \text{ и}$$

$$P(SF) = \frac{|E_{sf}|}{|V| * (|V| - 1) / 2} = \frac{18041237}{3279156 * (3279156 - 1) / 2} = 3,4 * 10^{-6}$$

– вероятности того, что между двумя случайно выбранными пользователями существует связь дружбы;

$$\bullet P(F | C) = \frac{P(C, F)}{P(C)} = \frac{|E_f \cap E_c|}{|E_c|} = \frac{3194403}{7113600} = 0,45 \text{ и}$$

$$P(SF | C) = \frac{P(SF, C)}{P(C)} = \frac{|E_{sf} \cap E_c|}{|E_c|} = \frac{2982964}{7113600} = 0,42$$

– вероятности того, что два пользователя связаны отношением дружбы, если между ними существует связь комментирования;

$$\bullet P(F | SC) = \frac{P(F, SC)}{P(SC)} = \frac{|E_f \cap E_{sc}|}{|E_{sc}|} = \frac{24386}{35713} = 0,68 \text{ и}$$

$$P(SF | SC) = \frac{P(SC, SF)}{P(SC)} = \frac{|E_{sf} \cap E_{sc}|}{|E_{sc}|} = \frac{23797}{35713} = 0,67$$

– вероятности того, что два пользователя связаны отношением дружбы, если между ними существует сильная связь комментирования.

Вероятности возрастают, следовательно, наличие связей комментирования свидетельствует о дружбе (повышает ее вероятность). Сильная связь комментирования повышает вероятность дружбы (в том числе сильной дружбы) более чем в 1,5 раза по сравнению с «обычной» связью комментирования.

В целом можно сделать следующие выводы:

- связь комментирования встречается гораздо реже, чем связь дружбы;

- если один пользователь комментирует другого, то в почти в половине случаев они являются друзьями;

- если один пользователь комментирует другого, то в двух из трех случаев они являются друзьями;

- если пользователи дружат, то лишь в одном случае из 25 между ними есть связь комментирования;

- если пользователи дружат и есть общий друг, то в одном случае из шести между ними есть связь комментирования.

Таким образом, на поставленный во введении вопрос можно дать следующий ответ: менее чем в половине случаев комментирование означает наличие связи дружбы. Поэтому для изучения распространения информации в сети через комментирование учет связей дружбы является недостаточным.

4. Заключение

В данной статье рассмотрена сеть связей дружбы пользователей социальной сети Facebook. Определены общие характеристики сети, дано определение сильным связям дружбы и проанализирована сила связей дружбы пользователей Facebook, исследована связность друзей пользователя (например, оказывается, что в целом большая часть друзей пользователя находится в одной компоненте, а остальные друзья пользователя изолированы от других друзей), выявлена зависимость между количеством друзей у пользователя и у его друзей, рассмотрена взаимосвязь связей дружбы и комментирования. Планируется в дальнейшем использовать полученные результаты для моделирования поведения пользователей социальной сети.

Литература

1. ЕВИН И.А., ХАБИБУЛЛИН Т.Ф. *Социальные сети* // Компьютерные исследования и моделирование. – 2012. – Т. 4, №2. – С. 423–430.
2. МИТИН Н.А., ПОДЛАЗОВ А.Б., ЩЕТИНИНА Д.П. *Исследование сетевых свойств Живого Журнала* // Препринты ИПМ им. М.В. Келдыша РАН. – 2012. – №76. – 16 с.
3. ALBERT R., BARABÁSI A.-L. *Statistical mechanics of complex networks* // *Reviews of Modern Physics*. – 2002. – №74. – P. 47–97.
4. BARNES J. *Class and Committees in a Norwegian Island Parish* // *Human Relations* – 1954. – Vol. 7. – P. 39–58.
5. GRANOVETTER M.S. *The Strength of Weak Ties* // *American Journal of Sociology*. – 1973. – №78. – P. 1360–1380.

6. GRANOVETTER M.S. *The Strength of Weak Ties: A Network Theory Revisited* // Sociological Theory. – 1983. – Vol.1. – P. 201–233.
7. KADUSHIN C. *Understanding social networks: Theories, concepts, and findings*. – Oxford University Press, 2012. – 252 p.
8. LAZARSFELD P.F., MERTON R.K. *Friendship as a Social Process: A Substantive and Methodological Analysis* // In: Freedom and Control in Modern Society, Morroe Berger, Theodore Abel, and Charles H. Page, eds. – New York: Van Nostrand, 1954. – P. 18–66.
9. MORENO J.L., JENNINGS H.H. *Statistics of Social Configurations* // Sociometry – 1938. – Vol. 1 – P. 342–374.
10. NEWMAN M.E.J. *Mixing patterns in networks* // Physical Review. – 2003. – Vol. 67, Issue 2. – P. 026126.
11. NEWMAN M.E.J. *Networks: An Introduction*. – Oxford University Press, 2010. – 720 p.
12. WASSERMAN S. *Social Network Analysis: Methods and Applications*. – Cambridge University Press, 1994. – 825 p.

FRIENDSHIP AND COMMENTING RELATIONS OF FACEBOOK USERS

Dmitry Gubanov, Institute of Control Sciences of RAS, Moscow, Candidate of Science, senior researcher (dmitry.a.g@gmail.com).

Alexander Chkhartishvili, Institute of Control Sciences of RAS, Moscow, Doctor of Science, leading researcher (sandro_ch@mail.ru).

Abstract: Friendship relations among Facebook users are analyzed and their interrelation with commenting relations is studied. We characterize general features of a friendship network, give a definition of strong friendship relations, analyze strength of friendship relations of Facebook users, investigate connectivity of friends of a user, reveal interrelation between the friends' count of a user, and the friends' count of his/her friends, and consider interrelation of friendship relation with commenting relations.

Keywords: social network, friendship and commenting links, strong and weak ties.

*Статья представлена к публикации
членом редакционной коллегии Д.А. Новиковым*

*Поступила в редакцию 21.07.2014.
Опубликована 30.11.2014.*