

DOI:

## ИССЛЕДОВАНИЕ ВАРИАНТОВ ФОРМАЛИЗАЦИИ ВЛИЯНИЯ ПОЛЬЗОВАТЕЛЕЙ В АКЦИОНАЛЬНОЙ МОДЕЛИ

Губанов Д. А

*Институт проблем управления им. В.А. Трапезникова РАН, Россия, г. Москва*

*ул. Профсоюзная д.65*

*dmitry.a.g@gmail.com*

*Аннотация. Рассматриваются и исследуются различные варианты формализации влияния и влиятельности пользователей онлайн-социальных сетей на основе акциональной модели влияния. Приводятся примеры расчетов влиятельности пользователей онлайн-социальной сети, сравниваются разные варианты формализации влиятельности.*

Ключевые слова: социальная сеть, влияние, акциональная модель, варианты формализации влияния.

### Введение

Влияние в социальных группах и сетях является предметом исследования различных наук и научных направлений (в частности, социальной психологии, теории коммуникаций и теории социально-экономических систем) еще с начала прошлого века. А уже со второй половины XX века предпринимаются попытки формализации влияния и влиятельности индивидов, входящих в состав социальных сетей. Так, например, в теории социально-сетевых анализа влиятельность узла определяется занимаемой им позицией в структуре связей статической сети. На настоящий момент существует сотни мер центральности, которые характеризуют значимость этой позиции (близость узла, степень узла, посредничество узла и др. – см., например, [1, 2]). Однако такая оценка влияния члена сети не позволяет учесть его роль в информационных процессах, протекающих в социальной сети. В теории моделирования информационных процессов в социальных сетях влиятельность узла сети определяется естественным образом: влиятельными являются пользователи, прямое и косвенное влияние которых вызывает наибольшее изменение состояния социальной сети [3, 4, 5]. Сложность с такими моделями микроуровня возникает при идентификации их параметров, особенно для больших онлайн-социальных сетей с сотнями и миллиардами взаимодействующих между собой пользователей. В частности, для моделей динамики мнений сложно количественно оценить как мнения индивидов, так и величину их изменения за короткий промежуток времени. Для больших сетей чаще всего используются нестрогие, зачастую слабо обоснованные, но вычислительно эффективные эвристические методы расчета влияния, в которых учитывается специфика конкретной онлайн-социальной сети. Для расчета влияния применяются как простые показатели (например, число друзей), так и сложные, основанные на машинном обучении (например, решается задача ранжирования пользователей на основе обучающей выборки) [6, 7]. При наличии соответствующих исходных данных эти методы достаточно легко могут быть применены на практике, однако нельзя считать доказанной обоснованность их использования для оценки влиятельности пользователей. В работах [8, 9] предложен акциональный подход к расчету влияния, который основан на эмпирических данных (на действиях пользователей социальной сети) и интересах управляющего субъекта. В данной работе исследуются различные варианты формализации влияния, основанные на акциональной модели. В разделе 1 описана модель распространения действий в социальной сети, в разделе 2 определены понятия влияния и влиятельности, а также разные случаи формализации влияния и влиятельности, в разделе 3 рассмотрены примеры расчета случаев функций влияния/влиятельности пользователей реальной социальной сети Reddit.

### 1 Распространение действий в социальной сети

В формальной модели распространения действий в социальной сети [9] базовым элементом анализа является действие, совершенное агентом (пользователем сети). Кратко опишем эту модель.

Пусть участниками сети являются агенты из множества

$$N = \{1, 2, \dots, n\}$$

которые совершают действия того или иного вида из фиксированного множества

$$K = \{1, 2, \dots, k\}$$

в те или иные моменты времени из интервала  $T$ . Видом действия может быть, например, написание поста, создание учетной записи и создание связи подписки. Обозначим конечное множество действий

(написание конкретного поста, комментария и т.д.) через  $\Delta$ . Каждое действие  $a \in \Delta$  характеризуется тремя параметрами: совершившим его агентом, видом действия и моментом времени совершения действия  $t$ :

$$a(i, j, t), i \in N, j \in K, t \in T.$$

Определим функцию  $\alpha(a)$ , которая каждому действию  $a \in \Delta$  ставит в соответствие совершившего его агента  $\alpha \in N$ .

Далее, пусть на множестве действий задано бинарное отношение частичного порядка « $a$  является причиной  $b$ » (или « $b$  является последствием  $a$ »), обозначаемое следующим образом:

$$a \rightarrow b.$$

Приведем пример такого отношения в реальной онлайн-социальной сети:  $a$  – создание учетной записи,  $b$  – создание связи подписки к этой учетной записи.

Будем считать, что бинарное отношение удовлетворяет следующим свойствам.

1. Рефлексивность: для любого  $a \in \Delta$  справедливо  $a \rightarrow a$ .
2. Транзитивность: если  $a \rightarrow b$  и  $b \rightarrow c$ , то  $a \rightarrow c$ .
3. Антисимметричность: если  $a \rightarrow b$  и  $b \rightarrow a$ , то  $a$  и  $b$  совпадают.

Если  $a \rightarrow b$  и  $a \neq b$ , но при этом не существует такого  $c \in \Delta$ , что  $a \rightarrow c$  и  $c \rightarrow b$ , то будем говорить, что  $a$  является *непосредственной причиной*  $b$  (или  $b$  является *прямым последствием*  $a$ ). Это позволяет выделить класс *однозначных* бинарных отношений, в которых у каждого действия существует не более одной непосредственной причины. Приведем пример неоднозначного бинарного отношения. Пусть  $a$  – пост,  $b$  – комментарий к этому посту,  $c$  – другой пост, при этом комментарий  $b$  содержит ссылку на пост  $c$ . Тогда, если считать справедливым  $a \rightarrow b$  и  $c \rightarrow b$ , бинарное отношение является неоднозначным.

Если задано множество  $A \subseteq \Delta$ , то можно определить множество всех действий, являющихся последствиями действий из  $A$ :

$$\pi(A) = \{b \in \Delta \mid \exists a \in A a \rightarrow b\}.$$

Отметим, что для всех множеств  $A \subseteq \Delta$  выполняется включение  $A \subseteq \pi(A)$  в силу рефлексивности бинарного отношения.

Среди всех действий  $\Delta$  выделим множество  $\Delta^0$  *начальных действий*, которые не являются последствиями какого-либо другого действия:

$$\Delta^0 = \{a \in \Delta \mid \forall b \in \Delta (b \rightarrow a) \Rightarrow (a = b)\}.$$

Для однозначных бинарных отношений у каждого действия существует ровно одно начальное действие, являющееся его причиной. Поэтому множества  $\pi(A)$  и  $\pi(B)$  не пересекаются для любых непересекающихся  $A, B \in \Delta^0$ .

Рассмотрим теперь проблему расчета влиятельности с точки зрения некоего управляющего органа (*центра*). Пусть центр определяет исходя из собственных предпочтений, какие именно действия агентов в социальной сети являются значимыми. Для формализации точки зрения центра введем в рассмотрение *значимость множества действий* – функцию  $\Phi(S)$ , которая каждому множеству действий  $S \subseteq \Delta$  ставит в соответствие неотрицательное вещественное число:

$$\Phi: 2^\Delta \rightarrow [0, +\infty).$$

Естественно предположить, что если к некоторому множеству действий добавить еще действия, то значимость множества увеличится (по крайней мере, не уменьшится). Поэтому будем считать, что значимость множества действий является монотонно возрастающей функцией:

$$(1) \text{ если } A \subseteq B, \text{ то } \Phi(A) \leq \Phi(B).$$

Кроме того, примем естественное предположение о том, что хотя бы какие-то действия обладают положительной значимостью:  $\Phi(\Delta) > 0$ .

Для решения конкретных прикладных задач значимость  $\Phi$  должна быть корректно определена, также должны существовать эффективные алгоритмы расчета ее значения.

## 2 Варианты формализации влияния агентов и мета-агентов

Определим влияние и влиятельность мета-агентов на основе акциональной модели. Мета-агентом будем называть любое непустое подмножество пользователей сети. Мета-агентом является как каждый отдельный агент  $i \in N$  (одноэлементное подмножество  $\{i\}$ ), так и множество всех агентов  $N$ . Для каждого мета-агента  $I \subseteq N$  определим множество  $\delta \subseteq \Delta$  всех совершенных им действий

$$\delta_I = \{a \in \Delta | \alpha(a) \in I\},$$

а также множество совершенных им начальных действий

$$\delta_I^0 = \{a \in \Delta^0 | \alpha(a) \in I\}.$$

Неформальное понимание влияния можно сформулировать следующим образом: влияние мета-агента  $I \subseteq N$  на мета-агента  $J \subseteq N$  велико, если деятельность агентов из множества  $J$  в достаточно большой степени обусловлена деятельностью агентов из множества  $I$ . Формализовать это понимание можно различным образом в зависимости от решаемой практической задачи.

Далее рассмотрим три вопроса, в зависимости от ответа на каждый из которых понятие влияние мета-агента следует формализовать несколько различным образом.

Вопрос А. Оказывают ли влияние все действия (А1), либо только начальные (А2)? Предположим, что некий пользователь реальной онлайн-социальной сети не привносит в сеть какого-либо нового содержания и не пишет оригинальных постов, однако его репосты обладают большой популярностью. Если центр выбирает ответ А1, то такой пользователь является влиятельным, если же ответ А2 – не является.

Вопрос В. Интересует ли центр воздействие на действия мета-агента (В1) или воздействие на последствия действий мета-агента (В2)? Пусть, например, пользователь  $j$  написал несколько постов, и все они получили широкое распространение исключительно благодаря тому, что пользователь  $i$  сделал их репосты. Если центр выбирает ответ В1, то пользователь  $i$  не влияет на пользователя  $j$ , который может даже не догадываться о его существовании. Но если центр выбирает ответ В2, то влияние является большим: если бы не пользователь  $i$ , то о пользователе  $j$ , возможно, никто бы не узнал.

Вопрос С. Считает ли центр, что максимально возможное влияние на любого мета-агента принимает одно и то же значение, то есть влияние является нормированной величиной (С1), либо нет (С2)? Ответ С1 в некотором смысле «уравнивает» пользователей с разным уровнем активности в сети. Это актуально в случае, когда центр интересуется не столько количество действий в сети, сколько количество совершивших их пользователей.

Комбинируя различные ответы на вопросы А–С, можно получить разные случаи и, соответственно, разные варианты формализации понятия влияния. Далее для функции влияния мета-агента  $I$  на мета-агента  $J$  будем использовать одно и то же обозначение  $\chi(I, J)$ , при необходимости каждый раз будем явно указывать на то, какой именно случай рассматривается.

Рассмотрим каждый из случаев.

Случай 1. Сочетанию А1, В1, С1 отвечает следующая функция влияния:

$$\chi(I, J) = \begin{cases} \frac{\Phi(\pi(\delta_I) \cap \delta_J)}{\Phi(\delta_J)}, & \Phi(\delta_J) > 0; \\ 0, & \Phi(\delta_J) = 0. \end{cases}$$

Нетрудно видеть, что в этом случае для любого  $J \subseteq N$  такого, что  $\Phi(\delta_J) > 0$ , справедливо соотношение

$$\chi(I, J) \leq \frac{\Phi(\pi(\delta_N) \cap \delta_J)}{\Phi(\delta_J)} = \chi(N, J) = 1.$$

Аналогичное соотношение  $\chi(I, J) \leq \chi(N, J) = 1$  справедливо для приведенных ниже случаев 3, 5 и 7. Кроме того, для любого  $I \subseteq N$  такого, что  $\Phi(\delta_I) > 0$ , справедливо  $\chi(I, I) = 1$ , т.е. каждый агент влияет на самого себя (это справедливо также для случая 3).

Случай 2. Сочетанию А1, В1, С2 отвечает следующая функция влияния:

$$\chi(I, J) = \Phi(\pi(\delta_I) \cap \delta_J).$$

Случай 3. Сочетанию А1, В2, С1 отвечает следующая функция влияния:

$$\chi(I, J) = \begin{cases} \frac{\Phi(\pi(\delta_I) \cap \pi(\delta_J))}{\Phi(\pi(\delta_J))}, & \Phi(\pi(\delta_J)) > 0; \\ 0, & \Phi(\pi(\delta_J)) = 0. \end{cases}$$

Случай 4. Сочетанию А1, В2, С2 отвечает следующая функция влияния:

$$\chi(I, J) = \Phi(\pi(\delta_I) \cap \pi(\delta_J)).$$

Случай 5. Сочетанию A2, B1, C1 отвечает следующая функция влияния:

$$\chi(I, J) = \begin{cases} \frac{\Phi(\pi(\delta_I^0) \cap \delta_J)}{\Phi(\delta_J)}, & \Phi(\delta_J) > 0; \\ 0, & \Phi(\delta_J) = 0. \end{cases}$$

Случай 6. Сочетанию A2, B1, C2 отвечает следующая функция влияния:

$$\chi(I, J) = \Phi(\pi(\delta_I^0) \cap \delta_J).$$

Случай 7. Сочетанию A2, B2, C1 отвечает следующая функция влияния:

$$\chi(I, J) = \begin{cases} \frac{\Phi(\pi(\delta_I^0) \cap \pi(\delta_J))}{\Phi(\pi(\delta_J))}, & \Phi(\pi(\delta_J)) > 0; \\ 0, & \Phi(\pi(\delta_J)) = 0. \end{cases}$$

Случай 8. Сочетанию A2, B2, C2 отвечает следующая функция влияния:

$$\chi(I, J) = \Phi(\pi(\delta_I^0) \cap \pi(\delta_J)).$$

Отметим важный частный случай, когда мета-агент  $J$  совпадает со всем множеством агентов (т.е.  $J = N$ ) и функция влияния характеризует влияние мета-агента  $I$  на всю сеть, которое назовем *влиятельностью* и обозначим  $\varepsilon(I)$ :

$$\varepsilon(I) = \chi(I, N).$$

Тогда  $\delta_J = \pi(\delta_J) = \Delta$  и ответ на вопрос В неважен. Выпишем влиятельность для каждого из восьми перечисленных выше случаев:

Случай 1, случай 3:  $\varepsilon(I) = \frac{\Phi(\pi(\delta_I))}{\Phi(\Delta)}$ .

Случай 2, случай 4:  $\varepsilon(I) = \Phi(\pi(\delta_I))$ .

Случай 5, случай 7:  $\varepsilon(I) = \frac{\Phi(\pi(\delta_I^0))}{\Phi(\Delta)}$ .

Случай 6, случай 8:  $\varepsilon(I) = \Phi(\pi(\delta_I^0))$ .

В зависимости от поставленной практической задачи влияние и влиятельность могут быть вычислены в соответствии с одним из описанных случаев.

### 3 Расчеты влияния и влиятельности

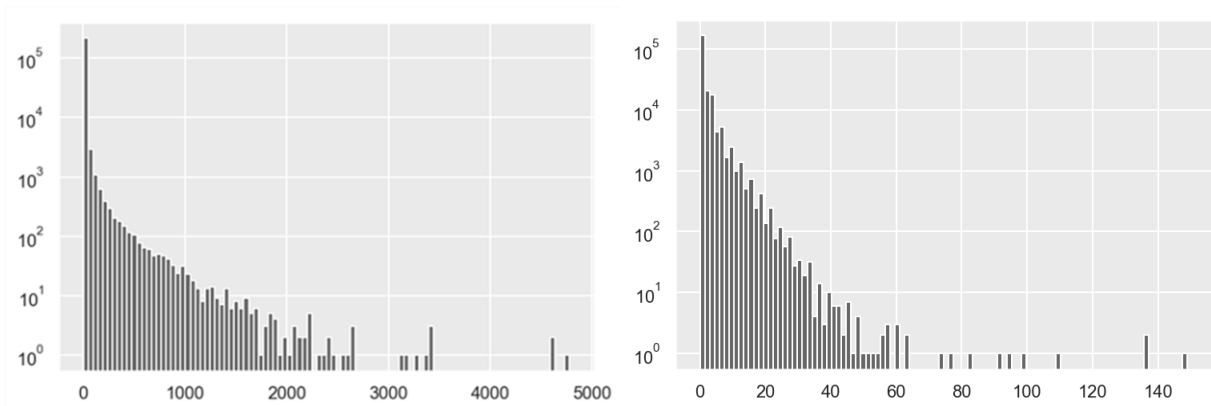
В данном разделе рассматриваются примеры расчета влияния и влиятельности пользователей онлайн-социальной сети Reddit (reddit.com). Для анализа использованы посты и комментарии участников форума politics за 2010 год (224 тысячи постов и 2,2 миллиона комментариев) [10]. На рис. 1а показано распределение деревьев обсуждений пользователей форума Reddit, на рис. 1б показано распределение деревьев по их высоте, на рис. 1в приведен центральный фрагмент одного из деревьев обсуждений.

Из рис. 1 следует, что на форум politics является довольно активным, дискуссии являются объемными, а цепочки обсуждений имеют существенную длину.

Предположим теперь, что значимыми для центра являются посты и комментарии форума politics сети Reddit. При этом центр интересуют только действия, совершенные в течение интервала времени  $\tau$ : с начала по конец 2010 года. Тогда можно ограничиться рассмотрением следующих видов действий:

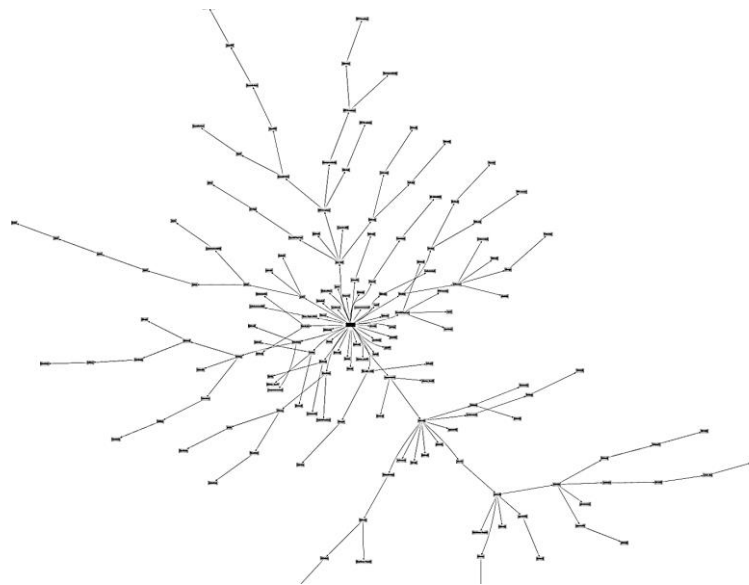
- 1) создание поста,
- 2) создание комментария к посту или комментарий.

Тем самым, множество  $N$  состоит из двух элементов:  $N = \{1, 2\}$ .



(a)

(б)



(в)

Рис. 1. Распределение деревьев обсуждений по числу сообщений (а), по высоте деревьев (б), пример фрагмента дерева обсуждений в сети Reddit (в): центральной вершиной является пост, все остальные вершины являются комментариями к посту или комментариями к комментариям

Будем считать, что однозначное бинарное отношение причинности  $a \rightarrow b$  выполнено в следующих случаях:

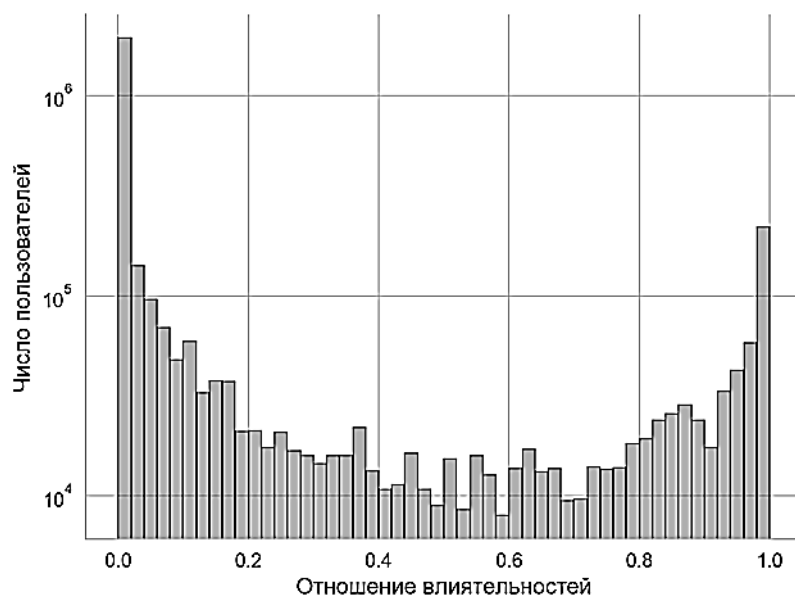
- $a$  – пост,  $b$  – комментарий к нему;
- $a$  – комментарий,  $b$  – комментарий к нему;
- $a$  и  $b$  совпадают.

Будем считать, что значимость совокупности действий зависит аддитивно от каждого из них:

$$\Phi(S) = \sum_{a \in S} \Phi(a),$$

где  $\Phi(a) = 1$ , если  $a$  – пост или комментарий форума politics, созданный в интервале  $\tau$ , иначе  $\Phi(a) = 0$ . Далее расчет влияния можно осуществить по ранее приведенным формулам влияния и значимости.

Различие между вариантами влияния проиллюстрировано на рис. 2 (для влияния существенным является только вопрос А, т. е. варианты А1 и А2).

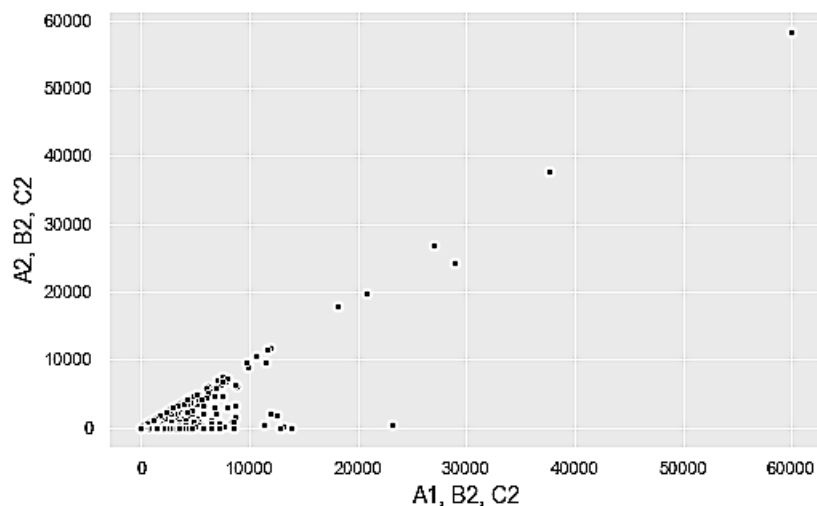


*Рис. 2. Распределение числа пользователей по отношению влиятельности по начальным действиям к влиятельности по всем действиям*

Рис. 2 показывает довольно четкое разделение пользователей на тех, влиятельность которых обусловлена их начальными действиями, («писателей») и тех, влиятельность которых обусловлена их действиями, не являющихся начальными («комментаторов»).

Далее приведем примеры различий между вариантами формализации влияния. На рис. 3–5 представлены диаграммы рассеяния, точками которых являются выявленные связи влияния между агентами (пользователями Reddit). По осям откладываются значения различных вариантов влияния.

На рис. 3 показаны различия между вариантами A1, B2, C2 и A2, B2, C2. В данном случае различие между ними обусловлено тем, что первый вариант расчета влияния учитывает все действия агента, оказывающего влияние, а второй – только его начальные действия. Коэффициент ранговой корреляции Спирмена между рассматриваемыми вариантами равен 0,066.



*Рис. 3. Диаграмма рассеяния для вариантов влияния A1, B2, C2 и A2, B2, C2*

На рис. 4 показаны различия между вариантами A1, B2, C2 и A1, B1, C2. Первый вариант расчета влияния рассматривает воздействие на действия агента, а второй – только воздействие на последствия его действий. Коэффициент ранговой корреляции Спирмена между рассматриваемыми вариантами составляет 0,314.

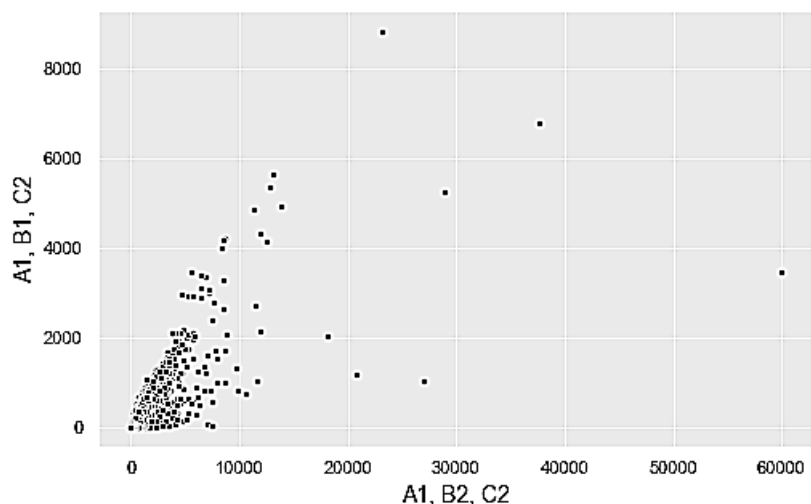


Рис. 4. Диаграмма рассеяния для вариантов влияния  $A1, B2, C2$  и  $A1, B1, C2$

На рис. 5 различия между вариантами  $A1, B2, C2$  и  $A1, B2, C1$  обусловлены тем, что первый вариант рассчитывает ненормированное влияние, а второй – нормированное. Коэффициент ранговой корреляции Спирмена между вариантами влияния равен 0,312. Отметим, что основная масса точек сосредоточена в левой части графика, в правой части графика располагаются точки, соответствующие таким связям влияния, для которых испытывающий влияние агент подвергается воздействию ровно со стороны одного агента и последствия такого воздействия значимы.

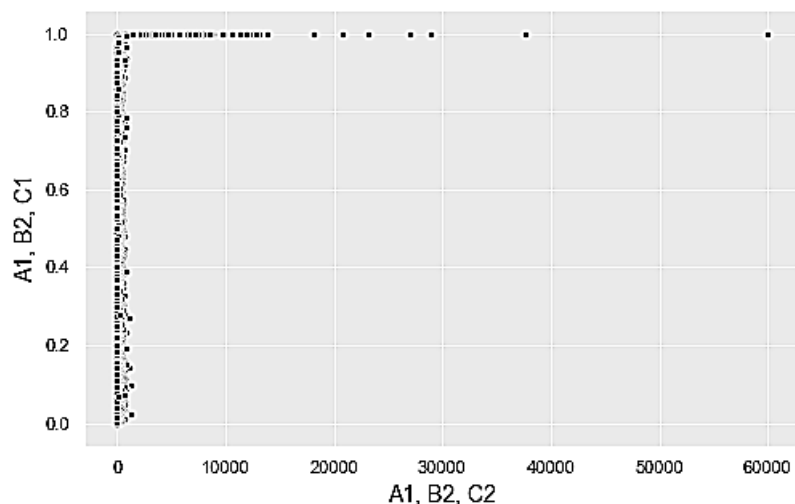


Рис. 5. Диаграмма рассеяния для вариантов влияния  $A1, B2, C2$  и  $A1, B2, C1$

Таким образом на примере сети Reddit можно заключить, что между предложенными вариантами влияния и влиятельностью на практике имеются существенные различия.

Работа выполнена при частичной поддержке РФФИ (проект 18-29-22042, 20-07-00190).

## Литература

1. Csató L. Measuring centrality by a generalization of degree // Central European Journal of Operations Research. – 2017. – Vol. 25, No. 4. – P. 771–790.
2. Skibski O., Rahwan T. Attachment Centrality: An Axiomatic Approach to Connectivity in Networks // Proc. of the 15th Int. Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016). – 2016. – P. 168–176.
3. Губанов Д.А., Новиков Д.А., Чхартушвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства. – М.: МЦНМО, 2018. – 224 с.
4. Новиков Д.А., Бреер В.В., Рогаткин А.Д. Управление толпой: математические модели порогового коллективного поведения. – М.: ЛЕНАНД, 2016. – 168 с.
5. Kempe D., Kleinberg J., Tardos E. Maximizing the Spread of Influence through a Social Network // Theory of Computing. – 2015. – Vol. 11, No. 4. – P. 105–147.

6. Rao A. et al. Klout score: Measuring influence across multiple social networks // IEEE Int. Conference on Big Data (Big Data) – 2015. – P. 2282–2289.
7. Riquelme F., González-Cantergiani P. Measuring user influence on Twitter: A survey // Information Processing & Management. – 2016. – No. 5(52). – P. 949–975.
8. Губанов Д.А., Чхартишвили А.Г. Акциональная модель влияния пользователей социальной сети // Проблемы управления. – 2014. – №4. – С. 20–25.
9. Губанов Д.А., Чхартишвили А.Г. Влиятельность пользователей и метапользователей социальной сети // Проблемы управления. – 2016. – №6. – С. 12–17.
10. *Stuck\_In\_the\_Matrix*. I have every publicly available Reddit comment for research. ~ 1.7 billion comments @ 250 GB compressed. Any interest in this? – URL: [https://www.reddit.com/r/datasets/comments/3bxlg7/i\\_have\\_every\\_publicly\\_available\\_reddit\\_comment/](https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/) (дата обращения: 05.01.2020).