

DOI:

## **ЭВОЛЮЦИЯ СОЦИАЛЬНЫХ СВЯЗЕЙ ПОЛЬЗОВАТЕЛЕЙ ВКОНТАКТЕ И ИНФОРМАЦИОННЫЕ ПУЗЫРИ**

**Козицин И.В.**

*Институт проблем управления им. В.А. Трапезникова РАН, Россия, г. Москва  
ул. Профсоюзная д.65*

*Московский физико-технический институт (национальный исследовательский университет), Россия, г. Долгопрудный*

*Институтский пер., 9*

*kozitsin.ivan@mail.ru*

**Щербинина А.А.**

*Московский государственный университет имени М.В. Ломоносова, Россия, г. Москва*

*Ленинские горы, д. 1*

*shcherbinina.aa.97@gmail.com*

**Бабкина Т.С., Гильмутдинов Б.Н.**

*Московский физико-технический институт (национальный исследовательский университет), Россия, г. Долгопрудный*

*Институтский пер., 9*

*babkinats@yandex.ru, bulat.gilmutdinov@phystech.edu*

**Гойко В.Л., Сайфулин Э.Р.**

*Национальный исследовательский Томский государственный университет, Россия, г. Томск  
пр. Ленина, 36*

*goiko.slava@gmail.com, sayfulin@data.tsu.ru*

*Аннотация: В данной работе рассматривается эволюция графа дружбы пользователей онлайн-социальной сети ВКонтакте. Мы анализируем, как процессы удаления старых связей и появления новых связаны с мнениями пользователей. Полученные результаты свидетельствуют о том, что новые связи являются механизмом, сдерживающим формирование эхо-камер.*

*Ключевые слова: эхо-камеры, гомофилия, динамика социальных сетей.*

### **Введение**

Социальные сети играют огромную роль в жизни общества, выступая проводником благ, информации, идей, мнений, инноваций и многих других важных сущностей [1]. От структуры социальной сети во многом зависит то, как эти сущности будут по ней распространяться [2, 3]. В связи с этим, чрезвычайно важно понимать законы, которым подчиняются социальные сети.

Эмпирические исследования позволили установить ряд свойств, присущих подавляющему большинству социальных сетей: степенное распределение степеней вершин, малый диаметр, высокий коэффициент кластеризации [4]. Кроме того, было обнаружено, что соседние вершины в социальных графах с высокой вероятностью имеют схожие свойства (явление гомофилии) [4 – 6]. Сюда относятся такая сугубо структурная характеристика, как степень вершин (“знаменитости общаются со знаменитостями”), а также другие атрибуты: пол человека, его возраст, цвет кожи, вероисповедание, достаток, а также мнение. Тот факт, что индивиды стремятся общаться с теми, кто разделяет их собственные взгляды, получил огромное внимание в социальной психологии и нашел свое отражение, к примеру, в теории предвзятого подтверждения (confirmation bias) [7] и в теории социального суждения (social judgment theory) [8, 9].

Учитывая, что под действием социального влияния мнения индивидов в результате общения становятся похожими друг на друга [5], коэволюция мнений индивидов и социальных связей между ними может привести к ситуации, когда люди будут общаться только лишь с теми, у кого такие же взгляды [10], то есть окажутся в эхо-камерах (информационных пузырях) [11, 12]. Особую роль здесь могут сыграть онлайн-социальные сети (Facebook, Twitter, ВКонтакте и др.), поскольку алгоритмы ранжирования и рекомендательные системы, которые используют данные медиа-платформы, стремятся окружить пользователя той информацией, которая ему заведомо понравится (для этого они используют историю действий пользователя), что в итоге может увеличить вероятность оказаться в информационном пузыре [13].

Цель данной работы – понять, как эволюция связей “онлайн дружба” между пользователями социальной сети ВКонтакте связана с мнениями пользователей и каким образом первая корректирует идеологическую напряженность (различия во мнениях относительно политической повестки) между пользователями и их онлайн друзьями.

## 1 Методы

### 1.1 Данные

Была использована предсказательная модель (далее – ПМ) для оценивания политических взглядов российских пользователей ВКонтакте [14]. Данному пользователю  $i$ , основываясь на информации с его аккаунта (а именно, его подписках и манифестированными им политических взглядах), выгруженной в момент времени  $t$ , она ставит в соответствие число  $x_i(t) \in [0,1]$ , выражающее степень поддержки им В.В. Путина (которое мы далее будем называть оценкой его политических взглядов).

Рассматривалась выборка пользователей ВКонтакте  $U$ , где  $n = |U| = 6,533,120$ . Пользователи этой выборки удовлетворяют следующим ограничениям:

1. Аккаунт пользователя не удален, не заблокирован и не закрыт полностью настройками приватности;
2. Пользователь посещал ВКонтакте каждую неделю в течение рассматриваемого промежутка времени;
3. Пользователь из РФ;
4. У пользователя указан возраст;
5. Пользователь старше 17-ли лет на момент начала наблюдений;
6. Пользователь имеет более 10 и менее 300 элементов в поле “подписки”;
7. Пользователь не принадлежит множеству аккаунтов, на основании которых ПМ конструирует  $x_i(t)$ .

Основываясь на информации, выгруженной в марте 2019 года ( $t_1$ ), был построен вектор  $\mathbf{x}(t_1) = [x_1(t_1) \dots x_n(t_1)]^T \in \mathbb{R}^n$ , оценивающий политические взгляды пользователей в указанный момент времени.

Кроме того, для каждого пользователя  $i \in U$  в марте 2019 года и в августе 2019 года ( $t_2$ ) были выгружены множества его онлайн друзей ВКонтакте  $V_i(t_1) \subset U$  и  $V_i(t_2) \subset U$  соответственно (далее – просто друзей). На основании этой информации были построены две матрицы смежности:  $\mathbf{A}(t_1) = [a_{ij}(t_1)] \in \{0,1\}^{n \times n}$  и  $\mathbf{A}(t_2) = [a_{ij}(t_2)] \in \{0,1\}^{n \times n}$ . Компоненты этих матриц подчиняются условию  $j \in V_i(t_k) \Leftrightarrow a_{ij}(t_k) = 1$ . При этом  $|V_i(t_k)| = \sum_{j=1}^n a_{ij}(t_k)$ . Отметим, что эти матрицы не являются симметричными (несмотря на то, что связь “дружба” в ВКонтакте является двусторонней), поскольку пользователи могут скрывать некоторых или всех своих друзей отдельными настройками приватности. К примеру, из равенств  $a_{ij}(t_1) = 1$ ,  $a_{ji}(t_1) = 0$  следует, что пользователи  $i$  и  $j$  являются друзьями в момент времени  $t_1$ , но пользователь  $j$  скрыл своими настройками приватности дружбу с пользователем  $i$ . Заметим, что равенства  $a_{ij}(t_1) = 0$ ,  $a_{ji}(t_1) = 0$  не гарантируют отсутствие дружбы между пользователями  $i$  и  $j$ , так как возможна ситуация, при которой они оба скрыли настройками приватности факт своей дружбы. Мы снизили число пропусков в данных, симметризовав матрицы  $\mathbf{A}(t_1)$  и  $\mathbf{A}(t_2)$ . Далее, говоря о множестве друзей  $V_i(t_k)$  пользователя  $i$ , мы будем понимать его с учетом симметризации.

Обозначим  $V_i^- = V_i(t_1)/V_i(t_2)$  и  $V_i^+ = V_i(t_2)/V_i(t_1)$ . Иными словами, множество  $V_i^-$  содержит в себе пользователей, с кем пользователь  $i$  потерял дружбу, а множество  $V_i^+$  тех, с кем он ее приобрел. Пусть  $j \in V_i^-$ . Тогда в силу симметрии матриц  $\mathbf{A}(t_1)$  и  $\mathbf{A}(t_2)$  выполнено  $i \in V_j^-$ . Это означает, что связь  $(i, j)$  исчезла за рассматриваемый промежуток времени, что может быть связано с тем, что один из пользователей отписался от другого, разорвав тем самым связь, или же с введением настроек приватности. В первом случае один из агентов является прямым виновником разрыва связи, однако имеющиеся в нашем распоряжении данные не позволяют идентифицировать таких пользователей. Далее, для простоты, мы будем говорить, что пользователь  $i$  удалил пользователя  $j$  из друзей. Отметим, что словосочетание “удалил из друзей” является общепринятым термином в социальной сети ВКонтакте (однако обычно оно применяется при разговоре о прямом виновнике исчезновения связи). Пусть  $j \in V_i^+$ . Появление связи  $(i, j)$  может быть связано с тем, что один из этих двух пользователей стал инициатором дружбы, отправив соответствующую заявку, а второй принял эту заявку, согласившись стать другом первого или же с тем, что были сняты соответствующие настройки приватности. Имеющиеся в нашем распоряжении данные не позволяют определить, что произошло на

самом деле. Далее для простоты мы будем говорить, что пользователь  $i$  добавил в друзья пользователя  $j$  (словосочетание “добавил в друзья” также является общепринятым термином в социальной сети ВКонтакте).

Введем матрицу  $\Delta \mathbf{A} = \mathbf{A}(t_2) - \mathbf{A}(t_1)$ , которую можно представить в виде

$$\Delta \mathbf{A} = \Delta \mathbf{A}_+ - \Delta \mathbf{A}_-,$$

где  $\Delta \mathbf{A}_+, \Delta \mathbf{A}_- \in \{0,1\}^{n \times n}$  определяются исходя из следующих соотношений:

$$[\Delta \mathbf{A}_-]_{ij} = 1 \Leftrightarrow j \in V_i^-,$$

$$[\Delta \mathbf{A}_+]_{ij} = 1 \Leftrightarrow j \in V_i^+.$$

Иными словами, матрица  $\Delta \mathbf{A}_-$  описывает исчезнувшие связи, а  $\Delta \mathbf{A}_+$  хранит информацию о новых связях.

## 1.2 Различия во мнениях между пользователями и их друзьями

Через  $[k]$ , где  $k \in \mathbb{N}$  будем обозначать множество натуральных чисел от единицы до  $k$ . Введем функцию  $d: \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}$ , где  $k \in \mathbb{N}$  – некоторое натуральное число:

$$d\left(x, \{x_j\}_{j \in [k]}\right) = \frac{\sum_{j=1}^k |x - x_j|}{k}.$$

Различия во мнениях между двумя пользователями будем описывать с помощью следующей конструкции.

**Определение 1.** Разногласиями между пользователями  $i \in U$  и  $j \in U$  в момент времени  $t$  будем называть величину

$$d_{ij}(t) = d\left(x_i(t), x_j(t)\right) = |x_i(t) - x_j(t)|.$$

Заметим, что  $d_{ij}(t) = d_{ji}(t)$ .

Множество разногласий между пользователем  $i \in U$ , для которого  $|V_i(t)| > 0$ , и его друзьями в момент времени  $t$  обозначим через  $X_i(t) = \{d_{ij}(t) \mid j \in V_i(t)\}$ . Пусть  $j_0 \in V_i(t)$  и пусть  $|X_i(t)| = m + 1$ . Расположим элементы множества  $X_i(t)$  в порядке возрастания:

$$(1) \quad d_{ij_1}(t) < \dots < d_{ij_0}(t) < \dots < d_{ij_m}(t).$$

**Определение 2.** Рангом разногласий между пользователями  $i \in U$  и  $j_0 \in V_i(t)$  будем называть позицию  $p_{ij_0}(t) \in [m + 1]$ , которую число  $d_{ij_0}(t)$  занимает в последовательности (1).

Чем больше ранг разногласий  $p_{ij}(t)$ , тем выше разногласия  $d_{ij}(t)$  и наоборот. Заметим, что кардинальное число множества  $X_i(t)$  лимитировано числом друзей пользователя  $i$ :  $m + 1 \leq |V_i(t)|$ .

Введем величину, описывающую идеологическую напряженность, имеющую место между пользователем  $i$  и его друзьями.

**Определение 3.** Пусть пользователь  $i \in U$  имеет хотя бы одного друга ( $|V_i(t)| > 0$ ). Разногласиями пользователя  $i$  с его друзьями в момент времени  $t$  назовем

$$(2) \quad d_i(t) = d\left(x_i(t), \{x_j(t)\}_{j \in V_i(t)}\right) = \frac{1}{|V_i(t)|} \sum_{j \in V_i(t)} d_{ij}(t).$$

Справедливо следующее утверждение.

**Утверждение 1.**  $d_i(t) \in [0,1]$ .  $d_i(t) = 1$  тогда и только тогда, когда разногласия пользователя  $i$  с каждым из его друзей максимальны (равны единице), а  $d_i(t) = 0$  тогда и только тогда, когда мнение пользователя  $i$  совпадает со мнением каждого из его друзей (все они равны в совокупности).

Если  $d_i(t) = 0$ , то тогда можно утверждать, что пользователь  $i$  оказался в эхо-камере: среди его друзей нет тех, чьи мнения хоть сколько-нибудь отличались бы от его собственного. Диаметрально противоположная ситуация  $d_i(t) = 1$  свидетельствует о том, что  $x_i(t) \in \{0,1\}$ , а друзья пользователя  $i$  придерживаются полярной позиции.

## 1.3 Стандартизированный ранг исчезнувшей связи

Рассмотрим множество

$$U_- = \{i \in U \mid (|V_i^-| > 0) \& (|V_i(t_1)| > 1) \& (|V_i(t_2)| > 0)\},$$

включающее в себя пользователей, которые за рассматриваемый промежуток времени удалили хотя бы одного из своих друзей. Ограничение  $|V_i(t_2)| > 0$  позволяет снизить вероятность того, что

исчезновение связи (связей) связано с настройками приватности пользователя  $i$ , а неравенство  $|V_i(t_1)| > 1$  отбрасывает случай, когда пользователь  $i$  удалил из друзей единственного своего друга.

Пусть  $i_0 \in U_-$ . Тогда существует  $j_0 \in U$  такое, что  $j_0 \in V_{i_0}^-$ . Отметим, что в силу симметричности матриц  $\mathbf{A}(t_1)$  и  $\mathbf{A}(t_2)$  выполнено  $j_0 \in U_-$ . Исчезнувшей связи  $(i_0, j_0)$  мы ставим в соответствие величину

$$(3) \quad p_{i_0 j_0}^* = \left( \frac{p_{i_0 j_0}(t_1)}{m+1} - \frac{1}{m+1} \sum_{k=1}^{m+1} \frac{k}{m+1} \right) / \frac{m}{2(m+1)},$$

которую будем называть *стандартизированным рангом* этой связи. Легко доказать следующее утверждение.

**Утверждение 2.** Стандартизированный ранг связи является строго монотонной функцией  $p_{i_0 j_0}$ , принимает значения из промежутка  $[-1, 1]$ . При этом,  $p_{i_0 j_0}^* = -1$  тогда и только тогда, когда  $p_{i_0 j_0}(t_1) = 1$ ,  $p_{i_0 j_0}^* = 1$  тогда и только тогда, когда  $p_{i_0 j_0}(t_1) = m + 1$ .

Отметим, что несмотря на симметричность связи “дружба”, метрика (3) в общем случае не удовлетворяет условию

$$p_{i_0 j_0}^* = p_{j_0 i_0}^*.$$

В связи с этим, для каждой исчезнувшей связи  $(i_0, j_0)$  мы отдельно находили значения  $p_{i_0 j_0}^*$  и  $p_{j_0 i_0}^*$ . Если пользователь  $i_0 \in U_-$  удалил из друзей несколько пользователей, например  $j_0 \in U_-$  и  $k_0 \in U_-$ , то отдельно высчитывались показатели  $p_{i_0 j_0}^*$  и  $p_{i_0 k_0}^*$  (и, соответственно,  $p_{j_0 i_0}^*$  и  $p_{k_0 i_0}^*$ ).

При фиксированном  $m$ , чем больше значение  $p_{ij}^*$ , тем выше разногласия между пользователями  $i$  и  $j$  по сравнению с разногласиями между пользователем  $i$  и другими его друзьями. Напротив, малое значение  $p_{ij}^*$  говорит, о том, что пользователь  $i$  удалил из друзей пользователя, с которым у него были относительно небольшие разногласия.

**Пример 1.** Пусть  $i_0 \in U_-$ ,  $V_{i_0}(t_1) = \{j_0, j_1, j_2, j_3, j_4\}$ ,  $j_0 \in V_{i_0}^-$ ,  $x_{i_0}(t_1) = 0.2$ ,  $x_{j_0}(t_1) = 0.6$ ,  $x_{j_1}(t_1) = 0.1$ ,  $x_{j_2}(t_1) = 0.4$ ,  $x_{j_3}(t_1) = 0.55$ ,  $x_{j_4}(t_1) = 0.65$ . Тогда  $d_{i_0 j_0}(t_1) = 0.4$ ,  $d_{i_0 j_1}(t_1) = 0.1$ ,  $d_{i_0 j_2}(t_1) = 0.2$ ,  $d_{i_0 j_3}(t_1) = 0.4$ ,  $d_{i_0 j_4}(t_1) = 0.45$  и  $X_{i_0}(t_1) = \{0.4, 0.1, 0.2, 0.35, 0.45\}$ . В этом случае  $m = 4$  и  $p_{i_0 j_0} = 4$ . Тогда в соответствии с (3) имеем  $p_{i_0 j_0}^* = 0.5$ .

Если факт исчезновения дружбы между пользователями никак не связан с их разногласиями, то тогда стоит ожидать, что распределение величины  $p_{ij}^*$  будет симметричным со средним выборочным, равным нулю. Однако данное соображение верно только в том случае, если для любого  $i \in U_-$  выполнено

$$(4) \quad |X_i(t_1)| = |V_i(t_1)|$$

(как в примере 1), что наглядно иллюстрирует следующий пример.

**Пример 2.** Пусть  $i_0 \in U_-$ ,  $V_{i_0}(t_1) = \{j_0, j_1, j_2, j_3, j_4\}$ ,  $j_0 \in V_{i_0}^-$ ,  $x_{i_0}(t_1) = 0.2$ ,  $x_{j_0}(t_1) = 0.6$ ,  $x_{j_1}(t_1) = 0.1$ ,  $x_{j_2}(t_1) = 0.4$ ,  $x_{j_3}(t_1) = 0.6$ ,  $x_{j_4}(t_1) = 0.6$ . Тогда  $d_{i_0 j_0}(t_1) = 0.4$ ,  $d_{i_0 j_1}(t_1) = 0.1$ ,  $d_{i_0 j_2}(t_1) = 0.2$ ,  $d_{i_0 j_3}(t_1) = 0.4$ ,  $d_{i_0 j_4}(t_1) = 0.45$  и  $X_{i_0}(t_1) = \{0.4, 0.1, 0.2\}$ . В этом случае  $m = 2$  и  $p_{i_0 j_0} = 3$ . Тогда  $p_{i_0 j_0}^* = 1$ . Однако в случае, если пользователь  $i_0$  случайного выбирает, с кем потерять связь, вероятность выбрать пользователя, с которым у него разногласия 0.4 равняется 0.5 и если провести значительное число испытаний со случайным выбором, то среднее выборочное величины  $p_{ij}^*$  будет больше 0.

В связи с этим, показатели (3) рассчитывались только для тех пользователей  $i \in U_-$ , для которых выполнено ограничение (4). Множество таких пользователей обозначим  $U_-^{nr}$ . Отметим, что  $\frac{|U_-^{nr}|}{|U_-|} = 1$  с точностью до трех знаков после запятой. Дополнительно был проведен эксперимент, в рамках которого для каждого пользователя  $i \in U_-^{nr}$  и для каждого  $j \in V_i^-$  случайно выбирался элемент  $j_{rand}$  из множества  $V_i(t_1)$  и высчитывался показатель  $p_{i j_{rand}}^*$ . В результате, имитировалась ситуация, при которой пользователь “случайно выбирает”, с кем из друзей разорвать связь.

#### 1.4 Влияние связей на разногласия

Рассмотрим пользователя  $i_0 \in U_-$ . Пусть  $j_0 \in V_{i_0}^-$ . Потеря дружбы с пользователем  $j_0$  сделает разногласия пользователя  $i_0$  с его друзьями равными

$$d\left(x_{i_0}(t_1), \{x_j(t_1)\}_{j \in V_{i_0}(t_1) \setminus \{j_0\}}\right) = \frac{1}{|V_{i_0}(t_1)|-1} \sum_{j \in V_{i_0}(t_1) \setminus \{j_0\}} d_{i_0j}(t_1),$$

однако при условии, что остальные связи и мнения пользователей останутся неизменными. Исчезнувшей связи  $(i_0, j_0)$  мы ставим в соответствие величину

$$(5) \quad \Delta d_{i_0j_0}^- = \frac{1}{|V_{i_0}(t_1)|-1} \sum_{j \in V_{i_0}(t_1) \setminus \{j_0\}} d_{i_0j}(t_1) - d_{i_0}(t_1),$$

описывающую, как изменились бы разногласия пользователя  $i_0$  с его друзьями в результате потери дружбы с пользователем  $j_0$  при условии, что остальные связи и мнения пользователей остались бы неизменными.

Метрика (5) удовлетворяет условию  $\Delta d_{ij}^- \in [-1, 1]$ . Она положительна, если исчезновение связи  $(i_0, j_0)$  привело бы к увеличению уровня разногласий и отрицательна, если исчезновение связи привело бы к уменьшению уровня разногласий. Отметим, что в общем случае  $\Delta d_{ij}^- \neq \Delta d_{ji}^-$ . Метрика (5) была посчитана для каждого  $i \in U_-$  и каждого  $j \in V_i^-$ . Кроме того, как и в случае с метрикой (3), для каждого  $i \in U_-$  и каждого  $j \in V_i^-$  случайно выбирался элемент  $j_{rand}$  из множества  $V_i(t_1)$  и высчитывался показатель  $\Delta d_{ij_{rand}}^-$ .

Рассмотрим множество

$$U_+ = \{i \in U \mid (|V_i^+| > 0) \& (|V_i(t_1)| > 0)\}.$$

Это множество состоит из пользователей, которые за рассматриваемый промежуток времени обзавелись хотя бы одним новым другом. Ограничение  $|V_i(t_1)| > 0$  позволяет снизить вероятность того, что это появление связано с отключением настроек приватности пользователя, а также сопоставлять нового друга с уже имеющимися.

Пусть  $i \in U$  и  $|V_i(t_1)| > 0$  и  $j_0 \notin V_i(t_1)$ . Величина

$$d\left(x_{i_0}(t_1), \{x_j(t_1)\}_{j \in V_{i_0}(t_1) \cup \{j_0\}}\right) = \frac{1}{|V_i(t_1)|+1} \left( \sum_{j \in V_i(t_1)} d_{ij}(t_1) + d_{ij_0}(t_1) \right),$$

принимая значения из промежутка  $[0, 1]$ , описывает, чему были бы равны разногласия пользователя  $i_0$  с его друзьями в случае, если множество его друзей пополнилось бы пользователем  $j_0$  при условии, что остальные связи и мнения пользователей остались бы неизменными.

Пусть  $i_0 \in U_+$ . Тогда в силу симметричности матриц  $\mathbf{A}(t_1)$  и  $\mathbf{A}(t_2)$  существует  $j_0 \in U_+$  такое, что  $j_0 \in V_{i_0}^+$ , что означает появление связи  $(i_0, j_0)$  за рассматриваемый промежуток времени. Этой связи мы ставим в соответствие величину

$$(6) \quad \Delta d_{i_0j_0}^+ \stackrel{\text{def}}{=} \frac{1}{|V_{i_0}(t_1)|+1} \left( \sum_{j \in V_{i_0}(t_1)} d_{i_0j}(t_1) + d_{i_0j_0}(t_1) \right) - d_{i_0}(t_1),$$

которая описывает изменение разногласий пользователя  $i_0$  с его друзьями, связанное с приобретением дружбы с пользователем  $j_0$  при условии, что остальные связи и мнения пользователей остались неизменными.

Метрика (6) удовлетворяет условию  $\Delta d_{i_0j_0}^+ \in [-1, 1]$ . Она положительна, если появление связи  $(i_0, j_0)$  привело бы к увеличению уровня разногласий и отрицательна, если результатом появления связи стало бы уменьшение уровня разногласий. Значение величины (6) было посчитано для каждого  $i \in U_+$  и каждого  $j \in V_i^+$ .

## 1.5 Динамика разногласий пользователей

Рассмотрим пользователя  $i \in U$ , для которого выполнено  $|V_i(t_1)| > 0$  и  $|V_i(t_2)| > 0$ . Введем следующую метрику:

$$(7) \quad \Delta d_i \stackrel{\text{def}}{=} d\left(x_i(t_1), \{x_j(t_1)\}_{j \in V_i(t_2)}\right) - d_i(t_1).$$

Метрика (7) описывает изменение разногласий пользователя  $i$  с его друзьями в результате эволюции множества его друзей  $V_i(t_1) \rightarrow V_i(t_2)$  при условии, что мнения пользователей не меняются. Отметим, что  $\Delta d_i \in [-1, 1]$ .

## 1.6 Коэффициент ассортативности и идеологическая гомофилия

Рассмотрим простой неориентированный невзвешенный граф  $G$ , который описывается матрицей смежности  $A = [a_{ij}] \in \{0,1\}^{n \times n}$ , а некоторая характеристика его вершин задается вектором  $u = [u_1 \dots u_n]^T \in \mathbb{R}^n$ . Введем метрику, которая позволит оценить степень гомофилии в графе  $G$  [4].

**Определение 4.** Коэффициентом ассортативности графа  $G$  будем называть следующую величину:

$$r(G) = r(A, u) = \frac{\sum_{i,j}(a_{ij}-k_i k_j/2m)u_i u_j}{\sum_{i,j}(k_i \delta_{ij}-k_i k_j/2m)u_i u_j}$$

где  $k_i$  – степень вершины  $i$  ( $k_i = \sum_{j=1}^n a_{ij}$ ),  $m$  – число ребер в графе  $G$ , а  $\delta_{ij}$  – символ Кронекера.

Коэффициент ассортативности  $r(G)$  принимает значения из промежутка  $[-1,1]$  и описывает, каким образом коррелируют характеристики смежных вершин графа  $G$ . Таким образом, коэффициент ассортативности также можно интерпретировать как меру идеологической напряженности между пользователями и их друзьями.

### Результаты

На рисунке 1 (слева) представлены гистограммы, показывающие распределения стандартизованного ранга исчезнувшей связи для случаев реальных данных  $p_{ij}^*$  (серые столбцы) и случайно сгенерированных данных  $p_{ij}^*_{rand}$  (черные столбцы). Статистики распределений  $p_{ij}^*$  и  $p_{ij}^*_{rand}$  приведены в таблице 1 (с точностью до четырех знаков после запятой). Заметим, что в зоне отрицательных значений преимущество у показателя  $p_{ij}^*_{rand}$  (черные столбцы выше серых), а в зоне положительных значений – у  $p_{ij}^*$  (серые столбцы выше черных). U-критерий Манна-Уитни свидетельствует о том, что распределение величины  $p_{ij}^*$  находится правее распределения величины  $p_{ij}^*_{rand}$ .

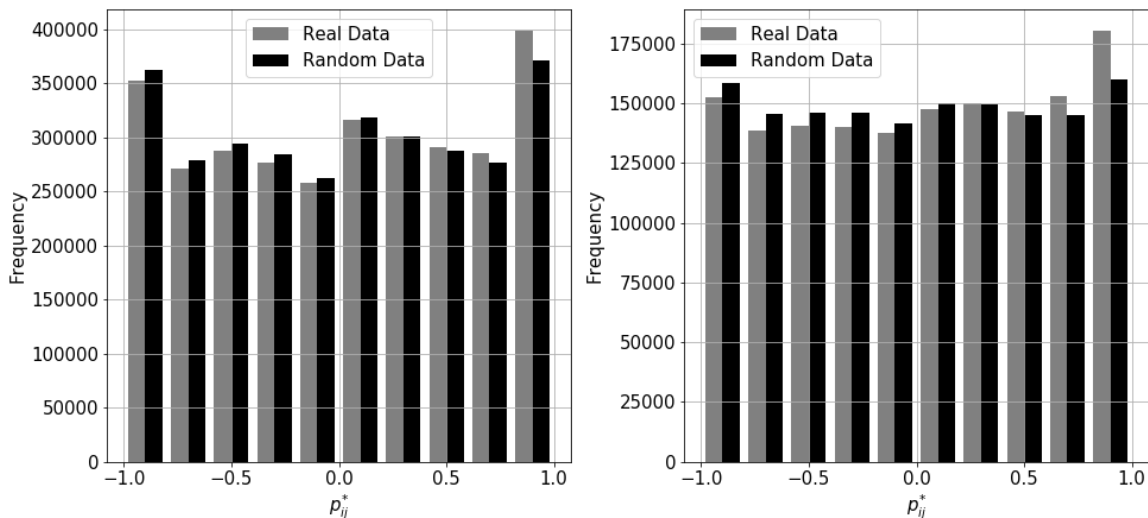


Рис. 1. Распределение  $p_{ij}^*$  (серые столбцы) и  $p_{ij}^*_{rand}$  (черные столбцы) (слева). Распределение  $p_{ij}^*$  (серые столбцы) и  $p_{ij}^*_{rand}$  (черные столбцы) для пользователей, у кого  $|V_{i_0}(t_1)| > 30$  (справа).

Таблица 1. Статистики распределений  $p_{ij}^*$  и  $p_{ij}^*_{rand}$

Статистика	Распределение $p_{ij}^*$	Распределение $p_{ij}^*_{rand}$
Среднее выборочное	0.0177	0.0005
Стандартное отклонение	0.613	0.6096
Минимальное значение	-1	-1
Медиана	0	0
Максимальное значение	1	1

Особый интерес на рисунке 1 (слева) представляют крайние два пика, а также срединный пик. Их присутствие связано с тем, что в случае пользователей с малым числом друзей величина  $p_{ij}^*$  с высокой

вероятностью попадает в один из этих столбцов. Если же построить аналогичные гистограммы для пользователей, у кого не менее 30 друзей (см. рисунок 1, (справа)), то они будут иметь более “гладкую форму”. При этом, правый пик, соответствующий пользователям, потерявшим дружбу с теми, с кем у них были максимальные разногласия, будет все так же выделяться своей высотой.

Распределение величины  $\Delta d_{ij}^-$  представлено на рисунке 2 (слева). Статистики распределений  $\Delta d_{ij}^-$  и  $\Delta d_{ijrand}^-$  представлены в таблице 2 (с точностью до четырех знаков после запятой). U-критерий Манна-Уитни свидетельствует о том, что распределение величины  $d_{i_0}(t_1)$  из (5) находится правее распределения величины  $\frac{1}{|V_{i_0}(t_1)|-1} \sum_{j \in V_{i_0}(t_1)/\{j_0\}} d_{i_0j}(t_1)$ , а распределение величины  $\Delta d_{ij}^-$  находится левее распределения величины  $\Delta d_{ijrand}^-$ .

Таблица 2. Статистики распределений  $\Delta d_{ij}^-$  и  $\Delta d_{ijrand}^-$

Статистика	Распределение $\Delta d_{ij}^-$	Распределение $\Delta d_{ijrand}^-$
Среднее выборочное	-0.0001	0
Стандартное отклонение	0.0128	0.0128
Минимальное значение	-0.4694	-0.4534
Медиана	0.0003	0.0004
Максимальное значение	0.4694	0.4694

Распределение величины  $\Delta d_{ij}^+$  представлено на рисунке 2 (справа). Статистики распределения  $\Delta d_{ij}^+$  представлены в таблице 3 (с точностью до четырех знаков после запятой). U-критерий Манна-Уитни свидетельствует о том, что распределение величины  $d_{i_0}(t_1)$  из (6) находится левее распределения величины  $\frac{1}{|V_{i_0}(t_1)|+1} (\sum_{j \in V_{i_0}(t_1)} d_{i_0j}(t_1) + d_{i_0j_0}(t_1))$ .

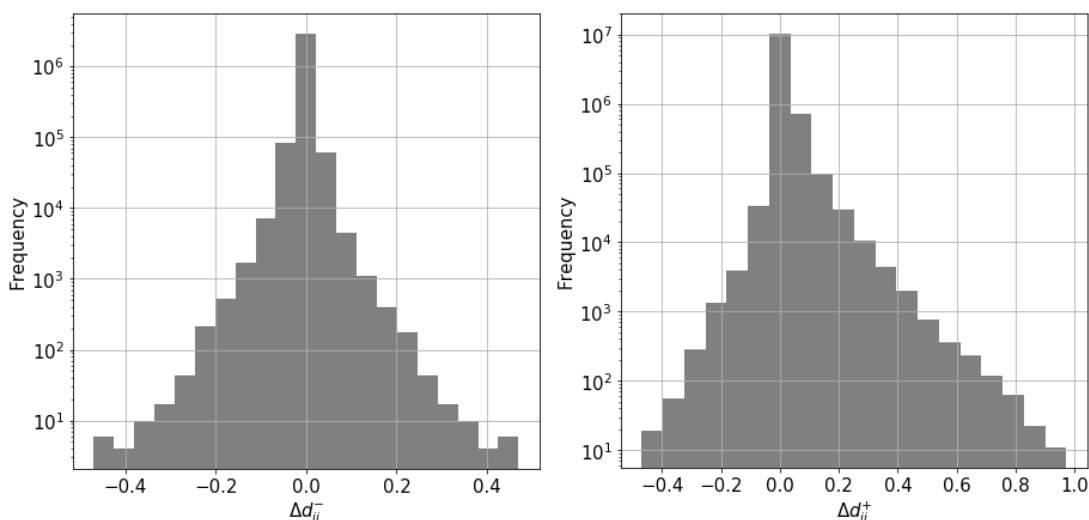


Рис. 2. Распределение величины  $\Delta d_{ij}^-$  (слева). Распределение величины  $\Delta d_{ij}^+$  (справа).

Таблица 3. Статистики распределений  $\Delta d_{ij}^+$  и  $\Delta d_i$

Статистика	Распределение $\Delta d_{ij}^+$	Распределение $\Delta d_i$
Среднее выборочное	0.0107	-0.0002
Стандартное отклонение	0.0268	0.0235
Минимальное значение	-0.4697	-0.6463
Медиана	0.0039	0
Максимальное значение	0.9722	0.6565

На рисунке 3 представлено распределение величины  $\Delta d_i$ . Статистики распределения  $\Delta d_i$  представлены в таблице 3 (с точностью до четырех знаков после запятой). Согласно U-критерию Манна-Уитни, распределение величины  $d_i(t_1)$  из (7) находится правее распределения величины  $d(x_i(t_1), \{x_j(t_1)\}_{j \in V_i(t_2)})$ .

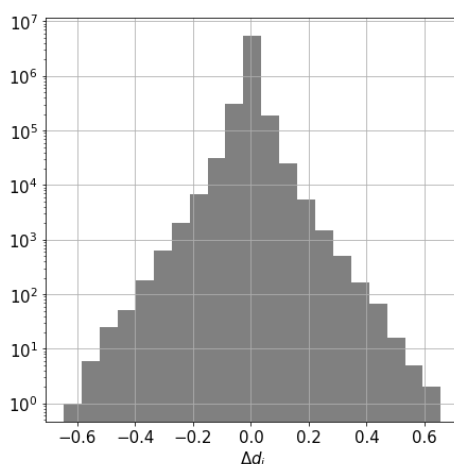


Рис. 3. Распределение величины  $\Delta d_i$ .

В конце данного раздела приведем результаты измерения коэффициента ассортативности (с точностью до четырех знаков после запятой):

$$\begin{aligned}
 r(A(t_1), x(t_1)) &= 0.2499, \\
 r(A(t_1) - \Delta A_-, x(t_1)) &= 0.2502, \\
 r(A(t_1) + \Delta A_+, x(t_1)) &= 0.2464, \\
 r(A(t_2), x(t_1)) &= 0.2467.
 \end{aligned}$$

### Обсуждение результатов

Рассматриваемый социальный граф обладает выраженным свойством гомофилии, поскольку коэффициент ассортативности к моменту начала наблюдений за системой равен 0.2499. Как было отмечено в [10], данная система могла прийти к такому состоянию за счет следующих механизмов: (1) в силу социального влияния индивиды в ходе общения снижают разногласия друг с другом, в результате чего их мнения становятся все ближе и ближе друг к другу; (2) люди формируют новые социальные связи и разрывают старые таким образом, чтобы общаться только с теми, чьи взгляды похожи на их собственные. При этом, согласно [10], каждый из этих механизмов должен был внести определенный вклад в итоговый результат.

Из представленных результатов, можно сделать вывод, что пользователи реже, чем следовало бы ожидать при случайном выборе, удаляют из друзей тех, с кем у них относительно небольшие разногласия и чаще, чем следовало бы ожидать при случайном выборе, теряют дружбу с теми, с кем у них относительно большие разногласия. Данный вывод подтверждает ту часть механизма 2, которая связана с удалением связей. Тот факт, что удалив мысленно все исчезнувшие к моменту времени  $t_2$  связи из графа дружбы, каким он был на момент времени  $t_1$ , оставив остальные связи и мнения пользователей неизменными, мы увеличим коэффициент ассортативности, также служит подтверждением сделанному выводу. Следовательно, то, как пользователи, разрывают связи дружба друг с другом, может привести к формированию эхо-камер, в рамках которых пользователи будут иметь в друзьях только тех, чьи мнения похожи на их собственные.

Распределение величины  $\Delta d_{ij}^{\pm}$  свидетельствует о том, что появление новых связей ведет к увеличению разногласий между пользователями и их друзьями. Кроме того, мысленное добавление всех новых связей, появившихся к моменту времени  $t_2$ , к графу дружбы, каким он был на момент времени  $t_1$ , уменьшит коэффициент ассортативности. Отсюда можно сделать вывод, что появление новых связей приводит к росту идеологической напряженности между пользователями и их друзьями и поэтому является механизмом, препятствующим возникновению информационных пузырей. Данный вывод не противоречит предположению о том, что пользователи чаще формируют связи с тем, с кем у них похожие мнения, поскольку в рамках нашего исследования мы сравниваем мнения новых друзей пользователя с мнениями его текущих друзей, а не с мнениями тех, кто мог бы быть его другом (одной из причин этому является то, что данное множество трудно идентифицируемо). Вполне может быть, что пользователи, формируя новые связи, выбирают в итоге тех, чьи взгляды похожи на их собственные, однако этого недостаточно для снижения разногласий с друзьями и роста коэффициента ассортативности.



Анализ композиции процессов появления и удаления связей между пользователями более затруднен, поскольку, с одной стороны, полученные результаты свидетельствуют о том, что разногласия пользователей с их друзьями снизились, но при этом коэффициент ассортативности также уменьшился, что свидетельствует о росте разногласий. Здесь необходимо отметить, что эти две метрики по-разному оценивают систему, поэтому полученные расхождения не должны вводить в заблуждение. Однако, порядок изменений данных величин различен: в то время как среднее  $\Delta d_i$  составляет  $-0.0002$ , коэффициент ассортативности изменился на  $r(\mathbf{A}(t_2), \mathbf{x}(t_1)) - r(\mathbf{A}(t_1), \mathbf{x}(t_1)) = -0.0032$ , что на порядок больше. Изменение коэффициента ассортативности не может выйти за пределы промежутка  $[-2, 2]$ , а величина  $\Delta d_i$  лежит в пределах от минус единицы до единицы, поэтому такое сравнение уместно. Мы делаем вывод, что эволюция графа дружбы при неизменных мнениях пользователей приводит к росту уровня идеологической напряженности между пользователями и их друзьями. В результате, риск для пользователя оказаться в информационном пузыре снижается.

В данном исследовании мы опираемся на данные о двух измерениях состояний системы. Для полноты картины необходимо воспроизвести полученные результаты в рамках эксперимента с большим числом измерений. Кроме того, мы рассматривали фиксированную выборку пользователей, которые были активны в течение всего процесса наблюдения. Вместе с тем, число пользователей ВКонтакте постоянно меняется: некоторые аккаунты перестают быть активными, появляются новые аккаунты, что должно быть учтено в следующих исследованиях. Также не стоит забывать про динамику мнений пользователей, которая также может привести к снижению разногласий между пользователями и их друзьями [3, 5, 6, 15]. Следующим важным шагом будет понять, какой из процессов (эволюция графа дружбы или мнений пользователей) оказывает большее влияние на систему.

Работа выполнена при частичной поддержке РФФИ (проект 18-29-22042).

## Литература

1. Chandrasekhar A.G., Jackson M.O. A network formation model based on subgraphs // Available at SSRN 2660381. – 2016.
2. Currarini S., Jackson M.O., Pin P. An economic model of friendship: Homophily, minorities, and segregation // *Econometrica*. – 2009. – Vol. 77. – №. 4. – P. 1003-1045.
3. Castellano C., Fortunato S., Loreto V. Statistical physics of social dynamics // *Reviews of modern physics*. – 2009. – Vol. 81. – №. 2. – P. 591.
4. Newman M. *Networks*. – Oxford university press, 2018.
5. Flache A. et al. Models of social influence: Towards the next frontiers // *Journal of Artificial Societies and Social Simulation*. – 2017. – Vol. 20. – №. 4.
6. Proskurnikov A.V., Tempo R. A tutorial on modeling and analysis of dynamic social networks. Part I // *Annual Reviews in Control*. – 2017. – Vol. 43. – P. 65-79.
7. Nickerson R.S. Confirmation bias: A ubiquitous phenomenon in many guises // *Review of general psychology*. – 1998. – Vol. 2. – №. 2. – P. 175-220.
8. Sherif M., Hovland C.I. *Social judgment: Assimilation and contrast effects in communication and attitude change*. – 1961.
9. Sherif M., Sherif C.W. Attitude as the individual's own categories: The social judgment-involvement approach to attitude and attitude change // *Attitude, ego-involvement, and change*. – 1967. – P. 105-139.
10. Holme P., Newman M. Nonequilibrium phase transition in the coevolution of networks and opinions // *Physical Review E*. – 2006. – Vol. 74. – №. 5. – P. 056108.
11. Sasahara K. et al. On the inevitability of online echo chambers // arXiv preprint arXiv:1905.03919. – 2019.
12. Petrov A.P., Proncheva O.G. Modeling Propaganda Battle: Decision-Making, Homophily, and Echo Chambers // *Conference on Artificial Intelligence and Natural Language*. – Springer, Cham, 2018. – P. 197-209.
13. Dandekar P., Goel A., Lee D.T. Biased assimilation, homophily, and the dynamics of polarization // *Proceedings of the National Academy of Sciences*. – 2013. – Vol. 110. – №. 15. – P. 5791-5796.
14. Kozitsin I.V. et al., 'Modeling Political Preferences of Russian Users Exemplified by the Social Network Vkontakte' // *Mathematical Models and Computer Simulations*, – 2020. – Vol. 12. – №. 12 – P. 185–194.
15. Kozitsin I.V., Belolipetskii A.A. Opinion convergence in the Krasnoshchekov model // *The Journal of Mathematical Sociology*. – 2019. – Vol. 43. – №. 2. – P. 104-121.