

О ПОДХОДАХ К ВЫЯВЛЕНИЮ КАНАЛОВ РАСПРОСТРАНЕНИЯ ИНФОРМАЦИИ В ОНЛАЙНОВЫХ СОЦИАЛЬНЫХ СЕТЯХ

Губанов Д.А., Чхартишвили А.Г.

Институт проблем управления им. В.А. Трапезникова РАН

Россия, Москва, ул. Профсоюзная, д. 65

dmitry.a.g@gmail.com, sandro_ch@mail.ru

Аннотация: В работе предлагается подход к выявлению каналов распространения информации в онлайн-социальных сетях. Подход использует акциональную модель влияния для определения наиболее значимых каналов.

Ключевые слова: онлайн-социальные сети, акциональная модель влияния, каналы распространения информации.

Введение

В последнее время заметно возросла роль онлайн-социальных сетей (Facebook, Twitter и др.) в жизни общества. Важным аспектом информационно-аналитической работы с ними [1] является оценка влияния пользователей (см., напр., [2-6]), поскольку влиятельные пользователи (лидеры общественного мнения) в существенной мере определяют тематику и содержание информационных потоков в социальной сети.

Однако наряду с пользователями влиянием может обладать и связь между пользователями. Например, тот факт, что один пользователь регулярно реагирует на посты (сообщения в социальной сети) другого пользователя может оказаться значимым с точки зрения распространения информации. В докладе предлагается подход к выявлению таких связей и состоящих из них каналов распространения информации.

1 Акциональная модель влияния

В данном разделе мы кратко опишем, в основном следуя [2], акциональную модель распространения действий в социальной сети.

Участниками сети будем считать агентов из фиксированного множества $N = \{1, 2, \dots, n\}$, которые совершают действия из фиксированного множества возможных видов действий $K = \{1, 2, \dots, k\}$ в те или иные моменты времени из интервала T . Видом действия может быть создание (написание) поста, создание комментария к посту и т. д. Множество действий (создание конкретного поста, комментария и т. д.) обозначим за Δ и далее будем считать конечным.

Каждое действие $a \in \Delta$ характеризуется тремя параметрами – совершившим его агентом, видом действия и моментом времени t , в который действие было совершено. Функция $\alpha(a)$, которая каждому действию $a \in \Delta$ ставит в соответствие совершившего его агента $\alpha \in N$.

Пусть на множестве действий задано бинарное отношение частичного порядка « a является причиной b » (или, что будем далее считать эквивалентным, « b является последствием a »), обозначаемое следующим образом: $a \rightarrow b$. 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 что a и b являются *последовательными действиями*. Будем обозначать это следующим образом: $a \downarrow b$.

Для каждого действия $a \in \Delta$ определено множество действий $\pi(a)$, являющихся его последствием: $\pi(a) = \{b \in \Delta \mid a \rightarrow b\}$.

Значимость множества действий задана функцией $\Phi(S): 2^\Delta \rightarrow [0, +\infty)$. Будем считать, что значимость является монотонной функцией (если $A \subseteq B$, то $\Phi(A) \leq \Phi(B)$), и что хотя бы какие-то действия обладают положительной значимостью ($\Phi(\Delta) > 0$).

2 Каналы распространения информации

Каждая последовательность действий a_1, a_2, \dots, a_m , таких, что каждое действие начиная со второго является прямым последствием предыдущего (т.е. $a_1 \downarrow a_2 \downarrow \dots \downarrow a_m$), определяет канал распространения информации – последовательность агентов, совершивших данные действия: $i_1 = \alpha(a_1), \dots, i_m = \alpha(a_m)$.

Определим множество действий, совершенных j -м агентом под прямым влиянием i -го агента:

$$\delta_{ij} = \{b \in \delta_j \mid \exists a \in \delta_i \ a \downarrow b\}.$$

Значимость *непосредственной связи* между i -м и j -м агентами определим как нормированную значимость последствий действий, совершенных j -м агентом под прямым влиянием i -го агента:

$$\xi_{ij} = \frac{\Phi(\pi(\delta_{ij}))}{\Phi(\Delta)}.$$

Каналом распространения можно считать такую последовательность агентов i_1, i_2, \dots, i_m таких, что, во-первых, агенты в последовательности различны, а, во-вторых, значимость связи между каждыми двумя соседними агентами в последовательности достаточно велика:

$$\xi_{12} \geq h, \xi_{23} \geq h, \dots, \xi_{m-1,m} \geq h$$

для некоторого параметра h . При этом величину

$$H = \min \{ \xi_{12}, \xi_{23}, \dots, \xi_{m-1,m} \}$$

назовем *интенсивностью* канала i_1, i_2, \dots, i_m .

Увеличивая значение параметра h и находя соответствующие (непосредственные) связи между агентами, можно выявлять все более интенсивные каналы распространения информации.

3 Пример расчета каналов распространения информации

Для дальнейших расчетов введем следующее определение. Назовем *связной структурой каналов с интенсивностью H* – максимальное по включению множество агентов сети, в котором любая пара агентов включена в некоторый канал с интенсивностью не меньшей H .

Рассмотрим пример выявления связанных структур каналов в онлайн-социальной сети ВКонтакте (vk.com)¹¹⁵. Будем считать, что значимыми для центра являются посты, в которых содержится ключевое слово «Назарбаев» (в любых падежных формах), а также их репосты, комментарии и лайки к ним. В качестве интервала T будем рассматривать 2015 год (т.е. T – это промежуток от 0 часов 0 минут 0 секунд 1 января 2015 года до 23 часов 59 минут 59 секунд 31 декабря 2015 года). Будем считать, что значимость всех действий (постов, репостов, комментариев и лайков) для центра одинакова. В соответствии с формулами раздела 2 можно рассчитать значимость непосредственных связей между пользователями.

На рис. 1 приведено кумулятивное распределение количества связей для полученной выборки по их значимости.

Из рис. 1 видно, что лишь небольшое число связей обладает существенной значимостью, в частности, значимость 99% связей меньше 10.

Для интенсивности $H = 250$ найдены связанные структуры каналов, которые приведены на рис. 2. (величина узла нелинейно отражает влияние пользователя по начальным действиям (см. [2]), а толщина связи – ее значимость).

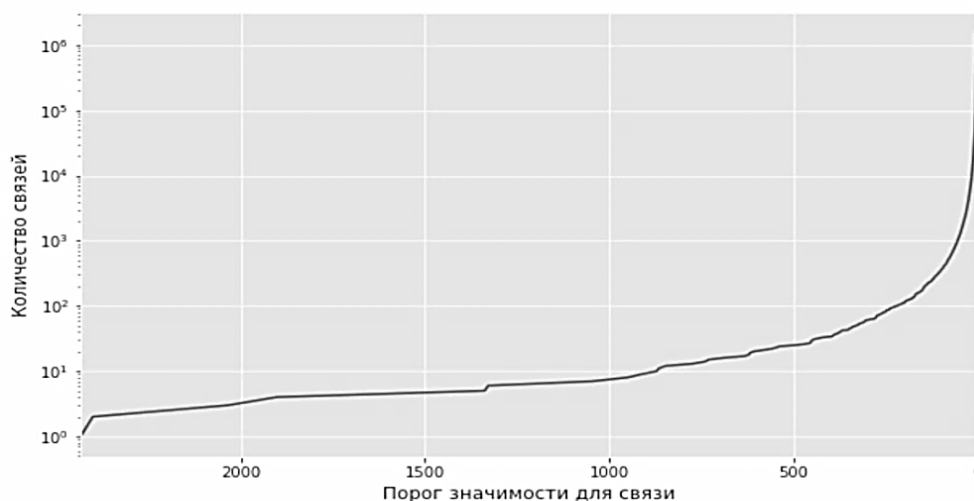


Рис. 1. Кумулятивное распределение количества связей со значимостью не меньшей порога h

Рис. 2 показывает, что связанные структуры каналов в данном случае в основном имеют простую форму: состоят из двух узлов. Однако встречаются и более сложные структуры, содержательное объяснение появления которых является перспективным направлением дальнейших исследований.

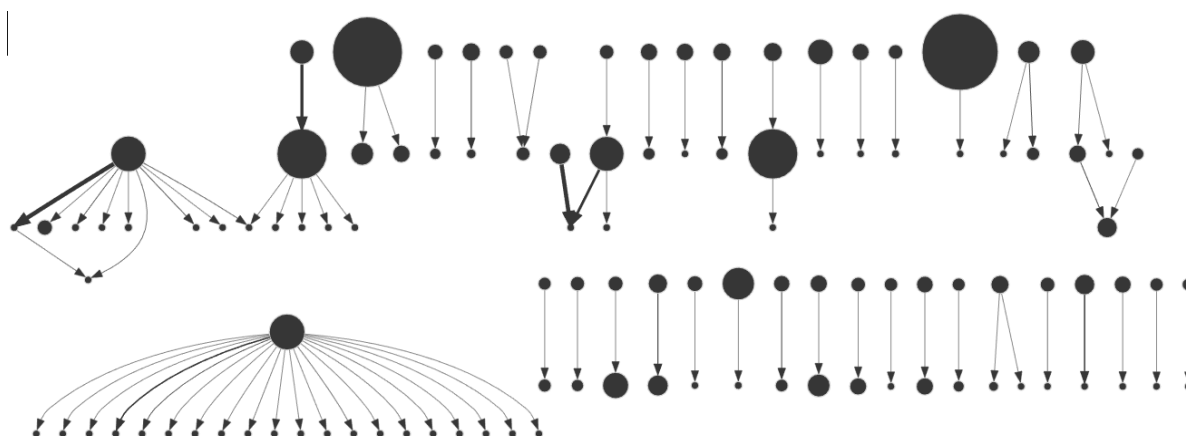


Рис. 2. Графы связанных структур каналов ($H = 250$)

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

¹¹⁵ Анонимизированные данные были предоставлены для исследований компанией DSS Lab (dss-lab.ru).

Литература

1. Губанов Д.А., Чхртишвили А.Г. Концептуальный подход к анализу онлайн-социальных сетей // Управление большими системами. 2013. Вып. 45. С. 222-236.
2. Губанов Д.А., Чхртишвили А.Г. Влияемость пользователей и метапользователей социальной сети // Проблемы управления. 2016. №6. С. 12-17.
3. Губанов Д.А., Новиков Д.А., Чхртишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства. М.: МЦНМО, 2018. 224 с.
4. Aggarwal C.C. Social Network Data Analytics. Boston/Dordrecht/London: Kluwer Academic Publishers, 2011. 502p.
5. Kempe D., Kleinberg J., Tardos E. Maximizing the spread of influence through a social network // Proc. of the 9th ACM SIGKDD Intern. Conf. on Knowledge discovery and data mining. 2003. P. 137—146.
6. Weng J., Lim E.-P., Jiang J., He Q. Twitterrank: finding topic-sensitive influential twitterers // Proc. of the Third Int. Conf. on Web Search and Web Data Mining. 2010. P. 261—270.