Инженерный вестник Дона, №1 (2026)
ivdon.ru/ru/magazine/archive/n1y2026/10638

Максимизация влиятельности коалиции пользователей социальной сети на основе аппроксимированных метрик центральности и жадного алгоритма

Б.А. Торопов

Академия управления МВД России

Аннотация: Статья посвящена проблеме оценки влиятельности пользователей социальных сетей и их коалиций в ситуациях распространения информационного влияния на аудиторию. Формирование оптимального подмножества пользователей сети для инициации информационных влияний представляет собой вычислительно сложную задачу со стохастически неопределенным результатом. Существующие метрики центральности предполагают, как правило, поиск в графе всех кратчайших путей, либо решение масштабных систем уравнений. В настоящей работе предлагаются аппроксимированные оценки влиятельности отдельных участников сетей и их подмножеств, которые основаны на модификациях алгоритма Флажоле-Мартена. Также рассматривается жадный алгоритм, основанный на полученных метриках, и позволяющий формировать коалицию, итеративно дополняя ее квазиоптимальными элементами. Полученные результаты могут найти применение в задачах анализа, прогнозирования и планирования информационных влияний в социальных сетях.
Ключевые слова: социальная сеть, информационное влияние, коалиция, влиятельность, граф связей, метрика центральности, центральность по близости, кратчайший путь, маршрут, аппроксимированная оценка, жадный алгоритм

Введение

Интернет-сервисы в последние годы играют все более важную роль в общественных отношениях, позволяя людям удовлетворять свои не только информационные, но и социально-экономические и потребности. Среди различных Интернет-сервисов, в свою очередь, необходимо выделить социальные сети, которые получили достаточно широкое распространение и популярность. Изначально предназначенные для общения между людьми, данные сетевые структуры все чаще используются их участниками для получения актуальной новостной информации, коммерческих взаимодействий, самоорганизации в сообщества по интересам, последующей организации совместной активности в реальном мире. Точки зрения по различным вопросам, высказываемые пользователями этих Интернет-сервисов, способны воздействовать на сознание отдельных людей, а, следовательно, формировать актуальную информационную повестку и оказывать влияние на общественное мнение по дискутируемым вопросам.

Это свойство социальных сетей используется как маркетологами (см., например, [1]), так и политтехнологами (см., например, [2]) для продвижения определенных (желательных) точек зрения, в первом случае, относительно привлекательности тех или иных товаров и услуг, во втором случае, относительно идей об устройстве общества и государства. В названных сферах деятельности социальные сети играют любопытную роль, позволяя при относительно небольших затратах существенно увеличить охват аудитории, воспринявшей распространяемую информацию, и примкнувшей к желательной точке зрения. Существует специфический термин – посев информации, означающий мотивацию отдельных, как правило, немногочисленных пользователей к размещению распространяемой информации, которая затем предположительно будет распространяться стихийным способом, передаваясь сначала аудитории этих мотивированных пользователей, а затем далее по сетевой структуре по принципу «сарафанного радио».

Научный интерес представляют модели, методы и алгоритмы, позволяющие сформировать такую коалицию мотивированных пользователей, которая бы обеспечила наибольший охват аудитории.

Постановка задачи

Пусть моделью социальной сети выступает взвешенный ориентированный граф $G(V, g)$, где $V = \{v_1, v_2, \ldots, v_{|V|}\}$ – множество вершин, каждая из которых отображает одного участника сети; $g = \{g_{ij}\}^{|V|\times|V|}$ – матрица смежности графа, определяющая связи между ними, каждая дуга g_{ij} графа G обладает весом $g_{ij} \in [0, 1]$. Веса дуг могут определяться на основе анализа предшествующих взаимодействий между пользователями сети. Моделирование социальных взаимодействий является сложной научной задачей, как отмечал, например, Ю.А. Жданов [3], поэтому в настоящей работе будем исходить из того, что дуги графа взвешены на основе совокупности реакций в виде репостов, лайков и комментариев в адрес друг друга.

На графе G формируется подмножество вершин $S \subset G$, это те самые мотивированные пользователи, которые будут выступать инициаторами информационного влияния. В результате информационного влияния формируется подмножество $W \subset G$, пользователей, которые восприняли данное влияние и примкнули к продвигаемой точке зрения. Сам процесс влияния носит стохастический характер, ведь, гарантии того, что между пользователями сохранится прежняя интенсивность взаимодействий, не существует. Одновременно, как правило, неизвестны и индивидуальные характеристики отдельных пользователей, их априорные точки зрения относительно предмета информационного влияния, то есть заранее невозможно предсказать апостериорные точки зрения. Таким образом, мощность подмножества W – некая стохастическая функция от структуры графа G и состава подмножества S:

$$|W| = F\,(G,\, S)$$

Каким образом можно определить влиятельность S, полезность каждой вершины v_i, уже вошедшей в него и перспективность включения каждой вершины v_i, еще не вошедшей для включения в S?

Подходы к определению центральности вершин графа

Существует большое количество так называемых метрик центральности, определяющих значимость вершин графа в том или ином смысле на основе топологии графа и положении вершины в нем:

$$c_i = F(G, v_i)$$

Простейшая такая метрика – центральность по степени (degree centrality) [4, C. 51], отражающая количество инцидентных дуг, в рассматриваемом случае интерес будет представлять количество входящих дуг: $c_i^{(degree)} = \sum_{j=1}^{|V|} g_{ij}$. Для масштабных графов реальных социальных сетей такая метрика малоинформативна и, кроме того, манипулируема. Так, например, искусственное увеличение количества подписок, «накрутка» комментариев и лайков при помощи роботов, все эти мероприятия увеличивают именно центральность по степени.

Значительно более перспективными представляются метрики, опирающиеся на длины кратчайших путей, соединяющих каждую вершину графа с каждой прочей. К этой категории относятся центральность по близости [5], центральность по промежуточности [6], центральность распада [7]. Недостаток данных метрик проистекает из того обстоятельства, что, с одной стороны, информационные влияния в графе распространяются далеко не только по кратчайшим путям, но и по бесконечному количеству маршрутов произвольной длины, а, с другой стороны, поиск всех кратчайших путей в графе задача вычислительно сложная.

Другой класс метрик центральности опирается на предпосылку о том, что влиятельность каждого отдельного элемента сетевой структуры зависит от влиятельности соседних элементов. Центральность по собственному вектору (подробный обзор работ, посвященных которому содержится в [8]), центральность Каца-Бонасича [9], влиятельность по Дегруту [10], рассчитываемая на основе матрицы смежности. Данные метрики, скорее всего, наиболее информативны, поскольку в них учитываются все маршруты распространения влияния. Однако существенный недостаток таких оценок заключается в необходимости решения больших систем уравнений размерности равной составу графа $|V|$ для поиска собственного вектора или расчета центральности Каца-Бонасича, либо возведения в некоторую заранее неизвестную степень матрицы g. Таким образом, вычислительная сложность названных метрик для графов большой размерности чрезвычайно велика.

Аппроксимированные метрики центральности на основе алгоритма Флажоле-Мартена

Однако существуют, аппроксимированные методы оценки количества вершин графа, находящихся на расстоянии кратчайшего пути, не превышающем r, от исследуемой вершины v_i. К этой категории относится метод, опирающийся на использование алгоритма Флажоле-Мартена [11].

Для каждой вершины графа v_i случайным образом формируется бинарная кодовая последовательность $M(i,0)$ длины не менее $\lceil log_2|V|\rceil$. Каждый k-й ($k \in [0, \lceil log_2|V|\rceil]$) разряд которой принимает с вероятностью $p = 2^{-(k+1)}$ значение 1 и вероятностью $q = 1 - p$ значение 0.

На каждой итерации r выполнения алгоритма каждая кодовая строка $M(i,r)$ рассчитывается как дизъюнкция (логическое «ИЛИ» – $\vee$) между собственным значением на предыдущей итерации и значениями кодовых строк вершин – непосредственных соседей на предыдущей итерации.

$$M(i,r) = M(i,r-1) \vee_{\substack{j=1 \\ g_{ij} \neq 0}}^{|V|} M(j,r-1)$$

Затем рассчитывается оценка количества вершин находящихся на расстоянии r от v_i:

$$N(i,r) = \frac{2^b}{0,77359}$$

где: b – положение первого нулевого разряда в $M(i,r)$;

$0,77359$ – поправочный коэффициент Флажоле-Мартена.

На какой-то итерации r^* кодовые комбинации всех вершин графа полностью совпадают, на этом выполнение алгоритма закончено и можно оценить центральность каждой вершины v_i как:

$$c_i^{(FM)} = \sum_{r=1}^{r^*} \frac{1}{r}(N(i,r) - N(i,r-1))$$

Результаты расчета будут аппроксимацией центральности по близости, поскольку $N(i,r)$ дает оценку количества уникальных вершин, то есть разнообразие маршрутов, по которым распространяется влияние, не учитывается.

Частично компенсировать это обстоятельство и одновременно учесть веса дуг g_{ij} графа позволит незначительная модификация алгоритма:

,

$$M(i,r) = M(i,r-1) \vee_{j=1}^{|V|} \xi_{ijk}, \qquad \xi_{ijk} = \begin{cases} p = g_{ij}: & M_k(j,r-1), \\ q = 1-p: & 0. \end{cases}$$

где: индекс k используется для адресации разрядов $M(i,r)$, (теперь дизъюнкция выполняется поразрядно);

ξ_{ijk} – дискретно-вероятоностная передаточная функция, определяющая с вероятностью $p = g_{ij}$ то, произойдет ли операция дизъюнкции по отношению к k-му разряду $M(i,r)$.

Очевидно, что при такой модификации алгоритма возможна ситуация, в которой кодовые комбинации вершин не совпадут никогда $(r^*\to\infty)$. Поэтому предлагается выбрать фиксированное значение $r^*=2d$, где d – диаметр графа G, то есть длина кратчайшего пути между двумя наиболее удаленными друг от друга вершинами.

Результаты носят приближенный характер. Но, тем не менее, сопоставляя их с иными метриками центральности можно прийти к выводу, что это достаточно точная аппроксимация центральности по близости, что было показано в ряде работ (см, например, [12]), в том числе, в одной из предыдущих работ автора [13].

Развивая рассмотренную идею, следует обратить внимание на концепцию метрик групповой центральности и применить для расчета аппроксимированной оценки центральности по близости группы вершин S кодовые комбинации Флажоле-Мартена. Кодовые комбинации коалиции образуются путем дизъюнкции *M(i,r)* всех вершин $v_i \in S$:

$$M(S,r) = \vee_{\substack{i=1 \\ v_i \in S}}^{|V|} M(i,r)$$

На основе алгоритмически рассчитываемой метрики центральности каждой отдельной вершины и коалиции произвольного их количества далее можно предложить жадный алгоритм, позволяющий сформировать на графе *G* коалицию вершин $S \subset V$. На каждой итерации алгоритма в *S* включается такая v_i, что:

$$v_i = \arg\max_i (c_{S \cup \{v_i\}}^{(FM)} - c_S^{(FM)})$$

Блок-схема представлена на рисунке 1. В левой части блок-схемы происходит расчет кодовых комбинаций *M(i,r)* для всех вершин $v_i \in V$ графа *G*. В правой части реализуется жадный поиск. В изначально пустую коалицию *S = {}*, поочередно добавляется та вершина, которая на каждой длине кратчайшего пути от *r* до *2d* привносит наибольшую полезность в коалицию, то есть сильнее всего смещает положение первого нулевого разряда в кодовой комбинации коалиции *M(S,r)*.

Выводы

Предложенные в работе аппроксимированные метрики центральности позволяют приближенно оценивать центральность, как отдельных вершин графа, так и их подмножеств. Рассмотренный жадный алгоритм, безусловно, не всегда позволяет найти оптимальное сочетание пользователей. Однако сама форма используемых кодовых комбинаций предполагает, что находя на каждой итерации вершину графа, сильнее всего улучшающую кодовую комбинацию формируемого подмножества, получается добавить в это подмножество квазиоптимальный элемент.

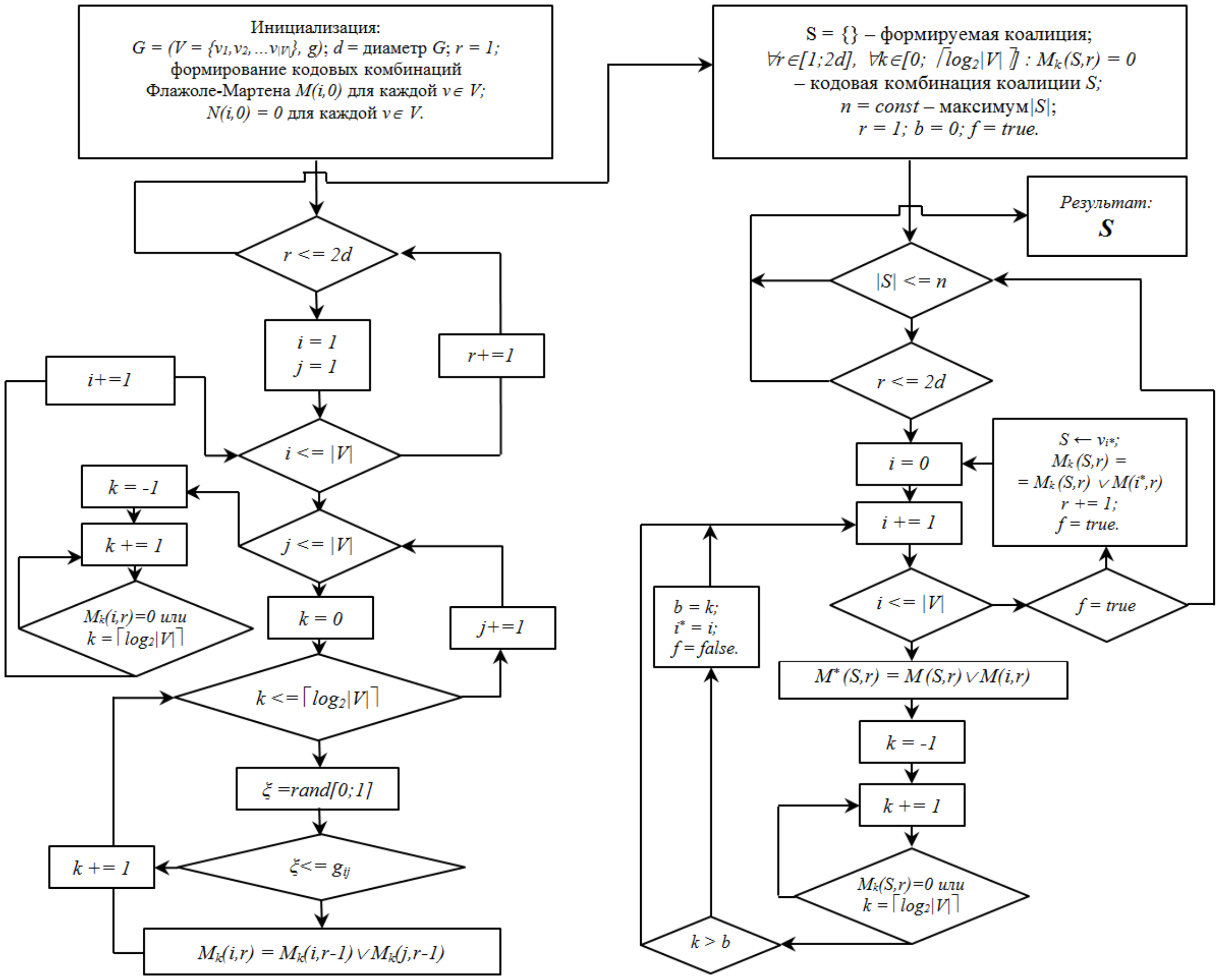

Рис. 1. – Блок-схема жадного алгоритма формирования коалиции вершин $S \subset V$ с максимальной групповой центральностью по близости

Представленные результаты позволяют ставить и решать оптимизационные задачи, связанные с формированием коалиций участников социальных сетей для инициации информационных влияний, а также принимать решения в ситуациях информационного противоборства.

Литература

1. Подопригора М.Г., Маслова А.А. «Flocktory» как инструмент реферального маркетинга в России // Инженерный вестник Дона. 2014. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2014/2421.

2. Овчинский А.С., Борзунов К.К. Информационные взаимодействия и противоборство в политической деятельности // Информация и безопасность. 2024. Т. 27, № 4. С. 509-520.

3. Розин М.Д., Свечкарев В.П. Научное наследие Ю.А. Жданова и современные проблемы моделирования сложных социосистем // Инженерный вестник Дона. 2014. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2014/2437.

4. Jackson M.O. Social and Economic Networks. Princeton University Press. 2008. 520 p.

5. Freeman L.C. Centrality in social networks: Conceptual clarification. Social Networks. 1978. V 1. pp. 215-239.

6. Brandes, U. A faster algorithm for betweenness centrality. The Journal of Mathematical Sociology. 2001. V. 25(2). pp. 163–177.

7. Jackson M.O., Wolinsky A. A Strategic Model of Social and Economic Networks. Journal of Economic Theory. 1996. V. 71. pp. 44-74.

8. Iacobucci D., Mcbride R., Popovich D. Eigenvector Centrality: Illustrations Supporting the Utility of Extracting More Than One Eigenvector to Obtain Additional Insights into Networks and Interdependent Structures. Journal of Social Structure. 2017. V. 18(1). pp. 1-21.

9. Bonacich P. Eigenvector-like measures of centrality for asymmetric relations. Social Networks. 2001. V. 23. pp. 191-201.

10. Degroot, M.H. Reaching a Consensus. Journal of the American Statistical Association. – 1974. V. 69. pp. 118-121.

11. Flajolet P., Martin G.N. Probabilistic counting algorithms for data base applications. Journal of Computer and System Sciences. 1985. V. 31(2). pp. 182–209.

12. Zhao J., John C.S. Lui, Towsley D., Guan X.H. Measuring and Maximizing Group Closeness Centrality over Disk-Resident Graphs. Proceedings of the 23rd International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland. 2014. pp. 689–694.

13. Торопов Б.А. Центральность распада в социальных графах и адаптированный алгоритм Флажоле-Мартена для ее расчета // International Journal of Open Information Technologies. 2017. Т. 5, № 9. С. 27-33.

References

1. Podoprigora M.G., Maslova A.A. Inzhenernyj vestnik Dona. 2014. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2014/2421.

2. Ovchinskij A.S., Borzunov K.K. Informaciya i bezopasnost'. 2024. V. 27(4). pp. 509-520.

3. Rozin M.D., Svechkarev V.P. Nauchnoe nasledie Yu.A. Inzhenernyj vestnik Dona. 2014. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2014/2437.

4. Jackson M.O. Social and Economic Networks. Princeton University Press. 2008. 520 p.

5. Freeman L.C. Centrality in social networks: Conceptual clarification. Social Networks. 1978. V 1. pp. 215-239.

6. Brandes U. The Journal of Mathematical Sociology. 2001. V. 25(2). pp. 163–177.

7. Jackson M.O., Wolinsky A.A Journal of Economic Theory. 1996. V. 71. pp. 44-74.

8. Iacobucci D., Mcbride R., Popovich D. Journal of Social Structure. 2017. V. 18(1). pp. 1-21.

9. Bonacich P. Social Networks. 2001. V. 23. pp. 191-201.

10. Degroot, M.H. Journal of the American Statistical Association. 1974. V. 69. pp. 118-121.

11. Flajolet P., Martin G.N. Probabilistic counting algorithms for data base applications. Journal of Computer and System Sciences. 1985. V. 31(2). pp. 182–209.

12. Zhao J., John C.S. Lui, Towsley D., Guan X.H. Measuring and Maximizing Group Closeness Centrality over Disk-Resident Graphs. Proceedings of the 23rd International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland. 2014. pp. 689–694.

13. Toropov B.A. International Journal of Open Information Technologies. 2017. V. 9(5). pp. 27-33.

Авторы согласны на обработку и хранение персональных данных.

Дата поступления: 11.11.2025

Дата публикации: 25.12.2025