

## Прогнозирование редких событий на основе анализа графлетов взаимодействия в социальных сетях

*К.А. Хализев, А.В. Осин*

*Московский технический университет связи и информатики*

**Аннотация:** Развитие информационных технологий и активное использование социальных сетей делают возможным прогнозирование редких событий на основе анализа взаимодействий пользователей. В данной работе рассматривается метод прогнозирования редких событий с использованием теории графов, в частности графлетов. Социальная сеть «ВКонтакте» используется как источник данных, а алгоритм ORCA применяется для выявления характерных графовых структур. Проведен анализ взаимодействий пользователей, отобраны предвестники редких событий и оценена точность прогноза. Полученные результаты показывают перспективность предложенного метода, его применимость для мониторинга угроз и возможности дальнейшего совершенствования моделей предсказания на основе графов.

**Ключевые слова:** социальные сети, событие безопасности, прогнозирование событий, теория графов, графлет, анализ взаимодействий, анализ временных рядов, корреляционный анализ, обработка данных, аномальная активность.

### Введение

Информационные технологии всё глубже проникают в жизнь человека, в результате чего социальные сети становятся полноценным отражением разнообразных сторон современного мира. Редкие события, такие как неожиданные изменения в политике и экономике или инциденты, связанные с безопасностью, находят своё отражение в социальных сетях. Их анализ позволяет выявить своеобразные сигналы-предвестники – факты, признаки или индикаторы, свидетельствующие о том, что какое-то событие имеет высокую вероятность наступления. Цель прогнозирования редких событий – не только своевременно предупреждать об опасностях, но и разрабатывать эффективные стратегии и меры по снижению потенциальных рисков.

Существуют различные подходы к анализу взаимодействий в социальных сетях, в том числе на основе теории графов. Так, например, можно прибегнуть к использованию метаграфов [1]. Одним из новых

подходов является применение графлетов для анализа взаимодействий, в частности метода прогнозирования редких событий на основе графлетов.

Социальная сеть «ВКонтакте» — одна из самых популярных в России, поэтому в ней отражается множество событий, происходящих в стране, что делает получаемые из неё данные особенно релевантными и пригодными для анализа.

Таким образом, если получится выявить взаимосвязь между сигналами-предвестниками, получаемыми этим методом и самим событием, то тогда применение метода прогнозирования редких событий на основе графлетов для данных о взаимодействиях в социальной сети «ВКонтакте» позволит предсказывать эти события.

### **Связь социальных сетей с редкими событиями**

Одним из методов прогнозирования редких событий является определение сигналов-предвестников, в частности, слабых сигналов — понятия, широко рассмотренного в работах И. Ансоффа [2], М. Годе [3], Б. Коффмана [4], Е. Хилтунен [5] и К. Вельца [6]. На основании этих работ можно сформулировать следующее определение слабого сигнала: это неточная и неполная, часто остающаяся без внимания информация, которая указывает на значительные изменения в будущем. К ним могут относиться необычная сетевая активность, аномальное поведение пользователей или изменения в конфигурации системы. Выявление и анализ слабых сигналов помогают обнаруживать потенциальные угрозы до того, как они перерастут в серьёзные инциденты.

Обнаружение сигналов-предвестников в социальных сетях может быть сложной задачей из-за большого объема данных, которые необходимо обработать, а также их неоднородности. Для выявления таких сигналов применяются различные подходы: анализ социальных сетей (SNA) [7], анализ текста и обработка естественного языка [9, 10], а также методы

---

машинного обучения [11]. Некоторые исследователи также пытаются выявить специфические паттерны в сетях, которые можно рассматривать как предвестники событий. Например, был представлен [12] метод изучения корреляций в графах предстоящих землетрясений с использованием алгоритмов теории графов. Результаты этого исследования позволили выдвинуть гипотезу о том, что применение теории графов, в частности графлетов, может быть эффективным для обнаружения сигналов, предвещающих различные события.

Графлеты представляют собой малые связанные подграфы фиксированного размера от 2 до 5 узлов, которые отображают локальные структуры взаимодействий в сети (рис. 1).

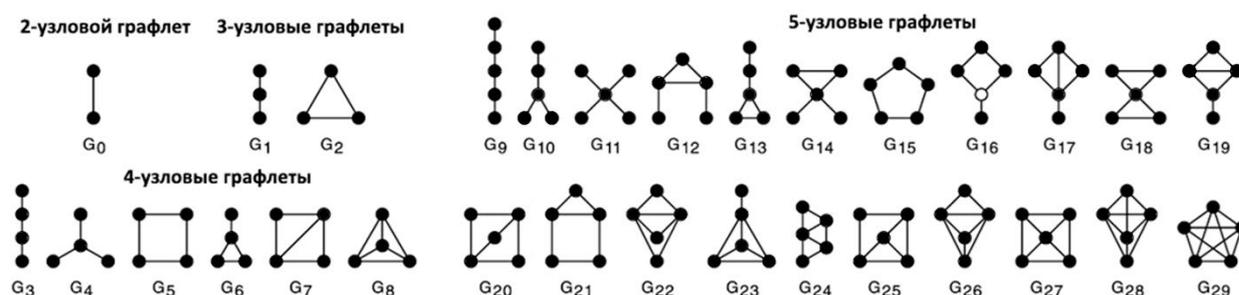


Рис. 1. – Виды графлетов

Социальная сеть «ВКонтакте» предоставляет инструмент VK API, который позволяет извлекать большой объём данных для анализа взаимодействий пользователей и администраторов сообществ. Для изучения графовых структур наиболее ценным представляется анализ комментариев под публикациями пользователей и сообществ, так как они позволяют фиксировать изменения во взаимодействиях с течением времени с помощью инструмента VK API.

### Общая схема метода

Для прогнозирования редких событий на основе графлетов был разработан метод, включающий несколько последовательных этапов анализа данных (рис. 2):

- Сбор и предварительная обработка данных из социальной сети «ВКонтакте» для дальнейшего анализа;
- Выделение предвестников событий с использованием метода на основе графлетов на основе алгоритма обнаружения предвестников, предложенного Х. Абу Джамра, М. Савонне и Е. Леклерк [13], примененного к данным из «ВКонтакте»;
- Оценка полученных результатов и их интерпретация.

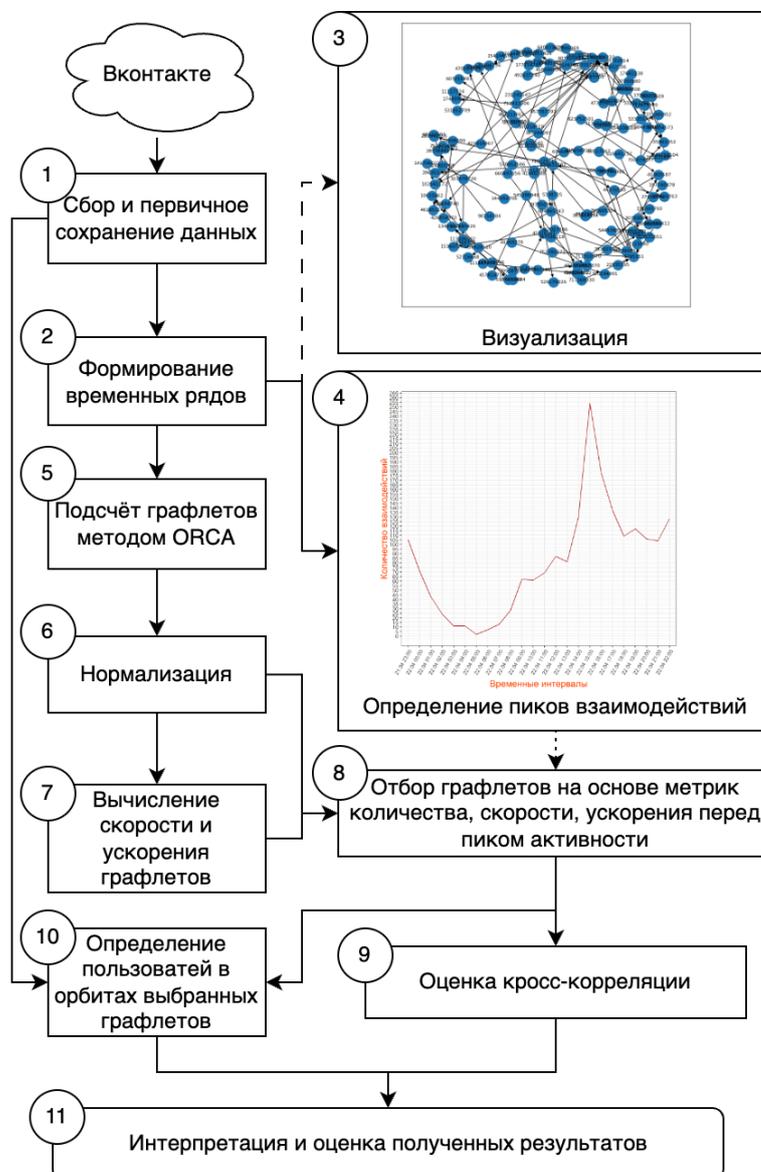


Рис. 2. – Структурная схема метода

Алгоритмы, описанные в данной работе, были программно реализованы на языках R и Python, что позволило создать трёхмодульную систему.

### **Описание данных**

Собранный набор данных представляет собой временную последовательность взаимодействий между пользователями «ВКонтакте» в комментариях к публикациям в сообществах (группах) и в профилях других пользователей социальной сети. Данные охватывают комментарии в период с 3:00 21 апреля 2023 по 19:00 22 апреля 2023 в специально отобранных сообществах и аккаунтах, связанных с Белгородской областью.

В результате сбора комментариев была получена последовательность взаимодействий между пользователями за указанный временной промежуток. Всего набор данных содержит 8376 комментариев, каждый из которых отображает уникальное взаимодействие между пользователями в группе «ВКонтакте». Данные в результате сбора и предобработки позволяют сформировать граф, представляющий собой взаимодействия между пользователями или администраторами сообществ «ВКонтакте».

### **Отбор графлетов и анализ их динамики**

Отбор и оценка графлетов произведены следующим образом:

1. На основе анализа активности пользователей на собранных данных было выбрано целевое событие для прогнозирования: максимальные значения активности пользователей зафиксированы 22 апреля 2023 14:00 – 15:00, что соответствует событию – объявлению эвакуации.
2. Подсчёт графлетов при помощи алгоритма ORCA на временных интервалах с разницей в 1 час. Получена таблица, где столбцы представляют различные типы графлетов, строки соответствуют

временным интервалам, значения в таблице показывают количество графлетов определённого типа в заданный интервал (таблица 1).

3. Производится подсчёт метрик нормированного количества, скорости и ускорения графлетов за 3 часа до начала события (с 8:00 до 11:00). Среди графлетов отбираются те, в которых наблюдаются наибольшие изменения метрик по сравнению с началом события (таблица 2).

Таблица № 1

#### Фрагмент результата подсчёта графлетов

(результаты для интервалов 22.04.2023 15:00 – 22.04.2023 20:00)

	G0	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
22.04.2023 17:00	137	368	6	276	1009	0	65	2	1	144	1307	2332
22.04.2023 18:00	109	352	0	204	1572	0	0	0	0	215	1560	6389
22.04.2023 19:00	117	413	0	150	2466	0	0	0	0	64	1037	12851
22.04.2023 20:00	106	350	1	151	1473	0	1	0	0	58	1081	4986
22.04.2023 21:00	104	170	2	34	309	0	4	0	0	4	58	472
22.04.2023 22:00	128	681	2	240	4169	0	23	0	0	163	1772	21297

Было выявлено, что в четырех типах графлетов - G1, G4, G5, G11, наблюдаются наибольшие колебания показателей метрик количества, скорости и ускорения в указанный период времени, следовательно, именно они рассматриваются как потенциальные предвестники.

4. Проведена оценка кросс-корреляции между временными рядами графлетов и временными рядами общего количества взаимодействий пользователей, что позволяет определить степень их взаимосвязи во времени.

Анализ кросс-корреляции (рис. 3) показал, что графлеты G1, G4, G5 и G11 достигают максимальной корреляции (~0.6) при смещении в 9 часов. Это указывает на сильную связь между их динамикой и общим количеством взаимодействий задолго до события, что позволяет рассматривать изменения

их метрик как предвестники. Совокупность отобранных графлетов (G1, G4, G5, G11) и времени смещения (9 часов) формирует шаблон для прогнозирования редких событий.

Таблица № 2

Результаты для некоторых графлетов на интервалах до наступления события

Графлет	08:00 – 09:00 22.04			09:00 – 10:00 22.04			10:00 – 11:00 22.04			12:00 – 13:00 22.04		
	$\overline{G_X^E}$	$\overline{V_X^E}$	$\overline{A_X^E}$									
G1 	-0,275	0,411	-0,047	-0,047	-0,047	-0,603	0,194	0,661	0,852	-0,047	0,948	0,948
G4 	-0,347	0,137	-0,129	-0,129	-0,129	-0,216	0,208	0,635	0,715	-0,129	0,940	0,940
G5 	0,117	0,401	-0,284	-0,284	-0,284	-0,802	-0,284	0,000	0,401	-0,284	0,000	0,000
G11 	-0,329	0,036	-0,188	-0,188	-0,188	-0,060	0,112	0,464	0,488	-0,188	0,637	0,637
G13 	-0,224	0,052	-0,276	-0,276	-0,276	-0,105	-0,276	0,000	0,052	-0,276	-0,244	-0,244
G18 	-0,234	0,000	-0,234	-0,234	-0,234	0,000	-0,234	0,000	0,000	-0,234	0,000	0,000

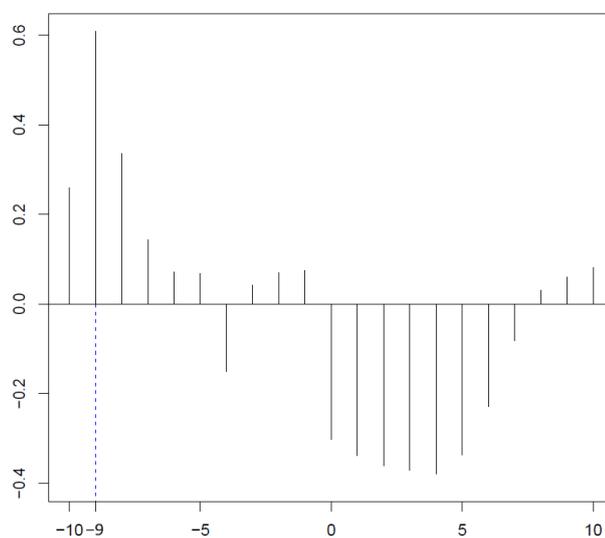


Рис. 3. – График кросс-корреляции для графлета G11

### Оценка и интерпретация результатов

В результате анализа для ранее отобранных графлетов G1, G4, G5 и G11 были определены их вершины — пользователи, находящиеся в составных частях графлетов, называемых орбитами (группами симметрии).

Для каждого графлета выделены соответствующие орбиты: O1 и O2 для G1, O6 и O7 для G4, O8 для G5, O22 и O23 для G11. Было подсчитано количество вхождений пользователей в соответствующие орбиты графлетов, что позволило получить результаты, представленные в таблице 3.

Таблица № 3

Пользователи орбит графлетов G4, G11

Интервал	Пользователь	O6	O7	O22	O23
22.04.2023 8:00	283969460	1	0	0	1
22.04.2023 8:00	630250267	1	0	1	0
22.04.2023 8:00	639631882	2	0	0	1
22.04.2023 8:00	283969460	1	0	0	1
22.04.2023 8:00	186311942	0	1	0	0
22.04.2023 8:00	427941838	0	1	0	0
22.04.2023 8:00	-35961352	1	0	0	0
22.04.2023 8:00	245962695	1	0	0	0
22.04.2023 8:00	64355195	1	0	0	0

Идентифицируя пользователей по их уникальному идентификатору (ID), выделяются следующие страницы «ВКонтакте»: официальный профиль представителя Белгородской области (639631882) и городского сообщества «Белгород – это интересно» (-35961352). При этом они не выступают инициаторами взаимодействий, а становятся объектом внимания – под их публикациями массово оставляют комментарии другие пользователи.

Для оценки точности прогнозирования было выбрано ещё четыре случайных события, для которых были собраны свои наборы данных о взаимодействиях пользователей соцсети. Для каждого из них был сформирован набор графлетов и подсчитано время смещения, в результате чего получилось пять шаблонов для прогнозирования событий. Качество сформированных шаблонов было проверено на отдельном наборе данных за период с 25 апреля по 25 мая 2023 года, имеющем множество событий,

вызывавших всплеск пользовательской активности. Для каждого шаблона были зафиксированы моменты, когда наблюдалось изменение метрик соответствующих графлетов. Если изменение метрик при заданном в шаблоне смещении указывало на появление события, что проявлялось в повышенной активности пользователей, то такой прогноз считался верным. В результате для каждого шаблона было выявлено от 12 до 41 случая изменения метрик графлетов, каждый случай был сопоставлен с фактом наличия события. Итоговые результаты и составляющие шаблонов приведены в таблице 4.

Считая допустимой погрешностью  $\pm 2$  часа, прогноз совпадает с реальным временем повышенной пользовательской активности в 94% случаев. В 6% случаев прогноз оказывается значительно удалённым от реальных событий.

Таблица № 4

#### Оценка прогноза

Шаблон		Событие присутствует	Событие отсутствует
Набор графлетов	Смещение прогноза (ч)		
G1, G4, G5, G11	9	34	1
G2, G4, G11	6	10	2
G1, G5, G23	7	26	1
G1, G3, G10	8	29	3
G1, G2, G5, G11	9	39	2
Итого		94%	6%

#### Выводы

Оценка результатов работы метода, основанная на применении шаблонов к данным взаимодействий, позволяет судить о наличии перспектив развития прогнозирования на основе графлетов. В ходе оценки результатов

работы разработанного ПО была получена точность работы – 94% (с ошибкой в предсказании до двух часов). Совершенствование полученной системы, в частности улучшение механизмов сбора данных и формирования картины пользовательских взаимодействий позволит получить более качественные показатели работы метода, рассмотренного в данной работе. Способность выявлять и предсказывать редкие события даёт возможность заранее подготовиться к ним, что важно, учитывая влияние безопасности на экономическую устойчивость предприятий и организаций [14].

### Литература

1. Астанин С.В., Драгныш Н.В., Жуковская Н.К. Вложенные метаграфы как модели сложных объектов // Инженерный вестник Дона, 2012, №4(2). URL: [ivdon.ru/ru/magazine/archive/n4p2y2012/1434](http://ivdon.ru/ru/magazine/archive/n4p2y2012/1434).
2. Ansoff I. Managing strategic surprise by response to weak signals. *California Management Review*, 1975, Vol. 18, No. 2, pp. 21–33.
3. Godet M. *From anticipation to action: a handbook of strategic prospective*. UNESCO Publishing. 1994. 283 p.
4. Coffman B. Weak signal research, part I: Introduction. *Journal of Transition Management*. 1997. Vol. 2, No. 1. pp. 21–33.
5. Hiltunen E. *Weak Signals in Organisational Futures*. Aalto University School of Economics, Aalto, 2010. 265 p.
6. Welz K., Brecht L., Pengl A., Kaueldt J.V., Schallmo D.R.A. Weak signals detection: Criteria for social media monitoring tools. *ISPIM Innovation Symposium. The International Society for Professional Innovation Management (ISPIM)*, 2012, pp. 1.
7. Charitonidis C., Rashid A., Taylor P.J. Weak signals as predictors of real-world phenomena in social media. 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). Paris, France, 2015. pp. 864–871

8. Yoon J. Detecting weak signals for long-term business opportunities using text mining of web news. *Expert Systems with Applications*. 2012. Vol. 39, No. 16, pp. 12543–12550.

9. Rashid A., Baron A., Rayson P., May-Chahal C., Greenwood P., Walkerdine J. Who am I? analyzing digital personas in cybercrime investigations. *Computer*. 2013. Vol. 46, No. 4, pp. 54–61.

10. Taylor P., Dando C., Ormerod T., Ball L., Jenkins M., Sandham A., Menacere T. Detecting insider threats to organizations through language change. *Law and Human Behavior*. 2013. Vol. 37, No. 4, pp. 267–275.

11. Georgiadis D., Raubal M. An interdisciplinary review on weak signal detection. Singapore-ETH Centre, 2020. 35 p.

12. Baiesi M. Scaling and precursor motifs in earthquake networks. *Physica A: Statistical Mechanics and its Applications*. 2006. Vol. 360, No. 2. pp. 534–542.

13. Abou Jamra H., Savonnet M., Leclercq E. Detection of Event Precursors in Social Networks: A Graphlet-Based Method. *Research Challenges in Information Science*. Springer International Publishing, 2021, pp. 205–220.

14. Ибрагимова З.М., Батчаева З.Б., Ткаченко А.Л. Информационная безопасность как элемент экономической безопасности // *Инженерный вестник Дона*. 2022. №11. URL: [ivdon.ru/ru/magazine/archive/n11y2022/8010](http://ivdon.ru/ru/magazine/archive/n11y2022/8010).

### References

1. Astanin S.V., Dragnysh N.V., Zhukovskaya N.K. *Inzhenernyj vestnik Dona*. 2012. №4(2). URL: [ivdon.ru/ru/magazine/archive/n4p2y2012/1434](http://ivdon.ru/ru/magazine/archive/n4p2y2012/1434).

2. Ansoff I. Managing strategic surprise by response to weak signals. *California Management Review*, 1975, Vol. 18, No. 2, pp. 21–33.

3. Godet M. From anticipation to action: a handbook of strategic prospective. UNESCO Publishing. 1994. 283 p.



4. Coffman B. Weak signal research, part I: Introduction. *Journal of Transition Management*. 1997. Vol. 2, No. 1. pp. 21–33.
5. Hiltunen E. *Weak Signals in Organisational Futures*. Aalto University School of Economics, Aalto, 2010. 265 p.
6. Welz K., Brecht L., Pengl A., Kaueldt J.V., Schallmo D.R.A. Weak signals detection: Criteria for social media monitoring tools. *ISPIM Innovation Symposium. The International Society for Professional Innovation Management (ISPIM)*, 2012, pp. 1.
7. Charitonidis C., Rashid A., Taylor P.J. Weak signals as predictors of real-world phenomena in social media. *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. Paris, France, 2015. pp. 864–871
8. Yoon J. Detecting weak signals for long-term business opportunities using text mining of web news. *Expert Systems with Applications*. 2012. Vol. 39, No. 16, pp. 12543–12550.
9. Rashid A., Baron A., Rayson P., May-Chahal C., Greenwood P., Walkerdine J. Who am I? analyzing digital personas in cybercrime investigations. *Computer*. 2013. Vol. 46, No. 4, pp. 54–61.
10. Taylor P., Dando C., Ormerod T., Ball L., Jenkins M., Sandham A., Menacere T. Detecting insider threats to organizations through language change. *Law and Human Behavior*. 2013. Vol. 37, No. 4, pp. 267–275.
11. Georgiadis D., Raubal M. *An interdisciplinary review on weak signal detection*. Singapore-ETH Centre, 2020. 35 p.
12. Baiesi M. Scaling and precursor motifs in earthquake networks. *Physica A: Statistical Mechanics and its Applications*. 2006. Vol. 360, No. 2. pp. 534–542.



13. Abou Jamra H., Savonnet M., Leclercq E. Detection of Event Precursors in Social Networks: A Graphlet-Based Method. Research Challenges in Information Science. Springer International Publishing, 2021, pp. 205–220.

14. Ibragimova Z.M., Batchaeva Z.B., Tkachenko A.L. Inzhenernyj vestnik Dona. 2022. №11. URL: [ivdon.ru/ru/magazine/archive/n11y2022/8010](http://ivdon.ru/ru/magazine/archive/n11y2022/8010).

**Дата поступления: 16.02.2025**

**Дата публикации: 26.03.2025**