

# Модели социального влияния: психология и математика

Козицин Иван Владимирович  
кандидат физико-математических наук  
старший научный сотрудник, Институт проблем управления РАН

Москва, 2023

# Общая постановка



Агенты (люди),  
их характеристики

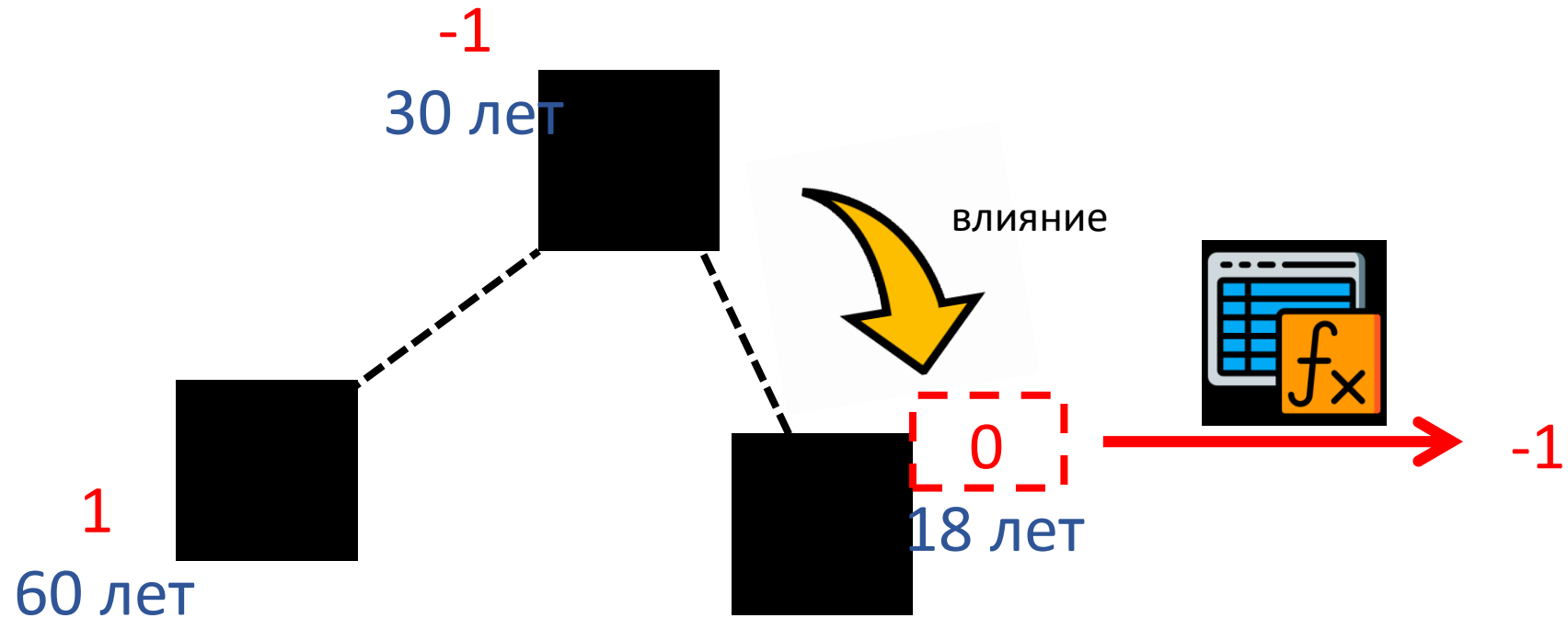
(**мнения**, другие параметры)



Структура взаимодействий  
(кто с кем общается,  
интенсивность)



Протокол взаимодействий  
(результат общения)



# О мнениях...

“Мнение – это то, что изменяется в результате социального влияния”

R. Axelrod (1997)

(Carpentras et al., 2022)	Дискретные	Непрерывные
Эмпирические (имеющие правильный ответ)	“Земля круглая?” Да / Нет	“Какова высота г. Эверест в метрах?” Положительное число
Нормативные (не имеющие правильного ответа)	“За какую партию Вы отдадите голос на выборах?” А / В / С	“Как Вы относитесь к ношению оружия?” Число от -1 до 1 (“-1” – очень плохо, “1” – очень хорошо)

# Задачи

- К какому распределению мнений сойдется система?
- Скорость сходимости мнений?
- Влияние структуры взаимодействий на динамику мнений?
- Влиятельность отдельных агентов?
- Управление (маркетинг, полит технологии, информационные войны)

# Что мы уже умеем?



Измерять мнения  
(opinion mining)



Измерять структуру  
(взаимодействий)

Отношение к вакцинации: положительное  
(0.89 из промежутка [-1, 1], “-1” – очень плохое, “1” -  
очень хорошее)

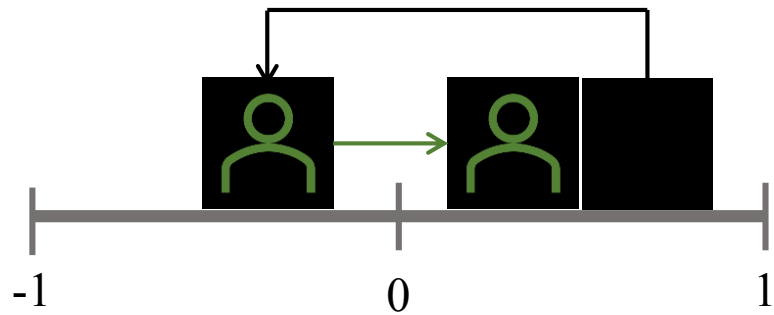


Строить прогноз,  
находить управление

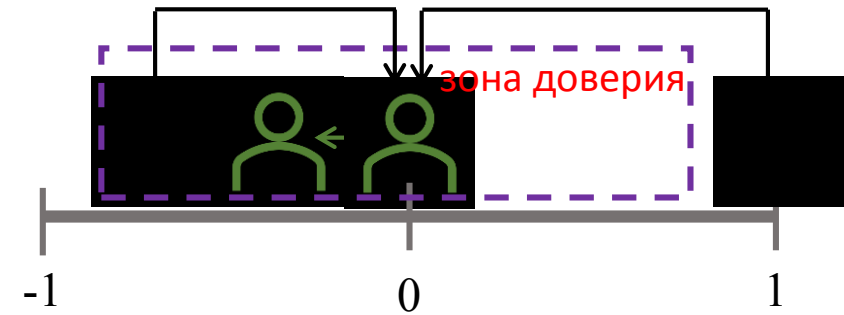


# Обзор микро-механизмов социального влияния (непрерывные одномерные мнения)

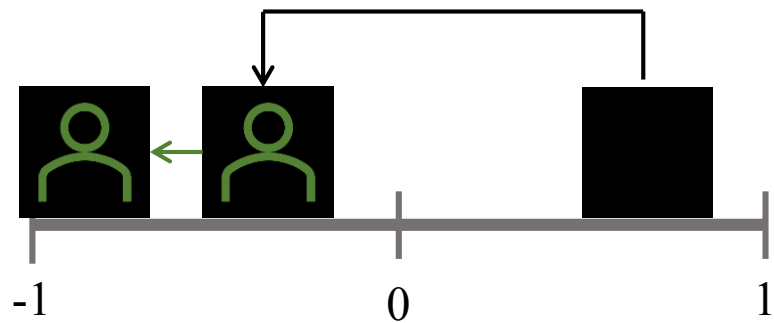
Ассимилятивное влияние



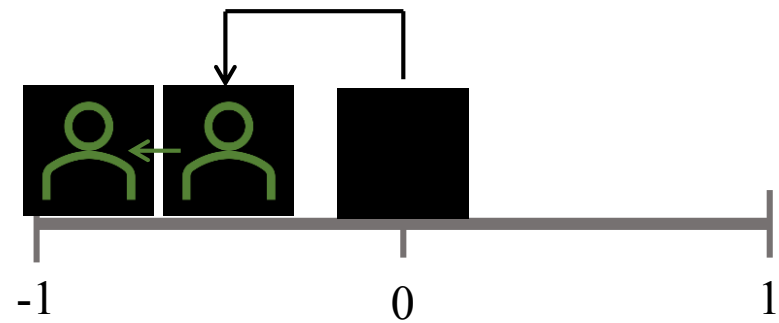
Ограниченное доверие (к информации)



Диссимилятивное влияние



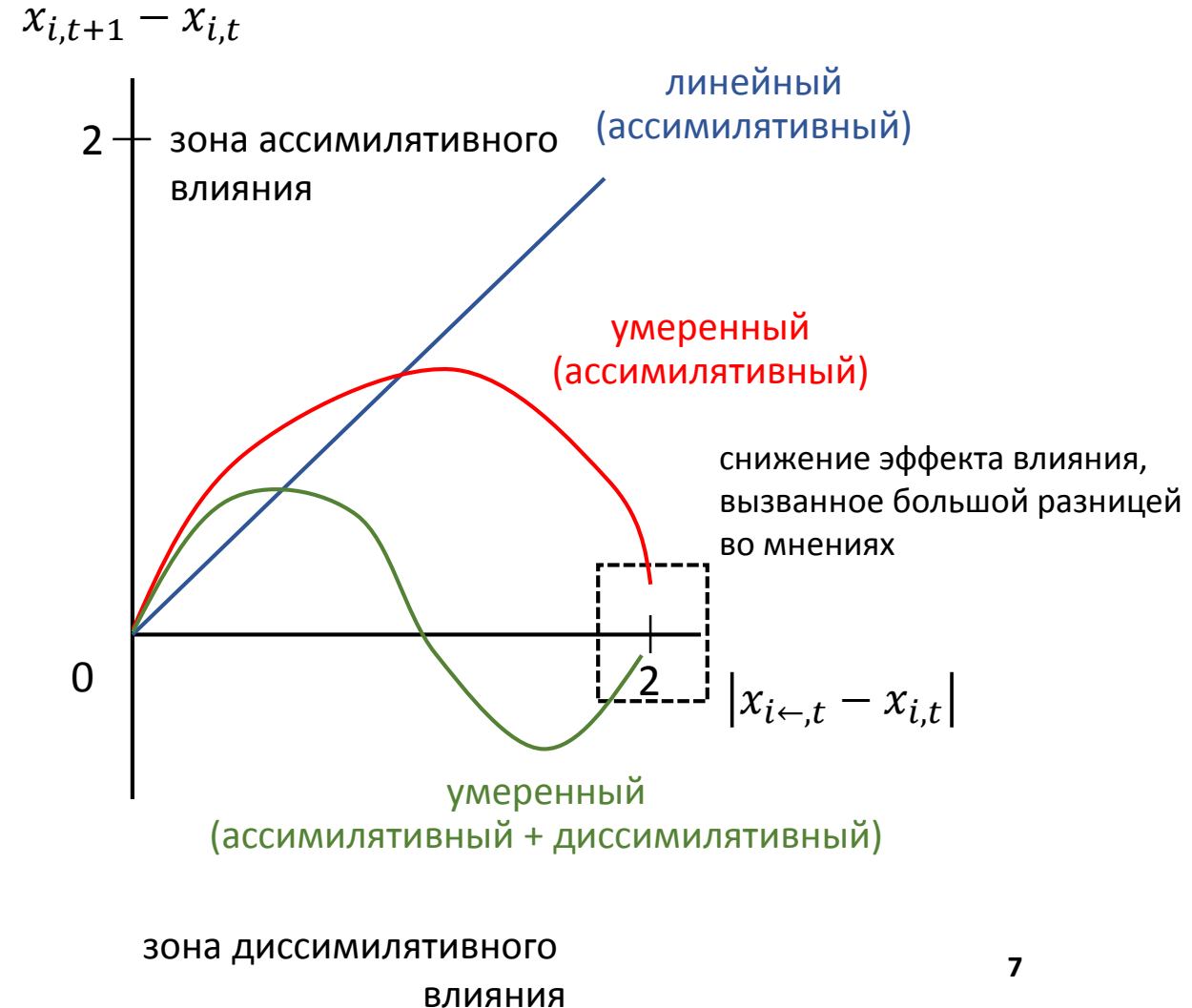
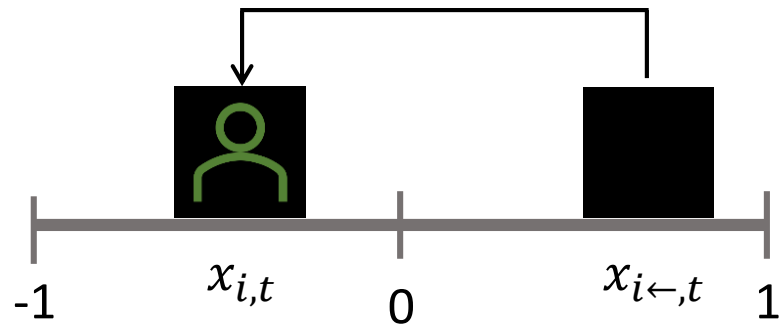
Искаженное восприятие (biased assimilation)



Нейтральные/"колеблющиеся" аргументы  
могут трактоваться искаженным образом

# Математическая формализация механизмов (influence-response functions)

(Takács et al., 2016)



Механизм	Описание
Линейный (ассимилятивный)	Изменение мнения пропорционально абсолютной разнице во мнениях между объектом и источником влияния
Умеренный (ассимилятивный)	Слишком большая разница во мнениях не вызывает отклика – ограниченное доверие (умеренная форма).
Умеренный (ассимилятивный) + диссимилятивный)	С ростом разницы во мнениях ассимилятивное влияние уступает место диссимилятивному, которое, в свою очередь, сходит на нет при дальнейшем росте уровня когнитивного диссонанса

# Классические модели и “прорывные” работы

## Модель ДеГроота

$$x(t + 1) = Wx(t)$$

$$x = \begin{bmatrix} x_1 \\ \dots \\ x_N \end{bmatrix}$$

– вектор мнений

$$W = \begin{bmatrix} w_{11} & \dots & w_{1N} \\ \dots & \dots & \dots \\ w_{N1} & \dots & w_{NN} \end{bmatrix}$$

– строчно-стохастическая матрица (влияний)

Механизмы:



- ассимилятивный



- ограниченное доверие



- диссимилятивный

## Козволюция мнений и социального графа

Мнения:



- дискретные



- непрерывные



- на основе аргументов

## “Проблема” разнообразия мнений

Социальная  
власть

(Rashevsky, 1939)

3

(French Jr, 1956)

7

(Abelson, 1964)

13

(Schelling, 1969)

15

(Granovetter, 1978)

(Clifford & Sudbury, 1973)

(DeGroot, 1974)

(Galam, 1986)

(Friedkin & Johnsen, 1990)

44

(Axelrod, 1997)

184

(Macy et al., 2003)

(Hegselmann & Krause, 2002)

(Holme & Newman, 2006)

554

(Deffuant et al., 2000)

(Martins, 2008)

1550

(Mäs & Flache, 2013)

(Altafini, 2013)

(Parsegov et al., 2017)

3420

Теория аргументных взаимодействий

Логически связанные мнения

Непрерывные мнения / дискретные действия

Фейковые новости

Интернет, социальные сети

Алгоритмы ранжирования

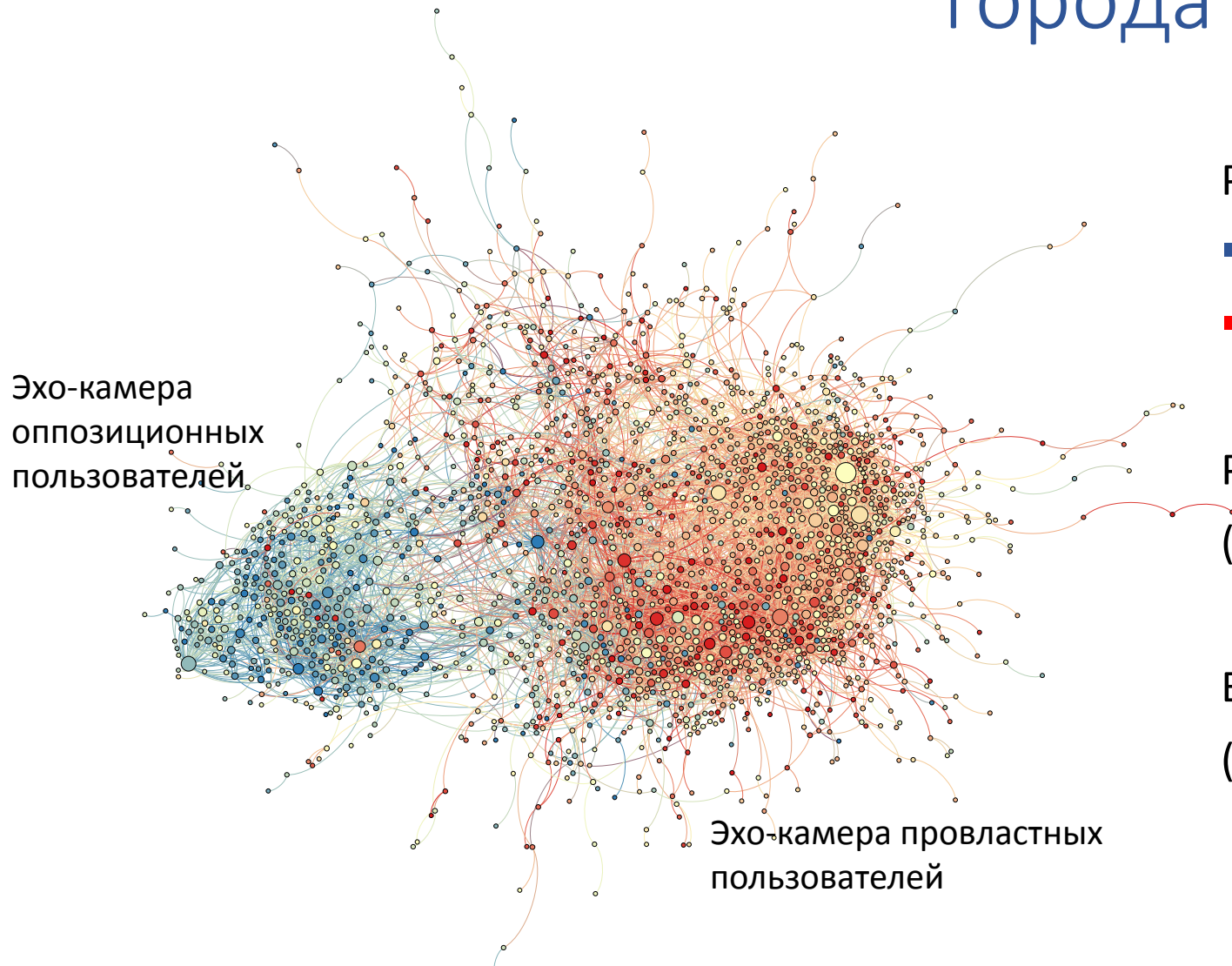
Поляризация, эхо-камеры

6510





# Фрагмент социального графа ВКонтакте – активные аккаунты жителей Подмосковного города



Раскраска вершин соответствует их мнениям:

- Синие вершины – оппозиция
- Красные – провластные взгляды

Размер вершины соответствует ее степени (числу друзей)

Вершины группируются в два кластера (эхо-камеры) с противоположными взглядами

# Некоторые исследования в ИПУ РАН

(DeGroot, 1974)

(Schelling, 1969)

(Granovetter, 1978)

(Friedkin & Johnsen, 1990)

Теория  
линейных многоагентных систем

Чеботарев, П. Ю., & Агаев, Р. П.  
(2009)

Управление

Губанов, Д. А., Новиков, Д. А., &  
Чхартишвили, А. Г. (2010)

Управление

Бреер, В. В., Новиков, Д. А.,  
& Рогаткин, А. Д. (2016)

Parsegov, S. E., Proskurnikov,  
A. V., Tempo, R., & Friedkin,  
N. E. (2016).

Федянин, Д. Н., &  
Чхартишвили, А. Г.  
(2016)

Логически связанные  
мнения

Gubanov, D., & Petrov, I.

ММО-модели

Нейробиология

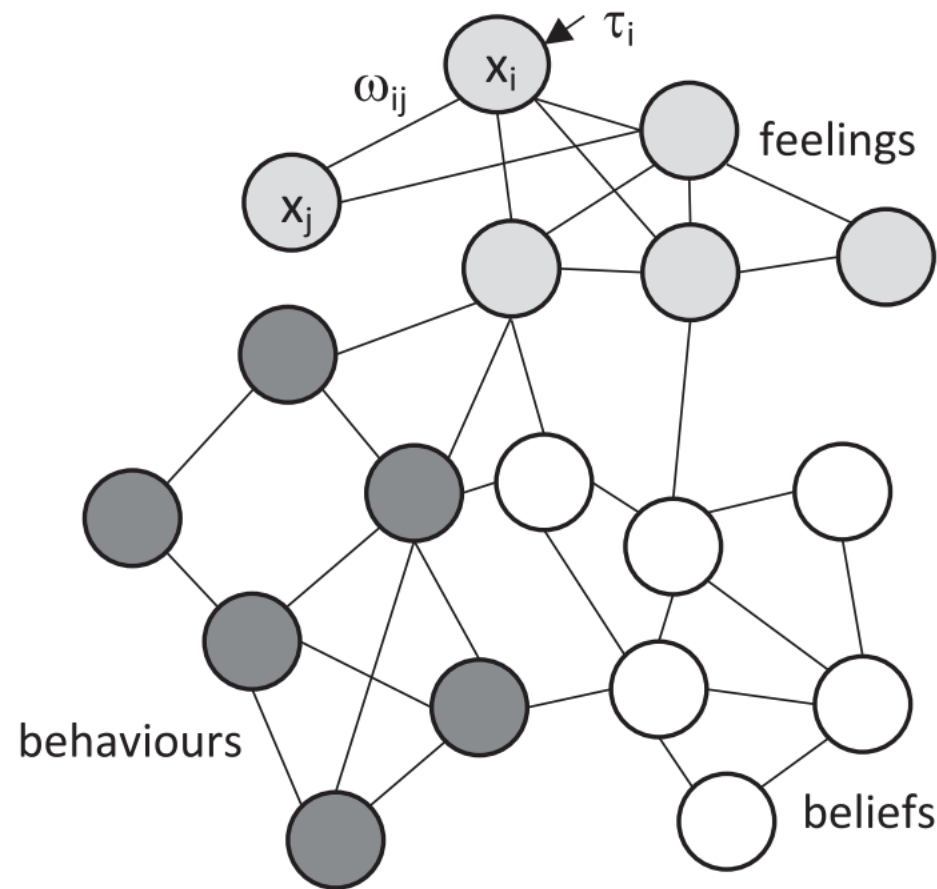
Активация агентов (2019)

Губанов, Д. А. (2022)

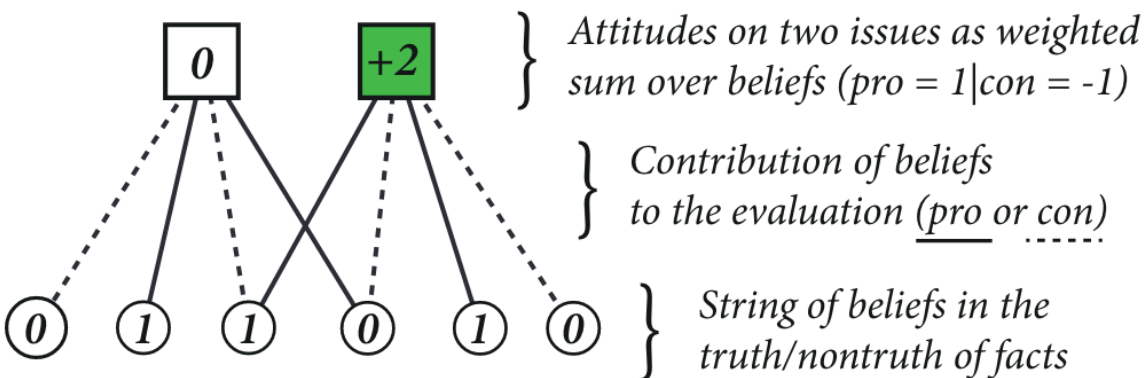
Жилякова, Л. Ю. (2022)

# Тренды

- Многомерные пространства мнений (van der Maas et al., 2020)
- Теория аргументных взаимодействий (Mäs & Flache, 2013)
- Эхо-камеры, поляризация, фейковые новости (Currin et al., 2021)
- Внедрение особенностей онлайн-взаимодействий: информационные каскады, алгоритмы ранжирования (Perra & Rocha, 2019)
- Исследования устройства онлайн-ландшафта с помощью социальных ботов (Chen et al., 2021)



(van der Maas et al., 2020)



(Banisch & Olbrich, 2021)

# Проблемы

Специалисты в области физики	Специалисты в области теории управления + ПСИХОЛОГИ
Physica A: Statistical Mechanics and its Applications	IEEE Transactions on Automatic Control
International Journal of Modern Physics C	Automatica
Physical Review E	American Control Conference
PLOS ONE	IEEE Conference on Decision and Control
The European Physical Journal B	SIAM Journal on Control and Optimization

“Кто где публикуется” – (Mastroeni et al., 2019)

- Быстро растущее число моделей
- Разобщенность различных групп исследователей
- Низкий уровень систематизации
- Слабая эмпирическая база

(Proskurnikov & Tempo, 2017, 2018)

(Flache et al., 2017)

“Публикации, игнорирующие шаг валидации модели, собирают больше цитат”

(Chattoe-Brown, 2022)

# Результаты эмпирических исследований

Публикация	Основной результат (механизм)	Число наблюдений / тип эксперимента
(Moussaïd et al., 2013)	Умеренный (ассимилятивный)	52 и 59 / лабораторный
(Liu & Srivastava, 2015)	Ассимилятивный + диссимилятивный (без конкретизации математической формы)	276 / натурный
(Kerckhove et al., 2016)	Линейный (ассимилятивный)	861 / лабораторный
(Takács et al., 2016)	Линейный (ассимилятивный)	89 и 100 / лабораторный
(Clemm von Hohenberg et al., 2017)	Линейный ассимилятивный	1,504, 1,331, and 781,044 / первые два – лабораторных, третий – натурный (на макро-уровне !)
(Friedkin & Bullo, 2017)	Линейный ассимилятивный	161 / лабораторный
<del>(Friedkin et al., 2021)</del>	<del>Линейный ассимилятивный</del>	<del>107 и 108 / лабораторный</del>
(Kozitsin, 2020), (Kozitsin, 2021)	Ассимилятивный + диссимилятивный (линейный + умеренный)	1,660,927 / натурный

# Анализ динамики мнений пользователей социальной сети ВКонтакте

- **Выборка:** ~1.6 млн аутентичных российских пользователей VK
- **Мнения** от нуля (оппозиция) до единицы (поддержка власти) – 3 снимота (февраль, июль и декабрь 2018 года)
- **Оценка мнений** – на основании подписок на информационные источники ВКонтакте (модель логистической регрессии)
- **Граф** социальных связей (связь “дружба” ВКонтакте) – один снимот (июль 2018)

Тип сдвига:

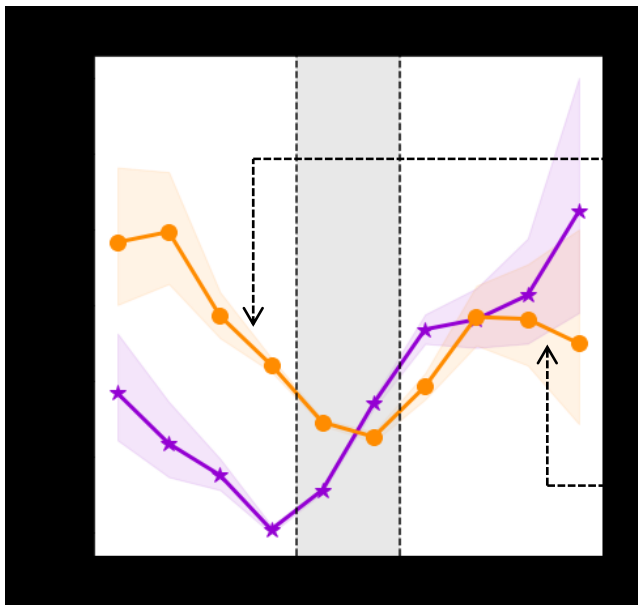


ассимилятивный



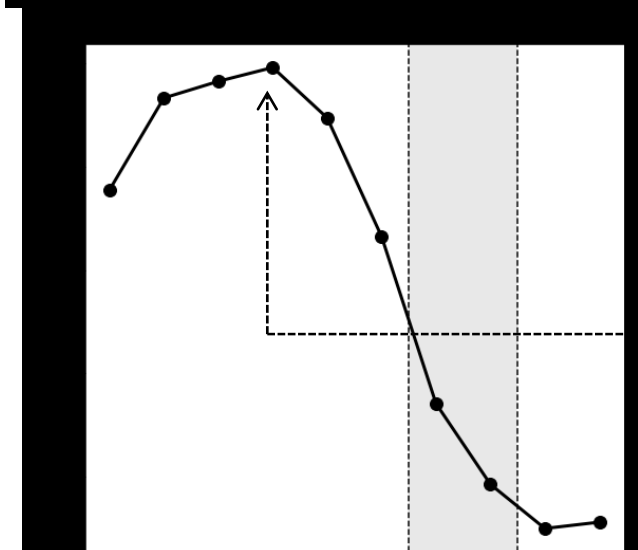
диссимилятивный

# Некоторые результаты



линейная зависимость

спад, который можно интерпретировать как проявление ограниченного доверия



пик зависимости определяет позицию  $x_{-i} \approx 0.35$ , которая характеризуется наибольшей вероятностью вызвать ассимилятивный отклик

- Пользователи редко меняют свои взгляды ( $\sim$  в 1 из 10 случаев)
- Вероятность изменения мнения растет с ростом разницы во мнениях между пользователем и его друзьями как в случае ассимилятивных, так и в случае диссимилятивных сдвигов
- Характер зависимостей не относится ни к одному из основных классов функций отклика, однако кривые имеют четко выраженные линейные и перевернутые U-образные фрагменты
- Баланс между ассимилятивными и диссимилятивными сдвигами зависит от величины  $|x_i - x_{-i}|$  – вероятность ассимилятивного сдвига наиболее велика, если абсолютная разница во мнениях между объектом и источником влияния находится в “умеренной” зоне – не слишком мала и не слишком велика.

# Перспективные направления для дальнейших исследований

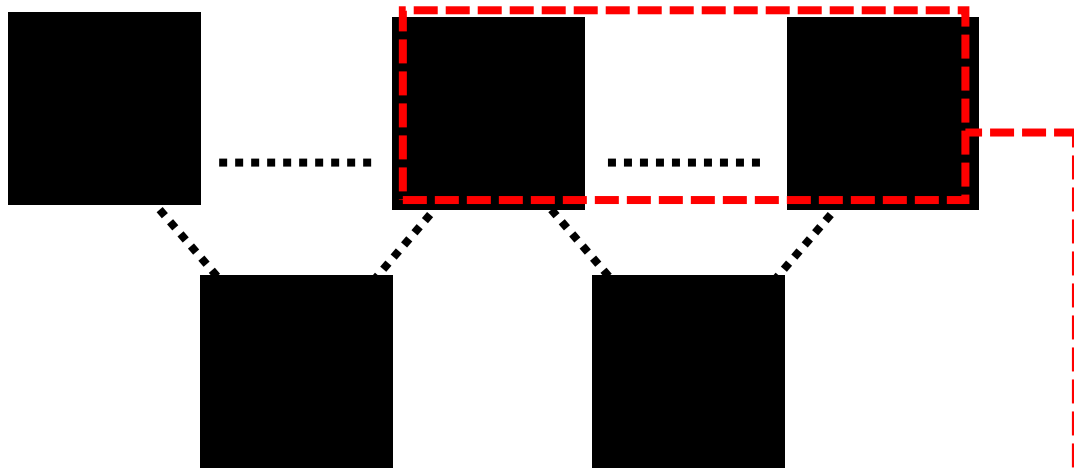
- Дальнейший сбор, систематизация эмпирических данных. Определение **контекстов**, при которых возникает диссимильное влияние
- Детализация пространства мнений: аргументы + **правдивость информации**
- Динамика структуры взаимодействий – внедрение адекватных моделей динамики социальных сетей, учитывающих не только селективность по взглядам, но и другие ключевые детерминанты формирования социальных сетей: предпочтительное присоединение, транзитивность, селективность по возрасту и т.д.
- Систематизация, сравнение существующих моделей, приведение к общей теоретической базе, построение и развитие унифицирующих (**!минимальных**) моделей



Спасибо за внимание!

# SCARDO модель – “табличный” подход к описанию процессов социального влияния

Структурированная социальная система



Локальные последовательные взаимодействия



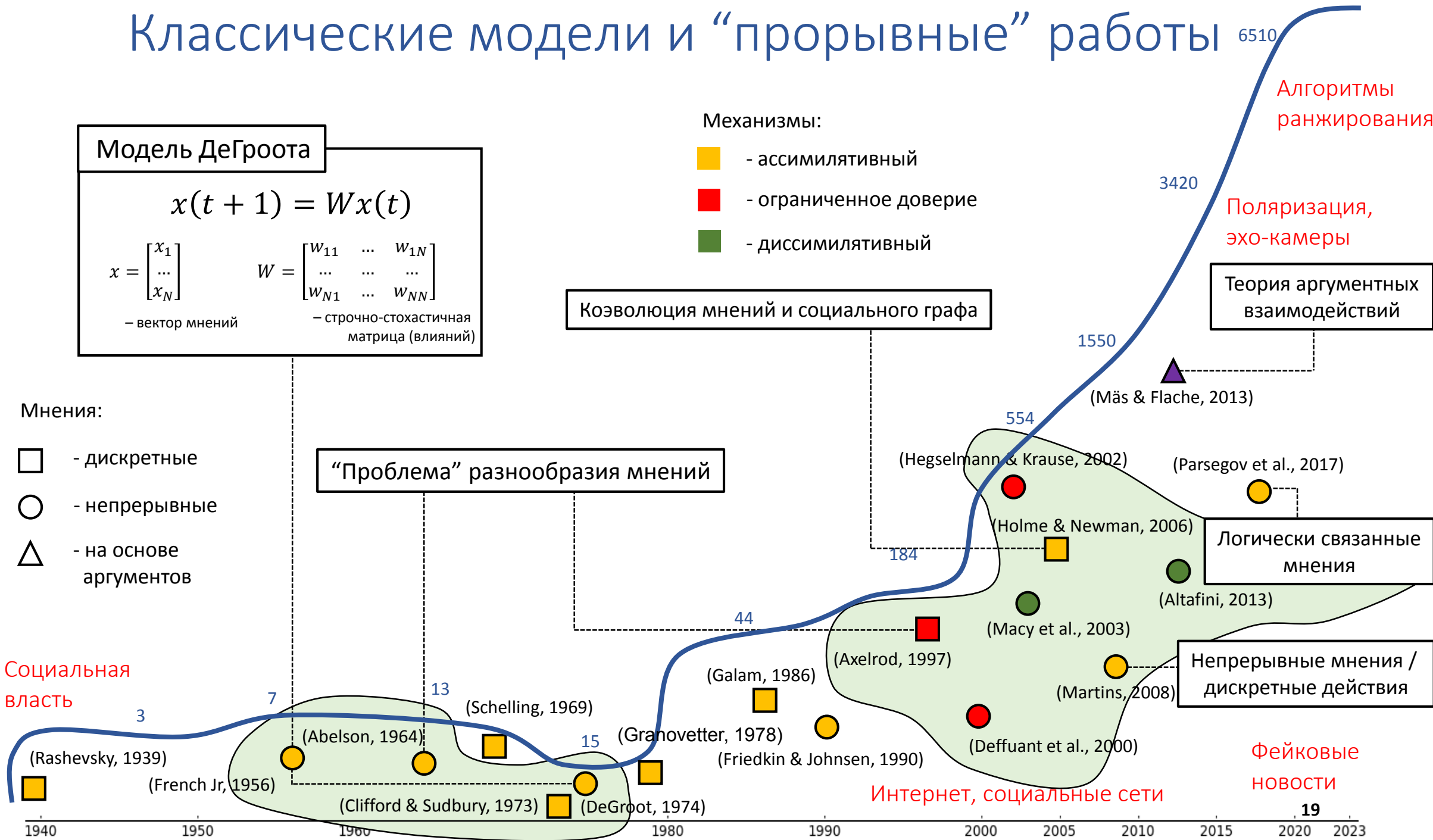
Азбука мнений:  $\Xi = \{\Xi_1, \dots, \Xi_m\}$

$\Xi_1$	$\Xi_2$	...	$\Xi_m$
■	■■■	■■■■■	■

Исход взаимодействия определяется стохастически на основании таблицы переходов

$\Xi_s \rightarrow \Xi_1$	$p_{s,l,1}$
...	...
$\Xi_s \rightarrow \Xi_m$	$p_{s,l,m}$

# Классические модели и “прорывные” работы



# Приближение среднего поля

(число агентов велико, граф знакомств – полный граф)

Пусть  $y_f(t)$  - число агентов в момент времени  $t$ , придерживающихся позиции  $\Xi_f$ . Для

масштабированного времени  $\tau = \frac{t}{N}$ ,  $\delta\tau = \frac{1}{N}$  получим задачу Коши:

$$\frac{dy_1(\tau)}{d\tau} = \sum_{s,l,k} y_s(\tau)y_l(\tau)p_{s,l,k}(\delta_{k,1} - \delta_{s,1}),$$

...

$$\frac{dy_m(\tau)}{d\tau} = \sum_{s,l,k} y_s(\tau)y_l(\tau)p_{s,l,k}(\delta_{k,m} - \delta_{s,m}),$$

$$y_1(0) = q_1, \dots, y_m(0) = q_m,$$

где  $q_1 + \dots + q_m = 1$  и все  $q_i \geq 0$ .

(1)

# Приближение среднего поля

(число агентов велико, граф знакомств – полный граф)

**Утверждение 1.** Функция  $y_1 + \dots + y_m$  является первым интегралом системы (1).

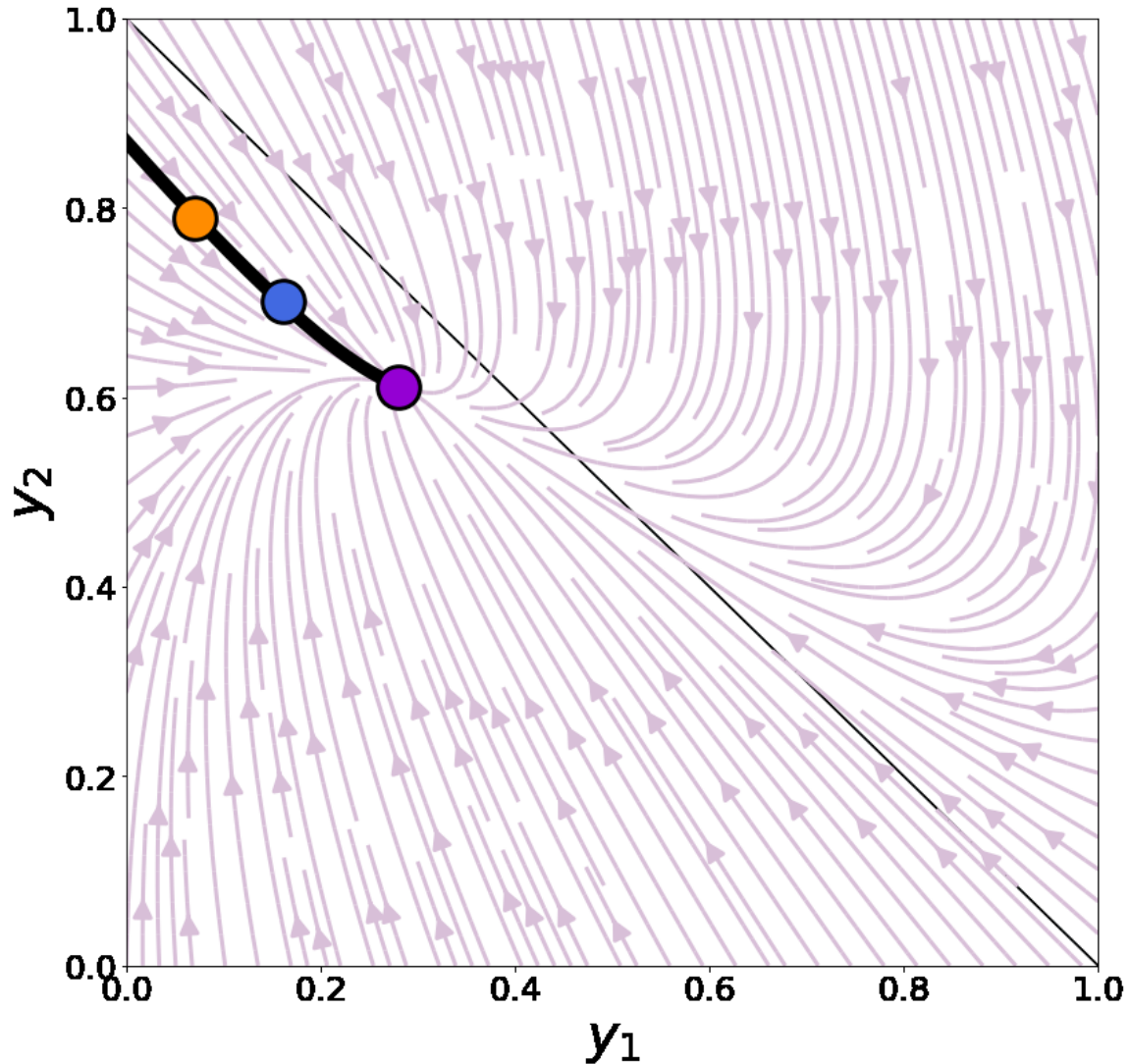
**Утверждение 2.** Уравнения системы (1) можно переписать в виде

$$\frac{dy_f(\tau)}{d\tau} = \sum_{s,l} y_s(\tau)y_l(\tau)p_{s,l,k} - y_f(\tau).$$

**Утверждение 3.** Фазовые переменные в (1) неотрицательны.

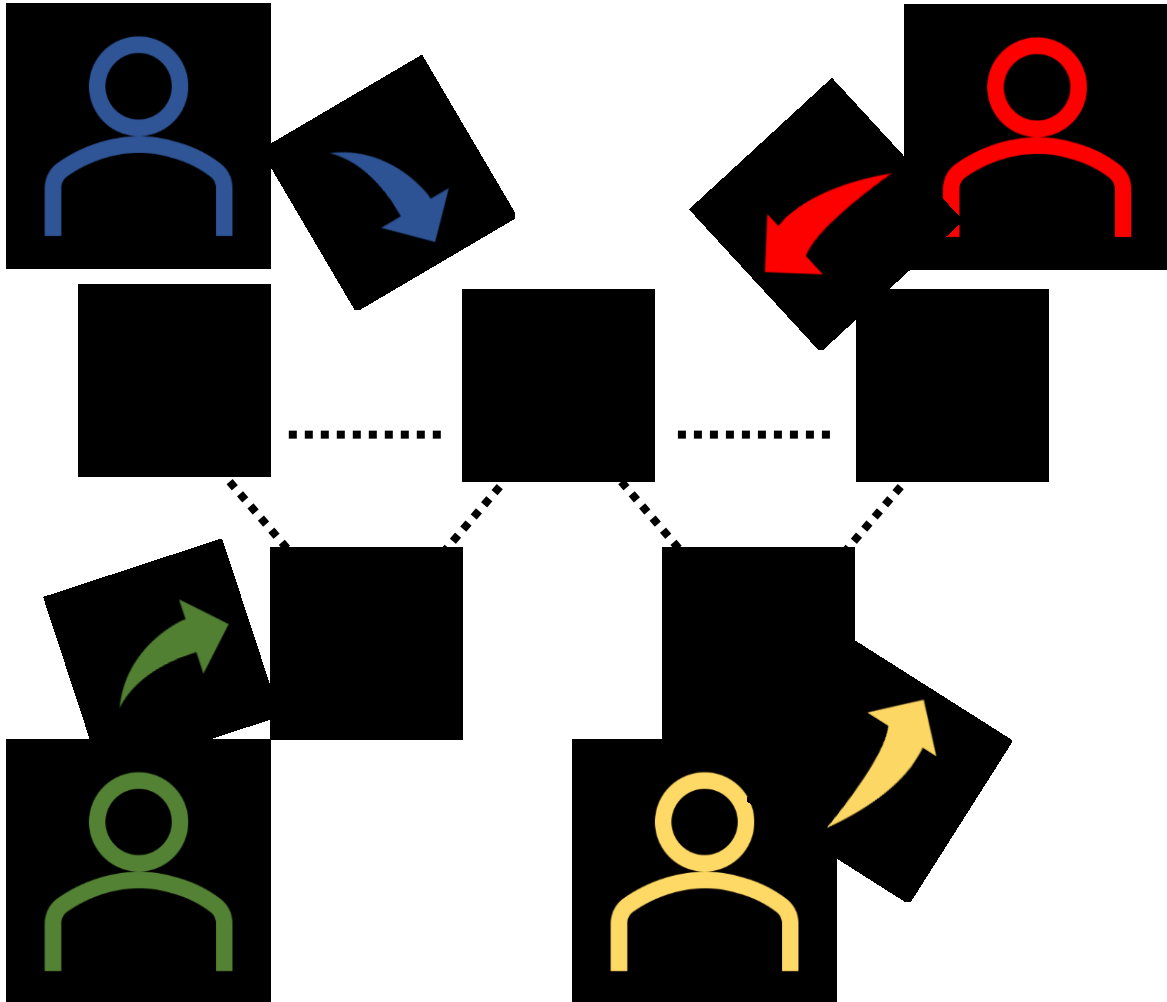
**Утверждение 4.** Решение Задачи Коши (1) существует, единственно и продолжимо на всю числовую ось. Является аналитической функцией параметров модели и начальных условий.

# Фазовый портрет динамической системы (социальная сеть ВКонтакте)



- Синяя точка – положение эмпирической системы ( $y_1 = 0.161, y_2 = 0.702, y_3 = 0.137$ ) в Феврале 2018 года
- Фиолетовая точка – положение равновесия ( $y_1^* = 0.28, y_2^* = 0.611, y_3^* = 0.109$ )
- Положение равновесия – устойчивый узел (спектр матрицы Якоби состоит из двух различных отрицательных чисел)

# Расширение SCARDO модели: социальные боты



- Система включает два типа агентов: обычные агенты и социальные боты
- Динамика описывается двумя матрицами переходов:
- Матрица  $P = [p_{s,l,k}]_{s,l,k=1}^m$  описывает взаимодействия между обычными агентами
- Матрица  $U = [u_{s,l,k}]_{s,l,k=1}^m$  формализует влияние ботов на обычных агентов
- Мнения ботов неизменны и являются экзогенными параметрами

**Задача управления: настроить мнения ботов таким образом, чтобы изменить мнения обычных агентов**