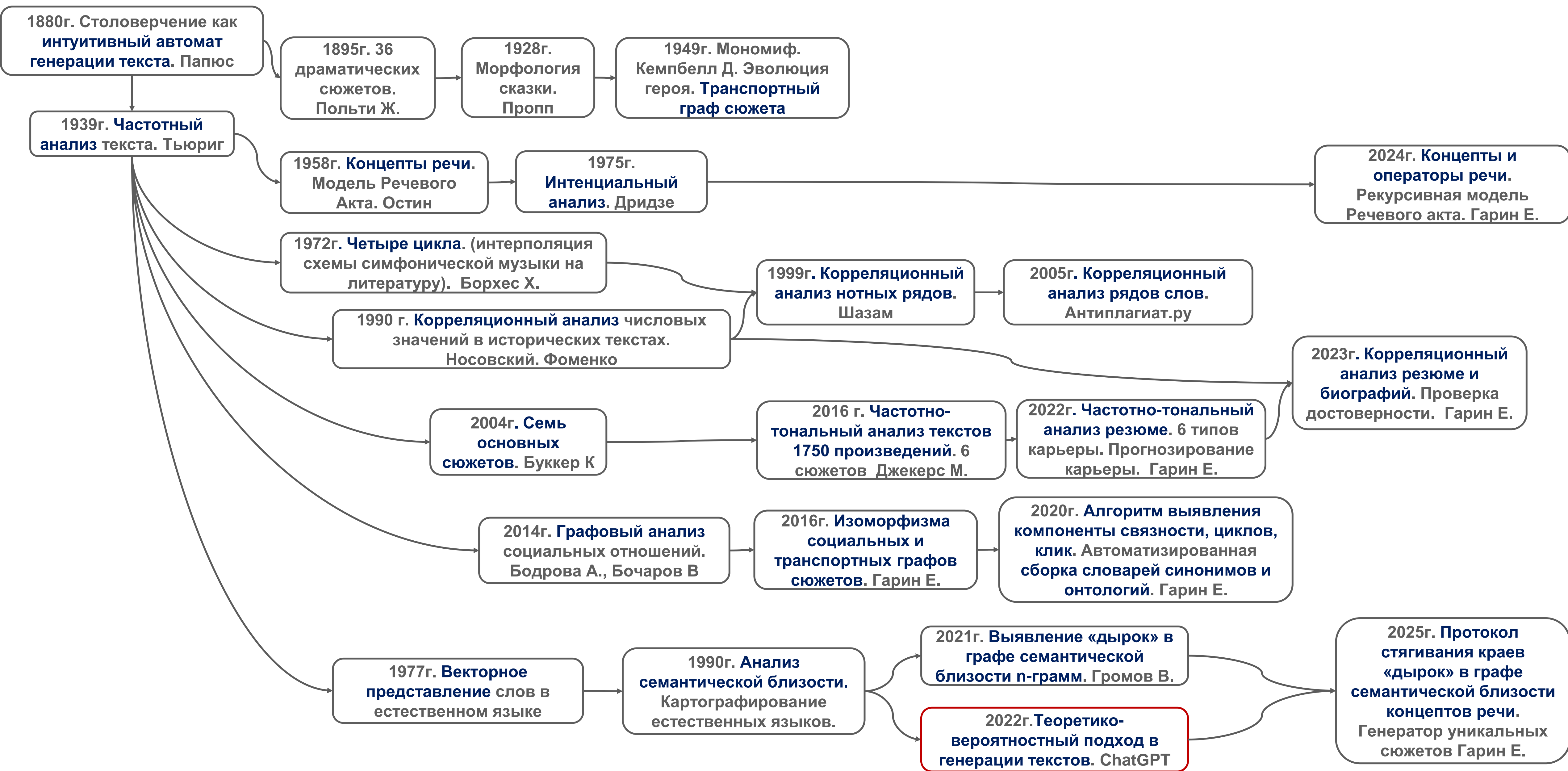


Ключевые слова: теория графов, математическое моделирование социума, выделение апеллятивов и топонимов в тексте и семантическом графе, построение транспортного и социального графа, сравнение сюжетов методом изоморфизма.

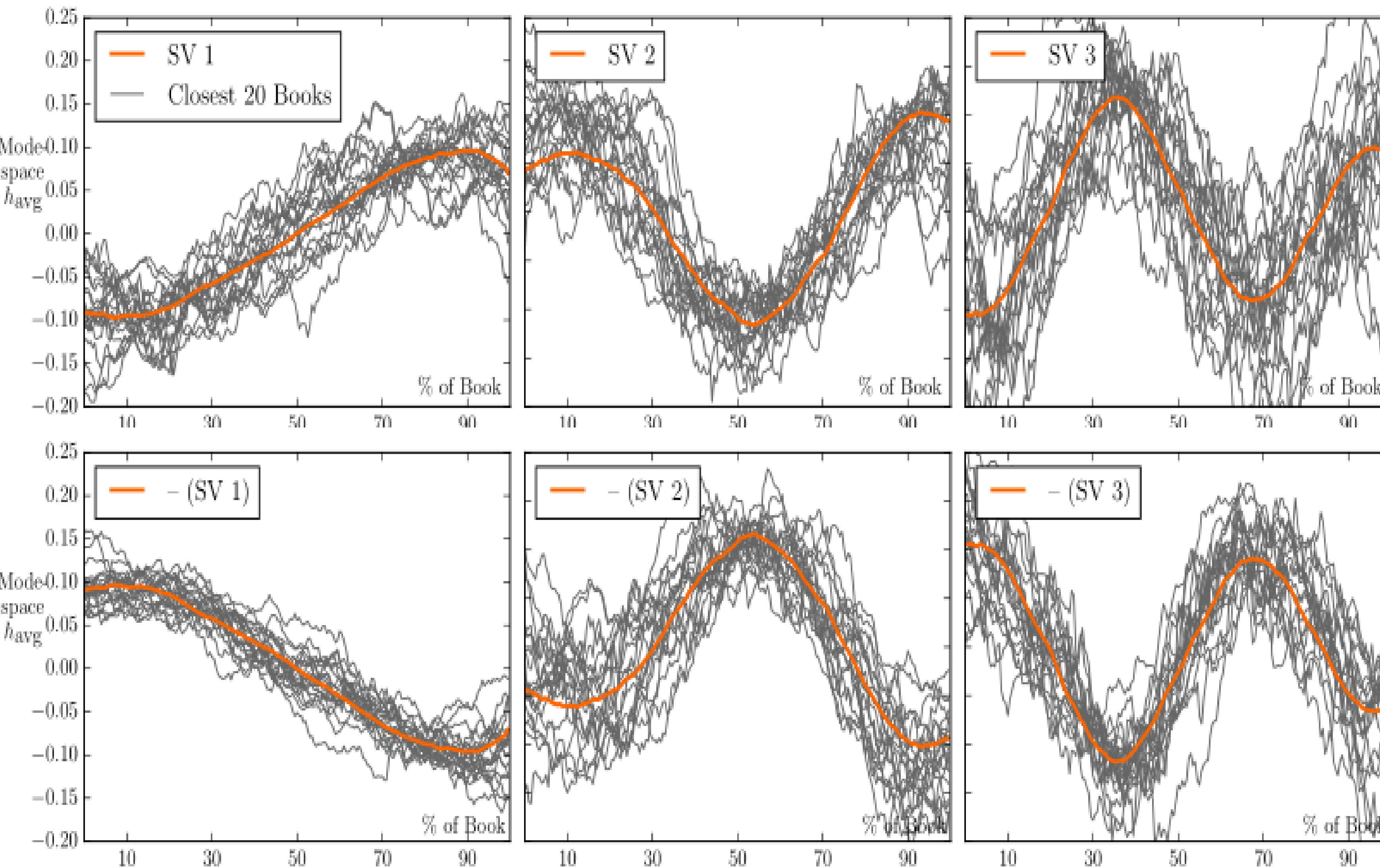
Аннотация: Классификация сюжетов как объект исследования литературоведения сводилась к сведению сюжетов в группы по схожести из совокупного корпуса текстов, с которыми был знаком исследователь. При этом каждый литературовед, классифицирующий сюжеты, давал свое субъективное определение схожести. Поэтому на сегодняшний момент существует несколько устоявшихся классификаций уникальных сюжетов из 36, 16, 8, 6 и 1 сюжета. Иными словами, в зависимости от метода сравнения сюжеты мировой литературы можно свести к ограниченному количеству групп – вплоть до 1 группы! В отличие от всех остальных классификация сюжетов на 6 групп методом частотно-фонового анализа является плодом междисциплинарного исследования на стыке литературоведения, психологии и комбинаторики. Фактически это первая классификация, полученная с применением математических методов компьютерной лингвистики. В данной статье представлен новый подход представления сюжетных сцен как социальных и транспортных графов, сюжетных поворотов как операций над графами, а самого сюжета как эволюции социального и транспортного графа. Новый подход позволил классифицировать сюжеты на группы на основе метода изоморфизма графов, разработать метод выявления заимствования сюжетов, классифицировать методы сокрытия заимствования сюжетов, построить зависимость количества уникальных сюжетов (уникальных графов) от количества персонажей (количества вершин в графе) и ответить на вопрос является ли количество уникальных сюжетов конечной либо бесконечной величиной.

Исторический обзор подходов к классификации сюжетов



Частотно-тональный анализ литературных текстов. Джекерс

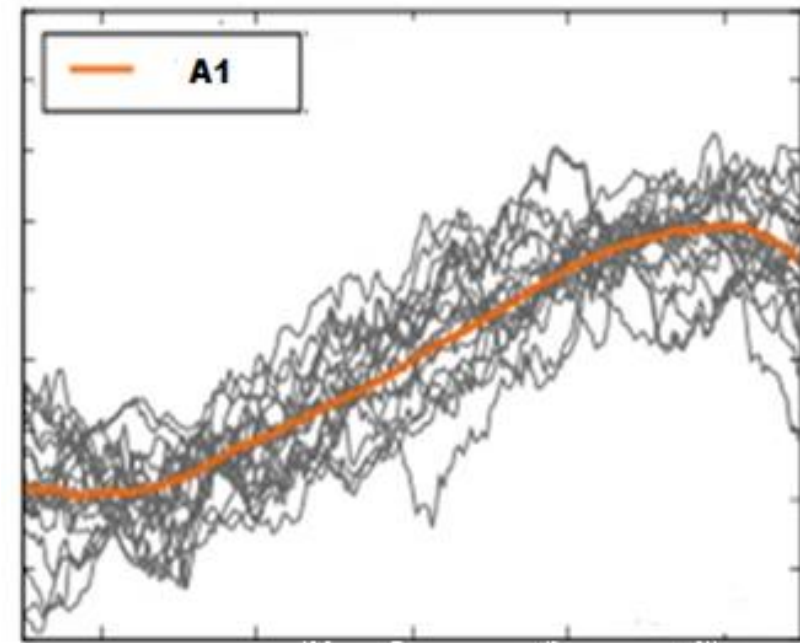
Частотно-тональный анализ текстов резюме. Гарин



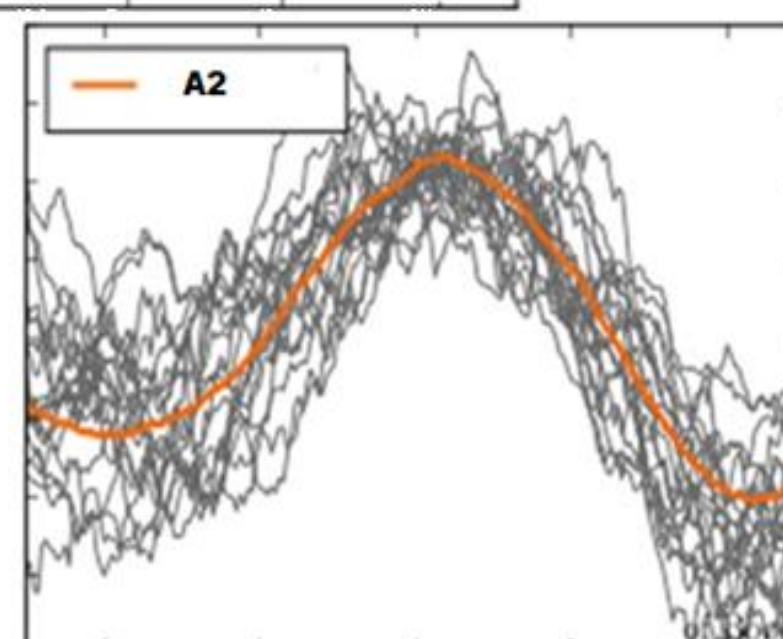
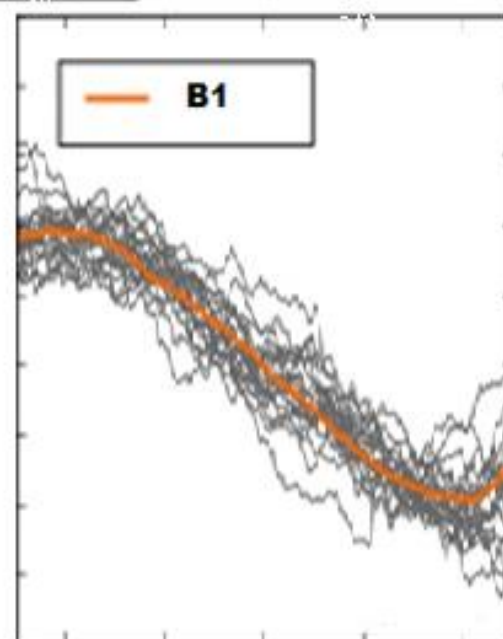
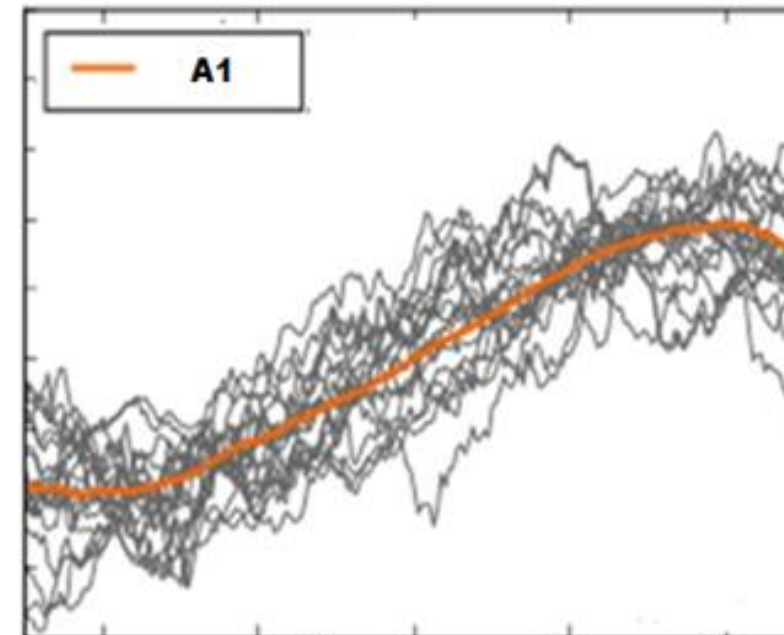
85% образовательно-трудовых траекторий и дифференцируются на 6 типов, совпадающих с основными поведенческими архетипами, мировой литературы

Вероятность девиантного поведения коррелирует с отклонением образовательно-трудовой траектории от нормального распределения для каждого из 6 типов развития карьеры

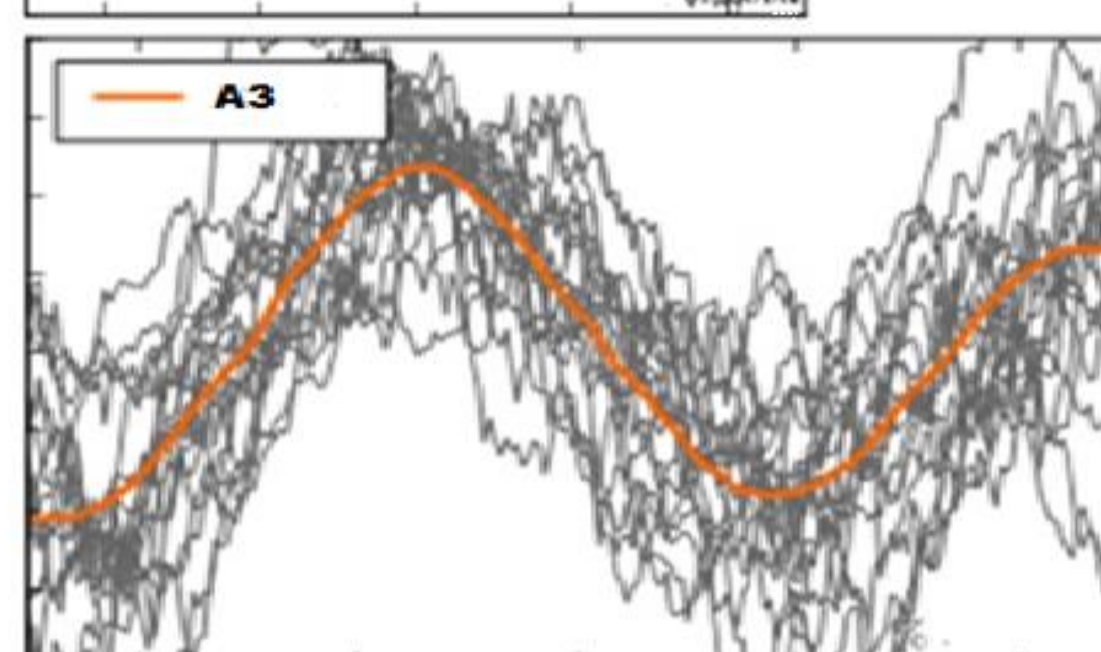
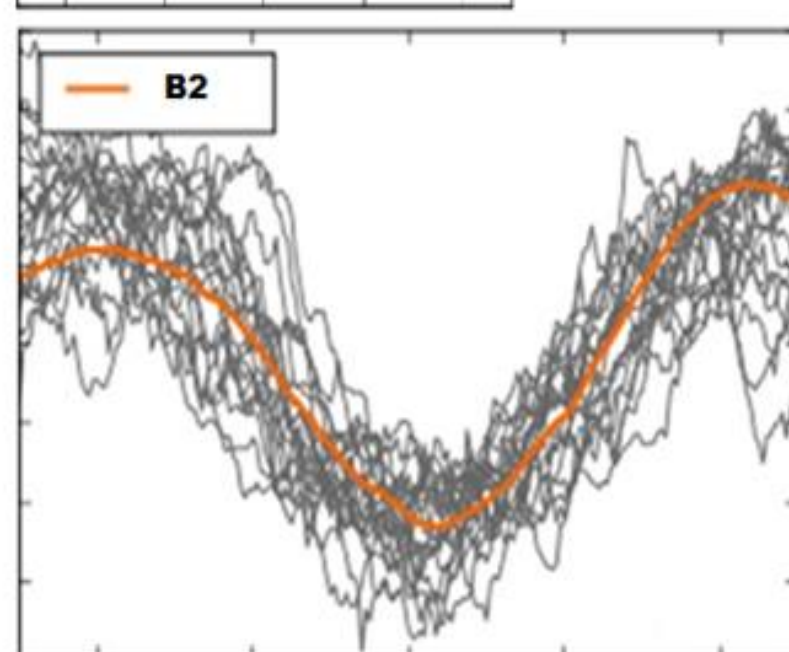
Частотно-тональный анализ текстов резюме. Гарин



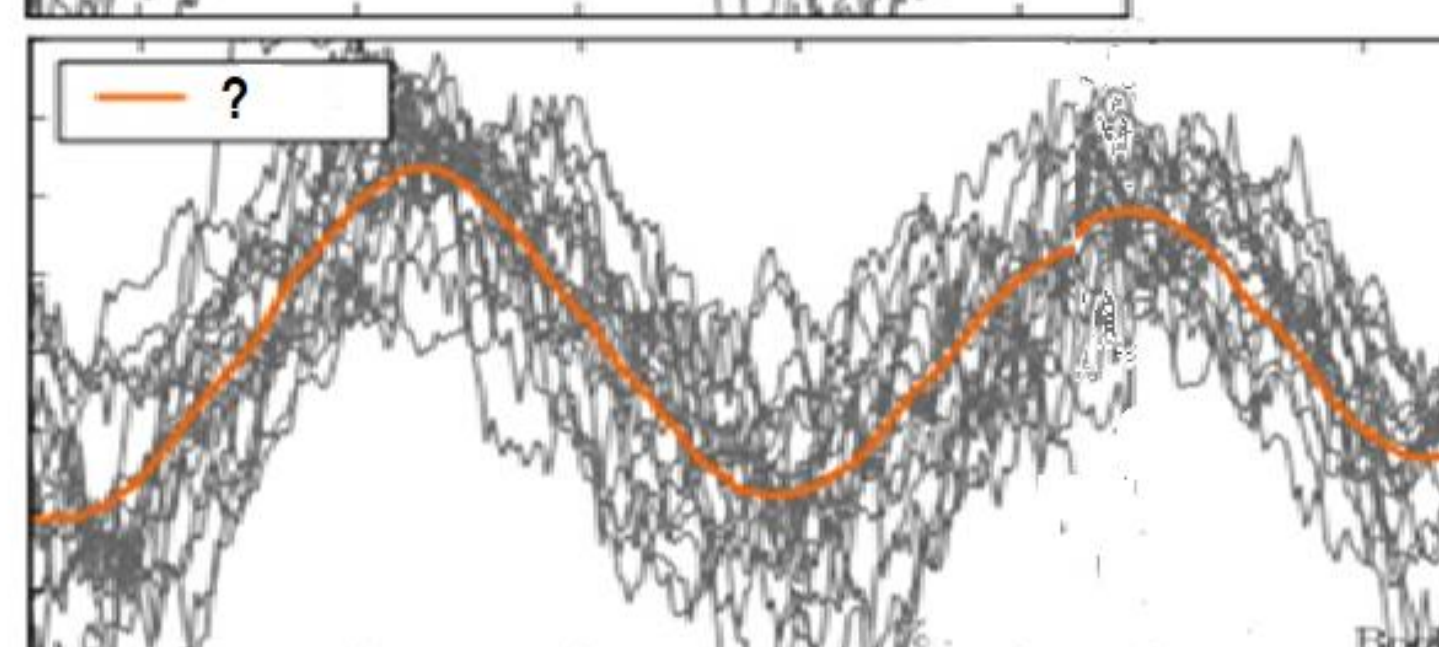
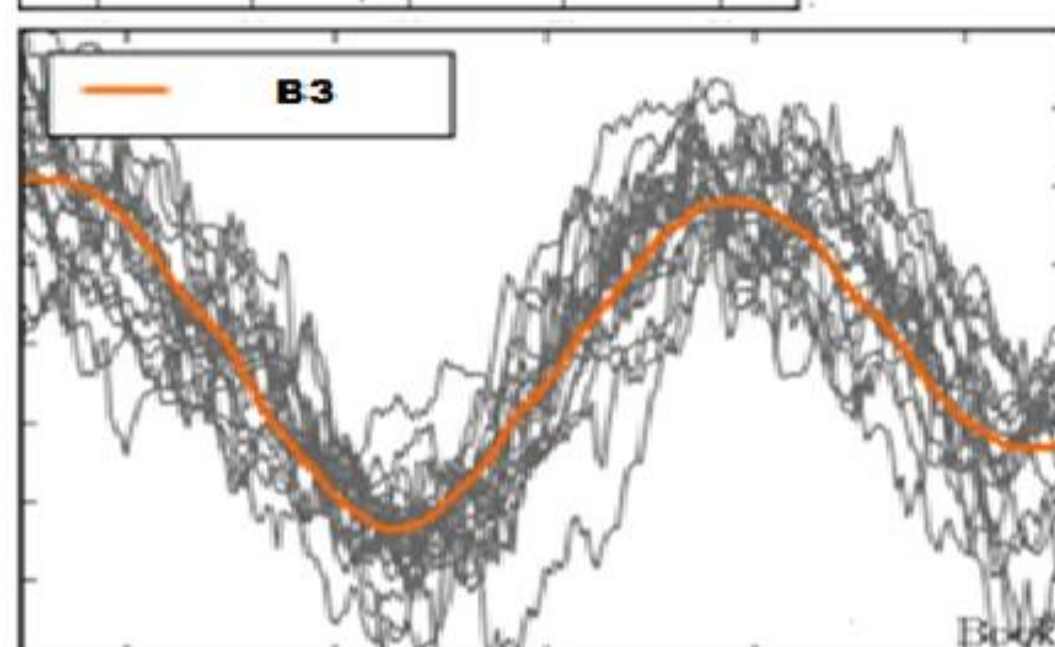
A1 – основной тип образовательно-трудовых траекторий



Все остальные типы образовательно-трудовых траекторий не самостоятельны, несут признаки срежессированности и зависимости от основного типа карьерой траектории A1

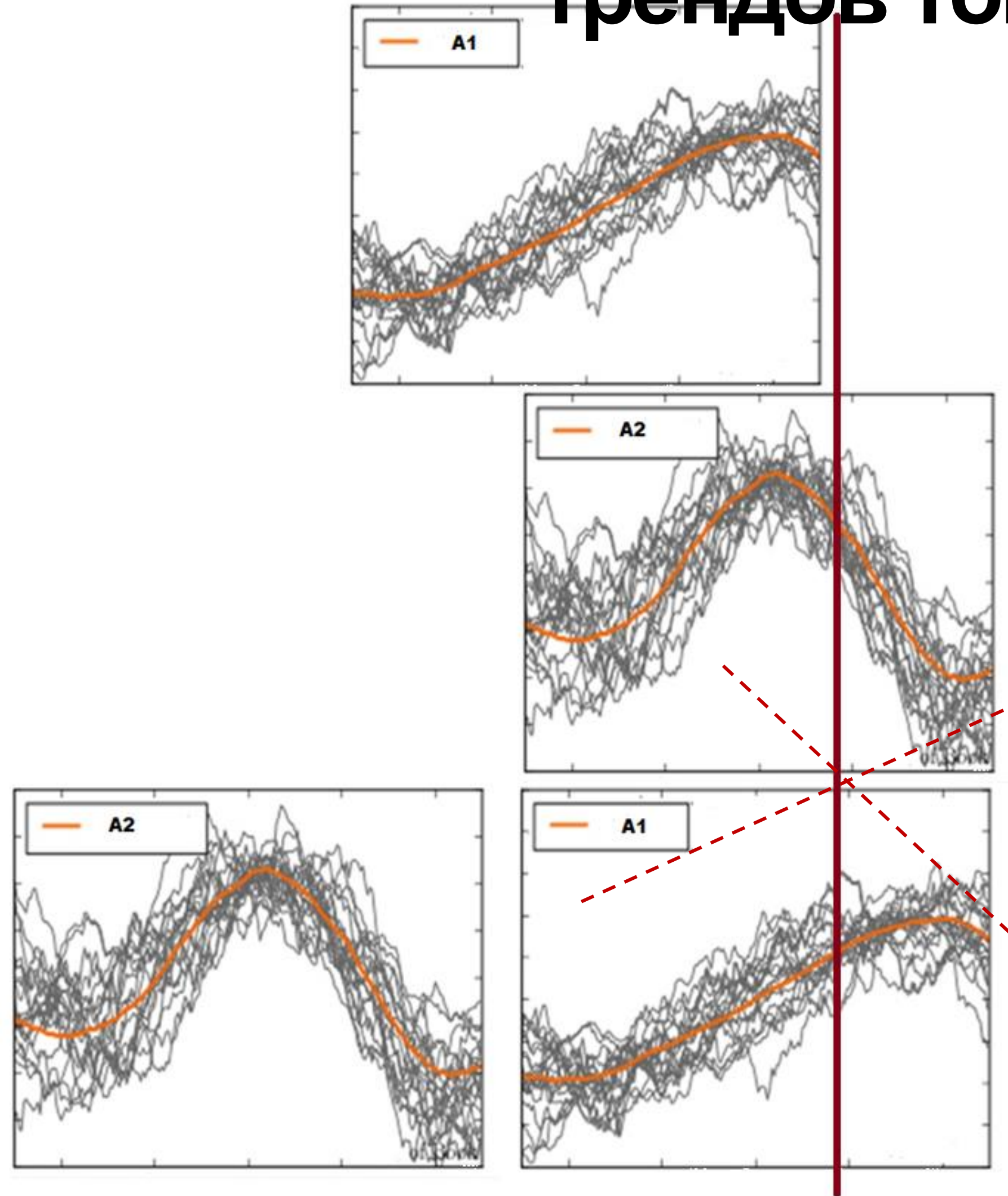


Карьерные траектории A3 и B3 появились из-за роста продолжительности и качества жизни

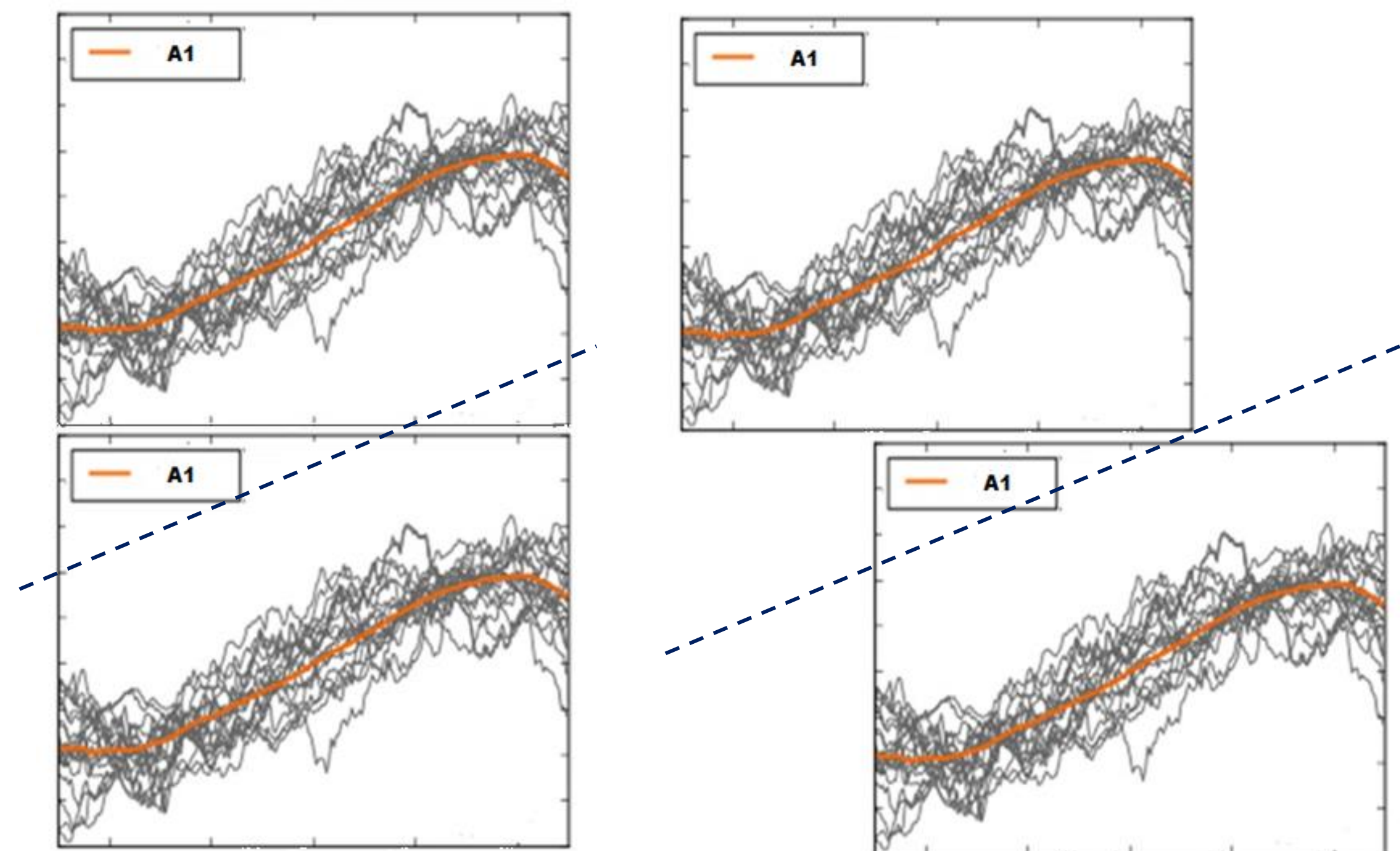


Дальнейший рост продолжительности и качества жизни привет к росту количества синусоидальных карьерных траекторий

Проверка сюжетов литературных произведений на достоверность трендов тональности пар персонажей

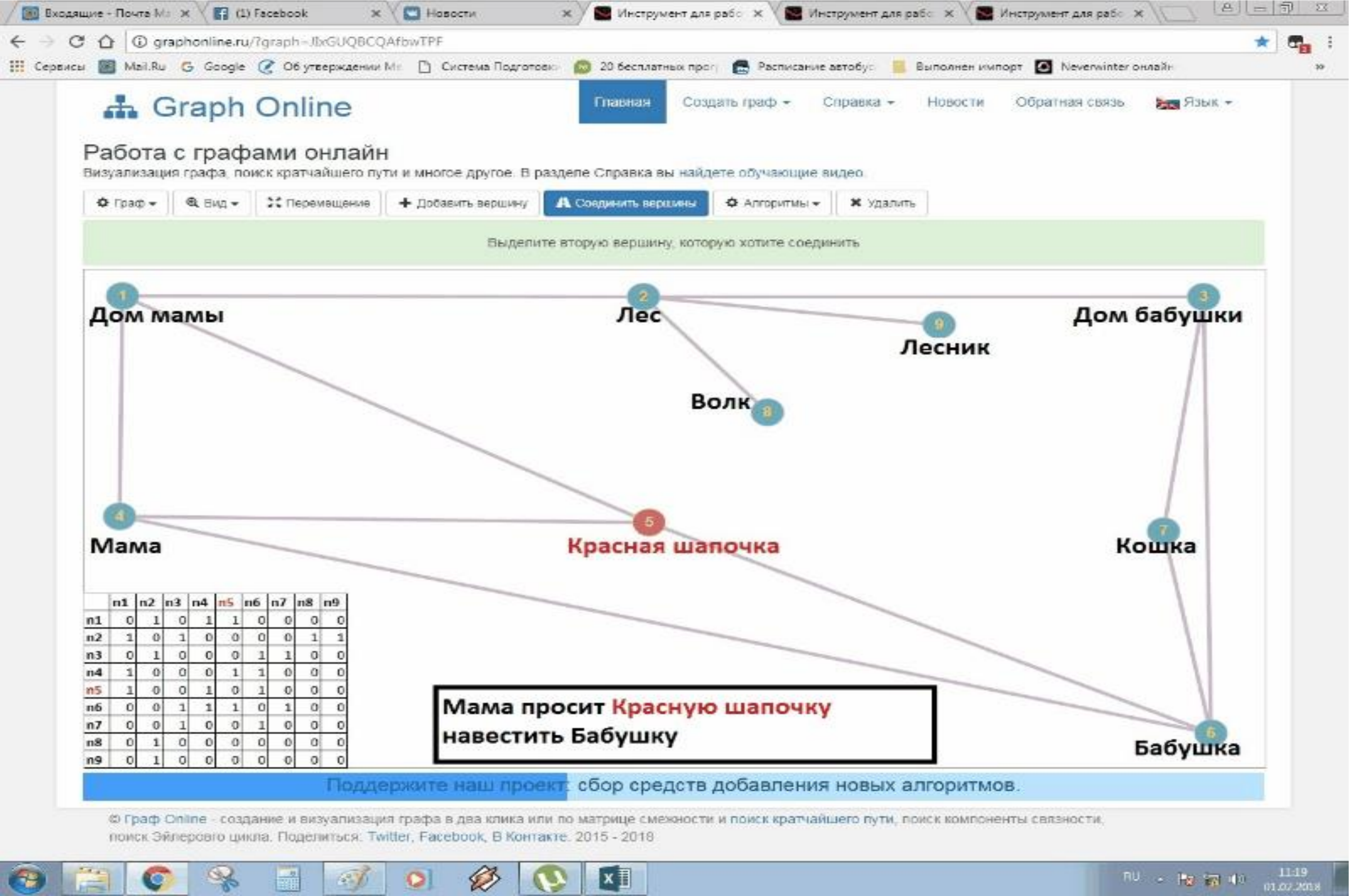


Причина большинства разводов заключается в рассогласованности трендов образовательно-трудовых траекторий



Пример идеальной согласованности трендов образовательно-трудовых траекторий

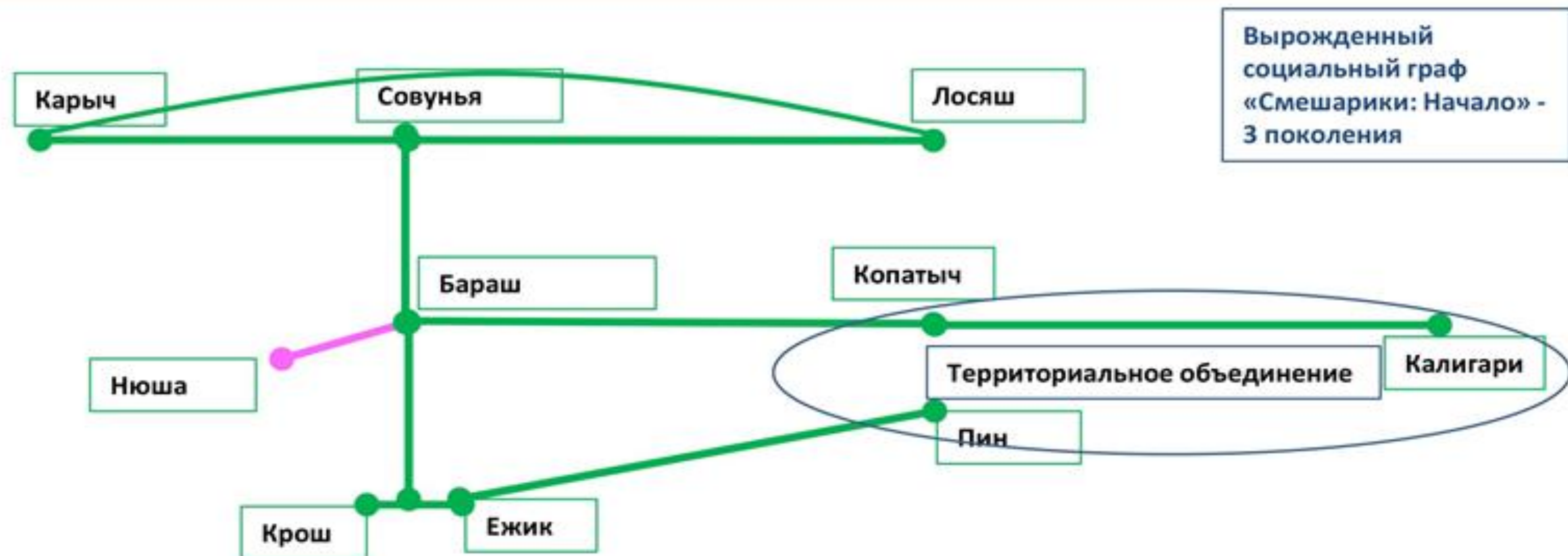
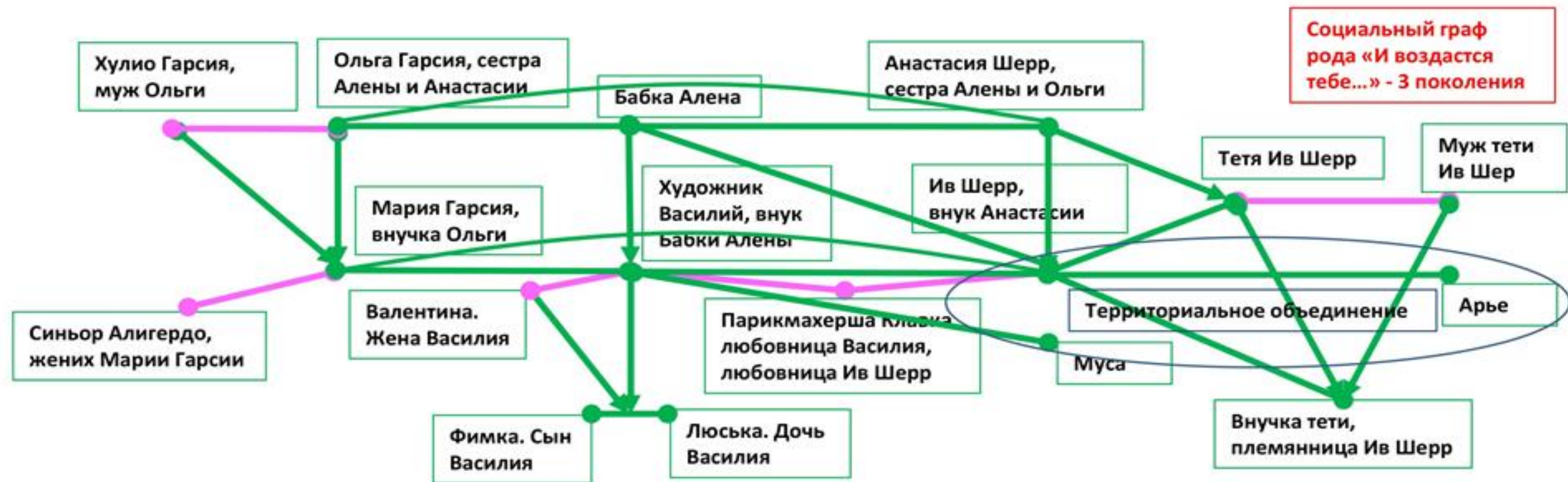
Пример представления сюжета литературного произведения в виде эволюции социального-транспортного графа



Произведение:
Шарль Перо. Красная шапочка.

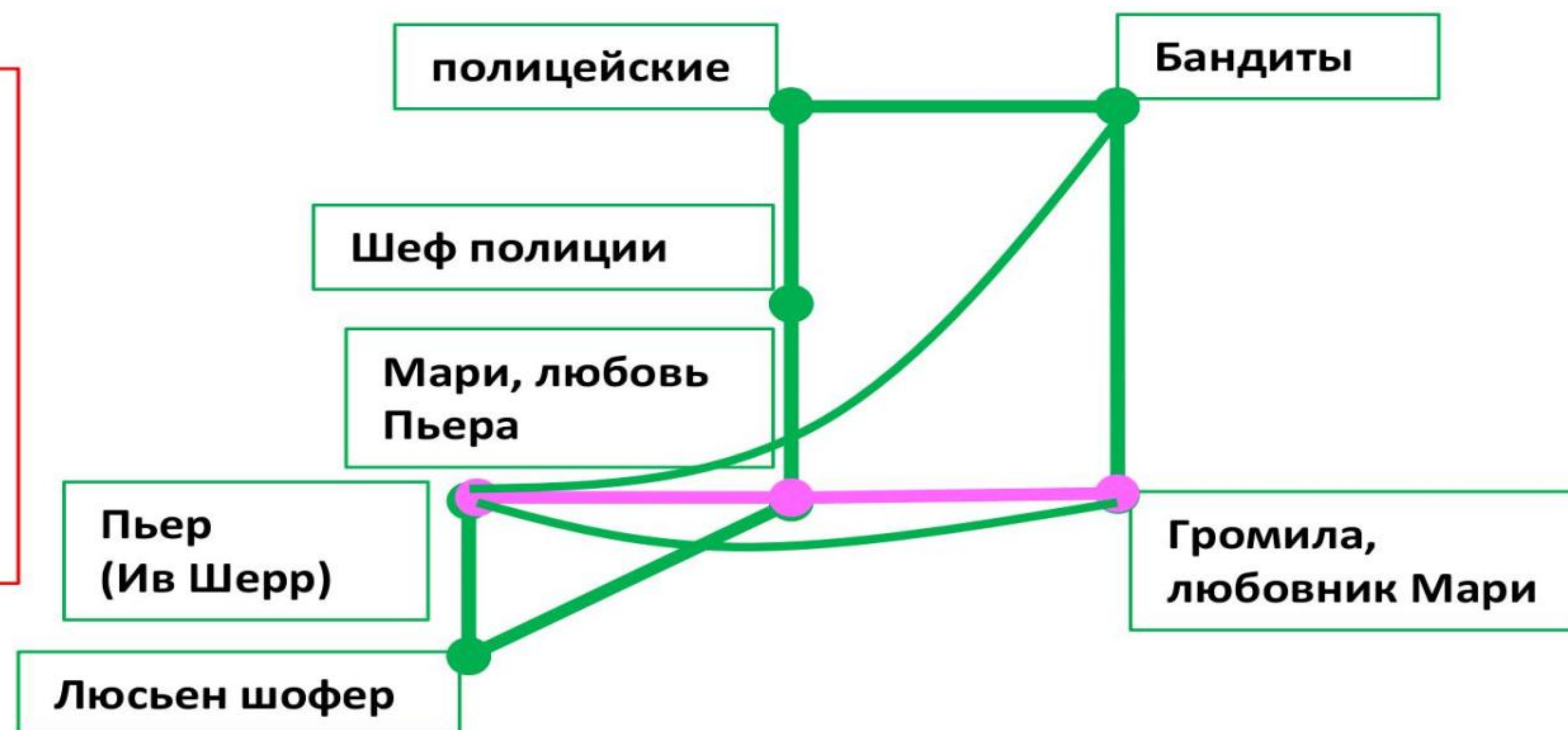
Размер в формате txt
14612 байт
Представление в
виде социального-
транспортного графа
182 бита (9*9*18
матриц)

Коэффициент сжатия
для задач выявления
заимствования
сюжета и проверки
на логическую
достоверность ≈81



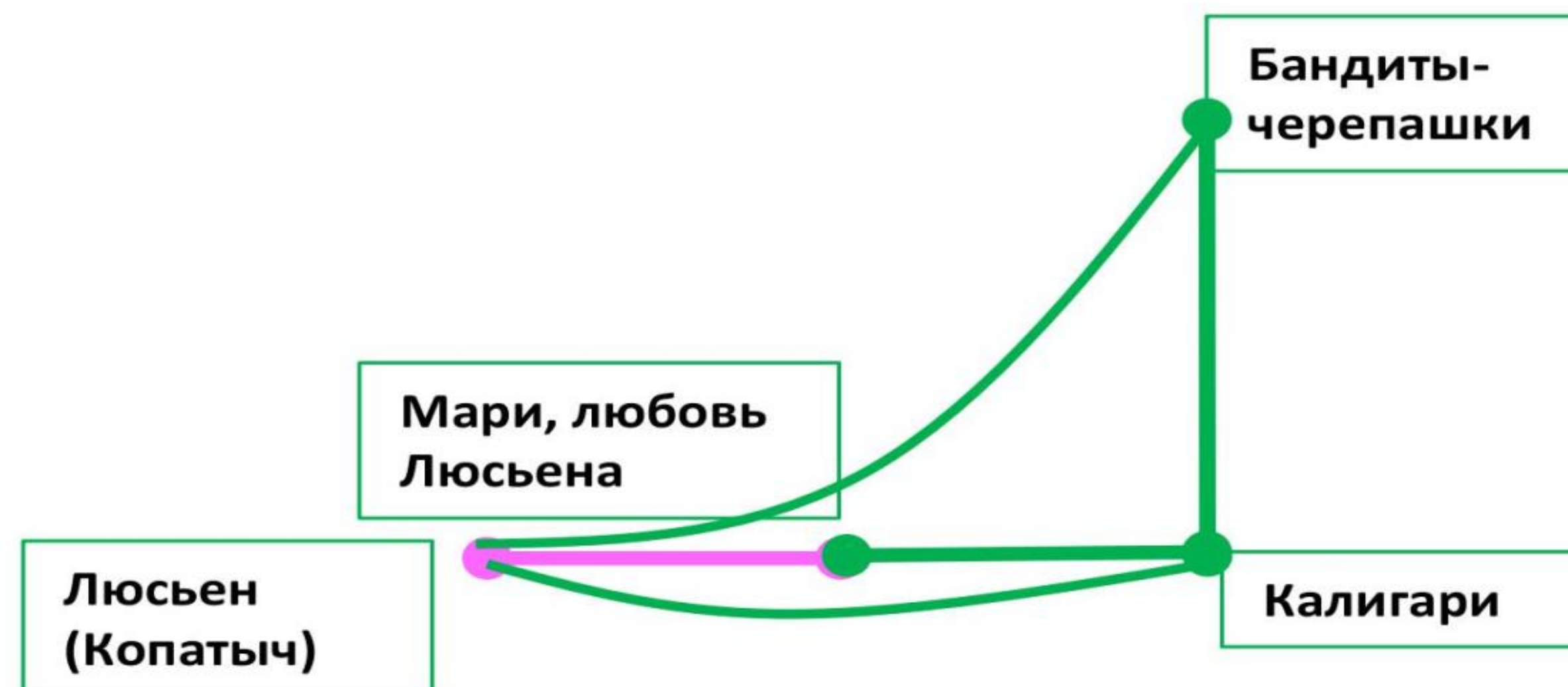
Вырожденный граф кино «Спаситель» в «Самородок»:

- 1) Разрушается связь Мари и Громилы
- 2) Умирает Пьер – пропадает его вершина со всеми связями
- 3) Умирают бандиты – пропадает их вершина со всеми связями
- 4) Умирает Громила – пропадает его вершина со всеми связями



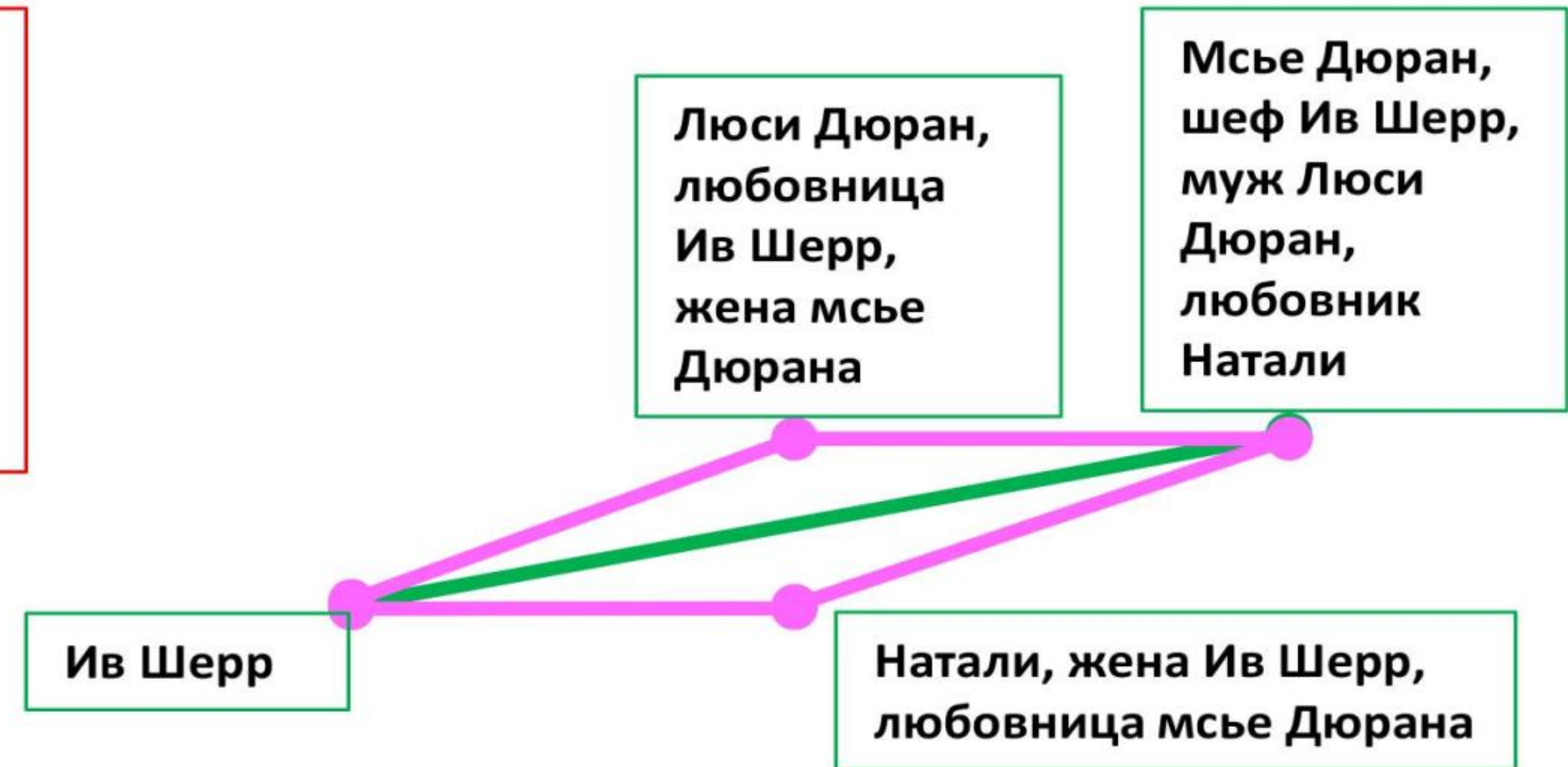
Вырожденный граф кино «Титаник» в «Смешарики : Начало»:

- 1) Разрушается связь Мари и Калигари
- 2) Умирает Мари – пропадает ее вершина со всеми связями
- 3) Повержены бандиты-черепашки – пропадает их вершина со всеми связями
- 4)...



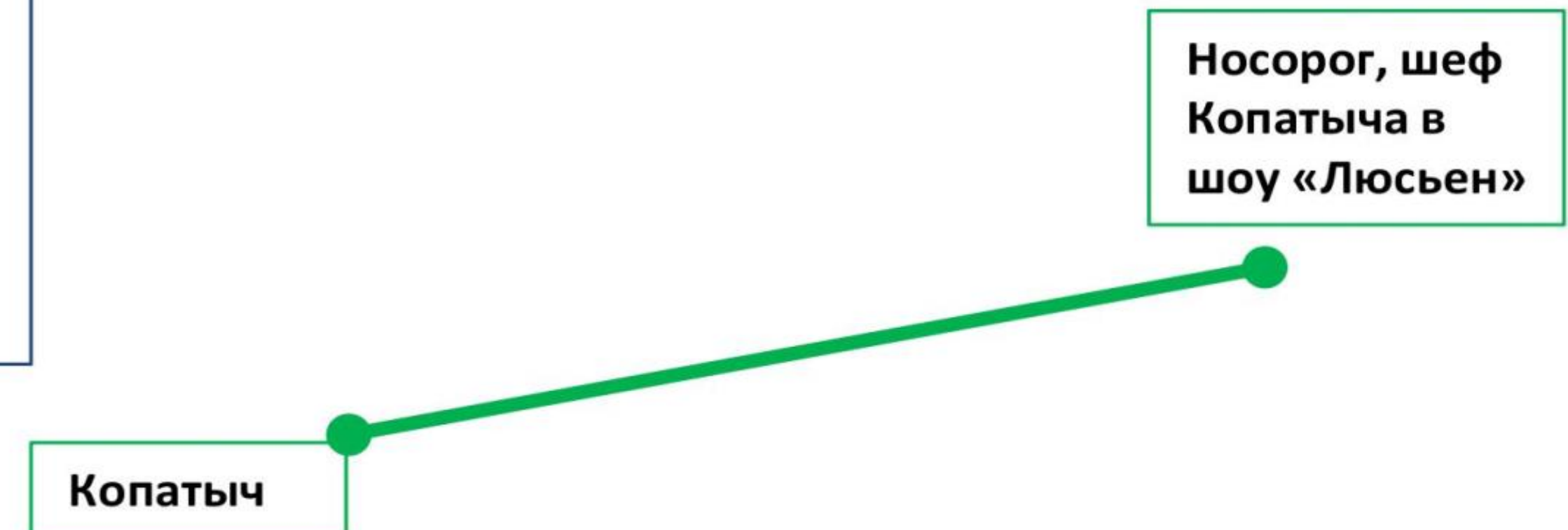
Социальный граф ближайшего социального окружения ИВ Шерр в «Самородке»:

- 1) Разрушается связь Ив Шерр с Люси
- 2) Разрушается связь Ив Шерр с Натали
- 3) Разрушается связь Ив Шерр с мсье Дюраном
- 4) Умирает мсье Дюран – пропадает его вершина со всеми связями



Вырожденный социальный граф ближайшего социального окружения Копатча в «Смешарики: Начало»:

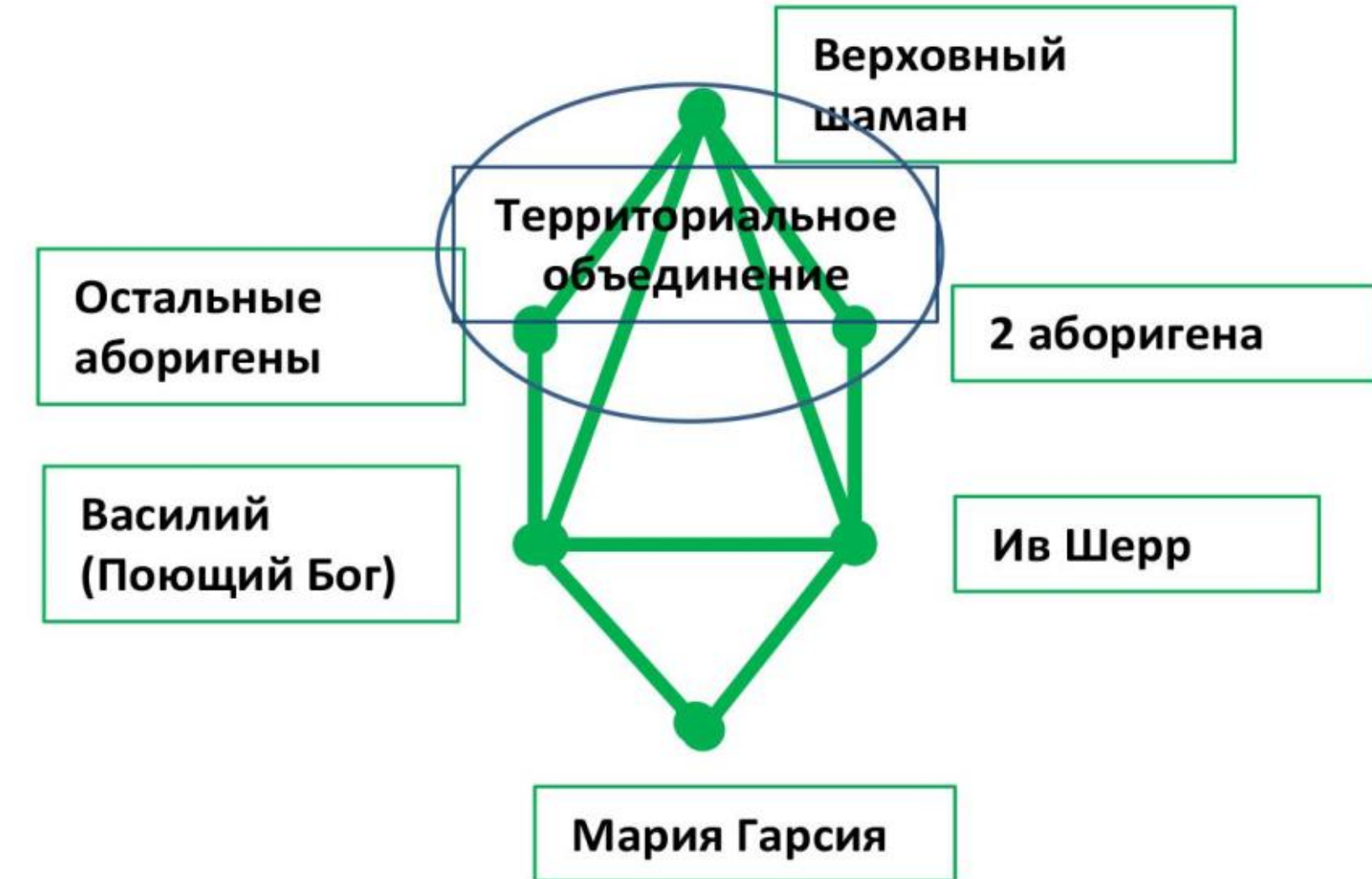
- 1) Разрушается связь Копатыча с Носорогом
- 2)...
- 3)...
- 4) Носорога уличают в связи с музейными ворами – его вершина не пропадает, но ее социальные связи с социумом рвутся



Эволюция основного социального графа 2 главы «Семейная тайна» в:

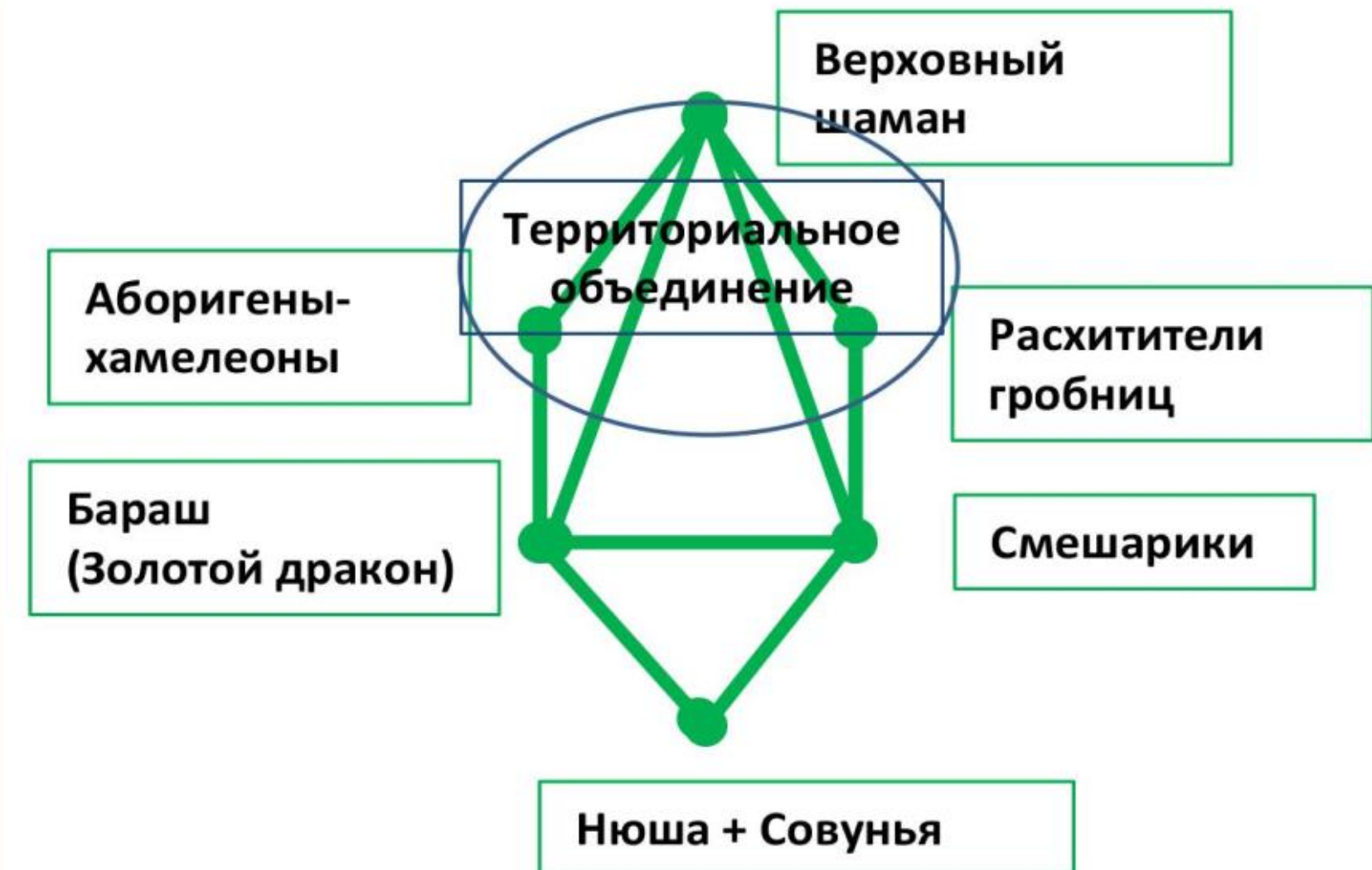
«И воздастся тебе»:

- 1) Разрушается связи Марии Гарсии с Василием и Ив Шерр (те прыгают с парашютами).
- 2) Разрушается связь между Василием и Ив Шерр (Их разделяет Верховный Шаман)
- 3) Аборигены разделяются на 2 группы. Одна контактирует с Василием. Другая с Ив Шерр
- 4) Совершаются 3 безуспешные попытки удаления вершины - убийства Ив Шерр
- 5) Удаляется вершина с 2 аборигенами (те погибают)
- 6) Совершаются безуспешная попытка удаления вершины Василия (ритуальное сожжение)
- 7) Восстанавливается связь Марии Гарсии с Ив Шерр (она прилетает за ним на вертолете)
- 8) Восстанавливается связь Марии Гарсии и Ив Шерр с Василием (они прилетает за ним на вертолете)
- 9) Удаляется вершина Верховного Шамана – пропадает его связь с Остальными аборигенами



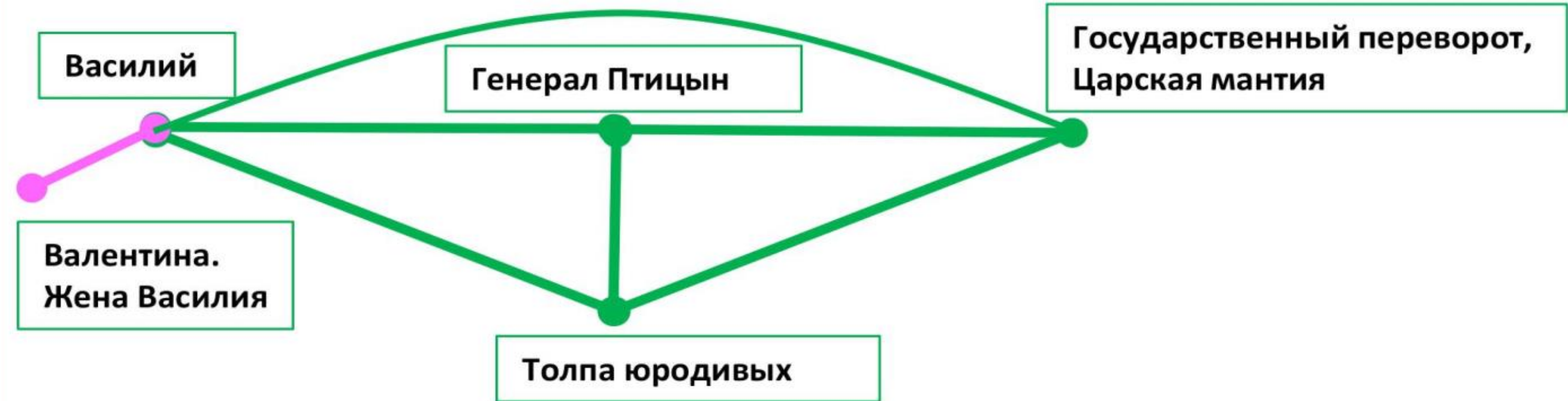
Эволюция социального графа «Смешарики: Легенда о золотом Драконе»:

- 1) Разрушается связи Нюши и Совуньи со Смешариками (Смешарики разделяются: Нюша и Совунья ищут гусеницу, Смешарики идут за телом Бараша)
- 2) Разрушается связь между Барашем и Смешариками (Они разделяются когда прыгают с парашютами, далее их разделяет Верховный Шаман)
- 3) Аборигены-хамелеоны это одна группа, а два расхитителя гробниц – другая обособленная группа. Обе группы объединены территориальным признаком.
- 4) Совершаются 3 безуспешные попытки удаления вершины - убийства Бараша
- 5) Удаляется вершина с расхитителями гробниц (их машина взрывается)
- 6) Совершаются безуспешная попытка удаления вершины Смешариков (ритуальное сожжение)
- 7) Восстанавливается связь Бараша со Смешариками (Бараш открывает дверь клетки)
- 8) Восстанавливается связь Нюши и Совуньи со Смешариками (Нюша и Совунья прилетают за Смешариками на вертолете)
- 9) Под угрозой удаления вершины аборигенов и Верховного Шамана – но все заканчивается хеппи-эндом.



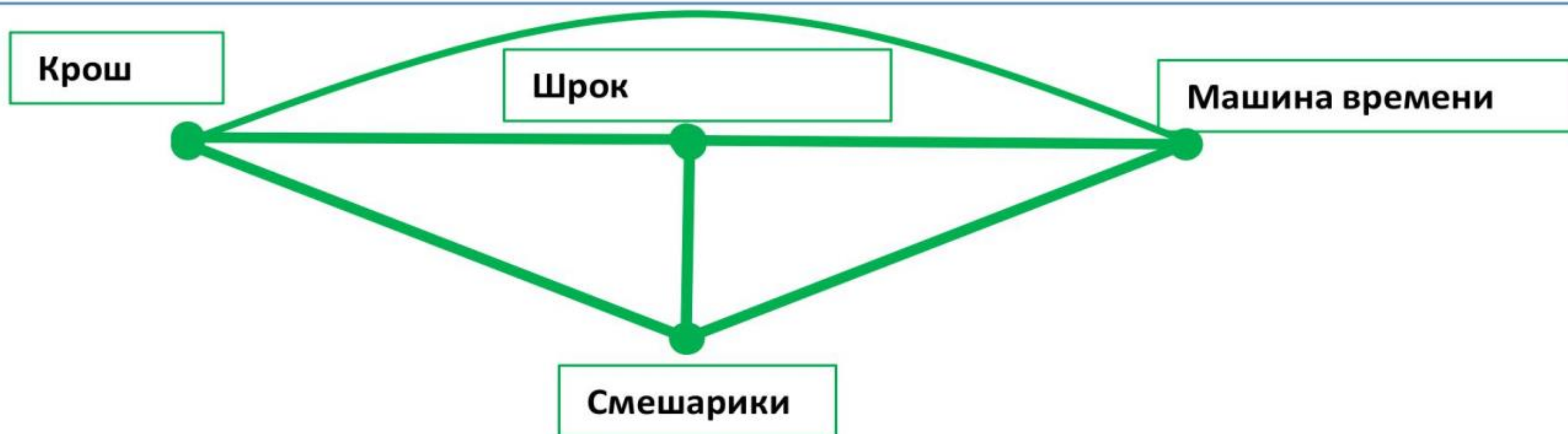
Граф сна Василия из Эпилога «И воздастся тебе...»:

- 1) У Василия обнаруживается новая связь – Генерал Птицын.
- 2) Генерал Птицын соединяет Василия с Царской мантией, а потом выводит к толпе юродивых
- 3) Умирает генерал Птицын – пропадает его вершина со всеми связями.
- 4) Пропадает Царская мантия – пропадает ее вершина со всеми связями
- 5) Василий просыпается – пропадает его связь с Юродивыми
- 6) Василия просыпается – разрушается социальный граф сна и восстанавливается социальный граф «И воздастся тебе»



Граф путешествий Кроша во времени в «Смешарики: ДежаВю»:

- 1) У Кроша обнаруживается новая связь – Шрок.
- 2) Шрок помогает Крошу вновь объединиться со Смешариками.
- 3) Умирает Шрок – пропадает его вершина со всеми связями.
- 4) Уничтожается Машина Времени – пропадает ее вершина со всеми связями
- 5) Крош попадает обратно в свою не изменённую реальность – альтернативные Смешарики пропадают
- 6) Восстанавливаются старые Смешарики – разрушается социальный граф альтернативной реальности и восстанавливается социальный граф «Смешариков: Начало»



Расчет количества уникальных сюжетов литературных произведений

Сюжет – это эволюция социального/транспортного графа произведения

Количество уникальных сюжетных сцен рассчитывается как сумма сочетаний:

$$N1 = \sum_{k=1}^n C\binom{k}{n} = \left(\frac{n!}{1!(n-1)!}\right) + \left(\frac{n!}{2!(n-2)!}\right) + \dots + \left(\frac{n!}{n!(n-n)!}\right) - IG(p)$$

где $n=(p^2-p)/2$, p – количество персонажей (для $p=500$, $N1>10^5$), IG – количество изоморфных (конформных) инвариантов графа из p вершин ($IG \ll C\binom{k}{n}$)

Так как сюжет в свою очередь это последовательность сюжетных сцен (статичных социальных графов), то общее количество сюжетов является сочетанием с повторениями и рассчитывается по формуле:

$$N2 = C\binom{s}{N1} = \left(\frac{N1!}{s! (N1 - 1)!}\right)$$

где $N1$ – количество уникальных сюжетных сцен, а s – количество сцен в произведении.

Пример графового представления сочетания основных концептов научных произведений

1									2									3									4									5													
	1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8	
1	0	1	0	1	0	0	0	0		1	0	1	0	1	0	0	1	1		1	0	0	1	1	0	1	0	0		1	0	1	0	1	0	0	0		1	0	0	1	1	0	0	0	0		
2	1	0	0	1	0	0	0	0		2	1	0	0	1	0	0	1	1		2	0	0	0	0	0	0	0	0		2	1	0	0	1	0	1	0	0		2	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0		3	0	0	0	0	0	0	0	0		3	1	0	0	1	0	1	0	0		3	0	0	0	0	0	0	0	0		3	1	0	0	1	1	0	0	0	
4	1	1	0	0	0	0	0	0		4	1	1	0	0	0	0	0	1	1		4	1	0	1	0	0	1	0	0		4	1	1	0	0	0	1	0	0		4	1	0	1	0	1	0	0	0
5	0	0	0	0	0	0	0	0		5	0	0	0	0	0	0	0	0	0		5	0	0	0	0	0	0	0	0		5	0	0	0	0	0	0	0	0		5	1	0	1	1	0	0	0	0
6	0	0	0	0	0	0	0	0		6	0	0	0	0	0	0	0	0	0		6	1	0	1	1	0	0	0	0		6	1	1	0	1	0	0	0	0		6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0		7	1	0	0	1	0	0	0	0	1		7	0	0	0	0	0	0	0	0		7	0	0	0	0	0	0	0	0		7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0		8	0	0	0	0	0	0	0	0	0		8	0	0	0	0	0	0	0	0		8	0	0	0	0	0	0	0	0		8	0	0	0	0	0	0	0	0
6									7									8									9									10													
	1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8	
1	0	0	1	1	1	0	1	1		1	0	0	0	0	0	0	0	0	0		1	0	0	1	1	1	0	1	1		1	0	0	1	1	1	0	1	1		1	0	0	1	1	1	0	1	1
2	0	0	0	0	0	0	0	0		2	0	0	0	0	0	0	0	0	0		2	0	0	0	0	0	0	0	0		2	0	0	0	0	0	0	0	0		2	0	0	0	0	0	0	0	0
3	1	0	0	1	1	0	1	1		3	0	0	0	1	1	1	1	1	1		3	1	0	0	1	1	0	1	0		3	1	0	0	1	1	0	1	1		3	1	0	0	1	1	0	1	1
4	1	0	1	0	1	0	1	1		4	0	0	1	0	1	1	1	1	1		4	1	0	1	0	1	0	1	0		4	1	0	1	0	1	1	0	1	1		4	1	0	1	0	1	1	1
5	1	0	1	1	0	0	1	1		5	0	0	1	1	0	1	1	1	1		5	1	0	1	1	0	0	1	0		5	1	0	1	1	0	0	1	1		5	1	0	1	1	0	0	1	1
6	0	0	0	0	0	0	0	0		6	0	0	1	1	1	0	1	1	1		6	0	0	0	0	0	0	0	0		6	0	0	1	1	1	0	0	0		6	0	0	0	0	0	0	0	0
7	1	0	1	1	1	0	0	1		7	0	0	1	1	1	1	0	1	1		7	1	0	1	1	1	0	0	0		7	1	0	1	1	1	0	0	1		7	1	0	1	1	1	0	0	1
8	1	0	1	1	1	0	1	0		8	0	0	1	1	1	1	1	1	0		8	0	0	0	0	0	0	0	0		8	0	0	0	0	0	0	0	0		8	1	0	1	1	1	0	1	0
11									12									13									14									1 - 14													
	1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8			1	2	3	4	5	6	7	8	
1	0	1	0	1	1	1	1	1		1	0	1	0	1	1	1	0	0		1	0	0	1	1	1	1	1	1		1	0	5	6	11	8	6	7	6		1	0	5	6	11	8	6	7	6	
2	1	0	0	1	1	1	0	1		2	1	0	0	1	1	1	0	0		2	0	0	0	0	0	0	0	0		2	5	0	0	5	2	3	1	2		2	5	0	0	5	2	3	1	2	
3	0	0	0	0	0	0	0	0		3	0	0	0	0	0	0	0	0		3	1	0	0	1	1	1	1	1		3	6	0	0	8	7	4	5	4		3	6	0	0	8	7	4	5	4	
4	1	1	0	0	1	1	1	1		4	1	1	0	0	1	1	0	0		4	1	0	1	0	1	1	1	1		4	11	5	8	0	9	7	7	6		4	11	5	8	0	9	7	7	6	
5	1	1	0	1	0	1	1	1		5	1	1	0	1	0	1	0	0		5	1	0	1	1	0	1	1	1		5	8	2	7	9	0	6	7	6		5	8	2	7	9	0	6	7	6	
6	1	1	0	1	1	0	1	1		6	1	1	0	1	1	0	0	0		6	1	0	1	1	1	0	1	1		6	6	3	4	7	6	0	4	4		6	6	3	4	7	6	0	4	4	
7	1	1	0	1	1	1	0	1		7	0	0	0	0	0	0	0	0		7	1	0	1	1	1	1	0	1		7	7	1	5	7	7	4	0	7		7	7	1	5	7	7	4	0	7	
8	1	1	0	1	1	1	1	0		8	0	0	0	0	0	0	0	0		8	1	0	1	1	1	1	1	0		8	6	2	4	6	6	4	7	0		8	6	2	4	6	6	4	7	0	

Сочетания концептов в научных текстах

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Свойство	Модель Эрдэша-Ранье	Модель предпочтительного присоединения Барабаши-Альберт	Модель Боллобаши-Риордан	Модель Бакли-Остгуса	Модель Гречникова	Кооперативная модель Джексона-Волински	Модель Шелинга	Модель копирования	Модель Уоттса-Строгатца	Модель Чайлз-Боргс с равновесием Нэша	Эпидемиологическая модель	Реляционно-иерархическая модель со строго ориентированным социальным графом Гарина	Модель стоячих аплодисментов Грановеттера	Модель социального тела проблемы Тихонова	Реляционно-иерархическая модель с обратной связью Гарина
1	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+
2	+	+	-	+	-	-	-	-	-	-	+	+	-	-	+
3	-	-	+	-	+	+	+	+	+	+	-	-	+	-	-
4	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
5	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+
6	-	-	+	+	-	-	+	-	-	-	+	+	+	+	+
7	-	+	-	-	-	+	+	+	+	+	+	-	+	+	+
8	-	+	-	-	-	+	+	-	-	+	+	-	+	+	+

Эмпирически обнаруженные свойства социума – концепты, в виде ограничительных свойств модели социума:

1) Социальный граф социума (социокультурное тело с общей проблемой видового выживания) представляет собой граф с количеством вершин в несколько миллиардов ($n \geq 8 \cdot 10^9$). При большой величине социальный граф связный, имеет одну гигантскую компоненту связности. Для самого большого изученного подграфа социума как соцсеть Facebook гигантская компонента составляет 99,91%;

2) Распределение вершин социального графа подчиняется степенному закону;

3) Среднее число связей между двумя случайными вершинами в социальном графе колеблется от 2 до 10, а среднее его значение от 4,4 до 6,6 что позволяет сделать предположение о маленьком диаметре социального графа: $10 \leq D \leq 10^2$;

4) Социальный граф сильно разрежен (является примером вырождения полного графа). При n -вершин, в социальном графе $k \cdot n$ -ребер, где k - константа. При этом теоретически количество ребер в полном графе равно $n \cdot (n-1)/2$;

5) Социальный граф имеет большой коэффициент кластеризации;

Теоретико-игровой подход обогатил список ограничений модели еще тремя теоретически предопределёнными, но экспериментально не обоснованными свойствами:

6) Социальный граф ориентированный;

7) Вероятность образования новой связи между двумя случайными вершинами социального графа прямо зависит от количества совпадающих клик, в которые входят эти вершины;

8) Вероятность изменения связей вершины социального графа прямо зависит от количества изменений связей в ее ближайшем социальном окружении.