

Структурные дыры и распространение информации в сети

Р. Ю. Москотина, Н. В. С. Сидоров

Киевский национальный университет имени Т. Шевченко, Украина
Corresponding author. E-mail: rmoskotina@ukr.net

Paper received 30.07.20; Accepted for publication 21.08.20.

<https://doi.org/10.31174/SEND-HS2020-237VIII41-13>

Аннотация. В статье представлено исследование, целью которого является определение того, как наличие/отсутствие структурных дыр влияет на эффективность распространения информации в сети. Оно показало, что эффективность распространения информации не возрастает с увеличением числа структурных дыр в сети. А вот удаление или добавление связей может привести к неравенству позиций агентов в сети. Это снижает эффективность распространения информации в сравнении с сетями, где агенты занимают одинаковые позиции.

Ключевые слова: распространение информации, структурные дыры, агенты, сеть.

Распространение информации является темой, актуальной для разных отраслей наук: социологии, экономики, компьютерных наук. В фокусе внимания социологии находится и вопрос о том, как структура связей в социальной сети (графе) может определять процесс распространения информации. Одним из тех, кто уделит внимание этому вопросу, был М. Грановеттер [1; 6], который предложил концепцию слабых связей. Слабые связи являются нерегулярными, в то время как сильные – это тесные, интенсивные связи. Среди примеров сильных связей можно назвать отношения между близкими друзьями, а примером слабых связей могут быть отношения между бывшими одноклассниками. Согласно Грановеттеру, слабые связи играют более важную роль в распространении информации, чем сильные. Сильные связи чаще всего являются транзитивными¹, способствуют созданию замкнутых групп, внутри которых циркулирует одна и та же информация. Слабые связи часто нетранзитивны, могут быть мостами между отдельными индивидами или группами. Благодаря этому свойству слабых связей, индивиды имеют доступ к новой для них информации. Более того, слабые связи помогают одним индивидам достичь других, социально отдаленных индивидов. А «люди с небольшим количеством слабых связей будут лишены информации из отдаленных частей социальной системы и будут ограничены периферийными новостями и мнениями близких друзей» [6, с. 202].

В концепции М. Грановеттера присутствует недостаток: дело в том, что он смешивает собственно силу связей и их транзитивность. В то же время Р. Барт [3; 4], автор теории структурных дыр, разделяет эти два момента, делая акцент именно на транзитивности связей. Он считает, что «новизна связана с закрытостью триад² независимо от силы связи» [2]. В самом простом случае наличие в сети структурной дыры предполагает отсутствие прямой связи между парой вершин так, что образуется одна или несколько открытых триад. В свою очередь, наличие открытых триад в сети приводит к появлению связей-мостов и посредников. Посредники связывают между собой индивидов и/или

группы. Также посредники могут транслировать информацию от одних индивидов и/или групп к другим индивидам и/или группам. В сетях, где много закрытых триад, структурных дыр и посредников очень мало либо же они вовсе отсутствуют, поэтому индивиды имеют доступ, преимущественно, к одной и той же информации. Р. Барт считает, что «социальный капитал существует там, где люди имеют преимущество благодаря своему положению в социальной структуре» [3, с. 351]. Таким образом, структурные дыры являются социальным капиталом, ведь они обеспечивают посредникам доступ к информации от тех индивидов и/или групп, которые они соединяют. Более того, «без посредников информация не сможет распространяться, а возможности для сотрудничества, обмена и инноваций будут утеряны» [5, с. 110].

Исходя из вышесказанного, можно предположить, что чем больше структурных дыр и, соответственно, посредников в сети, тем лучше распространяется информация. Это мы и проверим в данной статье. Таким образом, *цель* статьи заключается в том, чтобы определить, как наличие/отсутствие структурных дыр влияет на эффективность распространения информации в сети. В исследовании мы использовали линейную пороговую модель. Охарактеризуем ее более подробно.

Основные положения линейной пороговой модели изначально были сформулированы еще М. Грановеттером [7] в предложенной им пороговой модели коллективного поведения. Основная ее идея заключается в том, что агенты³ склонны принимать решения, основываясь на действиях других агентов. Впоследствии это приводит к возникновению коллективного поведения. Предполагается, что каждый агент может принимать участие в коллективном поведении (быть активным) или же не принимать в нем участие (быть неактивным). На каждого из неактивных агентов оказывается давление со стороны активных агентов. Давление выражено через процент активных агентов и изменяется в пределах от 0 до 100. Кроме того, модель предполагает, что каждый агент может противостоять этому давлению. Степень противостояния давлению со стороны других агентов называется порогом, значения которого также изменяются в пределах от 0 до 100. Порог 0 имеют инициаторы коллективного поведения.

¹ Транзитивность предполагает следующее: если между А и В и между В и С есть попарные связи, то такая связь существует между А и С.

² Транзитивные связи образуют закрытые триады, которые состоят из трех вершин и трех связей между ними. Нетранзитивные связи образуют открытые триады, состоящие из трех вершин и двух связей между ними.

³ Агенты в данном случае рассматриваются как действующие субъекты, искусственно созданные исследователем. В линейной пороговой модели агенты это также вершины сети.

А порог 100 имеют те агенты, которые ни под каким предлогом не захотят быть вовлечены в коллективное поведение [7, с. 1422]. Для значений порога, которые больше 0, но меньше 100, действует следующее условие: если давление не меньше порога, агент действует, в противном случае он бездействует. Процесс завершается тогда, когда уже ни одного из агентов не удастся вовлечь в коллективное поведение. Для лучшего понимания сути пороговой модели Грановеттера приведем пример. Пусть есть 100 агентов. Первым трем агентам присвоили значения порогов от 0 до 2 (каждому уникальное значение порога). Четвертому и пятому агенту присвоили значение порога 4. Всем остальным агентам присвоили значения порогов от 5 до 99; у каждого из них уникальное значение порога. Агент с порогом 0 – это инициатор, то есть он будет первым вовлечен в коллективное поведение. К нему присоединится агент с порогом 1. К этим двух агентам присоединится агент с порогом 2. И на этом процесс активации агентов завершается: для вовлечения четвертого агента в коллективное поведение нужно 4 агента, а всего активных агентов трое.

М. Грановеттер считает, что его модель может быть использована для анализа таких типов коллективного поведения как распространение инноваций, слухов или болезней, электоральное поведение, выбор будущей профессии, миграция [7, с. 1423-1424]. Но данная модель имеет существенный недостаток. Он заключается в том, что не совсем понятно, кто должен оказывать давление на неактивных агентов. Ведь это могут быть как те активные агенты, которые непосредственно связаны с неактивными агентами, так и те, которые связаны с ними опосредовано. С целью устранения этого недостатка Д. Уоттс [9; 10] немного модифицировал оригинальную модель М. Грановеттера. Он предложил изобразить агентов и взаимодействия между ними в виде сети и выдвинул предположение, что на неактивных агентов давление могут оказывать только их активные соседи, то есть активные агенты, которые непосредственно связаны с неактивными агентами. Наиболее «завершенную» версию линейной пороговой модели, а также соответствующее название предложил Д. Кемпе [8]. Он, также как и Д. Уоттс, представил агентов и взаимодействия между ними в виде графа. Активация агентов происходит в несколько шагов (шаг – это целое неотрицательное число). Изначально (на нулевом шаге) определены активные агенты, которые и являются агентами с нулевым порогом. Все остальные агенты неактивны, им может присваиваться порог в пределах 1 до 100^4 . На каждого неактивного агента оказывают давление его активные соседи. В невзвешенном графе давление выражено как отношение числа активных соседей агента к общему числу его соседей, умноженное на 100. Активация агента происходит тогда, когда давление не меньше значения порога агента. Процесс завершается тогда, когда уже ни одно-

го из агентов в сети не удастся активизировать. Отметим, что активный агент не может вернуться в состояние бездействия (то есть стать неактивным). Д. Кемпе [8] считает, что линейная пороговая модель может быть использована для изучения распространения идей или инноваций. Из этого заключаем, что линейная пороговая модель может быть использована для анализа распространения информации в сети.

В рамках исследования проверялись следующие гипотезы:

Гипотеза 1. Если в сети, состоящей лишь из закрытых триад, инициировать структурные дыры (а именно – удалить прямую связь между двумя агентами так, чтобы образовались открытые триады), эффективность распространения информации либо возрастет, либо не изменится.

Гипотеза 2. Если в сети, состоящей лишь из открытых триад, удалить структурные дыры (а именно – добавить связь между парой агентов, таким образом, чтобы образовалась закрытая триада), эффективность распространения информации снизится.

Гипотеза 3. Сети, состоящие лишь из закрытых триад, характеризуются меньшей эффективностью распространения информации, в сравнении с сетями, состоящими лишь из открытых триад.

Проведение исследования предполагает следующие этапы:

1) генерация 4 сетей (сеть 1, сеть 2, сеть 3, сеть 4). Все сети являлись неориентированными и невзвешенными графами и состояли из 101 агента⁵. Сеть 1 представляла собой полный граф, сгенерированный с помощью функции `make_full_graph` в R⁶ (пакет “igraph”). Сеть 2 была модифицированным вариантом сети 1, из которой удалили связь между произвольно выбранной парой агентов. Сеть 3 генерировалась с помощью функции `sample_smallworld` в R (пакет “igraph”). Каждый агент был связан с двумя другими агентами в сети. Сеть 4 представляла собой модифицированный вариант сети 3, куда добавили связь между двумя агентами таким образом, чтобы образовалась одна закрытая триада. Сеть 1 состояла лишь из закрытых триад, в ней не было структурных дыр и агентов-посредников, а в сеть 3 – лишь из открытых триад, она богата на структурные дыры и в ней все агенты являлись посредниками. Для всех сетей рассчитывался показатель транзитивности с помощью функции `transitivity` в R (пакет “igraph”). Он изменяется в пределах от 0 до 1. Если показатель транзитивности равен 0, сеть состоит лишь из открытых триад. Если показатель транзитивности равен 1, сеть состоит лишь из закрытых триад. Чем ближе показатель транзитивности к 1, тем больше в сети закрытых триад. Кроме того, для всех сетей был рассчитан показатель среднего расстояния между двумя агентами. Он рассчитывался с помощью функции

⁵ Уже было сказано, что значение порога 1 говорит об отсутствии сопротивления давлению со стороны неактивных агентов. Тогда значение давления со стороны активных агентов в полном графе должно быть не меньше 1, чтобы у активного агента, которого мы выбираем на нулевом шаге, была возможность активизировать неактивных агентов. Это достигается тогда, когда сеть состоит не больше, чем из 101 агента.

⁶ R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>

⁴ Значение порога 1 указывает на то, что агент не противодействует давлению со стороны активных агентов-соседей. Значение порога 100 указывает на максимально возможное сопротивление давлению со стороны активных агентов-соседей. Но в отличие от пороговой модели Грановеттера, в модели, которая будет использована в исследовании, даже при значении порога 100 есть возможность активизировать неактивных агентов.

mean_distance в R (пакет “igraph”). Минимальное его значение равно 1, которое указывает на то, что в сети все связи являются прямыми и в ней нет связей-мостов. Чем больше значение среднего расстояния между двумя агентами, тем больше «длина» связей-мостов в сети;

2) для каждой из сетей выбирали одного активного агента на нулевом шаге, при этом, в качестве активных, мы поочередно выбирали каждого из агентов. Также определяли значение порога, которое являлось натуральным числом, изменяющимся в пределах от 1 до 100. Всем агентам присваивали одинаковое значение порога. Подробнее о выборе оптимального значения порога см. ниже;

3) с помощью линейной пороговой модели провели эксперименты с каждой из четырех сетей. Это позволило сделать выводы об эффективности распространения информации в них. Предполагалось сравнение результатов экспериментов для: а) сетей 1 и 2 с целью проверки гипотезы 1, б) сетей 3 и 4 для проверки гипотезы 2, в) сетей 1 и 3 (проверка гипотезы 3).

В свою очередь, эффективность распространения информации в сети измерялась с помощью трех показателей:

1) среднее количество агентов, среди которых была распространена информация. На нулевом шаге мы поочередно выбираем по одному агенту в сети в качестве активного. Проведя эксперименты, получаем 101 числовое значение, обозначающие общее количество активных агентов после завершения процесса активации (без учета агента, выбранного на нулевом шаге). После этого рассчитываем среднее арифметическое этих числовых значений. Это и будет среднее количество агентов, среди которых была распространена информация. Максимальное значение этого показателя равно 100. То есть каждому из агентов, если его выбирали в качестве активного, удалось распространить информацию среди всех неактивных агентов. Минимальное значение данного показателя равно 0; то есть агенты, если их выбирали в качестве активных, не могли никого активизировать;

2) средняя скорость распространения информации. На нулевом шаге мы поочередно выбираем по одному агенту в сети в качестве активного. После проведения экспериментов получаем 101 числовое значение, обозначающие количество шагов активации агентов (без учета нулевого шага). Рассчитываем среднее арифметическое этих числовых значений. Полученный в результате показатель и будет средней скоростью распространения информации. Чем большее значение имеет этот показатель, тем медленнее распространяется информация в сети. Если значение данного показателя равно 1, мы имеем дело с максимальной средней скоростью распространения информации;

3) эффективность преодоления сопротивления со стороны неактивных агентов. Поочередно выбираем по одному агенту и для каждого из них подбираем максимальное значение порога, после которого уже не происходит активация. Таким образом, получаем 101 числовое значение и находим максимальное из них. Искомое значение и будет эффективностью преодоления сопротивления со стороны неактивных агентов для конкретной сети. Чем больше значение этого показателя, тем лучше преодолевается сопротивление со

стороны неактивных агентов. Исходя из значений порогов, которые попадают под определение эффективности преодоления сопротивления со стороны неактивных агентов, мы проводим эксперименты с сетями и находим среднюю скорость распространения информации и среднее количество агентов, среди которых была распространена информация.

Приступим к описанию результатов экспериментов с сетями. Вначале более подробно остановимся на сравнении сети 1 и сети 2. Их свойства представлены в Таблице 1. Как видим, значение показателя транзитивности в сети 2 несколько ниже, чем в сети 1, а значение среднего расстояния между двумя агентами несколько больше в сети 2, чем в сети 1⁷.

Таблица 1. Свойства сетей 1 и 2

	Сеть 1	Сеть 2
Количество агентов	101	101
Количество связей	5050	5049
Транзитивность	1	0,9998
Среднее расстояние между двумя агентами	1	1,0002

Результаты экспериментов с этими сетями представлены в Таблице 2.

Таблица 2. Результаты экспериментов с сетями 1 и 2

	Сеть 1	Сеть 2
Среднее количество агентов, среди которых была распространена информация	100	100
Средняя скорость распространения информации	1	1,02
Эффективность преодоления сопротивления со стороны неактивных агентов	1	1

Можем сделать следующие выводы. В сети 1 один активный агент может распространить информацию в среднем среди 100 остальных агентов. Средняя скорость распространения информации равна 1. По сути, любой агент, если его выбрать в качестве активного, имеет возможность распространить информацию среди остальных 100 агентов за 1 шаг. Эффективность преодоления сопротивления со стороны неактивных агентов равна 1. То есть в сети, где имеются лишь закрытые триады, один активный агент может распространять информацию в том случае, когда сопротивление со стороны неактивных агентов отсутствует.

В сети 2, так же как и в сети 1, активный агент имеет возможность распространить информацию в среднем среди 100 агентов. То есть любой агент, выбранный в качестве активного, имеет возможность распространить информацию среди остальных 100 неактивных агентов. Эффективность преодоления сопротивления со стороны неактивных агентов, так же, как и в сети 1, равна 1. Но средняя скорость распространения информации немного меньше, чем в сети 1 и составляет 1,02. Все дело в том, что удаление даже одной связи из сети 1 способствует появлению агентов-посредников. 99 агентов в сети 2 выступают посредниками между теми двумя агентами, между которыми была удалена связь.

⁷ Если числа представляли собой десятичные дроби с более, чем двумя знаками после запятой, они округлялись с точностью до сотых. Исключением является данный случай, поскольку нужно было проиллюстрировать разницу между сетью 1 и сетью 2 по показателям транзитивности и среднего расстояния между двумя агентами.

Они могут активизировать всех агентов за 1 шаг. А те агенты, между которыми была удалена связь, могут активизировать остальных агентов лишь за 2 шага. Каждый из них на первом шаге активизирует 99 агентов, с которыми есть прямые связи. А на втором шаге активизируется тот агент, с которым прямая связь была удалена. То есть удаление всего лишь одной связи из полного графа продуцирует неравенство положения агентов в сети. При этом агенты-посредники занимают более выгодное положение, в сравнении с теми агентами, между которыми была удалена прямая связь. Неравенство положения агентов в сети 2 приводит к незначительному снижению эффективности распространения информации в ней по сравнению с сетью 1. Последняя выигрывает за счет скорости распространения информации. То есть гипотеза 1 не подтвердилась.

Теперь сравним между собой сеть 3 и сеть 4, свойства которых представлены в Таблице 3. Как видим, значение показателя транзитивности в сети 4 больше, чем в сети 3. Напротив, среднее расстояние между двумя агентами в сети 4 меньше, чем в сети 3.

Таблица 3. Свойства сетей 3 и 4

	Сеть 3	Сеть 4
Количество агентов	101	101
Количество связей	101	102
Транзитивность	0	0,03
Среднее расстояние между двумя агентами	25,5	25,26

Результаты экспериментов с сетями 3 и 4 представлены в Таблице 4.

Таблица 4. Результаты экспериментов с сетями 3 и 4

	Сеть 3	Сеть 4
Среднее количество агентов, среди которых была распространена информация	100	96,10
Средняя скорость распространения информации	50	71,82
Эффективность преодоления сопротивления со стороны неактивных агентов	50	50

По результатам экспериментов можем сделать следующие выводы: в сети 3 один активный агент мог распространить информацию в среднем среди 100 агентов, средняя скорость распространения информации в сети 3 равна 50. Собственно, любой из агентов этой сети, выбранный в качестве активного, может распространить информацию среди 100 остальных агентов за 50 шагов. Эффективность преодоления сопротивления со стороны неактивных агентов достаточно высокая и равна 50.

В сети 4 среднее количество агентов, среди которых может быть распространена информация, равно 96,10; оно меньше, чем в сети 3. Также средняя скорость распространения информации в этой сети заметно меньше, чем в сети 3 (71,82 против 50). Лишь эффективность преодоления сопротивления со стороны неактивных агентов такая же, как в сети 3. То есть, можем сделать вывод о том, что эффективность распространения информации в сети 4 ниже, чем в сети 3.

Если проанализировать распространение информации в сети 4 более подробно, получаем следующее: за счет наличия замкнутой триады появляется один агент, который утрачивает возможность выступать посредником между двумя другими агентами, и он не может никого активизировать. Каждый из двух агентов, меж-

ду которыми добавили связь, может распространить информацию среди 100 остальных агентов за 50 шагов. Остальным агентам удается распространить информацию среди 97 агентов; они не имеют возможности распространить информацию среди агентов, составляющих закрытую триаду. Сделать это они могут минимум за 49, а максимум за 97 шагов. Причем чем ближе агент к замкнутой триаде, тем больше шагов ему потребуется, чтобы достичь этих 97 агентов. Таким образом, два агента, между которыми мы добавили связь, занимают самое выгодное положение в сети. Также достаточно выгодное положение в сети занимают те, кто максимально удален от замкнутой триады (то есть те агенты, которые распространяют информацию за 49 шагов). Самое невыгодное положение в сети занимает агент, который перестал быть посредником и находится в замкнутой триаде. Опять-таки, неравенство положения агентов в сети снижает эффективность распространения информации. Гипотеза 2 подтвердилась.

Если сравнивать эффективность распространения информации в сети 3 и в сети 1 (см. Табл. 2 и Табл. 4), можем заметить следующее: и в сети 1, и в сети 3 любой агент, если его выбрать в качестве активного, может распространить информацию среди остальных 100 агентов. Но сеть 3 проигрывает сети 1 в скорости распространения информации. Ведь в сети 3 каждый из агентов является посредником лишь между двумя агентами. Это значит, что первый агент может распространить информацию максимум среди двух агентов и передать им «эстафету» для дальнейшего распространения информации. Отсюда меньшая средняя скорость распространения информации, чем в том случае, когда каждый агент прямо связан с другими агентами в сети. А сеть 1 проигрывает сети 3 в плане эффективности преодоления сопротивления со стороны неактивных агентов. Ведь чем меньше связей имеет каждый агент, тем больше «вес» его мнения и действий. Таким образом, гипотеза 3 не подтвердилась. Каждая из сетей – и та, которая состоит лишь из закрытых триад, и та, которая состоит лишь из открытых триад – имеет свои преимущества и недостатки, если говорить об эффективности распространения информации.

Выводы. Как показало данное исследование, эффективность распространения информации не возрастает с увеличением числа структурных дыр в сети. Дело в том, что удаление или добавление прямых связей приводит к тому, что агенты начинают занимать неодинаковые позиции в сети⁸. Одни агенты получают больше преимуществ в плане возможностей распространения информации, чем другие. А это, в свою очередь, приводит к снижению эффективности распространения информации в сравнении с сетями, где агенты занимают одинаковые позиции. Если же сравнивать между собой сети, где все агенты занимают одинаковое положение, мы не можем сказать о том,

⁸ При условии, если до удаления/добавления связей все агенты занимали одинаковое положение в сети. Неодинаковые позиции в сети означают, что в сети есть как агенты-посредники, так и агенты, не являющиеся посредниками. Если же все агенты занимают одинаковое положение в сети, это означает, что либо все агенты являются посредниками (как в сети 3), либо все агенты не посредники (как в сети 1).

что в одной из них эффективность распространения информации выше, чем в другой. Сеть, где много структурных дыр и все агенты являются посредниками, проигрывает сети, где нет структурных дыр и посредников, в плане средней скорости распространения ин-

формации. В свою очередь, сеть, где нет структурных дыр и посредников, характеризуется меньшей эффективностью преодоления сопротивления со стороны неактивных агентов в сравнении с сетью, где много структурных дыр и все агенты посредники.

ЛИТЕРАТУРА

1. Грановеттер М. Сила слабых связей // Экономическая социология, 2009. Т. 10, № 4. С. 31-50.
2. Brashears M., Quintane E. The Weakness of Tie Strength // Social Networks, 2018. Vol. 55. P. 104-115. URL: <https://www.sciencedirect.com/science/article/pii/S0378873318300194>
3. Burt R. Structural Holes and Good Ideas // American Journal of Sociology, 2004. Vol. 110, No. 2, P. 349-399.
4. Burt R. Structural Holes: The Social Structure of Competition. Cambridge, Mass: Harvard University Press, 1992. – 313 p.
5. Centola D., Cook K., Hedström P. How Behavior Spreads: The Science of Complex Contagions. Princeton and Oxford: Princeton University Press, 2018. – 312 p.
6. Granovetter M. The Strength of Weak Ties: A Network Theory Revisited // Sociological Theory, 1983. Vol. 1. P. 201-233.
7. Granovetter M. Threshold Models of Collective Behavior // American Journal of Sociology, 1978. Vol. 83, No. 6. P. 1420-1443.
8. Kempe D., Kleinberg J., Tardos E. Maximizing the Spread of Influence through a Social Network // KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (August, 2003). P. 137-146. URL: <https://www.cs.cornell.edu/home/kleinber/kdd03-inf.pdf>
9. Watts D. A Simple Model of Global Cascades on Random Networks // Proceedings of the National Academy of Sciences of the United States of America, 2002. Vol. 99, No. 9. P. 5766-5771.
10. Watts D., Dodds P. Threshold Models of Social Influence // The Oxford Handbook of Analytical Sociology / Ed. by P. Hedström, P. Bearman. Oxford: Oxford University Press, 2009. P. 475-497.

REFERENCES

1. Granovetter, M. The Strength of Weak Ties // Economic sociology, 2009. Vol. 10. No. 4. P. 31-50. [In Russian]

Structural holes and information spread in the network

R. Yu. Moskotina, M. V. S. Sydorov

Abstract. The article presents a study in order to find out how the presence/absence of structural holes in the network influence to the effectiveness of information spread. The study shows that the effectiveness of information spread does not increase with increasing of a number of structural holes in the network. But deleting or adding ties may produce an inequality in the positions of agents in the network. It reduces the effectiveness of information spread in comparison with networks where agents occupy the same positions.

Keywords: information spread, structural holes, agents, network.