

УДК 519.24; 519.25

Зависимости и кластерность в эволюционирующих сетях

Н.М. Маркович*Институт проблем управления им. В.А. Трапезникова РАН*

Россия, 117997, Москва, Профсоюзная ул., 65

E-mail: markovic@ipu.ru

М.С. Рыжов*Институт проблем управления им. В.А. Трапезникова РАН*

Россия, 117997, Москва, Профсоюзная ул., 65

E-mail: maksim.ryzhov@frtk.ru

М.Р. Кулик*МИФИ*

Россия, 115409, Москва, Каширское ш., д. 31

E-mail: mishakulik2002@yandex.ru

Ключевые слова: Кластерный коэффициент, эволюция сети, предпочтительное и кластерное присоединения, число связей узла, число треугольников узла.

Аннотация: Изучается эволюция сети моделями предпочтительного, кластерного и смешанного присоединений для формирования связей новых узлов с существующими узлами. Моделированием на примерах сравнивается зависимость числа связей узлов и числа треугольников, т.е. троек связанных узлов, и поведение кластерных коэффициентов узлов. Оценивается тяжесть хвоста распределения для числа связей и треугольников. Смешанное присоединение предлагается впервые.

1. Введение

Рассматривается эволюция ненаправленных графов. Пусть такой граф на шаге эволюции t будет $G_t = (V_t, E_t)$, V_t – множество вершин, E_t – множество ребер, $\|\cdot\|$ – мощность множества. Удаление существующих узлов или связей во время эволюции не предполагается. Изучается зависимость трех характеристик случайных сетей: числа связей узлов, числа треугольников, т.е. числа троек связанных узлов, и кластерных коэффициентов узлов. Моделированием на примерах сравниваются зависимости при трех моделях эволюции сети: предпочтительном (ПП), кластерном (КП) и смешанном кластерно-предпочтительном (КПП) присоединениях для формирования связей новых узлов с существующими узлами. КПП предлагается впервые. ПП моделирует присоединение новых, вновь появившихся в сети узлов, с большей вероятностью к наиболее влиятельным узлам, так как вероятность

присоединения пропорциональна числу связей узла. Новый узел присоединяется случайно к существующему узлу i ненаправленного графа с вероятностью

$$(1) \quad P_{PA}(i, t) = d_{i,t} / \sum_{s \in V_t} d_{s,t},$$

$d_{i,t}$ – число связей узла $i \in V_t$. ПП был предложен для обоснования тяжелых хвостов распределений характеристик влиятельности узлов таких, как число их связей и Пейджранги. ПП выражается в стремительном росте связей небольшого числа узлов.

КП предложен в [1] для моделирования локальных сетей, когда присоединение нового узла происходит лишь к узлам, вовлеченным в треугольники. Например, в социальных сетях люди могут иметь друзей не среди популярных людей, а в своем окружении, формируя тесные сообщества. Вместо гигантских узлов с большими $d_{i,t}$, как для ПП, КП ведет к сообществам, порождаемым новыми узлами. В [1] показано без доказательства, что КП порождает легкие хвосты распределений характеристик влиятельности узлов. КП моделирует эволюцию социальных сетей и сетей, где новые узлы присоединяются не к большим узлам, а к сильно связанным сообществам. Вероятность присоединения нового узла к узлу $i \in V_t$ определяется как [2]

$$(2) \quad P_{CA}(i, t) = \left(c_{i,t}^\alpha + \epsilon \right) / \left(\sum_{s \in V_t} c_{s,t}^\alpha + \epsilon \|V_t\| \right).$$

$$c_{i,t} = \begin{cases} 0, & d_{i,t} = 0 \text{ или } d_{i,t} = 1, \\ 2\Delta_{i,t} / (d_{i,t}(d_{i,t} - 1)), & d_{i,t} \geq 2, \end{cases}$$

где $c_{i,t}$ – кластерный коэффициент, $\Delta_{i,t}$ – число треугольников, в которые вовлечен узел i в момент t , ϵ – некоторая постоянная вероятность присоединения, которая может быть нулевой, $\alpha \geq 0$ – параметр модели.

Предложим модель КПП, в которой новый узел присоединяется к существующим узлам с помощью ПП и КП, что больше соответствует реальности. Примером служит общение коллег через их сайты и присоединение к сайтам известных ученых. Вероятность присоединения нового узла к узлу $i \in V_t$ определяется как

$$(3) \quad P_{CPA}(i, t) = \begin{cases} P_{PA}(i, t), & \text{с вероятностью } p, \\ P_{CA}(i, t), & \text{с вероятностью } 1 - p. \end{cases}$$

Кроме того, изучим поведение среднего кластерного коэффициента узлов сети:

$$(4) \quad \bar{C}_t = \frac{1}{\|V_t\|} \sum_{i \in V_t} c_{i,t}.$$

для ПП, КП и КПП. \bar{C}_t отражает тенденцию узлов формировать треугольники.

Секции 2. и 3. содержат описание эксперимента и результаты моделирования для сравнения зависимости числа треугольников и связей узлов, поведения среднего кластерного коэффициента для ПП, КП и КПП. Секция 4. содержит выводы.

2. Описание экспериментальных данных

В качестве начального графа, с которого начинается эволюция с ПП, КП и КПП, выберем треугольник связанных узлов. На каждом шаге эволюции t новый

узел соединяется двумя связями с узлами из V_t , которые выбираются с вероятностью соответствующей модели. Для КПП один из узлов может выбираться ПП, а второй КП. КП берется с параметрами $(\alpha, \epsilon) = \{(0, 0), (1, 0), (1, 1)\}$. ПП не содержит параметров по (1). КПП использует параметры КП. Моделируются по 20 графов с КП, ПП, КПП для дальнейшего усреднения по всем графам среднего кластерного коэффициента \bar{C}_t (4):

$$(5) \quad \tilde{C}_t = \sum_{j=1}^{20} \bar{C}_t^{(j)}, \quad t \in \{0, 10^4\}.$$

Для распределений числа треугольников и связей узлов оцениваются индексы экстремальных значений оценками моментов, УН [3] и mixed moment [4].

3. Основные результаты

На Рис. 1а, 1с, 1е, 1г показана степенная зависимость числа треугольников $\{\Delta_{i,t}\}$ от числа связей $\{d_{i,t}\}$. При $p = 0.8$ в (3) эта зависимость близка для ПП и КПП, при $p = 0.2$ - для КП и КПП. На Рис. 1б, 1д, 1ф, 1h показаны убывающие последовательности усредненных кластерных коэффициентов \tilde{C}_t , $t \geq 0$. Для КПП и КП параметры $(\alpha, \epsilon) = \{(0, 0), (1, 1)\}$ приводят к \tilde{C}_t , близкому со случаем ПП, поскольку присоединение к новым узлам происходит практически независимо от наличия треугольников. Для $(\alpha, \epsilon) = (1, 0)$ \tilde{C}_t убывает быстрее остальных случаев.

В Таб. 1 для проверки тяжести хвоста распределений приводятся оценки индекса экстремальной величины $\hat{\gamma}$ для $\{\Delta_{i,t}\}$ и $\{d_{i,t}\}$. Для ПП $\hat{\gamma} > 0$, что соответствует тяжелым хвостам распределений для $\{\Delta_{i,t}\}$ и $\{d_{i,t}\}$, с более тяжелыми для $\{\Delta_{i,t}\}$. Для КПП и КП $\hat{\gamma}$ могут быть отрицательными или положительными близкими к 0 для $\{d_{i,t}\}$, что означает легкие хвосты распределений. Для КПП и КП распределения $\{\Delta_{i,t}\}$ имеют тяжелые хвосты при $\epsilon = 0$. При $\epsilon = 0$ новый узел присоединяется к узлам с $c_{i,t} > 0$. Треугольники возникают у малой части узлов, что влияет на тяжесть хвоста распределения.

Оценим меру зависимости $\{\Delta_{i,t}\}$ и $\{d_{i,t}\}$ с помощью корреляции расстояний (distance correlation). В Таб.2 показано, что имеется зависимость для моделей КП и всех видов моделей КПП для $((\alpha, \epsilon)) = (0,0), (1,0)$, максимальное значение наблюдается в случае КПП $p = 0.2$ для $((\alpha, \epsilon)) = (1,0)$, коэффициент корреляции расстояний оказался равен 0.923, что показывает наличие зависимости в данных. Обратный эффект наблюдается у оставшихся моделей. Видна тенденция: при переходе от КП к КПП и от КПП $p = 0.2$ к КПП $p = 0.8$ видно, что dCor для $((\alpha, \epsilon)) = (1,1)$ растет, однако каждое значение заметно меньше, чем любое из описанных выше, что говорит о малой зависимости в данных.

4. Заключение

Предложена и исследована смешанная модель эволюции случайной сети, когда новый узел присоединяется к существующим узлам с вероятностью, определяемой КП или ПП. Главный результат в том, что примешивание КП к ПП приводит к утере тяжелых хвостов распределений для числа связей, свойственных ПП. Число треугольников при всех моделях присоединения имеет тяжелый хвост.

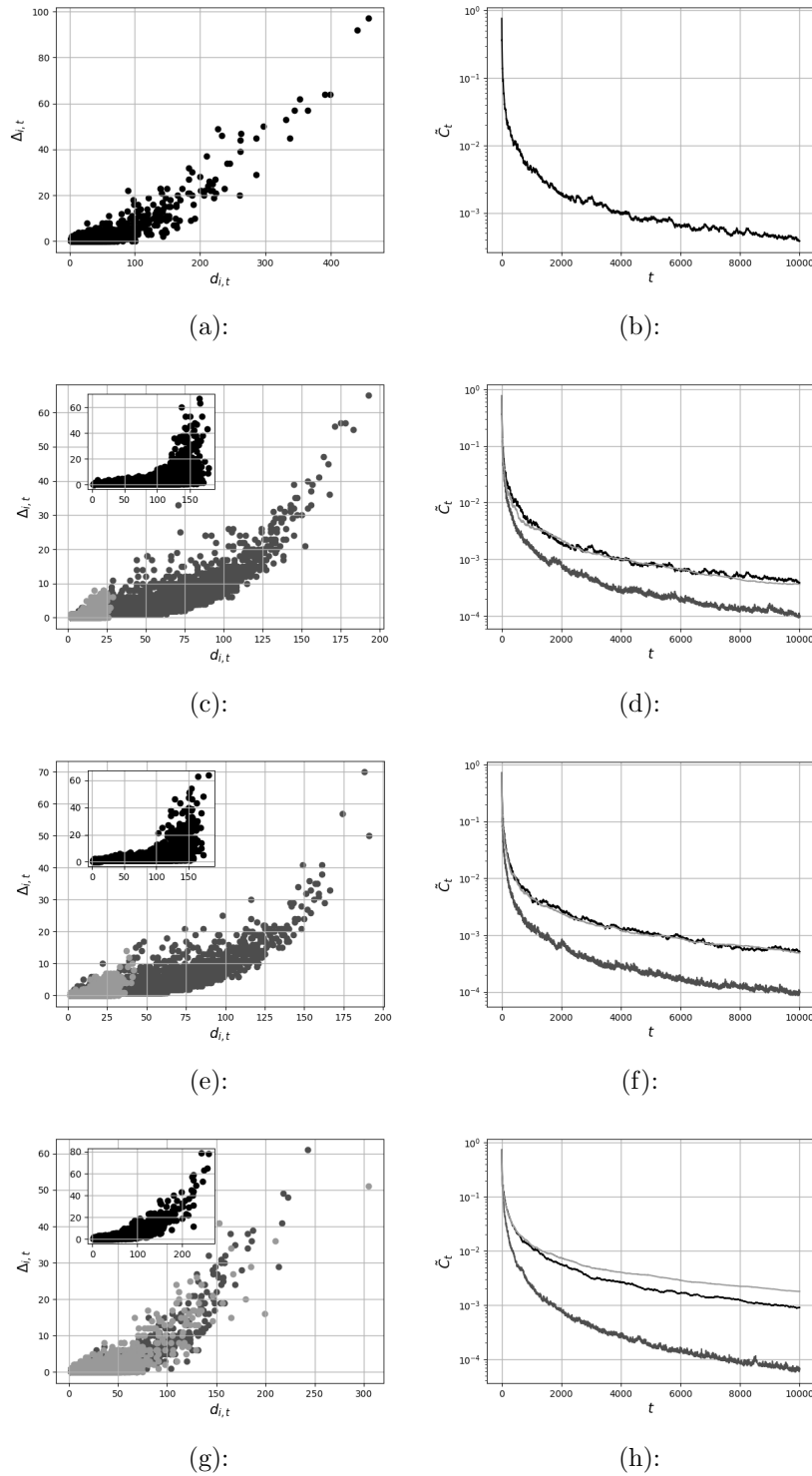


Рис. 1: Зависимость числа треугольников $\{\Delta_{i,t}\}$ от числа связей $\{d_{i,t}\}$, $i \geq 1$ для узлов графа после $t = 10^4$ шагов эволюции ПП (Рис. 1а), КП (Рис. 1с), КПП для $p = 0.2$ (Рис. 1е) и $p = 0.8$ (Рис. 1г); эволюция усредненного по 20 графам среднего кластерного коэффициента \tilde{C}_t , см. (5) для соответствующих методов присоединения (правый столбец). Для КП и КПП взяты параметры $(\alpha, \epsilon) \in \{(0, 0), (1, 0), (1, 1)\}$ (черные, темно-серые, серые точки и линии, соответственно).

Таблица 1: Оценки индекса экстремальной величины с 95% доверительными бутстреп-интервалами для числа связей $\{d_{i,t}\}$ и треугольников $\{\Delta_{i,t}\}$.

Модель эволюции	(α, ϵ)	Оценка mixed moment		Оценка моментов		Оценка УН	
		$\{d_{i,t}\}$	$\{\Delta_{i,t}\}$	$\{d_{i,t}\}$	$\{\Delta_{i,t}\}$	$\{d_{i,t}\}$	$\{\Delta_{i,t}\}$
ПП		0.192 (0.098, 0.84)	0.38 (0.529, 1.478)	0.103 (0.026, 0.68)	0.331 (0.167, 1.095)	0.064 (0.03, 0.578)	0.271 (0.151, 0.914)
КП	(0, 0)	-0.204 (-0.835, -0.036)	0.243 (0.026, 0.337)	-0.263 (-0.909, -0.079)	0.17 (0.063, 0.785)	-0.236 (-0.84, -0.188)	0.148 (0.124, 0.63)
	(1, 0)	-0.089 (-0.537, 0.052)	0.162 (0.074, 0.77)	-0.126 (-1.01, -0.109)	0.192 (0.176, 0.597)	-0.076 (-0.377, 0.282)	0.141 (0.136, 0.53)
	(1, 1)	0.078 (-0.106, 0.198)	-0.232 (-0.982, -0.176)	0.106 (-0.626, 0.244)	-0.028 (-0.418, 0.069)	-0.133 (-0.884, -0.122)	-0.04 (-0.462, 0.008)
КПП с $p = 0.2$	(0, 0)	-0.237 (-0.88, 0.053)	0.241 (-0.089, 0.814)	-0.318 (-0.807, 0.059)	0.121 (0.089, 0.748)	-0.255 (-0.889, -0.167)	0.135 (0.125, 0.637)
	(1, 0)	-0.066 (-0.407, 0.032)	0.195 (0.124, 0.686)	-0.07 (-0.096, 0.304)	0.245 (-0.008, 0.63)	-0.114 (-0.554, -0.03)	0.147 (0.034, 0.536)
	(1, 1)	-0.0178 (-0.064, 0.209)	0.07 (-0.053, 0.279)	0.054 (0.01, 0.199)	0.374 (0.123, 0.742)	-0.001 (-0.625, 0.353)	0.03 (-0.572, 0.87)
КПП с $p = 0.8$	(0, 0)	0.03 (-0.273, 0.05)	0.363 (0.244, 1.298)	-0.049 (-0.067, 0.349)	0.39 (0.244, 0.928)	-0.097 (-0.409, -0.075)	0.309 (0.083, 0.769)
	(1, 0)	-0.026 (-0.316, 0.084)	0.15 (0.074, 0.265)	-0.149 (-0.777, -0.111)	0.358 (0.22, 0.847)	-0.057 (-0.463, 0.088)	0.267 (-0.094, 0.336)
	(1, 1)	0.046 (-0.174, 0.078)	0.097 (0.091, 0.345)	-0.094 (-0.133, 0.278)	0.214 (0.175, 0.336)	-0.014 (-0.502, 0.474)	0.213 (0.18, 0.344)

Таблица 2: Значения индекса корреляции расстояний с 95% доверительными интервалами для числа связей $\{d_{i,t}\}$ и треугольников $\{\Delta_{i,t}\}$.

Модель	ПП	КП (α, ϵ)			КПП $p = 0.2$ (α, ϵ)		
		(0,0)	(1,0)	(1,1)	(0,0)	(1,0)	(1,1)
dCor	0.404 (0.395, 0.413)	0.896 (0.894, 0.898)	0.884 (0.806, 0.962)	0.192 (0.185, 0.199)	0.898 (0.896, 0.9)	0.923 (0.919, 0.927)	0.222 (0.214, 0.23)
Модель		КПП $p = 0.5$ (α, ϵ)			КПП $p = 0.8$ (α, ϵ)		
dCor		(0,0)	(1,0)	(1,1)	(0,0)	(1,0)	(1,1)
		0.874 (0.87, 0.878)	0.904 (0.9, 0.908)	0.276 (0.271, 0.281)	0.72 (0.715, 0.725)	0.779 (0.774, 0.784)	0.351 (0.344, 0.358)

Исследование выполнено Маркович Н.М. и Рыжовым М.С. за счет гранта Российского научного фонда №22-21-00177.

Список литературы

1. Bagrow J., Brockmann D. Natural Emergence of Clusters and Bursts in Network Evolution // Physical Review X. 2012. Vol. 3, No. 2. P. 021016.
2. Markovich N.M., Vaičiulis M. Extreme Value Statistics for Evolving Random Networks // Mathematics. 2023. Vol. 11, No. 9. P. 2171; <https://doi.org/10.3390/math11092171>.
3. Markovich N.M. Nonparametric Analysis of Univariate Heavy-Tailed data: Research and Practice. Chichester, West Sussex: Wiley, 2007.
4. Fraga Alves M.I., Gomes M.I., de Haan L. et al. Mixed moment estimator and location invariant alternatives. Extremes. 2009. Vol. 12. P. 149–185.