

Сравнение результатов вычислений моего кода и известных мне статей (анализ тестов)

Владимир Ивашкин

18 марта 2018 г.

1 Введение

Захотелось написать большой развернутый отчет о всех тестах, которые я провожу. Возможно, это поможет найти наши ошибки и в будущем быть уверенными в результатах.

2 Chebotarev: Studying new classes of graph metrics

Ссылка: <https://arxiv.org/abs/1305.7514>

Здесь нам интересен Fig. 1. На графе "цепочка" можно прогнать такие же метрики при тех же параметрах. В обозначениях ниже я имею в виду, что вершины графа названы слева направо цифрами от 0 до 3. Важно:

- это расстояния, не близости
- расстояния здесь нормированы на то, чтобы сумма $D[0, 1] + D[1, 2] + D[2, 3] = 3$

Достаточно будет сравнивать расстояния $D[0, 1]$, $D[1, 2]$, $D[0, 2]$, $D[0, 3]$. Будем считать, что расстояния не соответствуют друг другу, если хотя бы одна соответствующая пара расстояний различается на 0.04 в абсолютной величине. True — результат из статьи, Test — посчитанные результаты, Diff — абсолютная разница. Итак, вот результаты:

		$D[0, 1]$	$D[1, 2]$	$D[0, 2]$	$D[0, 3]$	Passed
Shortest path, Resistance distance	True	1,000	1,000	2,000	3,000	TRUE
	Test	1,000	1,000	2,000	3,000	
	Diff	0,000	0,000	0,000	0,000	
Walk, $\alpha = 1$	True	1,025	0,950	1,975	3,000	FALSE
	Test	1,013	0,975	1,406	1,732	
	Diff	0,012	0,025	0,569	1,268	
log Forest, $\alpha = 2$	True	0,959	1,081	2,040	3,000	FALSE
	Test	0,980	1,040	1,429	1,733	
	Diff	0,021	0,041	0,611	1,267	
Forest, $\alpha = 1$	True	1,026	0,947	1,500	1,895	FALSE
	Test	1,013	0,973	1,225	1,377	
	Diff	0,013	0,026	0,275	0,518	
Square-rooted resistance	True	1,000	1,000	1,414	1,732	TRUE
	Test	1,000	1,000	1,414	1,732	
	Diff	0,000	0,000	0,000	0,000	
Communicability (не указан параметр!)	True	0,964	1,072	1,492	1,564	TRUE
	Test	0,964	1,072	1,492	1,564	
	Diff	0,000	0,000	0,000	0,000	
plain Walk $\alpha = 4.5$	True	1,025	0,950	1,541	1,466	FALSE
	Test	1,013	0,975	1,242	1,211	
	Diff	0,012	0,025	0,299	0,255	
plain Walk, $\alpha = 1$	True	0,988	1,025	1,379	1,416	FALSE
	Test	0,994	1,012	1,174	1,190	
	Diff	0,006	0,013	0,205	0,226	

Таблица 1: Результаты воспроизведения результатов из Fig. 1

Итоги более коротко: результаты совпали для Shortest path, Resistance, Square-rooted resistance, Communicability, и не совпали для Walk, plain Walk, Forest, log Forest.

Давайте рассмотрим процесс получения матрицы расстояний на примере log Forest:

1. Считаем $H0 = (I + tL)^{-1}$. t — параметр метрики, в данном случае передается напрямую.
2. Поэлементно логарифмируем $H = \text{element-wise log}(H0)$
3. Превращаем ядро в матрицу расстояний с помощью $D = (h * 1^T + 1 * h^T - H - H^T)/2$, h — диагональ H .

Исследуя разные гипотезы и перепроверяя все что только можно, я наткнулся на следующий интересный факт: ответы начинают совпадать, если конечную матрицу расстояний возвести в квадрат. Причем не для всех метрик, а только для тех, результаты для которых не совпадали раньше. Это странно. В чем здесь может быть дело?

С корректировками таблица выглядит так:

		$D[0, 1]$	$D[1, 2]$	$D[0, 2]$	$D[0, 3]$	Passed
Shortest path, Resistance distance	True	1,000	1,000	2,000	3,000	TRUE
	Test	1,000	1,000	2,000	3,000	
	Diff	0,000	0,000	0,000	0,000	
Walk, $\alpha = 1$	True	1,025	0,950	1,975	3,000	TRUE
	Test	1,025	0,950	1,975	3,000	
	Diff	0,000	0,000	0,000	0,000	
log Forest, $\alpha = 2$	True	0,959	1,081	2,040	3,000	TRUE
	Test	0,959	1,081	2,041	3,000	
	Diff	0,000	0,000	0,001	0,000	
Forest, $\alpha = 1$	True	1,026	0,947	1,500	1,895	TRUE
	Test	1,026	0,947	1,500	1,895	
	Diff	0,000	0,000	0,000	0,000	
Square-rooted resistance	True	1,000	1,000	1,414	1,732	TRUE
	Test	1,000	1,000	1,414	1,732	
	Diff	0,000	0,000	0,000	0,000	
Communicability (не указан параметр!)	True	0,964	1,072	1,492	1,564	TRUE
	Test	0,964	1,072	1,492	1,564	
	Diff	0,000	0,000	0,000	0,000	
plain Walk, $\alpha = 4.5$	True	1,025	0,950	1,541	1,466	TRUE
	Test	1,025	0,950	1,541	1,466	
	Diff	0,000	0,000	0,000	0,000	
plain Walk, $\alpha = 1$	True	0,988	1,025	1,379	1,416	TRUE
	Test	0,988	1,025	1,379	1,416	
	Diff	0,000	0,000	0,000	0,000	

Таблица 2: Результаты после возведения некоторых расстояний в квадрат

Что, помимо вопроса насчет квадрата расстояния, можно вынести из этой истории? Все меры из этой статьи реализованы верно, функция преобразования параметра $\alpha \rightarrow t$ реализована верно. Также верно реализованы функции-переходы $H0 \rightarrow H$ и $H \rightarrow D$.

3 Chebotarev: The Walk Distances in Graphs

Ссылка: <https://arxiv.org/abs/1103.2059>

Думая о том, почему не совпадают результаты в предыдущей статье, я нашел также и вот эту статью. В ней есть Table 1, в которой те же метрики при тех же параметрах, но используются отношения расстояний. Воспроизвел также и эти результаты. Проблема здесь точно такая же: часть мер нужно возводить в квадрат, чтобы получился такой же ответ. Результаты в таблице 3.

		$\frac{D[0,1]}{D[1,2]}$	$\frac{D[0,1]+D[1,2]}{D[0,2]}$	$\frac{D[0,3]}{D[0,2]}$	Passed
Shortest path, Resistance distance	True	1.000	1.000	1.500	TRUE
	Test	1.000	1.000	1.500	
	Diff	0.000	0.000	0.000	
Walk, $\alpha = 1$	True	1.080	1.000	1.520	TRUE
	Test	1.080	1.000	1.519	
	Diff	0.000	0.000	0.001	
Log forest, $\alpha = 2$	True	0.890	1.000	1.470	TRUE
	Test	0.887	1.000	1.470	
	Diff	0.003	0.000	0.000	
Forest, $\alpha = 1$	True	1.080	1.320	1.260	TRUE
	Test	1.083	1.316	1.263	
	Diff	0.003	0.004	0.003	
Plain walk, $\alpha = 4.5$	True	1.080	1.280	0.950	TRUE
	Test	1.079	1.281	0.951	
	Diff	0.001	0.001	0.001	
Plain walk, $\alpha = 1$	True	0.960	1.460	1.030	TRUE
	Test	0.964	1.459	1.027	
	Diff	0.004	0.001	0.003	

Таблица 3: Результаты после возведения некоторых расстояний в квадрат

Скорее всего, это одни и те же данные, но я подумал, что если здесь есть ошибка, то она могла быть исправлена для одной стати и не исправлена для другой.

4 Kivimaki: Developments in the theory of randomized shortest paths with a article comparison of graph node distances

Ссылка: <https://arxiv.org/abs/1212.1666>

Здесь мы можем использовать два источника: это Figure 2, а также Table 2 с оптимальными значениями из Table 1.

4.1 Figure 2