Distance matrix

In mathematics and [graph](#) theory, a distance matrix is a matrix A containing the pairwise distances (*d*) of a set of points $a_{1...n}$ in the form $[[d_{ij}]]$. An example from every-day life are road distance tables that list the pairwise distances between towns (cf. illustration).
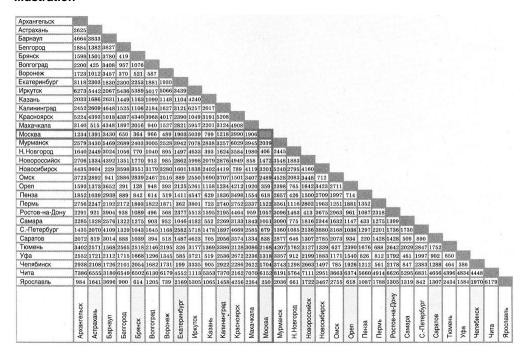
Illustration

	Арх	Астр	Барн	Белг	Брян	Волг	Ворон	Екат	Ирк	Каз	Калин	Красн	Махач	Моск	Мурм	Н.Новг	Новор	Новос	Омск	Орел	Пенз	Пермь	Рост	Сам	С.-Пб	Сарат	Тюм	Уфа	Челяб	Чита
Архангельск																														
Астрахань	2625																													
Барнаул	4664	3833																												
Белгород	1884	1382	3827																											
Брянск	1598	1501	3780	419																										
Волгоград	2200	425	3408	957	1076																									
Воронеж	1723	1012	3457	370	521	587																								
Екатеринбург	3118	2303	1830	2300	2253	1881	1930																							
Иркутск	6273	5442	2067	5436	5389	5017	5066	3439																						
Казань	2033	1686	2631	1449	1163	1099	1148	1104	4240																					
Калининград	2452	2609	4648	1525	1106	2184	1627	3121	6257	2017																				
Красноярск	5224	4393	1018	4387	4340	3968	4017	2390	1049	3191	5208																			
Махачкала	3140	515	4348	1897	2016	940	1527	2821	5957	2201	3124	4908																		
Москва	1234	1391	3430	650	364	966	489	1903	5039	799	1218	3990	1906																	
Мурманск	2579	3430	5469	2689	2403	3005	2528	3942	7078	2838	3257	6029	3945	2039																
Н.Новгород	1640	2449	3024	1056	770	1040	895	1497	4633	393	1624	3584	1980	406	2445															
Новороссийск	2706	1334	4392	1351	1770	913	985	2862	5998	2079	2876	4949	858	1472	3548	1883														
Новосибирск	4435	3604	229	3598	3551	3179	3280	1601	1838	2402	4419	789	4119	3201	5240	2795	4160													
Омск	3723	2892	941	2886	2839	2467	2516	889	2550	1690	3707	1501	3407	2489	4528	2083	3448	712												
Орел	1593	1373	3652	291	128	948	393	2125	5261	1158	1234	4212	1920	359	2398	765	1642	3423	2711											
Пенза	1852	1039	2938	889	842	614	519	1411	4547	629	1836	3498	1554	618	2657	426	1500	2709	1997	714										
Пермь	2756	2247	2192	2172	1893	1822	1871	362	3801	723	2740	2752	2337	1522	3561	1116	2802	1963	1251	1881	1352									
Ростов-на-Дону	2291	921	3904	938	1089	496	568	2377	5513	1595	2195	4464	959	1057	3096	1463	413	3675	2963	961	1087	2318								
Самара	2285	1328	2576	1322	1275	903	952	1046	4182	552	2269	3133	1843	1051	3090	775	1816	2344	1632	1147	433	1275	1399							
С.-Петербург	1435	2070	4109	1329	1043	1645	1168	2582	5718	1478	1897	4669	2585	679	1360	1085	2136	3880	3168	1038	1297	2201	1736	1730						
Саратов	2072	819	3014	888	1039	394	518	1487	4623	705	2056	3574	1334	838	2877	646	1307	2785	2073	934	220	1428	1428	509	890					
Тюмень	3402	2571	1568	2565	2518	2146	2195	326	3177	1369	3386	2128	3086	2168	4207	1762	3127	1339	627	2390	1676	688	2642	2029	2847	1752				
Уфа	2552	1721	2112	1715	1668	1296	1345	585	3721	519	2536	2672	2236	1318	3357	912	2199	1883	1171	1540	826	812	1792	461	1997	902	850			
Челябинск	2938	2108	1736	2101	2054	1682	1731	199	3335	905	2922	2286	2622	1704	3743	1298	2663	1497	785	1926	1212	561	2178	847	2383	1288	464	386		
Чита	7386	6555	3180	6549	6502	6130	6179	4552	1113	5353	7370	2162	7070	6152	8191	5764	7111	2951	3663	6374	5660	4914	6626	5295	6831	4656	4396	4834	4448	
Ярославль	984	1641	3696	900	614	1205	739	2169	5305	1065	1458	4256	2264	250	2036	661	1722	3467	2755	618	1087	1788	1305	1319	842	1307	2434	1584	1970	6179

Fig.1. An every-day example of a distance matrix listing road distances (in km) between important Russian tows.

In the case of [stemmatics](#), the points $a_{1...n}$ are text-samples from *n* witnesses. The mapping from a space of text-strings (w) to the distance matrix may be described as f: L(w×n) → (n×n). For n points this matrix will thus be an n×n-matrix. The distance between the points may be measured by any metric on the underlying space on which the points are defined. Therefore, the mapping f depends on the metric used, i.e. the kind of distance measure one chooses to use between the points. A simple example would be to define any change of a letter as a distance 1, so the distance between a='hortus' and b='ortus' would be $d_{ab} = 1$.

From the definition of a metric follow some basic properties: the matrix is symmetrical (the distance from *a* to *b* is by definition equal to the one from *b* to *a*); its trace is 0 (the distance between *a* and *a* must be 0 for all *a*); all entries $d_{ab} \geq 0$, and equal to 0 if and only if *a* = *b*; for its entries the triangle inequality holds ($d_{ac} \leq d_{ab} + d_{bc}$). As the matrix is by definition symmetrical, often only its lower left half is given (as in the illustration above) as the upper right half would be its mirror-image.

For manuscript traditions an adequate metric may be defined (one that reflects the psychological "distance" between readings) and a distance matrix calculated. The aim here should be to assign a large distance to changes that are unlikely to happen independently and may be hard to revert by a thinking [copyist](#) (cf. *Leitfehler*), whereas trivial changes (like orthographic ones) should be assigned a very small distance. Cf. *Leitfehler-based* [method](#).

There is standard software to approximate the optimal [tree](#) from a given distance matrix, e.g. in the open-source [PHYLIP](#) package. Note that no oriented tree (i.e. a real [stemma](#)) can be gained from a distance matrix without additional information as the latter does not contain orientation. Cf. [polarisation](#).

Cf. [distance-based methods](#).

In other languages

DE: Distanzmatrix
FR: matrice de distance
IT: matrice di distanza

[PR](#)