Supplementary Material for submission 6505 (How Search Algorithms Help Designing Fair Railway Networks)

Appendix 1: NP-hardness proof

Proposition 1. *For any fixed $p \in \mathbb{N}^+ \cup \{+\infty\}$, the p-railway design problem is strongly NP-complete, even when starting from an empty network.*

Proof. We give a polynomial reduction from the NP-hard problem 3-SAT. An instance $\mathcal{I} = (X, D)$ of 3-SAT is a set of Boolean variables X such that $|X| = n$, and a collection of clauses D such that $|D| = m$, where a clause is a disjunction of exactly 3 literals taken from X (a variable or a negation of a variable). The 3-SAT problem asks whether there exists an assignment of each variable of X such that each clause is satisfied.

Given an instance $\mathcal{I}$ of 3-SAT, we build an instance of the decision version of the p-railway design problem as $\mathcal{J} = (C, E, \ell, \tau, B, K, E_0)$ as follows (see Figure 1). We define $k = mn + 2n + 1$.

$$
\begin{aligned}
C &= \{v_x, p_x, n_x \mid x \in X\} \cup \{c_d \mid d \in D\} \cup \{a\} \\
E &= \quad \{(a, p_x), (a, n_x), (v_x, p_x), (v_x, n_x) \mid x \in X\} \\
&\quad \cup \{(c_d, p_x) \mid d \in D, x \in d\} \\
&\quad \cup \{(c_d, n_x) \mid d \in D, \neg x \in d\} \\
\ell_e &= \begin{cases} k & \text{for } e \in \{(a, p_x), (a, n_x) \mid x \in X\} \\ 1 & \text{for } e \in E \setminus \{(a, p_x), (a, n_x) \mid x \in X\} \end{cases} \\
\tau_e &= \begin{cases} 1 & \text{if } e \in \{(a, v_x), (a, c_d) \mid x \in X, d \in D\} \\ 0 & \text{otherwise} \end{cases} \\
B &= 2n + 3m + nk \\
K &= +\infty \\
E_0 &= \emptyset
\end{aligned}
$$

We also fix $\mathrm{SW} = (m + n)^{1/p}(k + 1)$.

Informally, we have a classical incidence graph of the 3-SAT instance. We add moreover a vertex for each variable, connected to both its literals. We also add a vertex a, connected to all literal vertices, but with a high cost. The demands are only from vertex a to all clause vertices and from a to all variable vertices. The budget is set-up such that all edges of weight 1 can be built $(2n + 3m)$ in addition to one edge from a to exactly one literal for each variable (kn). Note that when p is set to $+\infty$, SW equals $k + 1$.

We claim that $\mathcal{I}$ is satisfiable if and only if there is a $R \subseteq E$ respecting the budget B and having a social cost of at most SW.

Assume that $\mathcal{I} = (X, D)$ is satisfiable and let f be an assignment to the variables of X. We build $R \subseteq E$ as follows : add all edges with $l_e = 1$, and add the edge (a, p_x) if $f(x)$ is true, otherwise add the edge (a, n_x). One can easily check that it costs exactly $2n + 3m + kn = B$. The social cost for each demand $(a, v_x), x \in X$ is of $(k + 1)^p$ (with path going through either p_x or n_x according to the assignment of x), and for each $(a, c_d), d \in D$ is of $(k + 1)^p$, with a path going

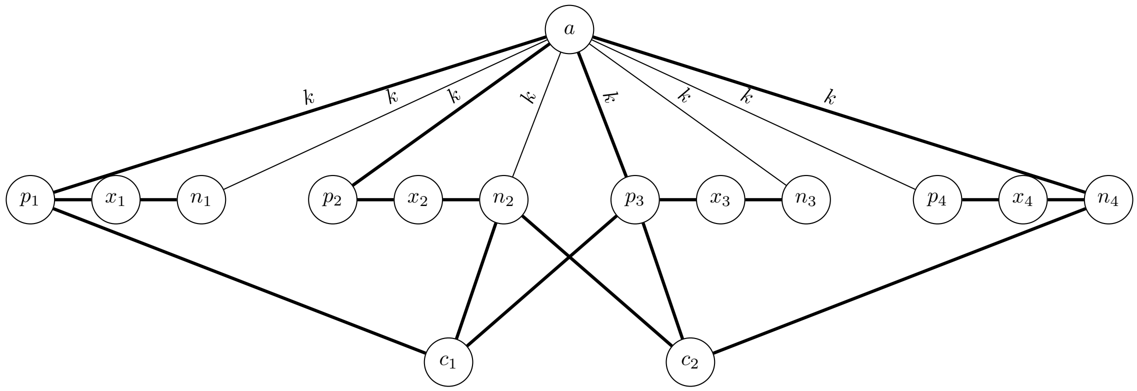


Figure 1: Example of the reduction from a toy instance $(x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_2 \vee x_3 \vee \neg x_4)$. Edges without label are of cost 1. Bold edges correspond to the built solution considering $x_1 = x_2 = x_3 = True$ and $x_4 = False$.

through one literal satisfying d, which exists since $\mathcal{I}$ is satisfiable. Therefore, the total social cost is $(m(k+1)^p + n(k+1)^p)^{1/p} = \mathrm{SW}$

For the reverse direction, assume that $R \subseteq E$ is a solution to $\mathcal{J}$. First, observe that the length of a path between all pairs of cities with positive demand is at least $k+1$, and that it is also the maximum possible length in order to respect SW – therefore, the length of this path must be exactly $k+1$ in any positive solution. From the previous observation, we can deduce that for every demand $(a, v_x), x \in X$, we must have exactly one of the edges $(a, p_x), (a, n_x), x \in X$. Indeed, having none would gives a path of length strictly greater than $k+1$ and thus exceed the social cost SW and having both would exceed the budget. Therefore, we can assign x to True if $(a, p_x) \in R$, and x to False otherwise.

For the demands $(a, c_d), d \in D$, the length of the corresponding path must also be $k+1$, which is possible if and only if there is an edge (a, l_x) in the solution, for some literal $l_x \in d$. By our previous assignment of the variables, such path exists only if the assignment satisfies the clause.

Since all weights are at most k, which is polynomially bounded by the size of the SAT formula, it concludes the claimed result of strong NP-hardness.

Appendix 2: Figures: fairness

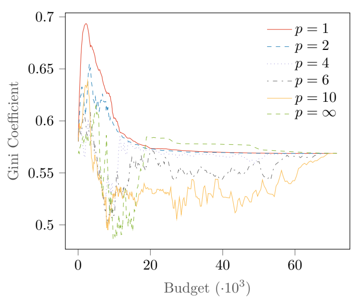
Gini index

The Gini index [Gini(1936)] is a statistical tool that measures economic inequality between individuals in a group (such a country or a region). In economics, the Gini index is often applied to wealth or revenue. Here we apply it to travel distances. The Gini index of a network R under instance $\mathcal{I}$ defined as follow:

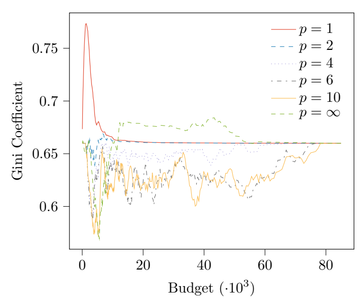
$$G = \frac{\sum_{1 \le i < j \le n} \sum_{1 \le k < l \le n} \left| t_{i,j}^R - t_{k,l}^R \right| \tau_{i,j} \tau_{k,l}}{2 \sum_{1 \le i < j \le n} \sum_{1 \le k < l \le n} t_{k,l}^R \tau_{i,j} \tau_{k,l}}$$

The values of the Gini index range from 0 (perfect equality) to 1 (extreme inequality).

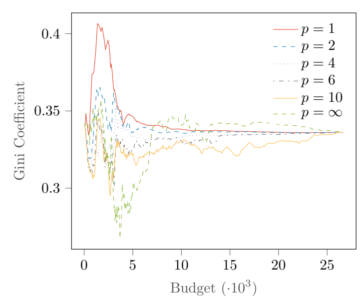
We give below the figures for all 9 countries considered, that show the evolution if the Gini index or the output network with a budget varying from small to large, and for different values of the fairness parameter p: Figures 2a for Brazil, 2b for Canada, 2c for France, 2d for Germany 2e for Italy 2f for Russia 2g for Spain 2h for the UK, and 2i for the USA, with $E_0 = \emptyset$.



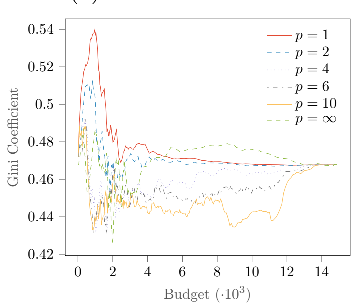
(a) Gini index: Brazil



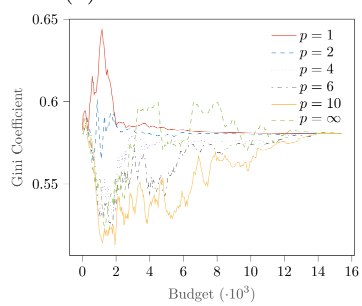
(b) Gini index: Canada



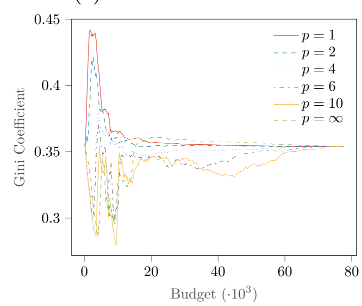
(c) Gini index: France



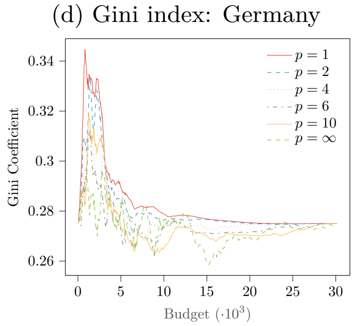
(d) Gini index: Germany



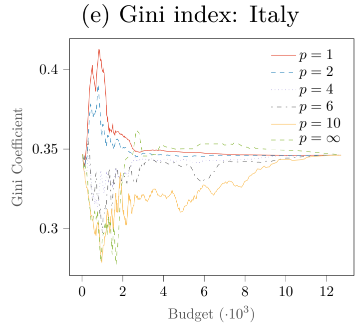
(e) Gini index: Italy



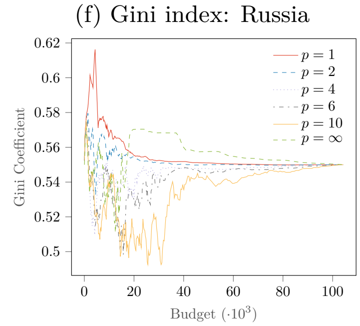
(f) Gini index: Russia



(g) Gini index: Spain



(h) Gini index: The UK



(i) Gini index: The USA

Figure 2: Evolution of the Gini index with budgets for different values of p

Appendix 3: Output networks

Now we present a selection of output networks, for each of the 9 countries considered, and for different values of p, B, and K, with $E_0 = \emptyset$.

Note that some of the maps (especially Canada) look very different from what they usually do. This is because we always normalize the position to a square; if we don't then we will get very unreadable flat maps where all edges are squeezed together.

US

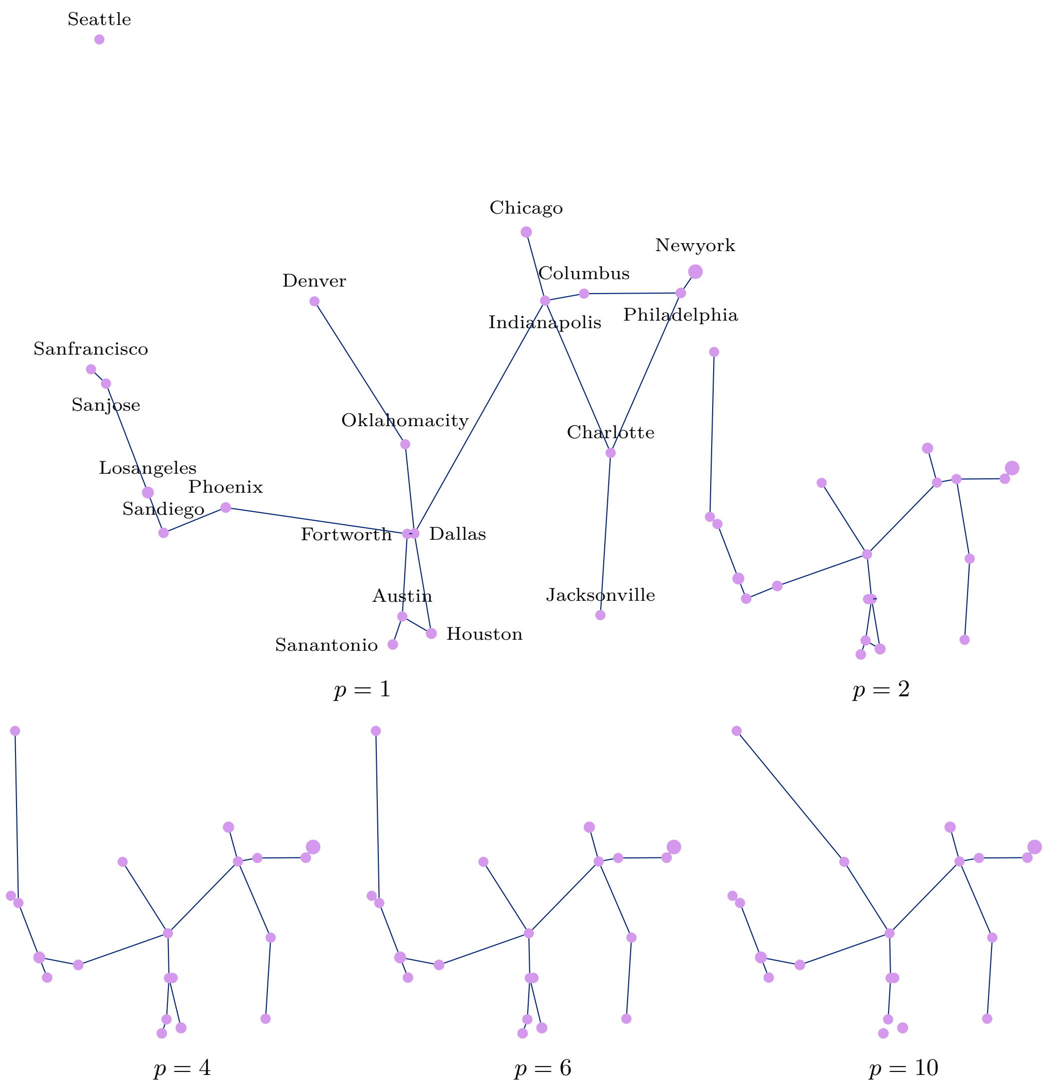


Figure 3: Solution networks for The USA with budget 9443, $K = 3$

Canada

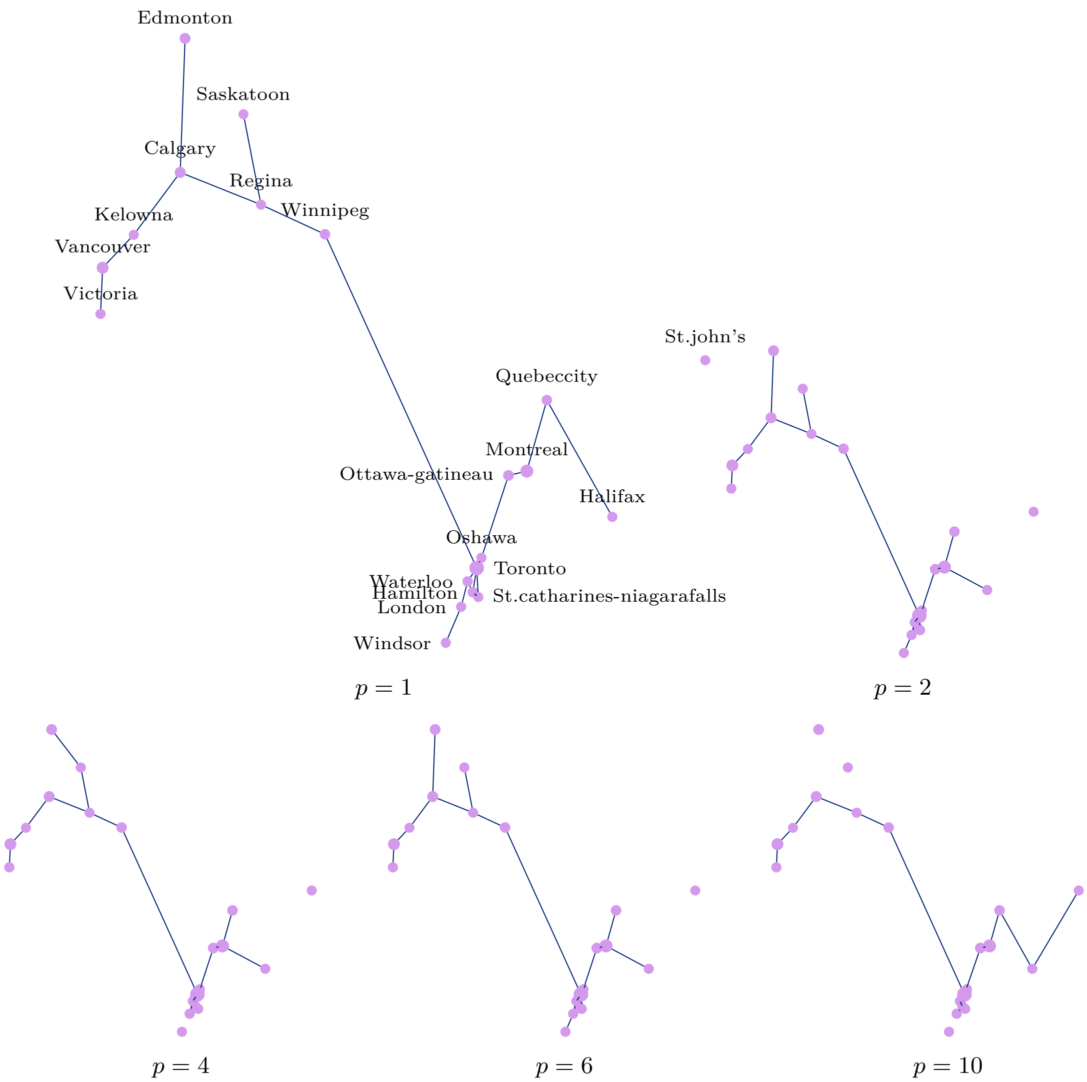


Figure 4: Solution networks for Canada with budget 6003, $K = 3$

Russia

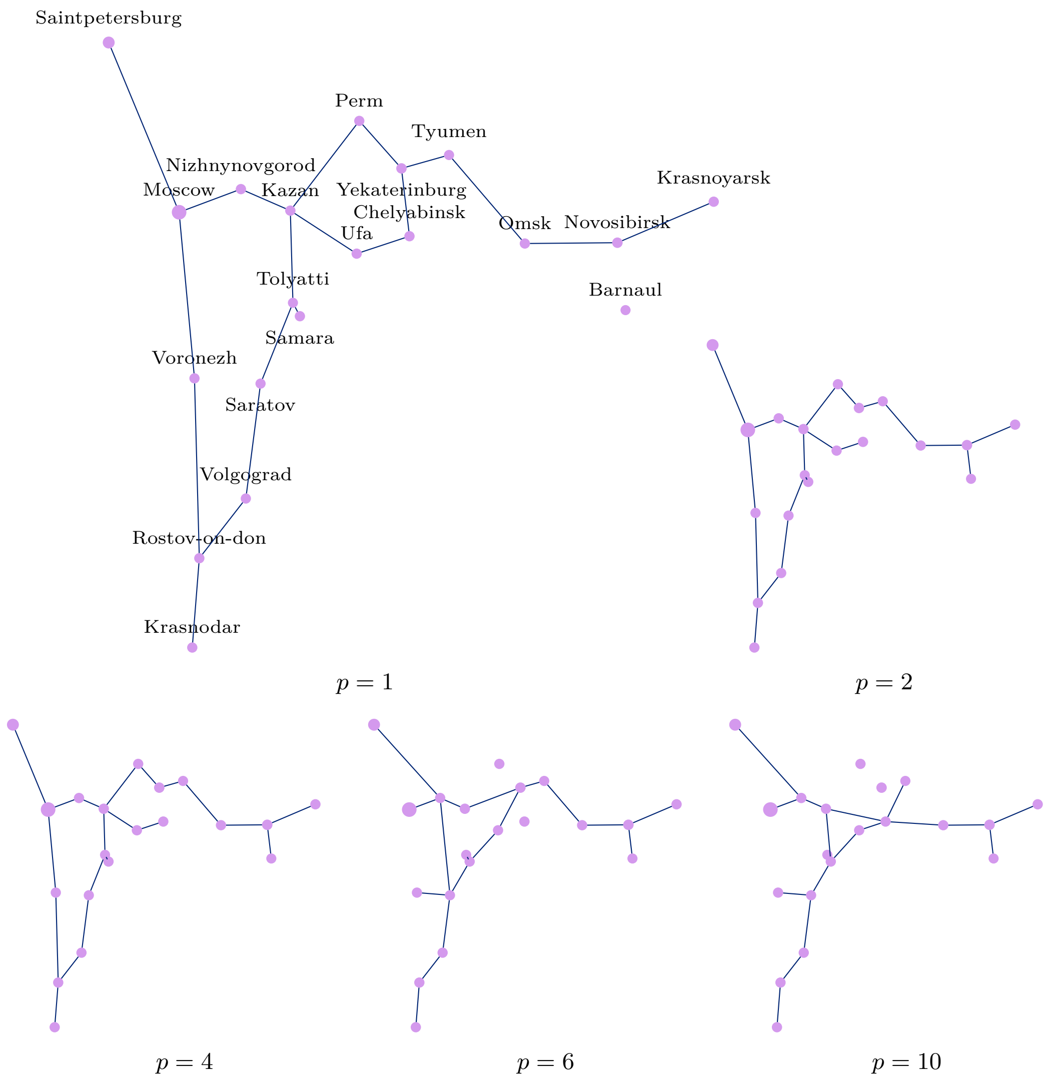


Figure 5: Solution networks for Russia with budget 7828, $K = 3$

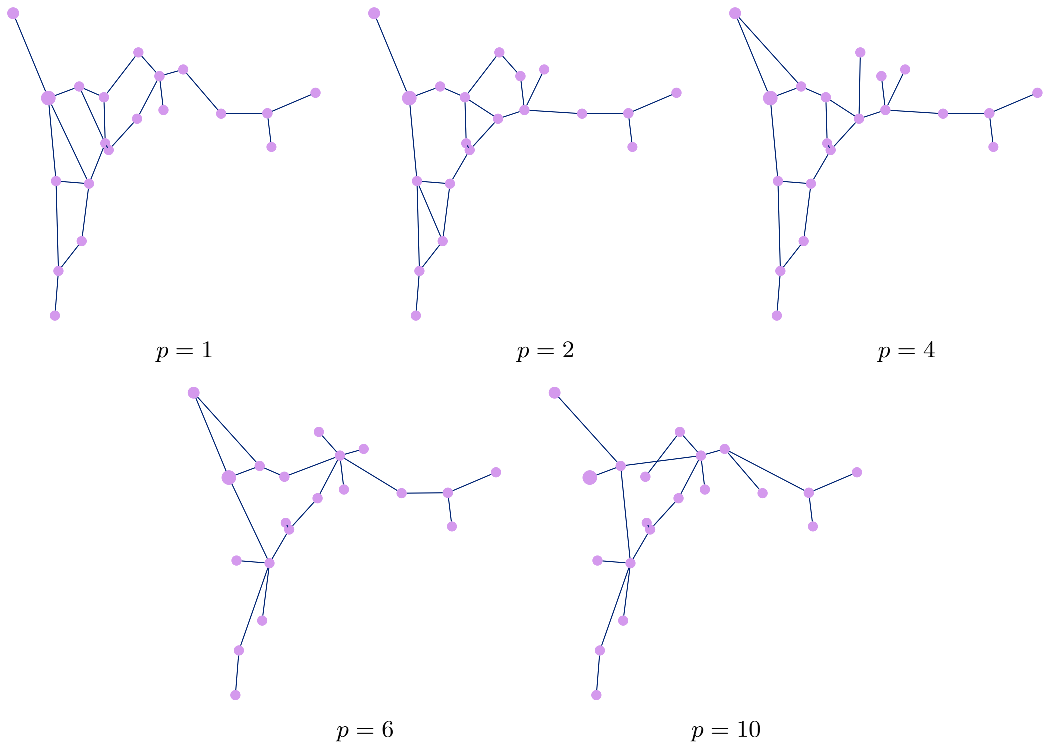


Figure 6: Solution networks for Russia with budget 9676, $K = 3$

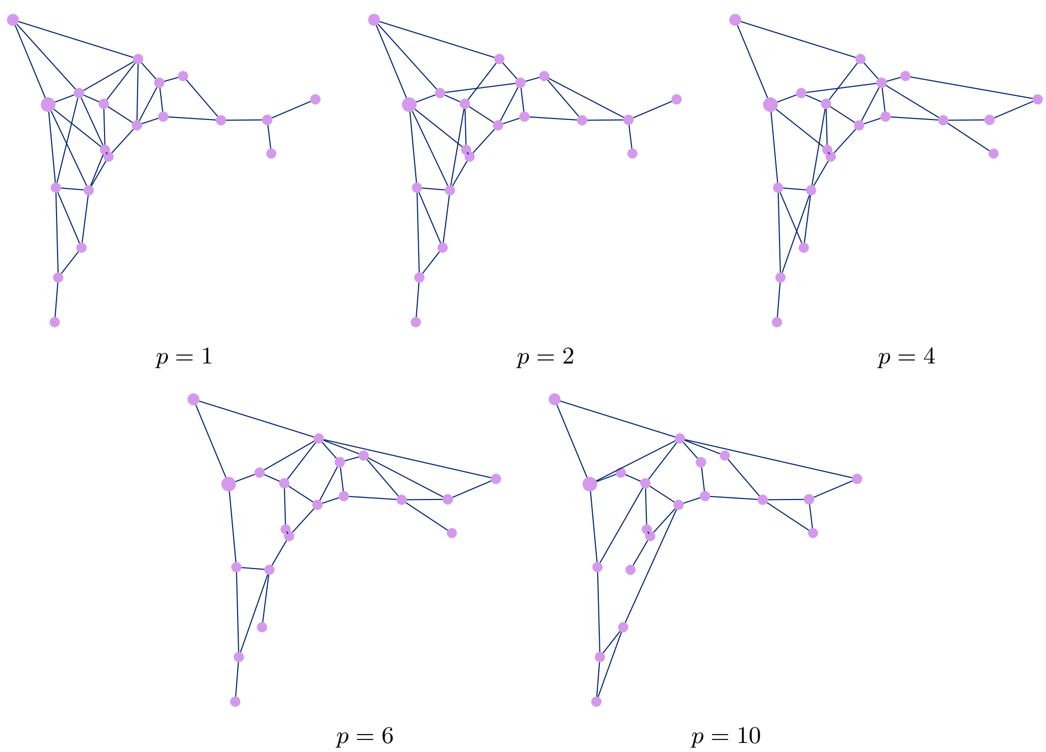


Figure 7: Solution networks for Russia with budget 17085, $K = 3$

Brazil

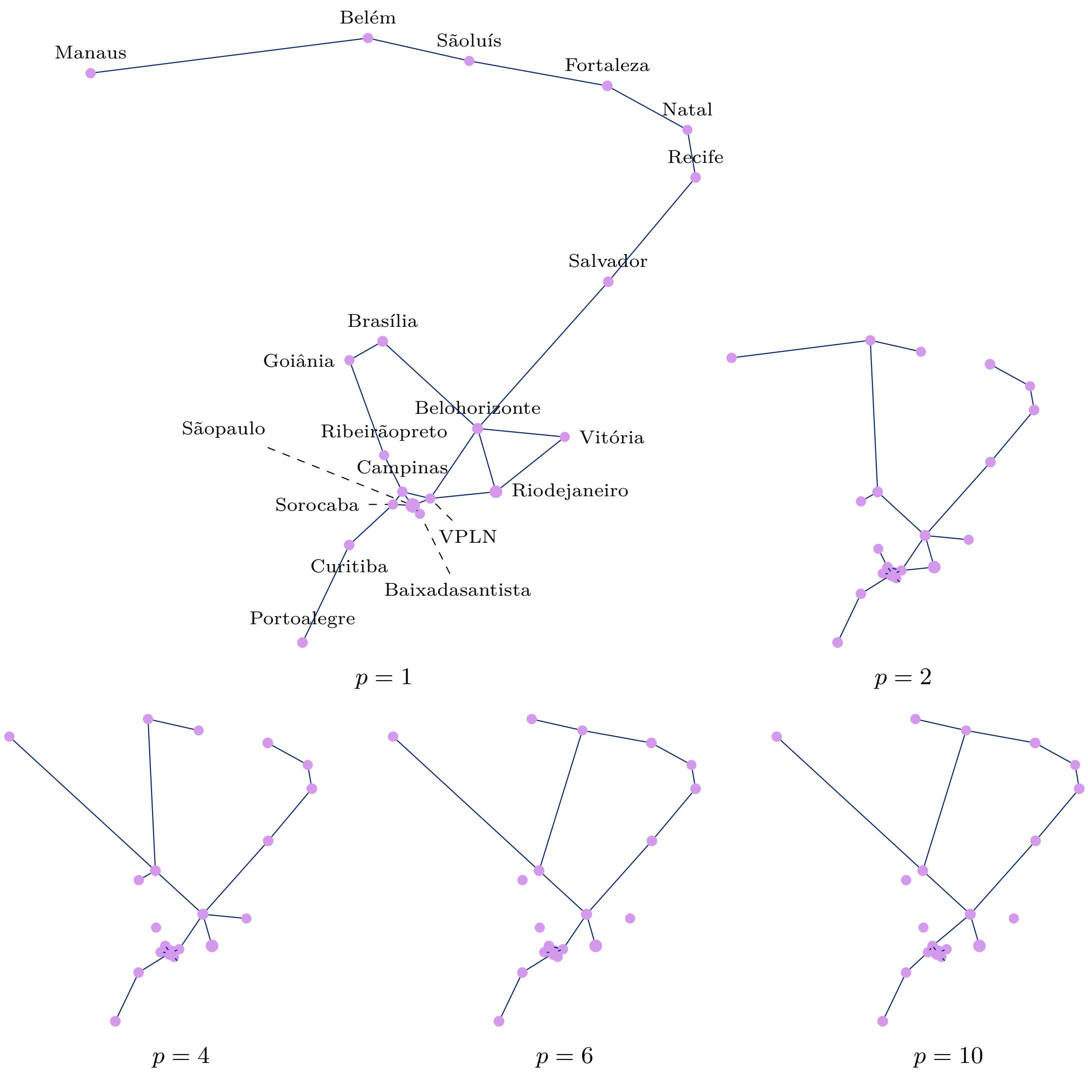


Figure 8: Solution networks for Brazil with budget 9510, $K = 3$

France

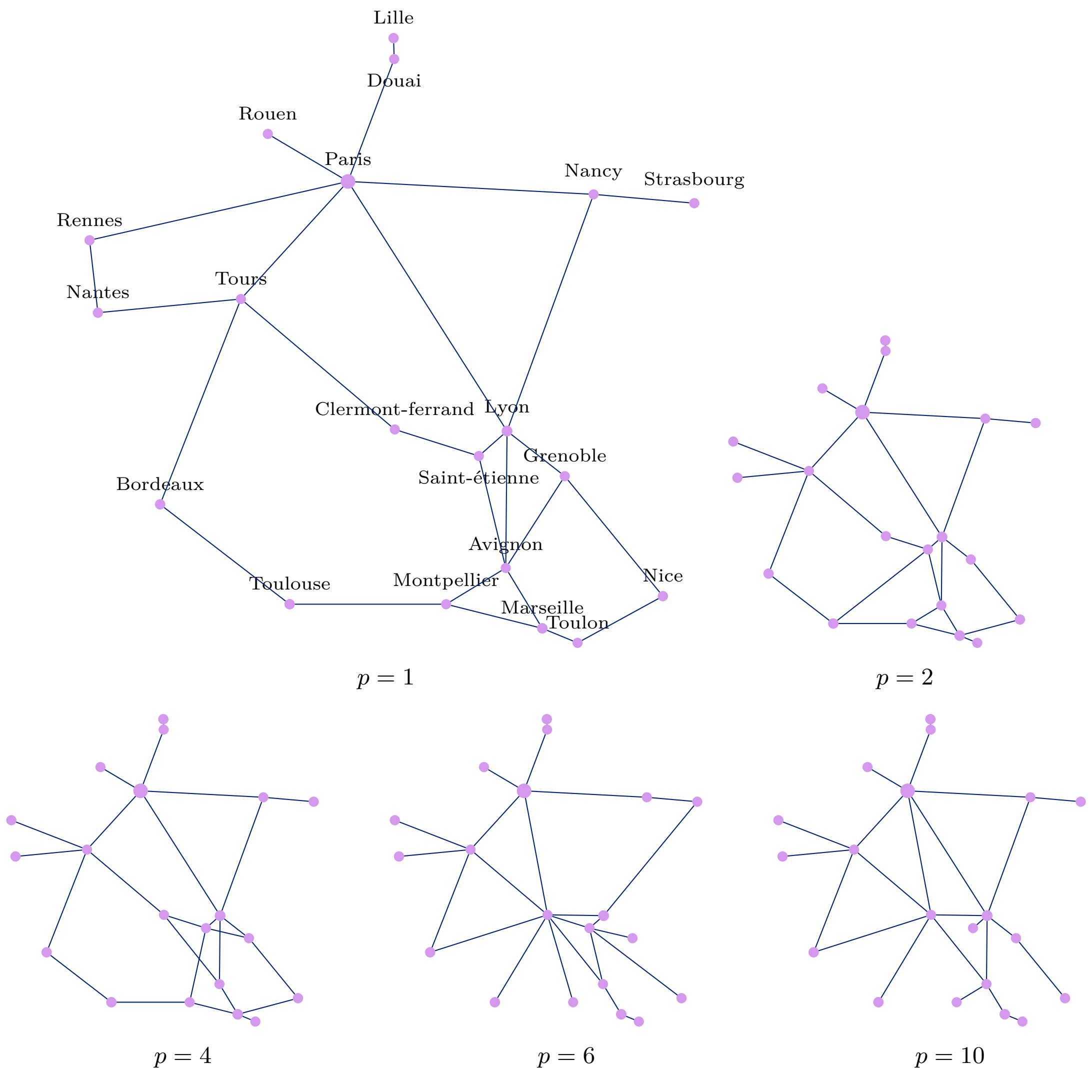


Figure 9: Solution networks for France with budget 4642, $K = 3$

Spain

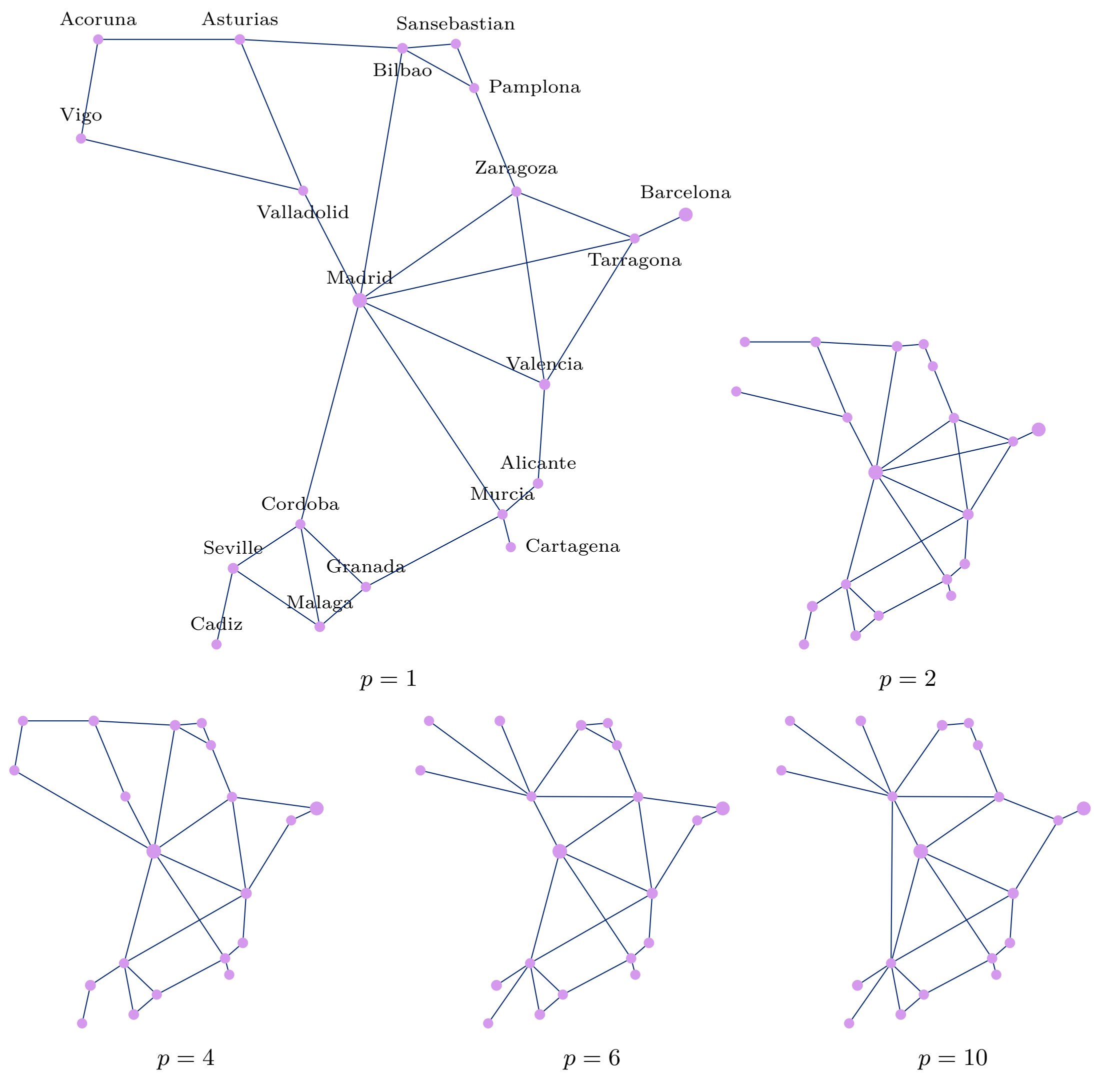


Figure 10: Solution networks for Spain with budget 5632, $K = 3$

Germany

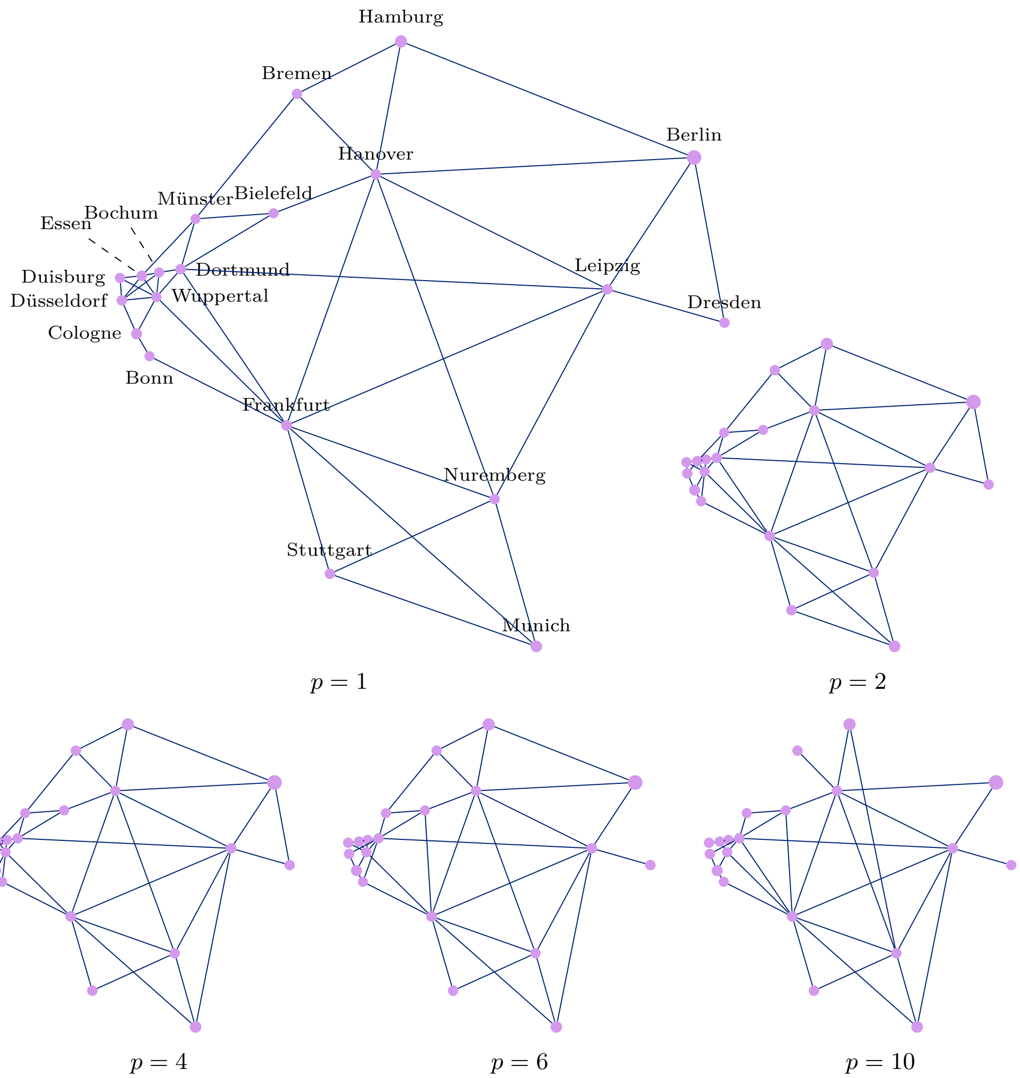


Figure 11: Solution networks for Germany with budget 5446, $K = 3$

UK

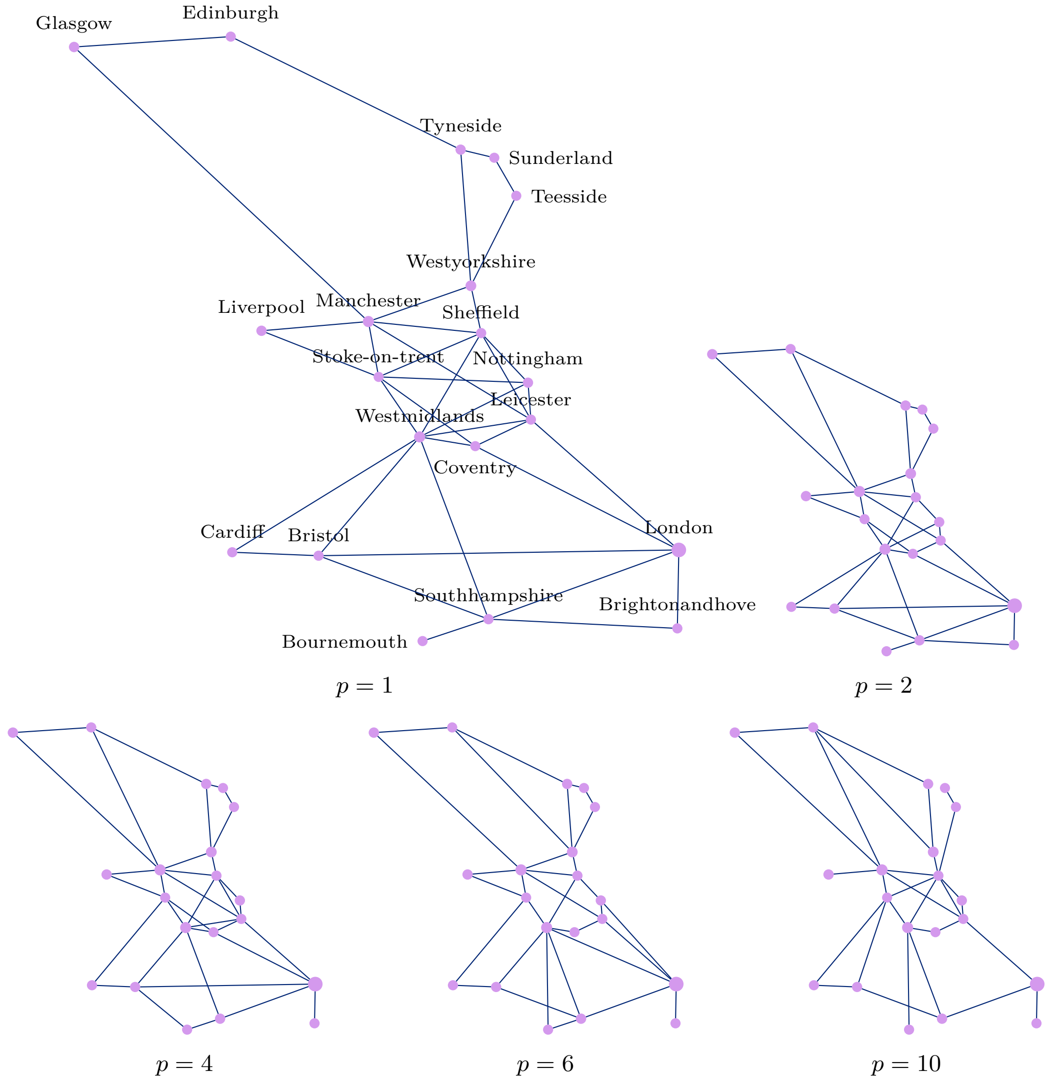


Figure 12: Solution networks for the UK with budget 3352, $K = 3$

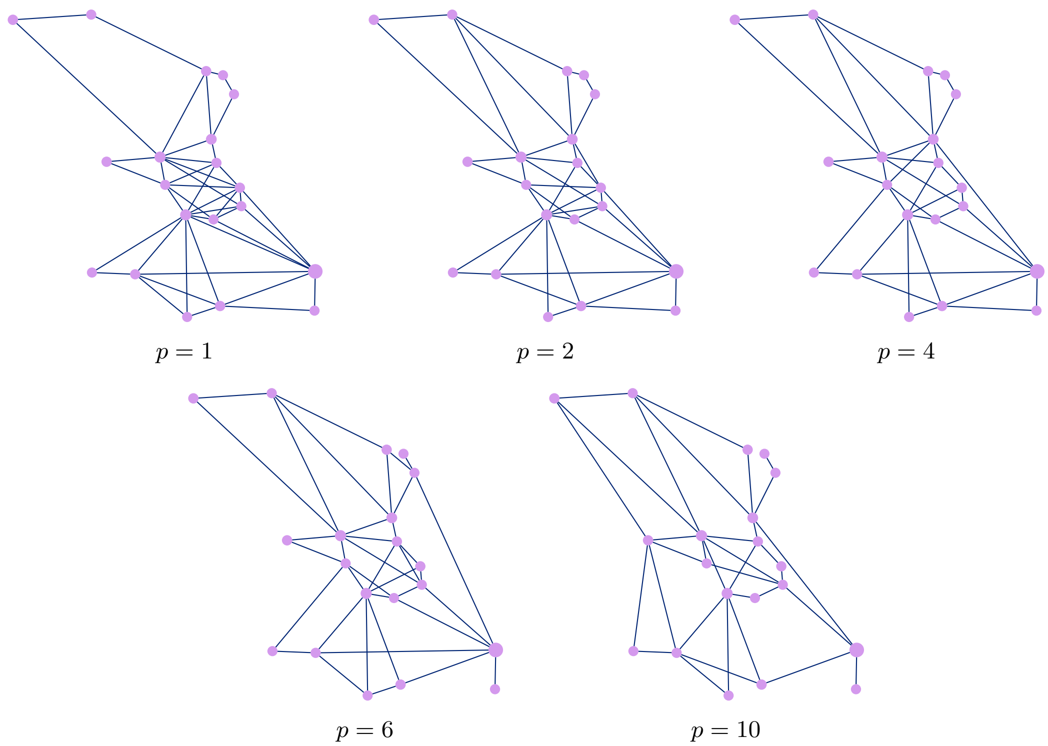


Figure 13: Solution networks for the UK with budget 4212, $K = 3$

Italy

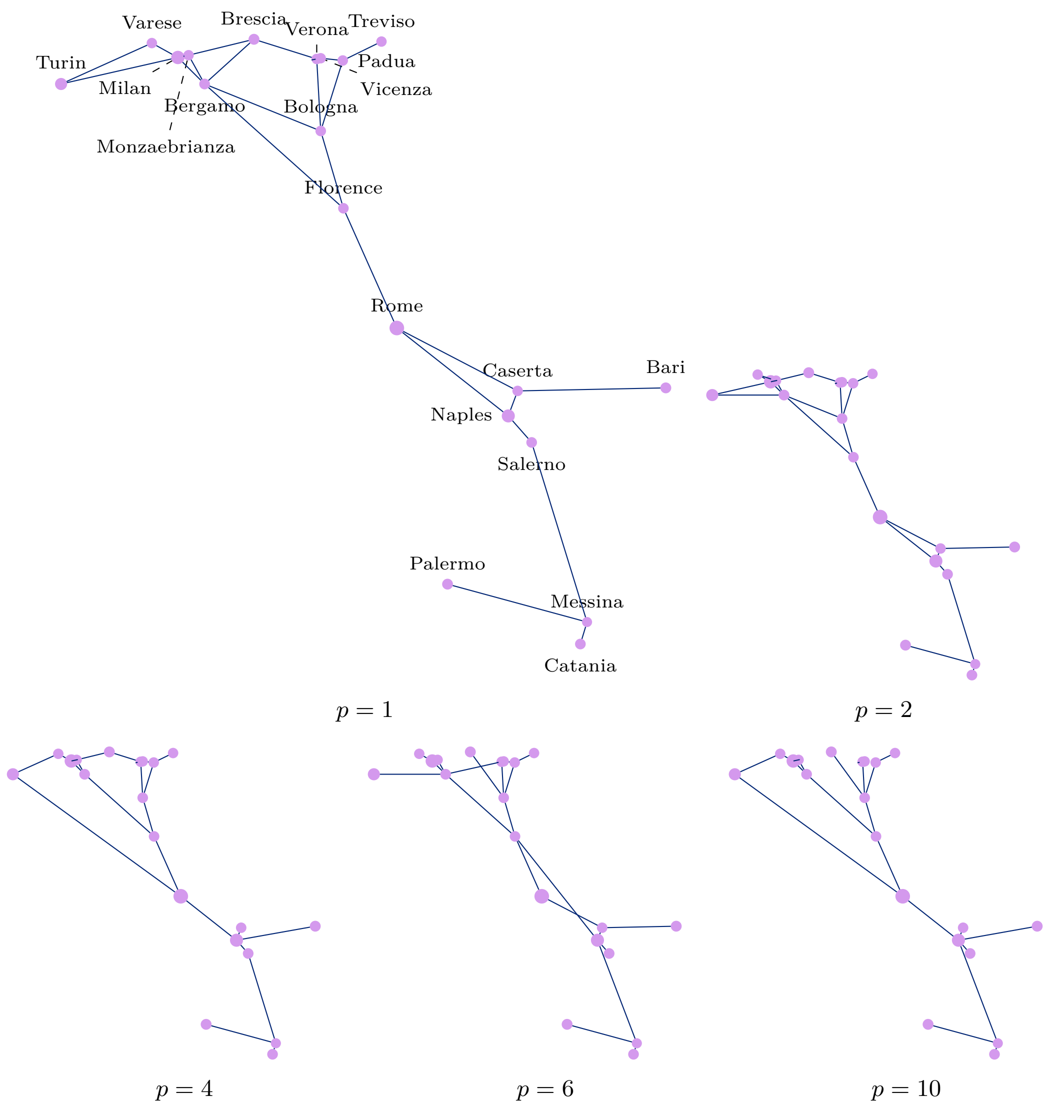
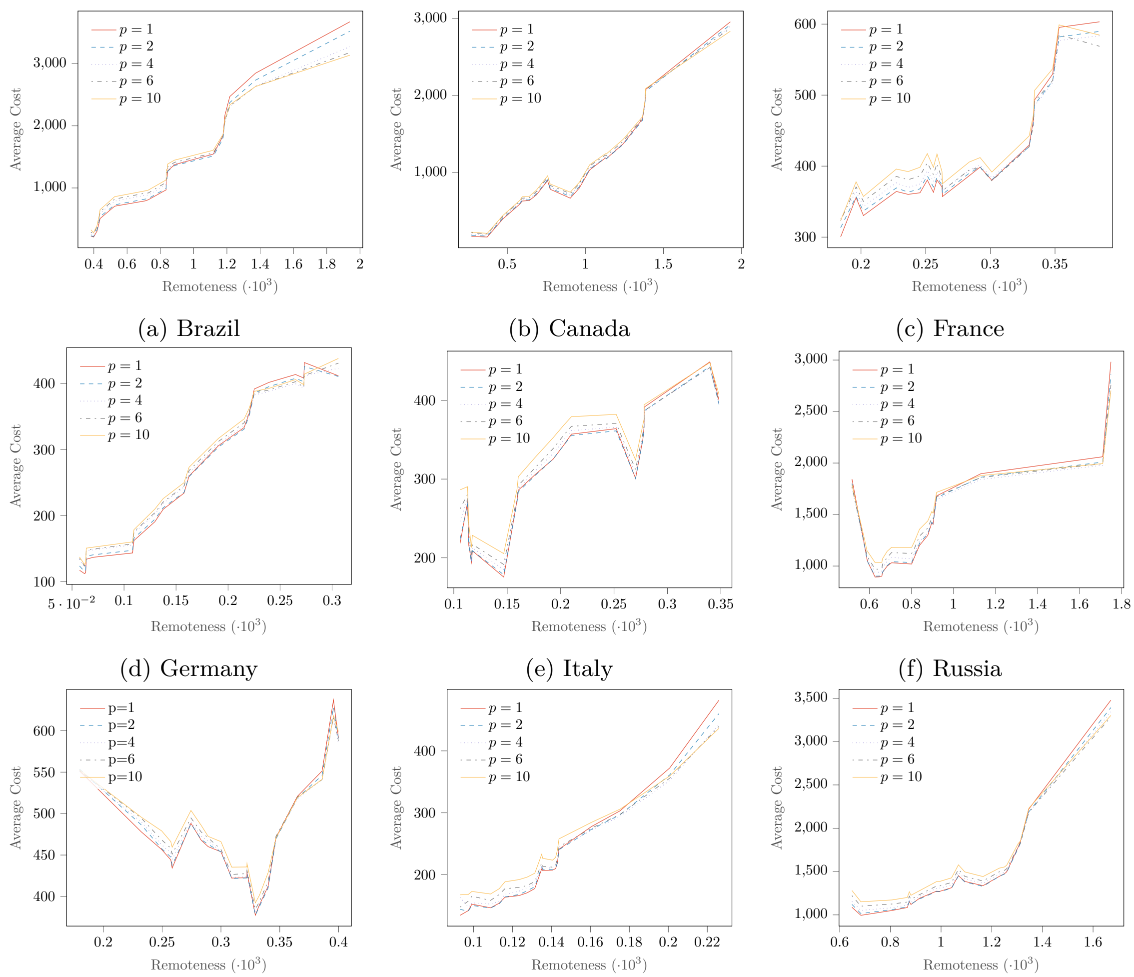


Figure 14: Solution networks for the Italy with budget 2813, $K = 3$

Appendix 4: Figures: influence of remoteness

Now we present figures that show the influence of centrality (or remoteness) of cities on the average travel cost of the individuals who travel from/to that city, depending on the fairness parameter p.



(a) Brazil (b) Canada (c) France

(d) Germany (e) Italy (f) Russia

(g) Spain (h) The UK (i) The USA

Figure 15: The change of average cost with remoteness under different values of p

Appendix 5: average travel cost for all cities

Here we show, for each country considered, the average travel cost for all cities. The average is taken over all citizens traveling to/from the city and over all considered budgets. We also show the population of each city to make it easier to verify that larger cities tend to benefit more from small values of p than smaller cities.

City	Population	$p = 1$	$p = 4$	$p = 10$
Paris	1257	409	416	432
Lyon	231	352	358	398
Marseille	176	414	412	440
Toulouse	135	582	568	607
Bordeaux	123	601	579	606
Lille	119	300	323	326
Nice	101	539	542	575
Nantes	96	477	483	512
Rennes	73	478	492	518
Nancy	44	453	439	456
Grenoble	69	369	388	413
Rouen	67	231	268	268
Toulon	63	354	357	385
Montpellier	61	421	426	467
Douai	54	234	251	253
Avignon	53	353	354	370
Saint-étienne	52	272	281	298
Tours	49	327	325	348
Clermont-ferrand	48	408	410	344

Table 1: Average travel costs: France

City	Population	$p = 1$	$p = 4$	$p = 10$
São Paulo	21734	351	387	443
Rio de Janeiro	12763	634	752	893
Belo Horizonte	5961	669	690	756
Brasília	4627	1051	1067	1101
Porto Alegre	4340	1398	1494	1463
Fortaleza	4106	2068	2078	2210
Recife	4079	1742	1774	1927
Salvador	3929	1457	1478	1605
Curitiba	3654	675	724	782
Campinas	3264	246	275	291
Manaus	2676	3521	3272	3136
Goiânia	2560	999	1044	1096
Vale do Paraíba	2552	230	254	326
Belém	2510	2612	2562	2549
Sorocaba	2143	240	270	314
Vitória	1979	1005	1075	1161
Baixada Santista	1865	184	211	227
Ribeirão Preto	1720	616	678	673
São Luís	1633	2212	2184	2172
Natal	1604	1578	1633	1753

Table 2: Average travel costs: Brazil

City	Population	$p = 1$	$p = 4$	$p = 10$
London	978	231	235	262
Manchester	255	160	169	180
West Midlands	244	153	170	180
West Yorkshire	177	168	168	181
Glasgow	95	482	443	437
Liverpool	86	176	189	212
South Hampshire	85	186	194	227
Tyneside	77	249	243	255
Nottingham	72	151	145	181
Sheffield	68	141	134	155
Bristol	61	226	234	250
Leicester	50	135	131	150
Edinburgh	48	404	379	379
Brighton and Hove	47	166	172	178
Bournemouth	46	219	225	260
Cardiff	44	261	267	281
Teesside	37	223	224	240
Stoke-on-Trent	37	134	164	168
Coventry	35	111	125	136
Sunderland	33	201	199	211

Table 3: Average travel costs: the UK

City	Population	$p = 1$	$p = 4$	$p = 10$
Toronto	5647	361	388	422
Montreal	3675	671	724	753
Vancouver	2426	1707	1697	1756
Calgary	1305	1503	1486	1532
Edmonton	1151	1574	1579	1687
Ottawa-Gatineau	1068	436	471	533
Winnipeg	758	1758	1723	1752
Quebec City	733	690	750	786
Hamilton	729	171	196	227
Waterloo	522	218	242	268
London	423	352	393	418
Victoria	363	711	727	799
Halifax	348	1638	1663	1706
Oshawa	335	150	162	175
Windsor	306	663	718	756
Saskatoon	264	1589	1598	1657
St. Catharines-Niagara Falls	242	171	214	231
Regina	224	1589	1583	1597
St. John's	185	2959	2900	2837
Kelowna	181	1143	1132	1159

Table 4: Average travel costs: Canada

City	Population	$p = 1$	$p = 4$	$p = 10$
Moscow	13010	1000	1052	1174
Saint Petersburg	5601	1117	1195	1306
Novosibirsk	1633	2084	1974	1974
Yekaterinburg	1544	1061	1063	1161
Kazan	1308	829	870	1025
Nizhny Novgorod	1228	686	730	811
Chelyabinsk	1189	1113	1147	1177
Krasnoyarsk	1187	2983	2766	2704
Samara	1173	794	846	909
Ufa	1144	1085	1145	1178
Rostov-on-Don	1142	1226	1271	1414
Omsk	1125	1795	1727	1739
Krasnodar	1099	1504	1539	1669
Voronezh	1057	836	888	1062
Perm	1034	1056	1124	1295
Volgograd	1028	1221	1279	1334
Saratov	901	974	1040	1095
Tyumen	847	1285	1276	1442
Tolyatti	684	673	723	806
Barnaul	630	1844	1760	1788

Table 5: Average travel costs: Russia

City	Population	$p = 1$	$p = 4$	$p = 10$
New York	8258	1194	1216	1274
Los Angeles	3820	1532	1537	1630
Chicago	2664	1501	1537	1614
Houston	2314	1287	1373	1520
Phoenix	1650	1682	1664	1773
Philadelphia	1550	454	467	507
San Antonio	1495	1087	1154	1281
San Diego	1388	1031	1084	1157
Dallas	1302	797	815	882
Jacksonville	985	2123	2200	2321
Austin	979	865	917	1014
Fort Worth	978	720	737	795
San Jose	969	1488	1456	1565
Columbus	913	1147	1170	1271
Charlotte	911	1566	1634	1734
Indianapolis	879	1100	1108	1163
San Francisco	808	1548	1518	1613
Seattle	755	3479	3343	3303
Denver	716	2132	2143	2073
Oklahoma City	702	1375	1417	1389

Table 6: Average travel costs: the US

City	Population	$p = 1$	$p = 4$	$p = 10$
Madrid	615	446	454	457
Barcelona	517	519	523	543
Valencia	164	364	395	395
Seville	130	525	523	533
Bilbao	98	482	495	505
Malaga	94	500	507	518
Asturias	84	592	584	599
Alicante	79	361	390	394
Zaragoza	63	389	353	418
Murcia	62	358	387	392
Granada	44	466	468	474
Vigo	41	800	750	711
Cartagena	40	382	412	417
Cadiz	40	552	551	553
San Sebastian	39	433	438	463
A Coruña	37	760	729	727
Valladolid	36	362	361	353
Tarragona	32	244	248	259
Cordoba	31	400	396	396
Pamplona	28	416	403	446

Table 7: Average travel costs: Spain

City	Population	$p = 1$	$p = 4$	$p = 10$
Berlin	367	412	423	438
Hamburg	190	356	371	390
Munich	148	558	521	514
Cologne	107	163	176	184
Frankfurt	75	329	341	346
Stuttgart	62	460	453	450
Düsseldorf	61	124	142	149
Leipzig	60	326	291	291
Dortmund	58	138	144	148
Essen	57	107	116	120
Bremen	56	306	318	339
Dresden	55	395	382	397
Hanover	53	256	255	259
Nuremberg	51	390	359	360
Duisburg	49	117	134	137
Bochum	36	102	109	113
Wuppertal	35	117	124	127
Bielefeld	33	234	237	242
Bonn	33	156	169	177
Münster	31	214	234	233

Table 8: Average travel costs: Germany

City	Population	$p = 1$	$p = 4$	$p = 10$
Rome	422	400	395	407
Milan	322	152	162	184
Naples	298	285	280	290
Turin	220	388	401	431
Brescia	125	218	246	286
Bari	122	540	541	555
Palermo	120	849	804	787
Bergamo	110	186	194	212
Catania	107	508	486	479
Salerno	106	253	248	259
Bologna	101	275	293	317
Florence	98	325	325	338
Padua	93	192	210	241
Verona	92	99	110	129
Caserta	90	217	222	231
Varese	87	169	185	215
Treviso	87	293	320	358
Monza e Brianza	87	84	92	105
Vicenza	85	95	104	122
Messina	60	407	388	382

Table 9: Average travel costs: Italy

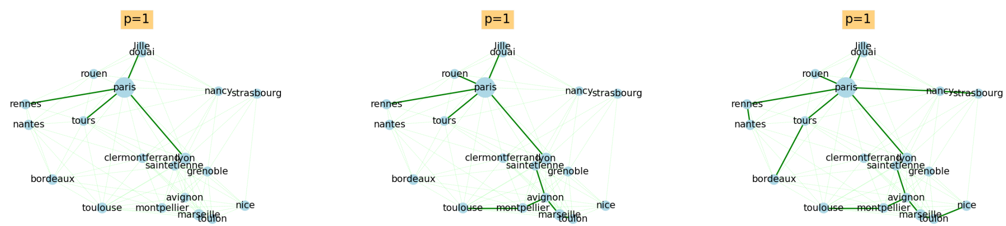
Appendix 6: Output with pre-existing network

As a proof of concept, we ran our algorithm with different pre-existing networks, for different values of p and budget on the France network.

To start with realistic pre-existing networks, we took the actual high-speed railway networks existing at different years and initialized E_0 with the lines existing at that time.

We provide as supplementary material the .gif animation of the different networks with increasing budgets for 3 different values of p in the attached files, starting with network of year 1991 for E_0.

For illustration purposes, we plot on Figure 16 the original network as it was in 1991, with two other networks with two different budgets, for $p = 1$. We remark that the evolution of the network over time has significant similarities with what happened in reality (with a few differences, still).



$p = 1$, initial network $p = 1$, budget 1887 $p = 1$, budget 2810

Figure 16

References

[Gini(1936)] Gini, C. 1936. On the Measure of Concentration with Special Reference to Income and Statistics. *Colorado College Publication, General Series*, (208): 73–79.