

## HeteWalk 6 datasets

1. Gene (proteins) interaction network: No problem found

```
In [8]: for_nx
```

```
Out[8]:
```

	Interactor 1 HPRD Id	Interactor 2 HPRD Id
0	1	1
1	2761	7
2	16000	17
3	1513	115
4	3311	1281
5	3995	1281
6	4199	1281
7	4404	1281
8	5961	1281
9	6090	1281

2. miRNA similarity network: miRNA similarity vectors are given.  
Association is calculated by misim score. Direct association is not given.

1	0.0553	0	0.2082	0.1875	0.15	0	0.1256	0	0	0.054	0.0592	0.0235	0.1186	0.0909	0.1228	0.1557	0.0928	0.0213
0.0553	1	0	0.0259	0.1051	0.2634	0.0303	0.287	0	0.4	0.2088	0.3003	0.0134	0.4568	0.4136	0.2345	0.0333	0.4136	0.03
0	0	1	0	0	0	0.3333	0	0	0.0794	0	0.6667	0	0	0.0099	0	0.5	0	0.4109
0.2082	0.0259	0	1	0.1496	0.1392	0	0.0283	0	0	0.0408	0.0255	0.0112	0.0937	0.0425	0.0839	0.3957	0.0435	0.0135
0.1875	0.1051	0	0	0.1496	1	0.3652	0.0588	0.0855	0.0882	0	0.2632	0.1154	0.0457	0.1886	0.1724	0.2817	0.1075	0.1765
0.15	0.2634	0	0	0.1392	0.3652	1	0.0196	0.1519	0.0221	0	0.2472	0.3281	0.2219	0.2303	0.1186	0.5098	0.2925	0.3274
0	0.0303	0	0	0.0588	0.0196	1	0.026	0.6667	0.0606	0.0335	0	0	0	0.0541	0	0.0231	0.0196	0.0353
0.1256	0.287	0	0.3333	0.0283	0.0855	0.1519	0.026	1	0	0.3333	0.4225	0.0498	0.3483	0.1658	0.0809	0.2107	0.0458	0.03
0	0	0	0	0.0882	0.0221	0.6667	0	1	0	0.0221	0	0	0	0.016	0	0.0221	0	0.01
0	0.4	0	0	0	0.0606	0.3333	0	1	0.1902	0	0	0	0.0128	0	0	0.0319	0	0.0075
0.054	0.2088	0	0.0794	0.0408	0.2632	0.2472	0.0335	0.4225	0.0221	0.1902	1	0.0568	0.1155	0.1122	0.0483	0.4597	0.0583	0.0515
0.0592	0.3003	0	0	0.0255	0.1154	0.3281	0	0.0498	0	0.0568	1	0.0411	0.1399	0.0423	0.4153	0.3659	0.5084	0.0515
0.0235	0.0134	0	0.6667	0.0112	0.0457	0.2219	0	0.3483	0	0	0.1155	0.0411	1	0.0557	0.0175	0.1095	0.0274	0.018
0.1186	0.4568	0	0.0937	0.1886	0.2303	0	0.1658	0	0	0.1122	0.1399	0.0557	1	0.2463	0.2221	0.1223	0.1124	0.057
0.0909	0.4136	0	0	0.0425	0.1724	0.1186	0	0.0809	0	0	0.0483	0.0423	0.0175	0.2463	1	0.0957	0.0391	0.0682
0.1228	0.2345	0.0099	0.0839	0.2817	0.5098	0.0541	0.2107	0.016	0.0128	0.4597	0.4153	0.1095	0.2221	0.0957	1	0.2532	1	0.2532
0.1557	0.0333	0	0.3957	0.1075	0.2925	0	0.0458	0	0.0583	0.3659	0.0274	0.1223	0.0391	0.2532	1	0.04	0.1337	0.1337
0.0928	0.4136	0	0.0435	0.1765	0.3274	0	0.0396	0	0.0495	0.5084	0.018	0.1124	0.0682	0.2477	0.04	1	0.0181	0.0181
0.0213	0.0336	0	0.0135	0.0366	0.0926	0.0231	0.1767	0	0.0319	0.1125	0.0515	0.0507	0.057	0.0176	0.2394	0.1337	0.01	0.01
0.1194	0.2427	0.25	0.0556	0.3009	0.3993	0.0196	0.2673	0.0221	0	0.3713	0.2903	0.2621	0.1523	0.0656	0.6484	0.1	0.28	0.28
0.1007	0.4022	0	0.0769	0.1631	0.4191	0	0.1756	0	0.126	0.0935	0.4714	0.6099	0.201	0.2255	0.0922	0.0939	0.06	0.06
0.1566	0.1924	0.5	0.1269	0.5857	0.3599	0.0353	0.4276	0.0441	0	0.3048	0.1218	0.4476	0.4046	0.2828	0.2964	0.1157	0.14	0.14
0.2037	0.1149	0	0.2657	0.3062	0.6176	0.0294	0.2702	0.016	0.0075	0.2289	0.1498	0.0909	0.3568	0.1145	0.4973	0.3456	0.3456	0.3456
0.0344	0.0558	0.4109	0.0172	0.0194	0.0578	0.1667	0.3618	0	0.075	0.2537	0.0288	0.3599	0.0305	0.0094	0.1953	0.0327	0.0327	0.0327
0.2104	0.123	0	0.2774	0.3414	0.5989	0.016	0.2833	0.0176	0	0.2499	0.1572	0.2386	0.3219	0.1289	0.4926	0.2498	0.11	0.11
0	1	0	0	0	0.3333	0	0.0794	0	0.6667	0	0.0099	0	0	0.25	0	0.5	0	0.4109
0.1758	0.1373	0	0.1751	0.1827	0.5882	0.0288	0.2796	0.0071	0.0053	0.2668	0.234	0.1696	0.1948	0.0679	0.591	0.3178	0.3178	0.3178
0.2093	0.0278	0	0.0233	0.1414	0.0985	0	0.0864	0	0.0464	0.0628	0.0355	0.3165	0.0457	0.1069	0.0436	0.0466	0.04	0.04
0	1	0	0	0	0.3333	0	0.0794	0	0.6667	0	0.0099	0	0	0.25	0	0.5	0	0.4109
0	0	0	0.0882	0.0221	0.6667	0	1	0.0221	0	0	0	0.016	0	0	0.0221	0	0.0441	0.016
0.1296	0.2449	0	0.0589	0.12	0.4175	0	0.0754	0	0.0869	0.2944	0.3056	0.2741	0.0449	0.2681	0.0922	0.356	0.05	0.05
0.0167	0.0403	0	0.0079	0.0321	0.0858	0	0.2822	0	0.2902	0.032	0.0358	0.0774	0.0123	0.3026	0.0221	0.0126	0.16	0.16
0.1362	0.306	0	0.0313	0.1387	0.1683	0	0.1132	0	0.0712	0.1042	0.0546	0.5849	0.518	0.1688	0.0761	0.0525	0.06	0.06
0.1355	0.2933	0.3333	0.0641	0.3717	0.3112	0.0512	0.5111	0.0294	0.3333	0.3786	0.1349	0.3332	0.1381	0.0761	0.4847	0.11	0.11	0.11
0.1392	0.1807	0	0.2145	0.2244	0.5414	0.0084	0.3364	0.0088	0	0.3008	0.2868	0.1511	0.2815	0.1098	0.6566	0.3786	0.17	0.17
0	0	0	0.0882	0.0221	0.6667	0	1	0.0221	0	0	0.016	0	0	0.0221	0	0.0441	0.016	0.01
0.0417	0.3087	0	0.0272	0.0159	0.0817	0.1208	0	0.131	0	0.2863	0.0512	0.0532	0.554	0.0284	0.2754	0.0392	0.029	0.029
0	0	0	0.0882	0.0221	0.6667	0	1	0.0221	0	0	0.016	0	0	0.0221	0	0.0441	0.016	0.01
0.2195	0.2341	0	0.1789	0.3429	0.2427	0	0.2041	0	0.1013	0.1386	0.0565	0.5212	0.4795	0.1946	0.1275	0.2113	0.04	0.04
0.0606	0.678	0	0.0283	0.1149	0.1054	0.0454	0.355	0	0.6667	0.212	0.0338	0.0132	0.1971	0.6667	0.0939	0.0326	0.04	0.04

- Disease phenotype similarity network: Column A is the id for disease. Vectors are given for diseases. How associations are created by paper is not given.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
100050	1	0.0890344	0.361651	0.266263	0.368639	0.197142	0.0678806	0.203997	0.238039	0.210782	0.15004	0.368094	0.408045	0.444959	0.357986	0.428706
100070	0.0890344	1	0.0873663	0.0267854	0.148069	0.233403	0.125448	0.0727986	0.0895563	0.0539512	0.15665	0.0315105	0.132105	0.076906	0.0358535	0.127289
100100	0.361651	0.0873663	1	0.0423538	0.269991	0.112267	0.0739974	0.2932	0.170992	0.0852404	0.120332	0.19298	0.339762	0.219626	0.194553	0.231472
100200	0.266263	0.0267854	0.0423538	1	0.215814	0.189442	0	0.108829	0.0935921	0.210522	0.137005	0.120919	0.185725	0.25726	0.237812	0.218111
100300	0.368639	0.148069	0.269991	0.215814	1	0.232661	0.0748263	0.243982	0.297043	0.107865	0.227114	0.342189	0.439217	0.435494	0.419952	0.383896
100600	0.197142	0.233403	0.112267	0.189442	0.232661	1	0.0776241	0.145725	0.214755	0.162785	0.206732	0.163983	0.252712	0.200448	0.280754	0.243524
100675	0.0678806	0.125448	0.0739974	0	0.0748263	0.0776241	1	0	0	0	0.0935284	0	0.0459457	0	0	0.0567755
100700	0.203997	0.0727986	0.2932	0.108829	0.243982	0.145725	0	1	0.215415	0.0734534	0.125023	0.263145	0.233984	0.321808	0.27193	0.19588
100800	0.238039	0.0895563	0.170992	0.0935921	0.297043	0.214755	0	0.215415	1	0.135493	0.307134	0.204123	0.374577	0.315296	0.379911	0.357129
100820	0.210782	0.0539512	0.0852404	0.210522	0.107865	0.162785	0	0.0734534	0.135493	1	0.150131	0.116732	0.170499	0.139049	0.16466	0.218523
101000	0.15004	0.15665	0.120332	0.137005	0.227114	0.206732	0.0935284	0.125023	0.307134	0.150131	1	0.115617	0.250209	0.208914	0.200225	0.248309
101120	0.368094	0.0315105	0.19298	0.120919	0.342189	0.163983	0	0.263145	0.204123	0.116732	0.115617	1	0.443734	0.537081	0.557657	0.441082
101200	0.408045	0.132105	0.339762	0.185725	0.439217	0.252712	0.0459457	0.233984	0.374577	0.170499	0.250209	0.443734	1	0.558585	0.580955	0.551508
101400	0.444959	0.076906	0.219626	0.25726	0.435494	0.200448	0	0.321808	0.315296	0.139049	0.208914	0.537081	0.558585	1	0.604536	0.492488
101600	0.357986	0.0358535	0.194553	0.237812	0.419952	0.280754	0	0.27193	0.379911	0.16466	0.200225	0.557657	0.580955	0.604536	1	0.504854
101800	0.428706	0.127289	0.231472	0.218111	0.383896	0.243524	0.0567755	0.19588	0.357129	0.218523	0.248309	0.441082	0.551508	0.492488	0.504854	1
101805	0.367318	0.0488374	0.268054	0.0459609	0.318692	0.155837	0	0.268287	0.154322	0.0463446	0.0850573	0.335687	0.333356	0.339157	0.315272	0.335732
101840	0.171618	0.152571	0.14174	0.0724564	0.260278	0.277464	0.150537	0.131717	0.106831	0.0729385	0.110731	0.173489	0.178914	0.13746	0.161523	0.198267
101850	0.135908	0.125154	0.152875	0.0389649	0.207483	0.203991	0.177894	0.106552	0.112115	0.0392769	0.0855087	0.157447	0.171088	0.12391	0.151588	0.183018

- Gene-Disease association network: Firstly filtered geneld-diseaseld table:

```
In [10]: pd.merge(df_for_ge, df2[["MedGen_name", "MIM"]
```

```
Out[10]:
```

	geneld	MIM_number
0	1	229700
1	1	312750
2	1	602535
3	1	230900
4	1	241500
5	1	606054
6	1	300373
7	1	617936
8	1	617936
9	1	608836

To get the OMIM id for disease, another table is joined with another table but there are too many entries: 21330796 rows.

5. Gene-miRNA interaction network: Found less edges than specified in paper. Most of the RNAs are named like this in the dataset: "hsa-mir-942-3p", "hsa-mir-942-5p" i.e. 4 dash separated words. Other dataset has 3 dash separated names for RNA like 'hsa-mir-217'

```
: rna_gene.loc[rna_gene["miRNA"].isin(x)]
```

	miRNA	Target Gene (Entrez Gene ID)
45	hsa-mir-429	9839
46	hsa-mir-429	9839
47	hsa-mir-429	9839
48	hsa-mir-429	9839
49	hsa-mir-429	9839
50	hsa-mir-375	1678
51	hsa-mir-375	1678
52	hsa-mir-375	136319
53	hsa-mir-375	5163
54	hsa-mir-375	5163

6. miRNA-Disease association network: according to the paper 2 datasets are combined. 1st one:

```
In [4]: df1
```

	microRNA ID	OMIM ID
0	hsa-mir-93	186860
1	hsa-mir-93	137215
2	hsa-mir-93	254500
3	hsa-mir-92a	114550
4	hsa-mir-92a	186860
5	hsa-mir-9	155255
6	hsa-mir-9	236000
7	hsa-mir-9	604370
8	hsa-mir-7	114480
9	hsa-mir-7	137800

In the 2nd one to get the OMIM id, another table is joined but found too many entries: 75721 rows (more edges than specified in the paper):

```
In [32]: pd.merge(df2, df3, left_on='disease', right_on='MedGen_name')[["mirnet", "MIM_number"]]
```

Out[32]:

	mirnet	MIM_number
0	miRNet-dis-10	612143
1	miRNet-dis-10	114550
2	miRNet-dis-10	615619
3	miRNet-dis-10	614112
4	miRNet-dis-10	167300
5	miRNet-dis-10	612093
6	miRNet-dis-10	605386
7	miRNet-dis-10	612091
8	miRNet-dis-10	610942
9	miRNet-dis-10	610783