Supplemental Material for "Systematic comparison of graph embedding methods in practical tasks"

Yi-Jiao Zhang,[1] Kai-Cheng Yang,[2] and Filippo Radicchi[2]

[1]*Institute of Computational Physics and Complex Systems, Lanzhou University, Lanzhou, Gansu 730000, China*

[2]*Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana 47408*

I. REAL-WORLD NETWORKS

We list all the real-world networks used in this study together with their basic information in Table S1. Although not all real-world networks display a power-law degree distribution, the application of some hyperbolic embedding methods requires the power-law exponent γ. We therefore use the program shared by Broido *et al.* [1] to estimate a suitable γ value for every network and list the values in Table S1. If the estimated γ value is smaller than 2.1, we set $\gamma = 2.1$.

TABLE S1: **Basic information of real-world networks.** From left to right, we report: name of the network, number of nodes N, number of edges E, average degree $\langle k \rangle$, estimated power-law exponent γ, and network type. The networks marked with * are only used to test the running time and performance on large networks for Node2vec and community embedding.

Network	N	E	$\langle k \rangle$	γ	Type
Social 3 [2]	32	80	5.00	3.46	social
Karate club [3]	34	78	4.59	2.16	social
Protein 2 [2]	53	123	4.64	10.50	biological
Dolphins [4]	62	159	5.13	7.71	social
Social 1 [2]	67	142	4.24	8.56	social
Les Miserables [5]	77	254	6.60	2.10	social
Human brain, layer 2 [6]	78	218	5.59	5.17	biological
Human brain, layer 1 [6]	85	230	5.41	5.14	biological
Protein 1 [2]	95	213	4.48	9.13	biological
E. Coli, transcription [7]	97	212	4.37	5.18	biological
Political books [8]	105	441	8.40	2.62	social
David Copperfield [9]	112	425	7.59	3.62	word adjacency
College football [10]	115	613	10.66	10.50	social
S 208 [2]	122	189	3.10	4.16	electronic circuits
High school, 2011 [11]	126	1,709	27.13	2.10	social
Bay Wet [12]	128	2,075	32.42	2.10	biological
Bay Dry [12]	128	2,106	32.91	2.10	biological
Radoslaw Email [12, 13]	167	3,250	38.92	2.10	social
High school, 2012 [11]	180	2,220	24.67	5.82	social
Little Rock Lake [12, 14]	183	2,434	26.60	2.10	biological
Jazz [15]	198	2,742	27.70	3.07	social
S 420 [2]	252	399	3.17	4.03	electronic circuits
C. Elegans, neural [16]	297	2,148	14.46	3.34	biological
Network Science [9]	379	914	4.82	3.36	social
Dublin [12, 17]	410	2,765	13.49	6.42	social
US Air Transportation [18]	500	2,980	11.92	2.10	transportation
S 838 [2]	512	819	3.20	3.97	electronic circuits
Yeast, transcription [19]	662	1,062	3.21	2.27	biological
European Road [20]	1,039	1,305	2.51	6.36	transportation
URV email [21]	1,133	5,451	9.62	4.89	social
Political blogs [8]	1,222	16,714	27.36	2.38	social
Air traffic [12]	1,226	2,408	3.93	3.71	transportation
Yeast, protein [22]	1,458	1,948	2.67	3.03	biological
Petster, hamster [12]	1,788	12,476	13.96	2.46	social
UC Irvine [12, 23]	1,893	13,835	14.62	2.80	social
Yeast, protein [24]	2,224	6,609	5.94	3.26	biological
Japanese [2]	2,698	7,995	5.93	2.11	word adjacency
Open flights [12, 25]	2,905	15,645	10.77	2.10	transportation
Air Transportation [26]	3,618	14,142	7.82	2.10	transportation
GR-QC, 1993-2003 [27]	4,158	13,422	6.46	2.10	social

US Power grid [16]	4,941	6,594	2.67	7.63	power grid
IPv6 Internet [6]	5,143	13,446	5.23	2.30	Internet
HT09 [17]	5,352	18,481	6.91	2.22	social
Hep-Th, 1995-1999 [28]	5,835	13,815	4.74	5.26	social
Jung [12, 29]	6,120	50,290	16.43	2.34	technological
Gnutella, Aug. 8, 2002 [27, 30]	6,299	20,776	6.60	4.74	p2p
JDK [12]	6,434	53,658	16.68	2.36	technological
AS Oregon Internet [31]	6,474	12,572	3.88	2.10	Internet
English [2]	7,377	44,205	11.98	2.19	word adjacency
Gnutella, Aug. 9, 2002 [27, 30]	8,104	26,008	6.42	4.73	p2p
French [2]	8,308	23,832	5.74	2.29	word adjacency
Hep-Th, 1993-2003 [27]	8,638	24,806	5.74	2.75	social
Gnutella, Aug. 6, 2002 [27, 30]	8,717	31,525	7.23	4.62	p2p
Gnutella, Aug. 5, 2002 [27, 30]	8,842	31,837	7.20	4.65	p2p
PGP [32]	10,680	24,316	4.55	2.24	social
Gnutella, Aug. 4, 2002 [27, 30]	10,876	39,994	7.35	4.58	p2p
Spanish [2]	11,558	43,050	7.45	2.16	word adjacency
DBLP, citations [12, 33]	12,495	49,563	7.93	3.36	social
Spanish [12]	12,643	55,019	8.70	2.17	social
Cond-Mat, 1995-1999 [28]	13,861	44,619	6.44	3.58	social
Cond-Mat, 1993-2003 [27]	21,363	91,286	8.55	3.35	social
Gnutella, Aug. 25, 2002 [27, 30]	22,663	54,693	4.83	7.12	p2p
Internet	22,963	48,436	4.22	2.10	Internet
Cora [12, 34]	23,166	89,157	7.70	3.30	social
AS Internet [35]	23,748	58,414	4.92	2.27	Internet
AS Caida [31]	26,475	53,381	4.03	2.10	Internet
Gnutella, Aug. 24, 2002 [27, 30]	26,498	65,359	4.93	6.27	p2p
Cond-Mat, 1995-2003 [28]	27,519	116,181	8.44	3.62	social
Digg [12, 36]	29,652	84,781	5.72	2.79	social
Cond-Mat, 1995-2005 [28]	36,458	171,735	9.42	2.91	social
Gnutella, Aug. 30, 2002 [27, 30]	36,646	88,303	4.82	4.90	p2p
IPv4 Internet [6]	37,542	95,019	5.06	2.10	Internet
YouTube friend* [12, 37]	1,134,890	2,987,624	5.27	2.14	social
AS Skitter* [31]	1,694,616	11,094,209	13.09	2.25	Internet

II. DETAILED RESULTS FOR DOWNSTREAM TASKS ON REAL-WORLD NETWORKS

A. Mapping accuracy

We only report the aggregate results for mapping accuracy in the main text. Here we list the Spearman's ρ between the pairwise shortest path length in the original graph and the distance in the embedding space for all embedding methods and networks in Table S2. Note that HyperLink cannot process certain networks, so the corresponding results are not shown.

TABLE S2: **Mapping accuracy results on real-world networks.** From left to right, we report: name of the network, the Spearman's ρ obtained by Node2vec, LE, HOPE, Isomap, HyperMap, Mercator, Poincaré maps (PMaps), HyperLink, Hydra, and Community embedding with Infomap and Louvain algorithm. We highlight in bold face the best method for each network. Some results for HyperLink are not shown since it cannot process the corresponding networks.

Network	Node2vec	LE	HOPE	Isomap	HyperMap	Mercator	PMaps	HyperLink	Hydra	Infomap	Louvain
Social 3	0.563	0.387	0.679	**0.925**	0.542	0.655	0.725	0.677	0.751	0.495	0.554
Karate club	0.686	0.421	0.686	**0.933**	0.689	0.681	0.806	0.657	0.889	0.683	0.723
Protein 2	0.913	0.438	0.791	**0.984**	0.614	0.784	0.937	-	0.923	0.878	0.880
Dolphins	0.861	0.418	0.801	**0.976**	0.428	0.645	0.894	0.642	0.818	0.728	0.744
Social 1	0.832	0.540	0.701	**0.956**	0.460	0.713	0.713	0.646	0.777	0.605	0.576
Les Miserables	0.723	0.510	0.766	**0.937**	0.682	0.714	0.671	0.656	0.923	0.681	0.691
Human brain, layer 2	0.887	0.589	0.783	**0.974**	0.623	0.793	0.862	0.740	0.865	0.760	0.736
Human brain, layer 1	0.913	0.382	0.669	**0.983**	0.497	0.723	0.889	0.700	0.795	0.779	0.736
Protein 1	0.943	0.136	0.884	**0.995**	0.467	0.633	0.952	-	0.953	0.946	0.943
E. Coli, transcription	0.960	0.370	0.833	**0.990**	0.557	0.713	0.952	0.607	0.847	0.888	0.848
Political books	0.896	0.241	0.851	**0.966**	0.581	0.587	0.868	0.576	0.669	0.741	0.695
David Copperfield	0.543	0.601	0.749	**0.908**	0.475	0.659	0.587	0.635	0.841	0.460	0.552
College football	0.733	0.062	0.707	**0.871**	0.145	0.493	0.584	-	0.540	0.564	0.564

S 208	0.919	0.348	0.728	**0.974**	0.198	0.590	0.769	0.551	0.685	0.674	0.632
High school, 2011	0.648	0.481	0.671	**0.851**	0.697	0.727	0.725	-	0.685	0.638	0.638
Bay Wet	0.338	0.419	0.418	**0.726**	0.376	0.542	0.493	0.531	0.691	0.418	0.470
Bay Dry	0.338	0.420	0.420	**0.727**	0.420	0.546	0.494	0.533	0.692	0.414	0.481
Radoslaw Email	-0.047	0.735	0.794	**0.876**	0.647	0.789	0.715	0.793	0.868	0.490	0.480
High school, 2012	0.689	0.478	0.692	**0.861**	0.575	0.693	0.667	0.680	0.698	0.610	0.610
Little Rock Lake	0.638	0.505	0.617	**0.863**	0.598	0.657	0.650	0.631	0.804	0.569	0.615
Jazz	0.723	0.575	0.786	**0.900**	0.583	0.716	0.767	0.729	0.808	0.660	0.614
S 420	0.912	0.309	0.717	**0.975**	0.166	0.585	0.748	0.440	0.679	0.650	0.535
C. Elegans, neural	0.605	0.418	0.687	**0.879**	0.380	0.512	0.511	0.532	0.748	0.464	0.396
Network Science	0.811	0.092	0.616	**0.990**	0.353	0.516	0.736	0.422	0.897	0.868	0.786
Dublin	0.878	0.372	0.726	**0.967**	0.572	0.700	0.842	0.623	0.803	0.763	0.737
US Air Transportation	0.758	0.565	0.822	**0.941**	0.612	0.537	0.754	0.608	0.919	0.733	0.738
S 838	0.868	0.333	0.622	**0.979**	0.080	0.596	0.717	0.423	0.659	0.703	0.569
Yeast, transcription	0.862	0.238	0.640	**0.969**	0.369	0.447	0.637	0.424	0.719	0.752	0.637
European Road	0.418	0.437	0.404	**0.996**	-0.010	0.590	0.684	0.214	0.615	0.934	0.861
URV email	0.569	0.586	0.725	**0.909**	0.410	0.685	0.594	0.647	0.806	0.568	0.540
Political blogs	0.246	0.597	0.803	**0.905**	0.616	0.701	0.713	0.686	0.803	0.600	0.612
Air traffic	0.870	0.608	0.600	**0.970**	0.237	0.553	0.698	0.488	0.585	0.699	0.568
Yeast, protein	0.809	0.584	0.412	**0.973**	0.226	0.367	0.604	0.340	0.839	0.695	0.485
Petster, hamster	0.543	0.619	0.709	**0.926**	0.479	0.615	0.654	0.618	0.761	0.608	0.573
UC Irvine	-0.002	0.582	0.795	**0.873**	0.488	0.655	0.590	0.635	0.833	0.521	0.541
Yeast, protein	0.622	0.586	0.617	**0.924**	0.421	0.578	0.584	0.567	0.839	0.622	0.540
Japanese	0.528	0.542	0.774	**0.882**	0.407	0.464	0.497	0.445	0.863	0.467	0.430
Open flights	0.722	0.664	0.451	**0.958**	0.596	0.610	0.704	0.630	0.884	0.753	0.724
Air Transportation	0.776	0.716	0.436	**0.966**	0.533	0.507	0.730	0.520	0.890	0.772	0.695
GR-QC, 1993-2003	0.637	0.593	0.183	**0.955**	0.311	0.404	0.596	0.383	0.813	0.695	0.439
US Power grid	0.165	0.264	0.100	**0.996**	-0.051	0.423	0.444	0.090	0.629	0.900	0.794
IPv6 Internet	0.663	0.541	0.749	**0.908**	0.498	0.555	0.555	0.485	0.903	0.576	0.477
HT09	0.436	0.351	0.692	**0.784**	0.397	0.447	0.398	0.403	0.758	0.317	0.277
Hep-Th, 1995-1999	0.570	0.671	0.183	**0.961**	0.228	0.437	0.646	0.396	0.854	0.730	0.501
Jung	0.354	0.315	0.505	**0.542**	0.113	0.077	0.216	0.076	0.538	0.211	0.164
Gnutella, Aug. 8, 2002	0.334	0.521	0.328	**0.839**	0.343	0.639	0.552	0.587	0.810	0.597	0.500
JDK	0.365	0.328	0.524	**0.559**	0.123	0.104	0.238	0.097	0.557	0.221	0.138
AS Oregon Internet	0.776	0.659	0.733	**0.943**	0.414	0.396	0.620	0.447	0.928	0.625	0.543
English	0.195	0.373	0.793	**0.852**	0.436	0.537	0.469	0.483	0.849	0.405	0.454
Gnutella, Aug. 9, 2002	0.288	0.491	0.300	**0.823**	0.342	0.644	0.556	0.584	0.807	0.580	0.504
French	0.545	0.455	0.721	**0.879**	0.408	0.491	0.504	0.403	0.861	0.437	0.377
Hep-Th, 1993-2003	0.528	0.623	0.199	**0.938**	0.372	0.466	0.619	0.483	0.820	0.675	0.501
Gnutella, Aug. 6, 2002	0.216	0.469	0.391	**0.797**	0.335	0.645	0.561	0.590	0.790	0.516	0.491
Gnutella, Aug. 5, 2002	0.225	0.486	0.382	**0.794**	0.332	0.645	0.569	0.589	0.790	0.524	0.462
PGP	0.438	0.630	0.106	**0.910**	0.134	0.387	0.648	-	0.896	0.747	0.569
Gnutella, Aug. 4, 2002	0.194	0.468	0.510	0.572	0.201	0.653	0.586	-	**0.788**	0.516	0.476
Spanish	0.380	0.331	0.730	0.645	0.448	0.527	0.488	-	**0.810**	0.432	0.422
DBLP, citations	0.361	0.546	0.563	0.732	0.464	0.597	0.564	-	**0.887**	0.629	0.573
Spanish	0.617	0.258	**0.914**	0.767	0.433	0.344	0.640	-	0.892	0.551	0.576
Cond-Mat, 1995-1999	0.391	0.597	0.183	0.734	0.199	0.453	0.621	-	**0.828**	0.704	0.436
Cond-Mat, 1993-2003	0.385	0.527	0.243	0.781	0.355	0.512	0.609	-	**0.869**	0.637	0.473
Gnutella, Aug. 25, 2002	0.258	0.351	0.381	0.589	0.146	0.555	0.474	-	**0.771**	0.548	0.518
Internet	0.710	0.623	0.650	0.837	0.260	0.387	0.563	-	**0.921**	0.607	0.522
Cora	0.441	0.540	0.327	0.751	0.226	0.521	0.604	-	**0.761**	0.645	0.506
AS Internet	0.615	0.511	0.627	0.737	0.476	0.458	0.461	-	**0.901**	0.535	0.475
AS Caida	0.694	0.597	0.644	0.824	0.313	0.392	0.562	-	**0.893**	0.626	0.517
Gnutella, Aug. 24, 2002	0.243	0.304	0.357	0.584	0.139	0.543	0.453	-	**0.801**	0.528	0.503
Cond-Mat, 1995-2003	0.340	0.548	0.163	0.800	0.253	0.503	0.622	-	**0.848**	0.669	0.474
Digg	0.363	0.247	0.623	0.740	0.443	0.610	0.576	-	**0.869**	0.520	0.499
Cond-Mat, 1995-2005	0.340	0.526	0.161	0.817	0.371	0.517	0.634	-	**0.831**	0.666	0.493
Gnutella, Aug. 30, 2002	0.221	0.292	0.302	0.592	0.193	0.543	0.477	-	**0.782**	0.571	0.536
IPv4 Internet	0.593	0.465	0.547	0.682	0.474	0.462	0.429	-	**0.899**	0.558	0.485

We study the effect of embedding dimension on the embedding quality for Isomap, Hydra, and Poincaré maps.

Since all three methods aim to preserve the shortest path length, we use mapping accuracy to directy measure the quality of their embedding results. In Fig. S1(a) we show the CCDF of Spearman's ρ for Isomap with $d = 2$ and $d = 128$, Hydra with $d = 2$ and $d = 128$, and Poincaré maps. Isomap performs better than Hydra when $d = 128$; however, for $d = 2$, both Poincaré maps and Hydra perform better than Isomap. This might be because the two-dimensional Euclidean space is not big enough for large networks. To confirm this hypothesis, we plot the Spearman's ρ against network size for different methods in Fig. S1(b). We can see that the performance of Isomap with $d = 2$ quickly drops as the network size increases. On the other hand, Hydra with $d = 2$ and Poincaré maps are less sensitive to changes in network size.

B. Greedy routing

Similarly, we list the greedy routing scores (GR score) for all embedding methods and networks in the Table S3. We exclude the results for HyperLink on some networks since the embedding cannot be obtained.

TABLE S3: **Greedy routing results on real-world networks.** From left to right, we report: name of the network, the GR score obtained by Node2vec, LE, HOPE, Isomap, HyperMap, Mercator, Poincaré maps (PMaps), HyperLink, Hydra, and Community embedding with Infomap and Louvain algorithm. We highlight in bold face the best method for each network. Some results for HyperLink are not shown since it cannot process the corresponding networks.

Network	Node2vec	LE	HOPE	Isomap	HyperMap	Mercator	PMaps	HyperLink	Hydra	Infomap	Louvain
Social 3	0.818	0.715	0.939	**0.998**	0.911	0.855	0.905	0.897	0.911	0.853	0.882
Karate club	0.894	0.826	0.896	**0.998**	0.933	0.940	0.959	0.957	0.988	0.967	0.974
Protein 2	0.914	0.435	0.941	**0.994**	0.698	0.888	0.960	-	0.935	0.776	0.799
Dolphins	0.872	0.356	0.849	**0.995**	0.675	0.655	0.894	0.688	0.897	0.595	0.641
Social 1	0.944	0.378	0.915	**0.991**	0.816	0.732	0.849	0.710	0.747	0.688	0.681
Les Miserables	0.956	0.704	0.870	**0.999**	0.928	0.983	0.778	0.931	0.990	0.926	0.933
Human brain, layer 2	0.920	0.328	0.777	**0.996**	0.787	0.849	0.901	0.655	0.857	0.723	0.720
Human brain, layer 1	0.945	0.312	0.861	**0.997**	0.715	0.810	0.907	0.653	0.826	0.653	0.599
Protein 1	0.820	0.199	0.947	**0.993**	0.478	0.566	0.820	-	0.797	0.451	0.512
E. Coli, transcription	0.942	0.209	0.760	**0.992**	0.393	0.354	0.915	0.434	0.731	0.418	0.307
Political books	0.911	0.360	0.882	**0.988**	0.521	0.586	0.869	0.627	0.748	0.676	0.658
David Copperfield	0.852	0.595	0.863	**0.992**	0.854	0.868	0.746	0.852	0.899	0.596	0.730
College football	0.903	0.309	0.908	**0.978**	0.642	0.793	0.835	-	0.781	0.767	0.768
S 208	0.951	0.206	0.891	**0.981**	0.481	0.604	0.741	0.624	0.486	0.692	0.551
High school, 2011	0.870	0.624	0.873	**0.992**	0.904	0.942	0.905	-	0.962	0.884	0.882
Bay Wet	0.851	0.909	0.876	**0.989**	0.933	0.968	0.899	0.943	0.965	0.907	0.918
Bay Dry	0.852	0.908	0.878	**0.990**	0.927	0.956	0.887	0.939	0.965	0.908	0.912
Radoslaw Email	0.769	0.985	0.756	**1.000**	0.886	0.970	0.788	0.827	0.999	0.839	0.850
High school, 2012	0.870	0.822	0.871	**0.983**	0.865	0.900	0.840	0.900	0.832	0.844	0.844
Little Rock Lake	0.926	0.701	0.864	**0.985**	0.902	0.946	0.928	0.925	0.959	0.901	0.915
Jazz	0.923	0.876	0.906	**0.996**	0.923	0.894	0.857	0.934	0.918	0.783	0.783
S 420	0.944	0.274	0.852	**0.970**	0.328	0.559	0.738	0.460	0.367	0.608	0.370
C. Elegans, neural	0.929	0.811	0.865	**0.974**	0.829	0.796	0.701	0.821	0.811	0.657	0.628
Network Science	0.779	0.223	0.513	**0.971**	0.473	0.503	0.553	0.500	0.799	0.645	0.577
Dublin	0.908	0.365	0.784	**0.974**	0.621	0.792	0.796	0.666	0.668	0.624	0.489
US Air Transportation	0.856	0.817	0.583	**0.992**	0.544	0.781	0.663	0.856	0.992	0.574	0.661
S 838	0.896	0.249	0.668	**0.954**	0.175	0.477	0.716	0.378	0.203	0.553	0.292
Yeast, transcription	0.946	0.291	0.769	**0.979**	0.441	0.474	0.637	0.526	0.449	0.675	0.455
European Road	0.337	0.143	0.246	**0.879**	0.055	0.109	0.325	0.094	0.063	0.254	0.062
URV email	0.913	0.837	0.931	**0.960**	0.657	0.520	0.446	0.621	0.621	0.557	0.364
Political blogs	0.673	0.901	0.770	**0.970**	0.806	0.741	0.478	0.715	0.889	0.631	0.617
Air traffic	0.894	0.579	0.705	**0.939**	0.375	0.303	0.499	0.404	0.255	0.515	0.229
Yeast, protein	0.842	0.619	0.594	**0.957**	0.297	0.202	0.368	0.293	0.375	0.586	0.271
Petster, hamster	0.879	0.836	0.817	**0.962**	0.715	0.599	0.401	0.687	0.687	0.538	0.424
UC Irvine	0.752	0.886	0.902	**0.959**	0.763	0.623	0.266	0.521	0.780	0.337	0.425
Yeast, protein	0.913	0.761	0.887	**0.957**	0.578	0.331	0.255	0.461	0.533	0.589	0.307
Japanese	0.936	0.899	0.812	**0.959**	0.861	0.768	0.432	0.745	0.953	0.751	0.588
Open flights	0.834	0.775	0.570	**0.929**	0.629	0.650	0.454	0.628	0.854	0.591	0.329
Air Transportation	0.857	0.755	0.563	0.878	0.603	0.649	0.518	0.734	**0.949**	0.631	0.323
GR-QC, 1993-2003	0.684	0.593	0.475	**0.889**	0.261	0.123	0.244	0.245	0.313	0.547	0.198
US Power grid	0.140	0.150	0.066	**0.432**	0.019	0.035	0.028	0.025	0.099	0.162	0.018
IPv6 Internet	0.941	0.837	0.718	0.902	0.818	0.735	0.428	0.804	**0.968**	0.657	0.510
HT09	**0.986**	0.950	0.968	0.950	0.825	0.859	0.524	0.594	0.875	0.810	0.512
Hep-Th, 1995-1999	0.589	0.588	0.428	**0.841**	0.218	0.071	0.179	0.175	0.229	0.416	0.115

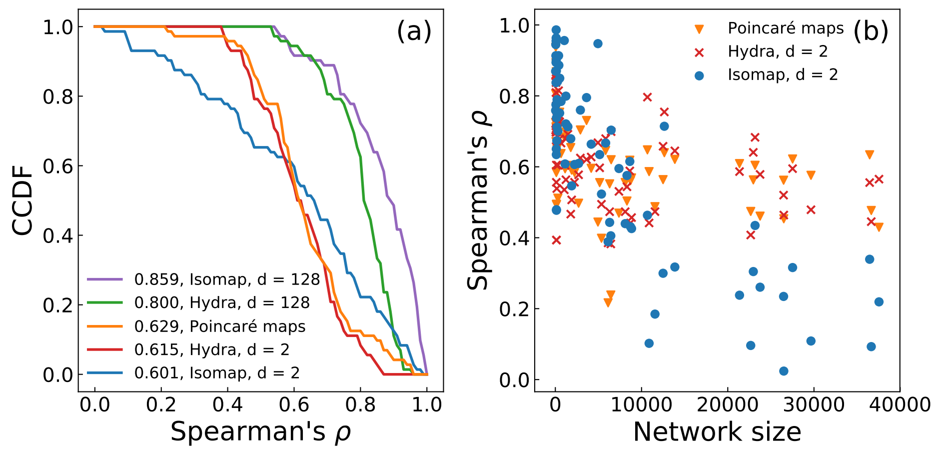

FIG. S1. (a) The CCDF of Spearman's ρ for Isomap with $d = 2$ and 128, Hydra with $d = 2$ and 128, and Poincaré maps. The numbers shown within the legend are the area under the corresponding CCDF curves (CCDF-AUC). (b) The scatter plot of Spearman's ρ and network sizes for all the real-world networks obtained by Poincaré maps, Hydra with $d = 2$, and Isomap with $d = 2$.

Jung	**0.977**	0.969	0.823	0.956	0.965	0.941	0.689	0.963	0.966	0.933	0.935
Gnutella, Aug. 8, 2002	0.833	0.593	0.596	**0.843**	0.397	0.154	0.094	0.165	0.367	0.358	0.114
JDK	**0.975**	0.966	0.824	0.952	0.970	0.940	0.697	0.964	0.969	0.928	0.930
AS Oregon Internet	**0.937**	0.807	0.709	0.874	0.818	0.625	0.513	0.789	0.869	0.652	0.479
English	0.909	0.649	0.675	0.927	0.786	0.719	0.321	0.479	**0.971**	0.519	0.617
Gnutella, Aug. 9, 2002	**0.799**	0.533	0.545	0.785	0.352	0.137	0.079	0.148	0.346	0.388	0.097
French	**0.956**	0.739	0.764	0.896	0.825	0.634	0.357	0.438	0.911	0.625	0.492
Hep-Th, 1993-2003	0.672	0.500	0.545	**0.845**	0.311	0.073	0.141	0.190	0.212	0.436	0.122
Gnutella, Aug. 6, 2002	**0.792**	0.494	0.691	0.734	0.387	0.137	0.076	0.130	0.451	0.397	0.085
Gnutella, Aug. 5, 2002	**0.777**	0.494	0.654	0.742	0.358	0.131	0.074	0.124	0.434	0.393	0.084
PGP	0.365	**0.535**	0.220	0.426	0.237	0.153	0.202	-	0.375	0.433	0.141
Gnutella, Aug. 4, 2002	**0.751**	0.441	0.667	0.412	0.235	0.111	0.059	-	0.542	0.412	0.069
Spanish	0.956	0.650	0.699	0.906	0.792	0.758	0.378	-	**0.960**	0.865	0.598
DBLP, citations	**0.843**	0.634	0.717	0.740	0.589	0.343	0.187	-	0.581	0.509	0.233
Spanish	0.838	0.824	0.813	**0.896**	0.733	0.763	0.620	-	0.872	0.790	0.692
Cond-Mat, 1995-1999	0.425	**0.448**	0.340	0.267	0.192	0.053	0.145	-	0.217	0.430	0.116
Cond-Mat, 1993-2003	**0.655**	0.313	0.459	0.556	0.344	0.065	0.086	-	0.408	0.476	0.140
Gnutella, Aug. 25, 2002	**0.717**	0.230	0.505	0.500	0.133	0.046	0.028	-	0.597	0.489	0.057
Internet	**0.918**	0.667	0.654	0.851	0.693	0.511	0.409	-	0.899	0.712	0.403
Cora	**0.530**	0.401	0.446	0.451	0.189	0.088	0.212	-	0.167	0.356	0.116
AS Internet	**0.966**	0.719	0.673	0.889	0.828	0.617	0.416	-	0.959	0.782	0.416
AS Caida	**0.919**	0.633	0.647	0.841	0.717	0.512	0.410	-	0.865	0.722	0.408
Gnutella, Aug. 24, 2002	**0.721**	0.194	0.565	0.579	0.147	0.053	0.023	-	0.623	0.518	0.051
Cond-Mat, 1995-2003	**0.525**	0.296	0.377	0.493	0.255	0.054	0.082	-	0.357	0.434	0.108
Digg	**0.861**	0.117	0.765	0.701	0.491	0.119	0.025	-	0.602	0.634	0.076
Cond-Mat, 1995-2005	0.560	0.209	0.402	**0.573**	0.265	0.052	0.082	-	0.264	0.428	0.099
Gnutella, Aug. 30, 2002	**0.654**	0.122	0.393	0.534	0.112	0.037	0.018	-	0.545	0.415	0.034
IPv4 Internet	**0.972**	0.679	0.643	0.869	0.832	0.590	0.379	-	0.960	0.789	0.379

C. Link prediction

Again, we list the detailed results, i.e., the ROC-AUC scores, for the link prediction task in Table S4. This task requires removing some edges from the original networks, which might substantially distort the network structure. We therefore only use networks with size $N > 300$. Some results are excluded since the corresponding embeddings are inaccessible.

TABLE S4: **Link prediction results on real-world networks.** From left to right, we report: name of the network, the link prediction ROC-AUC obtained by Node2vec, LE, HOPE, Isomap, HyperMap, Mercator, Poincaré maps (PMaps), HyperLink, Hydra, and Community embedding with Infomap and Louvain algorithm. We highlight in bold face the best method for each network. Some results for HyperLink are not shown since it cannot process the corresponding networks.

Network	Node2vec	LE	HOPE	Isomap	HyperMap	Mercator	PMaps	HyperLink	Hydra	Infomap	Louvain
Network Science	0.958	0.639	0.974	0.973	0.904	0.948	0.964	0.942	0.949	**0.977**	0.966
Dublin	0.901	0.700	0.941	**0.951**	0.845	0.934	0.948	0.923	0.794	0.939	0.937
US Air Transportation	0.854	0.893	0.891	0.954	0.933	0.957	0.905	**0.969**	0.929	0.948	0.954
S 838	0.839	0.476	0.665	0.834	0.598	0.779	0.835	0.666	0.734	**0.856**	0.816
Yeast, transcription	0.747	0.557	0.739	0.839	0.807	**0.874**	0.790	0.865	0.752	0.864	0.864
URV email	0.829	0.796	0.857	**0.911**	0.742	0.886	0.860	0.873	0.817	0.879	0.883
Political blogs	0.880	0.844	0.822	0.922	0.894	0.927	0.909	0.931	0.890	0.931	**0.936**
Air traffic	**0.854**	0.691	0.798	0.840	0.665	0.766	0.795	0.741	0.680	0.846	0.819
Petster, hamster	0.827	0.848	0.873	**0.934**	0.877	0.910	0.893	0.931	0.849	0.907	0.915
UC Irvine	0.726	0.811	0.704	0.883	0.842	**0.909**	0.851	0.893	0.898	0.902	0.896
Yeast, protein	0.808	0.812	0.797	**0.905**	0.817	0.882	0.812	0.886	0.841	0.879	0.888
Japanese	0.539	0.641	0.634	0.850	0.925	**0.927**	0.744	0.920	0.889	0.863	0.922
Open flights	0.969	0.914	0.956	0.981	0.966	0.980	0.939	**0.986**	0.937	0.981	0.980
Air Transportation	0.975	0.912	0.965	0.978	0.965	0.979	0.927	**0.985**	0.949	0.984	0.977
GR-QC, 1993-2003	0.968	0.946	0.958	**0.978**	0.917	0.924	0.903	0.935	0.891	0.977	0.959
IPv6 Internet	0.738	0.700	0.713	0.890	0.930	**0.951**	0.741	0.940	0.907	0.934	0.950
HT09	0.637	0.567	0.636	0.872	0.970	**0.986**	0.779	0.978	0.907	0.952	0.978
Hep-Th, 1995-1999	0.954	0.913	0.932	0.959	0.816	0.878	0.868	0.894	0.864	**0.964**	0.934
Jung	0.798	0.749	0.696	0.684	0.927	0.962	0.827	**0.964**	0.578	0.960	0.956
Gnutella, Aug. 8, 2002	0.638	0.712	0.641	0.809	0.686	**0.826**	0.742	0.793	0.816	0.757	0.823
JDK	0.813	0.761	0.706	0.684	0.931	0.964	0.844	**0.967**	0.594	0.960	0.957
AS Oregon Internet	0.775	0.672	0.645	0.822	0.908	0.926	0.702	0.925	0.787	0.895	**0.929**
English	0.622	0.747	0.683	0.881	0.942	**0.951**	0.848	0.943	0.918	0.903	0.947
Gnutella, Aug. 9, 2002	0.622	0.694	0.625	0.816	0.679	**0.833**	0.735	0.794	0.825	0.758	0.830
French	0.568	0.659	0.612	0.859	0.912	**0.913**	0.732	0.910	0.883	0.870	0.912
Hep-Th, 1993-2003	0.946	0.871	0.913	0.944	0.854	0.865	0.869	0.900	0.858	**0.960**	0.924
Gnutella, Aug. 6, 2002	0.593	0.669	0.575	0.778	0.647	**0.790**	0.722	0.752	0.787	0.741	0.782
Gnutella, Aug. 5, 2002	0.587	0.670	0.581	0.778	0.650	0.788	0.728	0.749	**0.789**	0.733	0.787
PGP	0.986	0.962	0.965	0.975	0.948	0.960	0.940	-	0.939	**0.991**	0.984
Gnutella, Aug. 4, 2002	0.602	0.659	0.581	0.698	0.607	0.793	0.731	-	**0.795**	0.747	0.778
Spanish	0.553	0.696	0.665	0.676	0.953	**0.963**	0.812	-	0.916	0.904	0.962
DBLP, citations	**0.956**	0.866	0.903	0.853	0.884	0.948	0.898	-	0.913	0.956	0.948
Spanish	0.694	0.556	0.632	0.459	0.948	0.957	0.790	-	0.758	0.890	**0.962**
Cond-Mat, 1995-1999	0.976	0.920	0.963	0.965	0.876	0.904	0.908	-	0.892	**0.984**	0.955
Cond-Mat, 1993-2003	0.961	0.854	0.946	0.954	0.879	0.898	0.905	-	0.891	**0.979**	0.935
Gnutella, Aug. 25, 2002	0.553	0.625	0.573	0.747	0.588	0.824	0.650	-	**0.839**	0.776	0.839
Internet	0.862	0.750	0.731	0.766	0.936	0.939	0.786	-	0.864	0.945	**0.966**
Cora	0.974	0.917	0.957	0.922	0.839	0.907	0.933	-	0.804	**0.976**	0.954
AS Internet	0.846	0.737	0.770	0.788	0.953	0.955	0.787	-	0.915	0.961	**0.974**
AS Caida	0.850	0.746	0.693	0.772	0.937	0.935	0.775	-	0.838	0.943	**0.969**
Gnutella, Aug. 24, 2002	0.534	0.607	0.536	0.663	0.598	0.827	0.646	-	0.804	0.778	**0.839**
Cond-Mat, 1995-2003	0.970	0.882	0.957	0.968	0.882	0.907	0.918	-	0.901	**0.985**	0.945
Digg	0.541	0.588	0.605	0.804	0.834	**0.882**	0.760	-	0.876	0.818	0.868
Cond-Mat, 1995-2005	0.965	0.870	0.955	0.968	0.900	0.906	0.917	-	0.894	**0.984**	0.943
Gnutella, Aug. 30, 2002	0.590	0.614	0.603	0.774	0.625	0.841	0.678	-	0.851	0.796	**0.852**
IPv4 Internet	0.860	0.755	0.766	0.782	0.966	0.961	0.802	-	0.933	0.970	**0.976**

D. AUPR and Precision results for link prediction

Other than ROC-AUC score, the area under the precision-recall curve (AUPR) and precision are also common metrics to evaluate the performance of link prediction. We measure the AUPR and precision value for each method on real-world networks with size $300 < N < 10,000$ in our dataset. The results are qualitatively similar with those using ROC-AUC score. The CCDF-AUC value of ROC-AUC, AUPR and precision for each method are reported in Table S5.

E. Robustness check on a subset of the networks

The CCDF-AUC value depends on the number of networks considered. Since the results are not accessible for HyperLink on some networks, the direct comparison of CCDF-AUC obtained by aggregating all available results in the main text is not completely fair. To make sure the findings are robust, here we re-calculate the CCDF-AUC values of the three tasks using the same subset of real-world networks that all methods can properly embed in Table S6. The results are qualitatively similar to those in the main text.

TABLE S5. The CCDF-AUC value of AUPR, Precision, and ROC-AUC on the real-world networks with size $300 < N < 10,000$. From left to right, we report: name of the method, CCDF-AUC of AUPR, CCDF-AUC of Precision, and CCDF-AUC of ROC-AUC. For each metric, we highlight in bold face the CCDF-AUC values of the top three embedding methods.).

Method	AUPR	Precision	ROC-AUC
Node2vec	0.818	0.738	0.784
HOPE	0.753	0.720	0.767
LE	0.742	0.685	0.741
Isomap	0.865	0.811	0.876
HyperMap	0.834	0.774	0.841
Mercator	**0.895**	**0.835**	**0.899**
Poincaré maps	0.825	0.763	0.827
HyperLink	0.895	0.823	0.891
Hydra	0.818	0.770	0.833
Comm. (Infomap)	**0.901**	**0.827**	**0.895**
Comm. (Louvain)	**0.905**	**0.842**	**0.908**

TABLE S6. The comparison of CCDF-AUC for mapping accuary, greedy routing, and link prediction on the subset of real-world networks that all embedding methods can properly process. From left to right, we report: name of the method, CCDF-AUC for mapping accuracy, CCDF-AUC for greedy routing, and CCDF-AUC for link prediction. For each task, we highlight in bold face the CCDF-AUC values of the top three embedding methods. The CCDF-AUC values are generated by aggregating the performance on 50 real-world networks for mapping accuracy and greedy routing and on 28 real-world networks for link prediction (networks with $N < 300$ are excluded).

Method	Mapping accuracy	Greedy routing	Link prediction
Node2vec	0.591	**0.845**	0.783
HOPE	0.606	**0.739**	0.770
LE	0.478	0.621	0.745
Isomap	**0.900**	**0.931**	0.875
HyperMap	0.410	0.630	0.840
Mercator	0.570	0.600	**0.900**
Poincaré maps	**0.641**	0.566	0.829
HyperLink	0.526	0.596	0.891
Hydra	**0.784**	0.698	0.831
Comm. (Infomap)	0.616	0.630	**0.896**
Comm. (Louvain)	0.567	0.512	**0.907**

III. PARAMETERS SELECTION

Most of the embedding methods considered in this study have multiple free parameters that could affect the embedding results (see Table S7). In theory, it is possible to use techniques like grid search to find a combination that yields the best results for a given task and network. But this process can be extremely time-consuming and not feasible in practical use. For simplicity and fairness, we deploy a hybrid strategy by combining heuristics and grid search to find a set of parameters for each method that yield decent results in general for all tasks. We then fix these parameters for different experiments.

Following the literature and the experience we gain in our experiments, we first isolate the parameters that have little affect on the embedding result and use the default values for them. Examples inlcude the bias parameters

TABLE S7. **The free parameters of different embedding methods.** From left to right, we report: the name of the embedding method, number of free parameters (Num.) and the parameters. For the parameters that have little affect on the embedding performance or that can be clearly explained, we directly point out in the table. The parameters that shown in bold face need to be analyzed specifically. In the table, d is embedding dimension, T is temperature, p and q are bias parameters of the random walk process, β is a free parameter to calculate the fitness for community-based embedding.

Method	Num.	Parameters
Node2vec	5	$p = 1$, $q = 1$, window size = 10, $d = \min\{N, 128\}$, **walk length** = 100 or 10
HOPE	1	$d = \min\{N, 128\}$
LE	1	$d = \min\{N, 128\}$
Isomap	1	$d = \min\{N, 128\}$
HyperMap	1	$\boldsymbol{T} = 0.5$
Mercator	0	-
Poincaré maps	2	Gaussian kernel width $\sigma_P = 1$, scaling parameter $\gamma_P = 2$
HyperLink	3	number of node layers $m = 20$, grid multiplier $g = 1$, $\boldsymbol{T} = 0.3$
Hydra	1	$d = \min\{N, 128\}$
Comm. (Infomap)	1	$\boldsymbol{\beta} = 0.3$
Comm. (Louvain)	1	$\boldsymbol{\beta} = 0.3$

p, q and window size for Node2vec, the Gaussian kernel width σ_P and scaling parameter γ_P for Poincaré maps, the number of node layers m, and the grid multiplier g that controls the resolution of angular coordinates inference for HyperLink.

Other parameters may substantially alter the embedding. A common parameter shared by many embedding methods is the embedding dimension d. A larger d value usually leads to better embedding quality, but the gain quickly saturates as d increases. To balance the tradeoff between embedding quality and computational complexity, we set dimension $d = \min\{N, 128\}$ in our experiments for all methods that can work with high dimensional spaces. The parameter N is the network size. The remaining critical parameters are temperature T for HyperMap and HyperLink, β for community-based embedding, and walk length for Node2vec. We discuss their effect on the embeddings and how we select the values separately.

A. Effect of temperature T on HyperMap

To find a value of T that can provide best overall performance for different networks and tasks, we use all the networks with size small than 10,000 in our dataset as the test set, scan the T and obtain the results of three

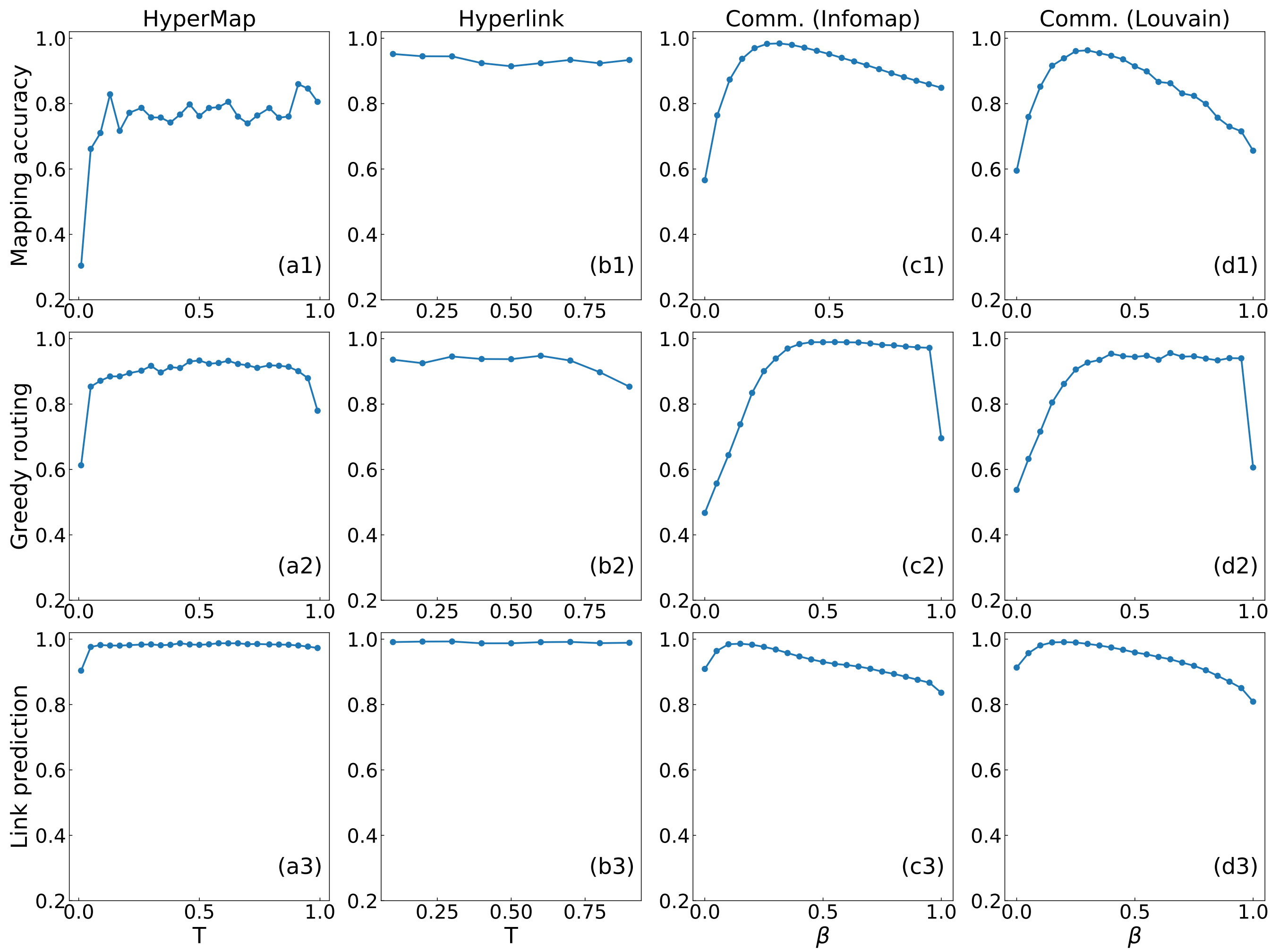

FIG. S2. The average performance of HyperMap (temperature, T), HyperLink (T), and community embedding with Infomap and Louvain algorithm (β) on the three downstream tasks as the function of the tunable parameters for each embedding methods. The results for HyperMap and community embeddings are obtained on all networks with size N smaller than $10,000$ in our dataset. The results for Hyperlink are obtained on all networks with $N < 5,000$ in our dataset.

downstream tasks. Then we calculate the average performance for each T value and task using the following formula

$$\langle S(T)\rangle = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\frac{S_x(T)}{\max_{T'} S_x(T')}\ , \tag{S1}$$

where $S_x(T)$ is the performance obtained by HyperMap in a given task on network x with temperature T, $\mathcal{X}$ is the set of the networks we use.

The average performance $\langle S(T)\rangle$ against T for HyperMap in three downstream tasks are shown in Fig. S2 (a1-a3). We find that the average performance of all the three tasks are good for HyperMap if the temperature is not too large or too small. Therefore, we set $T = 0.5$ for HyperMap.

B. Effect of temperature T on HyperLink

Following the same approach described in Sec. III A, the effects of the temperature T on the average performance of the downstream tasks for HyperLink are shown in Fig. S2 (b1-b3). The results in Fig S2 (b1-b3) are obtained on all networks smaller than $5,000$ in our dataset. We can see that temperature T do not affect the performance of the downstream tasks too much. But $T = 0.3$ provides the overall best results for all three tasks. Therefore we set $T = 0.3$ for HyperLink in all the experiments.

C. Effect of β on community-based embedding

We follow the same approach in Sec. III A and get the average performance of community-based embeddings (with Infomap and Louvain algorithms) against β on the three downstream tasks (Fig. S2(c1-c3) and Fig. S2(d1-d3)). The optimal value of β depends on the downstream task. For simplicity, we set β to 0.3 for all tasks since it yields optimal or near-optimal results for all tasks at the same time.

D. Effect of walk length on Node2vec

Our experiments based on three real-world networks show that increasing walk length has opposite effects in different tasks, see Fig. S3. Specially, mapping accuracy and greedy routing benefit from larger walk length while link prediction suffers from it. Therefore, we set walk length to 100 for mapping accuracy and greedy routing, and walk length to 10 for link prediction in our experiment.

E. Effect of distance calculation on Node2vec and HOPE

There are different approaches to calculate distance in Euclidean space. Although this is not an ordinary parameter, different choices do lead to different embedding results for some embedding methods. The definition of the algorithms suggests that dot product should be used for Node2vec and HOPE regardless of the tasks. However, we find that using Euclidean distance produces way better results for mapping accuracy and greedy routing (Fig. S4). We therefore use Euclidean distance in mapping accuracy and greedy routing for Node2vec and HOPE; dot product is only used in link prediction.

IV. EFFECT OF NETWORK CHARACTERISTICS ON EMBEDDING RESULTS

Here we show the effect of the average clustering coefficient of PSO model, the modularity of LFR model, the power-law exponent of power-law networks and spatial network, and the average degree of Poisson networks on the performance of three downstream tasks.

We plot the average values of Spearman's ρ, GR score, and ROC-AUC for different embedding methods while varying the parameters of the PSO model, LFR model, power-law networks, and spatial networks in Fig. S5. We find that the overall performance of all the methods is heavily affected by the the varying network characteristics. But the rankings of different methods are relatively stable.

We also explore the relation between the performance of link prediction and greedy routing for different methods on synthetic networks. The results are shown in Fig. S6. The observed patterns are qualitatively similar to those in the main text.

V. CONNECTION TO CENTRALITY MEASURES

The radial coordinates or the distance of nodes from the geometric center r of the embeddings usually convey centrality information. Here we use some real-world networks (networks with size $1000 < N < 5000$ in our dataset) to empirically study their connection. The average value of the Spearman's correlation coefficient are shown in Fig. S7. Clearly, the radial coordinates of HyperMap, Mercator, and HyperLink represent the degree of nodes. The distance to the geometric center r of Isomap, Hydra, and Poincaré maps are highly correlated with their closeness centrality. For HOPE and LE, the closeness and eigenvector centrality both show highl correlation with r. We don't find obvious relation between node centrality and the distance r in Node2vec embedding.

[1] A. D. Broido and A. Clauset, Nature communications **10**, 1017 (2019).

[2] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science **303**, 1538 (2004), https://science.sciencemag.org/content/303/5663/1538.full.pdf.

[3] W. W. Zachary, Journal of anthropological research **33**, 452 (1977).

[4] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, Behavioral Ecology and Sociobiology **54**, 396 (2003).

[5] D. E. Knuth, *The Stanford GraphBase: a platform for combinatorial computing*, Vol. 1 (Association for Computing Machinery, New York, NY, USA, 1993).

[6] K.-K. Kleineberg, M. Boguná, M. Á. Serrano, and F. Papadopoulos, Nature Physics **12**, 1076 (2016).

[7] S. Mangan and U. Alon, Proceedings of the National Academy of Sciences **100**, 11980 (2003), https://www.pnas.org/content/100/21/11980.full.pdf.

[8] L. A. Adamic and N. Glance, in *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD '05 (Association for Computing Machinery, New York, NY, USA, 2005) p. 3643.

[9] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[10] M. Girvan and M. E. J. Newman, Proceedings of the National Academy of Sciences **99**, 7821 (2002), https://www.pnas.org/content/99/12/7821.full.pdf.

[11] J. Fournet and A. Barrat, PLOS ONE **9**, 1 (2014).

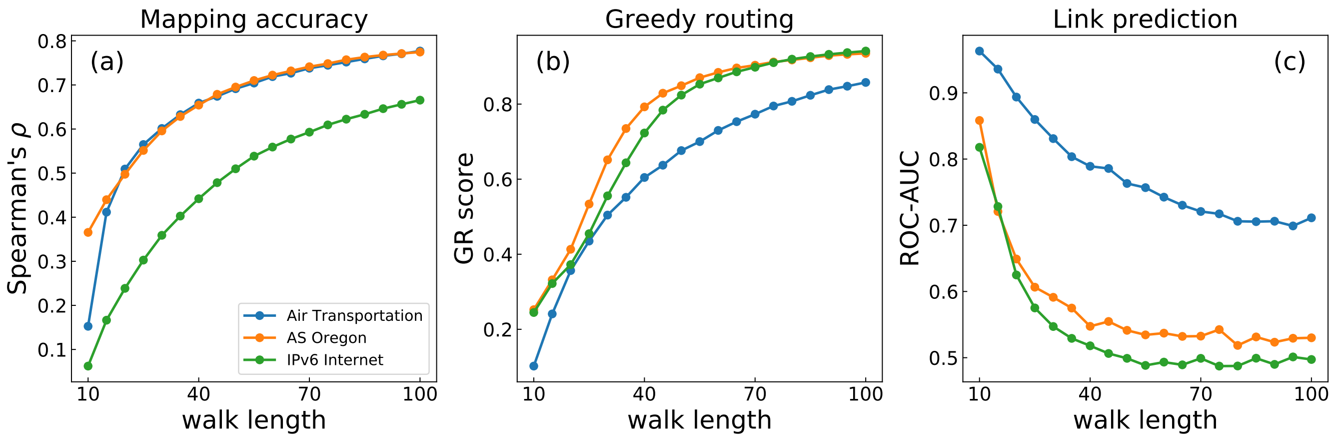

FIG. S3. (a) Spearman's ρ (b) GR score (c) Link prediction ROC-AUC as the function of the walk length of node2vec on Air Transportation, AS Oregon network, and IPv6 Internet network.

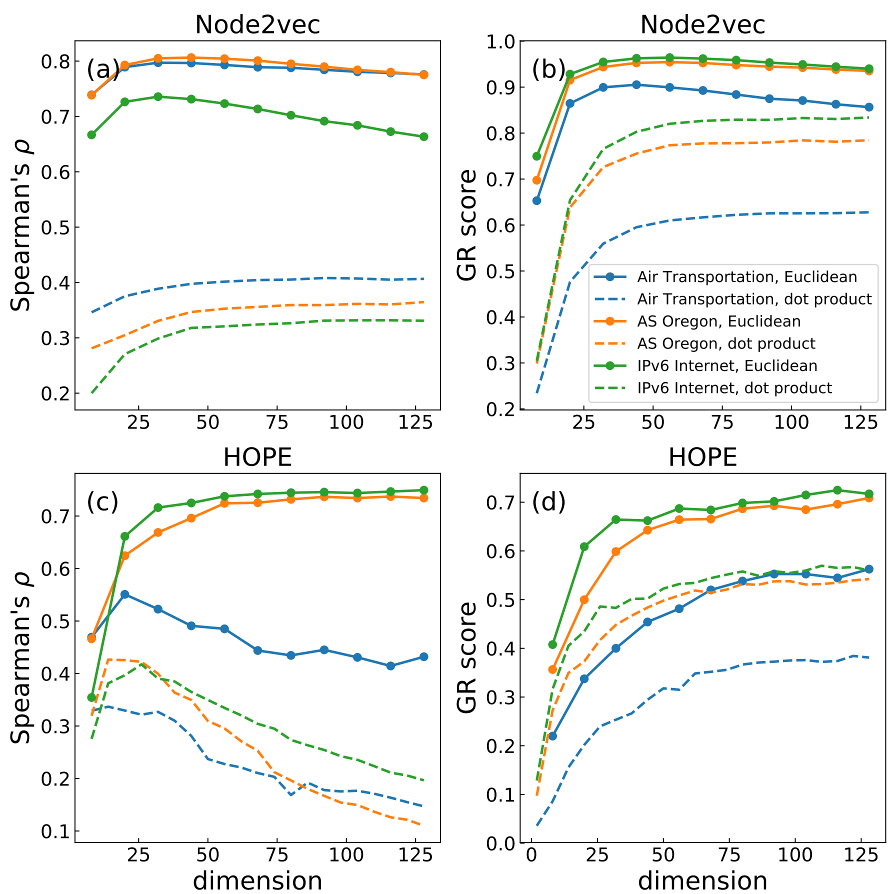

FIG. S4. (a) Spearman's ρ (b) GR score as the function of the dimension of node2vec, (c) Spearman's ρ (b) GR score as the function of the dimension of HOPE on Air Transportation, AS Oregon network, and IPv6 Internet network with different distance calculation approaches, i.e., Euclidean distance and dot product.

[12] J. Kunegis, in *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion (Association for Computing Machinery, New York, NY, USA, 2013) p. 13431350.

[13] R. Michalski, S. Palus, and P. Kazienko, in *Business Information Systems*, edited by W. Abramowicz (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011) pp. 197–206.

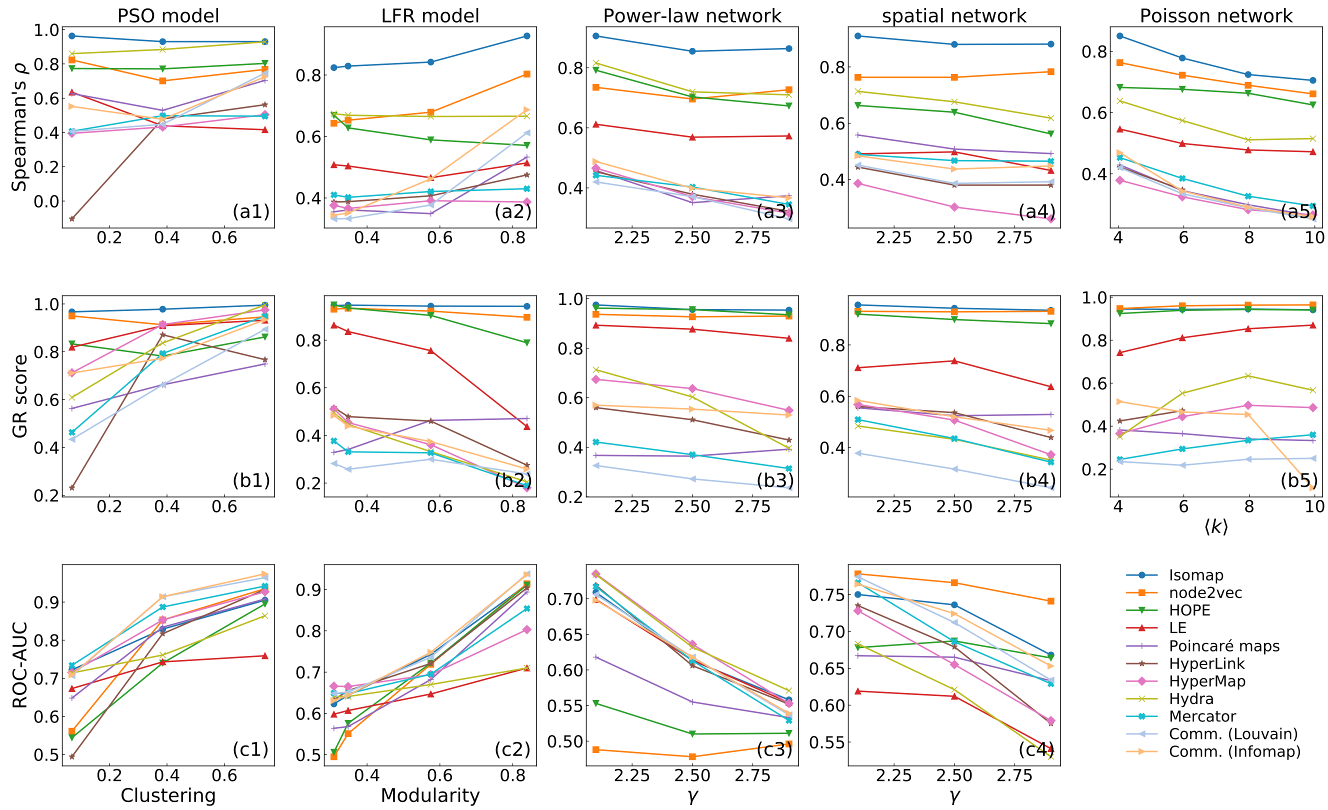

FIG. S5. (a1)-(a5) The Spearman's ρ, (b1)-(b5) GR score, (c1)-(c4) Link prediction ROC-AUC obtained by different embedding methods on PSO model networks, LFR model networks, Power-law networks, spatial networks, and Poisson networks with varying network parameters, respectively.

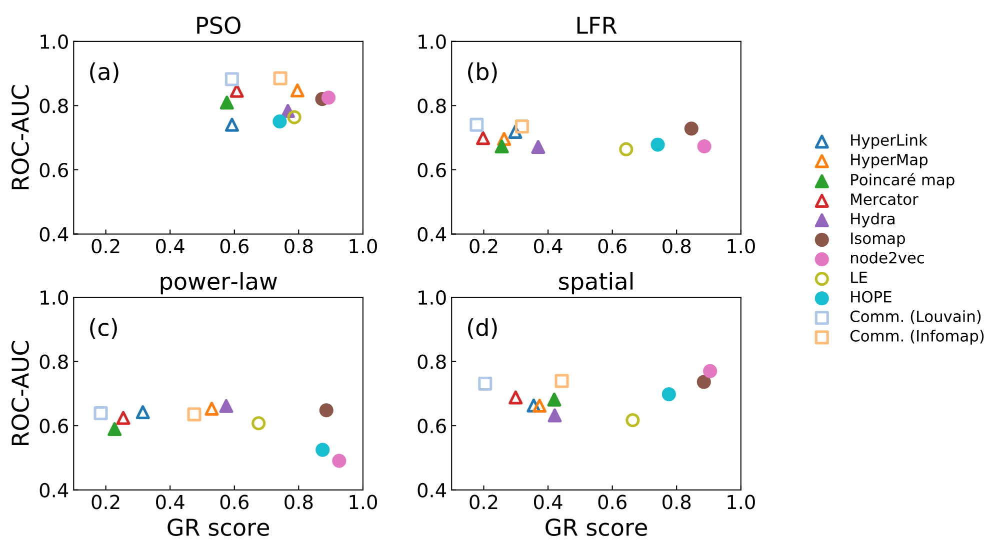

FIG. S6. The CCDF-AUC of ROC-AUC against GR score for different embedding methods on (a) PSO model networks, (b) LFR model networks, (c) power-law networks, and (d) spatial networks. Circles, triangles and squares represent Euclidean-, hyperbolic- and community-based embeddings, respectively. The hollow and solid symbols represent methods that preserve local and global network structure information, respectively.

[14] N. D. Martinez, Ecological Monographs **61**, 367 (1991), https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/2937047.

[15] P. M. GLEISER and L. DANON, Advances in Complex Systems **06**, 565 (2003),

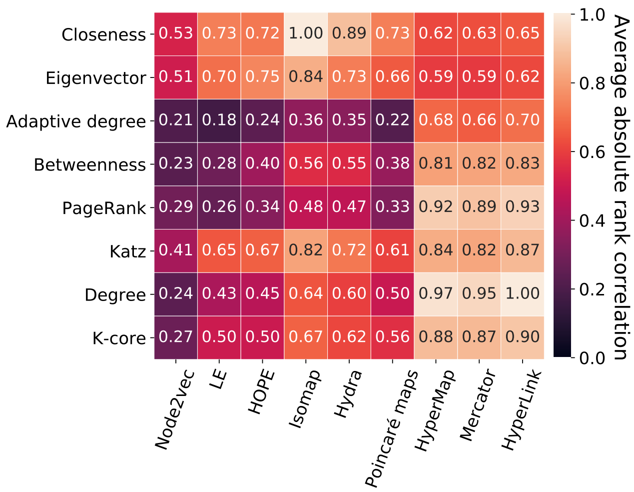

FIG. S7. Pairwise Spearman's correlation coefficient between different centrality measures and the distance of a node from the geometric center of different embeddings. The values are the average results of 13 real-world networks (networks with size $1000 < N < 5000$ in our dataset).

https://doi.org/10.1142/S0219525903001067.

[16] D. J. Watts and S. H. Strogatz, nature **393**, 440 (1998).

[17] L. Isella, J. Stehl, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, Journal of Theoretical Biology **271**, 166 (2011).

[18] V. Colizza, R. Pastor-Satorras, and A. Vespignani, Nature Physics **3**, 276 (2007).

[19] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002), https://science.sciencemag.org/content/298/5594/824.full.pdf.

[20] L. Šubelj and M. Bajec, The European Physical Journal B **81**, 353 (2011).

[21] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Phys. Rev. E **68**, 065103 (2003).

[22] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature **411**, 41 (2001).

[23] T. Opsahl and P. Panzarasa, Social Networks **31**, 155 (2009).

[24] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen, Nucleic Acids Research **31**, 2443 (2003), https://academic.oup.com/nar/article-pdf/31/9/2443/3773904/gkg340.pdf.

[25] T. Opsahl, F. Agneessens, and J. Skvoretz, Social Networks **32**, 245 (2010).

[26] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, Proceedings of the National Academy of Sciences **102**, 7794 (2005).

[27] J. Leskovec, J. Kleinberg, and C. Faloutsos, ACM Trans. Knowl. Discov. Data **1**, 2es (2007).

[28] M. Newman, "The structure of scientific collaboration networks," in *The Structure and Dynamics of Networks* (Princeton University Press, 2011) pp. 221–226.

[29] L. Šubelj and M. Bajec, in *Proceedings of the First International Workshop on Software Mining*, SoftwareMining '12 (Association for Computing Machinery, New York, NY, USA, 2012) p. 916.

[30] M. Ripeanu and I. T. Foster, in *Peer-to-Peer Systems*, edited by P. Druschel, M. F. Kaashoek, and A. I. T. Rowstron (Springer, Berlin, Heidelberg, 2002) pp. 85–93.

[31] J. Leskovec, J. M. Kleinberg, and C. Faloutsos, in *Proceedings of the Eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, edited by R. Grossman, R. J. Bayardo, and K. P. Bennett (ACM, New York, NY, 2005) pp. 177–187.

[32] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, Phys. Rev. E **70**, 056122 (2004).

[33] M. Ley, in *String Processing and Information Retrieval*, edited by A. H. F. Laender and A. L. Oliveira (Springer Berlin Heidelberg, Berlin, Heidelberg, 2002) pp. 1–10.

[34] L. Šubelj and M. Bajec, in *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion (Association for Computing Machinery, New York, NY, USA, 2013) p. 527530.

[35] M. Boguná, F. Papadopoulos, and D. Krioukov, Nature communications **1**, 62 (2010).

[36] M. De Choudhury, H. Sundaram, A. John, and D. D. Seligmann, in *2009 International Conference on Computational Science and Engineering*, Vol. 4 (2009) pp. 151–158.

[37] J. Yang and J. Leskovec, Knowledge and Information Systems **42**, 181 (2015).