

Bacterial protein interaction networks: connectivity is ruled by gene conservation, essentiality and function.

Maddalena Dilucca¹, Giulio Cimini^{2,3}, Andrea Giansanti^{1,4}

1 Dipartimento di Fisica, Sapienza University of Rome, Rome, Italy

2 IMT School for Advanced Studies, Lucca, Italy

3 Istituto dei Sistemi Complessi (ISC)-CNR, Rome, Italy

4 INFN Roma1 unit, Rome, Italy

Abstract

Protein-protein interaction (PPI) networks are the backbone of all processes in living cells. In this work we study how conservation, essentiality and functional repertoire of a gene relate to the connectivity k of the corresponding protein in the PPI networks. Focusing on a set of 42 mostly distantly related bacterial species, we investigate three issues: i) whether the distribution of connectivity values changes between PPI subnetworks of essential and nonessential genes; ii) how gene conservation, measured both by the evolutionary retention index (ERI) and by evolutionary pressure (as represented by the ratio K_a/K_s) is related to the connectivity of the corresponding protein; iii) how PPI connectivities are modulated by evolutionary and functional relationships, as represented by the Clusters of Orthologous Proteins (COGs). We show that conservation, essentiality and functional specialization of genes control in a universal way the topology of the emerging bacterial PPI networks. Noteworthy, a structural transition in the network is observed such that, for connectivities $k \geq 40$, bacterial PPI networks are mostly populated by genes that are conserved, essential and which, in most cases, belong to the COG cluster J, related to ribosomal functions and to the processing of genetic information.

Introduction

To operate biological activities in living cells, proteins work in association with other proteins, possibly assembled in large complexes. Hence, knowing the interactions of a protein is important to understand its cellular functions. Moreover, a comprehensive description of the stable and transient protein-protein interactions (PPIs) within a cell would facilitate the functional annotation of all gene products, and provide insight into the higher-order organization of the proteome [1, 2]. Several methodologies have been developed to detect PPIs, and have been adapted to chart interactions at a proteome-wide scale. These methods, that combine different technologies with complementary experiments and computational analyses, were shown to generate high-confidence PPI networks, enabling the assignment of several proteins to functional categories [3, 4].

Moreover, the statistical study of bacterial PPIs over several species (meta-interactomes) has brought important knowledge about protein functions and cellular processes [5, 6]. This work contributes to this line of research. Our aim is to shed light on the relationship between conservation, essentiality and functional annotation at the genetic level with the connectivity patterns of the PPI networks. We

extend previous observations which suggested a strong correlation between codon bias and the topology of PPI networks on the one hand, and between codon bias and gene conservation/essentiality on the other hand [7, 8]. It is worth, in the next two paragraphs, to make more precise what is usually meant by gene essentiality and gene conservation.

Individual genes in the genome contribute differentially to the survival of an organism. According to their known functional profiles and based on experimental evidence, genes can be divided into two categories: essential and nonessential ones [9, 10]. Essential genes are not dispensable for the survival of an organism in the environment it lives in [10, 11]. Nonessential genes are instead those which are dispensable [12], being related to functions that can be silenced without compromising the survival of the organism. Naturally, each species has adapted to one or more evolving environments and, plausibly, genes that are essential for one species may be not essential for another one.

It has been argued many times that essential genes are more conserved than nonessential ones [13–17]. The term “conservation” has however at least two meanings. On the one hand, a gene is conserved if orthologous copies are found in the genomes of many species, as measured by the Evolutionary Retention Index (ERI) [9, 18]. On the other hand, a gene is (evolutionarily) conserved when it is subject to a purifying evolutionary pressure which disfavors mutations. This pressure is usually measured by K_a/K_s , the ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site. In this second meaning a conserved gene is, in a nutshell, a slowly evolving gene [13, 19].

In this work we show that bacterial PPI networks display an interesting topological-functional transition, ruled by protein connectivity k and with a threshold between $k = 40$ and $k = 50$. Proteins with high PPI network connectivities (hubs) likely correspond to genes that are conserved and essential. Conversely, genes that correspond to hub proteins in the PPI network are likely to be essential and conserved. Additionally, below the threshold the functional repertoire of proteins is heterogeneous, whereas, above the threshold there is a quite strict functional specialization.

Materials and Methods

We consider a set of 42 selected bacterial genomes (that we have previously investigated in [8]), reported in Table 1. These species were chosen in order to have the largest possible coverage of available data for what concerns conservation, essentiality and selective pressure of genes. Nucleotide sequences were downloaded from the FTP server of the National Center for Biotechnology Information. [20].

Gene Conservation

We use the Evolutionary Retention Index (ERI) [9] as a first way of measuring conservation of a gene. The ERI of a gene is the fraction of genomes (among those reported in Table 1) that have at least an ortholog of the given gene. Then, a low ERI value means that a gene is specific, common to a small number of genomes, whereas high ERI is a characteristic of highly shared, putatively universal and essential genes.

We also make reference to another notion of gene conservation. Conserved genes are those which are subject to a purifying, conservative evolutionary pressure. To discriminate between genes subject to purifying selection and genes subject to positive selective Darwinian evolution, we use a classic but still widely used indicator, the ratio K_a/K_s between the number of nonsynonymous substitutions per nonsynonymous site (K_a) and the number of synonymous substitutions per synonymous site (K_s) [19].

Conserved genes are characterized by $K_a/K_s < 1$. We used K_a/K_s estimates by Luo [15] that are based on the method by Nej and Gojobori [21].

Gene Essentiality

We used the Database of Essential Genes (DEG, www.essentialgene.org) [15], which classifies a gene as either essential or nonessential on the basis of a combination of experimental evidence (null mutations or transposons) and general functional considerations. DEG collects genomes from Bacteria, Archea and Eukarya, with different degrees of coverage [22,23]. Of the 42 bacterial genomes we consider, only 23 are covered—in toto or partially—by DEG, as indicated in Table 1.

Protein-Protein Interaction Network

PPIs are obtained from the STRING database (Known and Predicted Protein-Protein Interactions, <https://string-db.org/>) [24]. We have chosen STRING because it is the only available database with 1678 bacteria species, thus useful to extend analysis such as those performed in [7] to multiple species. The problem with STRING (functional associations might be used to represent both physical protein-protein molecular interactions and more abstract meta-interactome interactions with it our best to modulate our results using the parameter w , offered by STRING to its users. In STRING, each interaction is assigned with a confidence level or probability w , evaluated by comparing predictions obtained by different techniques [25–27] with a set of reference associations, namely the functional groups of KEGG (Kyoto Encyclopedia of Genes and Genomes) [28]. In this way, interactions with high w are likely to be true positives, whereas, a low w possibly corresponds to a false positive. As usually done in the literature, we consider only interactions with $w \geq \Theta$ and select a stringent cut-off $\Theta = 0.9$ that allows for a fair balance between coverage and interaction reliability (see for instance the case of *E.coli* reported in [7]).

We denote by k the *degree* (number of connections) associated to each proteins in each PPI network after the thresholding procedure. Note also that after applying the cut-off we are left, for each network, with a number of isolated proteins (with no connections) that grows as \sqrt{n} (where n is the number of proteins in the genome). These proteins are not considered in the network analysis and are regarded as stemming from statistical noise.

Clusters of orthologous proteins

We use functional annotation given by the popular database of orthologous groups of proteins (COGs) from Koonin's group, available at <http://ncbi.nlm.nih.gov/COG/> [29,30]. We consider 15 functional COG categories (see Table 2), excluding the generic categories R and S for which functional annotation is too general or missing.

Results and discussion

Degree distribution of PPI networks. We start by studying the degree distributions $P(k)$ observed in bacterial PPIs. We first recall that such a distribution was found to be scale-free in *E.coli* [7,31–33], meaning that the corresponding PPI network features a large number of poorly connected proteins, and a relatively small number of highly connected hubs. In order to assess the generality of this observation, we compute $P(k)$ for each genome of Table 1 (plots are reported in figures S2-S3 of the

Supplementary Information). Note that, despite the fact that PPIs of different bacteria have different sizes and densities, their average connectivity and the support of their $P(k)$ are very similar (as shown in figure S1 of the Supplementary Information). Thus, we can superpose all the considered bacterial degree distributions without the need to normalise the support of each $P(k)$. When doing so, we observe two distinct regimes (see figure 1). For low values of k , the distribution has a scale-free shape $P(k) \propto k^{-\gamma}$. This result is consistent with previous findings for yeast, worm and fly [34] and for co-conserved PPIs in some bacteria [35]. For higher values of k , however, the distribution deviates from a power law, and a bump with a Gaussian-like shape emerges. Interestingly, this feature is almost undetectable taking individual species alone as *E.coli* (see fig 2 in supporting materials of Dilucca [7]), but clearly emerges when the statistics is enriched by adding together several species. The bump emerging for $k \geq 40$ is reasonably due to the contribution of proteins belonging to complexes [36]. Indeed, if we consider the separate contribution of essential and nonessential genes to the $P(k)$ (for DEG-annotated genomes), we see that the superposed peak is present only in the degree distribution of essential genes. Moreover, the degree distributions for essential and nonessential genes are well separated and the average degree is systematically higher for essential genes than for nonessential ones—consistently with previous findings [34].

Fig 1. Probability distribution $P(k)$ for the number of connections k of each protein, averaged over the bacterial species considered in Table 1. Inset: $P(k)$ for essential (E) and nonessential (NE) genes, averaged over DEG-annotated genomes. Note that the average degree is higher for essential genes than for nonessential ones, and the two probability distributions are quite distinct. The region of the curve for low k can be well approximated by a power law [37].

PPI connectivity and gene conservation We now investigate whether there is a correlation between the degree of conservation (as measured by ERI) of genes and the connectivity k of the corresponding proteins in PPI networks. Figure 2 shows that highly connected proteins are also highly conserved among the bacterial species we consider, that constitute a reasonably wide sample of different evolutionary adaptations. This observation is a strong signature of the existence of an invariant structure of conserved hubs in all bacterial PPI networks. Indeed, we observe the existence of a structural transition for bacterial proteins with $k > 40$, which have ERI close to 1 almost surely, and are thus highly conserved among the species. Proteins that are less connected, on the contrary, have a wide range of ERI values. Interestingly, as shown in the inset of figure 2, the fluctuations in ERI as a function of k abruptly decrease for connectivities above the threshold.

Fig 2. Average ERI values of bacterial genes as a function of the degrees k of the corresponding proteins, for all the considered genomes. Error bars are standard deviations of ERI values associated to a given k value. Inset: amplitude of the error bar (Δ ERI) as a function of k .

Fig 3. Dependence of selective pressure in terms of K_a/K_s for a gene on the degree k of the corresponding protein, for all the considered genomes.

We then look at the evolutionary pressure exerted on genes whose proteins have different connectivities. The graph in figure 3 shows the ratio K_a/K_s for groups of genes binned by the connectivity k of the corresponding proteins, for all the 42 bacterial

species considered here. We see that the more connected proteins correspond to genes which are subject to an increasing purifying evolutionary pressure. Indeed, values of the K_a/K_s systematically decrease until they become zero, as a function of k . This result point to the fact that the more proteins are connected in the PPI networks the more the genes encoding them are subject to a purifying evolutionary pressure.

PPI and Essentiality. To further investigate the relationship between gene essentiality and protein connectivities, we consider DEG-annotated genomes and classify interactions between proteins (links) making references to the essentiality of the corresponding genes. We distinguish three sets of links: ee (linking proteins from two essential genes), $\bar{e}\bar{e}$ (from two nonessential genes) and $e\bar{e}$ (from an essential gene and a nonessential one). We then compute the *density* of these sets of links respectively as:

$$\rho_{ee} = \frac{|ee|}{\frac{1}{2}E(E-1)}, \quad \rho_{\bar{e}\bar{e}} = \frac{|\bar{e}\bar{e}|}{\frac{1}{2}NE(NE-1)}, \quad \rho_{e\bar{e}} = \frac{|e\bar{e}|}{\frac{1}{2}E \cdot NE}, \quad (1)$$

where E and NE denote the number of essential and nonessential genes, respectively (self-connections are excluded in our analysis). Such densities are then compared with the overall density of the network—restricted to genes classified as either essential or nonessential:

$$\langle \rho \rangle = \frac{|ee| + |\bar{e}\bar{e}| + |e\bar{e}|}{\frac{1}{2}(E + NE)(E + NE - 1)}. \quad (2)$$

We use the ratios $r_{ee} = \rho_{ee}/\langle \rho \rangle$, $r_{\bar{e}\bar{e}} = \rho_{\bar{e}\bar{e}}/\langle \rho \rangle$ and $r_{e\bar{e}} = \rho_{e\bar{e}}/\langle \rho \rangle$ to assess the level of connectivity of the subnetworks with respect to the overall connectivity. Table 3 shows that subnetworks of essential genes are far denser than the overall networks, and that, in general, essential and nonessential genes tend to form network components that are weakly interconnected. This happens because many essential genes encode for ribosomal proteins, which in turn are localized in the ribosome so that they have a major probability to interact [38]. Figures S4-S5 of the Supplementary Information display such network features for each individual species.

PPI connectivity and functional specialization. For each PPI network, we define the conditional probability that a protein with degree k belongs to a given COG as:

$$P(\text{COG}|k) = P(k|\text{COG})P(\text{COG})/P(k), \quad (3)$$

where $P(k)$ is the degree distribution in the PPI network, $P(\text{COG})$ is the frequency of that COG in the proteome, and $P(k|\text{COG})$ is the degree distribution restricted to that COGs. Figure 4 shows the COG spectrum as a function of k over all bacteria species considered. Interestingly, we again note a marked transition. Below $k \simeq 40$ the COG spectrum is quite heterogeneous: genes corresponding to proteins with low connectivity are spread over several COGs which correspond to different functions (see Table 2). Genes whose proteins are highly connected ($k \geq 40$) are instead mainly concentrated in COG J, which encompasses translation processes and ribosomal functions. There are yet a handful of outliers, hubs with connectivity between 57 and 62, that belong to COG I (related to lipid transport and metabolism) and K and L (which, together with J, define the functional class of information storage and processing). The list of these outliers is reported in Table 4).

Fig 4. Probability distribution $P(\text{COG}|k)$ of belonging to a given COG for proteins with degree k , over all considered genomes. Proteins with low connectivity have a very heterogeneous COG composition, whereas, those with high k basically belong only to COG J.

Conclusions

Topological analysis of biological networks, such as protein-protein interaction or metabolic networks, has demonstrated that structural features of network subgraphs are correlated with biological functions [39,40]. For instance, it was shown that highly connected patterns of proteins in a PPI are fundamental to cell viability [41]. In this work we have shown the existence of a topological-functional transition in bacterial species, ruled by the connectivity of proteins in the PPI networks. The threshold in k of the transition is located between $k = 40$ and $k = 50$. Proteins that have connectivities above the threshold are mostly encoded by genes that are conserved (as measured both by ERI and K_a/K_s) and essential. Moreover the functional repertoire above the threshold focuses mainly on the COG J, with just a few interesting hubs belonging to COGs I, K and L.

Indeed, the PPI network of each bacterial species is characterized by a highly connected core of conserved ribosomal proteins, the components of multi-subunit complexes whose corresponding genes are mostly essential [31,35] and code for supra-molecular complexes, that pile up in the bump we have observed for the degree distribution (figure1). Hence, what we are seeing here is essentially the ribosome, and related protein complexes such as RNA Polymerase. Indeed, the ribosome is the only molecular machine in bacteria in which a given protein could legitimately have 40 or more protein binding partners, with the help of rRNA mediating interactions [42].

We believe that the observations we have presented here can have implications both for the prediction of gene essentiality, based on the knowledge of PPI networks, and for the prediction of interactions between proteins, based on genetic information [43,44]. It is interesting to note that our results are consistent with a previous study based on inferred bacterial co-conserved networks based on phylogenetic profiles [35]. The coupled flows of information from the genetic level up to the proteomic level and vice-versa should be further systematically investigated, taking into account archaeal and prokaryotic genomes in the search for emerging multi-layer structures that could offer basic theoretical grounds for clinical and systemic applications, for instance related to antimicrobial resistances [45–48].

References

1. Drewes G, Bouwmeester T. Global approaches to protein-protein interactions. *Current Opinion in Cell Biology*. 2003;15(2):199–205. doi:10.1016/S0955-0674(03)00005-X.
2. Golemis E, Adams PD. Protein-protein Interactions: A Molecular Cloning Manual. Cold Spring Harbor Laboratory Press; 2005.
3. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*. 2002;417:399–403. doi:10.1038/nature750.
4. Tong AHY, Drees B, Nardelli G, Bader GD, Brannetti B, Castagnoli L, et al. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science*. 2002;295(5553):321–324. doi:10.1126/science.1064987.
5. Shatsky M, Allen S, Gold BL, Liu NL, Juba TR, Reveco SA, et al. Bacterial Interactomes: Interacting Protein Partners Share Similar Function and Are Validated in Independent Assays More Frequently Than Previously Reported. *Mol Cell Proteomics*. 2016;15(5):1539–55. doi:10.1074/mcp.M115.054692.

Organisms	Abbr.	RefSeq	STRING	n
<i>Agrobacterium tumefaciens</i> (fabrum)	agtu	NC_003062	176299	2765
<i>Aquifex aeolicus</i> VF5	aqae	NC_000918	224324	1497
<i>Bacillus subtilis</i> 168	basu	NC_000964	224308	4175
<i>Bacteroides thetaiotaomicron</i> VPI-5482	bath	NC_004663	226186	4778
<i>Brucella melitensis</i> bv. 1 str. 16M	brme	NC_003317.1	224914	2059
<i>Burkholderia pseudomallei</i> K96243	bups	NC_006350	272560	3398
<i>Buchnera aphidicola</i> Sg uid57913	busg	NC_004061	198804	546
<i>Burkholderia thailandensis</i> E264	buth	NC_007651	271848	3276
<i>Caulobacter crescentus</i>	cacr	NC_011916	565050	3885
<i>Campylobacter jejuni</i>	caje	NC_002163	192222	1572
<i>Chlamydia trachomatis</i> D/UW-3/CX	chtr	NC_000117.1	272561	894
<i>Clostridium acetobutylicum</i> ATCC 824	clac	NC_003030.1	272562	3602
<i>Deinococcus radiodurans</i> R1	dera	NC_001263.1	243230	2629
<i>Escherichia Coli</i> K-12 MG1655	esco	NC_000913.3	511145	4004
<i>Francisella novicida</i> U112	frno	NC_008601	401614	1719
<i>Fusobacterium nucleatum</i> ATCC 25586	funu	NC_003454.1	190304	1983
<i>Haemophilus influenzae</i> Rd KW20	hain	NC_000907.1	71421	1610
<i>Helicobacter pylori</i> 26695	hepy	NC_000915	85962	1469
<i>Listeria monocytogenes</i> EGD-e	limo	NC_003210.1	169963	2867
<i>Mesorhizobium loti</i> MAFF303099	melo	NC_002678.2	266835	6743
<i>Mycoplasma genitalium</i> G37	myge	NC_000908	243273	475
<i>Mycoplasma pneumoniae</i> M129	mypn	NC_000912.1	272634	648
<i>Mycoplasma pulmonis</i> UAB CTIP	mypu	NC_002771	272635	782
<i>Mycobacterium tuberculosis</i> H37Rv	mytu	NC_000962.3	83332	3936
<i>Neisseria gonorrhoeae</i> FA 1090 uid57611	nego	NC_002946	242231	1894
<i>Porphyromonas gingivalis</i> ATCC 33277	pogi	NC_010729	431947	2089
<i>Pseudomonas aeruginosa</i> UCBPP-PA14	psae	NC_008463	208963	5892
<i>Ralstonia solanacearum</i> GMI1000	raso	NC_003295.1	267608	3436
<i>Rickettsia prowazekii</i> str. Madrid E	ripr	NC_000963.1	272947	8433
<i>Salmonella enterica</i> serovar Typhi	saen	NC_004631	209261	4352
<i>Shewanella oneidensis</i> MR-1	shon	NC_004347	211586	4065
<i>Sinorhizobium meliloti</i> 1021	sime	NC_003047.1	266834	3359
<i>Sphingomonas wittichii</i> RW1	spwi	NC_009511	392499	4850
<i>Staphylococcus aureus</i> N315	stau	NC_002745.2	158879	2582
<i>Staphylococcus aureus</i> NCTC 8325	stau_	NC_007795	93061	2767
<i>Streptococcus pyogenes</i> NZ131	stpy	NC_011375	471876	1700
<i>Streptococcus sanguinis</i>	stsa	NC_009009	388919	2270
<i>Synechocystis</i> sp. PCC 6803	syp	NC_000911.1	1148	3179
<i>Thermotoga maritima</i> MSB8	thma	NC_000853.1	243274	1858
<i>Treponema pallidum</i> Nichols	trpa	NC_000919.1	243276	1036
<i>Vibrio cholerae</i> N16961	vich	NC_002505	243277	2534
<i>Xylella fastidiosa</i> 9a5c	xyfa	NC_002488	160492	2766

Table 1. Summary of the selected bacterial dataset. Organism name, abbreviation, RefSeq, STRING code, size of genome (number of genes *n*). Genomes annotated in the Database of Essential Genes (DEG) are highlighted with bold fonts.

COG ID	Functional classification
<i>INFORMATION STORAGE AND PROCESSING</i>	
J	Translation, ribosomal structure and biogenesis
K	Transcription
L	Replication, recombination and repair
<i>CELLULAR PROCESSES AND SIGNALING</i>	
D	Cell cycle control, cell division, chromosome partitioning
T	Signal transduction mechanisms
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
O	Post-translational modification, protein turnover, chaperones
<i>METABOLISM</i>	
C	Energy production and conversion
G	Carbohydrate transport and metabolism
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
P	Inorganic ion transport and metabolism

Table 2. Functional classification of COG clusters.

Organisms	r_{ee}	$r_{\bar{e}\bar{e}}$	$r_{e\bar{e}}$
basu	44.46	0.80	0.11
bath	20.07	0.76	0.25
bups	6.21	0.83	0.27
buth	18.69	0.70	0.22
cacr	18.40	0.70	0.15
caje	3.65	0.82	0.32
esco	2.91	0.88	0.31
frno	9.84	0.52	0.18
hain	1.65	1.15	0.27
hepy	2.91	0.78	0.38
myge	1.42	0.29	0.08
mypu	3.42	0.22	0.12
mytu	8.09	0.78	0.23
pogi	11.03	0.41	0.21
psae	9.85	0.92	0.16
saen	28.80	0.81	0.12
shon	6.50	0.64	0.16
spwi	15.47	0.74	0.22
stau	23.05	0.58	0.23
stau_	21.89	0.64	0.16
stpy	9.30	0.73	0.23
stsa	30.65	0.61	0.22
vich	8.37	0.81	0.19

Table 3. Relative density values r for PPI subnetworks between essential genes (r_{ee}), between nonessential genes ($r_{\bar{e}\bar{e}}$) and between essential and nonessential genes ($r_{e\bar{e}}$), for each DEG-annotated bacterial genome.

6. Caufield JH, Wimble C, Shary S, Wuchty S, Uetz P. Bacterial protein meta-interactomes predict cross-species interactions and protein function. BMC

228
229

<i>k</i>	COG	Gene	Protein
57	1250I	paaH	3-hydroxyadipyl-CoA dehydrogenase, NADdependent
	0365I	acs	acetyl-CoA synthetase
58	1250I	paaH	3-hydroxyadipyl-CoA dehydrogenase, NADdependent
	0222J	rpIL	50S ribosomal subunit protein L7/L12
	0335J	rpIS	50S ribosomal subunit protein L19
	0267J	rpmG	50S ribosomal subunit protein L33
59	0365I	acs	acetyl-CoA synthetase
	0183I	paaJ	3-oxoadipyl-CoA3-oxo-5,6-dehydrosuberil-CoA thiolase
	1960I	ydiO	putative acyl-CoA dehydrogenase
60	0183I	atoB	acetyl-CoA acetyltransferase
	0197J	rpIP	50S ribosomal subunit protein L16
	0088J	rpID	50S ribosomal subunit protein L4
	0197J	rpIP	50S ribosomal subunit protein L16
	0087J	rpIC	50S ribosomal subunit protein L3
61	1960I	aidB	putative acyl-CoA dehydrogenase
	0085K	rpoB	RNA polymerase, beta subunit
62	0202K	rpoA	RNA polymerase, alpha subunit
	0087J	rpIC	50S ribosomal subunit protein L3
	0052J	rpsB	30S ribosomal subunit protein S2
	2965L	PriB	ribosomal replication protein

Table 4. Specific hubs. In this table we detail which proteins populate the few bins of connectivity around $k = 60$ in figure 4.

- Bioinformatics. 2017;18(1):171. doi:10.1186/s12859-017-1585-0. 230
7. Dilucca M, Cimini G, Semmoloni A, Deiana A, Giansanti A. Codon bias patterns of *E. coli*'s interacting proteins. PLoS ONE. 2015;10(11):e0142127. 231
doi:10.1371/journal.pone.0142127. 232
233
8. Dilucca M, Cimini G, Giansanti A. Essentiality, conservation, evolutionary pressure and codon bias in bacterial genomes. Gene. 2018;663:178 – 188. 234
doi:https://doi.org/10.1016/j.gene.2018.04.017. 235
236
9. Gerdes S, Scholle M, Campbell J, Balazsi G, Ravasz E, Daugherty M, et al. 237
Experimental determination and system level analysis of essential genes in 238
Escherichia coli MG1655. Journal of Bacteriology. 2003;185(19):5673–5684. 239
doi:10.1128/JB.185.19.5673-5684.2003. 240
10. Fang G, Rocha E, Danchin A. How essential are nonessential genes? Molecular 241
Biology and Evolution. 2005;22(11):2147–2156. doi:10.1093/molbev/msi211. 242
11. Peng C, Gao F. Protein localization analysis of essential genes in prokaryotes. 243
Scientific Reports. 2014;4:6001. doi:10.1038/srep06001. 244
12. Lin Y, Gao F, Zhang CT. Functionality of essential genes drives gene strand-bias 245
in bacterial genomes. Biochemical and Biophysical Research Communications. 246
2010;396(2):472–476. doi:10.1016/j.bbrc.2010.04.119. 247
13. Hurst LD, Smith NG. Do essential genes evolve slowly? Current Biology. 248
1999;9(14):747–750. doi:10.1016/S0960-9822(99)80334-0. 249

14. Jordan IK, Rogozin IB, Wolf YI, Koonin EV. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Research*. 2002;12(6):962–968. doi:10.1101/gr.87702. 250–252
15. Luo H, Gao F, Lin Y. Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes. *Scientific Reports*. 2015;5:13210. doi:10.1038/srep13210. 253–255
16. Ish-Am O, Kristensen DM, Ruppin E. Evolutionary conservation of bacterial essential metabolic genes across all bacterial culture media. *PLoS ONE*. 2015;10(4):e0123785. doi:10.1371/journal.pone.0123785. 256–258
17. Alvarez-Ponce D, Sabater-Muñoz B, Toft C, Ruiz-González MX, Fares MA. Essentiality is a strong determinant of protein rates of evolution during mutation accumulation experiments in *Escherichia coli*. *Genome Biology and Evolution*. 2016;8(9):2914–2927. doi:10.1093/gbe/evw205. 259–262
18. Bergmiller T, Ackermann M, Silander OK. Patterns of evolutionary conservation of essential genes correlate with their compensability. *PLoS Genetics*. 2012;8(6):e1002803. doi:10.1371/journal.pgen.1002803. 263–265
19. Hurst LD. The K_a/K_s ratio: diagnosing the form of sequence evolution. *Trends in Genetics*. 2002;18:486–487. doi:10.1016/S0168-9525(02)02722-1. 266–267
20. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Research*. 2013;41(D1):D36–D42. doi:10.1093/nar/gks1195. 268–270
21. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*. 1986;3(5):418. doi:10.1093/oxfordjournals.molbev.a040410. 271–273
22. Zhang R, Lin Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Research*. 2009;37(suppl 1):D455–D458. 274–275
23. Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Research*. 2014;42(D1):D574–D580. doi:10.1093/nar/gkt1131. 276–279
24. Szklarczyk MJ. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*. 2017;45(Database issue):D362–D368. doi:10.1093/nar/gkq973. 280–282
25. Chien CT, Bartel PL, Sternglanz R, Fields S. The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. *Proceedings of the National Academy of Science*. 1991;88(1):9578–9582. doi:10.1073/pnas.88.21.9578. 283–286
26. Phizicky EM, Fields S. Protein-protein interactions: Methods for detection and analysis. *Microbiological Reviews*. 1995;59(1):94–123. 287–288
27. Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, et al. The Tandem Affinity Purification (TAP) method: A general procedure of protein complex purification. *Methods*. 2001;24(3):218–229. doi:10.1006/meth.2001.1183. 289–291
28. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 2000;28(1):27. doi:10.1093/nar/28.1.27. 292–293

29. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*. 2001;29(1):22. doi:10.1093/nar/29.1.22. 294-297
30. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*. 2015;43(D1):D261. doi:10.1093/nar/gku1223. 298-300
31. Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, Canadien V, et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*. 2005;433:531–537. doi:10.1038/nature03239. 301-303
32. Jin Y, Turaev D, Weinmaier T, Rattei T, Makse HA. The Evolutionary Dynamics of Protein-Protein Interaction Networks Inferred from the Reconstruction of Ancient Networks. *PLoS ONE*. 2013;8:e58134. doi:10.1371/journal.pone.0058134. 304-306
33. Rajagopala SV, Sikorski P, Kumar A, Mosca R, Vlasblom J, Arnold R, et al. The binary protein-protein interaction landscape of *Escherichia coli*. *Nature Biotechnology*. 2014;32:285–290. doi:10.1038/nbt.2831. 307-309
34. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Molecular Biology and Evolution*. 2005;22(4):803–806. doi:10.1093/molbev/msi072. 310-312
35. Karimpour-Fard A, Leach SM, Hunter LE, Gill RT. The topology of the bacterial co-conserved protein network and its implications for predicting protein function. *BMC Genomics*. 2008;9(1):313. doi:10.1186/1471-2164-9-313. 313-315
36. Wuchty S, Uetz P. Protein-protein interaction networks of *E. coli* and *S. cerevisiae* are similar. *Scientific Reports*. 2014;4:7187. doi:10.1038/srep07187. 316-317
37. Annibale A, Coolen ACC, Planell-Morell N. Protein interaction networks and biology: Towards the connection; 2015. arXiv:1501.00662. 318-319
38. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4(2). doi:10.1186/1471-2105-4-2. 320-322
39. Srinivasan BS, Novak AF, Flannick JA, Batzoglou S, McAdams HH. In: Apostolico A, Guerra C, Istrail S, Pevzner PA, Waterman M, editors. *Integrated Protein Interaction Networks for 11 Microbes*. Springer Berlin Heidelberg; 2006. p. 1–14. 323-326
40. Rao VS, Srinivas K, Sujini GN, Kumar GNS. Protein-protein interaction detection: methods and analysis. *International Journal of Proteomics*. 2014;2014(147648). doi:10.1155/2014/147648. 327-329
41. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001;411(6833):41–42. doi:10.1038/35075138. 330-331
42. Fox GE. Origin and evolution of the ribosome. *Cold Spring Harbor Perspectives in Biology*. 2010;2(9). doi:10.1101/cshperspect.a003483. 332-333
43. Hwang YC, Lin CC, Chang JY, Mori H, Juan HF, Huang HC. Predicting essential genes based on network and sequence analysis. *Molecular BioSystems*. 2009;5:1672–1678. doi:10.1039/B900611G. 334-336

44. Wei W, Ning LW, Ye YN, Guo FB. Geptop: A gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. *PLoS ONE*. 2013;8(8):e72343. doi:10.1371/journal.pone.0072343. 337
338
339
45. Terradot L, Noirot-Gros MF. Bacterial protein interaction networks: puzzle stones from solved complex structures add to a clearer picture. *Integrative Biology*. 2011;3:645–652. doi:10.1039/C0IB00023J. 340
341
342
46. Zoraghi R, Reiner NE. Protein interaction networks as starting points to identify novel antimicrobial drug targets. *Curr Opin Microbiol*. 2013;16(5):566–72. doi:10.1016/j.mib.2013.07.010. 343
344
345
47. Sevimoglu T, Arga KY. The role of protein interaction networks in systems biomedicine. *Computational and Structural Biotechnology Journal*. 2014;11(18):22–27. doi:10.1016/j.csbj.2014.08.008. 346
347
348
48. Boccaletti S, Bianconi G, Criado R, del Genio CI, nes JGG, Romance M, et al. The structure and dynamics of multilayer networks. *Physics Reports*. 2014;544(1):1–122. doi:10.1016/j.physrep.2014.07.001. 349
350
351

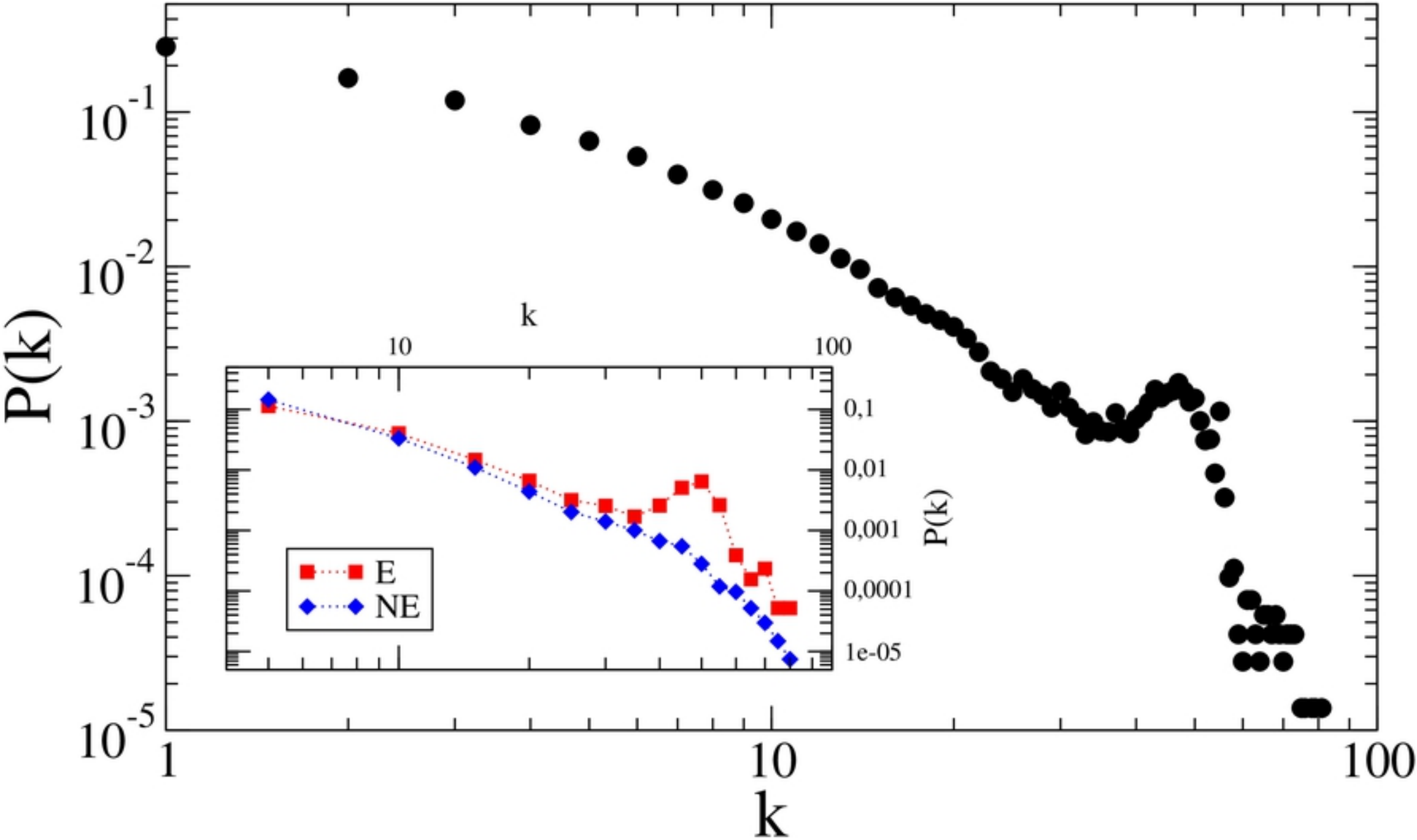


Fig.1

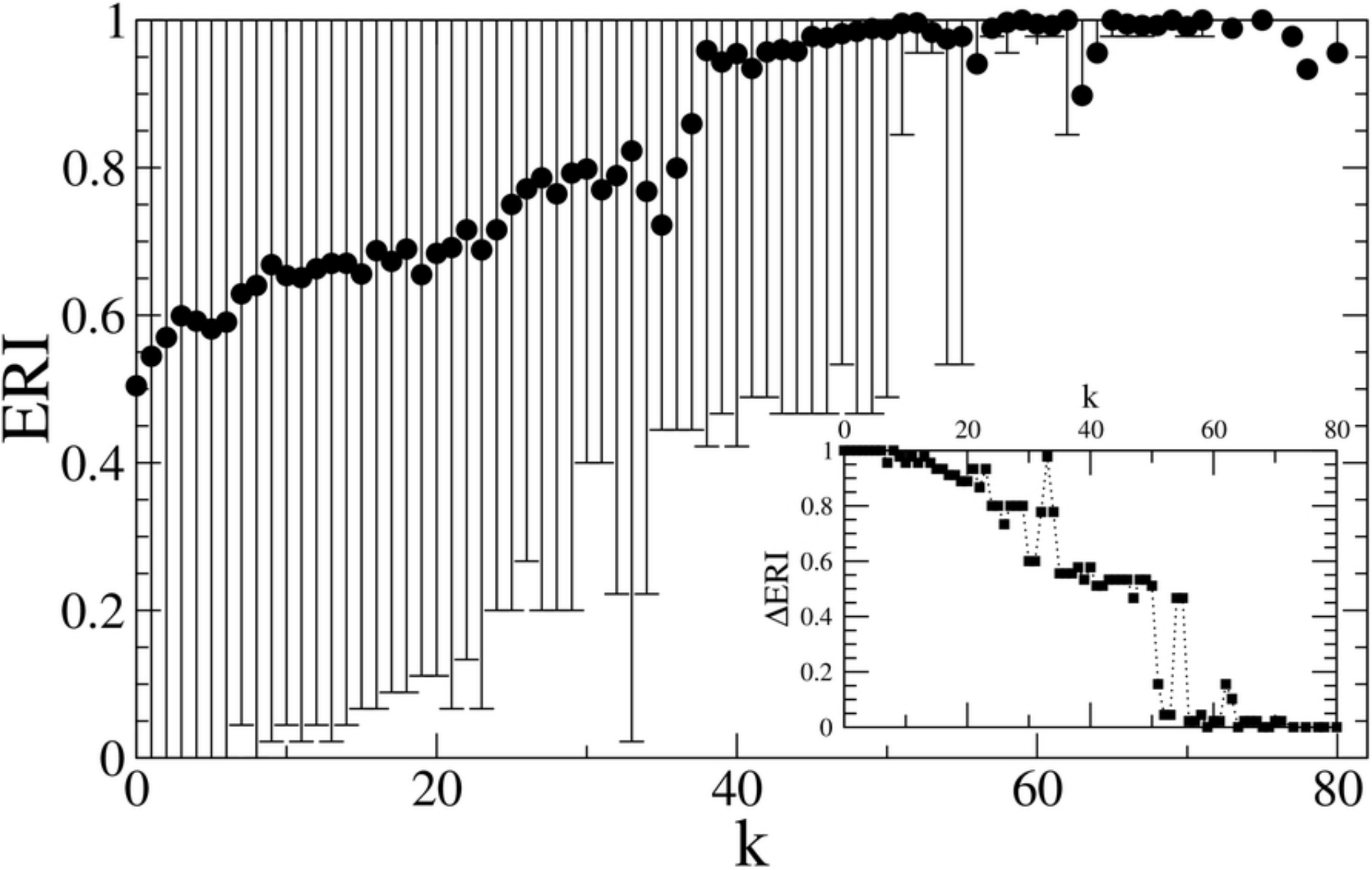


Fig.2

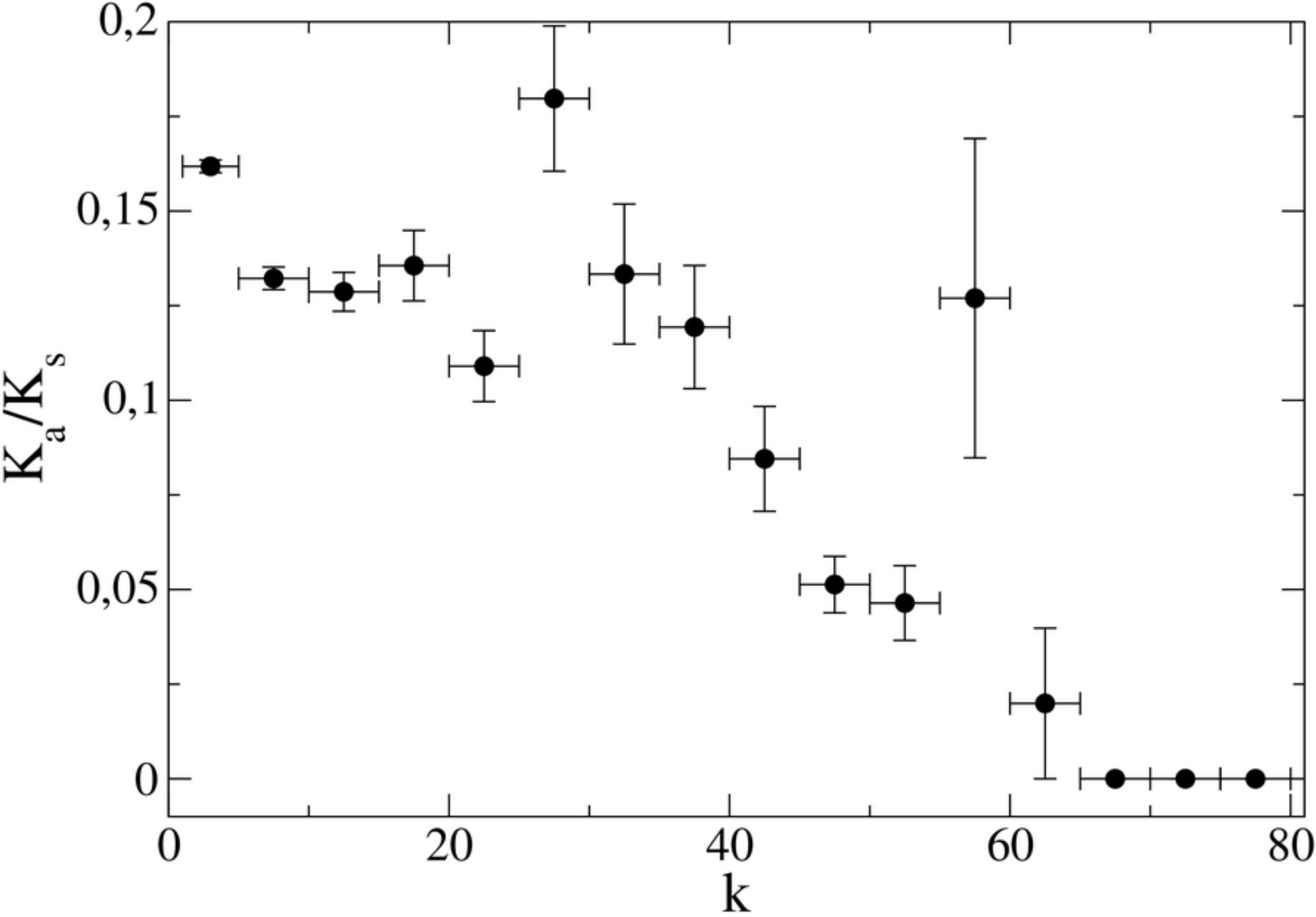


Fig.3

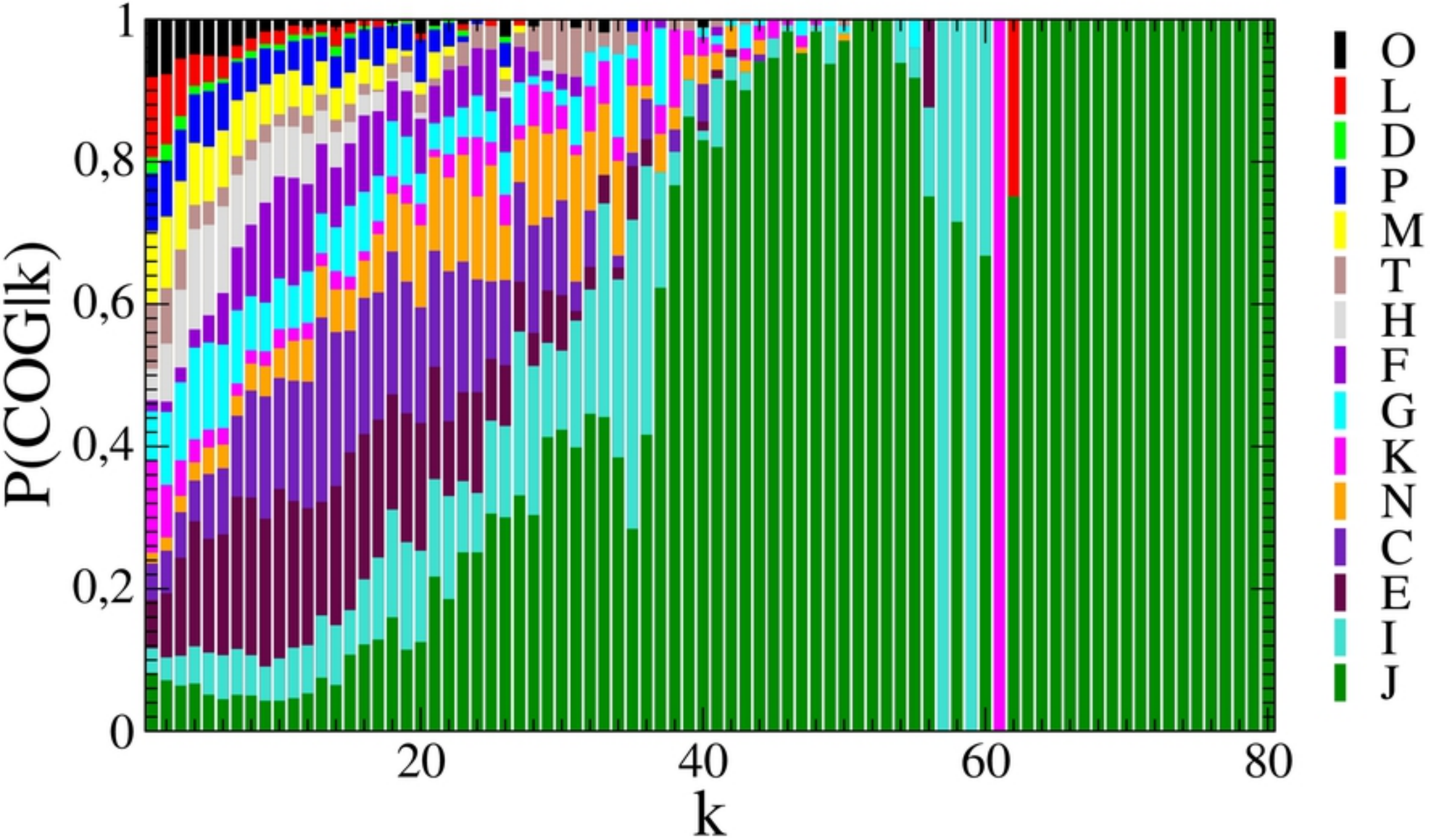


Fig.4