Scale-Free Networks in Biological Complexity

I. Scale-Free Network

Networks exist in all fields, with a wide range of size, time-scale, interaction mechanism and individual dynamics: Internet, network of move-actor collaborations, protein interaction network, network of world airport…etc [1]. Many of these real networks have been found to be scale free-- that is their connectivity of distribution obeys a power law $P(k) \propto k^{-\lambda}$ instead of an exponential law suggested by random graph model [2,3], see Fig1.

1.1 Barabasi-Albert's Model (BA Model)

To account for the scale-free power-law distribution of connectivity, A. -L.Barabasi and R.Albert proposed a model incorporating two key features of real networks: growth of network and preferential attachment [2]. They showed that a power-law distribution emerges automatically when new nodes (with connectivity fixed as a constant m) are added continuously into the open system and attach to the existing nodes with probability linearly proportional to the connectivity of the target nodes---the rich get richer. Their model leads to the connectivity distribution $P(k) = 2\frac{m^2}{k^3}$.

1.2 Generalization of BA model

BA model offers a successful mechanism to explain the scale-free nature of real networks, but perhaps too simple to account for all aspects of the existed networks. More sophisticated models vary the form of preferential attachment or include the effect of aging and capacity limitation [4,5,6]. Besides scale-free distribution, these generalized models predict exponential and truncated power-law in some parameter regime as well, in good agreement with the occurrence of the 3 different classes of small-world network presented by L.A.N.Amaral et al [4].

(1) Non-linear preferential attachment

Generally, the linking probability can have an arbitrary nonlinear form $\Pi_k \sim k^{\alpha}$ and it was found that quite different behaviors arise for $\alpha < 1, \alpha = 1$ and $\alpha > 1$ [5]. Obviously, the case $\alpha = 1$ corresponds to the linear preferential attachment assumed in BA model, and results in the power-law connectivity distribution $P_k \propto k^{-\gamma}$, γ tunable in the range $2 < \gamma < \infty$ by the finer details of dependence of connection kernels on k [5]. For $\alpha < 1$, the connectivity distribution P_k varies as a stretched exponential; for $\alpha > 1$, the system will be dominated by a single site connecting to nearly all other sites.

(2) Constrains for addition of new links

L.A.N.Amaral et al suggested that two constrain factors hinder the preferential attachment and might account for the emergence of different classes of network [4]: a) aging of vertices and b) limited capacity of vertices. With time going on, a vertex will get old, inactive or worn-out and a highly connected vertex will eventually stop receiving new links, though still a part of the whole network. Furthermore, the limited capacity (time, space etc) and physical cost won't allow too many links to attach to a given vertex. Simulations with models incorporating these factors yield very good agreement with data from real world network, and it was shown that when the constrains are strong

enough, no power law region is visible for the connectivity distribution (Figure2)[4]. A paper published earlier pointed out that if $\Pi_k \propto \tau^{-\alpha}$, with τ the age of the preexisted vertices, the network shows scaling behavior only in the region $\alpha < 1$; otherwise, the distribution P_k will just be exponential [6].

(3) Growth of connections

An important character of a small-world network is its diameter D or the characteristic length L as named in some literature, defined as the shortest distance averaged over all pairs of vertices. In *BA* model, the average connectivity of a vertex is fixed, which is plausible for almost all the non-biological networks examined [3]. With constant average connectivity, the diameter increases logarithmically with the addition of new vertices [7]. However, we can't exclude the possibility that in some systems, as the network evolves, not only new vertices are introduced but connections between established vertices increase as well. This will in turn modify the exponent γ and, in some cases, the growth of network diameter [3], see discussion in Part II 3.

II. Biological Complex Systems

Power law distribution of connectivity, the signature of scale-free networks, has been found in different kinds of networks in wide range such as the electric grid of Southern California, the network of move-actor collaborations, the world-wide-web and citation of scientific paper etc [4]. This signature is also imprinted in some biological complex systems [3,8,9] and in my personal opinion they might present more interesting phenomena than in other systems, because: a) The components of biological networks are living cells, proteins or organisms etc, which are highly active, mobile and self-adaptable, thus the resulting networks are generally rather changeful and complicated; b) these components possess both contextual properties as network members and specific functions as individuals. To better understand the dynamics of the biological network, these factors have to be incorporated together.

Recently, the research group lead by A.-L.Barabasi in University of Notre Dame has carried out a series of study on networks in biological complex systems such as protein interaction network[8] and metabolic interaction network [3]. They found that an inhomogeneous scale-free network structure emerges in both systems, suggesting the evolutionary selection of a common large-scale topology of biological network. Their results are shown below:

2.1 System Architecture

The S.*cerevisiae* protein-protein interaction network being investigated contains 1,847 proteins as nodes and 2,240 identified direct physical interactions as links, as shown in Fig3 (a) [8]. To find the architecture of this complex network, they first investigated the connectivity distribution $P(k)$ (defined as the possibility that a given yeast proteins interacts with k other yeast proteins). It was found that $P(k)$ follows a power law, indicating the emergence of a highly inhomogeneous scale-free network:

$$P(k) \propto \frac{1}{(k+k_0)^{\gamma}} e^{-k/k_c}$$

where the short-length scale correction $k_0 \approx 1$ and the exponential cut-off $k_c \approx 20$ (Fig 3 (b)).

Similar power-law connectivity distribution $P(k) \propto k^{-\gamma}$ was also found in a systematic comparative analysis of the metabolic network of 43 organisms, from all three domains of life: archae, bacterium and eukaryote [3] (Fig 4). The metabolic network is made up by nodes, the substrates, which are connected to each other though links, the actual metabolic reactions.

2.2 Robust while Fragile

An important consequence of the power-law connectivity distribution is that: the network is dominated by a few highly connected nodes and is fragile against the removal of these nodes. In both systems [3,8], when the most highly connected nodes are removed subsequently, the diameter of the network increases sharply. In contrast, when random chosen nodes are removed, the diameter is rarely affected; indicating that the networks are insensitive to random errors, see Fig 5.

2.3 Diameter of the network

The diameter of the network was believed to increase with the addition of new nodes [7], provided that the average connectivity is constant. For a metabolic network, this implies that a more complex system (with more nodes) will have a larger diameter than a simple system. However, it was found that for all the 43 organisms investigated, the diameter is the same! To maintain a constant diameter, the average connectivity of a substrate has to increase as new nodes are added and this was proved by the relationship between average link number and total number of nodes shown in Fig 6.

2.4 Discussion and suggestion for follow-up studies

Origin of preferential attachment

The power-law distribution emerges as a result of preferential attachment. At the level of protein interactions, we may ask why certain proteins are more likely to get attached to the newly introduced proteins than others are? Is the chance to become the most highly connected nodes equal for all members, determined by initial condition at the birth of the network, or related to the specific functions or structure of those nodes? It was suggested that proteins that make up the highly connected nodes in cellular networks might share common structural feature [10]. To get a solid proof, we need further investigations in and comparison among the architecture of different protein networks.

An encouraging support was provided in the study of metabolic systems. The researchers found that though only around 4% of total substrates are present in all 43 organisms, they turn out to be the most highly connected substrates found in any individual organism [3]. This discovery indicates that the metabolic network in all these 43 species is dominated the **same** highly connected substrates, so their high connectivity must result from their special features or functions, which may yield clues to the emergence of power-law distribution in these networks.

Lethality and centrality

In both systems, the lethality of the most highly connected nodes are revealed. Thus removing a node with high connectivity will be much more disruptive to the system function while removal of a less connected node will barely affect the whole system. These networks have high tolerance and resistance to random attack while quite fragile to targeted attacks. This phenomenon is characteristic of inhomogeneous network and has been observed in non-biological systems like the Internet [2] and biological systems like p53 network [9]. From a biomedical standpoint, this finding could be used in

drug development [10]. For example, if a highly connected protein is selected as a drug target, the resulting inactivation of that protein could be fatal to the normal functions of the cell system.

Exponential Tail:

In protein interaction system [8], an apparent cut-off is observed and for k large enough ($k \geq k_c$) the connectivity distribution no longer obey the power law and an exponential tail follows. The investigation into what factors actually bring about the exponential cut-off k_c (aging of proteins? Space limitation? Information storage limitation?) might be hard in such complicated coupling biological system. A possible guess, I think, is that there exist some competitive mechanisms and when the one that disfavor high connectivity takes the upper hand, the power law breaks down. For a better understanding, follow-up studies are needed in other protein-protein interacting systems.

On the other hand, just knowing the factor that an exponential cut-off exists in the protein interaction system already yield useful information. For example, if we want a drug protein to attach to a specific kind of proteins in the network, the saturation of connectivity of the target proteins has to be taken into consideration.

Conservation of Diameters

A most surprising feature of the metabolic network is the apparent conservation of diameter in all 43 living organisms, in contrast to the logarithmic growth with network size predicted by some theories of 'small-world network' and observation in non-biological networks [2,7]. The authors argue that this feature may represent an additional survival and growth advantage for living systems, since larger diameter will attenuate the organism's ability to respond efficiently to an external change or internal error. As I said in the beginning of part II, biological networks are more interesting because the nodes are highly active and will adjust their connectivity to maintain the stability and efficiency of the network. In contrast, in non-living systems, such a self-adjusting situation rarely arises.

Is the conservation of diameter, which up to now only reported from metabolic networks, a unique feature of biological networks or just a feature for some specific systems? We can expect that more follow-up research in other biological systems will give us a final answer.

Reference:

[1] Steven H. Strogatz, *Nature*, **410**, 268 (2001)

[2] A. -L. Barabasi & R.Albert, *Science*, **286**,509 (1999)

[3] H. Jeong, B. Tombor, R. Albert, Z.N.Oltvai & A. -L. Barabasi, *Nature*, **407**, 651 (2000)

[4] L. A. N.Amaral, A. Scala, M. Barthelemy & H.E. Stanley, *PNAS*, **97**, 11149 (2000)

[5] P.L.Krapivsky, S.Redner & F.Leyvraz, *Phys.Rev.Lett.* , **85**, 4629, (2000)

[6] S.N.Dorogovtesv & J.F.F.Mendes, *Phys.Rev.E*, **62**, 1842 (2000)

[7] D.J.Watts & S.H.Stogatz, *Nature*, **393**, 440 (1998)

[8] H. Jeong, S.P. Mason, A. –L. Barabasi & Z.N. Oltvai, *Nature*, **411**, 41 (2001)

[9] F. Slanina & M. Kotrla, *Physical Review E*, **62**, 6170 (2000)

[10] J.Hasty & J.J. Collins, *Nature*, **411**, 30 (2001)

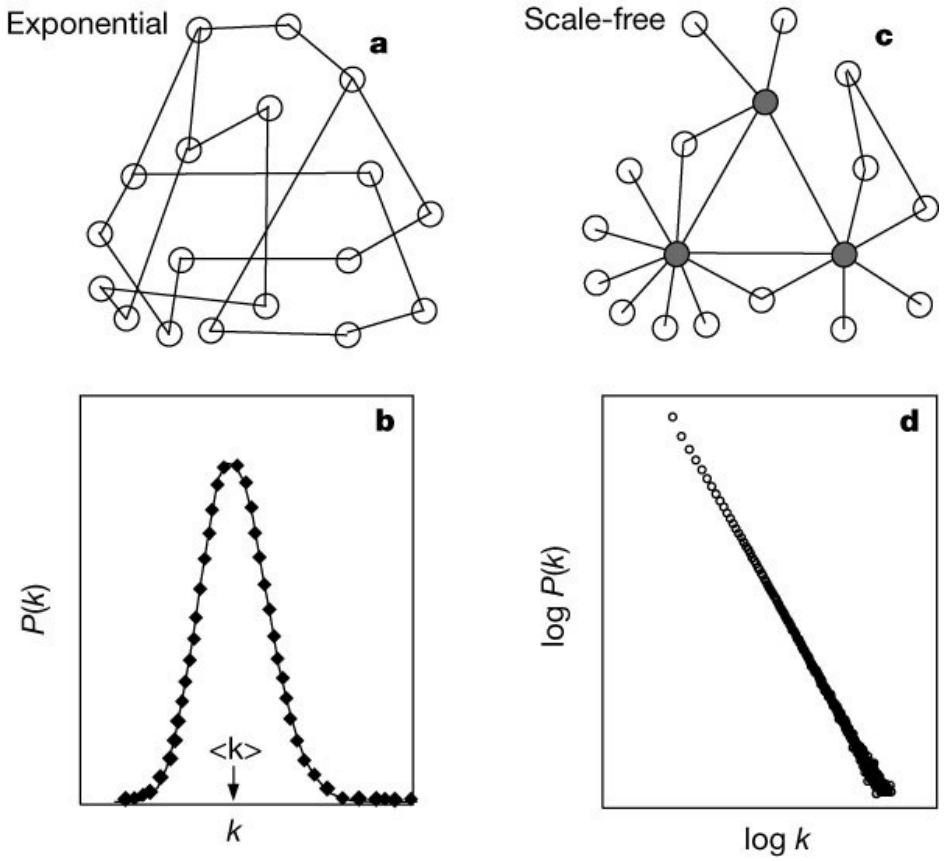

Fig 1. a) Representative structure of random network. Each pair of nodes in the network is connected randomly with a probability p, leading to a homogeneous network, in which most nodes have the same number of links <k>. b) The connectivity obeys a Poisson distribution peaking strongly at <k> and decays exponentially for k's far away from <k>.

c) Representative structure of scale-free networks. Most nodes have a few links while a few nodes have a very large number of links, leading to a heterogeneous network. d) The connectivity has no well-defined peak and decays as a power-law [3].

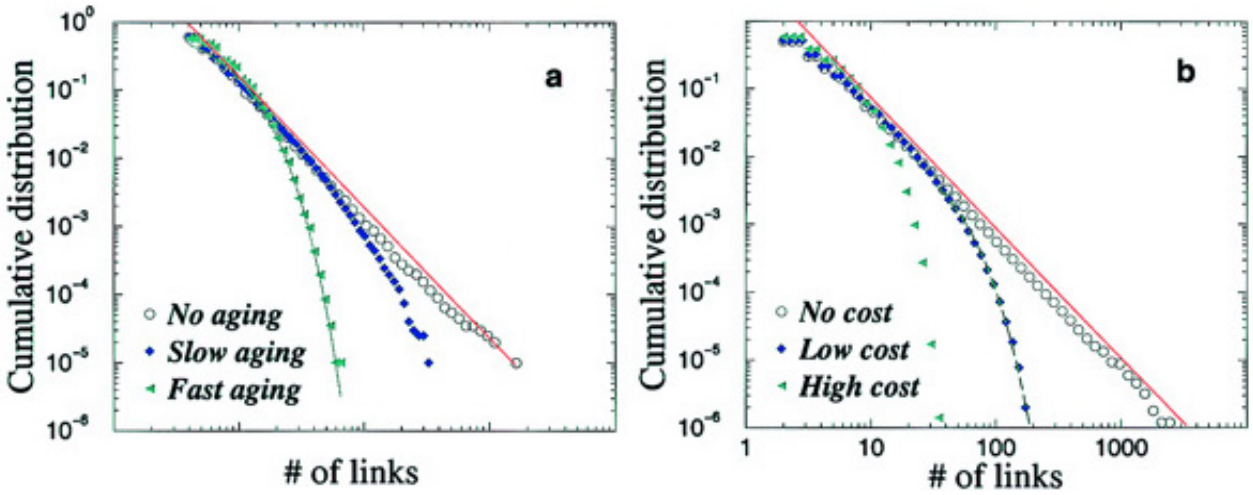

Fig.2 Truncation of scale-free connectivity by adding constrains in to the model of preferential attachment a) Effect of aging of vertices and b) effect of cost of adding linking on the connectivity distribution. As we can see, as the constrains become stronger, a cutoff of power law regime emerges and when the constrain is strong enough, the power law regime disappear altogether [4].

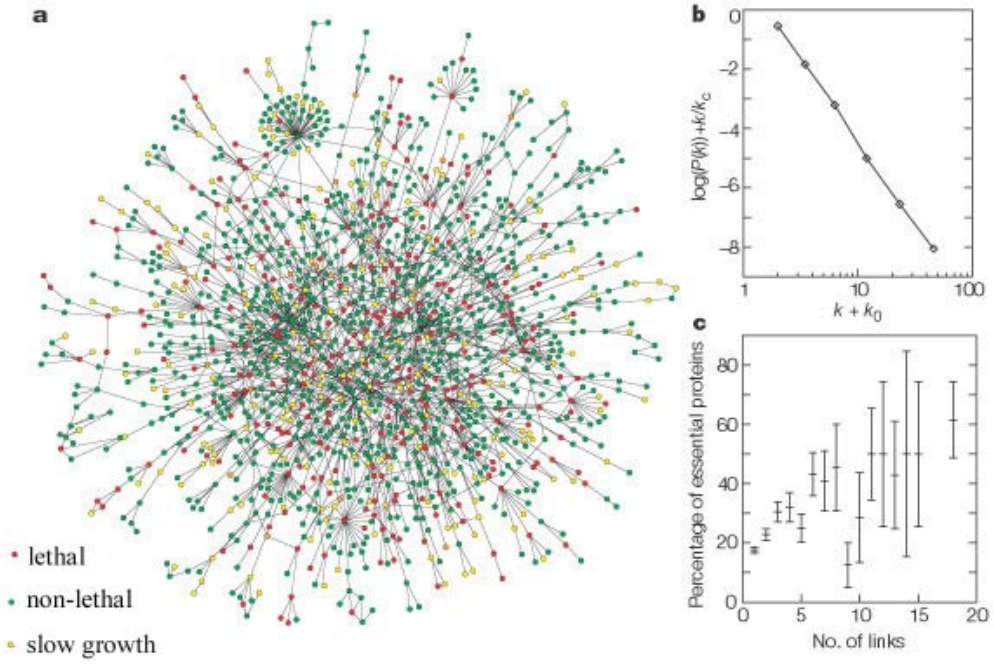

Fig 3 a, Map of S.cerevisiae protein-protein interaction. The phenotypic effect of removing a protein is indicated by the color of the corresponding node.

b, Connectivity distribution $P(k)$ of the interacting protein network.

c, The fraction of essential proteins with k links versus their connectivity k. (adapted from ref [8])

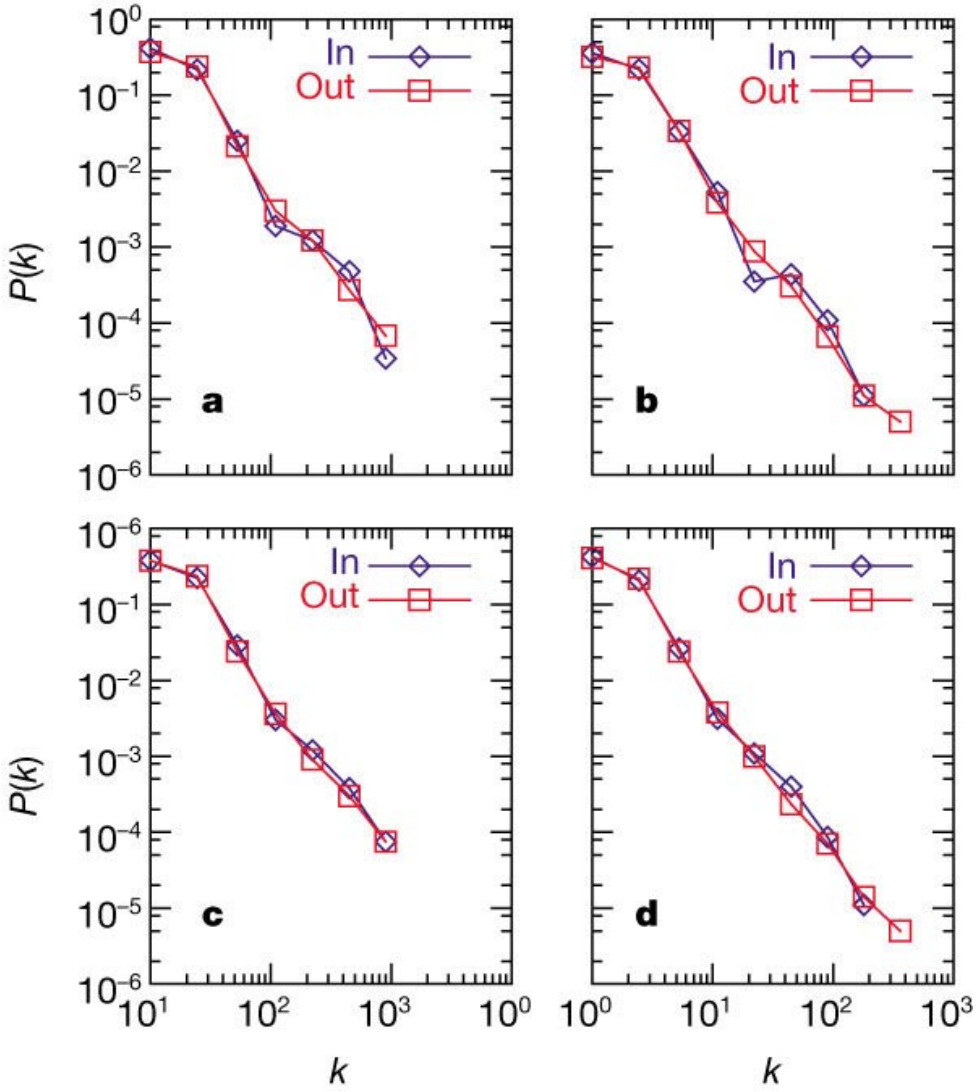

Fig 4. Connectivity distribution $P(k)$ for the metabolic networks. **a,** Archaeoglobus fulgidus (archae); **b**, E.colli (bacterium); **c**, Caenorhabditis elegans (eukaryote) **d**, The connectivity distribution averaged over all 43 organisms. Incoming and outcoming links are shown separately, corresponding to the number of reactions in which a substrate participates as product or an educt.

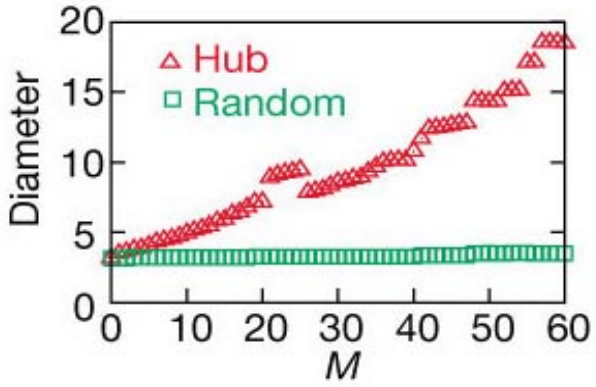

Fig 5. The effect of substrate removal on the metabolic network of E. coli. [3]. In the upper curve, the most connected substrates are removed first and the diameter increase rapidly with the number of substrates removed M ; In the bottom curve, substrates are randomly removed and the diameter is rarely disturbed. Similar phenomenon was also found in protein network [8]

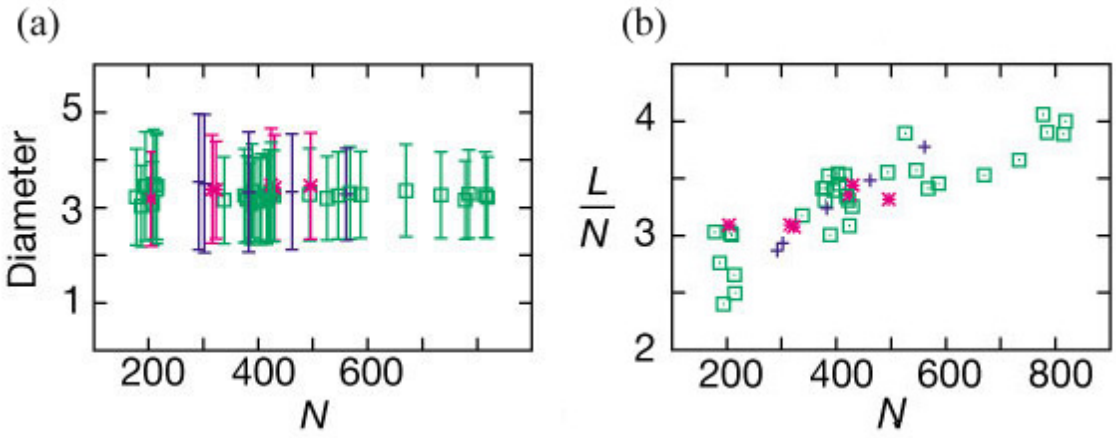

Fig 6. (a) The diameter of metabolic network for each of the 43 organisms [3]. Surprisingly, despite of the difference in total substrates present in each organism, the diameter is the same.

(b) Average incoming links per node for each of the 43 organisms. We can see that with addition of new nodes, the average connectivity increases instead of being a constant.

In both (a) and (b), color of each node indicates the domain the organism belongs to: archaea (magenta), Bacteria (green) and eukaryotes (blue)