



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

C. R. Biologies 326 (2003) 65–74



Biological modelling / Biomodélisation

Networks as constrained thermodynamic systems

Les réseaux : systèmes thermodynamiques contraints

Derek J. Raine^a, Yohann Grondin^{a,b}, Michel Thellier^c, Vic Norris^{c,*}

^a Department of Physics and Astronomy, University of Leicester, Leicester, LE1 7RH, UK

^b UFR « Science et Techniques », université de Cergy-Pontoise, 33, bd du Port, 95011 Cergy-Pontoise, France

^c Laboratoire des processus intégratifs cellulaires, UMR CNRS 6037, faculté des sciences et techniques de Rouen, 76821 Mont-Saint-Aignan, France

Received 10 June 2002; accepted 19 November 2002

Presented by Michel Thellier

Abstract

We show how a network of interconnections between nodes can be constructed to have a specified distribution of nodal degrees. This is achieved by treating the network as a thermodynamic system subject to constraints and then rewiring the system to maintain the constraints while increasing the entropy. The general construction is given and illustrated by the simple example of an exponential network. By considering the constraints as a cost function analogous to an internal energy, we obtain a characterisation of the correspondence between the intensive and extensive variables of the network. Applied to networks in living organisms, this approach may lead to macroscopic variables useful in characterising living systems. **To cite this article:** *D.J. Raine et al., C. R. Biologies 326 (2003).*

© 2003 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

Résumé

Nous montrons comment un réseau de nœuds interconnectés peut être construit de manière à présenter une distribution spécifique de ses degrés de connectivité. Ceci est réalisé en traitant le réseau comme un système thermodynamique qui, soumis à des contraintes, évolue par reconnections en maintenant les contraintes tout en augmentant l'entropie. La construction générale est donnée et illustrée en prenant l'exemple simple d'un réseau exponentiel. En considérant la contrainte comme une fonction de coût, analogue à une énergie interne, nous mettons en évidence la correspondance entre les variables intensives et extensives du réseau. Appliquée aux réseaux dans des organismes vivants, cette approche peut mener aux variables macroscopiques utiles à leur caractérisation. **Pour citer cet article :** *D.J. Raine et al., C. R. Biologies 326 (2003).*

© 2003 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.

Keywords: intensive and extensive variables; network; connectivity; thermodynamics; life; cell

Mots-clés : variables intensives et extensives ; réseau ; connectivité ; thermodynamique ; vie ; cellule

* Correspondence and reprints.

E-mail address: victor.norris@univ-rouen.fr (V. Norris).

Version française abrégée

L'explosion d'intérêt pour les graphes non aléatoires comme représentation des réseaux réels, constitués de nœuds interconnectés, a généré une importante et croissante littérature, aussi bien à partir de l'observation de ces réseaux qu'à partir de la construction de modèles spécifiques. Ces réseaux, pouvant représenter des réactions métaboliques, des régulations génétiques, des relations linguistiques, des graphes sociaux, des connexions Internet, etc., sont approchés par une relation générale entre la connectivité des nœuds, cette dernière étant définie comme le nombre de connexions (non dirigées) liées à un nœud. Cela permet ainsi de caractériser les modèles par la distribution de leurs degrés de connectivité. Par exemple, le nombre de nœuds n_r de connectivité r pour les réseaux *scale free* est donné par une loi de puissance $n_r \propto r^{-\beta}$. Bien qu'il existe plusieurs moyens d'obtenir des exemples spécifiques de réseaux non aléatoires, basés sur différentes méthodes d'évolution des connexions, il n'existe pas encore de voie systématique et générale permettant d'obtenir une distribution donnée, les modèles courants étant trouvés largement par un processus d'essai erreur. De plus, nombre de modèles dynamiques d'évolution de réseaux peuvent mener à une même distribution des degrés de connectivité, ces derniers ne se rapportant donc pas forcément à l'évolution présente.

Dans cet article, nous allons montrer par une approche thermodynamique qu'il est possible de caractériser un réseau présentant une configuration de nœuds ayant une entropie maximale et soumis à des contraintes. Ceci n'a pas un intérêt seulement technique : la contrainte peut en effet être interprétée comme une fonction de coût qui, en principe, représente le coût en énergie libre pour l'établissement du réseau et dont la forme contient implicitement des informations sur l'évolution du système. Nous espérons donc que cette approche permettra de mettre en évidence la structure du réseau, avec des applications possibles aux réseaux métaboliques, génétiques et protéiques. Pour construire un réseau présentant une distribution donnée des degrés de connectivité, nous avons besoin de déterminer la contrainte appropriée menant à cette distribution. Nous ferons alors évoluer les connexions du réseau, choisi arbitrairement, en

maximisant l'entropie. Le tout repose sur une dynamique d'évolution respectant la contrainte.

Nous illustrons cette approche avec une application au cas simple des réseaux exponentiels, qui sont d'une certaine manière analogues à un gaz parfait. Comme nous allons le montrer, cette analogie n'est pas complète, car la probabilité d'avoir une connexion entre nœuds de degrés donnés n'est pas indépendante des degrés. Nous estimons cet effet sur l'entropie du réseau en montrant qu'elle diffère de celle des réseaux aléatoires d'une petite quantité – ce qui, dans un sens, fait que les connexions sont presque aléatoires. Enfin, nous montrons comment la relation entre les variables thermodynamiques intensives et extensives du réseau apparaît naturellement par cette approche. Nous émettons l'idée que la variable intensive des réseaux exponentiels – et par extension pour tout réseau de classe similaire – est une mesure de la complexité du réseau, que nous avons appelé complexité β . Nous montrons comment ce type d'analyse des réseaux dans les organismes vivants promet de fournir les variables macroscopiques appropriées à la vie.

Considérons un système de N nœuds, avec n_r de degré r et ayant la probabilité p_{rs} qu'un nœud de degré r soit connecté à un nœud de degré s . L'entropie des nœuds du réseau (à une constante additive près) est :

$$\Omega = - \sum_r n_r \log n_r$$

qui sera maximisée selon la contrainte

$$\mathcal{C}(n_r) = \text{constante}$$

La méthode habituelle des multiplicateurs de Lagrange donne alors :

$$\frac{\partial \mathcal{C}}{\partial n_r} = -\frac{1}{\beta} \log n_r$$

où la constante additive liée au nombre total de nœuds est négligée, d'où :

$$\mathcal{C} = -\beta^{-1} \sum_r \int dn_r \log n_r$$

Ainsi, pour toute distribution n_r , nous pouvons déterminer la contrainte appropriée.

Pour construire un réseau ayant une contrainte donnée, nous partons d'un réseau arbitraire, un réseau aléatoire ou ordonné, par exemple. Nous faisons alors évoluer les connexions, tout en maintenant

la contrainte, et nous vérifions si cette action a augmenté l'entropie. Si c'est le cas, nous acceptons le nouveau réseau et nous continuons le processus. Sinon, nous acceptons la nouvelle configuration avec une probabilité correspondant à $\exp(-k \Delta \Omega)$, où k est une constante établie par essai erreur afin d'améliorer la convergence de la méthode. Ce processus continue jusqu'à ce que l'entropie, soumise à la contrainte, soit maximisée.

Il est à noter que la similarité avec l'algorithme de Metropolis n'est que superficielle. Nous ne suggérons pas que les propriétés de l'état stable du réseau sont déterminées par une moyenne de l'ensemble. Cette approche permet plutôt d'empêcher le réseau de se fixer dans un état correspondant à un maximum local d'entropie. Cependant, la solution n'est pas attendue pour être unique, dans la mesure où la distribution des degrés de connectivité ne caractérise pas complètement le réseau.

Pour des réseaux exponentiels :

$$n_r \propto r^{-\beta}$$

la contrainte requise est

$$\sum r n_r = \text{constante} \tag{3.1}$$

et la méthode se simplifie considérablement, car la contrainte exprime le fait que la connectivité moyenne d'un réseau exponentiel est constante. Elle peut donc être maintenue par des reconnections aléatoires, qui seront acceptées si elles augmentent l'entropie et acceptées avec une probabilité $\exp(-k \Delta \Omega)$ dans le cas contraire. Dans l'exemple numérique, nous choisissons $k = 10^6$, bien que nous ayons trouvé dans cet exemple que $k \rightarrow \infty$ est tout autant effectif. La valeur constante dans (3.1), c'est-à-dire la connectivité moyenne ν par nœuds, est bien sûr liée au multiplicateur de Lagrange, β . Une simulation numérique est présentée, partant d'un réseau aléatoire de 2000 nœuds, avec $\nu = 10$.

Il est intéressant de comparer la valeur du modèle « entropie » $\sum n_r \log n_r$ avec la vraie valeur statistique de l'entropie prenant en compte les corrélations entre nœuds. Désignons $n_{r,s}$ le nombre de nœuds de degré r connectés aux nœuds de degrés s . Si les connexions étaient aléatoires, la probabilité pour des liens d'un nœud fortement connecté d'être reconnecté à tout autre nœud et la probabilité pour des liens d'un nœud

de faible degré de connectivité d'être reconnecté à tout autre nœud devraient être les mêmes :

$$\frac{n_{rs}}{n_s} = \frac{n_{rp}}{n_p}$$

ce qui correspond à la probabilité de trouver une seule connexion indépendamment du degré des nœuds. Ceci est équivalent à :

$$\frac{n_{rs}}{n_{rp}} = \frac{n_s}{n_p}$$

d'où, par déduction pour un réseau non corrélé :

$$n_{rs} = Q n_r n_s \tag{4.1}$$

$Q = 2\nu/N$ étant approximativement la probabilité moyenne pour qu'une paire de nœuds aléatoirement choisie soit connectée (en prenant $N(N-1) \approx N^2$ pour N grand). Pour un réseau non corrélé, les connexions contribuent à une entropie :

$$-\frac{1}{\mathcal{N}} \sum_{\text{pairs}} [Q \log Q + (1-Q) \log(1-Q)] \tag{4.2}$$

où \mathcal{N} est le nombre de paire de nœuds ou, de manière équivalente, le nombre maximum de liens possibles. L'entropie pour des liens aléatoires est donc

$$-Q \log Q - (1-Q) \log(1-Q)$$

Pour un réseau général, l'entropie est encore donnée par (4.2), mais en utilisant maintenant (4.1) pour définir :

$$Q = Q_{rs} = \frac{n_{rs}}{n_r n_s}$$

D'où enfin le rapport entre l'entropie « des liens » et l'entropie « des nœuds » :

$$\frac{\sum_r n_r \log n_r}{\frac{1}{2NQ} \sum_{r,s} n_{rs} \log n_{rs} - \frac{N}{2} \log Q}$$

correspondant à l'unité pour un réseau aléatoire.

La fonction de coût joue le rôle d'une énergie interne pouvant être changée de deux manières : en reconnectant (ajout de chaleur), ce qui changera la distribution des nœuds, ou par croissance du réseau, sans changer la distribution des nœuds (lors d'un travail) :

$$\delta C = -\beta^{-1} \sum \log n_r \delta n_r - \beta^{-1} \sum \int \delta \log n_r dn_r \tag{5.1}$$

Le premier terme de droite correspond au changement de l'entropie des nœuds (multiplié par β^{-1}). Le second terme de droite représente, d'une certaine manière, le coût par nœuds, associé à la variation des paramètres extérieurs. Le paramètre extérieur pouvant être, par exemple, le nombre total de nœuds ou le nombre total de connexions ; pour chacun des choix X , il y aura une variable intensive x qui lui sera associée et définie par (5.1) :

$$x = \left(\frac{\partial C}{\partial X} \right)_S$$

L'approche thermodynamique fournit donc une caractérisation naturelle des variables macroscopiques associées à tout réseau. D'une certaine manière, c'est la fin de l'histoire, mais nous aimerions clairement acquérir une intuition quant à la nature des variables macroscopiques tout comme pour la pression d'un gaz ou la magnétisation d'un milieu.

Un nombre d'approches a été proposé pour définir la complexité des réseaux. Dans la théorie des graphes, elle est dérivée de la propriété des arbres. La complexité structurale peut être définie en fonction du nombre de paramètres requis pour définir le réseau. La complexité des liens a été quant à elle définie comme la variabilité du second plus court chemin entre nœuds. Toutes ces définitions ont en commun, en contraste avec la notion de complexité algorithmique en informatique, le fait que les systèmes ordonnés et aléatoires ont une complexité nulle.

Nous avons proposé une approche plutôt différente, issue de la considération suivante. En effet, une autre caractéristique commune aux réseaux aléatoires et ordonnés est que l'entropie par nœuds est indépendante du nombre de ceux-ci. Ceci peut être paraphrasé en disant que l'information locale est suffisante pour déterminer la large structure d'échelle des graphes. Nous voulons que la notion de complexité capture l'étendue où ce n'est pas le cas. Ainsi, cet aspect de la complexité des graphes est encodé dans la relation entre la distribution des nœuds liés par une connexion, deux connexions, etc. Les graphes aléatoires et ordonnés restent aléatoires et ordonnés par cette sorte de renormalisation. De manière similaire, les graphes exponentiels et *scale free* restent approximativement les mêmes par cette transformation. Cela confirme que la complexité de tels graphes peut être captée par n'importe quelle paire de nœuds. En d'autres termes, cette

forme de complexité de réseaux peut être exprimée par un simple paramètre. Par exemple, nous pouvons nous concentrer sur les premiers et les derniers liens, à savoir le paramètre de voisinage C et la longueur caractéristique L . Par exemple, pour le réseau de *small world* de Watts et Strogatz, nous montrons que le paramètre C/L a en effet le caractère d'une mesure de la complexité, que nous appelons la complexité β .

Pour les réseaux exponentiels, nous montrons que la complexité β varie selon β^{-3} . Si nous interprétons β comme l'inverse d'une température, cela signifie que la complexité décroît quand la température décroît. Ceci est attendu, car, pour de faibles valeurs de β , la distribution des degrés des nœuds approche une constante inférieure à $1/\beta$. De manière équivalente, le « coût » par nœud, tel qu'il est mesuré par la valeur de la contrainte, décroît comme β^{-1} .

Nous voulons maintenant comparer ceci avec la définition thermodynamique de la complexité des réseaux, qui, d'après nos hypothèses précédentes, sera de la forme :

$$\text{complexité} \propto \left(\frac{\partial C}{\partial X} \right)_S$$

où X est un paramètre extérieur approprié. Au regard des étapes de renormalisation discutées plus haut et du fait que la longueur caractéristique L est de l'ordre du nombre de pas requis pour parcourir un réseau complet, un choix naturel pour x sera $x = L$, c'est-à-dire que la complexité du réseau correspond au changement en coût par changement d'unité de « volume » – où le volume est unidimensionnel. Pour des réseaux exponentiels, cela donne une complexité proportionnelle à β^{-1} . Nous devrions ainsi avoir :

$$\text{complexité} = \beta^{-2} \left(\frac{\partial C}{\partial X} \right)_S$$

Il reste à voir si cela est compatible avec le comportement d'autres réseaux.

En résumé, nous avons montré que la méthode thermodynamique nous permettait de construire des réseaux ayant n'importe quelle distribution des degrés de connectivité. Nous donnerons d'autres exemples par ailleurs. Pour établir si cette approche permet de comprendre les propriétés des réseaux, nous avons besoin de regarder des applications correspondant des systèmes explicites. Par exemple, de nombreux processus biologiques peuvent être décrits par la loi

canonique simplifiée (SCL) de Mandelbrot :

$$n_r \propto \frac{1}{(r + \rho)^\beta}$$

Dans le langage des réseaux thermodynamiques, cette distribution correspond à une fonction de coût, qui décrit le système dans lequel l'information est transmise sous forme de mots avec un coût constant par bit. Cela suggère une correspondance avec le coût de synthèse d'une enzyme. Par ailleurs, le coût du réseau exponentiel croît linéairement avec le degré, ce qui exprime la conservation du nombre de liens dans un réseau. Cette approche pourrait nous permettre d'expliquer, par exemple, la connectivité différentielle dans la régulation des réseaux génétiques de la levure. L'approche thermodynamique nous permet de caractériser les variables macroscopiques d'un réseau et donc, par extension, de n'importe quel système pouvant être décrit en terme de réseau. En particulier donc, l'approche nous offre la possibilité d'une caractérisation macroscopique de la vie.

1. Introduction

The explosion of interest in non-random graphs as representations of real-world networks of connections between nodes has generated a large and growing literature, both in the characterisation of observed networks and on methods of construction of specific models [1]. In all of these cases, the specific contexts, in which networks may represent metabolic reactions [2], genetic regulation [3], linguistic relations [4], social graphs [5], internet connections [6] and so on, are abstracted into a general relation between the connectivity of the nodes. To this end, we define the degree of a node as the number of (undirected) connections to or from that node. The models can then be characterised by the distribution of nodal degrees. For example, in the now well-known, scale-free networks, the number of nodes n_r with r connections is given by a power law $n_r \propto r^{-\beta}$. Although many ways are known to obtain specific examples of non-random networks by evolving the connections according to various schemes [1], there is as yet in general no systematic way of constructing the connections in a network to obtain a given nodal distribution, the current models being found largely by trial and error. Furthermore,

since many dynamical models for the evolution of a network can lead to the same nodal distribution [7,8], the dynamical models may be unrelated to the actual evolution.

In this paper we shall show that a thermodynamic approach allows us to characterise a network as a maximum entropy configuration of nodes subject to an external constraint. This is not only of technical interest: the constraint can be interpreted as a cost function, which in principle represents the cost in free energy of establishing the network. The form of the cost function therefore implicitly contains information on the evolution of the network. We therefore expect this approach to illuminate the structure of networks with possible applications in metabolic, genetic and protein networks.

To construct a network with given nodal distribution, we need to derive the relevant constraints that will lead to the given distribution. This is done in section 2. Starting from a given arbitrary network, we then evolve the connections to maximise the entropy of the network. The trick here is to find an evolutionary dynamics that respects the constraints.

We illustrate the approach with an application to the simple case of exponential networks, which are somewhat analogous to a perfect gas. As we shall show, the analogy is not complete, because the probability of a connection between nodes of given degrees is not independent of the degrees. We estimate the effect of this on the network entropy, showing that it differs from random by a relatively small amount (i.e. that in this sense the connections are close to random). Finally, we show how a relation between intensive and extensive thermodynamic variables for the network arises naturally from this approach. We speculate that the intensive variable for exponential networks (and, by extension, for any similar classes of networks) is a measure of the complexity of the network that we have called β -complexity. We show how this type of analysis of the networks in living organisms this approach promises to provide us with the macroscopic variables appropriate to life.

2. The thermodynamic method

Let us consider a system of N nodes with n_r of degree r and with probability p_{rs} that a node of degree

r is connected to a node of degree s . The entropy of the nodes of the network (up to an additive constant) is:

$$\Omega = - \sum_r n_r \log n_r$$

which we are going to maximise subject to the constraint

$$\mathcal{C}(n_r) = \text{constant}$$

The usual method of Lagrange multipliers gives

$$\frac{\partial \mathcal{C}}{\partial n_r} = -\frac{1}{\beta} \log n_r$$

where we neglect an additive constant related to the total number of nodes, or

$$\mathcal{C} = -\beta^{-1} \sum_r \int dn_r \log n_r$$

Thus, for any distribution n_r , we can construct appropriate constraints.

To avoid confusion, note that the n_r values are not all independent, so the integration cannot be carried out unless they are known explicitly. For example, if we want an exponential network, we have $n_r = n_0 e^{-\beta r}$, hence $\mathcal{C} = \sum_r r n_r$ (up to a constant) as expected.

To construct a network with given constraint, we begin from an arbitrary network, for example a random network or an ordered one. We then make some rewiring, subject to maintenance of the constraints, and test to see if the entropy is increased by this action. If it is, we accept the new network and continue. If it is not, then we accept the new configuration with probability $\exp(-k_1 \Delta \Omega)$, where k is a constant set by trial and error to improve the convergence of the process. The rewiring is continued until the constrained entropy is maximised.

Note that the similarity here to the Metropolis algorithm [9] is superficial only. We are not suggesting that the steady-state properties of the network are determined by an average over this ensemble. Rather the approach is intended to prevent the network settling into configuration with a shallow local maximum of entropy. Even so, the solution is not expected to be unique in those cases where the distribution of nodal degrees does not completely characterise the network. (For example, various different networks with the same scale free node distribution are known [10].)

3. Application to exponential networks

For exponential networks

$$n_r \propto r^{-\beta}$$

the required constraint is:

$$\sum_r r n_r = \text{constant.} \quad (3.1)$$

and the method simplifies considerably. This arises because the constraint expresses the fact that the mean number of links in an exponential network is constant, and this can be maintained by random rewiring with probability unity at each rewiring step. The rewiring is accepted if it increases the entropy and accepted with probability $\exp(-k \Delta \Omega)$ otherwise. In the numerical examples, we choose $k = 10^6$, although we have found in this example that $k \rightarrow \infty$ is equally effective. The value of the constant in (3.1), i.e. the mean degree ν per node, is, of course, related to the Lagrange multiplier β . The results of this numerical computation, starting from a random network of 2000 nodes and $\nu = 10$, is shown in the figures. Fig. 1 shows the evolution of the entropy starting from a random network as it evolves towards exponential. Fig. 2 shows the degree distribution of the nodes, which is indeed exponential. Fig. 3 shows the distribution of the connection probabilities between nodes of degree r and s . Some comment on the form of this

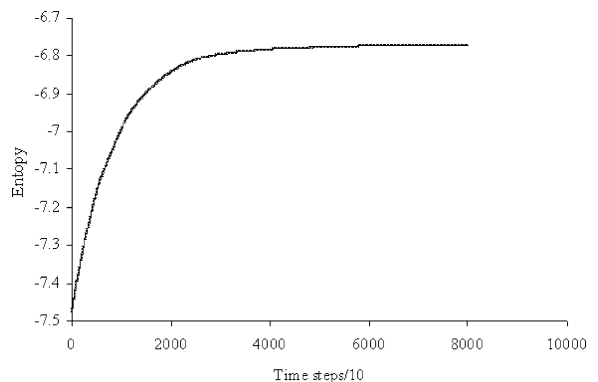


Fig. 1. Evolution of the entropy of the network during the process of maximising the entropy. The number of rewiring trials, after a pair of linked nodes has been randomly chosen, is about 1% of the total of the possible new linkages for one of those nodes. This condition is set to improve the convergence without affecting the random process. Then the total number of rewiring trials is about 1 600 000.

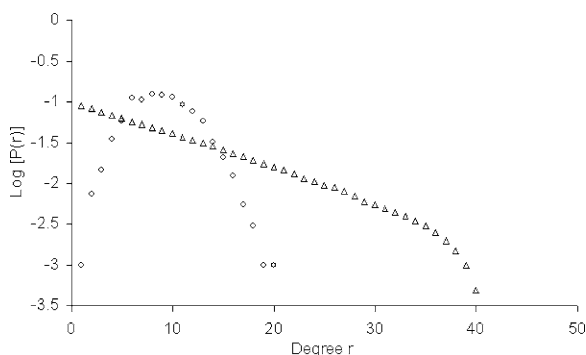


Fig. 2. The degree distribution of the nodes before and after the process of maximisation. The network contains 2000 nodes and is constructed as explained in the text. The empty circles are for the random network (prior to entropy maximisation) and the empty triangles are for the exponential network.

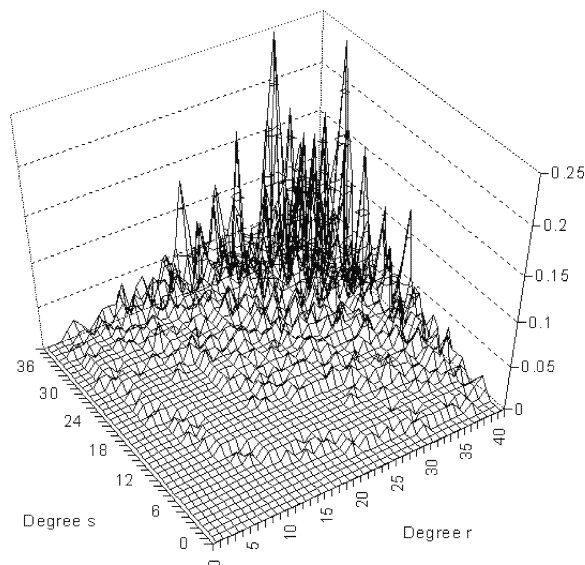


Fig. 3. Distribution of the connection probability in an exponentially connected network between nodes of degree r and s .

figure will be helpful. Note that there is apparently a large probability for nodes of high degree to be connected together, but this arises because of the way the figure has been constructed and results simply because the high-degree nodes have more connections. Most of the connections obviously involve low-degree nodes, because those of high degree are exponentially small in number. As we verify indirectly below, the distribution in Fig. 3 is in fact very close to that for a random network.

4. Entropy of an exponential network

It is of interest to compare the value of the model entropy $\sum n_r \log n_r$ associated with the degree distribution of the nodes with the true statistical entropy of the network, taking into account correlations between the connections of the nodes. Let n_{rs} be the number of nodes of degree r that are connected to nodes of degree s . The probability of finding a connection between a node of degree r and a randomly selected node of degree s is n_{rs}/n_s . If the connections were made randomly, this probability is independent of the degree of the node to which the connection is made, so we should have:

$$\frac{n_{rs}}{n_s} = \frac{n_{rp}}{n_p}$$

for any s and p . This is, equivalently,

$$\frac{n_{rs}}{n_{rp}} = \frac{n_s}{n_p}$$

from which we deduce that $n_{rs} \propto n_s$, hence, by symmetry $n_{rs} \propto n_r n_s$ and, finally:

$$n_{rs} = Q n_r n_s \tag{4.1}$$

for an uncorrelated network. Here $Q = 2v/N$ is approximately the average probability that a randomly chosen pair of nodes is connected (taking $N(N - 1) \approx N^2$ for large N). For an uncorrelated network, the edges contribute to an entropy:

$$-\frac{1}{\mathcal{N}} \sum_{\text{pairs}} [Q \log Q + (1 - Q) \log(1 - Q)] \tag{4.2}$$

where \mathcal{N} is the number of pairs of nodes, or, equivalently, the maximum possible number of links. The entropy from the random links is therefore

$$-Q \log Q - (1 - Q) \log(1 - Q)$$

For a general network, the entropy is still given by (4.2) but with (4.1) now used to define:

$$Q \equiv Q_{rs} = \frac{n_{rs}}{n_r n_s}$$

Fig. 4 shows how the ratio of the entropy contributed by the nodes to that of the links depends on the size and mean connectivity of the exponential network. The ratio is defined in such a way that it is unity

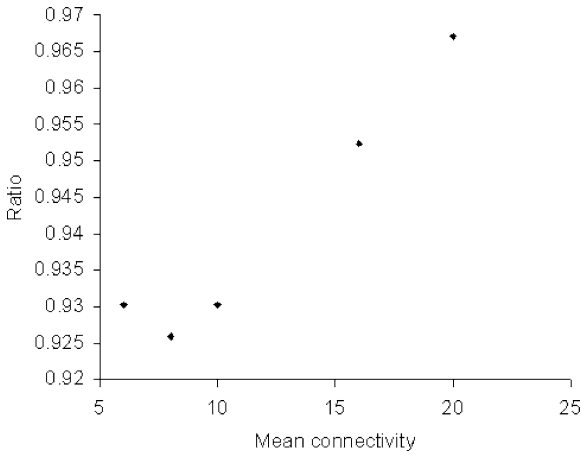


Fig. 4. Ratio of the entropy contributed by the nodes to that of the links plotted against the mean connectivity. This ratio is 1 for random networks.

for a random network, namely as:

$$\frac{\sum_r n_r \log n_r}{\frac{1}{2NQ} \sum_{r,s} n_{rs} \log n_{rs} - \frac{N}{2} \log Q}$$

The figure shows therefore that the links of an exponential network are connected close to random.

5. Intensive and extensive variables

The cost function plays the role of an internal energy, which can be changed in two ways: by rewiring (adding heat) which will change the node distribution, or by growth of the network without changing the node distribution (doing work):

$$\delta C = -\beta^{-1} \sum \log n_r \delta n_r - \beta^{-1} \sum \int \delta \log n_r \, dn_r \quad (5.1)$$

The first term on the right is just the change in entropy of the nodes (multiplied by β^{-1}). The second term on the right represents, in some sense, the cost per node associated with varying the external parameters. The external parameter might be, for example, the total number of nodes, or the total number of connections; for each choice, X , there will be an associated intensive variable, x , defined by (5.1):

$$x = \left(\frac{\partial C}{\partial X} \right)_S$$

The thermodynamic approach therefore provides a natural characterisation of the macroscopic variables associated with any network. In a sense, this is the end of the story, but we would clearly like to acquire an intuitive feeling for the nature of the macroscopic variables associated with a network, just as we derive an intuitive understanding of the pressure of a gas or the magnetisation of a medium. In the next section, we argue the case for regarding the quantity x as a measure of the complexity of a network, related to what we have called β -complexity [11].

6. Network complexity

A number of approaches to defining the complexity of networks have been proposed. In graph theory the complexity is derived from the properties of the spanning trees. Structural complexity [12] can be defined in terms of the number of parameters required to define the network. Edge complexity [13] has been defined in terms of the variability of the second shortest path between nodes. What all of these definitions have in common, and in contrast to the notion of algorithmic complexity in computer science, is that both ordered systems and random ones have zero complexity.

We have proposed a rather different approach to network complexity, which arises from the following consideration. Another feature that random and ordered networks have in common is that the entropy per node is independent of the number of nodes. This can be paraphrased by saying that local information is sufficient to determine the large-scale structure of the graphs. We want the notion of complexity to capture the extent to which this fails to be the case. Thus this aspect of the complexity of a graph is encoded in the relation between the distributions of nodes linked by one connection, two connections and so on. Random graphs and ordered graphs remain random and ordered under these ‘renormalisation’ transformations. Similarly, exponential graphs and scale-free graphs remain approximately exponential or scale-free respectively under these transformations. This suggests that the complexity of such graphs can be captured by any pair of links in this chain of connections. In other words, the complexity of networks of this form can be expressed by a single parameter. For example, we

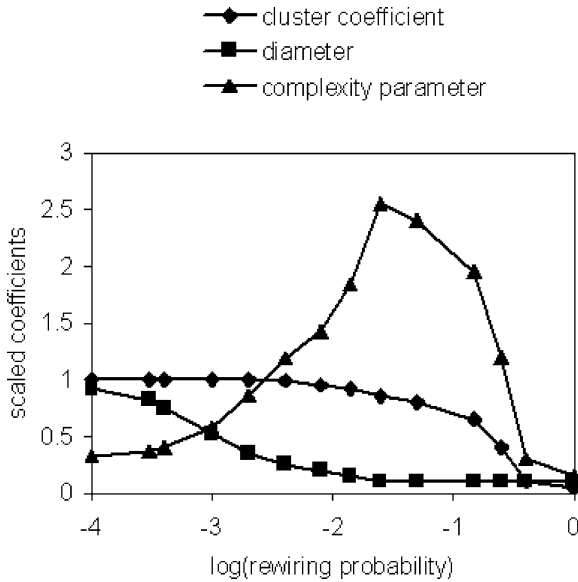


Fig. 5. The ratio of clustering coefficient to characteristic length scale (or diameter) of the small world network of Watts and Strogatz [13] behaves as a complexity parameter. The coefficients are scaled to 1 for $p = 0$.

can concentrate on the first and last links in the chain, namely the cliqueness parameter C and the characteristic path length L [13]. For example, for the small world network of Watts and Strogatz [14], we show that the parameter C/L does indeed have the character of a measure of complexity (Fig. 5), which we have called β -complexity.

For the exponential network we show the average value of C/L as a function of the mean connectivity ν of the nodes in Fig. 6. In terms of the parameter β , the curve is fitted fairly closely by $C/L \propto \beta^{-3}$. If we interpret β as an inverse temperature, then this says that the complexity decreases with decreasing temperature. This is to be expected since for small β the distribution of the degrees of the nodes approaches a constant for degrees $< 1/\beta$. Equivalently, the ‘cost’ per node, as measured by the value of the constraint, decreases with β^{-1} .

We now want to compare this with the thermodynamic definition of network complexity, which, according to our hypothesis above, will be of the form:

$$\text{complexity} \propto \left(\frac{\partial C}{\partial X} \right)_S$$

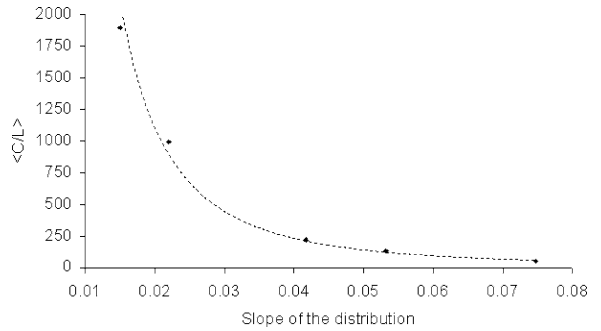


Fig. 6. Value C/L against the slope of the exponential distribution (in the log-linear plot of Fig. 2). The graph is plotted for the following mean connectivities, respectively 6, 8, 10, 16, 20. The curve is fitted closely by a power function with exponent close to -2.25 .

where X is an appropriate external parameter. In view of the renormalisation steps discussed above, and of the fact that the characteristic length L is of the order of the number of steps required to make a complete network, a natural choice for x is $x = L$, i.e. the complexity of a network is the change in cost per unit change in ‘volume’ (where ‘volume’ is here one-dimensional). For the exponential network, this gives a complexity $\propto \beta^{-1}$. Thus we should have:

$$\beta\text{-complexity} = \beta^{-2} \left(\frac{\partial C}{\partial X} \right)_S$$

in order to get a β -complexity of β^{-3} . It remains to be seen if this is compatible with the behaviour of other networks.

7. Conclusions

We have shown that the thermodynamic method allows us to construct networks of any given node distribution. We shall give some other examples elsewhere. To establish whether the thermodynamic ideas in this paper can provide any further insight into network properties, we need to look at applications to explicit systems. For example, many biological processes can be described by the Mandelbrot simplified canonical distribution (SCL) [15]:

$$n_r \propto \frac{1}{(r + \rho)^\beta}$$

In the language of thermodynamic networks, this distribution corresponds to a cost function that de-

scribes a system in which information is transmitted in words with a constant cost per bit. This suggests a correspondence with the cost of synthesising enzymes. On the other hand, the exponential network has a cost that grows linearly with rank expressing the conservation of the number of links in a network. This approach may enable us to explain, for example, the different connectivity in the regulation of the genetic network in yeast [3].

The thermodynamic approach allows us to characterise the macroscopic variables of a network, and, hence, by extension, of any system that can be described in network terms. In particular, therefore, the approach holds out the possibility of a macroscopic characterisation of living systems.

Acknowledgements

We have benefited from conversations with Camille Ripoll, Jeremy Ramsden and Steve Gurman.

References

- [1] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.* 74 (2002) 47–97.
- [2] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks, *Nature* 407 (2000) 651–654.
- [3] N. Guelzim, S. Bottani, P. Bourguin, F. Kepes, Topological and causal structure of the yeast transcriptional regulatory network, *Nat. Genet.* 31 (2002) 60–63.
- [4] I. Ferrer, R.F. Cancho, R.V. Sole, The small world of human language, *Proc. R. Soc. Lond. B Biol. Sci.* 268 (2001) 2261–2265.
- [5] F. Liljeros, C.R. Edling, L.A. Amaral, H.E. Stanley, Y. Aberg, The web of human sexual contacts, *Nature* 411 (2001) 907–908.
- [6] A. Medina, I. Matta, J. Byers, On the origin of power laws in internet topology, *SIGCOMM Comput. Commun. Rev.* 30 (2) (2000).
- [7] P.L. Krapivsky, S. Redner, F. Leyvraz, Connectivity of growing random networks, *Phys. Rev. Lett.* 85 (2000) 4629–4632.
- [8] K. Klemm, V.M. Eguiluz, Growing scale-free networks with small-world behaviour, *Phys. Rev. E* 65 (2002) 057102.
- [9] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* 21 (1953) 1087–1092.
- [10] P.L. Krapivsky, S. Redner, Organization of growing random networks, *Phys. Rev. E* 63 (2001) 066123.
- [11] D.J. Raine, V. Norris, Network complexity, in: P. Amar, F. Kepes, V. Norris, P. Tracqui (Eds.), *Modeling and simulation of biological processes in the context of the genome*, Conference Proceedings, Autrans, France, 2002, pp. 67–75.
- [12] J.P. Crutchfield, The calculi of emergence – computation, dynamics and induction, *Physica D* 75 (1994) 11–54.
- [13] D.J. Watts, *Small worlds: the dynamics of networks between order and randomness*, Princeton UP, NJ, USA, 1999.
- [14] D.J. Watts, S.H. Strogatz, Collective dynamics of ‘small-world’ networks, *Nature* 393 (1998) 440–442.
- [15] J.J. Ramsden, J. Vohradsky, Zipf-like behaviour in prokaryotic protein expression, *Phys. Rev. E* 58 (1998) 7777–7780.