**THE EUROPEAN
PHYSICAL JOURNAL B**

Regular Article

Assortativeness and information in scale-free networks

M. Piraveenan[1,2], M. Prokopenko[1,a], and A.Y. Zomaya[2]

[1] CSIRO Information and Communications Technology Centre, Locked Bag 17, North Ryde, NSW 1670, Australia
[2] School of Information Technologies, The University of Sydney, NSW 2006, Australia

Abstract. We analyze Shannon information of scale-free networks in terms of their assortativeness, and identify classes of networks according to the dependency of the joint remaining degree distribution on the assortativeness. We conjecture that these classes comprise minimalistic and maximalistic networks in terms of Shannon information. For the studied classes, the information is shown to depend non-linearly on the absolute value of the assortativeness, with the dominant term of the relationship being a power-law. We exemplify this dependency using a range of real-world networks. Optimization of scale-free networks according to information they contain depends on the landscape of parameters' search-space, and we identify two regions of interest: a slope region and a stability region. In the slope region, there is more freedom to generate and evaluate candidate networks since the information content can be changed easily by modifying only the assortativeness, while even a small change in the power-law's scaling exponent brings a reward in a higher rate of information change. This feature may explain why the exponents of real-world scale-free networks are within a certain range, defined by the slope and stability regions.

PACS. 89.75.Hc Networks and genealogical trees – 89.75.Da Systems obeying scaling laws – 89.70.Cf Entropy and other measures of information – 89.75.Fb Structures and organization in complex systems

1 Introduction

Various network growth models have been proposed and studied to emulate the features of the real world networks [1–6]. One prominent model is the preferential attachment model, which explains scale-free power law degree distributions observed in many real world networks [4–6]. In this model, the probability of a new node making a link to an existing node in the network is proportional to the degree of the target node. As pointed out by Newman [2], this model does not take into account the degree of the source node in influencing the attachment probability. In the real world, however, many networks show the tendency where highly connected nodes are more likely to make links with other highly connected links (i.e., to mix assortatively). The reverse is also true with some networks, whereby highly connected nodes are more likely to make links with more isolated, less connected, nodes (i.e., to mix disassortatively). In both cases, the likelihood of creating a link depends on the degrees of both nodes. Assortativeness is a measure that quantifies this tendency for preferential association [2]. It has been suggested that the extent of assortativeness affects network's resilience under node removal or percolation [2,7]. Both assortative and disassortative mixing should be contrasted with non-assortative mixing, where preferential connection cannot be established.

Naturally occurring networks display various extents of assortative mixing, and it is often possible to measure or calculate the level of assortativeness in these networks [1]. In this paper we analyze the relationship between network assortativeness and Shannon information within the network. Solé and Valverde [1] defined the information transfer simply as mutual information contained in the network, or the amount of general correlation between network nodes – we refer to this quantity as (Shannon) information content. Importantly, the maximum attainable information content defines the network's capacity, in analogy with the information-theoretic notion of channel capacity – the maximum mutual information for the channel over all possible distributions of a transmitted signal. In general, information content is a vital indicator of complex non-linear behavior in self-organizing systems, and can be associated with *predictive information*, richness of structure (i.e. *excess entropy*), and *physical complexity* [8]. We advance the conjecture that the information content within a scale-free network increases nonlinearly with the absolute value of the assortativeness.

2 Definitions

We study assortativeness in scale-free networks described by power law degree distributions, formally specified as $P(k) = Ak^{-\gamma}u(k/N_p)$ where u is a step function specifying a cut-off at $k = N_p$.

a e-mail: mikhail.prokopenko@csiro.au

To formally define assortativeness, we need to first define a number of concepts. The degree of a node is the number of other nodes to which it is connected to. Let us consider a network with N nodes (vertices) and M links (edges), and say that the probability of a randomly chosen node having degree k is p_k, where $1 \leq k \leq N_p$. The distribution of such probabilities is called the *degree distribution* of the network. However, if we reach a node by following a randomly chosen link, we will see that the remaining number of links (the remaining degree) of this node is not distributed according to p_k. Instead it is biased in favour of nodes of high degree, since more links end at a high-degree node than at a low-degree one [2]. The distribution of such remaining degrees is called the *remaining degree distribution*. It is related to the original degree distribution as follows:

$$q_k = \frac{(k+1)p_{k+1}}{\sum_j^{N_p} j p_j}, \quad 0 \leq k \leq N_p - 1 \qquad (1)$$

where p_k is the degree distribution of the network, and q_k is the remaining degree distribution of the network [2]. For scale-free networks, equation (1) yields that if $\gamma = 1$ (that is, $p(k) = A/k$ before the cut-off), the resulting remaining degree distribution is uniform, therefore symmetric. For any other value of γ, the remaining degree distribution is asymmetric.

Following Callaway et al. [3] and Newman [2], we can now define the quantity $e_{j,k}$ to be the joint probability distribution of the remaining degrees of the two nodes at either end of a randomly chosen link. As pointed out by Newman [2], this quantity is symmetric in its indices for an undirected graph, that is $e_{j,k} = e_{k,j}$, and it obeys the sum rules $\sum_{jk} e_{j,k} = 1$ and $\sum_j e_{j,k} = q_k$.

Following Newman [4], we formally define assortativeness as a correlation function which is zero for non-assortative mixing and positive or negative for assortative or disassortative mixing respectively. It is defined as

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{j,k} - q_j q_k) \qquad (2)$$

where σ_q^2 is the variance of q_k. Here r lies between -1 and 1, whereby $r = 1$ means perfect assortativeness, $r = -1$ means perfect disassortativeness, and $r = 0$ means no assortativeness (random linking).

If a network has perfect assortativeness ($r = 1$), then all nodes connect only with nodes with the same degree. For example, the joint distribution $e_{j,k} = q_k \delta_{j,k}$ where $\delta_{j,k}$ is the Kronecker delta function, produces a perfectly assortative network. If the network has no assortativeness ($r = 0$), then any node can randomly connect to any other node. A sufficiency condition for a non-assortative network is $e_{j,k} = q_j q_k$. This is not a necessary condition: other $e_{j,k}$ may also produce non-assortativeness[1].

---

[1] For example, for the uniform remaining degree distribution q_k, the distribution $e_{j,k} = [q_j \delta_{j,k} + q_j \delta_{j,(N_p-1-k)}]/2$ will produce a non-assortative network.

Newman noted that if a network is perfectly disassortative then every link connects two nodes of different degrees (types) [9]. However, this requirement is not sufficient to generate an $e_{j,k}$ resulting in $r = -1$. In fact, the $r = -1$ case is possible only for symmetric degree distributions where $q_k = q_{(N_p-1-k)}$, and $e_{j,k} = q_k \delta_{j,(N_p-1-k)}$. In other words, for a network with remaining degrees $0, \ldots, N_p - 1$, a node with a degree k must be linked to a node with a degree $N_p - 1 - k$. Nodes with identical degrees may still be connected in a perfectly disassortative network (e.g., when their degree j is precisely in the middle of the distribution q, i.e., N_p is odd and $j = (N_p - 1)/2$).

Perfect disassortativeness is not possible for non-symmetric degree distributions q, because the $e_{j,k}$ distribution must obey the rules $e_{j,k} = e_{k,j}$, as well as $\sum_j e_{j,k} = q_k$. We denote the maximum attainable disassortativeness as r_m, where $r_m < 0$ ($r_m = -1$ only for symmetric q_k). This limit and the corresponding $e_{j,k}^{(r=r_m)}$ can be obtained, given the distribution q_k, via a suitable minimization procedure by varying $e_{j,k}$ under its constraints.

In general, we would like to distinguish between difficulties in (i) constructing an $e_{j,k}$ distribution for a given degree distribution $p(k)$, and (ii) growing the network for the calculated link distribution $e_{j,k}$. When one is constructing an $e_{j,k}$ distribution for a given degree distribution $p(k)$, the cases of maximum disassortativity and maximum assortativity differ. On the other hand, when one is growing the network with the given $e_{j,k}$, it may also not be possible to achieve $r = 1$ for a given $p(k)$ or $q(k)$. This is despite the fact that the required $e_{j,k}$ can be obtained – the reason is that the network may not be large enough to accommodate all the necessary connections. Thus, the maximum limit of assortativity may also need to be considered for networks that can be actually constructed.

3 Information content of networks

Information Theory was originally developed by Shannon [10] for reliable transmission of information from a source X to a receiver Y over noisy communication channels. Put simply, it addresses the question of "how can we achieve perfect communication over an imperfect, noisy communication channel?" [11]. When dealing with outcomes of imperfect probabilistic processes, it is useful to define the information content of an outcome x which has the probability $P(x)$, as $\log_2 \frac{1}{P(x)}$ (it is measured in bits): improbable outcomes convey more information than probable outcomes. Given a probability distribution P over the outcomes $x \in \mathcal{X}$ (i.e., over a discrete random variable X representing the process), and defined by the probabilities $P(x) \equiv P(X = x)$ given for all $x \in \mathcal{X}$, the average Shannon information content of an outcome is determined by

$$H(X) = -\sum_{x \in \mathcal{X}} P(x) \log P(x), \qquad (3)$$

henceforth we omit the logarithm base 2. This quantity is known as *(information) entropy*. Intuitively, it measures, also in bits, the amount of freedom of choice (or the degree of randomness) contained in the process – a process with many possible outcomes has high entropy. This measure has some unique properties that make it specifically suitable for measuring "how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?" [10]. In answering this question, Shannon required the following properties for such a measure H:

– continuity: H should be continuous in the probabilities, i.e., changing the value of one of the probabilities by a small amount changes the entropy by a small amount;
– monotony: if all the choices are equally likely, e.g. if all the probabilities $P(x_i)$ are equal to $1/n$, where n is the size of the set $\mathcal{X} = \{x_1, \ldots, x_n\}$, then H should be a monotonic increasing function of n: "with equally likely events there is more choice, or uncertainty, when there are more possible events" [10];
– recursion: H is independent of how the process is divided into parts, i.e. "if a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H" [10],

proving that entropy function $-K \sum_{i=1}^{n} P(x_i) \log P(x_i)$, where a positive constant K represents a unit of measure, is the only function satisfying these three requirements.

The joint entropy of two (discrete) random variables X and Y is defined as the entropy of the joint distribution of X and Y:

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log P(x, y), \qquad (4)$$

where $P(x, y)$ is the joint probability. The conditional entropy of Y, given random variable X, is defined as follows:

$$H(Y|X) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log \frac{P(x)}{P(x, y)} = H(X, Y) - H(X).$$
$$(5)$$

This measures the average uncertainty that remains about $y \in \mathcal{Y}$ when $x \in \mathcal{X}$ is known [11].

Mutual information $I(X; Y)$ measures the amount of information that can be obtained about one random variable by observing another (it is symmetric in terms of these variables):

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \ \log \frac{P(x, y)}{P(x)P(y)}. \qquad (6)$$

Mutual information $I(X; Y)$ can also be expressed via the conditional entropy:

$$I(X; Y) = H(Y) - H(Y|X). \qquad (7)$$

The amount of information $I(X; Y)$ shared between transmitted X and received Y signals is often maximized by designers of communication channels, via choosing the best

possible transmitted signal X. Channel capacity is defined as the maximum mutual information for the channel over all possible distributions of the transmitted signal X (the source). The conditional entropy $H(Y|X)$ is also called the equivocation of Y about X, and thus, informally, the mutual information $I(X; Y)$ is equal to the difference between receiver's diversity $H(Y)$ and the equivocation of receiver about source $H(Y|X)$. Hence, the channel capacity is optimized when receiver's diversity is maximized, while its equivocation about the source is minimized.

Let us define the network's information content (called information transfer by Solé and Valverde [1]):

$$I(q) = H(q) - H(q|q') \qquad (8)$$

where the first term is the Shannon entropy of the network, $H(q) = - \sum_{k=0}^{N_p - 1} q_k \log(q_k)$, that measures the diversity of the degree distribution or the network's heterogeneity, and the second term is the conditional entropy defined via conditional probabilities of observing a node with k links leaving it, provided that the node at the other end of the chosen link has k' leaving links. Importantly, the conditional entropy $H(q|q')$ estimates correlations in the network created by connecting the nodes with dissimilar degrees – this component affects the overall diversity or the heterogeneity of the network, but does not contribute to the amount of information within it. Informally, information content within the network is the difference between network's heterogeneity and assortative noise within it [1].

In information-theoretic terms, $H(q|q')$ is the assortative noise within the network's information channel, i.e., it is the non-assortative extent to which the preferential (either assortative or disassortative) connections are obscured [8]. Given the joint remaining degree distributions, the information content can be expressed as:

$$I(q) = \sum_{j=0}^{N_p - 1} \sum_{k=0}^{N_p - 1} e_{j,k} \log \frac{e_{j,k}}{q_j q_k} \qquad (9)$$

Shannon information $I(q)$ is a better, more generic measure of dependence than the correlation functions that measure linear relations. Mutual information measures the general dependence and is thus a less biased statistic [1]. Shannon information (9) can also be seen as the Kullback-Leibler divergence $K(e_{j,k} \| q_j q_k)$ [12] (i.e., relative entropy) of the product of two marginal distributions q from the joint distribution e. This divergence amounts to the expected number of extra bits that must be transmitted in order to identify (on average) remaining degrees of connected nodes j and k of the link (j, k) if they are assigned using only the marginal distribution q, instead of the joint distribution $e_{j,k}$. It is evident that maximal information $I(q)$ is attained when the product $q_j q_k$ diverges the most from the joint distribution $e_{j,k}$, and minimal information $I(q)$ is attained when the product $q_j q_k$ and the joint distribution $e_{j,k}$ diverge the least.

The entropy and information content described above are defined with respect to the degree distribution and

joint degree distribution, and there are alternative definitions where the entropy of the network could be characterized by higher order correlations or community structure [15].

It should also be noted that Shannon information contains no inherent directionality, and various alternatives have been proposed. For example, transfer entropy [13] measures the average information contained in the source about the next state of the destination that was not already contained in the destination's past. It can be argued that transfer entropy is the appropriate measure for *predictive* information transfer in spatiotemporal systems [14]. In this work we follow Solé and Valverde [1] in using the mutual information to represent information content (and not a directional transfer) within a network.

Solé and Valverde [1] were among the first to empirically analyse the relationship between assortativeness and Shannon information, using a set of real world networks. Their conclusion was that the information (transfer) and assortativeness are correlated in a negative way: the extent of disassortativeness increases with mutual information (see Fig. 7 in [1]). First of all, we argue that networks with the same assortativeness r and the same distribution q_k could have different information contents I – because they may disagree on $e_{j,k}$. Secondly, we intend to demonstrate a more accurate dependency for networks with power law degree distributions: *the information content non-linearly and asymmetrically depends on the absolute value of the assortativeness*, i.e. mutual information increases when assortativeness varies from a critical point $\overline{r}$, in either positive or negative direction. This relationship subsumes the one implied by Solé and Valverde [1].

Most of the real world networks studied by Solé and Valverde [1] did not show perfect (or nearly perfect) assortativeness or disassortativeness: the observed values were between 0.4 and −0.2. Moreover, the compared networks did not agree on average degree, degree distribution, etc. These reasons obscured the conclusion reported in [1]. Figure 1 demonstrates that Shannon information is not negatively correlated with assortativeness (as conjectured by Solé and Valverde [1]), but is correlated with the absolute value of the assortativeness. The networks shown in Figure 1 include the scale-free networks considered by Solé and Valverde (Tab. 1 in [1], including some technological and biological networks), metabolic substrate networks (see Tab. 1), metabolic substrate networks without inorganic components [16], transcription networks (see Tab. 2), protein-protein interaction networks (see Tab. 3), Internet at autonomous systems (AS) level (see Tab. 4), citation networks (see Tab. 5), and collaboration networks (see Tab. 6). The correlation between assortativeness and Shannon information is non-linear, and we study this relationship for scale-free networks in the Results and Analysis section.

4 Classification

In establishing the relationship between $I(q)$ and r, we classify networks according to the dependency of the dis-
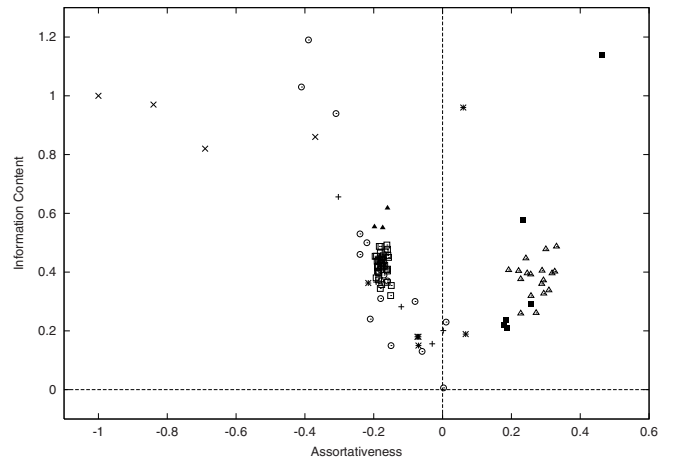

Fig. 1. Shannon information and assortativeness for different networks. Circles: networks considered by Solé and Valverde (Tab. 1 in [1]); squares: metabolic substrate networks; triangles: metabolic substrate networks without inorganic components [16]; crosses: transcription networks; stars: protein-protein interaction networks; filled triangles: Internet AS; pluses: citation networks; filled squares: collaboration networks.

tribution $e_{j,k}^{(r=r')}$ on the assortativeness r'. Within a class, the same distribution q_k and the same assortativeness r result in the same information content $I(r)$ (of course, when distribution q_k changes, the same r will typically correspond to different values I).

As noted, the assortativeness r is defined in terms of the joint distribution $e_{j,k}$. We assert that if the distribution $e_{j,k}$ is given by the linear decomposition (10) for a real number $r' > 0$, then the network assortativeness is precisely r':

$$e_{j,k}^{(r=r')} = r' \left(e_{j,k}^{(r=1)} - e_{j,k}^{(r=0)} \right) + e_{j,k}^{(r=0)} \qquad (10)$$

where $e_{j,k}^{(r=1)} = q_k \delta_{j,k}$ and $e_{j,k}^{(r=0)} = q_j q_k$. This is a sufficient but not necessary condition. A similar sufficient condition also exists for $r' < 0$:

$$e_{j,k}^{(r=r')} = -\frac{r'}{r_m} \left(e_{j,k}^{(r=r_m)} - e_{j,k}^{(r=0)} \right) + e_{j,k}^{(r=0)} \qquad (11)$$

where r_m is the maximum attainable disassortativeness. For symmetric distributions q_k, the condition reduces to

$$e_{j,k}^{(r=r')} = r' \left(e_{j,k}^{(r=-1)} - e_{j,k}^{(r=0)} \right) + e_{j,k}^{(r=0)} \qquad (12)$$

where $e_{j,k}^{(r=-1)} = q_k \delta_{j,(N_q-1-k)}$. These assertions can be verified by substituting templates (10)–(12) into equation (2). The templates (10)–(12) define a class of networks, *class A*. As intended, the same distribution q_k and the same assortativeness r result in the same value $I(r)$ within the class. This is so simply because the templates define a unique distribution $e_{j,k}^{(r=r')}$ for a given r', and the distribution $e_{j,k}^{(r=r')}$ yields a unique information $I(r')$

Table 1. Shannon information $I(r)$ computed for metabolic (substrate) networks and their corresponding class-A and class-B networks. N is the total number of substrates, temporary substrate-enzyme complexes, and enzymes [18].

Network	N	N_p	γ	r	$I_A(r)$	$I(r)$	$I_B(r)$
A. pernix	517	86	2.2	−0.181	0.14	0.34	0.75
A. fulgidus	1281	191	2.2	−0.173	0.14	0.42	0.66
M. thermoautotroph.	1138	167	2.2	−0.182	0.18	0.44	0.64
M. jannaschii	1103	160	2.2	−0.176	0.17	0.45	0.70
P. furiosus	790	114	2.1	−0.177	0.16	0.44	0.85
P. horikoshii	807	111	2.1	−0.176	0.22	0.43	0.80
A. aeolicus	1105	147	2.1	−0.193	0.19	0.38	0.67
C. pneumoniae	412	67	2.2	−0.151	0.11	0.32	0.68
C. trachomatis	467	80	2.3	−0.149	0.16	0.35	0.70
Synechocystis sp.	1486	233	2.1	−0.192	0.22	0.45	0.79
P. gingivalis	1052	161	2.2	−0.171	0.13	0.44	0.64
M. bovis	1102	193	2.2	−0.163	0.12	0.41	0.89
M. leprae	1106	177	2.2	−0.18	0.14	0.42	0.72
M. tuberculosis	1534	252	2.1	−0.179	0.20	0.43	0.86
B. subtilis	2217	410	2.1	−0.159	0.18	0.46	0.68
E. faecalis	1049	166	2.1	−0.186	0.18	0.40	0.90
C. acetobutylicum	1349	200	2.1	−0.187	0.17	0.44	0.86
M. genitalium	490	75	2.3	−0.184	0.13	0.49	0.42
M. pneumoniae	420	61	2.2	−0.189	0.05	0.40	0.50
S. pneumoniae	1116	180	2.1	−0.186	0.15	0.40	0.62
S. pyogenes	1087	176	2.1	−0.189	0.19	0.42	0.67
C. tepidum	953	136	2.1	−0.182	0.14	0.44	0.76
R. capsulatus	1808	283	2.1	−0.178	0.19	0.44	0.72
R. prowazekii	469	71	2.3	−0.161	0.08	0.37	0.67
N. gonorrhoeae	1104	169	2.1	−0.19	0.19	0.44	0.65
N. meningitidis	1032	160	2.2	−0.189	0.16	0.42	0.65
C. jejuni	993	153	2.2	−0.186	0.19	0.42	0.68
H. pylori	996	140	2.1	−0.196	0.21	0.45	0.74
E. coli	2316	430	2.1	−0.162	0.21	0.49	0.66
S. typhi	2403	444	2.2	−0.16	0.21	0.48	0.66
Y. pestis	1534	254	2.1	−0.168	0.19	0.42	0.86
A. actinomycetemcomit.	1046	154	2.1	−0.185	0.16	0.40	0.75
H. influenzae	1484	222	2.2	−0.179	0.20	0.48	0.63
P. aeruginosa	2023	364	2.1	−0.16	0.22	0.41	0.82
T. pallidum	506	87	2.2	−0.177	0.17	0.36	0.69
B. burgdorferi	433	78	2.3	−0.16	0.11	0.40	0.82
T. maritima	863	129	2.1	−0.186	0.19	0.37	0.67
D. radiodurans	2337	433	2.1	−0.157	0.24	0.45	0.60
E. nidulans	976	157	2.1	−0.177	0.14	0.43	1.00
S. cerevisiae	1559	260	2.1	−0.181	0.18	0.47	0.98
C. elegans	1207	208	2.1	−0.173	0.10	0.42	1.51
O. sativa	708	99	2.2	−0.167	0.09	0.47	1.50
A. thaliana	737	108	2.2	−0.172	0.09	0.46	1.51

Table 2. Shannon information $I(r)$ computed for transcription networks and their corresponding class-A and class-B networks. N is the total number of transcription factors [19].

Network	N	N_p	γ	r	$I_A(r)$	$I(r)$	$I_B(r)$
C. diptheria	71	63	8.49	−0.84	0.85	0.97	0.99
C. efficiens	50	27	8.5	−0.69	0.65	0.82	0.83
C. glucamitum	539	104	1.86	−0.37	0.44	0.86	0.88
C. jeikeium	52	51	8.49	−1.00	1.00	1.00	1.00

Table 3. Shannon information $I(r)$ computed for protein-protein interaction networks and their corresponding class-A and class-B networks. N is the total number of proteins [20].

Network	N	N_p	γ	r	$I_A(r)$	$I(r)$	$I_B(r)$
H. pylori	714	54	1.26	−0.216	0.18	0.36	0.71
M. musculus	502	12	1.96	−0.073	0.12	0.18	1.17
H. sapien	1529	39	1.62	0.067	0.09	0.19	1.10
D. melanogaster	7485	178	1.17	−0.07	0.06	0.15	0.83
S. cerevisiae	502	12	1.96	−0.07	0.12	0.18	1.17
E. coli	1861	152	1.15	0.06	0.04	0.96	1.03

Table 4. Shannon information $I(r)$ computed for Internet and their corresponding class-A and class-B networks. N is the total number of autonomous systems [21].

Network	N	N_p	γ	r	$I_A(r)$	$I(r)$	$I_B(r)$
AS 1998	3216	642	1.36	−0.198	0.20	0.55	0.58
AS 1999	4513	1018	1.21	−0.174	0.21	0.55	0.58
AS 2000	6474	1460	1.18	−0.16	0.18	0.62	0.83

Table 5. Shannon information $I(r)$ computed for citation networks and their corresponding class-A and class-B networks. N is the total number of cited papers [21].

Network	N	N_p	γ	r	$I_A(r)$	$I(r)$	$I_B(r)$
Scientometrics	2729	164	2.84	−0.03	0.03	0.16	0.88
Small & Griffith	1024	232	2.77	−0.193	0.08	0.37	0.76
Self-organizing maps	3773	740	2.88	−0.12	0.06	0.28	0.46
Small World	233	294	2.5	−0.303	0.15	0.66	0.72
Zewail	6652	331	2.63	0.002	0.03	0.20	0.21

Table 6. Shannon information $I(r)$ computed for collaboration networks and their corresponding class-A and class-B networks. N is the total number of authors [22].

Network	N	N_p	γ	r	$I_A(r)$	$I(r)$	$I_B(r)$
Astro Physics	16 046	360	2.71	0.235	0.16	0.58	0.58
Condensed matter	16 264	107	2.79	0.185	0.05	0.24	0.52
Condensed matt. 2003	30 460	202	2.74	0.178	0.09	0.22	0.40
Condensed matt. 2005	39 577	278	2.72	0.186	0.11	0.21	0.39
High-Energy Theory	7610	50	2.97	0.258	0.27	0.29	1.89

where $e_{j,k}^{(r=1)}$ and $e_{j,k}^{(r=r_m)}$, including $e_{j,k}^{(r=-1)}$ which replaces $e_{j,k}^{(r=r_m)}$ for symmetric distributions, are computed as for the class A templates. For a non-assortative class B network, the joint probability $e_{j,k}^{(r=0)}$ is the average between the corresponding probabilities of perfectly assortative and disassortative networks: $[e_{j,k}^{(r=1)} + e_{j,k}^{(r=r_m)}]/2$.

5 Results and analysis

5.1 Minimalistic and maximalistic networks

We computed Shannon information for a wide range of degree distributions by substituting the corresponding templates into equation (9). While a degree distribution can be characterized in terms of many properties, e.g. the average degree, the power law exponent γ, and the cut-off N_p, there are only two independent variables in any such characterization, and we choose the exponent γ and the

according to equation (9). In particular, Shannon information within a non-assortative class A network (i.e., $r' = 0$) is zero: $I(0) = 0$.

Among many other possible classes, we define another class, *class B*, by the following template:

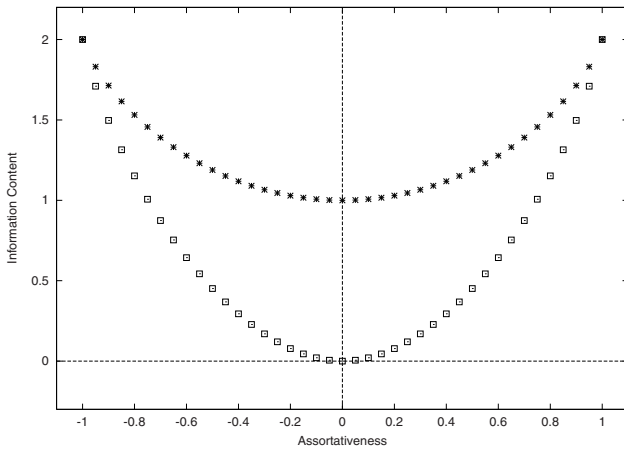$$e_{j,k}^{(r=r')} = \frac{r'+1}{2}\, e_{j,k}^{(r=1)} - \frac{r'-1}{2}\, e_{j,k}^{(r=r_m)} \qquad (13)$$

Fig. 2. Relationship between Shannon information and assortativeness for class-A (squares) and class-B (stars) networks, $\gamma = 1.0, N_p = 4$.
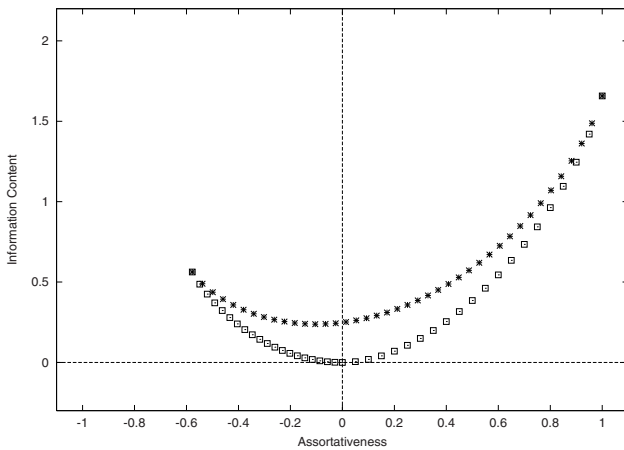

Fig. 3. Relationship between Shannon information and assortativeness for class-A (squares) and class-B (stars) networks, $\gamma = 2.3, N_p = 4$.

cut-off N_p as our independent variables[2]. It is worth pointing out that the constraints imposed by the connectivity structure of networks of finite size generate spontaneous correlations which in turn may introduce a structural cut-off N_p that possibly differs from the natural one [17].

Figure 2 (symmetric q_k with $\gamma = 1.0$) and Figure 3 (asymmetric q_k with $\gamma = 2.3$) show Shannon information for both class A and class B networks.

For $\gamma \neq 1$, the distribution q_k is not symmetric, preventing perfect disassortativeness, and therefore, it is not possible to get close to the ($r = -1$) case as shown by Figure 3.

Our conjecture is that the template defining the class A networks is the *minimalistic* linear template: that is, the information $I(r)$ for the class A is minimal for a given r. In other words, any real-world network with the same assortativeness r, and the degree distribution parameters γ and N_p, should have higher information $I(r)$. Similarly,
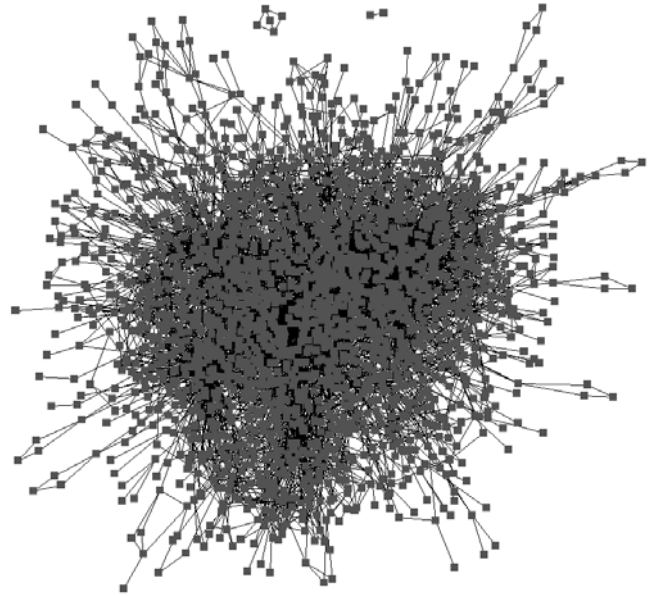

Fig. 4. The *Escherichia coli* metabolic network: $r = -0.162, I(r) = 0.49$ bits; $\gamma = 2.1, N_p = 430$. Figure is drawn with Cytoscape 2.5.1.

the template defining the class B networks is, we believe, the *maximalistic* linear template. That is, the information $I(r)$ for the class B is maximal for a given r, and real-world networks with the same r, γ and N_p, should have lower information $I(r)$.

We verify this conjecture using a range of real-world networks in the following section.

5.2 Shannon information of real-world networks

We computed Shannon information for a set of networks. For example, Table 1 is constructed using the cellular network data from Center for Complex Network Research, University of Notre Dame [18]. These results are augmented with information content computed for corresponding minimalistic and maximalistic networks, obtained as follows. For a metabolic network with given number of nodes N, assortativeness r, and the degree distribution parameters γ and N_p, we generate a minimalsitic class A network, using the template (10)–(12), so that it shares the parameters N, r, γ, N_p with the original metabolic network. Analogously, a corresponding maximalistic network is produced by using the template (13). For example, Figure 4 shows the metabolic network for *Escherichia coli* ($r = -0.162, I(r) = 0.49$ bits), while Figures 5 and 6 show its corresponding minimalistic ($r = -0.162, I(r) = 0.20$ bits) and maximalistic ($r = -0.162, I(r) = 0.68$ bits) networks[3]. It is evident that, although the three illustrated

---

[2] It is worth noting that a network's information content is independent of the network's size N.

[3] The networks shown in Figures 5 and 6 were constructed using a method described in the Appendix (Sect. 6). The method was used only to visualise the networks, while the information content and assortativeness were computed directly using the distribution $e_{j,k}$.
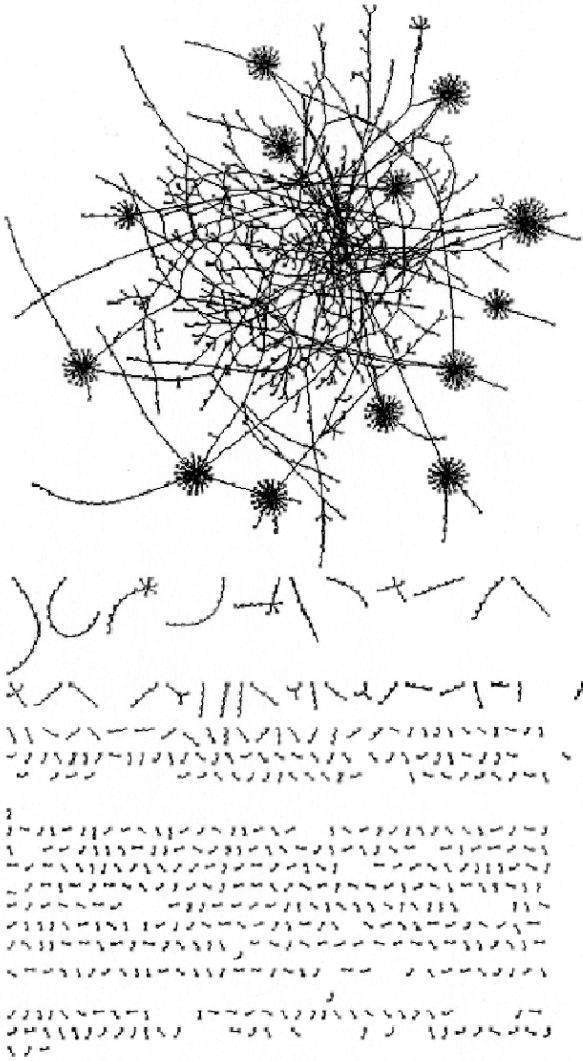
Fig. 5. Class A network corresponding to the *Escherichia coli* metabolic network: $r = -0.162$, $I(r) = 0.20$ bits; $\gamma = 2.1$, $N_p = 430$. Figure is drawn with Cytoscape 2.5.1.

networks have the same assortativeness and the scaling exponent of the power-law degree distribution, they differ topologically, and in terms of the information content. Table 1 demonstrates that information $I(r)$ is always within bounds defined by the information $I_A(r)$ for the corresponding minimalistic network and the information $I_B(r)$ for the corresponding maximalistic network. Analogously, Tables 2–6 empirically verify the respective bounds for other networks.

5.3 Power-law of information-assortativeness dependency

We studied the relationship $I(r)$ for class A networks, producing the following approximation:

$$I(r) = \begin{cases} a_1 \, r^{b_1} + c_1 \, e^{d_1 \, r} - c_1 & \text{if } r \geq 0 \\ a_2 \, |r|^{b_2} + c_2 \, e^{d_2 \, |r|} - c_2 & \text{if } r < 0 \end{cases} \quad (14)$$

Fig. 6. Class B network corresponding to the *Escherichia coli* metabolic network: $r = -0.162$, $I(r) = 0.68$ bits; $\gamma = 2.1$, $N_p = 430$. Figure is drawn with Cytoscape 2.5.1.

where $|r|$ denotes the absolute value of assortativeness r, and the coefficients a_i, b_i, c_i, d_i depend on variables γ and N_p. The critical assortativeness at which the respective $I(r)$ curve attains its minimum is denoted as $\overline{r}$. In general, $\overline{r}$ is specific for each degree distribution q_k, i.e. for each pair of γ and N_p, however for class A networks, $\overline{r} = 0$, and $I(0) = 0$ for all γ and N_p. For a symmetric distribution q_k, equation (14) reduces to

$$I(r) = a \, |r|^b + c \, e^{d \, |r|} - c. \quad (15)$$

Class B networks can be characterized in a similar way:

$$I(r) = \begin{cases} a_3 \, (r - \overline{r})^{b_3} + c_3 \, e^{d_3 \, (r - \overline{r})} + g_3 & \text{if } r \geq \overline{r} \\ a_4 \, |r - \overline{r}|^{b_4} + c_4 \, e^{d_4 \, |r - \overline{r}|} + g_4 & \text{if } r < \overline{r} \end{cases}$$

for a symmetric distribution q_k, $\overline{r} = 0$.

The main term of equations (14) and (15) is the *information* power-law $a_i \, |r|^{b_i}$ which dominates the *correction* term $c_i \, e^{d_i \, |r|}$. The *rate coefficient* $b_i(N_p, \gamma)$ is the scaling exponent of the information power-law, reflecting how the amount of Shannon information I would change with respect to a change in assortativeness r. Figure 7 shows the rate coefficient $b_1(N_p, \gamma)$ against N_p for various fixed exponents γ. This dependency can be approximated by a function which is dominated by a power law for small N_p:

$$b_1(N_p, \gamma) = \mu(\gamma) N_p^{\nu(\gamma)} + \lambda(\gamma) \quad (16)$$

where $\nu(\gamma) < 0$. For example, $b_1(N_p, 1.0) \approx 1.3 N_p^{-0.28} + 1.04$, and $b_1(N_p, 3.0) \approx 0.75 N_p^{-0.6} + 1.47$.
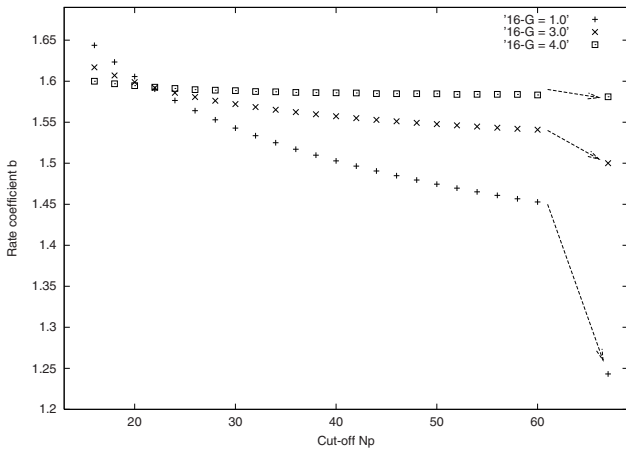
Fig. 7. The dependencies between the rate coefficient b_1 and cut-offs N_p, for different γ. The points indicated by arrows are coefficients b_1 for $N_p = 1000$.
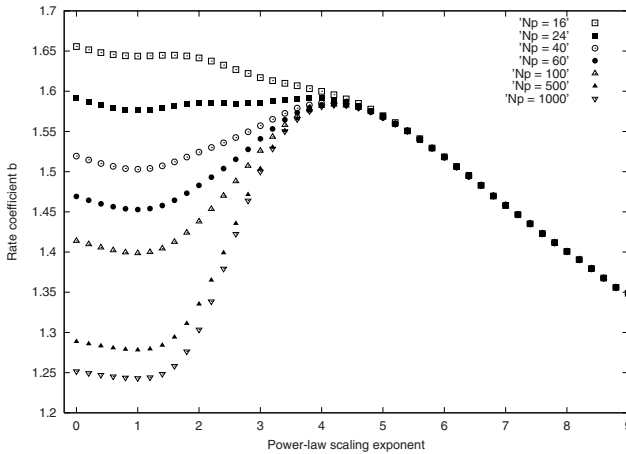

Fig. 8. The dependencies between the rate coefficient b_1 and power law exponents γ, for different N_p.

where $\nu(\gamma) < 0$. For example, $b_1(N_p, 1.0) \approx 1.3 N_p^{-0.28} + 1.04$, and $b_1(N_p, 3.0) \approx 0.75 N_p^{-0.6} + 1.47$.

Conversely, Figure 8 traces the rate coefficient $b_1(N_p, \gamma)$ against γ for various fixed cut-offs N_p. It can be observed that the rate $b_1(N_p, \gamma)$ tends to plateau when $4 < \gamma < 5$, and quickly diminishes when $\gamma > 5$. This creates a local "stability" region on the $N_p \times \gamma$ surface when $N_p > 20$ and $4 < \gamma < 5$. The stability region is also visible in Figure 9 that shows the dependency of the correction coefficient $d_1(N_p, \gamma)$ on γ for various fixed cut-offs N_p.

It can be also observed that the correction coefficient $d_1(N_p, \gamma)$ moves toward its minimum as N_p grows and the exponent reduces below $\gamma = 3.0$ (Fig. 9). This indicates that in many real-world networks that are typically characterized by larger N_p's and the range $2.0 < \gamma < 3.0$, the correction term plays a minor role.

We verified this conjecture by considering the assortativeness range $-0.4 \leq r \leq 0.6$, which corresponds to real-world networks (including social networks). In this case, it was observed that there is no need for a correction
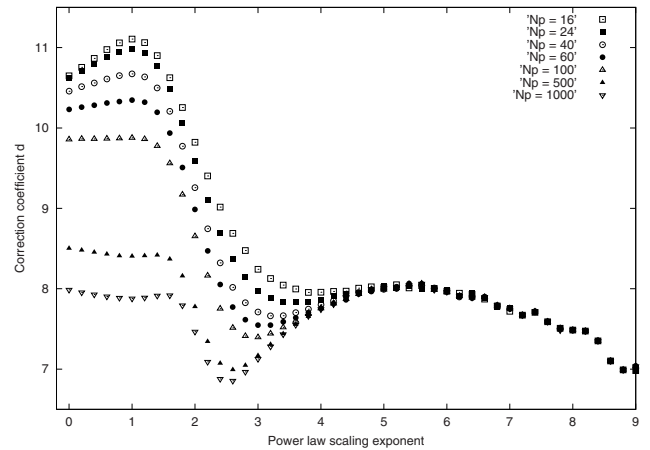

Fig. 9. The dependencies between the correction coefficient d_1 and power law exponents γ, for different N_p.

term at all, and the relationships (14) and (15) can be simplified as follows:

$$I(r) = \begin{cases} a_1 \, r^{b_1} & \text{if } r \geq 0 \\ a_2 \, |r|^{b_2} & \text{if } r < 0. \end{cases} \tag{17}$$

For a symmetric distribution q_k, equation (17) reduces to

$$I(r) = a \, |r|^b. \tag{18}$$

The resulting coefficients $b_1(N_p, \gamma)$ do not differ from the coefficients obtained by fitting the relationships (14) and (15) and shown in Figures 7 and 8 – more precisely, the difference is within 0.004 or 0.26%.

5.4 Slope and stability regions

If one attempts to optimize networks according to their information content (analogous to optimizing communication channels), assortativeness becomes the main factor: the less non-assortative is the network, the more information it can contain. The limit on maximally attainable disassortativeness r_m implies that it is easier to maximize the information content by increasing assortativeness toward $r = 1$, as $I(r = r_m) \leq I(r = 1)$.

In the following analysis (based on the rates for the lower bound, i.e. class A), we point out that new links and new node types (i.e., different degrees) in an evolving network affect assortativeness r much more than they do the degree distribution parameters N_p and γ. In other words, it is much easier to produce and explore a candidate network with a different r, rather than different N_p and/or γ.

When optimization or evolutionary processes explore the slope region, $2.0 < \gamma < 3.0$ (Fig. 8), there is more freedom to generate and evaluate candidate networks. In this region, even a small change in the scaling exponent γ brings a reward in a higher rate of information b. Thus, the information content can be changed easily by modifying only the assortativeness r, i.e. in the slope region,

the information content is most sensitive to assortativeness. Hence, the search becomes more efficient, and networks with higher information content are found more easily. This feature may explain why the exponents γ of real-world scale-free networks are mostly within $[2.0, 3.0]$ range.

The slope region is upper-bounded by the stability region, $3.8 < \gamma < 5.0$ (Fig. 8). Within the stability region, the different curves $I(r)$ tend to be close to each other when one varies N_p and γ. For example, the relationships between information and assortativeness for class-A network (analogous to the one shown in Fig. 3) for $\gamma = 4.0$ and $\gamma = 5.0$ would be very similar if $N_p > 20$. The stability region creates a further structure in the search-space defined by Shannon information. When networks evolve (or are explored in the search-space) by changing either or both the cut-off N_p and scaling exponent γ, the information content tends to stay constant in the stability region as long as the assortativeness is maintained at the same level. In other words, the informational fitness landscape of evolving networks is smoother in the stability region: the expense taken to modify N_p and/or γ is not rewarded with more freedom to produce a higher information content.

At the lower range of scaling exponent, $\gamma < 2.0$ (Fig. 8), the freedom to vary the information content is reduced as well – simply because the rate coefficient b is smaller for the high (e.g. real-world) cut-offs N_p.

6 Conclusion

We analyzed Shannon information of scale-free networks in terms of their assortativeness. Noting that the same assortativeness r could correspond to different information values I, we introduced a classification of networks according to the dependency of the distribution $e_{j,k}^{(r=r')}$ on the assortativeness r', with the intention that, within a class, the same distribution q_k and the same assortativeness r result in the same information $I(r)$. We observe that the two identified classes of networks provide lower and upper bounds, in terms of Shannon information, for the considered real world networks.

We also demonstrated that the information content of scale-free networks depends non-linearly (and asymmetrically) on the absolute value of the assortativeness. The identified dependency is symmetric when the corresponding remaining degree distribution is symmetric. We further studied class A assortative networks, and identified slope and stability regions on the $N_p \times \gamma$ surface. In the slope region, there is more freedom to generate and evaluate candidate networks since (i) the information content can be changed easily by modifying only the assortativeness r, and (ii) even a small change in the scaling exponent γ brings a reward in a higher rate of information b. This feature may explain why the exponents γ of real-world scale-free networks are within $[2.0, 3.0]$.

The optimization criteria defined according to information content of networks would allow us to advance research into networks resilience under node removal or percolation/diffusion of adverse conditions. For example, one may consider a task of *information-cloning* of a scale-free network [23], given its fragment and some topological properties of the original network. The "cloning", interpreted information-theoretically, would aim at attaining an equivalent information content of the resulting network which may disagree with the original one in terms of specific node to node connections.

Appendix A

In order to construct a network with a specific assortativeness value r, given a degree distribution p_k and a network size N, we used the *Assortative Preferential Attachment* (APA) method [23]. The remaining degree distribution q_k is obtained using equation (1).

We use the $e_{j,k}$ computed by templates (10)–(12) for *class A*, or (13) for *class B*, to grow a desired network. We start by creating a 'source pool' and 'target pool' of unconnected nodes, each of size $N_0 = N/2$, with the intention of sequentially adding the nodes from source pool to target pool. In the traditional preferential attachment method [4], the probability of a new link between a source and a target node depends only on the degree of the target node. In the APA method, however, the probability would depend on the degrees of both source and target nodes. We therefore, probabilistically assign an "intended degree" k to each node in both pools such that the resulting degree distribution is p_k.

Then we assign a probability distribution $\mu(k, j_0), \ldots, \mu(k, j_{N_p-1})$ to each target node with the degree k, where $\mu(k, j)$ is the probability of a source with degree j joining the target node with the degree k. The probability $\mu(k, j)$ is calculated as $\mu(k, j) = e_{j,k}/p_j$, then normalized such that $\sum_j \mu(k, j) = 1$. The distribution $\mu(k, j)$ has to be biased by division by p_j, because each source node with degree j does not occur in the source pool with the same probability. In other words, sequential addition would not maintain $e_{j,k}$, and the biased probability $\mu(k, j)$ accounts for that. Once $\mu(k, j)$ is generated, each source node with degree j is added to the target pool and forms a link to a target node with degree k with probability $\mu(k, j)$.

For example, if there are twice as many source nodes with degree j_2 than those with degree j_1 (i.e., $p(j_2) = 2\, p(j_1)$), while $e(k, j_2) = e(k, j_1)$, then the biased probabilities $\mu(k, j_1)$ and $\mu(k, j_2)$ would be such that $\mu(k, j_2) = e(k, j_2)/p(j_2)$ and $\mu(k, j_1) = e(k, j_1)/p(j_1) = 2\mu(k, j_2)$. This ensures that nodes with degree j_1 (represented twice as scarce as the nodes with degree j_2) would find it twice as easy to form a link with a target node with degree k.

When a target node with k degrees forms its last, k-th, link, all its probabilities $\mu(k, j)$ are set to zero (i.e., this node will not form any more links). The grown network will thus have the desired joint distribution $e_{j,k}$, and hence the desired assortativeness r'.

References

1. R.V. Solé, S. Valverde, in *Complex Networks*, edited by E. Ben-Naim, H. Frauenfelder, Z. Toroczkai (Springer, 2004), Vol. 650 of Lecture Notes in Physics
2. M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002)
3. D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E. Newman, S.H. Strogatz, Phys. Rev. E **64** (4 Pt 1) (2001)
4. T.R. Albert, A.L. Barabasi, Rev. Modern Phys. **74**, 47 (2002)
5. R. Albert, H. Jeong, A.L. Barabasi, Nature **401**, 130 (1999)
6. M. Faloutsos, P. Faloutsos, C. Faloutsos, Comp. Comm. Rev. **29**, 251 (1999)
7. M. Piraveenan, M. Prokopenko, A.Y. Zomaya, Europhys. Lett. **84**, 28002 (2008)
8. M. Prokopenko, F. Boschetti, A. Ryan, Complexity (accepted, 2008)
9. M.E.J. Newman, Phys. Rev. E **67**, 026126 (2003)
10. C.E. Shannon, The Bell System Technical J. **27**, 379 (1948)
11. D.J. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, 2003)
12. S. Kullback, R.A. Leibler, Ann. Math. Stat. **22**, 79 (1951)
13. T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000)
14. J.T. Lizier, M. Prokopenko, A.Y. Zomaya, Phys. Rev. E **77**, 026110 (2008)
15. G. Bianconi, Europhys. Lett. **81**, 28005 (2008)
16. R. Guimera, M. Sales-Pardo, L.A. Amaral, Nature Phys. **3**, 63 (2007)
17. M. Boguñá, R. Pastor-Satorras, A. Vespignani, Eur. Phys. J. B **38**, 205 (2004)
18. *Centre for Complex Network Research, University of Notre Dame* (2008), http://www.nd.edu/networks/index.htm
19. J. Baumbach, BMC Bioinformatics **8**, 429 (2007)
20. *Database of Interacting Proteins, University of California, Los Angeles* (2008), http://dip.doe-mbi.ucla.edu/dip/
21. V. Batagelj, A. Mrvar, *Pajek datasets* (2006), http://vlado.fmf.uni-lj.si/pub/networks/data/
22. M.E.J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001)
23. M. Piraveenan, M. Prokopenko, A. Zomaya, in *Advances in Artificial Life: 9th European Conference on Artificial Life (ECAL-2007), Lisbon, Portugal*, edited by F.A. e Costa, L.M. Rocha, E. Costa, A.C. I. Harvey (Springer, 2007), Vol. 4648 of Lecture Notes in Artificial Intelligence