Eur. Phys. J. B **50**, 617–630 (2006)
DOI: 10.1140/epjb/e2006-00170-5

THE EUROPEAN PHYSICAL JOURNAL B

Modeling network growth with assortative mixing*

A.P. Quayle[a], A.S. Siddiqui, and S.J.M. Jones

Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada

Received 14 June 2005 / Received in final form 21 November 2005
Published online 17 May 2006 –

Abstract. We propose a model of an underlying mechanism responsible for the formation of assortative mixing in networks between "similar" nodes or vertices based on generic vertex properties. Existing models focus on a particular type of assortative mixing, such as mixing by vertex degree, or present methods of generating a network with certain properties, rather than modeling a mechanism driving assortative mixing during network growth. The motivation is to model assortative mixing by non-topological vertex properties, and the influence of these non-topological properties on network topology. The model is studied in detail for discrete and hierarchical vertex properties, and we use simulations to study the topology of resulting networks. We show that assortative mixing by generic properties directly drives the formation of community structure beyond a threshold assortativity of $r \sim 0.5$, which in turn influences other topological properties. This direct relationship is demonstrated by introducing a new measure to characterise the correlation between assortative mixing and community structure in a network. Additionally, we introduce a novel type of assortative mixing in systems with hierarchical vertex properties, from which a hierarchical community structure is found to result.

PACS. 89.75.-k Complex systems – 89.75.Hc Networks and genealogical trees – 89.75.Fb Structures and organization in complex systems

1 Introduction

1.1 Aims and motivation

The modeling of networks and complex systems has recently developed into a highly-active area of research, with studies focusing on a range of properties of these systems. Some network properties have received particular attention, such as the degree distribution [1,2], clustering or transitivity [3,4], and the so-called "small-world effect" [5,6]. The models and theories developed can typically be applied to networks in a wide range of fields, including technological networks such as the Internet and the world-wide web [7,8], social networks of various kinds [9,10], and biological networks (e.g. genetic, metabolic and neural) [11–13].

The majority of models developed consider the individual elements which make up the system (nodes or vertices of the network) as abstract entities, and thus the models are easily extensible to many network types, as described above. However, the behaviour and dynamics of real networks is a function of both the network topology, and the attributes or properties of the elements of the system. It is well-known that network edges do not connect to vertices independent of their properties or attributes, but relatively few models explicitly take this into account. In recent years, this phenomenon has received more attention, and a measure of the tendency of vertices to be connected to others which are "similar" or like them in some way is often referred to as assortative mixing, a term originated in the ecology and epidemiology literature [14]. A closely related term is the homophily, which is often only used in the social literature, where this concept is most well recognised and understood [15]. Homophily measures the driving force behind this tendency, and assortativity measures the extent of mixing between "similar" vertices in a given network.

A specialised and well-recognised type of assortative mixing is that of mixing by vertex degree. Data on degree assortativity is readily available since no additional, non-topological data is required in order to calculate the degree assortativity of a network. It has been shown that the degree assortativity depends to a large extent on the type or classification of the network. Social networks tend to be assortative (where high degree vertices tend to connect to other high degree vertices, and vice versa), whereas biological and technological networks tend to be disassortative (where high degree vertices tend to connect to low degree vertices) [16]. The degree assortativity of a network has a number of interesting ramifications on the network robustness, which is a critical problem in many areas of research [17–19].

* Online supplementary materials are only available in electronic form at `http://www.eurphysj.org`

[a] e-mail: `aquayle@bcgsc.ca`

Many other vertex properties may also exhibit assortative mixing in a network, other than the vertex degree. For example, in social networks properties which may show assortative mixing include race, sex, language, and age [20,21]. Although less studied, assortative mixing occurs in many network types, such as by page content in the world-wide web [22], by protein function in a protein-protein interaction network [23], and by location for spatially distributed networks amongst others. Vertex properties may be either inherent and independent of network topology (e.g. world-wide web content, or social characteristics such as sex, age), while others are only meaningful in the context of the network (such as vertex degree and biological function, to a good approximation). Furthermore, other properties such as language and race although meaningful outside the context of the network, are primarily influenced by network topology, since the formation of communities has given rise to different languages and races. In the case of biological function, we anticipate that there are many inherent properties which show assortative mixing and influence network growth, but remain less clearly defined or understood than many vertex properties in social networks.

One of the fundamental premises of network research is that network topology reflects information or properties of the vertices in some way, but so far this aspect of real networks has recieved little attention. Without an explicit consideration of such vertex properties, network models cannot explain certain aspects of network topology which are due to vertex properties, nor can they explain the distribution or assortativity of these properties within a network. Recent work has showed that it is possible to derive meaningful values of vertex similarity, based on generic properties, from network topology for both simulated and real networks [24]. Our aim is to model the tendency of vertices to link to similar vertices based on their inherent non-topological properties, and study the influence of these properties on network topology. We show that this tendency or homophily leads to assortative mixing, which beyond a threshold leads to the formation of community structure, which in turn influences other topological properties. Furthermore we demonstrate the correlation between these non-topological properties and network topology in the resulting networks.

In summary, we present a model for a realistic mechanism by which assortative mixing arises during network growth. The model is presented in a generic form, such that it can potentially be applied to any type or classification of network, and for any type of assortative mixing. We also present a new type of assortative mixing for vertex properties which are represented using a hierarchical classification tree. The current model considers mixing by properties inherent to vertices, and whose values are static during network growth, which allows for a clearer characterisation of the influence of vertex properties and assortative mixing on network topology. A more complex model may allow vertex properties to change during network growth, or an alternative model could consider properties which are determined solely or primary by network topology. Note that we use the term "assortative mixing" to refer to mixing by generic, non-topological properties, as used for example by Newman [25]. Due to the prevalence of studies of degree assortativity, assortative mixing is often assumed to refer to this specific type of mixing, but we do not refer to degree assortativity when using this term, unless otherwise stated.

Due to the complexities of real networks, an application of such a generic model is a complex problem for a given real network or system of networks. Multiple properties typically influence topology in real networks, and we use the generic model to study how such properties may influence topology due to assortative mixing. In most areas of network science, a detailed understanding of which vertex properties are significant and how these may influence topology is many years away, for example in social networks or systems biology. A generic model can be used to help to develop the understanding of the dependencies between vertex properties and network topology for a given type of real network. We plan to apply the model initially to relatively simple systems in future work.

It is known that assortativity is one mechanism which can account for the formation of community structure in a network [26]. We reproduce this observation for the proposed model, and study the correlation between assortative mixing and community structure in detail. The extent of community structure in a network is measured using the algorithm of Newman and Girvan [27]. Community structure is an important aspect of network topology, and is responsible in part for the high clustering coefficients observed in many real network systems [28].

Some existing models of assortative mixing define a method to produce a network with a given assortativity, but do not model a realistic mechanism for how assortativity arises in a real network system. An advantage of such models is that they are often more analytically tractable, using the techniques of statistical physics for the analysis of network ensembles. For example, the method defined in [26] generates a network with a known mixing pattern, and for any specified degree distribution for each vertex type. If the mixing pattern is known, it may be possible to derive an exact expression for the assortativity, as has been shown for a symmetric binomial mixing pattern [16]. Kim et al. generalised a static model motivated for social networks in which a fixed number of vertices are assigned weights, which can generate networks with assortative mixing while reproducing a power-law degree distribution [29].

Other studies have defined more realistic mechanisms for the formation of assortative mixing in networks, but are specific to either a certain network type or classification (e.g. social networks), or specific to a type of assortativity (e.g. degree assortativity). Boguñá et al. presented a model of social networks using a preferential attachment rule between vertices based only on their social distance, which reproduced many characteristics of real social networks, including community structure [30]. Catanzaro et al. defined a model of assortative mixing in social networks which is closer to the model proposed in

this paper [31,32]. Their model was based on the Barabási-Albert (BA) preferential attachment rule [33], but there was no consideration of other aspects of network topology produced by the model, specifically the community structure. Yang et al. described a bipartite network model in which each type of vertex exhibits a power-law degree distribution, with tunable mixing between the two types [34].

Alternative mechanisms for the formation of assortative mixing by degree have also been developed, other than due to the homophily between vertices with similar degree. Onody and Castro studied a non-linear version of the BA preferential attachment rule, and showed that the resulting networks may be assortative or disassortative (or show no assortative mixing) depending on the power of the vertex degree in the preferential attachment rule [35]. Xulvi-Brunet and Sokolov added a rewiring step after each preferential attachment step, and were able to generate networks with any required assortativity, from no assortative mixing to totally assortative [36]. This type of rewiring was originally proposed to generate random networks with an exact degree distribution for comparison with hierarchical networks [37]. We use a similar idea in the current model when considering the assortativity of properties which are network dependent, and not inherent to a vertex.

2 Model and methods

The governing equation of the network growth model presented below is a combination of two preferential attachment rules, one of which is a function of vertex degree, and the other a function of vertex properties. For simplicity we implement the model using the standard BA preferential attachment rule for the function of vertex degree, and use the same basic implementation of network growth [33]. That is, starting with a small number m_0 of vertices, we add a vertex to the network at each time step, and add $m \leq m_0$ edges between the new vertex and existing vertices in the network. However, we define the model in terms of a generalised function of vertex degree, since it has been shown that alternative attachment rules provide more accurate models of many real networks. A non-linear preferential attachment [38], and the inclusion of vertex "attractiveness" such that new vertices may attach to isolated vertices with a non-zero probability [39] are just two examples of alternative attachment rules. Preferential attachment by vertex degree can be expressed as,

$$\Pi(k_j) = \frac{f(k_j)}{\sum_l f(k_l)}, \tag{1}$$

where Π is the probability that a new vertex i will link to an existing vertex j in a network, and k_j is the degree of vertex j, summing over all l vertices in the network. The BA preferential attachment rule given by,

$$\Pi(k_j) = \frac{k_j}{\sum_l k_l}, \tag{2}$$

is a specific case of the generalised preferential attachment by vertex degree given in equation (1). We define the preferential attachment by vertex properties in terms of the "similarity" x_{ij} of vertices i and j,

$$\Pi(x_{ij}) \propto e^{-\alpha(1-x_{ij})}, \tag{3}$$

where $\alpha > 0$ is a tunable parameter which we call the homophily, vertex similarity x_{ij} is defined within the range $0 < x_{ij} < 1$, and the probability Π is a function of vertex similarity. The precise definition of vertex similarity depends on the type of vertex property being considered, and is defined in detail in Sections 2.1 to 2.3. An exponential dependency is chosen in order to assign an appropriate weighting to the vertex similarity, and since this form of dependency has been used in a number of closely related studies, including a model of degree assortativity [31]. Watts et al. used an exponential variation to model social distance in the context of network search [40]. Bianconi and Barabási developed a fitness model with an exponential dependency on vertex fitness [41,42]. The expression used in their study is similar in form to the model defined in equation (3), but vertex fitness and similarity are very different properties, and influence network topology in correspondingly different ways. Combining equations (1) and (3) as a product of two preferential attachment rules and normalising gives,

$$\Pi(k_j, x_{ij}) = \frac{f(k_j)e^{-\alpha(1-x_{ij})}}{\sum_l f(k_l)e^{-\alpha(1-x_{il})}}, \tag{4}$$

where the probability Π is explicitly a function of two independent variables, the vertex degree k_j and vertex similarity x_{ij}. The preferential attachment rules are combined as a product such that they function independently and with equal weighting after normalising. Equation (4) is the governing equation describing network growth in the current model. More specific governing equations are given in Sections 2.1 to 2.3 for each distinct type of vertex property and assortative mixing. Note that equation (4) reduces to equation (1) if all the vertices are identical, and the similarity between all vertex pairs is one.

The model is motivated by networks where the dominating influence on the formation of assortative mixing is the vertex similarity, and provides a simple underlying mechanism capable of generating complex mixing patterns, particularly for hierarchical properties where mixing can occur across many generations of a specified hierarchy. Some networks may alternatively exhibit mixing patterns between particular vertex types (e.g. for discrete properties), where there is effectively a different homophily between specific vertex properties or types, which cannot be explained simply by the influence of vertex similarity. In this case, the governing equation can be generalised to the form,

$$\Pi(k_j, x_{ij}) = \frac{f(k_j)e^{-\alpha_{ij}(1-x_{ij})}}{\sum_l f(k_l)e^{-\alpha_{il}(1-x_{il})}}, \tag{5}$$

but this proves increasingly complex to utilise with more than a couple of vertex properties. In this case, alternative

models may be better suited to networks with certain very specific mixing patterns, such as re-wiring models which can generate an exact mixing pattern [26], but do not provide a realistic mechanism for network growth.

Although we focus on assortative mixing in this study, it is important to note that some vertex properties show disassortative mixing. For example, biological and technological networks tend to be disassortative by vertex degree [16], but networks can also be disassortative with respect to non-topological parameters.

To model network growth with a homophily between "dissimilar" vertices, we simply reverse the dependency on similarity in the governing equation, such that the governing equation for disassortative mixing is given by,

$$\Pi(k_j, x_{ij}) = \frac{f(k_j)e^{-\alpha x_{ij}}}{\sum_l f(k_l)e^{-\alpha x_{il}}}. \quad (6)$$

In Sections 2.1 to 2.3 we present specific implementations of the model for particular types of vertex properties. The behaviour of the model is investigated in detail for discrete properties and also for hierarchical properties in Sections 3.1 and 3.2 respectively. The influence of the homophily in the model can be studied effectively for discrete vertex properties where the influence of vertex similarity is relatively simple, and a model for hierarchical properties enables a more detailed investigation into the influence of vertex similarity.

2.1 Discrete properties

For a given vertex property the vertex similarity is defined between zero and one, such that there is no similarity between vertices i and j if $x_{ij} = 0$, and the vertices have the same property value if $x_{ij} = 1$. These are the only possible values of vertex similarity for discrete vertex properties, and thus in this case the model reduces in complexity to a tendency for vertices to link to identical vertices. For all other property types, the model explicitly considers a range of similarity values. For example in a social network model which accounts for the sex of individuals in the network, two individuals are either the same sex, and $x_{ij} = 1$, or different sex, and therefore $x_{ij} = 0$ with respect to this particular property. Hence the vertex similarity for discrete properties can be expressed in terms of the delta function,

$$x_{ij} = \delta_{p_i p_j}, \quad (7)$$

where p_i is the property of vertex i, and $\delta_{p_i p_j} = 1$ if $p_i = p_j$, otherwise $\delta_{p_i p_j} = 0$. The governing equation can then be rewritten as,

$$\Pi(k_j, \delta_{p_i p_j}) = \frac{f(k_j)e^{-\alpha(1-\delta_{p_i p_j})}}{\sum_l f(k_l)e^{-\alpha(1-\delta_{p_i p_l})}}. \quad (8)$$

We use the assortativity coefficient defined by Newman to measure the extent of assortative mixing in networks, which is calculated from the values of a mixing matrix [25]. A given value in a mixing matrix e_{st} is the fraction of edges in a network which link a vertex of property s with a vertex of property t, where we use the indices s and t to avoid confusion with the vertex indices. The matrix is symmetrical for undirected networks, and for any network satisfies the expressions,

$$\sum_{st} e_{st} = 1, \quad \sum_t e_{st} = a_s, \quad \sum_s e_{st} = b_t, \quad (9)$$

where the a and b values are the row and column sums respectively. The assortativity coefficient r is defined in the range $-1 < r < 1$ where $r = 0$ means no assortative mixing, and for discrete properties is given by the expression,

$$r = \frac{\sum_s e_{ss} - \sum_s a_s b_s}{1 - \sum_s a_s b_s}. \quad (10)$$

The numerator measures the difference between observed diagonal values in the matrix where edges connect vertices with the same property, and the diagonal coefficient values if there were no assortative mixing, as given by the product $a_s b_s$. A derivation of the resulting variation of the assortativity coefficient with the homophily and number of discrete properties is presented in the Appendix, and we show that,

$$r = \frac{1 - e^{-\alpha}}{1 + (n_p - 1)e^{-\alpha}}, \quad (11)$$

for assortative mixing, where n_p is the number of discrete properties. We briefly note that in some problems or applications it may be useful to determine a more localised measure of assortativity, which to our knowledge has so far not been considered. Assortativity in localised areas of a network may not be captured when averaged across an entire network. The definitions above and for other types of properties can be extended to communities or to individual vertices, where a mixing matrix is constructed for each vertex and considers only those edges incident on a given vertex. This also leads to the concept of an assortativity distribution, and we hope to investigate these ideas in future work.

2.2 Hierarchical properties

A natural extension of a classification by discrete properties is a hierarchical classification scheme. Such a classification is rarely used in network models, but we argue that this is a more appropriate and accurate method for modeling many vertex properties. A similar classification model was developed by Watts et al. to model search in social networks [40]. For example, social properties such as race and language may be more effectively modeled in this way. Languages are often organised into groups of related languages such as the Indo-European group, which can be subdivided into groups such as Celtic, Germanic, Slavic etc., and similarly with race. An example of a hierarchical classification is shown in Figure 1, which shows a small subset of the Google internet directory.

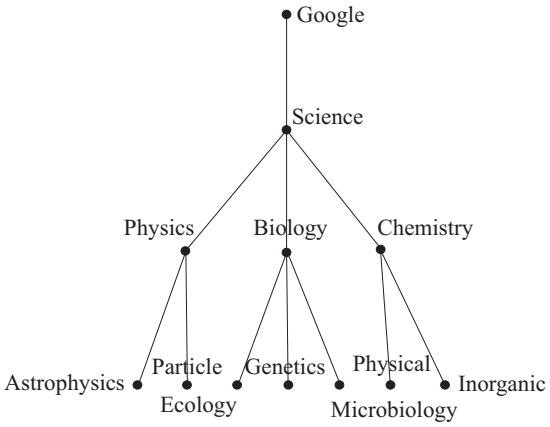

Fig. 1. Small subset of the Google internet directory.

This hierarchical classification is used to describe webpage content, but a similar classification could also be used to describe scientific specializations in a collaboration or citation network. Businesses typically operate using a hierarchical classification of employees, and also large corporations are often subdivided into smaller company units. A hierarchical classification can be reduced to a discrete classification at a particular generation of the hierarchy, which can be useful for certain types of analysis. However, a lot of detail may be lost using such an approach, and we postulate that assortative mixing extends across many generations of such classifications. For example, a webpage describing inorganic chemistry will tend to link to other webpages about inorganic chemistry, but may also link to pages describing physical chemistry with a greater probability than pages describing genetics. We allow a given vertex of a network (e.g. a webpage) to be associated with multiple classifications in the hierarchy, for example a webpage may describe both physical and inorganic chemistry.

We define the vertex similarity for hierarchical properties by the expression,

$$x_{ij} = 1 - \frac{g_{ij}}{G}, \tag{12}$$

where $g_{ij} = \max(d_i, d_j)$ is the number of generations back to the lowest common ancestor of vertex i and vertex j in the hierarchy, and d_i, d_j are the distances to the lowest common ancestor of i and j. G is the total number of generations in the tree, ignoring the root which is defined as generation zero. If the distances d_i and d_j are different, one of the vertices is defined at a lower generation than the other (a more detailed definition), in which case the maximum distance to the lowest common ancestor should be used.

Using this definition of vertex similarity, the governing equation becomes,

$$\Pi(k_j, g_{ij}) = \frac{f(k_j)e^{-\alpha g_{ij}/G}}{\sum_l f(k_l)e^{-\alpha g_{il}/G}}. \tag{13}$$

We define a new measure of assortativity for hierarchical vertex properties, such that the assortativity coefficient is given by the expression,

$$r = \frac{1}{G}\sum_{i=1}^{G} r_{g_i} \tag{14}$$

where r_{g_i} is the assortativity coefficient for discrete properties at the ith generation of the hierarchy and r_{g_0} is the root node. This is an average of the assortativity coefficient in each generation of the hierarchy and provides a measure of assortativity across all generations. If there is no additional mixing between properties in generations higher than the leaf classifications, this expression reduces to the assortativity coefficient for discrete properties at the leaf or lowest generation of the hierarchy.

2.3 Other properties

A network may also show assortative mixing by other types of vertex properties, notably including scalar or vector properties. For example, age or income in a social network, or hydrophobicity or pKa in biological networks are all scalar properties. The spatial distribution of vertices in a network is an example of a vector property, for example, individuals are more likely to know other individuals who live or work in the same or nearby location in a social network. Scalar properties may also be either discrete or continuous, depending on the property or the accuracy required. The most well studied type of assortativity is based on vertex degree, which is a discrete scalar property. Vertex degree is neither static nor inherent to vertices, however, and therefore other models are more appropriate for assortative mixing by degree.

There are a number of possible alternative definitions of vertex similarity based on scalar or vector properties, which may be suitable depending on the specific system being modeled. We suggest a simple form for the vertex similarity given by,

$$x_{ij} = \frac{1}{1 + |p_i - p_j|}, \tag{15}$$

where p_i and p_j are the properties of vertices i and j as defined previously (scalar or vector). A similar measure to the homophily between scalar properties is an affinity measure assigned to vertices in a paper by Gómez-Gardeñes [43]. In their case, a preferential attachment rule is restricted to some tunable "neighbourhood" of affinity values for a given vertex addition, which generates increased clustering. This approach is relatively coarse-grained and is highly restrictive for vertices not contained in a given neighbourhood. Such a model may be appropriate in some specialised applications, but we suggest that generally a more detailed account of vertex properties is required for an improved model.

With the dependency described in equation (15), the governing equation for assortative mixing is given by,

$$\Pi(k_j, p_i, p_j) = \frac{f(k_j)e^{-\alpha|p_i - p_j|/(1+|p_i - p_j|)}}{\sum_l f(k_l)e^{-\alpha|p_i - p_l|/(1+|p_i - p_l|)}}. \tag{16}$$

Assortative mixing by scalar properties is calculated in terms of a mixing matrix as for discrete properties described in Section 2.1, which must also satisfy equation (9). The assortativity coefficient is given by,

$$r = \frac{\sum_{st} st(e_{st} - a_s b_t)}{\sigma_a \sigma_b}, \tag{17}$$

where σ_a and σ_b are the standard deviations of the distributions of the row and column sums, a_s and b_t, and a similar expression can be used to calculate the coefficient for vector properties.

2.4 Community structure

A focus of the current work is on the topology of the networks produced by the model defined above, and in particular the community structure. A number of algorithms have been proposed in the literature for detecting community structure in a network [9,44]. We choose to use the method of Girvan and Newman [45], since it has been shown to perform favorably compared to alternative methods, and since a closely related method is able to measure the extent of community structure in a network, which is effectively a measure of the accuracy of a given division into communities [27]. A faster algorithm has also been developed by Newman which is useful for detecting community structure in very large networks [46].

The algorithm of Girvan and Newman removes edges successively with the highest betweenness, and recalculates betweenness values after each edge removal. This recalculation stage is crucial to the success of the algorithm in correctly detecting community structure. Betweenness measures the number of shortest paths which pass through a given edge or vertex, and edge betweenness can be expressed formally as,

$$C_B(e) = \sum_{i \neq j} \frac{\sigma_{ij}(e)}{\sigma_{ij}}, \tag{18}$$

where $\sigma_{ij}(e)$ is the number of shortest paths between vertices i and j passing along an edge e, and σ_{ij} is the total number of shortest paths between vertices i and j. Edge betweenness is an effective property to use for detecting community structure since those edges which connect vertices between communities will on average have a higher betweenness than edges connecting vertices in the same community.

The output of the algorithm is an edge removal sequence, which is commonly used to construct a dendrogram to represent a hierarchy of possible community divisions. In order to determine the highest quality community division specified by the resulting dendrogram, a measure called the modularity is introduced which represents the extent of community structure [27]. The modularity is calculated from the values of a mixing matrix using a similar principle to the calculation of assortativity coefficients described above. An element of this "community" mixing matrix e_{uv} represents the fraction of edges in the original network that link vertices in community u with vertices in community v in the current community division (as opposed to properties s and t in the case of the mixing matrix for assortative mixing). A high quality community structure division gives relatively large fractions for the values e_{uu} along the diagonal of the community mixing matrix. The modularity, Q, is given by,

$$Q = \sum_u \left(e_{uu} - a_u^2\right) \tag{19}$$

where a_u are the values of the row sums of the community mixing matrix. We use this measure extensively in simulations, in order to measure the extent of community structure as the homophily in the model is varied. For a given network, the modularity can be calculated for each community division specified in the resulting dendrogram, and the maximum modularity gives the highest quality community division obtained.

3 Results and discussion

3.1 Discrete properties

The behaviour of the model described in Section 2 was investigated using numerical simulations over a range of homophily values. To test the model for discrete properties, vertices were randomly assigned to one of 10 discrete properties or types, and simulations were run to generate networks with 200, 500 or 1000 vertices. Note that if the vertices are then grouped according to their property, this method produces approximately equal group sizes, with a bell-curve distribution (Binomial, approximated by Poisson). It is simple to run simulations of the model with any arbitrary group size distribution, by defining the required property distribution, and assigning properties to vertices using sampling without replacement.

The BA preferential attachment rule was used with $m = m_0 = 2$ which gives (undirected) networks with an average degree $k \sim 4$, where m is the number of edges added for each new vertex (see Section 2). Simulations were run for every integer homophily value from 0 to 20, which is sufficiently large such that all parameters of interest have reached asymptotic values. Figure 2 shows the variation of the assortativity coefficient and modularity with the homophily. Unless otherwise stated we ignore all community divisions other than the highest quality division according to the modularity value, for a given network and corresponding dendrogram. Hence we use the term modularity to refer to the modularity of the highest quality division. Every data point is calculated from an average of 100 network simulations in order to calculate the average behaviour of the network "ensemble" with the specified parameters, and error bars are plotted to illustrate the variation within the ensemble.

Figure 2 shows that the assortativity varies from approximately zero when the homophily is zero (no tendency for vertices to connect to similar vertices) to close to maximum assortativity at large homophily. The increase is also

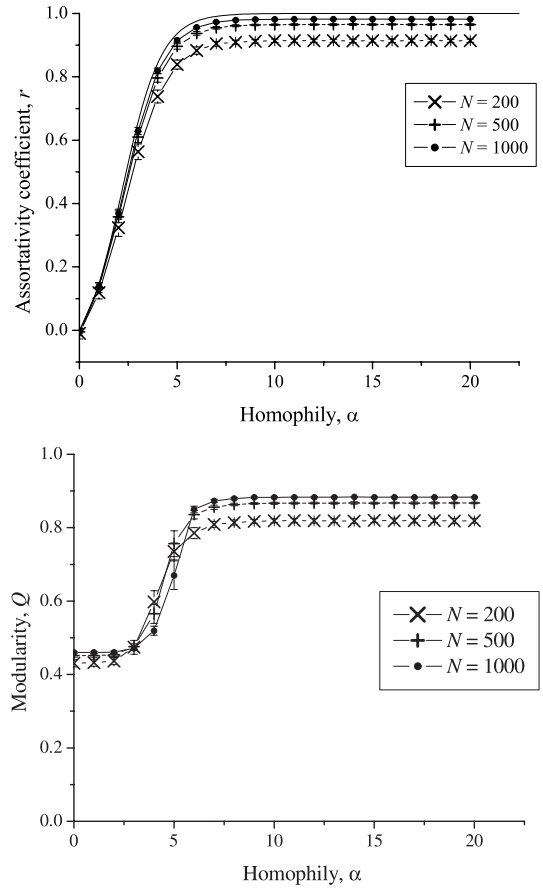

Fig. 2. Variation of the assortativity coefficient, r, (upper) and the modularity, Q, (lower) with the homophily, α, for varying N.

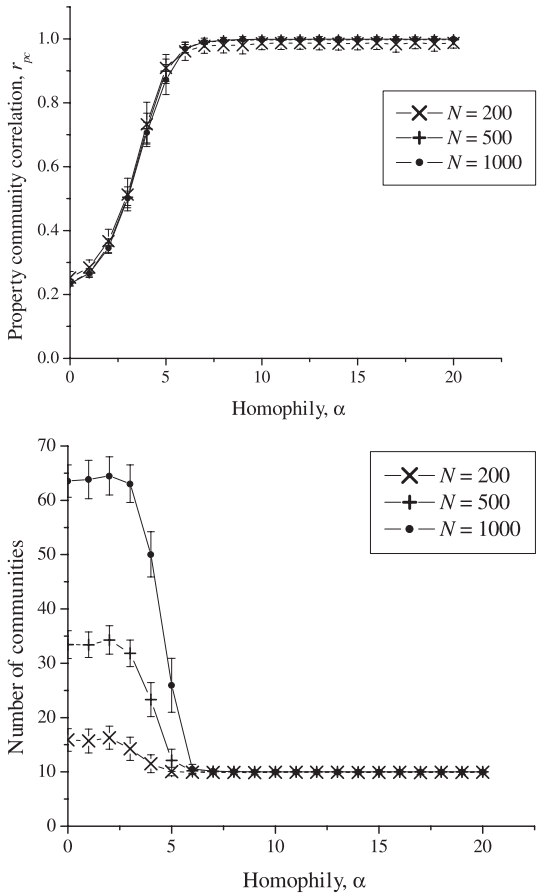

Fig. 3. Variation of the property community correlation, r_{pc}, (upper) and the number of communities (lower), with the homophily, α, for varying N.

rapid and immediate as the homophily is increased, and maximum assortativity is reached at a homophily value of around 7 or 8. Also plotted is the analytical function for the variation of the expected assortativity coefficient with the homophily given by equation (11), for $N \to \infty$, finite k and $n_p = 10$. The difference between the simulated results and the predicted variation for small N reflects the biases discussed in the Appendix.

The lower plot of Figure 2 shows that the modularity or extent of community structure in the networks also increases with the homophily, but the increase is less immediate than the increase in the assortativity. The two plots show that the formation of assortative mixing in the resulting networks drives the formation of community structure, and that a certain threshold assortativity is required before any significant community structure appears. The plots indicate that this threshold occurs at approximately $r \sim 0.5$. Recall that the method of Girvan and Newman for detecting community structure will always produce a hierarchical set of community divisions, since the output of the algorithm can be represented as a dendrogram. Modularity values are typically between 0.3 and 0.7 [27], which explains why networks grown with zero homophily have a non-zero modularity. The network size has only a minimal influence on the observed parameter variation with homophily. The asymptotic values of the assortativity and modularity are slightly less for smaller networks, since a greater fraction of edges connect between communities (and different properties) in smaller networks than larger networks.

These plots clearly show that the model is able to generate networks with both assortative mixing and community structure. Additionally, the plots indicate that the increased assortative mixing is directly responsible for the increased community structure in the networks. This is confirmed more clearly by the upper plot of Figure 3, which plots the correlation between vertex properties and community structure for the same networks used in Figure 2.

We define the property community correlation, r_{pc}, by the expression,

$$r_{pc} = \frac{\sum_c N_c}{N}, \tag{20}$$

where N is the network size or number of vertices, and N_c is the largest number of vertices with the same discrete property in a given community, summing over all communities. If the vertices in each community have the same property or type, then $r_{pc} = 1$, and if there is no linkage between vertex properties and community structure, the correlation takes some non-zero value, depending on the

number of defined properties. For example, if a model was considering the sex of individuals in a social network, the correlation would be greater than 0.5 even with no linkage. The plot shows that the correlation increases with the homophily, and that the asymptotic value is very close to one. This confirms that assortative mixing is directly responsible for the formation of community structure in the model, since otherwise the probability of the correlation being so close to the maximum value is extremely small. Also the increase in the correlation is not as immediate as that of assortative mixing, but more immediate than the formation of community structure, which is expected if the formation of assortative mixing drives the formation of community structure in the model. The observed correlation between vertex properties and community structure in our networks reflects the same type of correlation found in a wide range of real networks between vertex properties and network topology, as also shown by Leicht et al. [24].

The lower plot of Figure 3 provides an alternative visualisation of this correlation, since the average number of communities tends to an asymptotic value of 10 around a homophily of 7 or 8. This is the same as the number of discrete properties, confirming the observed correlation between property and community. The fact that larger networks are divided into more communities when there is no homophily is an artefact of the community structure algorithm. In this regime there is no significant community structure, and random variation in topology tends to result in larger networks being divided into more communities.

Note that assortative mixing is just one mechanism which can produce community structure, and we do not suggest this is the only cause of community structure formation in real networks. For example, many real networks may be best modeled with relatively low homophily, in the regime where there is insufficient assortativity to drive the formation of significant community structure. To understand the complete set of forces which influences the formation of community structure in networks requires more work in a number of directions.

The formation of community structure in networks has a direct influence on a number of other topological parameters, such as the clustering coefficient and the average shortest path. The clustering coefficient measures the fraction of transitive triples or triangles between nearest neighbours, and the clustering coefficient of vertex i, C_i, is given by,

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \tag{21}$$

where k_i is the degree or number of nearest neighbours of vertex i, and E_i is the number of edges connecting between these nearest neighbours [5,9]. The clustering coefficient of a network is calculated by averaging C_i over all vertices of the network. In a network with significant community structure, there are relatively few edges connecting between communities, and hence the average shortest path is greater than in a similar network with less community structure. The clustering coefficient is also greater in networks with more community structure for similar reasons.

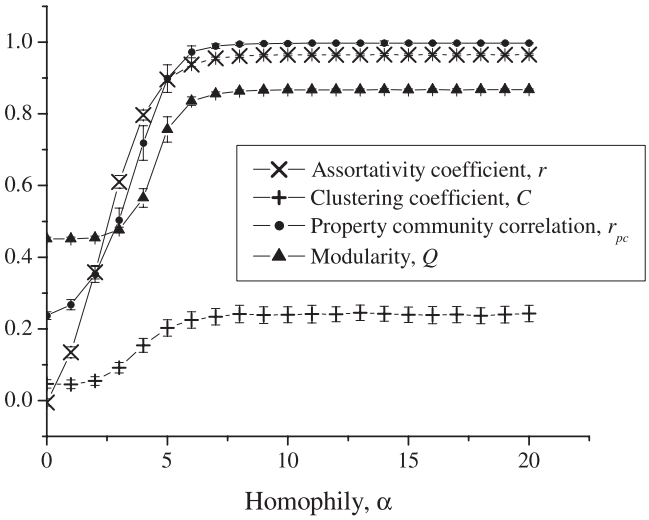

Fig. 4. Variation of related network properties with the homophily, α, for networks with $N = 500$.

Such dependencies are observed if the clustering coefficient and average shortest path are plotted against the homophily, as shown in the Online Supplementary Materials. These parameters show a similar dependency on the homophily to the community structure, and increase at the same threshold assortative mixing.

In the current model, the community structure can be thought of as the primary parameter influenced by assortative mixing, and other topological parameters as secondary parameters, which are influenced by the formation of community structure, rather than the formation of assortative mixing. This notion is supported by the significant variation found in the clustering coefficient and average shortest path in a given network ensemble, where networks contain the same modularity or extent of community structure.

Depending on the average degree and number of communities, the clustering coefficient of networks produced by the model may be relatively large ($C \sim 0.40$ is found with $N = 200$), and of the order of the high clustering values observed in many real networks [28]. We do not expect the formation of assortative mixing or community structure to be the primary influence on clustering in real networks, but we propose that these are contributing factors which often have some influence on the observed clustering. Other mechanisms have been suggested and studied which are likely to have more influence on clustering in most networks, including triad formation and vertex ageing [47,48]. Notice also that at high homophily, the average shortest path of the networks is still small relative to the network size, and the networks can still be considered to be "small-world" networks [5].

Figure 4 provides an additional illustration of the relationship between the parameters discussed, which plots the variation of selected parameters with the homophily for networks with $N = 500$.

The interdependence of the parameters discussed can be seen in Figure 4 by the homophily value at which each

of the respective parameters starts to increase. The assortativity increases immediately, which drives the correlation between property and community, which in turn drives the formation of community structure and related topological properties. Equivalent plots for networks with $N = 200$ and $N = 1000$ show the same qualitative parameter independence and variation in profile.

We have studied the variation of network parameters with the homophily while varying the network size, but it is also instructive to vary the network average degree, which is another primary parameter of the model. Equivalent plots to Figures 2 to 3 for varying average degree are presented in the Online Supplementary Materials section of this paper. The variation of the assortativity coefficient with average degree for a given homophily is minimal, and tends to the predicted behaviour as k decreases for fixed N, i.e. as $N/k \to \infty$. However, the average degree has a strong influence in the ability of the community structure algorithm to detect high quality divisions, such that at low homophily, the average modularity drops rapidly as k increases. This is intuitive, since if a network is highly connected, random variation in topology is less likely to produce community-like structures, and hence the quality of the divisions which are always produced by the algorithm is lower.

The average degree also influences the clustering coefficient and average shortest path of the networks produced by the model, where the clustering coefficient increases as k increases, and the average shortest path decreases as k increases. These variations are both intuitive, and the relationship between these parameters has already been studied in detail in a number of established network models [28,49].

Another important topological parameter is the degree distribution, which has been a central component of much network research since Barabási and Albert developed the so-called "scale-free model" [33], which is capable of reproducing the power-law degree distribution observed in many real networks [8,12]. Figure 5 shows the degree distribution for networks with $N = 1000$ produced from the current model, averaged over 50 networks for each homophily, using a log-log scale. A power-law degree distribution is given by,

$$P(k) \propto k^{-\gamma}, \tag{22}$$

and for the scale-free model, the exponent $\gamma = 3$. Hence the gradient of a plot of $\log(P_k)$ against $\log(k)$ for the scale-free model has a slope of –3.

Figure 5 shows that the degree distribution of networks with homophily up to $\alpha = 6$ is a power-law, with the exponent $\gamma = 3$ as predicted by the scale-free model. Recall that since we use a BA preferential attachment rule in simulations, in this case the model reduces to the scale-free model at $\alpha = 0$. As the homophily is increased beyond $\alpha = 6$ the distribution deviates from an exact power-law, and it can be shown that this deviation starts at the same homophily value independent of network size. The extent of deviation is smaller if the network size is large in comparison to the number of discrete properties. Since there

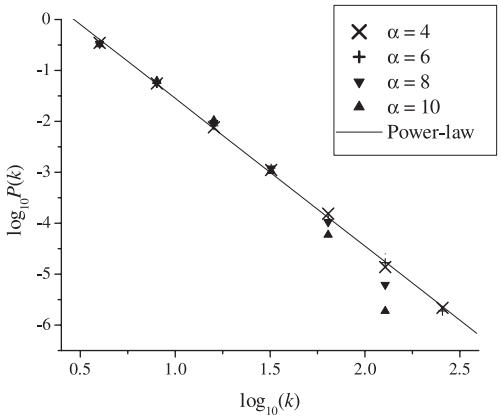

Fig. 5. Degree distribution for networks with $N = 1000$, and varying homophily, α, in the model.

is some deviation in the degree distribution at large homophily, we used the method of Maslov et al. [18] to give randomised versions of the generated networks in order to check that the change in degree distribution alone cannot explain the other changes in network topology discussed above. Their method works to maintain the exact degree distribution of the original network after randomisation. The randomised networks do not show increased assortativity and community structure as homophily is increased (see Online Supplementary Materials), proving that the degree distribution is not responsible for the observed changes in assortativity and topology. Rather, the degree distribution deviates from a power-law beyond $\alpha = 6$ due to other topological changes, and primarily the formation of community structure. The same result was found when this randomisation method was applied to networks generated with hierarchical properties (see Sect. 3.2).

The observed deviation from a power-law distribution at large homophily may model a constraint present in real networks. It has been shown that a model which incorporates the aging of vertices can produce highly clustered networks, but if the speed of aging is increased, the degree distribution of the networks also deviates from the power-law distribution observed with slower aging [6]. Many networks are known to approximate power-law distributions, but which exhibit some deviation in the tail of the distribution [6].

We focus on assortative mixing but it is interesting to test the behaviour of the model for properties which tend to connect to "dissimilar" vertices and show some degree of disassortative mixing. In this case, the parameter which has the strongest influence on the resulting behaviour is the number of discrete properties or vertex types. Results for disassortative mixing for a varying number of discrete properties are shown in the Online Supplementary Materials section. The assortativity coefficient follows the behaviour predicted by equation (31), where the extent of disassortative mixing increases with increasing homophily, but decreases rapidly as the number of properties increases.

More significantly, disassortative mixing does not result in any increase in the extent of community

structure unlike assortative mixing, and similarly has minimal influence on other topological parameters such as the clustering coefficient and average shortest path. Disassortative mixing has no tendency to produce groups of similar vertices, since vertices cannot group according to "dissimilarity" or the difference in properties. In fact, disassortative mixing has a small tendency to decrease any clustering that is present due to random variation when the homophily is zero, since this type of mixing tends to separate rather than group similar vertices. These results highlight a very important difference between assortative and disassortative mixing and their influence on network topology.

3.2 Hierarchical properties

The behaviour of the model was investigated for hierarchical vertex properties using simulations to generate networks of 200, 500 or 1000 vertices, with an average degree $k = 4$ and using the BA preferential attachment rule as in Section 3.1. A regular hierarchical classification with a branching ratio of 2 and five generations ($G = 5$) was chosen, and vertices were randomly assigned to one of the 32 "leaves" of this hierarchy. Note that it is simple to apply the model to non-regular hierarchies, and to assign network vertices to higher generations and multiple positions in the hierarchy. Figure 6 shows the variation of the assortativity coefficient and modularity with the homophily in the resulting networks.

The upper plot shows the variation of two types of assortativity coefficient, for both the coefficient for hierarchical properties defined in equation (14) and for an equivalent coefficient for discrete properties where each leaf of the hierarchy is considered to be a discrete property. The difference between these coefficient values for the same network size and homophily illustrates the additional assortative mixing present in higher generations of the hierarchy. If there was no additional mixing in the higher generations, the difference between these coefficient values would be minimal (and tends to zero as $N \to \infty$, as discussed). However, the difference between these values is significant in most regimes of the homophily, and demonstrates the importance of using a hierarchical model to capture the true extent of assortative mixing in such networks.

The assortativity increases with the homophily and network size as for discrete properties, but the increase is less rapid and occurs over a wider range of homophily. This reflects the influence of non-integer vertex similarity values (between zero and one). As the homophily increases, the extent of additional mixing in higher generations of the hierarchy decreases, and the behaviour of the model becomes more similar to that for discrete properties. The lower plot of Figure 6 confirms that assortative mixing between hierarchical properties in the model also drives the formation of community structure, and indicates that a similar threshold assortative mixing of approximately $r \sim 0.5$ is required for the formation of community structure. Interestingly, the threshold values for assortative

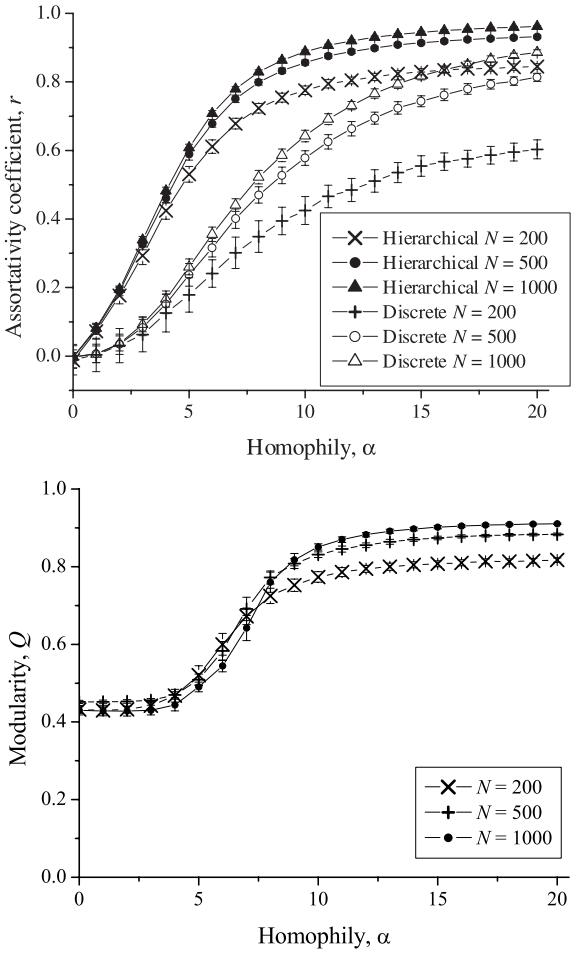

Fig. 6. Variation of the assortativity coefficient, r, (upper) and the modularity, Q, (lower) with the homophily, α, for varying N.

mixing between discrete and hierarchical properties are the same only if we account for mixing across all generations in the hierarchy. If we only calculate the assortative mixing at the lowest generation (the discrete measure in Fig. 6), the threshold value occurs around $0.15 < r < 0.20$. We speculate that the observed threshold may be universal and correspond to a phase transition, but more work is required to determine the exact nature of the observed threshold.

Other topological parameters such as the clustering coefficient and average shortest path show a similarly less rapid increase with homophily when compared with the same variation observed for discrete properties (see Online Supplementary Materials). The degree distribution of networks grown with assortative mixing of hierarchical properties also shows a very similar variation to that observed for discrete properties, where the distribution follows a power-law, also with a deviation in the tail of the distribution at high homophily.

It is also interesting to observe the variation of the property community correlation for hierarchical properties, where we apply the definition for discrete properties to each generation of a hierarchical property. In this case

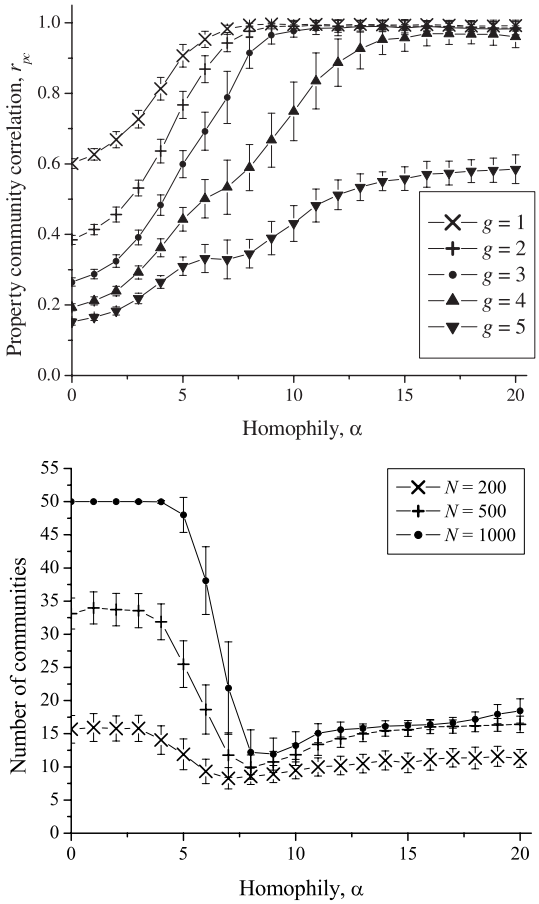

Fig. 7. Variation of the property community correlation, r_{pc}, for $N = 500$ (upper) and the number of communities (lower), with the homophily, α, for varying N.

the quantity N_c in equation (20) is the largest number of vertices with the same property for the specified generation of the hierarchy in a given community. For a given generation, the number of unique classifications or nodes is equal to the number of "discrete" properties. For example, for the hierarchical subset shown in Figure 1, r_{pc} for $g = 2$ is the r_{pc} value for the "discrete" properties Physics, Biology and Chemistry. Figure 7 shows the variation of the property community correlation for each generation of the current hierarchy with 5 generations for $N = 500$, together with the variation of the number of communities for varying network size.

Figure 7 shows that the property community correlation generally increases with homophily for each generation as expected, where $g = 5$ refers to the "leaf" generation, and $g = 1$ is the highest generation, ignoring the root node. The correlation for higher generations is generally greater, which reflects the increased assortative mixing in higher generations, as observed in Figure 6. Note also the correlation values at zero homophily are lower for lower generations in the hierarchy (larger g), since there are more "discrete" properties at lower generations, and so the correlation due to random mixing is reduced.

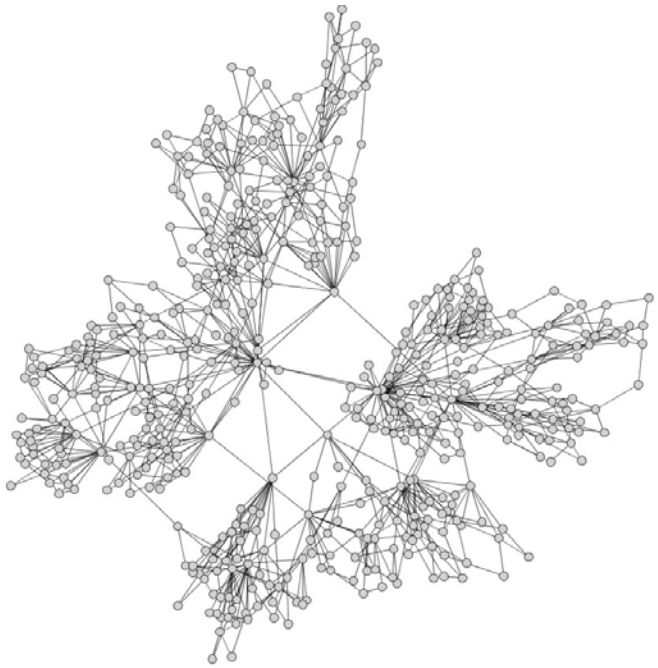

Fig. 8. Visualisation of an example network generated with hierarchical vertex properties, where $\alpha = 10$ and $N = 500$.

The correlation values are calculated for the optimum community structure based on the modularity value, and the lower plot of Figure 7 shows that the optimum community structure is found to have around 10–15 communities at high homophily. Therefore this "optimum community structure" does not reflect any real community structure due to assortative mixing at the lowest level of the hierarchy, where there are 32 different classifications. This explains why the value of the property community correlation for $g = 5$ at large homophily is not close to 1.0, unlike the values for $1 < g < 4$ in this case.

Since the community structure is driven by assortative mixing of hierarchical properties, it seems intuitive that the resulting community structure may also exhibit some kind of hierarchical structure. This type of community structure is indeed observed as the homophily increases, which can be shown by viewing the resulting networks using visualisation software. Figure 8 shows an example network with $\alpha = 10$ and $N = 500$, produced using the freely available Cytoscape software package [50].

The network visualisation indicates a hierarchical community structure, where a division into two communities is clear, which can be further subdivided into smaller communities, and so on. Naturally the divisions become less clearly defined the more the communities are sub-divided. Images in the Online Supplementary Materials section clarify the hierarchical community structure, where colors are used to depict vertex properties at each generation in hierarchy. The resulting images illustrate the topological community divisions, and show a strong correlation between vertex properties and community structure at equivalent hierarchical generations.

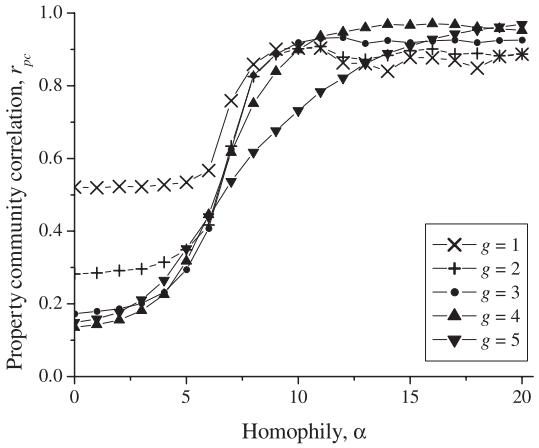

Fig. 9. Variation of the "modified" property community correlation, r_{pc}, with the homophily, α, for $N = 500$. Error bars are omitted for clarity.

Some measures of hierarchy have been proposed in the literature [51,37], but these measure hierarchy based on individual vertices rather than communities. We hope to develop a suitable measure of the extent of hierarchical community structure in future work, and encourage other research towards this goal. Many real networks are known to have a hierarchical community structure such as biological and communication networks [52,53], and a suitable measure for analysing hierarchical community structure would prove very useful for the analysis of these networks.

Since assortative mixing by hierarchical properties has been shown to produce hierarchical community structure, the "optimum" community structure as measured by the modularity is perhaps not the best community division to consider. A hierarchical community structure can be divided into communities at a number of possible levels, and the correlation between vertex properties and a hierarchical community structure can alternatively be based on the community structure corresponding to a specified generation of the hierarchy. For example in the current simulations, r_{pc} for $g = 1$ can be calculated based on the division of a resulting network into two communities. Similarly, r_{pc} for $g = 5$ can be calculated based on the division of a network into 32 communities. Figure 9 shows the variation of the property community correlation with the homophily when calculated using this alternative approach.

Figure 9 shows that the r_{pc} value for $g = 5$ using this modified definition has an asymptotic value close to the maximum, confirming that there is both community structure and property community correlation at the lowest generation in the hierarchy. The disadvantage of this definition of r_{pc} is that there is a large variation in the values in the ensemble at large homophily, since the nature of the community structure algorithm means that the exact community divisions may not always reflect the property correlations producing the community structure. We propose that the two alternative definitions of r_{pc} presented for hierarchical properties are both useful, depending on the particular problem of interest.

Disassortive mixing by hierarchical properties produces similar behaviour to that observed for discrete properties, where the extent of disassortative mixing increases with homophily, but the modularity is unchanged, and there is no measureable community structure. No type of disassortative mixing can drive the formation of community structure, in clear contrast to assortative mixing.

We have chosen to study the behaviour of the proposed model in detail for ensembles of model networks, where vertex properties are generic and are defined only by their type (e.g. discrete or hierarchical). In this way we can study the inter-relationship of model parameters, and the range of network topologies generated by the model. We emphasise that the model produces ensembles of networks with a range of topological parameters for a specified combination of input parameters, as shown by the error bars illustrated in most figures in this paper. Therefore, when the model is applied to real networks, we can expect to reproduce large-scale parameters such as the assortativity coefficient or modularity by tuning the homophily, but we would not expect to reproduce the exact mixing pattern or community structure, due to the random nature of the growth processes in the model.

As suggested by other authors [26], we expect that assortative mixing is one factor which drives the formation of community structure in real networks, but we would not expect this to be the sole driving force. For example, if we tune the homophily to reproduce the assortativity coefficient for a given vertex property in a network, we may observe a greater extent of community structure in the real network than in the modeled network. This may be due to both multiple assortativity types driving the community structure, or other forces which are currently unaccounted for. More research on problems relating to community structure in a wide range of network systems is needed before we understand the complete set of factors which can influence the formation of community structure. We will present a more detailed discussion of some of these issues, and applications of the model to systems of real networks in future work. We also hope to determine whether the observed assortative mixing threshold is consistent to all network types, and whether this corresponds to a phase transition point.

4 Summary

In summary, we have modeled a single mechanism which drives the formation of assortative mixing in networks based on generic, non-topological vertex properties, and to a varying extent based on the tunable parameter in the model, called the homophily. We have shown that the formation of assortative mixing beyond a threshold $r \sim 0.5$ directly drives the formation of community structure, which in turn influences a number of related topological parameters. The variation of network topology with homophily has been studied in detail, including the correlation between assortative mixing and community structure, and we have confirmed the direct relationship between these parameters in the model. We have studied a new type of assortative mixing by hierarchical properties, and introduced a new definition of the assortativity

coefficient to characterise this type of mixing. Additionally, we have shown that the type of community structure formed depends directly on the type of assortative mixing, whereby a hierarchical community structure results from assortative mixing by hierarchical vertex properties.

S.J. is a Michael Smith Foundation for Health Resarch Scholar. We thank Genome Canada and the British Columbia Cancer Foundation for funding, and Cenk Sahinalp and Misha Blenky for critical discussion.

Appendix: Assortative mixing by discrete properties

An analytical expression can be derived for the expected variation of the assortativity coefficient with the homophily in the limit as $N \to \infty$ for finite k, for discrete vertex properties using equations (8) to (10). Since the governing equation of the model describes the probability of edge formation between vertices, the exact network and corresponding parameter values produced by a given simulation cannot be determined. The set of possible networks grown by the model for a specified homophily, network size and average degree defines a network ensemble. Within this ensemble, some networks may be generated with a much greater probability than others, and we can predict parameter values of a typical or average network in the ensemble, or the expected parameter values.

The governing equation is defined in terms of a product of two distinct functions of vertex degree and vertex similarity with independent parameters, and hence the influence of these functions on the resulting system behaviour is independent of each other when averaged across the resulting network ensemble. This is true unless the vertex property being considered is actually the vertex degree, in which case the derivation presented below cannot be applied.

Therefore we can derive an expression for the variation of the ensemble assortativity coefficient with respect to the homophily, which applies for any function of vertex degree, $f(k)$. For finite networks, the resulting network topology is biased slightly from the probabilities defined in the governing equation, since in general there must always be a certain number of edges connecting vertices of different properties and between different communities in a network. However, if we let $N \to \infty$ as k is kept finite, these biases tend to zero, and an analytical derivation is possible in this regime.

The expected assortativity coefficient can be derived by considering the average or expected values of the mixing matrix e_{st} for a given homophily, which we call the ensemble mixing matrix. The exponential term in equation (8) can only take two possible values, either 1 if $p_i = p_j$ or $e^{-\alpha}$ otherwise. These are the two probability values influencing edge formation between vertices in the growing network. Therefore, the ratio of the number of edges between vertices with the same property, over the number of edges between vertices with different properties should be equal to the ratio of the respective probabilities, averaged across the network ensemble. If we let e_{st} refer to the values of the ensemble mixing matrix, we can write,

$$\frac{e_{st,\forall s\neq t}}{e_{st,\forall s=t}} = e^{-\alpha}, \tag{23}$$

where the notation $e_{st,\forall s\neq t}$ means those values e_{st} where $s \neq t$. If n_p is the number of discrete properties, the mixing matrix dimensions are n_p x n_p, and by symmetry the row and column sums of the ensemble mixing matrix are $1/n_p$ if vertices are assigned to properties at random. Using the notation of equation (9),

$$\sum_s e_{st} = \sum_t e_{st} = \frac{1}{n_p}. \tag{24}$$

Equations (23) and (24) define a set of constraints which determine the ensemble mixing matrix values. It can be shown that,

$$\begin{aligned} e_{st} &= \frac{1}{n_p(1+(n_p-1)e^{-\alpha})} \quad \forall s = t \\ e_{st} &= \frac{e^{-\alpha}}{n_p(1+(n_p-1)e^{-\alpha})} \quad \forall s \neq t, \end{aligned} \tag{25}$$

are the solutions which satisfy the constraints defined in equations (23) and (24) for an n_p x n_p mixing matrix. The assortativity coefficient of the ensemble is determined by substituting the values of the ensemble mixing matrix in equations (24) and (25) into equation (10) to give,

$$r = \frac{\frac{1}{(1+(n_p-1)e^{-\alpha})} - \frac{1}{n_p}}{1-\frac{1}{n_p}}, \tag{26}$$

and after rearranging,

$$r = \frac{1-e^{-\alpha}}{1+(n_p-1)e^{-\alpha}}. \tag{27}$$

Equation (27) describes the variation of the ensemble assortativity coefficient with the homophily, for a given number of discrete properties. Note that this expression is valid for $\alpha \geq 0$ and for integer $n_p \geq 2$ since we assume there is more than one discrete property. Alternatively, if we know the assortativity coefficient of a network and the corresponding number of discrete properties, we may wish to determine the best homophily value for modeling the network. If we assume that a network with an assortativity r is a typical network of the corresponding ensemble, then a suitable homophily for modeling the network growth is given by,

$$\alpha = \ln\left(\frac{1+r(n_p-1)}{1-r}\right). \tag{28}$$

We can use a similar approach to derive an expression for the ensemble assortativity coefficient for properties which show disassortative mixing, where edges tend to connect between dissimilar vertices. If a property can take many possible discrete values, disassortativity means that a vertex tends to connect preferentially to a vertex with any property other than the same property as itself. This is relatively rare for discrete properties, and is generally

more relevant to scalar properties, for example vertex degree, but we include the analysis for completeness as this only requires a simple extension of the above analysis. If vertices tend to connect to dissimilar vertices, then equation (23) becomes,

$$\frac{e_{st,\forall s=t}}{e_{st,\forall s\neq t}} = e^{-\alpha}. \tag{29}$$

Equation (24) is still valid, and using the constraints defined in equations (24) and (29) gives,

$$\begin{aligned} e_{st} &= \frac{e^{-\alpha}}{n_p(e^{-\alpha}+n_p-1)} \quad \forall s = t \\ e_{st} &= \frac{1}{n_p(e^{-\alpha}+n_p-1)} \quad \forall s \neq t. \end{aligned} \tag{30}$$

Finally, using these mixing matrix values and substituting equations (24) and (30) into equation (10), after simplification gives,

$$r = \frac{e^{-\alpha}-1}{e^{-\alpha}+n_p-1}, \tag{31}$$

$$\alpha = \ln\left(\frac{1-r}{r(n_p-1)+1}\right). \tag{32}$$

References

1. R. Albert, A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000)
2. B. Bollobás, O. Riordan, J. Spencer, G. Tusnády, *Random Structures and Algorithms* **18**, 279 (2001)
3. K. Klemm, V.M. Eguíluz, Phys. Rev. E **65**, 036123 (2002)
4. Z. Burda, J. Jurkiewicz, A. Krzywicki, Phys. Rev. E **69**, 026106 (2004)
5. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)
6. L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, *Proceedings of the National Academy of Sciences USA* **97**, 11149 (2000)
7. R. Pastor-Satorras, A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, UK, 2004)
8. R. Albert, H. Jeong, A.-L. Barabási, Nature **401**, 130 (1999)
9. S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, UK, 1994)
10. M.E.J. Newman, *Proceedings of the National Academy of Sciences USA* **98**, 404 (2001)
11. P. Smolen, D.A. Baxter, J.H. Byrne, Neuron **26**, 567 (2000)
12. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, Nature **407**, 651 (2000)
13. J.G. White, E. Southgate, J.N. Thompson, S. Brenner, *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* **314**, 1 (1986)
14. G.P. Garnett, R.M. Anderson, Journal of Infectious Diseases **174**, S150 (1996)
15. P.F. Lazarsfeld, R.K. Merton, *Freedom and Control in Modern Society*, edited by M. Berger (Van Nostrand, New York, USA, 1954), pp. 18–66
16. M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002)
17. R. Cohen, D. Erez, K. ben-Avraham, S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000)
18. S. Maslov, K. Sneppen, Science **296**, 910 (2002)
19. P. Holme, B.J. Kim, C.N. Yoon, S.K. Han, Phys. Rev. E **65**, 056109 (2002)
20. J. Gelissen, Social Science Research **33**, 361 (2004)
21. D. Lusseau, M.E.J. Newman, *Proceedings of the Royal Society of London Series B - Biological Sciences* **271**, S477 (2004)
22. F. Menczer, Eur. Phys. J. B **38**, 211 (2004)
23. S.-H. Yook, Z.N. Oltvai, A.-L. Barabási, Proteomics **4**, 928 (2004)
24. E.A. Leicht, P. Holme, M.E.J. Newman, Phys. Rev. E, 9:0510143 (2005)
25. M.E.J. Newman, Phys. Rev. E **67**, 026126 (2003)
26. M.E.J. Newman, M. Girvan, *Statistical Mechanics of Complex Networks*, edited by R. Pastor-Satorras, J. Rubi, A. Diaz-Guilera (Springer, Berlin, Germany, 2003), pp. 66–87
27. M.E.J. Newman, M. Girvan, Phys. Rev. E **69**, 026113 (2004)
28. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
29. D.-H. Kim, B. Kahng, D. Kim, Eur. Phys. J. B **38**, 305 (2004)
30. M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, A. Arenas, Phys. Rev. E **70**, 056122 (2004)
31. M. Catanzaro, G. Caldarelli, L. Pietronero, Phys. A **338**, 119 (2004)
32. M. Catanzaro, G. Caldarelli, L. Pietronero, Phys. Rev. E **70**, 037101 (2004)
33. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
34. Shi-Jie Yang, Hu Zhao, *Generating multi-scaling networks with different types of nodes*, e-print `arXiv:cond-mat/0508338` (2005)
35. R.N. Onody, P.A. de Castro, Phys. A **336**, 491 (2004)
36. R. Xulvi-Brunet, I.M. Sokolov, Phys. Rev. E **70**, 066102 (2004)
37. A. Trusina et al., Phys. Rev. Lett. **92**, 178702 (2004)
38. H. Jeong, Z. Néda, A.-L. Barabási, Europhys. Lett. **61**, 567 (2003)
39. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000)
40. D.J. Watts, P.S. Dodds, M.E.J. Newman, Science **296**, 1302 (2002)
41. G. Bianconi, A.-L. Barabási, Europhys. Lett. **54**, 436 (2001)
42. G. Bianconi, A.-L. Barabási, Phys. Rev. Lett. **86**, 5632 (2001)
43. J. Gómez-Gardenñes, Y. Moreno, Phys. Rev. E **69**, 0.37103 (2004)
44. C. Castellano et al., Eur. Phys. J. B **38**, 311 (2004)
45. M. Girvan, M.E.J. Newman, *Proceedings of the National Academy of Sciences USA* **99**, 7821 (2002)
46. M.E.J. Newman, Phys. Rev. E **69**, 066133 (2004)
47. P. Holme, B.J. Kim, Phys. Rev. E **65**, 026107 (2002)
48. A.F.J. van Raan, Scientometrics **47**, 347 (2000)
49. M.E.J. Newman, SIAM Rev. **45**, 67 (2003)
50. P. Shannon et al., Genome Research **13**, 2498 (2003)
51. V. Gol'dshtein, G.A. Koganov, G.I. Surdutovich, Vulnerability and hierarchy in complex networks, ArXiv Condensed Matter e-prints, p. 0409298 (September 2004)
52. E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.-L. Barabási, Science **297**, 1551 (2002)
53. T. Ozugur, IEEE Transactions on Wireless Communications **4**, 775 (2005)