Network Science **2** (3): 326–340, 2014. © Cambridge University Press 2014
doi:10.1017/nws.2014.24

The evolution and structure of social networks

WHITMAN RICHARDS
CSAIL- 32-364, Mass. Inst. of Technology, Cambridge, MA 02139, USA
(e-mail: wrichards@mit.edu*)*

NICHOLAS WORMALD
School of Mathematical Sciences, Monash University, Clayton VIC 3149, Australia
(e-mail: nicholas.wormald@monash.edu*)*

Abstract

As social networks evolve, new nodes are linked to the large-scale organization already in place. We show that the combination of two simple algorithms, one the Barabasi-Albert preferential attachment proposal and the other a neighbor attachment rule, successfully generate networks exhibiting both the local and global characteristics of empirical data on social network structures. Ideally, one might hope that some coarse features of this linking process and the form of the local patterns might enable the prediction of large-scale properties. We show that this is generally not the case. This might help explain the variety of local and global patterns in empirical networks.

Keywords: *network evolution, preferential attachment, social networks, degree distributions*

1 Introduction

Over the past decade there has been an increasing interest in understanding and modeling the structure and dynamics of social networks. When a new node (or agent) is added to a network, how are the links to current nodes selected? How do these choices affect the resulting large-scale structure of the network as it grows? Finally, what is the relation between the local structure in the network and its global form?

Barabasi & Albert (1999) and others (see Dorogovtsev and Mendes (2003)) have proposed that new attachments in a network favor nodes with higher degrees, which presumably would be the most popular choices. We explore a revision of this proposal, namely that after a new node makes its first connection, to a node u, attachments favor neighbors of u. The key motivation is that when an individual joins a social network, that individual tends to meet friends of his first acquaintance rather than arbitrary existing members. We explore variations on this theme related to preferential versus random (uniform) attachments. As we explain below, models related to this were studied by Kumar et al. (2000); Dorogovtsev & Mendes (2003); Blum et al. (2006); and Bebek et al. (2006).

In this article we examine variations of the following rule for iteratively attaching new nodes to an existing network.

Rule N: Choose a node, say v_j in the network G^*_{n-1} using a prescribed distribution D_1. Link the new node v_n to v_j. Then for a fixed $k \geqslant 1$ add k additional links from

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

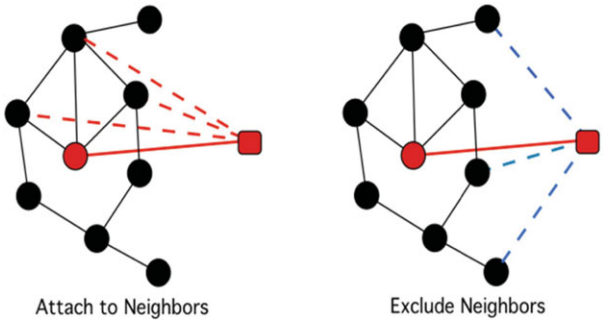

Fig. 1. Neighbor attachments of a new node (square) to a first node (red disc) increase the number of triangles, and consequently the cluster coefficients (left.) Non-neighbor attachments lead to smaller cluster coefficients. From Richards, 2011.

v_n to k neighbors of v_j chosen according to a prescribed distribution D_2 on those neighbors.

Invoking Rule N means that, after the first link, the remaining k nodes linked to must all be neighbors of the first.

We obtain variations in Rule N by specifying the fixed integer k and the distribution D_1 and D_2. We will examine in particular two versions of the distribution D_1: (1) U denotes a uniform distribution of all nodes; (2) P denotes that the node is chosen preferentially, with probability proportional to its degree. Thus P gives exactly the bias encountered in the Barabasi-Albert (BA) rule. We also use U (or P) to denote the analogous version of D_2, in which the k nodes are chosen repeatedly using uniform (or preferential) probabilities. This gives four versions of Rule N for a given k: UU, UP, PU, and PP, where the first letter applies to the first node chosen and the second for the remaining k nodes. To make our main points, we will focus on two of these, namely UU, PP, plus BA. Normally, we will not permit multiple edges in the U versions, but permit them in the P versions. (Permitting two nodes to have multiple links between them was not expressly permitted or forbidden by Barabasi & Albert.)

Note that if the neighbor constraint is dropped in the PP version, we then have the Barabasi-Albert (BA) rule. We believe that, like the BA rule, networks that evolve with Rule N will also lead to a power law for degree distributions and are hence scale-free. The intuition behind this is the fact that even for the UU version of Rule N, a node's likelihood of being chosen as a secondary node (not the first one chosen) is typically proportional to its degree. Using a preferential rule for D_1 or D_2 should only assist the effect. Moreover, we see Rule N as a natural way to achieve the scale-free behavior of social networks. It also has other significance, which we elaborate below with examples of evolving networks.

As shown in Figure 1, the obvious effect of including the neighbor constraint is that at least k triangles are created in G^*_{n-1}, improving the clustering among nodes (Dorogovsev & Mendes, 2003.) This is consistent with the high clustering coefficients that characterize many of the social networks we have studied (Macindoe, 2010; Richards & Macindoe, 2010.)

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

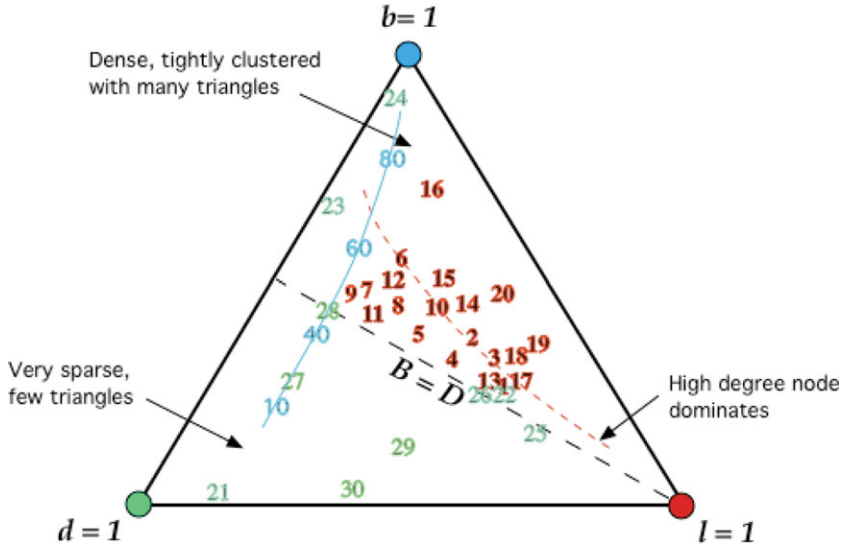

Fig. 2. Locations in the LBD simplex of 20 social networks are shown in red. Note these all lie above the B = D diagonal, and are quite distinct from the location of Erdos-Renyi random graphs (blue locus with link probabilities indicated) as well as "small world graphs" (number 23) and other random networks (green). See Table 1. (From Macindoe & Richards, 2010.)

2 Large scale characterization of networks

Pictures fail to provide vivid representations of large networks and hence are often difficult to differentiate the detailed form of these networks (see Read & Wilson, 1998 for the explosion in number of different graphs as number of nodes increases.) Hence researchers have chosen to characterize networks using parameterizations such as degree distributions, leadership measures, cluster coefficients, characteristic path lengths and a host of others (Newman 2003, 2006). One of the more popular parameterizations is the distribution of degrees of nodes. However, it is generally agreed that this characterization fails to capture the underlying essence of the network, with quite different evolutionary processes leading to nearly identical degree distributions, with different underlying local structures (Wormald & Farczadi, 2013). A new "LBD" parameterization (to be described shortly) provides more useful insight into network differences (Richards & Wormald, 2009). For example, all social networks we have studied lie in a limited portion of the space created by the LBD parameterization, as shown in Figure 2. These networks (red numbers) lie outside of the region of Erdos-Renyi random graphs (blue locus with link percent labels), and also lie above the B = D divider of the LBD simplex. The diagram also includes some of the BA networks (e.g. numbers 25 & 26). Note that these lie at the opposite side of the simplex from the ER networks, nearer L = 1.

Briefly, the LBD parameters reflect three important aspects of social networks: leadership, the bonding of agents, and their diversity. The definitions are as follows:

L: For any graph G_n, let d_i be the degree of node v_i. The leadership index is then:

$$L = \sum_{i=1}^{n} (d_{\max} - d_i)/((n-2)(n-1))$$

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

This relation sums the difference in the degree of a node with respect to the maximum degree in G_n, and normalizes this sum by the maximum possible (Freeman, 1978.)

B: The bonding index the number of triangles normalized by the maximum achievable by a graph with the same number of (directed) paths of length 2:

$$B = 6 * (\#\text{triangles})/(\#\text{paths\_length\_two})$$

This is one of two classical measures going by the name of clustering coefficient (Newman 2003, (3.4) and (3.6)). Note that if G_n is the fully connected graph K_n, then bonding ***B*** is maximal with value "1", whereas for the "star" graph S_n or for any tree T_n, the bonding index will be zero.

D: The diversity index counts the number of pairs of disjoint nonadjacent dipoles K_2 in G_n with $n \geqslant 4$. This count is divided by the number in the graph consisting of two disjoint cliques of size n/2, thus normalizing the measure to the interval [0,1]. The square root boosts low ratios (Richards & Wormald, 2009):

$$\mathrm{D} = \mathrm{Sqrt}\left[(\#\text{disjoint\_dipoles})\Big/\left(\frac{1}{2} * \frac{n}{2}\left(\frac{n}{2} - 1\right)\right)^2\right]$$

For convenience, it is useful to project the raw LBD values onto the <111> plane as follows to create "the simplex".

$$\begin{aligned} \boldsymbol{l} &= \boldsymbol{L}/(\boldsymbol{L}+\boldsymbol{B}+\boldsymbol{D}) \\ \boldsymbol{b} &= \boldsymbol{B}/(\boldsymbol{L}+\boldsymbol{B}+\boldsymbol{D}) \\ \boldsymbol{d} &= \boldsymbol{D}/(\boldsymbol{L}+\boldsymbol{B}+\boldsymbol{D}). \end{aligned}$$

Figure 2 illustrated this compression. At the top of the triangle $\boldsymbol{B} = 1$, (green node) which is the position of a dense graph topology corresponding to the complete graph. All complete graphs, regardless of size will be mapped to this point. Likewise, with $\boldsymbol{L} = 1$ ($\boldsymbol{B} = 0$, $\boldsymbol{D} = 0$), the topology of the graph will be a "star", with all star graphs regardless of size mapped to the red node at lower right. A ring topology is a simple example of $\boldsymbol{D} = 1$ with $\boldsymbol{B} = \boldsymbol{L} = 0$. Trees with nodes having roughly equal degrees will have high $\boldsymbol{D}$ values, with low $\boldsymbol{L} \sim 0$ and $\boldsymbol{B} = 0$. Hence when diversity becomes maximal, the graph will be located at the lower left. (See Macindoe & Richards, 2010; Macindoe, 2010; Richards & Wormald, 2009 for details.)

3 Network evolution

The LBD simplex is a convenient representation to compare how social networks might evolve. In particular, we use the LBD simplex to reveal substantial differences in the BA algorithm as well as variations in Rule N that arise from choosing preferential attachments by degree or simply uniform random choices. In Figure 3, four different trajectories are plotted that use different kinds of preferential choices. Obvious variables are (1) preferential attachments or not (2) whether neighbors are selected or not. For these simulations we keep the number of attachments for new nodes set at 4 for each iteration. A further parameter is the structure of the seed graph used to initiate the iterations. We have found that this last variable is not a

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

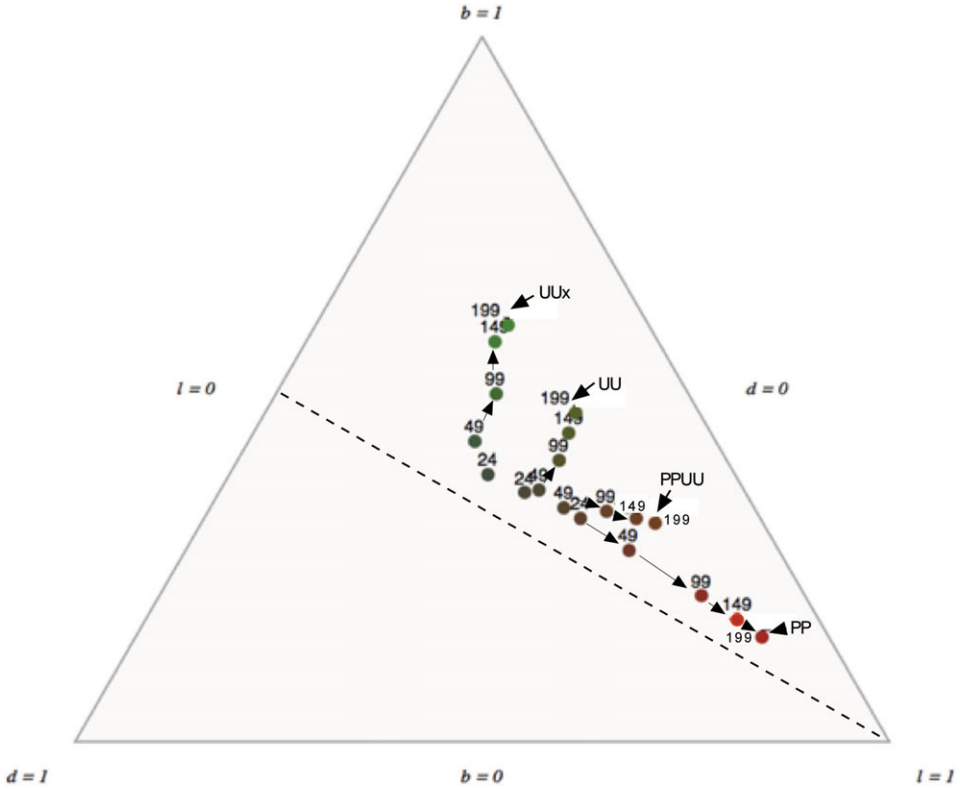

Fig. 3. Four different evolutions of social networks constructed using neighbor attachments. The labels next to the points indicate the number of iterations. UU does not use preferential degree choices, but PP does. The intermediate trajectory labeled PPUU is a 50–50 mixture of both. UUx is an example of one iteration of UU which helps to show the range of UU trajectories. Note that the lower bound of all these evolutions lies near the B = D divider (dashed line). (Color online)

critical factor for networks evolving past 50 nodes, provided the seed graph has say less than ten nodes.

All trajectories in Figure 3 invoke one of two variations in Rule N, namely a first node v_j is chosen in G^*, and then additional attachments for v_n are to k neighbors of v_j. For trajectory UU, there are no other preferential constraints. In particular, the first node v_j is chosen using a uniform distribution of all nodes in G^*_{n-1}. Likewise, the k additional attachments are chosen uniformly over the neighbors of v_j, with $k = 3$. The resultant locus lies near the border of the Erdős-Rényi random graphs with the UU trajectory headed toward $B = 1$ (i.e. high cluster counts).

In contrast, trajectory PP imposes preferential choices favoring the higher node degrees, both for the choice of the first node v_j as well as for the choices among the neighbors of v_j. Again, k is set at 3. This trajectory lies on the opposite side of the Simplex from the UU locus and heads toward $L = 1$. Hence mixtures of UU and PP can span the region of the simplex occupied by social networks. The third locus in Figure 3 labeled UUPP illustrates a 50–50 mixture where preferential by degree and uniform choices are made with equal probability for each link. A single run of

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

UU, labeled UUx, demonstrates the degree of variation in a typical run. Otherwise the trajectories are averages of 10 runs.

Note that all the evolutions lie above the B = D divider in the simplex, in accordance with the empirical results in Figure 2. We have yet to prove this result formally.

As mentioned, the choice of the (small) seed graph has no impact on the trajectories for networks larger than about 50 nodes. As the network grows from the seed, the evolutionary path moves first to the region near L = B = D (l = b = d), and then departs as illustrated in Figure 3. This is shown in detail in another paper for "start-up" networks that are initiated at $L = 1$ (Richards & Wormald, 2009).

To further span the region in the Simplex above B = D and between UU and PP, we can add non-neighbors to slow down or to truncate the evolution.

Simulations show that even a portion of non-neighbors of 40% of the additional k node choices, causes the trajectory to fall near or below the B = D divider until the network grows to 100 or more nodes. Our observations show that these corrupted networks tend to hover near L = B = D (l = b = d).

In Figure 4 we show the effect of eliminating the neighbor constraint from Rule N. For reference, we include the BA trajectory, for 200 iterations as well as PP (the latter with the neighbor constraint in place.) If the neighbor constraint is not invoked, then PP moves to PP4 and UU moves to UU4. Note that these latter two results lie well below the B = D divider, largely because nodes that are non-neighbors to the first choice nodes are now included in the generation process.

4 Relation between algorithm BA and neighbor rule N

The astute reader has already realized that algorithms BA and Rule N are closely related. After making some slight adjustments to the definitions, we can explicitly make them both extreme cases of a more general algorithm, as follows.

Let G^*_{n-1} be the evolving network after n – 1 iterations. Let v_n be the new node to be linked at the nth iteration to a node v_j in G^*_{n-1}, where v_j is chosen preferentially based on its degree. If the algorithm chooses k uniformly random neighbors of v_j in G^*_{n-1} then it is PU, whilst if it chooses preferentially without the neighbor restriction then it is BA. Now let us expressly permit multiple joins in both cases. That is, for each choice of a node, the previously chosen nodes are not excluded. Moreover, let us insist that during an iteration, probabilities are not recomputed; that is, the degrees of the vertices are used before any edges are added during the one iteration. To obtain a hybrid version, we introduce a parameter $r \leqslant k$, and force the algorithm to choose r neighbors uniformly and a further k – r nodes, without the neighbor restriction, preferentially. Doing this in every iteration, we get a mixture of (versions of) BA and PU that we call the r-hybrid model. We have the following result on the distribution of degree of any particular node.

Theorem: Consider the r-hybrid model for some fixed r, $0 \leqslant r \leqslant k$, and a given seed graph. Letting v be any particular node, the probabilistic distribution of the degree of v after n iterations does not depend at all on r.

The theorem's proof is only a minor variation of a proof in Farczadi and Wormald (2013), where it was shown that the distribution of the degree of a node v is the

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

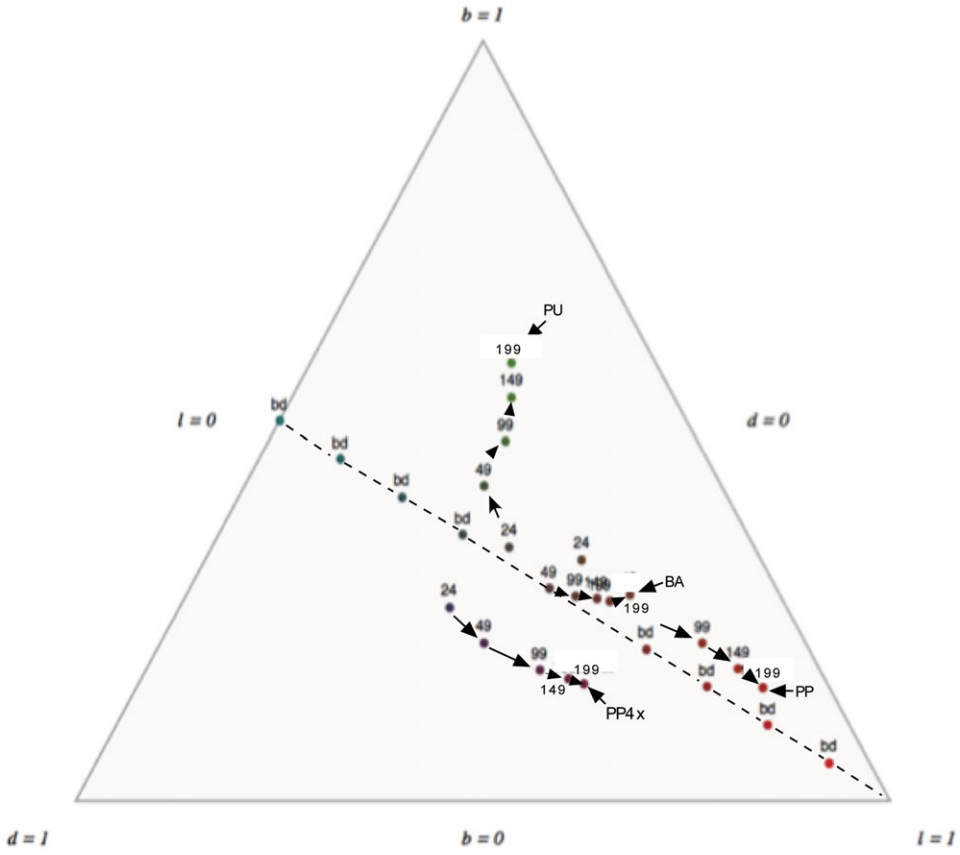

Fig. 4. Effect of including non-neighbors in evolutions. Trajectory PP4x avoids the neighbor constraint and permits non-neighbor attachments. For reference, BA and PP are included, as well as one trial for PU, the latter showing that PU trajectories lie near UU as shown on the LBD Simplex plot in Figure 3 .The location of the BD divider is indicated by the dashed lie and points labeled "bd". (Color online)

same in both BA and PU (though in that case not permitting multiple edges) in a certain sense that involves scaling the number of iterations. So we just sketch it here. Given a number, n, of iterations, let p be the probability that the first chosen node in the next iteration is v_j. Then p is precisely the degree of v_j divided by the total degree of all nodes, due to the preferential choice rule that applies in both versions for the first node. On the other hand, for any of the subsequent links, there are two possibilities. If it is a link chosen preferentially to any node, then the probability this link comes to v_j is again p. If it is a link chosen to a neighbor of the first-chosen node, then a short calculation again shows that the probability that this link comes to v_j is again p. Hence, regardless of r, the number of links joining to v_j has a distribution that depends only on p. By induction on n, the value of p is exactly the same for all r, and the theorem follows.

The above theorem only relates to the degree of any one particular node, showing that its distribution (and hence expected degree and variance of its degree) is exactly the same regardless of r. This is true for all n. Another viewpoint is achieved by considering the joint distribution of all nodes. This can be, and no doubt is, different

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

in the models having different values of r. For instance, in the PU model, consider two adjacent vertices u and v of the seed graph. If u has high degree after n steps, we would expect it to increase the chances that v does too, since each time u is the initial vertex chosen in any step, there is a high chance that v is chosen in the second lot of k vertices. In the BA model, this effect is missing.

The import of the above theorem is to show that the BA-PU mixtures generate large-scale networks that are correlated, although different. However, this says almost nothing about their local structures, which will be considered in the next section. Indeed, an interesting result is that the local structures for PU, UU and PP are very similar, but quite different from the structure of networks evolved using the BA algorithm. Hence the global LBD index appears heavily dependent on the choice of distribution P or U, and less so on the local structure of its components as influenced by enforcing or ignoring the neighbor constraint.

Although the theorem says nothing about PP and UU, it leads us to the following slightly imprecise conjecture for social networks using mixtures of the BA and N algorithms:

Conjecture: All networks of order (size) n generated using mixtures of the BA and N algorithms will have similar degree distribution.

Admittedly, the conjecture describes an imprecise trend rather than an exact statement that can be checked.

Extending to empirical networks, we have the following:

Prediction: All social networks of order (size) n will have similar degree distribution.

The degree distribution is of course easy to compute for a given empirical network; the difficulty would be to find different empirical networks of the same order. See Table 1 for a list of small social networks.

5 Small scale parameterization

To address the relation between the local (small scale) structure of the networks and its global (LBD simplex) characterization we define an index for local structures that captures most of those found by Stoica & Prieur, 2009; Milo, 2002; Palla et al, 2005. Our choice for local structure is the subgraph induced by the nodes adjacent to a given node. For instance, this could be a complete graph K_n or the degenerate star S_j. Figure 5 illustrates a range of such graphs for $n = 5$ and the R^* index used to characterize this local (small scale) measure. The measure makes explicit two extreme forms, namely the complete graph ($R^* = 1$) or the star ($R^* = 0$).

Definition: the small graph ratio R^* of a given node v, called the "hub", is the number of links joining pairs of nodes that are one link step from v divided by the maximum possible such links, which is $d(d-1)/2$ if v has degree d.

Note that R^* is the same as the "local value" of the clustering coefficient defined at eqn. (3.5) of Newman (2003).

In the previous section we considered degree distribution in BA and PU. To compare small-scale structures, we focus on PP and UU since PU is presumably intermediate between these. Figure 6 compares the small-scale structure for PP and

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

Table 1. *From Richards, W. & O. Macindoe (2010) Characteristics of small social networks; MIT Computer science and artificial intelligence Lab Tech report MIT-CSAIL-TR-2010-033. (Downloadable from that paper are the references to networks given in the last column.)*

Num	Graph	Vertices/Nodes	v	e	Edges	Ref.
1	Seed+	Start-up	6	7	Collaboration	34
2	SmallGroup	Start-up (avg)	25	42	Collaboration	34
3	LosAlamos	Scientists	30	78	Collaboration	31
4	Karate	Club Members	34	78	Friendship	42
5	Enron	Employee email	37	50	Email exchange	9
6	LesMiserable	Characters*	40	105	SceneCoappearance	18
7	HIV	Core Group	40	56	Friendship	32
8	Bright	Words	54	175	Free associations	34
9	Dolphins	Dolphins	62	159	Time in Proximity	22
10	Enron	Employee email	79	147	Email exchange	9
11	PolVotes	Senate 2009	99	356	Same Votes	19
12	PolBooks	Books	105	441	Purchased together	20
13	Adj-Noun	Adjectives & Nouns	112	425	Co-occurence in DavidCopperfield	27
14	SantaFe	Scientists	116	174	Collaboration	13
15	Enron	Enron Employees	143	623	Email exchange	9
16	JJATT	Terrorists	263	998	Known associate	3
17	C.Elegans	Neurons	297	2148	Neural connection	40
18	Linux2001	Kernel mailing list members	302	749	Email exchange	14
19	Linux2008	ditto	447	2122	ditto	14
20	PolBlogs	Political Blogs "Non-Cognitive"	1490	16715	Blog Hyperlinks	1
21	BinaryTree	Binary Tree	127	126		
22	InfxDisease	HIV spread	250	266	transmission	32
23	Football	College football tournament	115	613	Match played	13
24	SmallWorld	Ring seed	1000	10^3		
25	Barabasi_2	Multi-scale (2 Attach)	500	982	PrefAttach	5
26	Barabasi_5	Multi-scale (5 Attach)	500	2422	PrefAttach	5
27	Erdős-Rényi	Random graph (eP=0.1)	100	524		
28	Erdős-Rényi	Random graph (eP=0.4)	100	2083		
29	BuddedTree	Tree + Triangle buds	109	110		
30	RandomTree		100	99		

UU. To eliminate overlap between data sets, these curves, and others to follow, have been scaled by number of iterations, i.e. approximate network size (which is indicated by the label on curves). Although different in detail, there is still a striking similarity in their R^* distributions. In other words, although their large-scale LBD values are at quite different locations in the simplex (See Figure 2), their local structure is comparable. Both invoke the neighbor constraint. The principal difference is that for UU the choice of nodes is uniform over the neighbors, whereas for PP the choice is preferential, favoring higher degree nodes (hence the larger L value). The similarity in the two R^* distributions suggests that for less than 200 iterations the difference is small to first order.

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

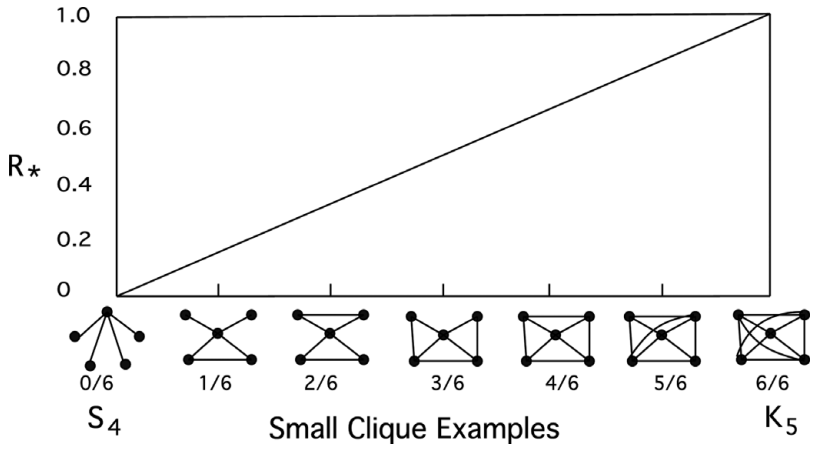

Fig. 5. Local structures can be characterized by the number of links between nodes in the rim (i.e. links to the central hub are excluded) to give a ratio R* that is the fraction of the maximum of such links. (See text.)

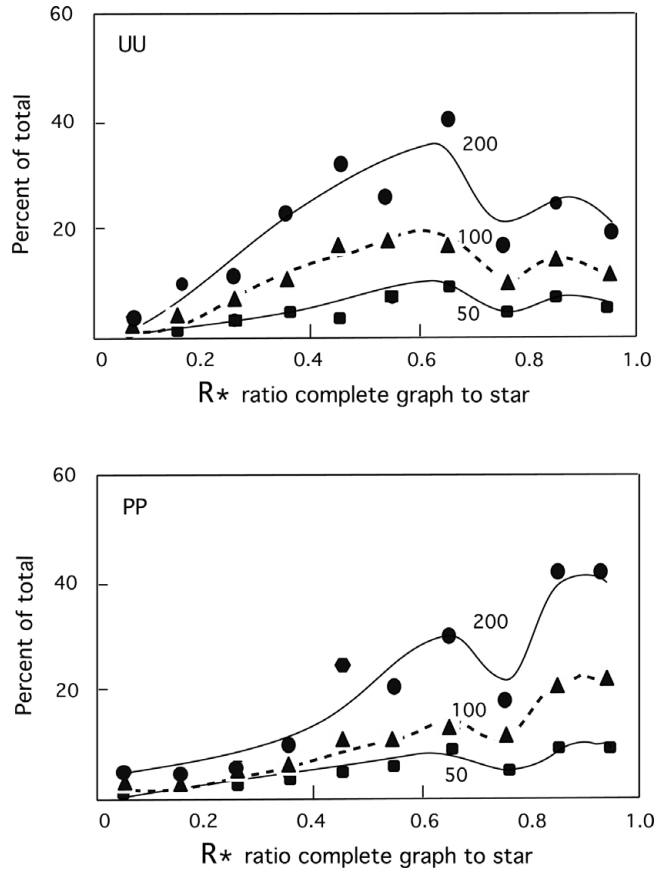

Fig. 6. Comparison of R* distributions for PP and UU. Note the similarity in shape, even though the trajectories are quite distant in the LBD simplex. See text for a note on how each curve is scaled.

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

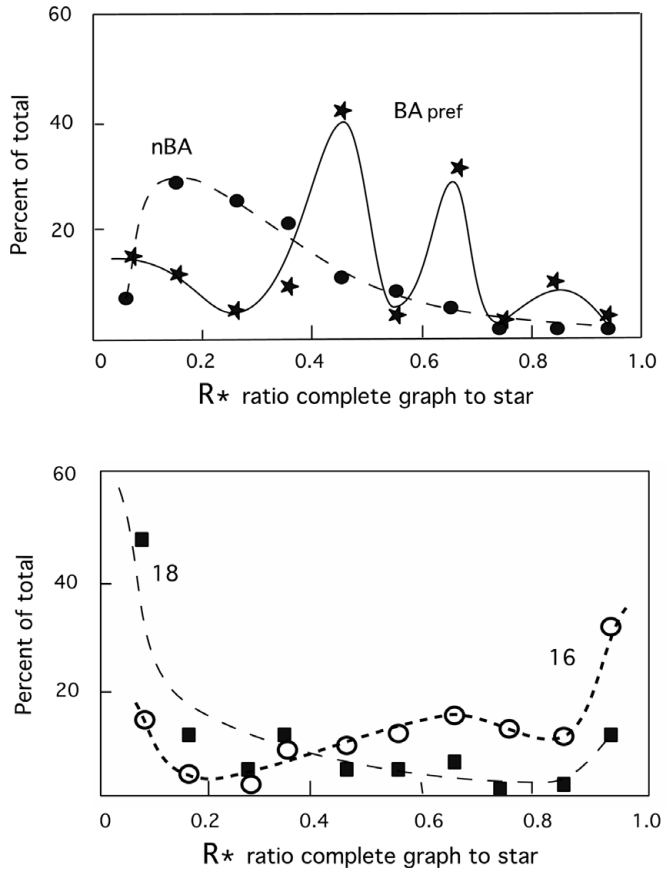

Fig. 7. Lower panel: R^* distribution for two social networks: Linux (16), JJATT (18). Upper panel: R^* distribution for algorithm BA (200 iterations) and also a hybrid of rule N and the BA algorithm (nBA).

In contrast, if we use the BA algorithm, the R^* pattern is quite different, as shown in the upper panel of Figure 7.

Furthermore, if the neighbor constraint is ignored, and non-neighbors to the first choice are chosen, then in both the UU and PP versions, the small scale distributions shift dramatically toward low R^* values. (Recall that when the neighbor constraint is dropped in the PP version, it becomes the BA rule.) In other words, whereas in Figure 6 there is significant dense clustering with large R^* values, without the neighbor constraint the clusters are sparse moving more toward the left of Figure 5 (i.e. "star-like" patterns). Furthermore the LBD values shift below the B = D divider. Hence with non-neighbor choices, not only does LBD change dramatically, but so does R^*. (There is also a small average number of triangles per node, namely about 1.5 in contrast to about 6 triangles per node with Rule N in force.) Once again, the effect of preferential attachments or uniform for the distributions D_1 and D_2 makes little difference.

To summarize, the use of Rule N or not is a huge categorical constraint that dramatically affects small scale structure, just as preferential or uniform node choice affects global LBD values.

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

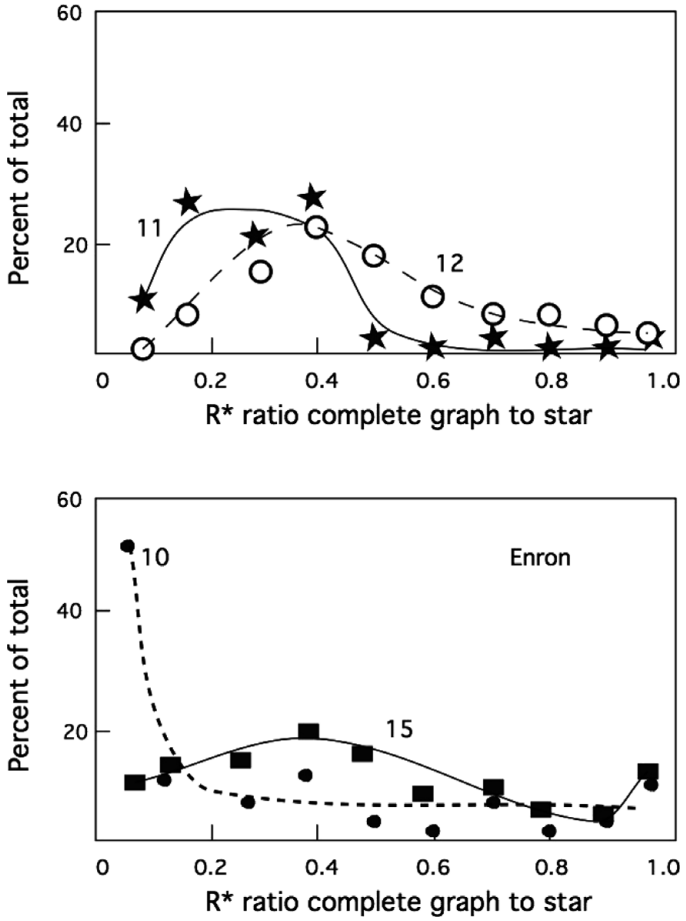

Fig. 8. Lower panel: R^* distribution for two Enron email networks: Circles, 79 employees (10), Squares, 143 employees (15). Upper panel: R^* distribution for two other networks: Senate (11), Polbooks (12). Points scale by 100/network size.

The lower panel of Figure 7 and the upper panel of Figure 8 show some R^* distributions for empirical data. Curve 18 is Linux, with high L values and curve 16 is JJATT (a terrorist network), whereas curve 11 is senate voting and 12 is political book selections. Curves 10 and 15 are two samples from an Enron network. These distributions suggest a mix of the simplified, ideal conditions, i.e. Rule N, preferential vs not, etc. To recreate these distributions we need a mixture of rules as the network evolves. We can take a step in this direction by mixtures of rules BA and N. Kumar et al. (2000) proposed a directed graph model related to this in the case of UU. It uses U to determine the initial node v_j to link to, and after that, for each link it uses a random outlink of v_j with probability p, and otherwise it links to any node chosen using U. All new links are directed outwards from the new node v_n.

For example, a simple mixture of rules BA and N (option UU) is shown in the upper panel of Figure 7 (curve labeled nBA.) Here the BA algorithm is used for three of the four attachments, reserving one to implement with the neighbor rule. The bottom panel of Figure 7 shows that a trade-off between BA and rule N are consistent with empirical data. Note that our "mixture" of BA and rule N labeled

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

nBA in the top panel of Figure 7 is similar to empirical network 11. Other networks are shown similarly in Figure 8.

Our proposal, then, is that most empirical networks use a combination of rule BA and rule N when evolving (we have no insight as to which option is chosen, UU or PU.) This makes sense in networks evolving by a social mechanism. An obvious first choice for an attachment is a friend in G^* of the new node, and its neighbors. But another realistic possibility is to choose a node from another clique in G^* that may have a second good friend, or simply because members of that clique are engaged in activities of interest to the agent associated with the new node (Granovetter, 1973). A mixture of BA and neighbor attachments would then be very common. If geometry is involved, similar effects might occur, based on proximity of nodes to the one first chosen.

6 Discussion

Our main result is that a simple mixture of the Barabasi-Albert preferential attachment algorithm (BA) together with Rule N (option UU) that enforces attachments to neighbors of a first choice (UU) can generate networks with LBD values that lie in the region of the simplex that is also the location of twenty social networks we have studied (Figure 2.)

A secondary, disappointing result is that there is no strong correlation between local structures and the global structure captured by its LBD value. Rather, a dominant factor in the resultant global structure, for those algorithms we study here, is whether or not preference of high degree vertices is included, but this has a much lesser effect on local structure. This failure to find strong correlations between a small-scale measure and the LBD location was also noted by Macindoe (2010), using a local LBD index.

To model network evolution, typically we will need a mix of strategies at various stages in the evolution of the network. One such stage is the completion of "start up" at about 25 nodes, where L∼B∼D (see Richards & Wormald (2009), Richards & Macindoe (2010)). Following this initial stage, different factors come into play. Another stage begins near 100 node or larger networks where the size of the possible node choices grows (whether neighbors or not). Here one must expect a high degree of arbitrariness in choice. These stages create enormous flexibility in the evolutionary process (e.g. Figure 8 lower panel shows that the local structure can change as the network evolves) and are one of the features that limits our ability to predict trajectories, but not necessarily our ability to duplicate them (Dunbar, 1992).

A strong positive benefit of BA-N mixtures is that both algorithms are highly motivated. A new member will favor a friend already in the network, as well as friends of his first choice. However, it also seems quite likely that a new member might also explore other cliques, or seek new attachments in areas that look intriguing. These choices are reflected in the combination of the two algorithms.

Acknowledgements

In his MS thesis at MIT, Owen Macindoe compiled link data of networks which is accessible from MIT Computer science and artificial intelligence Lab Tech report

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

MIT-CSAIL-TR-2010-033. We appreciate the availability of these data, and also helpful comments by Rajesh Kasturirangan. We also acknowledge helpful comments from Linda Farczadi, particularly with regard to existing literature. Support for earlier work was provided by AFOSR with the support and guidance of Terry Lyons. Finally, we thank one of the reviewers for a careful reading of the manuscript and comments. Nicholas Wormald was supported by the ARC via an Australian Laureate Fellowship.

References

Adamic, L. A., & Glance, N. (2005) The political blogosphere and the 2004 US election. *Proceedings WWW-2005 Workshop on sthe Webblogging Ecosystem.*

Atran, S., Bennett, S., Fatica, A., Magouik, J., Noricks, D., Sageman, M., … Wright, D. (2008) John Jay & ARTIS Transnational Terrorism (JJATT) dataset. http://doitapps.jjay.cuny.edu/jjatt/

Barabasi, A-L. (2002). *Linked: The New Science of Networks*. NY: Perseus Press.

Barabasi, A-L. and R. Albert (1999) Emergence of scaling in random networks. *Science,* **286**, 509–511.

Bebek, G, P. B., Cooper, C., Friedetzky, T., Nadeau, J., & Sahinalp, S. C. (2006). The degree distribution of the generalized duplication model. *Theoretical Computer Science*, **369**, 239–249.

Blum, A., Chan, T. H. H., & Rwebangira, M. R. (2006). A random-surfer web-graph model. *ANALCO,* **6**, 238–246.

Bollabas, B. (2001). *Random Graphs* (2nd ed.). Cambridge Univ. Press.

Bourassa, V., & Holt, F. (2003) SWAN: Small-world wide area networks. *Proceedings of International Conference on Advances in Infrastructure (SSGRR 2003w paper # 64)*, L'Aquila, Italy.

Caldarelli, G, Pastor-Satorras, R., & Vespignani, A. (2004) Structure of cycles and local ordering in complex networks. *European Physics Journal B*, **38**, 183–186.

Cohen, W. W. (2009). Enron email dataset. http://www.cs.cmu.edu/~enron/

Dorogovtsev, S. N., & Mendes, J. F. (2003). *Evolution of Networks*, Oxford Univ. Press.

Dunbar, R. (1992). Neocortical size as a constraint on group size in primates. *Journal Human Evolution,* **22**, 469–493.

Farczadi, L., & Wormald, N. (2013). On the degree distribution of a growing network model, *Mathematical arXiv:1401.0933.*

Freeman, L. C. (1978). Centrality in social networks: conceptual clarification. *Social Networks,* **1**, 215–239.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings National Academy Science*, **99**, 7821–7826.

Gnawali, O. D. (2009). Linus kernel email communication neworks from Jan 2001 and 2008. Personal communication.

Granovetter, M. (1973). The strength of weak ties. *American Journal Sociology,* **78**, 1360–1368.

Kasturirangan, R. (1999). Multiple scales in small world graphs. MIT-CSAIL technical reports: //publications.ai.mit.edu/ai-publications/pdf/AIM-1663.pdf

Knuth, D. E. (1993). *The Stanford GraphBase: A Platform for Combinatorial Computing*. Addison-Wesley, Reading Ma.

Krebs, V. (2003). Books about US politics dataset (unpublished.) http://www.orgnet.com/

Kumar, R., P. R., Rajagopalan, S., Sivakumar, D., Tomkins, A. & Upfal, E., (2000). Stochastic models for the web graph. *Proceedings Foundations of Computer Science, 41st Annual Symposium*, 57–65.

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press

Leskovec, J., Singh, A., & Kleinberg, J. (2006) Patterns of influence in a recommendation network. *Proceedings Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*.

Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. (2003) The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, **54**, 396–405.

Macindoe, O. (2010). Investigating the fine grained structure of networks. MS Thesis. Dept. Elec. Engr. and Comp. Sci. http://

Macindoe, O. & Richards, W. (2010) Graph comparison using fine structure analysis. *Proceedings IEEE Conference on Society Computing*, # 244.

Milo et al. (2002). *Network Motifs. Science*, **298**, 824–827.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review,* **45**, 167–256.

Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, **74**.

Newman, M. E. J. (2005). A measure of betweenness centrality based on random walk. *Social Networks,* **27**, 39–54.

Page, S.E. (2007). *The Difference: how the power of diversity creates better groups, firms, schools, and societies*. Princeton Univ. Press.

Palmer, E. M. (1985). *Graphical Evolution*, Wiley & Sons.

Palla, G. Derenyi, I. Farkes, I. & T. Viesek (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, **435**, 814–818.

Read, R. C. & Wilson, R. J. (1998). *An Atlas of Graphs*. Oxford Press.

Richards, W. & Macindoe, O. (2010). Decomposing Social networks. *Proceedings IEEE Conference on Society Computing*, # 205.

Richards, W. & Macindoe, O. (2010). Characteristics of small social networks. MIT Computer science and artificial intelligence Lab Tech report MIT-CSAIL-TR-2010-033.

Richards, W. & Wormald, N. (2009). Representing small group evolution. *Proceedings IEEE Conference on Society Computing*, # 232.

Schelp, R. H. & Thomason, A. (1998). A remark on the number of complete and empty subgraphs. *Combinatorics, Probability and Computing*, **7**, 217–219.

Stoica, A. & Prieur, C. (2009). Structure of neighborhoods in a large social network. *Proceedings IEEE Conference on Society Computing*, # 225.

Watts, D. (2002). A simple model of global cascades on random networks, *PNAS,* **99**, 5766–5771.

Watts, D. and Strogatz, S. (1998). Collective dynamics of small world networks, *Nature,* **393**, 440–442.

White, J. G., Southgate, E., Thompson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode c.elegans. *Philosophical Transactions of the Royal Society,* **314**, 1–340.

Wolfram, S (2002). *A New Kind of Science,* pp. 508–515. Wolfram Media, Inc.

Zachary, W. W. (1977) An information flow model for conflict and fission in small groups. *Journal Anthropological Research,* **33**, 452–473.

https://doi.org/10.1017/nws.2014.24 Published online by Cambridge University Press