The Number of Large Graphs with a Positive Density of Triangles

P. Collet[1] *and J.-P. Eckmann*[2,3]

[1]Centre de Physique Théorique, Laboratoire CNRS UMR 7644, Ecole Polytechnique, F-91128 Palaiseau Cedex, France
[2]Dépt. de Physique Théorique, Université de Genève, CH-1211 Genève 4, Switzerland
[3]Section de Mathématiques, Université de Genève, CH-1211 Genève 4, Switzerland

Abstract. We give upper and lower bounds on the number of graphs of fixed degree which have a positive density of triangles. In particular, we show that there are very few such graphs, when compared to the number of graphs without this restriction. We also show that in this case the triangles seem to cluster even at low density.

1. Introduction and statement of results

In a number of contexts involving large graphs (such as the World Wide Web (WWW) or Citation Networks) it has been observed that such graphs contain a large number of triangles, and probably a positive density of them (per node). We refer to [EM] for a detailed discussion. The importance of topology is also mentioned in [AB], where the authors say (p.41):

> *"But if the topology of these networks indeed deviates from a random graph, we need to develop tools and measurements to capture in quantitative terms the underlying organizing principles."*

On the other hand, it is well known [B, Chapter IV] that in several models of random graphs with a bounded number of links per node the probability of observing a large number of triangles is vanishingly small when the number of nodes diverges. A natural question is therefore to estimate more precisely the number of graphs with a large number of triangles. It will become clear from the discussion of our paper that this result is beyond the "large deviation bounds" which are found in the literature [VV, JLR].

In this paper we study the cardinality of sets of graphs with a *positive density* of triangles per node. We consider (random) graphs with sparse sets of links, *i.e.*, random graphs in which the number of links is bounded by a fixed constant times the number of nodes. We will consider three models of labeled graphs:

G) The model $\mathcal{G}_{n,k}$ comprises the graphs with n nodes and kn links. We call them k-general graphs.

O) The model $\mathcal{G}_{n,k-\mathrm{out}}$ is the set of all graphs with n nodes, and from each node there leave exactly k directed links (directed from that node). We call these graphs k-out.

R) The model $\mathcal{G}_{n,k-\mathrm{reg}}$ is the set of all graphs where at each node exactly k links meet. (This definition is only interesting if kn is even, which we tacitly assume in the sequel.) These graphs are called k-regular.

A well-studied question is that of the probability of finding triangles in such graphs, where the probability is relative to the uniform measure on the set of graphs, giving the same weight to each graph. For all of the above examples, it is known (see *e.g.* [B, JLR]) that the expected number of triangles in these graphs is *bounded* independently of n, by a quantity $\lambda = \mathcal{O}(k^3)$. Furthermore, for each $t \geq 1$, it has been shown (see *e.g.* [B, Theorem IV.1]) that the probability to find exactly t triangles is given, in the limit $n \to \infty$, by the Poisson distribution

$$P(t) = e^{-\lambda} \frac{\lambda^t}{t!} \ .$$

Note however, that this limit is not at all uniform in t, as will be illustrated by our results in Sec. 4. Further studies have greatly refined this result, giving very precise estimates on the tails of this distribution, as a sort of large deviation result. A very recent summary of these results can be found in [JLR]. Our study, in this paper, goes beyond that region, since we ask for the size of subsets of the three graph families with a **positive density** of triangles. We assume throughout that α is a fixed constant $\alpha > 0$ and we consider those graphs in the above classes which have αn triangles (or, more precisely $[\alpha n]$ triangles, where $[x]$ denotes the integer part of x). We denote these subsets by $\mathcal{G}_{n,k,\alpha}$, $\mathcal{G}_{n,k-\mathrm{out},\alpha}$, $\mathcal{G}_{n,k-\mathrm{reg},\alpha}$. If X is a finite set we denote by $|X|$ its cardinality. Our main result is the following

Theorem 1.1. *Fix $k \in \mathbf{N}$ and $\alpha > 0$. For the three graph families we have the bounds (valid when the lower bound is non-negative):*

$$\lim_{n\to\infty} \frac{\log|\mathcal{G}_{n,k,\alpha}|}{\log|\mathcal{G}_{n,k}|} = 1 \ , \tag{1.1}$$

$$\begin{aligned} 1 - \frac{3\alpha}{4(k^2-1)} &\leq \liminf_{n\to\infty} \frac{\log|\mathcal{G}_{n,k-\mathrm{out},\alpha}|}{\log|\mathcal{G}_{n,k-\mathrm{out}}|} \\ &\leq \limsup_{n\to\infty} \frac{\log|\mathcal{G}_{n,k-\mathrm{out},\alpha}|}{\log|\mathcal{G}_{n,k-\mathrm{out}}|} \leq 1 - \frac{\alpha}{2k^2(5k+1)} \ , \end{aligned} \tag{1.2}$$

$$\begin{aligned} 1 - \frac{12\alpha}{k^2-1} &\leq \liminf_{n\to\infty} \frac{\log|\mathcal{G}_{n,k-\mathrm{reg},\alpha}|}{\log|\mathcal{G}_{n,k-\mathrm{reg}}|} \\ &\leq \limsup_{n\to\infty} \frac{\log|\mathcal{G}_{n,k-\mathrm{reg},\alpha}|}{\log|\mathcal{G}_{n,k-\mathrm{reg}}|} \leq 1 - \frac{2\alpha}{k(k-1)} \ . \end{aligned} \tag{1.3}$$

We conjecture that in the statement above the limits exist (assuming, of course, that kn is even in the k-regular case).

Remark 1.2: *From the point of view of Information Theory or Statistical Mechanics/Large Deviations, the number of triangles is an extensive quantity relative to the number of nodes. But the logarithmic bounds we find are* **not** *extensive in the number of nodes: They are extensive and small on the scale (of the logarithms) of the number of graphs. This suggests that the presence of a positive density of triangles is a very strong information about the system.*

Indeed, imposing that the number of triangles is proportional to the number of nodes leads intuitively to the conclusion that if one considers two links emanating from a common node, there is a non zero probability that their ends are also linked.

Remark 1.3: *We prove more precise bounds in (3.2).*

Remark 1.4: *One should note that a k-regular graph is more like a $k/2$-out graph (because each link is counted twice).*

Remark 1.5: *The lower bounds are obtained by constructing graphs containing complete graphs of maximal size. We do not know whether these bounds are optimal. If they are, this would mean that complete graphs are "typical" among random graphs with a positive density of triangles.*

The reader should observe that the lower bound is a little surprising. Indeed, assume $\alpha > 0$ is very close to 0. Then, one might expect that since the density of triangles is very low, they will typically be (edge and node) disjoint in the set $\mathcal{G}_{n,k-\mathrm{out},\alpha}$ (and similarly for $\mathcal{G}_{n,k-\mathrm{reg},\alpha}$). For each such triangle, once one has placed 2 of the 3 links forming it, the third link is already determined when we close the triangle and thus, under the assumption of disjointness, only $nk - \alpha n$ links can be chosen freely [in the case of the regular graphs this number is $nk/2 - \alpha n$] leading to an upper bound of $n^{nk-\alpha n}$. But, our lower bound is larger than that. Therefore we conclude:

Remark 1.6: *In the families of graphs $\mathcal{G}_{n,k-\mathrm{out},\alpha}$ and $\mathcal{G}_{n,k-\mathrm{reg},\alpha}$ the triangles have a natural tendency to coagulate into clusters. In other words, restricting the random graphs to the subset of those with a* **positive** *density α of triangles automatically implies that we can expect those triangles to cluster into complete graphs (of size at most k) even if that density α is very low. Thus, it is statistically "advantageous" for the triangles to coagulate, even if there are very few of them, as soon as their density is positive.*

Another surprising result is that in the case of $\mathcal{G}_{k,n,\alpha}$ of Eq.(1.1), there is *no loss* in the number of graphs on the scale of n^n. This might seem all the more surprising in view of the Poisson distribution of the expected number of triangles (but is more credible in view of the non-uniform diagonal limit $t = \alpha n \to \infty$ which we are considering).

Remark 1.7: *Most results of this paper deal with graphs with the same valence k at each vertex. The graphs one encounters in the WWW have valencies which differ from node to node and one can ask how our results would extend to this more general case. It has been observed (see [AB]) that the distribution of the valencies satisfies an approximate power law of the form $P(k) = Z^{-1}k^{-\gamma}$ with $\gamma \approx 2.5$, Z a normalization and $P(k)$ the probability to find a vertex with k out-links. It is thus an interesting, but hard, problem, to extend the current work to such a case of variable valencies. A few calculations, using the clustering techniques of Section 4 as a guideline for orders of magnitude, leads us to conjecture the following picture: Without further restrictions, the probability to find a density of triangles among all graphs seems to be of order one, provided one allows for complete graphs of size $\mathcal{O}(n^{1/3})$ to appear (and perhaps imposing a condition like $\gamma < \gamma^*$ with $\gamma^* = 3$). This mathematical statement implies the occurrence of* *a density of triangles in a graph with variable valencies,* **but only** *if one allows for these giant complete graphs. For the case of the WWW this would mean that the whole density of triangles would have to come from about $n^{1/3}$ "phone books" with $n^{1/3}$ links each,* **all of which** *basically only connect among themselves. This seems unreasonable. Therefore, we conjecture that the observed density of triangles in the WWW is either a man-made feature, or can be captured through a more constrained probabilistic model (perhaps taking into account human behavior).*

The paper proceeds from the k-regular graphs via the k-out graphs to the k-general graphs. We prove first upper bounds and then the (easier) lower bounds.

2. Upper bound for k-regular graphs

This section should be considered as a warm-up for the next one. Therefore, many arguments are sketched, and the reader can find longer explanations in the next section. On the other hand, the general line of proof should be more transparent. The reader will also notice that the k-regular case is much less delicate than the k-out case.

We assume that nk is even because otherwise there are no k-regular graphs. For a k-regular graph, the general bound is [B]:

$$|\mathcal{G}_{n,k-\mathrm{reg}}| \approx C(k)^n n^{nk/2} . \tag{2.1}$$

A given link cannot be an edge in more than $k-1$ triangles, because exactly k links meet at each node (see also Lemma 3.3). For every link we say that it is s times occupied if it occurs in s triangles, and every triangle occupies (in this sense) 3 links. Thus the total occupation number is $3\alpha n$, and therefore the number of links involved in edges of triangles is at least $3\alpha n/(k-1)$ (and at most $3\alpha n$). We next bound the number of ways to draw αn triangles. Label the nodes, and for every node i let t_i be the number of triangles having i, j, m as corners with i the *smallest* of the three indices. The number of ways to choose the t_i is

$$\binom{\alpha n + n - 1}{n - 1} \leq 2^{(\alpha+1)n} , \tag{2.2}$$

which is negligible on the scale we consider. With the t_i fixed, we draw the triangles at $i = 1, 2, \dots$. *Note that in this process we will have to place at least $\ell = 3\alpha n/(k-1)$ links.* Now, if we draw a triangle, several things can happen. Either the triangle is already drawn, because its 3 sides have been placed as sides of triangles which have been drawn earlier. No link needs to be placed in this case. In all other cases, between one and 3 links need to be drawn. The least favorable case occurs when 3 links have to be placed. Then, there are at most $n(n-1)$ possibilities to choose the first 2 links and then at most 1 possibility for the third, and we get a factor of $n^{1-(1/3)}$ per link. If only one new link is used, and it starts at i (and the two others are already there), there are at most k^2 possible endpoints for that link and we get a factor $k^2 n^{1-1}$ in this case. If the one missing link is between two links (which are already there when this link has to be placed) we get at most $k(k-1)n^0/2$ possibilities. Finally, if two links are missing, there are 2 possibilities: Either it is two links starting at i, and this makes at most $nk = kn^{2-1}$ ways, or one link starting at i and one not starting at i which makes at most $nk = n^{2-1}k$ possibilities. Indeed, there are n possibilities to choose the end j of the link starting at i and then there are at most k possibilities for choosing the second link. Once two links are chosen, the triangle is completely determined since all its nodes are fixed.

Thus for all these links we get a bound of at most

$$k^{2\ell} n^{\ell-\ell/3} \tag{2.3}$$

possibilities, that is, $k^2 n^{2/3}$ *per link*. Finally, the remaining $kn/2 - \ell$ links can be put in at most $n^{kn/2-\ell}$ ways. Summing over the possible number of links (which is bounded by $3\alpha n/(k-1) \le \ell \le 3\alpha n$ and yields a factor which can be easily absorbed), and combining the two bounds, we get a bound

$$C^n n^{kn/2-3\alpha n/(3(k-1))} = C^n n^{kn/2-\alpha n/(k-1)} . \tag{2.4}$$

This completes the proof of the upper bound of (1.3). □

Remark. A second proof could be derived from a modification of the proof for the case of k-out graphs which we give below.

3. An upper bound for k-out graphs

In this section, we consider the set $\mathcal{G}_{n,k-\mathrm{out}}$ of graphs where each node has k out-links. The cardinality of this set is

$$|\mathcal{G}_{n,k-\mathrm{out}}| = \binom{n-1}{k}^n . \tag{3.1}$$

In other words, we allow for links which go back and forth between 2 nodes, but we do not allow double directed links in the same direction between 2 nodes. Also self-links (loops) are forbidden. We denote by $\mathcal{G}_{n,k-\mathrm{out},\alpha}$ the subset of $\mathcal{G}_{n,k-\mathrm{out}}$ with $[\alpha n]$ triangles, where triangles are counted as follows: Once the links are placed, their orientation is neglected and unoriented triangles are counted, including the multiplicity of the edges (which can be 1 or 2 by what we said above). For example, 3 nodes with the possible 6 directed links between them count as $8 = 2^3$ triangles. By and large, these distinctions are not very essential for the proofs we are going to give and other choices will work with similar proofs.

Since, for fixed j, one has

$$\frac{(m-j+1)^j}{j!} \le \binom{m}{j} \le \frac{m^j}{j!} ,$$

we see that, for fixed k,

$$\lim_{n\to\infty} \frac{\log|\mathcal{G}_{n,k-\mathrm{out}}|}{n\log n} = k ,$$

which we will sometimes write in the more suggestive form

$$|\mathcal{G}_{n,k-\mathrm{out}}| \approx n^{nk} .$$

To be more precise, we define the notation $F(n) \approx n^{nc}$ to mean that there are constants $C_1 > 0$ and C_2 independent of n (but not of k) such that

$$C_1^n n^{nc} \le F(n) \le C_2^n n^{nc} , \tag{3.2}$$

so that the error term in the limit is subexponential. Another way to say this is

$$\log |\mathcal{G}_{n,k-\mathrm{out}}| = n\bigl(k \log n + \mathcal{O}(1)\bigr) .$$

Define

$$\varrho(k) = \frac{1}{2k(5k+1)} .$$

Proposition 3.1. *There is a $C = C(\alpha, k) < \infty$ for which the quantity $|\mathcal{G}_{n,k-\mathrm{out},\alpha}|$ satisfies an upper bound of the form*

$$|\mathcal{G}_{n,k-\mathrm{out},\alpha}| \le C^n n^{n(k-\alpha\varrho(k))} . \tag{3.3}$$

Remark 3.2: *This is the upper bound of (1.2).*

To avoid the notation $[\alpha n]$, we assume henceforth that αn is an integer. We consider a configuration with αn triangles. The triangles which can occur in a k-out graph are of two types, which we call type R (for round) and type F for (for frustrated) depending on the relative orientation of the links.

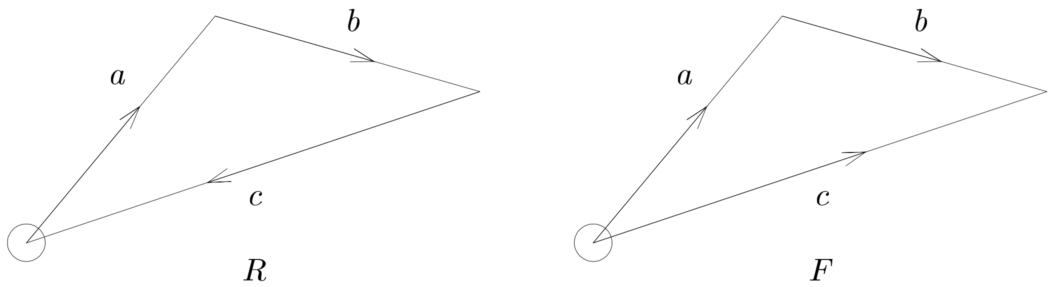

Fig. 1: "Round" and "frustrated" triangles. In the first case all links "follow each other" while in the second there is a "reverse" (frustrated) link, the link c. The corner with a circle is called the anchor of the triangle, and the links are then labeled in such a way that for a round triangle the a link leaves the anchor, and the others follow in order, while for the frustrated triangles, the a link leaves the anchor and the b link leaves the end of the a link. These rules determine a unique labeling of each triangle if we require the anchor for the round triangle to be at the node with lowest number.

We next consider the number of triangles in which a given edge can occur. Because of the k-out model, edges of type b can occur in arbitrary many triangles of type F, by just connecting 2 lines from any node to a given edge. In this respect, the k-out model is more complicated than the k-regular model. However, the other lines can occur only in a small number of triangles.

Lemma 3.3. *Bounds on the number triangles per link:*
- *A link can be an edge of type a in at most k triangles of type R and in at most $k-1$ triangles of type F.*
- *A link can be an edge of type b in at most k triangles of type R.*
- *A link can be an edge of type c in at most k triangles of type R and in at most $k-1$ triangles of type F.*

Proof. Consider first the case R. The edge a can occur in at most k triangles. To see this, note that once a is placed, there are k edges of type b leaving its end, and then the triangle must be closed, so there are at most k such triangles. Since R is round, the same reasoning can be done for the other edges. In the case of a triangle of type F, we have already remarked that there is no bound possible for the link b, but we claim the others cannot be part of more than $k-1$ triangles of type F. Indeed, once link a is fixed, we need to choose another out-link to become link c (and then the b link is fixed). This gives $k-1$ as a bound. Finally, link c can belong only to $k-1$ triangles of type F because for fixed c there remain only $k-1$ candidates for the edge of type a. □

An important consequence of Lemma 3.3 is that the number of edges belonging to at least one of the αn triangles grows proportionally with αn:

Lemma 3.4. *The number of edges ℓ_{triang} belonging to at least one of the αn triangles in a graph of type $\mathcal{G}_{n,k-\text{out}}$ is bounded by*

$$\frac{\alpha n}{2k} \leq \ell_{\text{triang}} \leq 3\alpha n \,. \tag{3.4}$$

Proof. The upper bound is obvious. To prove the lower bound, note that every triangle involves a link of type a. Since there are k links leaving from the far end of that link, the number of triangles for which this link is an a link is bounded above by $2k$ (k of type F and k of type R). Thus, at least $\frac{\alpha n}{2k}$ links are needed just to draw all a links. □

Remark 3.5: *Note that an a link can be also a b or c link for many other triangles, and so the above argument cannot be easily improved. When $k=2$ and $n=3$ the complete graph forms 8 triangles, but needs 6 links, instead of the 4 as given by the lower bound. For complete, directed, k-out graphs the asymptotic bound for $k \to \infty$ is $\frac{3\alpha n}{4k}$.*

To prove the bound of Proposition 3.1 we give an algorithm which constructs all the graphs with αn triangles, and perhaps a few more with more triangles, and we bound the number of ways in which this can be done.

To enumerate all the cases, we first label the nodes in an arbitrary fashion from 1 to n. Once this is done, we consider any configuration with αn triangles. We associate each triangle with a node as follows: Triangles of type F are associated with the node from which the a and c links originate. For triangles of type R we label the edges in such a way that the corner where the a and the c edges meet has the lowest label among the 3 corners. We call the point from which the a link leaves the *anchor* of the triangle.

Once this is done, there will be t_i triangles anchored at node i, for $i = 1, \ldots, n$. Furthermore, we denote by v_i the number of links arriving at node i *once the graph will have been completely constructed*. Both t_i and v_i indicate the values *at the end* of constructing the graph.

In order to construct all possible graphs with αn triangles, we start by choosing the t_i and the v_i. Clearly,

$$\sum_{i=1}^{n} t_i = \alpha n , \tag{3.5}$$

and therefore the number of ways of choosing the t_i is bounded by

$$A = \binom{\alpha n + (n-1)}{n-1} \le 2^{(\alpha+1)n} . \tag{3.6}$$

The v_i satisfy $\sum_i v_i = kn$, since each link arrives somewhere. The number of ways to distribute the ends of the kn links is therefore bounded above by

$$B = \binom{kn + (n-1)}{n-1} \le 2^{(k+1)n} . \tag{3.7}$$

We will need the following

Lemma 3.6. *The product of the $v_i + k$ satisfies*

$$\prod_{i=1}^{n} (v_i + k) \le (2k)^n . \tag{3.8}$$

Proof. Since the sum of the $v_i + k$ equals $2kn$, the maximal value of the product is $\bigl(2kn/n\bigr)^n$. □

Thus, once the t_i and v_i are fixed, we have lost a (negligible) combinatorial factor C_0^n, with $C_0 \le 2^{\alpha+k+2}(k+1)$.

Lemma 3.7. *At each node, at most k^2 round triangles and at most $k(k-1)$ frustrated triangles can be anchored.*

Proof. Consider first the round triangles. There are k outgoing links from a given node, and from each of their ends there are another k outgoing links and then the triangle must be closed, and so there are at most k^2 round triangles. For the frustrated triangles, we first choose a pair of outgoing links and then the direction of the link connecting their ends. □

Having fixed the t_i, v_i, we now place the triangles starting with all those anchored at node 1, proceeding to node 2, 3, and so on, until we arrive at node n. At each node i we construct first all the F triangles and then all the R triangles. Assume the first $s-1$ triangles have been drawn, and assume we are placing the next triangle anchored at node i. We will first make a choice of which links of the new triangle are assumed to be present. This gives $8 = 2^3$ choices. There are 2 more choices between type F and R, (in fact less since we insist on building first all the F before the R). For each of these choices, we bound the maximal number of ways a triangle can be placed. We call these bounds F_j for the frustrated triangles and R_j for the round triangles, $j = 1, \ldots, 8$. An upper bound on the number of ways to place a triangle (given its anchor) is then

$$16 \max(F_1, \ldots, F_8, R_1, \ldots, R_8) .$$

When we construct a triangle at i, it will be denoted by its corners (i, j, m). If it is round its links are $a = ij$, $b = jm$, and $c = pq$ with $p = m$, $q = i$. If it is frustrated, its links are $a = ij$, $b = jm$, and $c = pq$ with $p = i$, $q = m$.

Table I

case	new links	# links= δq	R	F	min. gain
1	abc	3	n^2	n^2	n^{-1}
2	ab	2	nv_i	nk	$(v_i + k)n^{-1}$
3	ac	2	nk	nk	kn^{-1}
4	a	1	n	n	$1 \star$
5	bc	2	nk	nk	kn^{-1}
6	b	1	kv_i	k^2	$(v_i + k)^2 n^{-1}$
7	c	1	k^2	k^2	$k^2 n^{-1}$
8	none	0	k^2	k^2	$k^2 \star$

The 16 cases are represented in Table I. The second column indicates which links are new in forming the triangle, and the next the number of these new links. The next two columns indicate the maximum number of ways the given case can appear. The last column will be explained below.

Proof of Table I. To prove the bounds on the multiplicative factors is just a verification. We indicate a few cases to guide the reader. In case 1, we place 3 links of which 2 can be chosen freely (the a link and then the b link), whereas the third link is then completely determined. Therefore, we get a factor $(n-1)(n-2)$, for both types of triangles, and we bound this by n^2 In case 2, the c link is already present. For a round triangle there are at most v_i possibilities for a c link to end in i. The a link can be chosen in n ways, and the b link must connect the end of a to one of the c, and this can be done in v_i ways. Thus we get a factor nv_i. In the case of the frustrated triangles, there are only $k-1$ possibilities for the c link (which now *originates* at i) since one link is used as the a link: The factor is therefore at most nk. All other cases are discussed similarly, for example, in case 7, the c link is missing, but the a and b links are present, and there are k possible ends for a and another k possible ends for each of the b attached to a. Finally, we explain case 4 which is the critical case. In it, the b link and the c link are given. Since one link cannot be placed in more than $n-1$ ways, the factor n is an upper bound. Finally, the last column of Table I is calculated as follows. Its entry is an upper bound on the sup of the R and F column, divided by $n^{\#\text{ links}}$. □

To explain the proof of Proposition 3.1 we will first consider the simpler situation where the cases 4 and 8 do not appear. Indeed, in these two cases, the combinatorial factor of the last column in Table I is *not* small when n is large. In this simplified case, each time we place a triangle, the number of links increases by δq and the number of possibilities is bounded by

$$16n^{\delta q}n^{-1}(v_i+k)^2 \ .$$

Since $1 \le \delta q \le 3$ we can bound this from above by

$$16n^{\delta q(1-1/3)}(v_i+k)^2 \ .$$

From Lemma 3.4, we know limits on ℓ_{triang}, so that $\sum(\delta q) = \ell_{\text{triang}}$, and $\ell_{\text{triang}} \in [\alpha n/(2k), 3\alpha n]$. We get, in the end, for constructing at least αn triangles with ℓ_{triang} links, an upper bound of

$$16^{\alpha n}n^{2\ell_{\text{triang}}/3}\prod_{i=1}^{n}(v_i+k)^{4k^2} \ .$$

The last factor is obtained by observing that there are at most $k^2+k(k-1)$ triangles anchored at a given node by Lemma 3.7. Placing the links not involved in making triangles gives at most a factor n per link, and therefore, using also Eqs.(3.6) and (3.7), we get an upper bound

$$n^{nk}n^{-\ell_{\text{triang}}/3}2^{(\alpha+1)n}16^{\alpha n}2^{(k+1)n}(2k)^{n4k^2} \ .$$

The sum over the possible values of ℓ_{triang} is bounded by $3\alpha n$ times the largest contribution (which occurs for $\ell_{\text{triang}} = \alpha n/(2k)$), and we get a bound

$$3\alpha n \cdot C^n n^{nk} n^{-\alpha n/(6k)} 2^{(\alpha+1)n} \ , \tag{3.9}$$

with

$$C \;=\; 16^{\alpha}2^{(k+1)}(2k)^{4k^2} \ .$$

What about the starred cases? As is visible from Table I, there is *no gain* in the cases 4 and 8. Since we count everything in terms of links, the case 8 is harmless: We have no gain, but we also place no link. Thus the bad case is 4. We will show that case 4 cannot occur too often, and thus a fixed minimal proportion of the cases *will* give a gain.

In the case 4 (for a frustrated triangle) we are in the process of drawing a triangle in which only an a link is missing. In this case, we observe the "history" of the c link. Note that the c link *originates* at i, but its other end has an index m which can be greater or less than i. We distinguish several cases:

F1) $m > i$: Then there are 2 subcases.

- F1a) The link $i \to m$ was placed when a triangle anchored at some node $i' < i$ was formed. Then it must have been placed as a b link.
- F1b) The link $i \to m$ was placed when another triangle anchored at i was formed. Since we begin with the frustrated triangles, this must have been a frustrated triangle. Note, however, that when the *first* frustrated triangle at i is being placed, this case cannot occur and we must begin with the case F1a (or with a case other than case 4).

F2) $m < i$: There are 4 subcases:

- F2a) The link $i \to m$ was a b link when it was placed.
- F2b) The link was placed as a c link of an R triangle anchored at m.
- F2c) The link was placed as a c link of another F triangle anchored at i.
- F2d) The link was placed as an a link of another frustrated triangle anchored at i. Note, however, that when the *first* frustrated triangle at i is being placed, this case cannot occur and we must begin with one of the cases F2a–F2c (or with a case other than case 4).

In the case 4 (for a round triangle) we complete a round triangle with a missing a link. In this case, we observe the "history" of the b link. There is only one possibility:

R) The b link of such a triangle will connect nodes $j \to m$ with both j and m greater than i. Therefore, it can only have been placed as a b link when we constructed a triangle anchored at a node with label $i' \leq i$.

To keep track of the conditions mentioned above, we introduce counters which "distribute" the gain which comes from placing a b or c link onto those further uses of this link in case 4, where no gain is possible. To do the bookkeeping, we introduce for every link ij a counter c_{ij}. Each of these is 0 as we start the inductive procedure to be described below:

$$c_{ij,0} = 0 \,,$$

for all $i \neq j \in \{1, \ldots, n\}$. If a link ij is placed for the first time *and* it is a b link or a c link as it is placed, we set $c_{ij} = 2k - 2$. Each time a link ij is used as a c link in one of the cases F1a or F2a–F2c (*and only in those*) the counter c_{ij} is reduced by one. The maximal number of uses in Case 4 of a link placed originally as a b link is $2k - 2$ (used $k - 1$ times for the c link of an F triangle and another $k - 1$ times for the b link of an R triangle). Similarly, a link placed as a c link can be used in Case 4 another $k - 2$ times as a c link in an F triangle. Since the number of uses is less than $2k - 2$, *none of the counters* c_{ij} *will ever become negative.*

Our last counters keep track of the occurrence of the number of times we are in case F1b or F2d at a given node i. At the beginning of the induction, we set $a_{i,0} = 0$ for all $i \in \{1, \ldots, n\}$. Each time we encounter a case among 1–3, 5–7, F1a, F2a–F2c, or R, at node i, we increase a_i

by k. Each time, we encounter case F1b or F2d, we decrease a_i by 1. Note that since these latter cases cannot occur more than k times, and they can not occur for the *first* triangle at node i, we conclude that *none of the counters a_i ever becomes negative*.

We now prove recursively that at any given step of the construction after adding triangle t, we have a bound on the total number of possibilities which is of the form

$$N_t \leq n^{q_t(1-\beta)}\, n^{-\sigma \sum_{rs} c_{rs,t}/(k-2)} n^{-\varrho \sum_s a_s/k} \prod_{s=1}^{t} \big(16(v_{i_s}+k)^2\big) , \tag{3.10}$$

where q_t is the number of links already drawn, i_s is the number of the node at which the s^{th} triangle is anchored and $\sigma > 0$, $\beta > 0$, and $\varrho > 0$ will be given later on. *If we can show that there is a positive β for which these inequalities hold, then we have shown a bound of the type of Proposition 3.1*, since none of the counters ever becomes negative.

The recursive proof starts when there is no link and all counters are equal to 0, hence the bound is trivially true, ($N_0 = 1$).

We now explain the action at node i. During the construction of the triangles anchored at i, some counters will be updated, and the bound on the combinatorial factor will evolve correspondingly. We now inspect the evolution of the bound during the different possible actions taken at step i. Assume that $t-1$ triangles have been placed, and that we are placing now triangle t which is anchored at $i_t = i$.

Case 1: According to column R or F of Table I, we have

$$N_t \leq N_{t-1} n^2 .$$

But according to the second and third columns of Table I, the number of links increases by 3, thus $q_t = q_{t-1} + 3$. The link jm was empty at time $t-1$ and will be filled at time t. Therefore, $c_{jm,t-1} = 0$ and $c_{jm,t} = 2k-2$, and similarly $c_{[m,i],t-1} = 0$ and $c_{[m,i],t} = 2k-2$. Finally, $a_{i,t} = a_{i,t-1} + k$. The other counters are unchanged. Therefore, we find

$$\begin{aligned} &n^2 \cdot n^{-\sigma c_{jm,t-1}/(2k-2)}\, n^{-\sigma c_{[m,i],t-1}/(2k-2)} n^{-\varrho a_{i,t-1}/k} \\ &\quad \leq n^{3(1-\beta)}\, n^{-\sigma c_{jm,t}/(2k-2)}\, n^{-\sigma c_{[m,i],t}/(2k-2)} n^{-\varrho a_{i,t}/k} \cdot n^{2-3(1-\beta)+2\sigma+\varrho} , \end{aligned} \tag{3.11}$$

where $[m,i] = mi$ or im according to the orientation of the link (R or F case).[1] Since $q_t = q_{t-1} + 3$, we see that N_t satisfies the inductive bound provided the last factor in

[1] Note that

$$-\sigma c_{jm,t-1}/(2k-2) = -\sigma c_{jm,t}/(2k-2) + \sigma ,$$

and also

$$-\sigma c_{[m,i],t-1}/(2k-2) = -\sigma c_{[m,i],t}/(2k-2) + \sigma ,$$

and this exactly compensates the factor $n^{2\sigma}$. Similarly,

$$-\varrho a_{i,t-1}/k = -\varrho a_{i,t}/k + \varrho ,$$

which compensates the factor n^{ϱ} at the end of (3.11).

(3.11) is ≤ 1, which is the case if

$$1 \ \ge \ 3\beta + 2\sigma + \varrho \,. \tag{3.12}$$

Case 2: Two new links (an a and a b) appear and the combinatorial factor is $\max(nv_i, nk) \le n(v_i + k)$. The counter c_{jm} is increased by $2k-2$ and the counter a_i is increased by k. Also, $q_t = q_{t-1} + 2$. The bound analogous to (3.11) is therefore

$$\begin{aligned} &n(v_i + k) \cdot n^{-\sigma c_{jm,t-1}/(2k-2)}\, n^{-\varrho a_{i,t-1}/k} \\ &\qquad \le \ n^{2(1-\beta)}\, n^{-\sigma c_{jm,t}/(2k-2)}\, n^{-\varrho a_{i,t}/k}\, n^{1-2(1-\beta)+\sigma+\varrho}(v_i + k) \,, \end{aligned}$$

which proves the inductive assumption if

$$1 \ \ge \ 2\beta + \sigma + \varrho \,. \tag{3.13}$$

Case 3: The counter $c_{[m,i]}$ is increased by $2k-2$ and a_i is increased by k, and the inductive bound is

$$\begin{aligned} &nk\, n^{-\sigma c_{[m,i],t-1}/(2k-2)}\, n^{-\varrho a_{i,t-1}/k} \\ &\qquad \le \ n^{2(1-\beta)}\, n^{-\sigma c_{[m,i],t}/(2k-2)}\, n^{-\varrho a_{i,t}/k}\, n^{1-2(1-\beta)+\sigma+\varrho}(v_i + k) \,, \end{aligned}$$

which proves the inductive assumption if (3.13) holds.

Case 5: The counters c_{jm} and $c_{[m,i]}$ are increased by $2k-2$, and a_i is increased by k. Therefore,

$$\begin{aligned} &nk \cdot n^{-\sigma c_{jm,t-1}/(2k-2)}\, n^{-\sigma c_{[m,i],t-1}/(2k-2)}\, n^{-\varrho a_{i,t-1}/k} \\ &\le n^{2(1-\beta)}\, n^{-\sigma c_{jm,t}/(2k-2)}\, n^{-\sigma c_{[m,i],t}/(2k-2)}\, n^{-\varrho a_{i,t}/k}\, n^{1-2(1-\beta)+2\sigma+\varrho}(v_i + k) \,, \end{aligned}$$

which proves the inductive assumption if

$$1 \ \ge \ 2\beta + 2\sigma + \varrho \,. \tag{3.14}$$

Case 6: The counter c_{jm} is increased by $2k-2$, and a_i is increased by k. Therefore we get

$$\begin{aligned} &n^0 (v_i + k)^2 \cdot n^{-\sigma c_{[jm],t-1}/(2k-2)}\, n^{-\varrho a_{i,t-1}/k} \\ &\qquad \le \ n^{1-\beta}\, n^{-\sigma c_{[jm],t}/(2k-2)}\, n^{-\varrho a_{i,t}/k}\, n^{-(1-\beta)+\sigma+\varrho}(v_i + k)^2 \,, \end{aligned}$$

which proves the inductive assumption if

$$1 \ \ge \ \beta + \sigma + \varrho \,. \tag{3.15}$$

Case 7: The counter $c_{[m,i]}$ is increased by $2k-2$, and a_i is increased by k. Therefore we get

$$\begin{aligned} &n^0 k^2 \cdot n^{-\sigma c_{[m,i],t-1}/(2k-2)}\, n^{-\varrho a_{i,t-1}/k} \\ &\qquad \le \ n^{1-\beta}\, n^{-\sigma c_{[m,i],t}/(2k-2)}\, n^{-\varrho a_{i,t}/k}\, n^{-(1-\beta)+\sigma+\varrho}(v_i + k)^2 \,, \end{aligned}$$

which proves the inductive assumption if (3.15) holds.

Case 8: This case occurs if we want to draw a triangle anchored at i which has appeared in an earlier phase of the construction (for example, if its sides are all sides of type b from triangles anchored at $i' < i$). In this case, no new link, but 1 new triangle and a factor appear

$$k^2 \le (v_i + k)^2 ,$$

and the inductive assumption evidently holds, since only the number of triangles increases, but not the number of links.

The conditions we require so far on β, σ, and ϱ are all satisfied if

$$1 \ge 3\beta + 2\sigma + \varrho . \tag{3.16}$$

We now come to

Case 4: Whenever one of the subcases F1a, F1b, or F2a–F2c applies, the link $i \to m$ was placed earlier as a b or a c link, and we decrease the corresponding counter c_{im} by one unit, and in the case R the counter c_{mi} is decreased. The counter a_i is unchanged in these cases. Therefore, we find that

$$n \cdot n^{-\sigma c_{[m,i],t-1}/(k-2)} \le n^{1-\beta} n^{-\sigma c_{[m,i],t}/(k-2)}\, n^{\beta-\sigma/(2k-2)} ,$$

and this proves the inductive assumption provided

$$\beta \le \sigma/(2k-2) . \tag{3.17}$$

The remaining cases are F1b and F2d. In these cases the counter a_i becomes useful: It is decreased by 1 and all other counters are unchanged. Therefore, we get

$$n \cdot n^{-\varrho a_{i,t-1}/k} \le n^{1-\beta} n^{-\varrho a_{i,t}/k}\, n^{\beta-\varrho/k} ,$$

and the inductive assumption holds provided

$$\beta \le \varrho/k . \tag{3.18}$$

We have now discussed all cases. It remains to see that the constants β, σ, and ϱ can be chosen consistently. They have to satisfy (3.16), (3.17), and (3.18). We find that the optimal solution is

$$\varrho = k\beta , \quad \sigma = (2k-2)\beta , \quad \text{and} \quad 1 = 3\beta + 2(2k-2)\beta + k\beta .$$

Thus, we find

$$\beta = \frac{1}{2k(5k+1)} ,$$

and the proof of the inductive assumption is complete.

Recall that to draw t triangles we need at least $t/(2k)$ links, and no more than $3t$ links. We therefore conclude (combining the bound on N_t with (3.6) and (3.7)) that the number of graphs with t triangles is bounded by

$$3t \cdot 16^t 2^{t+n} n^{nk-\beta t/(2k)} 2^{(k+1)n} (2k)^{n(2k^2-k)} ,$$

where the last factor follows from (3.10) and the observation that the number of triangles anchored at a node is bounded by $2k^2 - k$ by Lemma 3.7. The proof of Proposition 3.1 is complete. □

4. A lower bound for k-general graphs

Here, we consider the class $\mathcal{G}_{n,k}$ of graphs with n nodes and kn links which can be placed anywhere we please. (This model is close to the well known model $G_{n,p}$ where any of the links is chosen with probability $p = 2k/n$.)

Lemma 4.1. *Fix any $\alpha > 0$. The number of graphs with αn triangles in the class of graphs with n nodes and kn links is (for large enough n) at least*

$$|\mathcal{G}_{n,k,\alpha}| \geq e^{-\mathcal{O}(n^{2/3})} n^{-(\alpha n)^{2/3}} |\mathcal{G}_{n,k}| .$$

Remark 4.2: *This is clearly much larger than the bound of Proposition 3.1, and almost as large as n^{nk}, in fact*

$$\lim_{n\to\infty} \frac{\log |\mathcal{G}_{n,k,\alpha}|}{n \log n} = k .$$

This also proves (1.1), since

$$|\mathcal{G}_{n,k}| = \binom{\binom{n}{2}}{nk} \approx n^{nk} .$$

Proof. Among the n nodes choose $\alpha^{1/3} n^{1/3}$. This can be done in

$$\binom{n}{\alpha^{1/3} n^{1/3}} \geq 1 \tag{4.1}$$

ways. (Here and throughout the proof we do not worry about the integer parts.) With these nodes we build a complete graph, this consumes about $\alpha^{2/3} n^{2/3}$ links and gives αn triangles (and it can be done in 1 way). Now among the remaining $n - \alpha^{1/3} n^{1/3}$ nodes distribute the $kn - \alpha^{2/3} n^{2/3}$ links so that there is no triangle. The number of ways this can be done can be estimated from below using the well known result on the number of graphs without triangles (see [B]), namely this number is bounded below by

$$\mathcal{O}(1) \binom{\binom{n-\alpha^{1/3} n^{1/3}}{2}}{kn - \alpha^{2/3} n^{2/3}} \geq e^{-\mathcal{O}(n^{2/3})} n^{-(\alpha n)^{2/3}} \binom{\binom{n}{2}}{nk} . \tag{4.2}$$

Combining (4.1) and (4.2) we get the lower bound. □

Remark 4.3: *In some sense, this result can be viewed as a complement to the large deviation results of Vu[VV], see also [JLR]. Consider the polynomial associated with triangles in a graph with n nodes:*

$$Y = \sum_{1 \leq i < j < m \leq n} t_{ij} t_{jm} t_{mi} .$$

For this polynomial (in the case of the model $\mathcal{G}'_{n,p}$ where links appear with probability $p = k/n$) one has

$$\mathbf{E}(Y) \;=\; \mathcal{O}(1)\;, \quad \mathbf{E}'(Y) = \mathcal{O}(n^{-1})\;,$$

with $\mathbf{E}$ *the expectation in the random set. Taking Theorem 1.1 in [VV] and choosing $\lambda = \mathcal{O}(\alpha^{1/3}n^{1/2})$ one gets an upper bound of the form*

$$\mathbf{P}(|Y| \geq \alpha n) \;\leq\; e^{-\mathcal{O}(\alpha^{1/3}n^{1/2})}\;.$$

Note that this is consistent with the lower bound of Lemma 4.1.

5. A lower bound for k-out graphs (and for k-regular graphs)

Consider the graphs $\mathcal{G}_{n,k-\text{out}}$ which are of type k-out. Recall that

$$|\mathcal{G}_{n,k-\text{out}}| \;\approx\; n^{nk}\;.$$

We will consider graphs $\mathcal{K}_{k+1}$ which are complete in the sense that there are $k+1$ nodes, and from each node k out-links are leaving (to another node of $\mathcal{K}_{k+1}$). Counting in this case gives $k+1$ nodes, $k(k+1)$ links (one for each direction) and $8\binom{k+1}{3}$ triangles (the factor 8 accounting for the 8 ways to use the 6 links on each triangle, one back and one forth for each pair of nodes).

We now distribute the αn triangles into[2]

$$C_n \;=\; \frac{\alpha n}{8\binom{k+1}{3}}$$

disjoint complete graphs $\mathcal{K}_{k+1}$, and this leaves

$$R_n \;=\; n - (k+1)\frac{\alpha n}{8\binom{k+1}{3}}$$

nodes which will not have been used when making the complete graphs of type $\mathcal{K}_{k+1}$. All links originating from the nodes of the clusters are used up in forming the $\mathcal{K}_{k+1}$. The number of ways to place the C_n complete graphs is

$$\binom{n}{\frac{\alpha n(k+1)}{8\binom{k+1}{3}}}\frac{\left(\frac{\alpha n(k+1)}{8\binom{k+1}{3}}\right)!}{(k+1)!^{\alpha n/(8\binom{k+1}{3})}\left(\frac{\alpha n}{8\binom{k+1}{3}}\right)!} \;\approx\; n^{Q_n}\;, \tag{5.1}$$

with

$$Q_n \;=\; \frac{\alpha n(k+1)}{8\binom{k+1}{3}} - \frac{\alpha n}{8\binom{k+1}{3}}\;.$$

[2] To simplify the discussion, which is in terms of orders of magnitude, we assume that all quotients are integers.

The first factor in (5.1) counts the number of ways to choose the nodes involved, and the quotient counts the number of ways the $(k+1)C_n$ nodes are grouped into clusters of $k+1$ each. Since the graphs use k links per node, the graph we can construct with the remaining R_n nodes will be disjoint from the C_n clusters, and we want to bound the number of ways in connecting the remaining nodes *without* adding any triangles. A lower bound on the number of such graphs is obtained by constructing again a k-out bi-partite graph on the remaining R_n nodes.

The number of ways to place the remaining links is therefore bounded below by

$$\binom{R_n/2}{k}^{R_n} \approx n^{S_n} ,$$

with

$$S_n = k\left(n - \frac{\alpha n(k+1)}{8\binom{k+1}{3}}\right) .$$

Note that we do not insist on making a connected graph. So we find a lower bound of $E^n n^{T_n}$, where E depends only on k and

$$\begin{aligned} T_n &= Q_n + S_n = n\left(k - \frac{6\alpha}{8(k+1)k(k-1)}\big(k(k+1)-(k+1)+1\big)\right) \\ &= n\left(k - \frac{3\alpha}{4k}(1+\frac{1}{k^2-1})\right) = nk\left(1-\frac{3\alpha}{4(k^2-1)}\right) . \end{aligned} \tag{5.2}$$

Remark 5.1: *The above calculation proves the lower bound for (1.2). The lower bound for (1.3) is an easy variant, observing the fact that instead of 8 triangles in a complete graph on 3 nodes in the k-out model there is only 1 in the k-regular model.*

Acknowledgments. We thank E. Moses for helpful discussions and encouragements. This work was partially supported by the Fonds National Suisse.

References

[AB] R. Albert and A.-L. Barabási: Statistical mechanics of complex networks. Rev. Mod. Phys. **47**, 47–97 (2002).

[B] B. Bollobas: *Random Graphs*, New York, Academic Press (1985).

[EM] J.-P. Eckmann and E. Moses: Curvature of Co-Links Uncovers Hidden Thematic Layers in the World Wide Web. PNAS (in print).

[JLR] J. Janson, T. Łuczak, and A. Ruciński: An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In *Random Graphs 87*, (M. Karoński *et al.* eds.). Wiley, New York, 73–87 (1990).

[VV] H. Vu: On the concentration of multi-variate polynomials with small expectation. Random Structures and Algorithms **16**, 344–363 (2000).