Block Statistics in Subcritical Graph Classes

Dimbinaina Ralaivaosaona
Stellenbosch University, South Africa
naina@sun.ac.za

Clément Requilé
Technische Universität Wien, Austria
clement.requile@tuwien.ac.at

Stephan Wagner
Uppsala University, Sweden
Stellenbosch University, South Africa
stephan.wagner@math.uu.se

Abstract

We study block statistics in subcritical graph classes; these are statistics that can be defined as the sum of a certain weight function over all blocks. Examples include the number of edges, the number of blocks, and the logarithm of the number of spanning trees. The main result of this paper is a central limit theorem for statistics of this kind under fairly mild technical assumptions.

2012 ACM Subject Classification Mathematics of computing → Random graphs; Mathematics of computing → Generating functions

Keywords and phrases subcritical graph class, block statistic, number of blocks, number of edges, number of spanning trees

Digital Object Identifier 10.4230/LIPIcs.AofA.2020.24

Funding *Dimbinaina Ralaivaosaona*: Supported by Division for Research Development (DRD) of Stellenbosch University and the National Research Foundation (NRF) of South Africa.
Clément Requilé: Supported by the Special Research Program F50 *Algorithmic and Enumerative Combinatorics* of the Austrian Science Fund.
Stephan Wagner: Supported by the Knut and Alice Wallenberg Foundation.

1 Introduction

The detailed analytic study of *subcritical graph classes* was initiated by Drmota et al. in their seminal paper [4]; the formal definition, which will be given below, is based on properties of the generating function. Intuitively speaking, subcritical classes are “tree-like” in some sense, which is exhibited for instance by the fact that their scaling limit is the continuum random tree [12], meaning that the global structure is essentially determined by the block-cutpoint tree, while the blocks themselves are fairly small. Typical examples of subcritical graph classes are trees, cacti, block graphs, outerplanar graphs and series-parallel graphs. Unfortunately, there is probably no simple graph-theoretical characterisation of subcritical graph classes, as it was shown that every proper minor-closed family of graphs is contained in a subcritical family [9].

By a block statistic, we mean an invariant induced by a weight function w on all 2-connected graphs (blocks) of the specific graph class. Any graph G can be decomposed uniquely into maximal 2-connected subgraphs $B_1, B_2, \dots, B_k$ (that can only be joined at cutvertices), the so-called blocks of G. Using this decomposition, we define the block statistic S_w associated with w by

$$S_w(G) = \sum_{j=1}^{k} w(B_j).$$

31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020).
Editors: Michael Drmota and Clemens Heuberger; Article No. 24; pp. 24:1–24:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Let us give a few motivating examples of block statistics:

- The trivial weight function defined by $w(B) = 1$ for all possible blocks B yields the number of blocks.
- If we fix some block A and define

  $$w_A(B) = \begin{cases} 1 & B \simeq A, \\ 0 & \text{otherwise,} \end{cases}$$

  the associated block statistic is the number of (isomorphic) occurrences of A as a block.
- If the weight function $w(B)$ is the number of edges in B, then the associated block statistic $S_w(G)$ gives the number of edges of G.
- Let $\tau(B)$ be the number of spanning trees of a block B, and set $w(B) = \log \tau(B)$. Since every spanning tree of a connected graph decomposes uniquely into spanning trees on all the blocks, we have

  $$\tau(G) = \prod_{j=1}^{k} \tau(B_j)$$

  if $B_1, B_2, \dots, B_k$ are the blocks of a connected graph G. This translates to

  $$S_w(G) = \log \tau(G).$$

- Since the Tutte polynomial is also multiplicative over blocks, the previous example generalises to many others that are special values of the Tutte polynomial, specifically the (logarithm of the) number of subforests, spanning forests, connected spanning subgraphs, acyclic orientations and strongly connected orientations.
- The number of nontrivial complete subgraphs (i.e., complete subgraphs with more than one vertex) is also a block statistic in our sense, since every such subgraph needs to be contained entirely in one of the blocks.
- The number of occurrences of a fixed graph H as an induced subgraph, which was studied in [6], is not always a block statistic (since a copy of H may involve vertices of several blocks), but it becomes one if H is 2-connected.

The number of blocks and the number of edges were already shown by Drmota et al. [4] to satisfy a central limit theorem, the latter under the assumption (that was satisfied for all the examples studied in their paper) that the graphs are planar, so that the number of edges is necessarily linear in the number of vertices. However, this is not satisfied for all subcritical classes of graphs (block graphs, for example, are an exception), and there are also other statistics among the aforementioned for which the weight function can grow faster than linearly in the block size, for example the logarithm of the number of spanning trees, for which the weight can be as large as $w(B) = (|B| - 2) \log |B|$ when B is a complete graph. We are therefore interested in proving central limit theorems under weaker assumptions on the growth of the weights.

Before we formulate our main results, let us recall the formal definition of a subcritical graph class. For simplicity, we will restrict ourselves to the labelled case.

▶ **Definition 1.** *We call a class of graphs $\mathcal{G}$* block-stable *if it has the property that a graph G belongs to $\mathcal{G}$ if and only if each of its blocks belongs to $\mathcal{G}$. Now let $\mathcal{G}$ be a block-stable class of labelled graphs, and denote the subclasses of connected graphs and 2-connected graphs in $\mathcal{G}$ by $\mathcal{C}$ and $\mathcal{B}$ respectively. Since every graph can be seen as the union of its connected components, we have the symbolic decomposition*

$$\mathcal{G} = \mathrm{Set}(\mathcal{C}).$$

More importantly (for the definition of subcriticality), rooted connected graphs (indicated by $\mathcal{C}^\bullet$) can be decomposed as follows:

$$\mathcal{C}^\bullet = \mathcal{Z} \times \mathrm{Set}(\mathcal{B}' \circ \mathcal{C}^\bullet),$$

where $\mathcal{Z}$ stands for a single vertex, and $\mathcal{B}'$ for the class derived from $\mathcal{B}$ by not labelling one of the vertices. In words: a rooted connected graph decomposes into the root, the set of blocks that contain the root, and rooted connected graphs attached to all non-root vertices of the root blocks.

On the level of generating functions $G(z)$, $C(z)$ and $B(z)$ are associated with $\mathcal{G}$, $\mathcal{C}$ and $\mathcal{B}$ respectively, this yields

$$G(z) = \exp(C(z)) \tag{1}$$

and

$$C^\bullet(z) = z\exp(B'(C^\bullet(z)), \tag{2}$$

where $C^\bullet(z) = zC'(z)$ is the generating function for $\mathcal{C}^\bullet$. The class $\mathcal{G}$ is now said to be subcritical *if the radii of convergence ρ and η of C and B satisfy the inequality*

$$\gamma = C^\bullet(\rho) < \eta. \tag{3}$$

As it was shown in [4], the generating function $C^\bullet$ has a square root singularity for every subcritical class, which allows us to apply singularity analysis to derive asymptotic formulas for counting graphs of given order in $\mathcal{G}$ or $\mathcal{C}$.

▶ **Theorem 2** ([4]). *For every subcritical family of graphs, the generating function $C^\bullet$ is analytic in a region of the form*

$$\{z \in \mathbb{C} \,:\, |z| < r, |\operatorname{Arg}(z-\rho)| > \phi\}$$

for some $r > \rho$ and $\phi \in (0, \frac{\pi}{2})$. At the singularity ρ, it has an asymptotic expansion of the form

$$C^\bullet(z) = \gamma + \gamma_1(1-z/\rho)^{1/2} + \gamma_2(1-z/\rho) + \gamma_3(1-z/\rho)^{3/2} + O((1-z/\rho)^2). \tag{4}$$

Here, $\gamma > 0$ is the unique positive solution of the equation $\gamma B''(\gamma) = 1$, and $\rho = \gamma\exp(-B'(\gamma))$.

It is sometimes useful to have explicit expressions for γ_1 and γ_2. They are given by

$$\gamma_1 = -\sqrt{\frac{2\gamma^2}{1+\gamma^2 B'''(\gamma)}} \quad \text{and} \quad \gamma_2 = \frac{2\gamma - \gamma^4 B''''(\gamma)}{3(1+\gamma^2 B'''(\gamma))^2},$$

respectively, as one can see e.g. by comparing coefficients on the two sides of the functional equation (and using the identity $\gamma B''(\gamma) = 1$ to simplify).

2 The generating function for a block statistic

The functional equations (1) and (2) can be modified in a straightforward fashion to include the block statistic S_w. Let us define the bivariate function

$$C(z,t) = \sum_{C\in\mathcal{C}} \frac{z^{|C|}}{|C|!} e^{S_w(C)t},$$

and $G(z,t)$ in an analogous fashion. To keep notation simple, we do not indicate the dependence of the generating functions on w. Since S_w is additive over connected components, we clearly have

$$G(z,t) = \exp(C(z,t)).$$

Moreover, if we set

$$B(z,t) = \sum_{B \in \mathcal{B}} \frac{z^{|B|}}{|B|!} e^{w(B)t},$$

then (2) changes to

$$C^\bullet(z,t) = z \exp\big(B_z(C^\bullet(z,t),t)\big), \tag{5}$$

where B_z is the partial derivative with respect to z. Of course, when $t = 0$, everything simplifies to (1) and (2). It is important to notice that the sum defining B might not be convergent: if $w(B)$ has faster than linear growth for at least some blocks B, then the radius of convergence in z can become zero for all $t > 0$. Therefore, B and C are a priori only regarded as formal power series.

If, however, $w(B) = O(|B|)$, then the radius of convergence of B as a function of t changes continuously, so if t is close enough to zero, the inequality that defines a subcritical class remains true, and $C^\bullet$ still has a square root singularity, the position of which moves continuously with t. We are therefore in the scheme of [3, Theorems 2.21–2.23], and the quasi-power theorem ([10], see also [8, Section IX.5]) yields a central limit theorem almost automatically. This was in fact exploited in [4] to obtain the central limit theorems for number of edges and number of blocks.

However, not all interesting block statistics satisfy the condition $w(B) = O(|B|)$. The example of the logarithm of the number of spanning trees was mentioned earlier; others include the logarithm of the number of subforests, spanning forests or connected spanning subgraphs and the number of nontrivial complete subgraphs. Thus, we follow a slightly different route imposing somewhat milder conditions on the weight function w. Specifically, we will prove the following theorem:

▶ **Theorem 3.** *Consider a subcritical class of graphs with a weight function w on the blocks. Let W_n be the average of $w(B)^2$ over all blocks B on n vertices. Suppose that*

$$\limsup_{n\to\infty} W_n^{1/n} < \frac{\eta}{\gamma}, \tag{6}$$

with γ and η as in (3). Let C_n denote a random connected graph with n vertices in our subcritical class of graphs. The following statements on the distribution of $S_w(C_n)$ hold:

1. *There exist constants μ and λ such that the mean $\mathbb{E}(S_w(C_n))$ is asymptotically equal to $\mu n - \lambda + O(n^{-1})$.*
2. *There exists a constant $\sigma^2 \geq 0$ such that the variance $\mathbb{V}(S_w(C_n))$ is asymptotically equal to $\sigma^2 n + O(1)$. Moreover, we have $\sigma^2 > 0$ unless the weight function w is of the form $w(B) = c(|B| - 1)$ for some constant c.*
3. *If $\sigma^2 > 0$, then the distribution of $S_w(C_n)$ converges, suitably normalised by subtracting the mean and dividing by the standard deviation, weakly to a standard normal distribution.*

Intuitively, (6) states that the block generating function, with blocks weighted by $w(B)^2$, still satisfies the subcriticality condition. Most of the examples mentioned in the introduction satisfy the conditions of the theorem for all subcritical classes, since the growth of the weight function w is subexponential. Notable examples include the number of blocks, the number of edges and the (logarithm of the) number of spanning trees. It is possible that the condition is satisfied even if the weight grows exponentially in the block size, though. Importantly, blocks in random graphs from a subcritical class are typically small (the largest block only being logarithmic in size). This makes it possible that $\mathbb{E}(S_w(C_n))$ is linear in n even in cases where w can grow exponentially.

▶ Remark 4. While we are focusing on connected graphs in this paper, it would also be possible to transfer our results to arbitrary random graphs from the specific subcritical class of graphs.

▶ Remark 5. We remark that $S_w(C) = c(|C| - 1)$ holds deterministically for all connected graphs C in the "degenerate" case that $w(B) = c(|B| - 1)$, so that the variance is identically 0.

Several explicit examples are presented in detail in the appendix. The following table gives an overview:

Graph class	Block statistic	μ	σ^2	Sect.
Cacti	Number of blocks	0.64780	0.21218	A.1
	Number of edges	1.19149	0.06272	A.1
	Number of spanning trees (log)	0.24985	0.08007	A.1
	Number of connected spanning subgraphs (log)	0.29690	0.12113	A.1
Block graphs	Number of blocks	0.76322	0.12512	A.2
	Number of edges	1.28357	0.31267	A.2
	Number of spanning trees (log)	0.28580	0.23671	A.2
	Number of nontrivial complete subgraphs	1.69146	4.55177	A.2
Series-parallel graphs	Number of blocks	0.14937	0.14875	A.3
	Number of edges	1.61673	0.21125	A.3
	Number of spanning trees (log)	??	??	A.3

In the last example, numerical values of μ and σ^2 are surprisingly difficult to determine. This will be explained in Section A.3.

3 Mean and variance

It is somewhat easier for our calculations to consider rooted graphs. However, since every labelled graph with n vertices corresponds to precisely n rooted graphs, all distributional results that we obtain for rooted graphs in the following hold automatically for unrooted graphs as well.

In order to obtain asymptotic formulas for mean and variance, we consider the partial derivatives of $C^\bullet(z,t)$ with respect to t at $t = 0$. Differentiating (5) with respect to t yields

$$\begin{aligned} C_t^\bullet(z,t) &= z\exp\big(B_z(C^\bullet(z,t),t)\big)\Big(B_{zz}(C^\bullet(z,t),t)C_t^\bullet(z,t) + B_{zt}(C^\bullet(z,t),t)\Big) \\ &= C^\bullet(z,t)\Big(B_{zz}(C^\bullet(z,t),t)C_t^\bullet(z,t) + B_{zt}(C^\bullet(z,t),t)\Big). \end{aligned}$$

We solve for $C_t^\bullet(z,t)$, which gives us

$$C_t^\bullet(z,t) = \frac{C^\bullet(z,t)B_{zt}(C^\bullet(z,t),t)}{1 - C^\bullet(z,t)B_{zz}(C^\bullet(z,t),t)}.$$

In the same way, we can also differentiate with respect to z, which yields

$$C^\bullet_z(z,t) = \frac{C^\bullet(z,t)}{z(1 - C^\bullet(z,t)B_{zz}(C^\bullet(z,t),t))}.$$

Thus we have

$$C^\bullet_t(z,t) = zC^\bullet_z(z,t)B_{zt}(C^\bullet(z,t),t). \tag{7}$$

The second derivative is determined in a similar fashion. Differentiating (7) with respect to z and t respectively (and plugging in $t = 0$), we obtain

$$C^\bullet_{zt}(z,0) = C^\bullet_z(z,0)B_{zt}(C^\bullet(z,0),0)+zC^\bullet_{zz}(z,0)B_{zt}(C^\bullet(z,0),0)+zC^\bullet_z(z,0)^2B_{zzt}(C^\bullet(z,0),0)$$

and

$$\begin{aligned}C^\bullet_{tt}(z,0) &= zC^\bullet_{zt}(z,0)B_{zt}(C^\bullet(z,0),0) + zC^\bullet_z(z,0)C^\bullet_t(z,0)B_{zzt}(C^\bullet(z,0),0)\\ &\quad + zC^\bullet_z(z,0)B_{ztt}(C^\bullet(z,0),0).\end{aligned}$$

We plug the former equation into the latter, and also replace $C^\bullet_t(z,0)$ by the equation given in (7) to arrive at the following representation for $C^\bullet_{tt}(z,0)$:

$$\begin{aligned}C^\bullet_{tt}(z,0) &= z^2C^\bullet_{zz}(z,0)B_{zt}(C^\bullet(z,0),0)^2 + 2z^2C^\bullet_z(z,0)^2B_{zt}(C^\bullet(z,0),0)B_{zzt}(C^\bullet(z,0),0)\\ &\quad + zC^\bullet_z(z,0)\big(B_{zt}(C^\bullet(z,0),0)^2 + B_{ztt}(C^\bullet(z,0),0)\big).\end{aligned} \tag{8}$$

Note that this representation, like (7), only involves derivatives of $C^\bullet(z,t)$ with respect to z, so that we can use our knowledge of the behaviour of $C^\bullet(z,0)$ given in Theorem 2.

▶ **Theorem 6.** *Under the conditions stated in Theorem 3, the mean of the block statistic S_w over all graphs in $\mathcal{C}$ with n vertices is asymptotically*

$$\mathbb{E}(S_w(C_n)) = \mu n - \lambda + O(n^{-1}),$$

with $\mu = B_{zt}(\gamma,0)$ and

$$\lambda = \frac{3\gamma_2}{2}B_{zzt}(\gamma,0) + \frac{\gamma_1^2}{4}B_{zzzt}(\gamma,0).$$

Proof. Note that

$$B_{tt}(z,0) = \sum_{n\geq 2}\Big(\sum_{\substack{B\in\mathcal{B}\\ |B|=n}} w(B)^2\Big)\frac{z^n}{n!} = \sum_{n\geq 2} W_nB_n\frac{z^n}{n!},$$

where B_n is the number of blocks with n labelled vertices. The radius of convergence of $B(z,0)$ is $\eta = 1/\limsup_{n\to\infty}(B_n/n!)^{1/n}$, since the coefficient of z^n in $B(z,0)$ is $B_n/n!$. The technical condition (6) has been chosen in such a way that the radius of convergence of $B_{tt}(z,0)$, which is

$$\begin{aligned}\frac{1}{\limsup_{n\to\infty}(W_nB_n/n!)^{1/n}} &> \frac{1}{\limsup_{n\to\infty} W_n^{1/n}}\cdot\frac{1}{\limsup_{n\to\infty}(B_n/n!)^{1/n}}\\ &= \frac{\eta}{\limsup_{n\to\infty} W_n^{1/n}},\end{aligned}$$

is greater than γ. The radius of convergence of $B(z,0)$ is $\eta > \gamma$ by the definition of subcriticality, and the Cauchy-Schwarz inequality implies that

$$|B_t(z,0)|^2 = \Big| \sum_{B\in\mathcal{B}} \frac{z^{|B|}w(B)}{|B|!} \Big|^2 \le \sum_{B\in\mathcal{B}} \frac{|z|^{|B|}w(B)^2}{|B|!} \sum_{B\in\mathcal{B}} \frac{|z|^{|B|}}{|B|!} = B_{tt}(|z|,0)B(|z|,0).$$

This shows that $B_t(z,0)$ also has greater radius of convergence than γ. Thus $B_t(z,0)$, $B_{tt}(z,0)$ and all their derivatives with respect to z are analytic in a disk around 0 that includes γ. Now since $C^\bullet$ is amenable to singularity analysis, so is

$$C^\bullet_t(z,0) = zC^\bullet_z(z,0)B_{zt}(C^\bullet(z,0),0).$$

In particular, this function and other partial derivatives of $C^\bullet$ that we consider have a Puiseux expansion around the singularity ρ whose exponents are integers or half-integers. Specifically, in view of (4), we have

$$\begin{aligned} B_{zt}(C^\bullet(z,0),0) &= B_{zt}(\gamma,0) + B_{zzt}(\gamma,0)(C^\bullet(z,0)-\gamma) + \frac{B_{zzzt}(\gamma,0)}{2}(C^\bullet(z,0)-\gamma)^2 \\ &\quad + O\big((C^\bullet(z,0)-\gamma)^3\big) \\ &= B_{zt}(\gamma,0) + \gamma_1 B_{zzt}(\gamma,0)(1-z/\rho)^{1/2} \\ &\quad + \Big(\gamma_2 B_{zzt}(\gamma,0) + \frac{\gamma_1^2 B_{zzzt}(\gamma,0)}{2}\Big)(1-z/\rho) + O\big((1-z/\rho)^{3/2}\big). \end{aligned}$$

Thus we can represent $C^\bullet_t(z,0)$ as follows:

$$C^\bullet_t(z,0) = B_{zt}(\gamma,0)zC^\bullet_z(z,0) + \kappa_1 - \Big(\frac{3\gamma_2 B_{zzt}(\gamma,0)}{2} + \frac{\gamma_1^2 B_{zzzt}(\gamma,0)}{4}\Big)C^\bullet(z,0) + O\big(1-z/\rho\big)$$

for some constant κ_1. It would be possible to add further terms to the expansion. By the principles of singularity analysis, we obtain

$$[z^n]C^\bullet_t(z,0) = \mu[z^n]zC^\bullet_z(z,0) - \lambda[z^n]C^\bullet(z,0) + O(n^{-1}[z^n]C^\bullet(z,0))$$

with μ and λ as given in the statement of the theorem. Therefore,

$$\mathbb{E}(S_w(C_n)) = \frac{[z^n]C^\bullet_t(z,0)}{[z^n]C^\bullet(z,0)} = \mu n - \lambda + O(n^{-1}).$$ ◀

▶ **Theorem 7.** *Under the conditions stated in Theorem 3, the variance of the block statistic S_w over all graphs in $\mathcal{C}$ with n vertices is asymptotically $\mathbb{V}(S_w(C_n)) = \sigma^2 n + O(1)$, with*

$$\sigma^2 = B_{ztt}(\gamma,0) - \frac{\gamma^2 B_{zzt}(\gamma,0)^2}{1+\gamma^2 B_{zzz}(\gamma,0)}.$$

If the weight w is not of the form $w(B) = c(|B|-1)$ (where c is constant), then σ^2 is strictly positive.

Proof. The asymptotic formula for the variance is derived in a similar fashion as the mean. We now need to consider the second derivative with respect to t as well. The expression for $C^\bullet_{tt}(z,0)$ in (8) can be expanded around the dominant singularity ρ in the same way as $C^\bullet_t(z,0)$. Without going through the full calculation, let us just give the final result stating that

$$C^\bullet_{tt}(z,0) = \mu^2\big(z^2C^\bullet_{zz}(z,0) + zC^\bullet_z(z,0)\big) + (\sigma^2 - 2\lambda\mu)zC^\bullet_z(z,0) + O(1)$$

around the singularity, with μ, λ, σ^2 as defined above. Again, it would be possible to improve on the error term by including further terms. Now we get

$$[z^n]C^\bullet_{tt}(z,0) = \big(\mu^2 n^2 + (\sigma^2 - 2\lambda\mu)n + O(1)\big)[z^n]C^\bullet(z,0).$$

This gives us the second moment of $S_w(C_n)$ as $\mu^2 n^2 + (\sigma^2 - 2\lambda\mu)n + O(1)$, and subtracting the square of the mean yields the stated asymptotic formula for the variance.

It remains to prove that $\sigma^2 \neq 0$ except for trivial cases where $S_w(C)$ depends on the number of vertices of C only. To this end, recall first that γ is determined by the equation

$$\gamma B''(\gamma) = \gamma B_{zz}(\gamma, 0) = 1.$$

Thus we can rewrite the denominator in the expression for σ^2 as follows:

$$1 + \gamma^2 B_{zzz}(\gamma, 0) = \gamma B_{zz}(\gamma, 0) + \gamma^2 B_{zzz}(\gamma, 0) = \sum_{B \in \mathcal{B}} \frac{(|B|-1)^2 \gamma^{|B|-1}}{(|B|-1)!}.$$

The remaining two terms in the expression are

$$B_{ztt}(\gamma, 0) = \sum_{B \in \mathcal{B}} \frac{w(B)^2 \gamma^{|B|-1}}{(|B|-1)!}$$

and

$$\gamma^2 B_{zzt}(\gamma, 0)^2 = \Big(\sum_{B \in \mathcal{B}} \frac{w(B)(|B|-1)\gamma^{|B|-1}}{(|B|-1)!} \Big)^2,$$

respectively. Thus

$$\sigma^2 = \sum_{B \in \mathcal{B}} \frac{w(B)^2 \gamma^{|B|-1}}{(|B|-1)!} - \frac{\Big(\sum_{B \in \mathcal{B}} \frac{w(B)(|B|-1)\gamma^{|B|-1}}{(|B|-1)!} \Big)^2}{\sum_{B \in \mathcal{B}} \frac{(|B|-1)^2 \gamma^{|B|-1}}{(|B|-1)!}}. \tag{9}$$

The Cauchy-Schwarz inequality immediately shows that $\sigma^2 > 0$ unless $w(B)$ is a constant multiple of $|B| - 1$, in which case σ^2 is clearly 0. ◀

In order to illustrate the formulas for mean and variance, let us consider a concrete example that satisfies the conditions of Theorem 3 for all subcritical graph classes: the number of blocks. In this case, we have $B(z,t) = e^t B(z)$, which allows us to express μ and σ^2 in terms of B and γ only: the following formulas can also already been found in [4].

$$\mu = B'(\gamma) \quad \text{and} \quad \sigma^2 = B'(\gamma) - \frac{1}{1 + \gamma^2 B'''(\gamma)}. \tag{10}$$

4 Limit distribution

Next we derive a general central limit theorem for the block statistic S_w. As a first step, we consider the case where w is finitely supported, i.e., where $w(B) = 0$ for all but finitely many blocks. In this case, $B(z,t)$ differs from $B(z)$ only in finitely many terms: letting $\mathcal{B}_0$ be the set of blocks for which $w(B) = 0$, we have

$$B(z,t) = \sum_{B \in \mathcal{B}} \frac{z^{|B|}}{|B|!} e^{w(B)t} = \sum_{B \in \mathcal{B} \setminus \mathcal{B}_0} \frac{z^{|B|}}{|B|!} e^{w(B)t} + \sum_{B \in \mathcal{B}_0} \frac{z^{|B|}}{|B|!} = B(z) + \sum_{B \in \mathcal{B} \setminus \mathcal{B}_0} \frac{z^{|B|}}{|B|!} (e^{w(B)t} - 1).$$

The sum over $\mathcal{B} \setminus \mathcal{B}_0$ is finite and thus represents a function that is entire in both z and t. We are therefore in a position to apply a general result on perturbations of functional equations: by [3, Theorem 2.21], there exists a positive constant $\delta > 0$ such that $C^\bullet(z,t)$ still has a dominant square root singularity for $|t| < \delta$:

$$C^\bullet(z) = \gamma(t) + \gamma_1(t)(1 - z/\rho(t))^{1/2} + O(1 - z/\rho(t)),$$

where $\rho(t)$ is analytic as a function of t for $|t| < \delta$. Singularity analysis gives us an asymptotic formula for the moment generating function of $S_w(C_n)$:

$$\mathbb{E}\left(e^{tS_w(C_n)}\right) = \frac{[z^n]C^\bullet(z,t)}{[z^n]C^\bullet(z,0)} = \frac{\gamma_1(t)}{\gamma_1}\Big(\frac{\rho}{\rho(t)}\Big)^n\big(1 + O(n^{-1})\big).$$

Thus we can apply the quasi-power theorem ([10], [8, Section IX.5]), which proves that $S_w(C_n)$ satisfies a central limit theorem in the case that w has finite support:

$$\frac{S_w(C_n) - \mathbb{E}(S_w(C_n))}{\sqrt{\mathbb{V}(S_w(C_n))}} \xrightarrow{\mathrm{d}} N(0,1).$$

This line of reasoning does not apply if w grows too fast; with the weaker assumptions of Theorem 3, $C^\bullet(z,t)$ may no longer have a square root singularity for $t > 0$ (and might in fact have radius of convergence 0 as a power series in z). Therefore, we rather approximate S_w by considering truncated versions of the weight function w.

For a positive integer M, set

$$w^{(M)}(B) = \begin{cases} w(B) & \text{if } |B| \le M, \\ 0 & \text{otherwise.} \end{cases}$$

We observe that $w^{(M)}$ has finite support, so the block statistic $S_w^{(M)}$ associated with $w^{(M)}$ satisfies a central limit theorem as stated above. Clearly, every block statistic with finitely supported weight function satisfies the conditions of Theorem 3, thus in particular the statements on mean and variance in Theorem 6 and Theorem 7 apply:

- $\mathbb{E}(S_w^{(M)}(C_n)) = \mu_M n + O(1)$,
- $\mathbb{V}(S_w^{(M)}(C_n)) = \sigma_M^2 n + O(1)$,
- $\frac{S_w^{(M)}(C_n) - \mu_M n}{\sigma_M \sqrt{n}} \xrightarrow{\mathrm{d}} N(0,1)$, or equivalently $\frac{S_w^{(M)}(C_n) - \mathbb{E}(S_w^{(M)}(C_n))}{\sqrt{n}} \xrightarrow{\mathrm{d}} N(0,\sigma_M^2)$.

We can now apply the following lemma (see for instance [11, Theorem 4.28]):

▶ **Lemma 8.** *Let $(X_n)_{n\ge1}$ and $(W_{N,n})_{N,n\ge1}$ be sequences of random variables with mean 0. Assume that for some random variables W_N $(N \ge 1)$ and W, we have*

- *$W_{N,n} \xrightarrow{\mathrm{d}} W_N$ as $n \to \infty$ for every $N \ge 1$, and $W_N \xrightarrow{\mathrm{d}} W$ as $N \to \infty$.*
- *$\mathbb{V}(X_n - W_{N,n}) \le C_N$ for some constants C_N uniformly in n, and $C_N \to 0$ as $N \to \infty$.*

Then we also have $X_n \xrightarrow{\mathrm{d}} W$ as $n \to \infty$.

In our setting, we take

$$X_n = \frac{S_w(C_n) - \mathbb{E}(S_w(C_n))}{\sqrt{n}} \quad \text{and} \quad W_{N,n} = \frac{S_w^{(N)}(C_n) - \mathbb{E}(S_w^{(N)}(C_n))}{\sqrt{n}}.$$

Note that these random variables all have mean 0. Since the sums in the formula (9) for σ^2 converge by our assumptions on the weight function w, the constants σ_N^2 converge: $\lim_{N\to\infty} \sigma_N^2 = \sigma^2$. So we have $W_{N,n} \xrightarrow{\mathrm{d}} W_N = N(0,\sigma_N^2)$ as $n \to \infty$ for every N, and $W_N \xrightarrow{\mathrm{d}} W = N(0,\sigma^2)$ as $N \to \infty$.

Lastly, the conditions of Theorem 3 also apply to $S_w(C_n) - S_w^{(N)}(C_n)$, which is the block statistic associated with the weight function that is given by

$$w(B) - w^{(N)}(B) = \begin{cases} 0 & \text{if } |B| \le N, \\ w(B) & \text{otherwise.} \end{cases}$$

Thus the formula of Theorem 7 applies, which yields

$$\begin{aligned} \mathbb{V}(X_n - W_{N,n}) &= \mathbb{V}\Big(\frac{S_w(C_n) - S_w^{(N)}(C_n) - \mathbb{E}(S_w(C_n) - S_w^{(N)}(C_n))}{\sqrt{n}}\Big) \\ &= \frac{1}{n}\mathbb{V}\big(S_w(C_n) - S_w^{(N)}(C_n)\big) = \tau_N^2 + O(n^{-1}) \end{aligned}$$

for some constants τ_N^2 that satisfy $\lim_{N\to\infty} \tau_N^2 = 0$. One can verify that the O-constant can be chosen to depend only on the graph class and the weight function w, but not on N. Moreover, we clearly have $X_n = W_{N,n}$ and thus $\mathbb{V}(X_n - W_{N,n}) = 0$ for $n \le N$. Thus all conditions of Lemma 8 are satisfied, and we obtain the desired central limit theorem for $S_w(C_n)$:

$$\frac{S_w(C_n) - \mathbb{E}(S_w(C_n))}{\sqrt{n}} \xrightarrow{\mathrm{d}} N(0, \sigma^2)$$

as $n \to \infty$. This finally completes the proof of Theorem 3.

5 Conclusion

We obtained a central limit theorem for block statistics under rather mild conditions that cover many natural cases. It would be interesting to see if there are natural examples where the conditions fail and there is no central limit theorem. There are also many examples of statistics that are not block statistics, but of a similar nature, for example the number of (arbitrary, maximal or maximum) independent sets or matchings, see [5] for some examples. One would still expect a log-normal limit law to hold in these cases, akin to spanning trees.

References

1. Manuel Bodirsky, Omer Giménez, Mihyun Kang, and Marc Noy. Enumeration and limit laws for series-parallel graphs. *European Journal of Combinatorics*, 28(8):2091–2105, 2007.
2. Guillaume Chapuy, Éric Fusy, Mihyun Kang, and Bilyana Shoilekova. A complete grammar for decomposing a family of graphs into 3-connected components. *The Electronic Journal of Combinatorics*, 15(R148), 2008.
3. Michael Drmota. *Random Trees*. SpringerWienNewYork, Vienna, 2009.
4. Michael Drmota, Éric Fusy, Mihyun Kang, Veronika Kraus, and Juanjo Rué. Asymptotic study of subcritical graph classes. *SIAM Journal on Discrete Mathematics*, 25(4):1615–1651, 2011.
5. Michael Drmota, Lander Ramos, Clément Requilé, and Juanjo Rué. Maximal independent sets and maximal matchings in series-parallel and related graph classes. *The Electronic Journal of Combinatorics*, 27(P1.5), 2020.
6. Michael Drmota, Lander Ramos, and Juanjo Rué. Subgraph statistics in subcritical graph classes. *Random Structures & Algorithms*, 51(4):631–673, 2017.
7. Julia Ehrenmüller and Juanjo Rué. Spanning trees in random series-parallel graphs. *Advances in Applied Mathematics*, 75:18–55, 2016.
8. Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2009.

9 Agelos Georgakopoulos and Stephan Wagner. Subcritical graph classes containing all planar graphs. *Combinatorics, Probability and Computing*, 27(5):763–773, 2018.
10 Hsien-Kuei Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19(3):329–343, 1998.
11 Olav Kallenberg. *Foundations of Modern Probability*. Probability and its Applications. Springer-Verlag, New York, second edition, 2002.
12 Konstantinos Panagiotou, Benedikt Stufler, and Kerstin Weller. Scaling limits of random graphs from subcritical graph classes. *The Annals of Probability*, 44(5):3291–3334, 2016.

A Examples

A.1 Cacti

Cacti are graphs whose blocks are either single edges or cycles. Thus there are $\frac{(n-1)!}{2}$ (labelled) blocks on n vertices for every $n > 2$, and precisely one block on two vertices. We find that the block generating function is given by

$$B(z) = \frac{z^2}{2} + \sum_{n=3}^{\infty} \frac{(n-1)!}{2n!} z^n = -\frac{1}{2}\log(1-z) + \frac{z^2}{4} - \frac{z}{2}.$$

One finds that γ is the positive root of the polynomial $z^3 - 4z^2 + 6z - 2$, and

$$\rho = \gamma \exp\big(-B'(\gamma)\big) = \gamma \exp\Big(-\frac{\gamma(2-\gamma)}{2(1-\gamma)}\Big).$$

Numerically, $\gamma \approx 0.45631$ and $\rho \approx 0.23874$. Let us now determine the modified generating function $B(z,t)$ for different choices of the weight function.

Number of blocks

In this case, we have $B(z,t) = e^t B(z)$, and we can apply the formulas in (10), which give us

$$\mu = B'(\gamma) \approx 0.64780 \quad \text{and} \quad \sigma^2 = B'(\gamma) - \frac{1}{1+\gamma^2 B'''(\gamma)} \approx 0.21218.$$

Number of edges

Here, $w(B) = |B|$ for all blocks B other than a single edge. Thus we have $B(z,t) = B(ze^t) + \frac{z^2}{2}(e^t - e^{2t})$. It is straightforward to determine numerical values for μ and σ^2 using this explicit formula for $B(z,t)$: we have $\mu \approx 1.19149$ and $\sigma^2 \approx 0.06272$.

Number of spanning trees and number of connected spanning subgraphs

For the number of spanning trees (more precisely, its logarithm), the appropriate weight function is given by $w(B) = \log|B|$ for $|B| > 2$, since a cycle of length k has precisely k spanning trees, and $w(B) = 0$ for $|B| = 2$. Thus

$$B(z,t) = \frac{z^2}{2} + \sum_{n=3}^{\infty} \frac{(n-1)!}{2n!} z^n e^{t\log n} = \frac{z^2}{2} + \frac{1}{2}\sum_{n=3}^{\infty} n^{t-1} z^n.$$

Thus the logarithm of the number of spanning trees in cacti is asymptotically normally distributed, with mean and variance asymptotically equal to μn and $\sigma^2 n$ respectively, where

$$\mu = B_{zt}(\gamma, 0) = \frac{1}{2}\sum_{n=3}^{\infty} \gamma^{n-1}\log n \approx 0.24985$$

and $\sigma^2 \approx 0.08007$. The number of connected spanning subgraphs is very similar, except that $w(B) = \log(|B| + 1)$ for $|B| > 2$. We get an analogous result with $\mu \approx 0.29690$ and $\sigma^2 \approx 0.12113$.

A.2 Block graphs

Block graphs are similar to cacti: every block is a complete graph. Thus there is precisely one type of block for every size. Since there is only one way of labelling a complete graph, the block generating function is

$$B(z) = \sum_{n=2}^{\infty} \frac{1}{n!} z^n = e^z - z - 1.$$

Therefore, $\gamma \approx 0.56714$ is the positive real solution to the equation $ze^z = 1$, and $\rho \approx 0.26438$. We consider several block statistics again:

Number of blocks

Again, the formulas given in (10) apply, and we have $\mu \approx 0.76322$ and $\sigma^2 \approx 0.12512$ in Theorem 3.

Number of edges

For the number of edges, we now have to take $w(B) = \binom{|B|}{2}$. As a result, we obtain

$$\mu = B_{zt}(\gamma, 0) = \sum_{n=2}^{\infty} \binom{n}{2} \frac{1}{(n-1)!} \gamma^{n-1} = \Big(\gamma + \frac{\gamma^2}{2}\Big) e^{\gamma} = 1 + \frac{\gamma}{2} \approx 1.28357.$$

Similar calculations for higher order partial derivatives of B yield $\sigma^2 = \frac{\gamma(\gamma^2+2\gamma+2)}{4(\gamma+1)} \approx 0.31267$.

Number of spanning trees

For the (logarithm of the) number of spanning trees, we need to take $w(B) = (|B| - 2) \log |B|$, since a complete graph with b vertices has b^{b-2} spanning trees. It follows that

$$\mu = B_{zt}(\gamma, 0) = \sum_{n=2}^{\infty} \frac{(n-2) \log n}{(n-1)!} \gamma^{n-1} \approx 0.28580,$$

and we find the numerical value of σ^2 to be 0.23671.

Number of complete subgraphs

The number of complete subgraphs is an example of a block statistic whose weight function has exponential growth. However, since the block generating function has radius of convergence $\eta = \infty$ in this case, the conditions of Theorem 3 are still clearly satisfied. We have $w(B) = 2^{|B|} - |B| - 1$ in this case: recall here that we are only counting nontrivial complete subgraphs with at least two vertices – if we want to count all complete subgraphs, we only need to add the number of vertices, which is a deterministic quantity in our setting.

It follows that Theorem 3 applies with

$$\mu = B_{zt}(\gamma, 0) = 2e^{2\gamma} - (\gamma + 2)e^{\gamma} = \frac{2 - 2\gamma - \gamma^2}{\gamma^2} \approx 1.69146$$

and (by a similar calculation) $\sigma^2 = \frac{12\gamma^3 - 24\gamma^2 + 4\gamma + 4}{\gamma^4(\gamma+1)} \approx 4.55177$.

A.3 Series-parallel graphs

Series-parallel graphs are the most complicated example that we consider, since the block generating function can only be defined implicitly in terms of generating functions in two variables z and y, respectively marking the number of vertices and edges.

In fact, each block b with a distinguished vertex admits a tree-like decomposition $\tau(b)$ into components that are either of types *ring* or *multi-edge*, where the nodes of $\tau(b)$ correspond to the different components of b. We refer the reader to [2] for more details. Using then a vertex-distinguished version of the *dissymmetry theorem for tree-decomposable classes* (see [2, Section 5.3.3]), one can relate the generating function of blocks with a distinguished vertex to the generating functions $T^{(r)}(z,y)$, $T^{(m)}(z,y)$ and $T^{(rm)}(z,y)$ of their associated tree-decompositions, respectively rooted at a node of type ring, multi-edge or at an edge between a ring and a multi-edge node.

Those generating functions are in turn expressed in terms of $D(z,y)$, the generating function of series-parallel *networks*, i.e. 2-connected graphs with a pair of distinguished vertices called its *poles*. Such networks can either be the single edge, with generating function y, or of types *series*, with generating function $S(z,y)$, or *parallel*, with generating function $P(z,y)$, see [1] for a detailed exposition. Altogether, this gives:

$$\begin{aligned}
B_z(z,y) &= zy + T^{(r)}(z,y) + T^{(m)}(z,y) - T^{(rm)}(z,y),\\
T^{(r)}(z,y) &= zS(z,y)(D(z,y) - S(z,y))/2,\\
T^{(m)}(z,y) &= zP(z,y) - zyS(z,y) - zS^2(z,y)/2,\\
T^{(rm)}(z,y) &= zS(z,y)P(z,y),\\
D(z,y) &= y + S(z,y) + P(z,y),\\
S(z,y) &= zD(z,y)(D(z,y) - S(z,y)),\\
P(z,y) &= y\exp(S(z,y)) - y + \exp(S(z,y)) - S(z,y) - 1. \qquad (11)
\end{aligned}$$

From the last three equations of (11) we obtain an implicit equation defining $D(z,y)$. Furthermore, one can write $B_z(z,y)$ in terms of $D(z,y)$ only. This gives:

$$\begin{aligned}
D(z,y) &= (1+y)e^{\frac{zD(z,y)^2}{1+zD(z,y)}} - 1,\\
B_z(z,y) &= \frac{zD(z,y)(2 - zD(z,y)^2)}{2zD(z,y)+2}. \qquad (12)
\end{aligned}$$

So any partial derivative of $B_z(x,y)$ can be computed from the system (12). In particular, we get numerically that $\gamma \approx 0.12797$.

Number of blocks

For the number of blocks, one now sets $y = 1$ and $w(B) = 1$ in (12). The required conditions are satisfied, so we obtain a central limit theorem. In this case, the numerical values μ and σ^2 are 0.14937 and 0.14875 respectively.

Number of edges

Although the number of edges is now no longer just dependent on the number of vertices of a block, it is already controlled by the variable $y = e^t$ in the decomposition given in (12). We obtain a central limit theorem with $\mu \approx 1.61673$ and $\sigma^2 \approx 0.21125$ (cf. [1]).

Number of spanning trees

The number of spanning trees in series-parallel graphs was studied in a paper by Ehrenmüller and Rué [7]. They determined an asymptotic formula for the mean, but no limit distribution. As we find by means of our general result now, the distribution of the number of spanning trees in series-parallel graphs is asymptotically lognormal. Letting $\tau(G)$ be the number of spanning trees of a graph G, we simply set $w(B) = \log \tau(B)$ for all blocks (as for the other two graph classes), so that $S_w(G) = \log \tau(G)$. The conditions of Theorem 3 are clearly satisfied again, but the constants μ and σ^2 are rather difficult to evaluate in this example, as they can no longer be expressed directly by means of functional equations. Moreover, the infinite series

$$\mu = B_{zt}(\gamma, 0) = \sum_{B \in \mathcal{B}} \frac{w(B)\gamma^{|B|-1}}{(|B|-1)!}$$

converges poorly in this example, since $\gamma \approx 0.12797$ is only a little smaller than the radius of convergence of $B(z, 0)$, which is $\eta \approx 0.12800$. Therefore, the series representation is also not suitable to compute a numerical approximation, as determining $w(B) = \log \tau(B)$ for all blocks up to a certain size is rather costly.