Supplemental materials

Sebastian M. Krause and Vinko Zlatić
Theoretical Physics Division, Rudjer Bošković Institute, Zagreb, Croatia

Michael M. Danziger
Department of Physics, Bar Ilan University, Ramat Gan, Israel

List of variables

	Networks
N	Number of nodes
$\bar{k}$	Average degree
k_i	Degree of node i
p_k	Degree distribution
α	Exponent of scale free degree distribution
g_0	Generating function of degree
g_1	Generating function of excess degree
	Colors
C	Number of colors
$c \in 1, 2, \ldots C$	A color
r_c	Color distribution
n_{deg}	Degeneration of the highest color frequency
$\tilde{r}_{c,k}$	degree-dependent color distribution
	Standard percolation ingredients
$\mathcal{G}$	Set of nodes in the giant component (color blind)
u	Prob. of not being connected to giant comp. over a link
S	Size of giant component
$\phi_{\bar{c}}$	Fraction of nodes without color c
$\mathcal{G}_{\bar{c}}$	Set of nodes in the giant component avoiding color c
$u_{\bar{c}}$	Prob. of not being connected to color avoiding giant comp. over a link
$S_{\bar{c}}$	Size of color avoiding giant component
	Percolation over color avoiding paths
$\mathcal{G}_{\mathrm{color}}$	Set of nodes which can communicate avoiding all colors
S_{color}	Size of this component
$B_{k,k'}$	Prob. that out of k links k' connect to giant component
$M_{k',\vec{\kappa}}$	Prob. that out of k' links κ_1 connect to color 1 etc.
$P_{\vec{\kappa}}$	Success probability having neighbors of colors acc. to $\vec{\kappa}$
$U_{\bar{c}}$	Prob. that a link fails connecting to $\mathcal{G}_{\mathrm{color}}$ which already connects to $\mathcal{G}$
$S_{\mathrm{color},\infty}$	Size of the set of all nodes being connected to giant component over two links or more
β	Critical exponent
$\bar{k}_{\mathrm{crit}}$	Critical value of average degree
k_{step}	Degree above which all nodes have the same color
γ	Fraction of nodes with highest degree

I. SIZE OF GIANT AVOIDABLE COLORS COMPONENT IN THE CONFIGURATION MODEL

A. Graph ensembles with color distributions

B. Question and connection to percolation theory

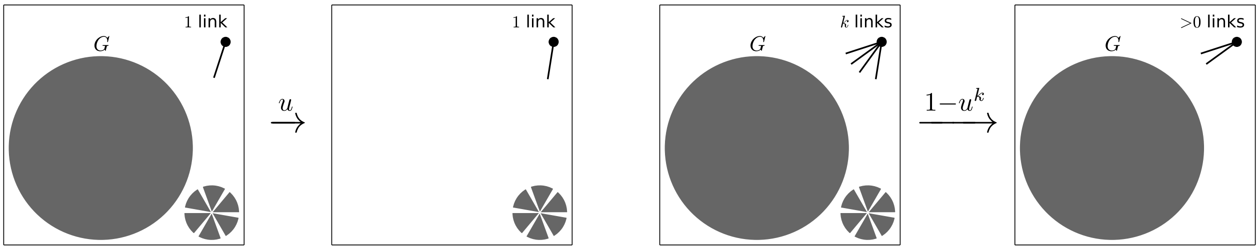

FIG. 1: We base our theory on the method to calculate the size of normal giant components, as illustrated in this figure. Using a self consistency equation, the probability u can be calculated. This is the probability, that a node is not connected to the giant component over a single link (see on the left). On the right, the probability for a node with k links is illustrated to have at least one link connecting to the giant component. u^k is the probability that all links fail.

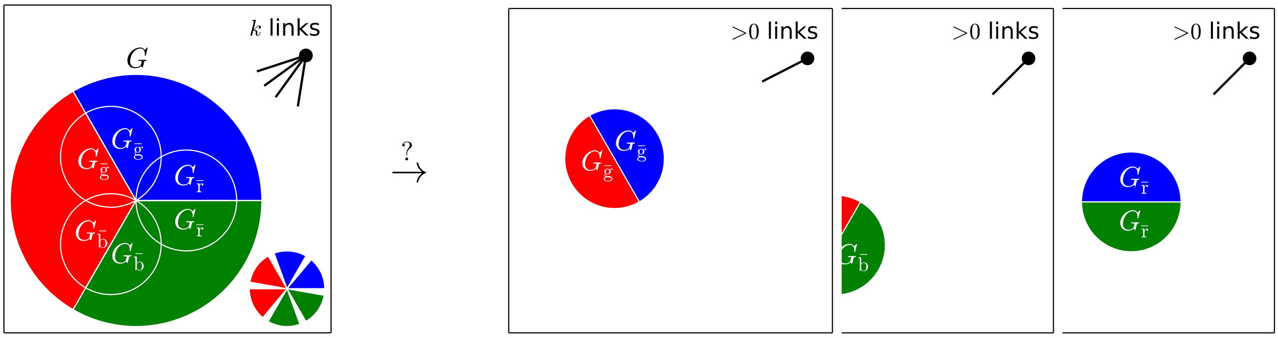

FIG. 2: We have to calculate the probability, if a node with k links is for every color c connected to the giant component $G_{\bar{c}}$ with deleted color c. All connections over at least one link have to exist at the same time. We illustrate this question with the three colors red ($c=$r), green ($c=$g) and blue ($c=$b). If a link connects to $G_{\bar{\text{g}}}$, it for sure does not connect to $G_{\bar{c}}$ for one of the other colors. This kind of dependence forces us to use a stepwise calculation with conditional probabilities.

For estimating the size of the largest avoidable colors component analytically, results from percolation theory can be used. First of all, the existence of a (colorblind) giant component clearly is a prerequisite for the existence of a macroscopic fraction of successful node pairs. With the generating functions of degree $g_0(z) = \sum_k p_k z^k$ and excess degree $g_1(z) = \sum_k q_k z^k$, the size of the giant component S can be calculated (assuming infinite networks which are locally treelike) using the average probability u, "that a vertex is not connected to the giant component via its connection to some particular neighboring vertex" (Newman, page 461):

$$u = g_1(u) \tag{1}$$

$$S = 1 - g_0(u). \tag{2}$$

Lets call the set of all nodes belonging to the giant component as $\mathcal{G}$.

As we have tested numerically that S_{color} can be used to describe the success of node pairs in connecting over paths with avoided colors, it is useful to assess this quantity analytically. We will do this assuming an infinite, locally treelike network. We calculate S_{color} as the probability, that a randomly chosen node belongs to $\mathcal{G}_{\text{color}}$.

Let us start with the standard percolation theory. We can rewrite eq. 2 for the size of the giant component as

$$S = \sum_{k=0}^{\infty} p_k \sum_{k'=0}^{k} B_{k,k'} \times (1 - \delta_{k',0}), \tag{3}$$

$$B_{k,k'} = \binom{k}{k'} (1-u)^{k'} u^{k-k'}, \tag{4}$$

where p_k is the probability that a randomly chosen node has exactly k links and $B_{k,k'}$ is the binomial probability that out of these links k' links connect to the giant component. The success probability $(1 - \delta_{k',0})$ is zero if there is no link connecting to the giant component and one else.

C. One link probabilities

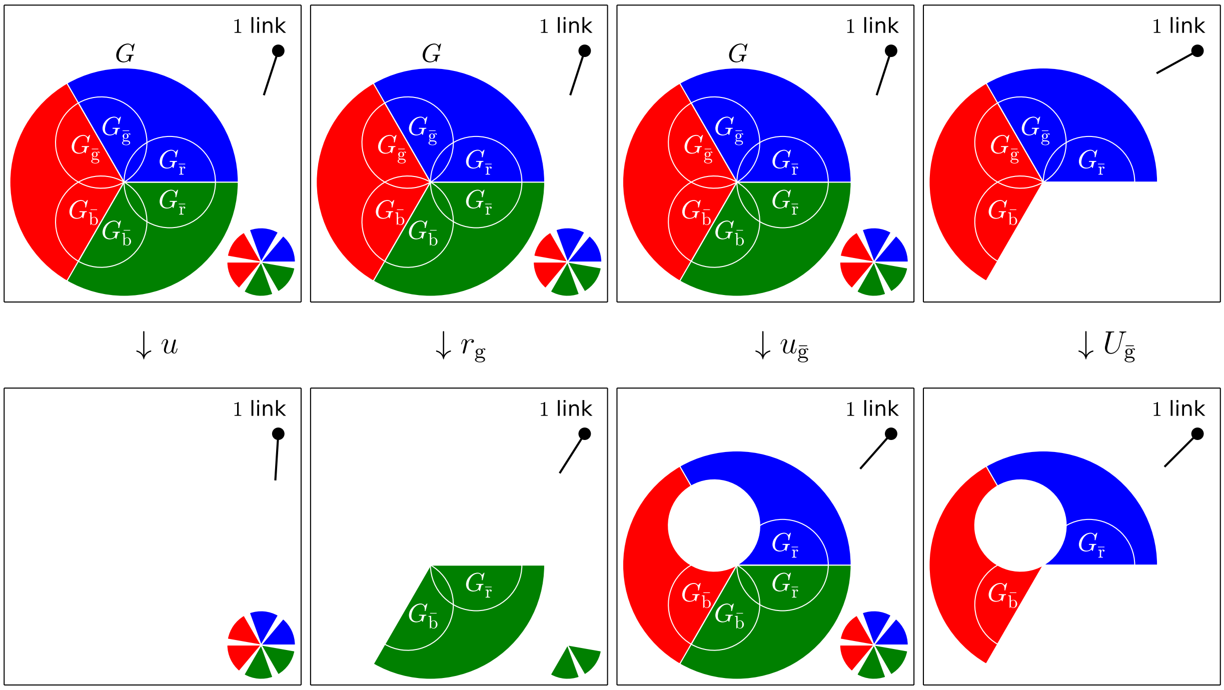

FIG. 3: Probabilities for a single link to connect to different parts of the network. We use these probabilities as primitives to calculate the probability for many links. While u, r_c and $u_{\bar{c}}$ can be calculated with standard methods invented for the configuration model before, the conditional probability $U_{\bar{c}}$ can be calculated as a combination of the others.

Another prerequisite is the existence of a giant component after deleting all nodes of one of the colors c. Lets call the analogue of u after all nodes of color c are deleted as $u_{\bar{c}}$, the set of all nodes in the remaining giant component as $\mathcal{G}_{\bar{c}}$, and its size as $S_{\bar{c}}$. We have

$$\phi_{\bar{c}} = 1 - r_c \tag{5}$$

$$u_{\bar{c}} = 1 - \phi_{\bar{c}} + \phi_{\bar{c}} g_1(u_{\bar{c}}) \tag{6}$$

$$S_{\bar{c}} = \phi_{\bar{c}}(1 - g_0(u_{\bar{c}})). \tag{7}$$

Finally a node pair is for sure successful, if for both nodes the following holds: For every color c there exists at least one neighbor belonging to $\mathcal{G}_{\bar{c}}$. The set of all nodes fulfilling this condition is L_{color}, and its size as S_{color}.

According to the choices we have made so far, those links connect to the giant component $\mathcal{G}$ and none of the nodes they are connecting to has color c. Therefore, a single of those links fails in connecting to $\mathcal{G}_{\bar{c}}$ with the conditional probability

$$U_{\bar{c}} = 1 - \frac{1 - u_{\bar{c}}}{(1-u)(1-r_c)}. \tag{8}$$

The last term is the probability, that over a single link a connection to $\mathcal{G}_{\bar{c}}$ is established, if this link already fulfills the following precondition: It connects to $\mathcal{G}$ and at the same time to a node without color c. This precondition has probability $(1-u)(1-r_c)$, as colors are randomly distributed and therefore are not correlated with the probability u or $1-u$. As the links connecting to $\mathcal{G}_{\bar{c}}$ are a subset of all links fulfilling the precondition, the conditional probability can be calculated by dividing with the probability of the precondition. Notice that the additional information of the explicit color, instead of only stating that the color is not c, does not alter the results, as a further restriction of the colors would meat the numerator and denominator identically and therefore would cancel out.

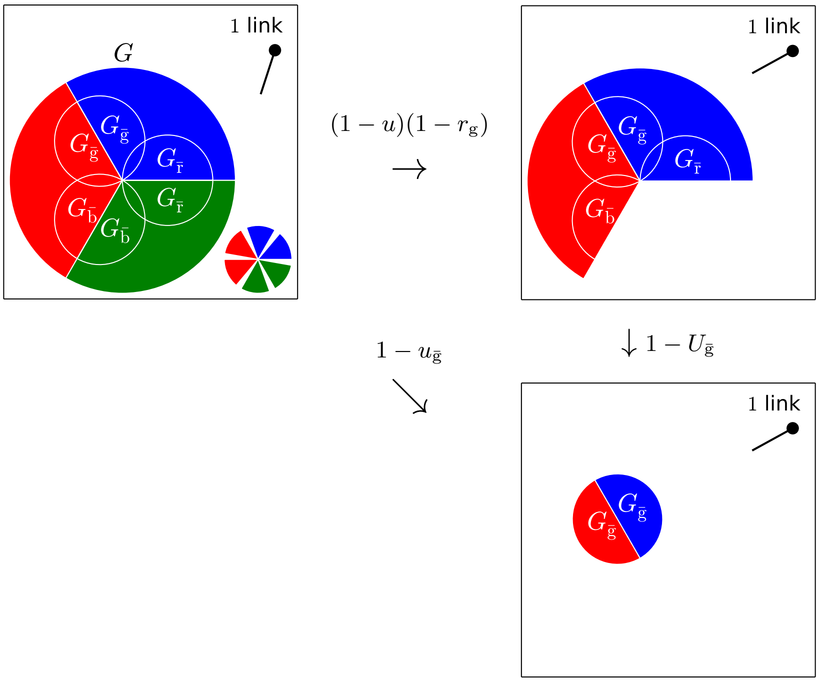

FIG. 4: This figure illustrates the calculation of $U_{\bar{g}}$ using the equality $(1-u)(1-r_{g})(1-U_{\bar{g}}) = 1-u_{\bar{g}}$. For that, we have assumed independence of the qualities of the link under consideration, especially of the color it connects to and if it connects to the giant component.

D. Averaging over link distributions

As we have tested numerically that S_{color} can be used to describe the success of node pairs in connecting over paths with avoided colors, it is useful to assess this quantity analytically. We will do this assuming an infinite, locally treelike network. We calculate S_{color} as the probability, that a randomly chosen node belongs to $\mathcal{G}_{\text{color}}$.

Let us start with the standard percolation theory. We can rewrite eq. 2 for the size of the giant component as

$$S = \sum_{k=0}^{\infty} p_k \sum_{k'=0}^{k} B_{k,k'} \times (1-\delta_{k',0}) \,, \tag{9}$$

$$B_{k,k'} = \binom{k}{k'} (1-u)^{k'} u^{k-k'}, \tag{10}$$

where p_k is the probability that a randomly chosen node has exactly k links and $B_{k,k'}$ is the binomial probability that out of these links k' links connect to the giant component. The success probability $(1-\delta_{k',0})$ is zero if there is no link connecting to the giant component and one else. For our problem, this last term has to be replaced. In order to calculate the probability that the k' links connect to all components $\mathcal{G}_{\bar{c}}$, we first have to consider the distribution of colors among the nodes these links connect to.

$$M_{k',\vec{\kappa}} = \frac{k'!}{\kappa_1! \times \cdots \times \kappa_C!} (r_1)^{\kappa_1} \times \cdots \times (r_C)^{\kappa_C} \, \delta_{k',\kappa_1+\cdots+\kappa_C} \tag{11}$$

denotes the multivariate probability that out of those k' links κ_1 connect to nodes of color 1, κ_2 links connect to nodes of color 2 etc. We define $P_{\vec{\kappa}}$ as the success probability to connect to all components $\mathcal{G}_{\bar{c}}$ given $\vec{\kappa}$. With this quantity, which will be evaluated below, we finally can write

$$S_{\text{color}} = \sum_{k=0}^{\infty} p_k \sum_{k'=0}^{k} B_{k,k'} \sum_{\kappa_1,\ldots,\kappa_C=0}^{k'} M_{k',\vec{\kappa}} P_{\vec{\kappa}}. \tag{12}$$

For evaluating $P_{\vec{\kappa}}$, lets first concentrate on one color c. We have $\sum_{c' \neq c} \kappa_{c'}$ links which potentially can connect to the desired component $\mathcal{G}_{\bar{c}}$. According to the choices we have made so far, those links connect to the giant component

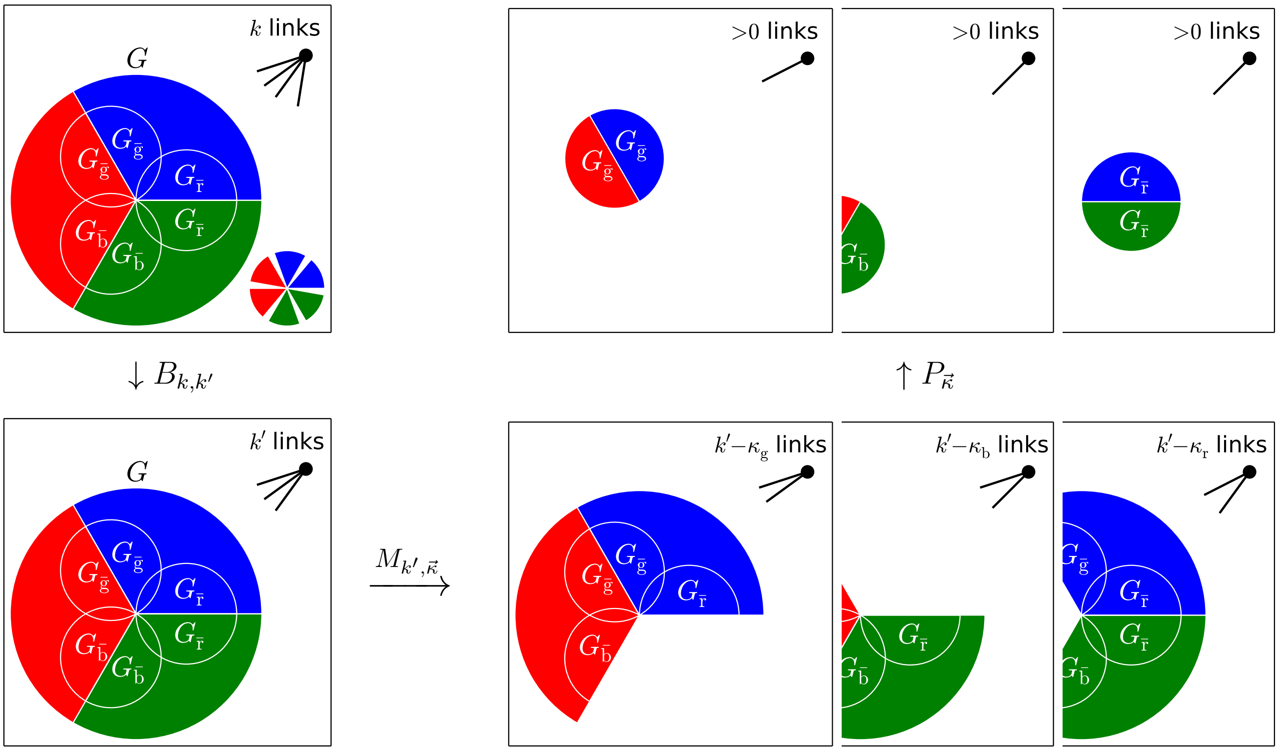

FIG. 5: For calculating the probability of a node with k links to belong to G_{color}, we have to average over different link constellations which this node might show. First, $B_{k,k'}$ is the probability that out of the k links k' connect to the giant component. It is calculated using u (compare figure 3 on the left). Second, $M_{k',\vec{\kappa}}$ gives the probability for a certain color distribution among the links. It is calculated using r_g etc. (compare figure 3, second from left). We assume that this second step is independent of the first step, what is confirmed with the final results. Third, $P_{\vec{\kappa}}$ gives the joint probability that for this color distribution $G_{\bar{r}}$, $G_{\bar{b}}$ and $G_{\bar{g}}$ are connected to at the same time. This is calculated using $U_{\bar{r}}$ etc. (compare figure 3 on the right). We tested numerically that e.g. $U_{\bar{r}}$ and $U_{\bar{b}}$ are independent for a link connecting to the giant component and color green.

$\mathcal{G}$ and none of the nodes they are connecting to has color c. Therefore, a single of those links fails in connecting to $\mathcal{G}_{\bar{c}}$ with the conditional probability

$$U_{\bar{c}} = 1 - \frac{1 - u_{\bar{c}}}{(1-u)(1-r_c)}. \tag{13}$$

The last term is the probability, that over a single link a connection to $\mathcal{G}_{\bar{c}}$ is established, if this link already fulfills the following precondition: It connects to $\mathcal{G}$ and at the same time to a node without color c. This precondition has probability $(1-u)(1-r_c)$, as colors are randomly distributed and therefore are not correlated with the probability u or $1-u$. As the links connecting to $\mathcal{G}_{\bar{c}}$ are a subset of all links fulfilling the precondition, the conditional probability can be calculated by dividing with the probability of the precondition. Notice that the additional information of the explicit color, instead of only stating that the color is not c, does not alter the results, as a further restriction of the colors would meat the numerator and denominator identically and therefore would cancel out. There is at least one link connecting to $\mathcal{G}_{\bar{c}}$ with probability $1-(U_{\bar{c}})^{\sum_{c'\neq c}\kappa_{c'}}$. The success probabilities for different colors have to be multiplied, as all $\mathcal{G}_{\bar{c}}$ have to be reached at the same time. Putting everything together we have

$$P_{\vec{\kappa}} = \prod_{c=1}^{C}[1-(U_{\bar{c}})^{\sum_{c'\neq c}\kappa_{c'}}]. \tag{14}$$

Results for Poisson graphs are shown in figure **??** with the red line, showing S^2_{color} as the probability of two nodes to be connected via all Components $G_{\bar{c}}$ simultaneously. Instead of evaluating the sums over k' and $\vec{\kappa}$ in eq. 12, we sampled 5000 events for every k.The outcome compares well with numerical results.

E. Degree-dependent color distributions

Our framework can be generalized to color distributions $\tilde{r}_{c,k}$ additionally depending on the degree k of the nodes. For an easier overview here all the needed equations. The following equations stay unchanged,

$$S_{\mathrm{color}} = \sum_{k=0}^{\infty} p_k \sum_{k'=0}^{k} B_{k,k'} \sum_{\kappa_1,\ldots,\kappa_C=0}^{k'} M_{k',\vec{\kappa}} P_{\vec{\kappa}}, \tag{15}$$

$$B_{k,k'} = \binom{k}{k'} (1-u)^{k'} u^{k-k'}, \tag{16}$$

$$u = g_1(u), \quad g_1(z) = \sum_k q_k z^k, \tag{17}$$

$$M_{k',\vec{\kappa}} = \frac{k'!}{\kappa_1! \times \cdots \times \kappa_C!} (r_1)^{\kappa_1} \times \cdots \times (r_C)^{\kappa_C} \, \delta_{k',\kappa_1+\cdots+\kappa_C}, \tag{18}$$

$$P_{\vec{\kappa}} = \prod_{c=1}^{C} [1 - (U_{\bar{c}})^{\sum_{c' \neq c} \kappa_{c'}}], \tag{19}$$

$$U_{\bar{c}} = 1 - \frac{1 - u_{\bar{c}}}{(1-u)(1-r_c)}, \tag{20}$$

while including modified quantities r_c and $u_{\bar{c}}$ according to

$$r_c = \sum_k k p_k \tilde{r}_{c,k} / \bar{k}, \tag{21}$$

$$u_{\bar{c}} = 1 - f_1(1) + f_1(u_{\bar{c}}), \quad f_1(z) = \sum_k q_k (1 - \tilde{r}_{c,k+1}) z^k. \tag{22}$$

r_c is the average probability to find color c over a link, including the fact that high degree nodes are reached more probably and therefore the colors on high degree nodes are found with higher probability. The self-consistency equation for $u_{\bar{c}}$ is well known from standard percolation theory.

II. LIMITING CASES OF THE THEORY

A. Limiting case of small color frequencies

In the limit of high numbers of colors C together with color frequencies $r_c \to 0$, the single paths have to avoid only a small part of nodes. Therefore we expect $U_{\bar{c}} \to 0$: If a link connects to the colorblind giant component, it will almost never fail to connect to the color avoiding component. Accordingly $P_{\vec{\kappa}}$ is close to one, if the needed links exist. We can use this idea to find a limiting case for S_{color} in eq. 12, and to compare to standard percolation. With the upper limit

$$\sum_{\kappa_1,\ldots,\kappa_C=0}^{k'} M_{k',\vec{\kappa}} P_{\vec{\kappa}} \leq 1 - \delta_{k',0} - \delta_{k',1} \tag{23}$$

and $\sum_k p_k \sum_{k'} B_{k,k'} \sum_{\kappa_1,\ldots,\kappa_C} M_{k',\vec{\kappa}} = 1$ (total probability is one) we finally get

$$S_{\mathrm{color}} \leq 1 - \sum_{k=0}^{\infty} p_k [u^k + k(1-u)u^{k-1}] \tag{24}$$

$$= 1 - g_0(u) - (1-u) \left. \frac{\mathrm{d}g_0(z)}{\mathrm{d}z} \right|_{z=u} \tag{25}$$

$$= S_{\mathrm{color},\infty}. \tag{26}$$

The upper limit $S_{\mathrm{color},\infty}$ for S_{color} reflects the fact, that every node has to be connected to the giant component at least over two links. It therefore includes a reduction compared to the standard percolation result $1 - g_0(u)$. In the limit of many colors and small probabilities r_c, S_{color} can come close to $S_{\text{color},\infty}$, as in this case $U_{\bar{c}}$ comes close to zero and only nodes fail which have less than two links connecting to the giant component. This result is closely connected to k-core percolation with $k = 2$. Note that k-core percolation shows a continuous phase transition for $k = 2$, and only for $k > 2$ has the well known discontinuous behavior.

B. Critical behavior for Poisson graphs

With general color distributions r_c ($\sum_c r_c = 1$), the color with the largest probability r_c dominates the behavior, as it corresponds to the largest conditional link failure probability $U_{\bar{c}}$ in equation 14. For Poisson graphs, $U_{\bar{c}}$ falls below one at $\bar{k} = 1/(1 - r_c)$, and as long as one $U_{\bar{c}}$ is one, equation 14 gives always zero. Therefor the critical value is

$$\bar{k}_{\text{crit}} = 1/(1 - \max_c r_c). \tag{27}$$

We can understand the critical exponent with expanding equation 14 using $U_{\bar{c}} = 1 - \varepsilon$ for the highest color frequency. We will show below using results from standard percolation that $\varepsilon \propto (\bar{k} - \bar{k}_{\text{crit}})$. First of all, we have $(U_{\bar{c}})^{\sum_{c' \neq c} \kappa_{c'}} \approx 1 - (\sum_{c' \neq c} \kappa_{c'}) \times \varepsilon$, and therefore we find $P_{\vec{\kappa}} \propto \varepsilon^{n_{\text{deg}}}$ (unless $P_{\vec{\kappa}} = 0$ in the case $\sum_{c' \neq c} \kappa_{c'} = 0$ for any c). As equation 12 is therefore a superposition of either vanishing terms or terms with leading order $\varepsilon^{n_{\text{deg}}}$, we have

$$\beta = n_{\text{deg}}. \tag{28}$$

To complete the discussion of the critical behavior, we have to show that $1 - U_{\bar{c}} \propto (\bar{k} - \bar{k}_{\text{crit}})$ for a certain region of critical behavior. We know from standard percolation theory for Poisson graphs that $S \propto (\bar{k} - 1)^1$, and therefore $1 - u(\bar{k}) \approx a \times (\bar{k} - 1)$ for $\bar{k}$ exceeding the critical value about small values. With inserting into equation 6 it can be shown that $u_{\bar{c}}(\bar{k}) = 1 - \phi_{\bar{c}} + \phi_{\bar{c}} u(\bar{k}\phi)$. Using this in equation 13 ($\phi_{\bar{c}} = 1 - r_c$), we find

$$U_{\bar{c}} = 1 - \frac{1 - u(\bar{k}\phi_{\bar{c}})}{1 - u(\bar{k})} \tag{29}$$

$$\approx 1 - \phi_{\bar{c}} \frac{\bar{k} - \bar{k}_{\text{crit}}}{\bar{k} - 1} \tag{30}$$

which drops linearly from one in the critical region

$$0 < \bar{k} - \bar{k}_{\text{crit}} \ll \bar{k}_{\text{crit}} - 1. \tag{31}$$

In figure **??** we have $\bar{k}_{\text{crit}} - 1 = 1/2$, and critical behavior up to about $\bar{k} - \bar{k}_{\text{crit}} = 1/10$.

As a final remark let us discuss the critical behavior for largest color frequencies which are not perfectly degenerated. Lets assume two colors with close by values $\bar{k}_1 < \bar{k}_2$, where $U_{\bar{c}}$ drops below one. With equation 14 and the definition $\varepsilon = \bar{k} - \bar{k}_2$ we have above $\bar{k}_2$ that $P_{\vec{\kappa}} \propto \varepsilon \times (\varepsilon + (\bar{k}_2 - \bar{k}_1))$. This is dominated by a linear term for small ε and by a quadratic term for larger values, the crossover is at about

$$\bar{k} - \bar{k}_2 = \bar{k}_2 - \bar{k}_1. \tag{32}$$

Exceeding the critical value $\bar{k}_{\text{crit}} = \bar{k}_2$ about more than the distance $\bar{k}_2 - \bar{k}_1$, S_{color} behaves as if the color frequencies would be degenerated.

With these results altogether we can understand the behavior of S_{color} for small frequencies of all colors. There is a deviation from $S_{\text{color},\infty}$ due to $S_{\text{color}} = 0$ below $\bar{k}_{\text{crit}}$, and above there is a region of slow critical growth with large critical exponent β. This region acts as an effective shift of the critical parameter. Further increasing $\bar{k}$, $U_{\bar{c}}$ saturates smoothly to zero for all colors. Accordingly, S_{color} has a smooth rise to finite values, but not governed by a certain critical exponent. It comes closer to $S_{\text{color},\infty}$ without reaching it, as there are other effects (e.g. nodes with exactly two links can connect to two nodes of the same color). The smaller the highest color frequencies are, the closer S_{color} is to $S_{\text{color},\infty}$.