Message passing problem on random graphs

Sebastian M. Krause

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. QUESTION

Lets assume the generalized configuration model graph ensamble with N nodes, where each degree sequences $\{k_i\}$ occurs with probability $\prod_i p_{k_i}$ with the degree distribution p_k. (see M.E.J. Newman: Networks, an Introduction; 2010; Eq. (13.30). The generalization of the configuration model is important for numerics with small networks, where many network realizations are sampled for averiging.) Lets additionally assign to every node i a color $c_i \in 1, 2, \ldots, C$. The color sequence $\{c_i\}$ has probability $\prod_i r_{c_i}$ with the color distribution r_c. How large is the fraction of node pairs, which can be connected via a set of paths, such that for every color there exists a path avoiding this color?

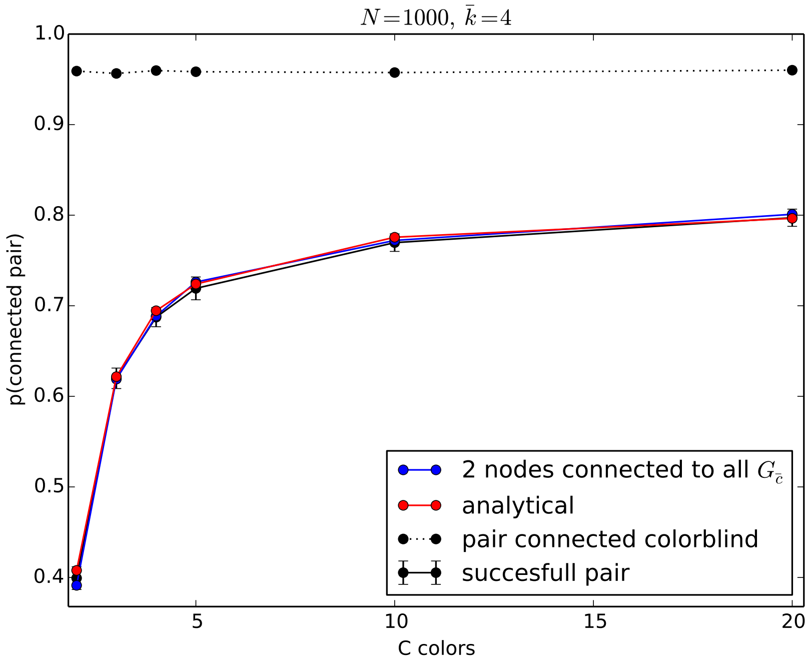

FIG. 1: Fraction of node pairs connected with color avoiding paths as described in the text. Colors are distributed randomly on Poisson graphs with average degree $\bar{k} = 4$. Black symbols with error bars show results for networks of size $N = 1000$, with samples of 200 node pairs, averaged over 50 networks. Connected pairs are suppressed compared to the existence of colorblind paths, as indicated with black symbols. Blue symbols show the fraction of node pairs where each node is connected to color avoiding giant components, confirming the understanding using percolation theory. Red symbols show analytical results of percolation theory.

For numerical results, we generated Poisson graphs with $N = 1000$ (we used the Poisson distribution as an approximation to the binomial distribution in ER-graphs), and distributed C colors over the nodes, with $r_c = 1/C$. To check for a single pair, if it can be connected in the way as described above, we removed all nodes with the first color (except for the node pair under consideration) and searched for the shortest path (using a library function returning an empty set, if no path exists). Then we took the original graph and removed all nodes with the second color etc. If for all colors the shortest path exists, we considered the pair as successful. The case of a successful pair with a smaller number of paths, where e.g. one path avoids two colors, is included, as such a path could just be counted twice with our procedure. We repeated this 200 times with node pairs randomly chosen, and calculated the fraction of successful node pairs. See the black straight line in figure 1, with averages over 50 network realizations. The dotted line shows the fraction of node pairs, which are connected with a path which can have all colors.

II. THEORY

A. Connection to percolation

For estimating the fraction of successful pairs 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 $\mathcal{G}$. 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

$$\phi_{\bar{c}} = 1 - r_c \tag{3}$$
$$u_{\bar{c}} = 1 - \phi_{\bar{c}} + \phi_{\bar{c}} g_1(u_{\bar{c}}) \tag{4}$$
$$S_{\bar{c}} = \phi_{\bar{c}}(1 - g_0(u_{\bar{c}})). \tag{5}$$

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 $\mathcal{G}_{\bar{c}}$. Lets call the set of all nodes fulfilling this condition as $\mathcal{G}_{\text{color}}$, and its size as S_{color}. The fraction of successful node pairs should be approximately S^2_{color}.

We tested this hypothesis with numerical results. For every color c, we labeled all the nodes with a boolean variable, if they belong to $\mathcal{G}_{\bar{c}}$ (more precise, if they belong to the largest component, as the networks are finite). A single node can belong to different components $\mathcal{G}_{\bar{c}}$. Then we searched for nodes which can be successful in a node pair. For a single node, we iterated over all the neighbors and collected all the memberships $\mathcal{G}_{\bar{c}}$ present. If for all colors c there is at least one membership $\mathcal{G}_{\bar{c}}$ among the neighbors, the node belongs to $\mathcal{G}_{\text{color}}$ and is potentially successful in a pair. We calculated the fraction S_{color} over all nodes, and by squaring this we got an estimate for successful node pairs. This procedure ignores paths in small components, but is a good estimate, as can be seen with the blue line in figure 1. This procedure is much faster as well.

B. Analytical results for the percolation problem

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

$$S = \sum_{k=0}^{\infty} p_k \sum_{k'=0}^{k} B_{k,k'} \times (1 - \delta_{k',0}), \tag{6}$$
$$B_{k,k'} = \binom{k}{k'} (1-u)^{k'} u^{k-k'}, \tag{7}$$

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.

$$M_{k',\vec{\kappa}} = \frac{k'!}{\kappa_1! \times \cdots \times \kappa_C!} (r_1)^{\kappa_1} \times \cdots \times (r_C)^{\kappa_C} \, \delta_{k',\kappa_1+\cdots+\kappa_C} \tag{8}$$

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

$$S_{\text{color}} = \sum_{k=0}^{\infty} p_k \sum_{k'=0}^{k} B_{k,k'} \sum_{\kappa_1,\ldots,\kappa_C=0}^{k'} M_{k',\vec{\kappa}} P_{\vec{\kappa}}. \tag{9}$$

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

$\mathcal{G}$ 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

$$U_{\bar{c}} = 1 - \frac{1 - u_{\bar{c}}}{(1-u)(1-r_c)}. \tag{10}$$

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

$$P_{\vec{\kappa}} = \prod_{c=1}^{C} [1 - (U_{\bar{c}})^{\sum_{c' \neq c} \kappa_{c'}}]. \tag{11}$$

Results for Poisson graphs are shown in figure 1 with the red line, showing S_{color}^2 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. 9, we sampled 5000 events for every k.The outcome compares well with numerical results.

C. 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. 9, and to compare to standard percolation. With the upper limit

$$\sum_{\kappa_1, \ldots, \kappa_C = 0}^{k'} M_{k', \vec{\kappa}} P_{\vec{\kappa}} \leq 1 - \delta_{k',0} - \delta_{k',1} \tag{12}$$

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

$$\begin{aligned} S_{\text{color}} &\leq 1 - \sum_{k=0}^{\infty} p_k [u^k + k(1-u)u^{k-1}] && (13) \\ &= 1 - g_0(u) - (1-u) \left. \frac{\mathrm{d}g_0(z)}{\mathrm{d}z} \right|_{z=u} && (14) \\ &= S_{\text{color},\infty}. && (15) \end{aligned}$$

The upper limit $S_{\text{color},\infty}$ 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.

D. 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,

$$S_{\text{color}} = \sum_{k=0}^{\infty} p_k \sum_{k'=0}^{k} B_{k,k'} \sum_{\kappa_1,\ldots,\kappa_C=0}^{k'} M_{k',\vec{\kappa}} P_{\vec{\kappa}}, \tag{16}$$

$$B_{k,k'} = \binom{k}{k'} (1-u)^{k'} u^{k-k'}, \tag{17}$$

$$u = g_1(u), \quad g_1(z) = \sum_k q_k z^k, \tag{18}$$

$$M_{k',\vec{\kappa}} = \frac{k'!}{\kappa_1! \times \cdots \times \kappa_C!} (r_1)^{\kappa_1} \times \cdots \times (r_C)^{\kappa_C} \, \delta_{k',\kappa_1+\cdots+\kappa_C}, \tag{19}$$

$$P_{\vec{\kappa}} = \prod_{c=1}^{C} [1 - (U_{\bar{c}})^{\sum_{c' \neq c} \kappa_{c'}}], \tag{20}$$

$$U_{\bar{c}} = 1 - \frac{1 - u_{\bar{c}}}{(1-u)(1-r_c)}, \tag{21}$$

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

$$r_c = \sum_k k p_k \tilde{r}_{c,k} / \bar{k}, \tag{22}$$

$$u_{\bar{c}} = 1 - f_1(1) + f_1(u_{\bar{c}}), \quad f_1(z) = \sum_k q_k (1 - \tilde{r}_{c,k+1}) z^k. \tag{23}$$

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.

III. RESULTS

A. Poisson graphs

In figure 2 the dependence of S_{color} on the average degree is shown for different numbers of colors C ($r_c = 1/C$). Comparing to the standard giant component size S (dashed black line in the figure), the percolation sets in at increasing $\bar{k}$ with smaller numbers of colors, and the component size grows slower to the saturation value of one. The symbols show numerical results with $N = 1000$ and 100 network realizations, the lines show results of equation 9, both correspond well.

The suppression of the number of connected nodes can be understood as a combination of two effects. The first effect is purely topological and can be understood with $S_{\text{color},\infty}$ of eq. 15 (shown with dotted red line). It means that only nodes can belong to S_{color}, which are connected to the colorblind giant component over at least two links. We can confirm that S_{color} comes close to $S_{\text{color},\infty}$ for high numbers of colors C with the results for $C = 10$. $S_{\text{color},\infty}$ is remarkably reduced compared to S for small k, but has the same critical parameter. For the Poisson graph we have $g_0(z) = \exp(\bar{k}(z-1))$ and $S = 1 - u$ and therefore $S_{\text{color},\infty} = S - \bar{k}S(1-S)$, and for small positive $\bar{k} - 1$ the giant component grows approximately with $S \approx 2(\bar{k}-1)/\bar{k}^2$. Therefore

$$S_{\text{color},\infty} \approx 2(\bar{k}-1)^2/\bar{k}, \tag{24}$$

which grows slowly for small parameter $\bar{k} - 1$.

The second effect is connected to finite color frequencies r_c which further reduces the percolating fraction of nodes. This also changes the critical value $\bar{k}_{\text{crit}}$ and the critical exponent β. We will discuss the critical behavior for the more general case of heterogeneous color distributions r_c.

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 11. For Poisson graphs, $U_{\bar{c}}$ falls

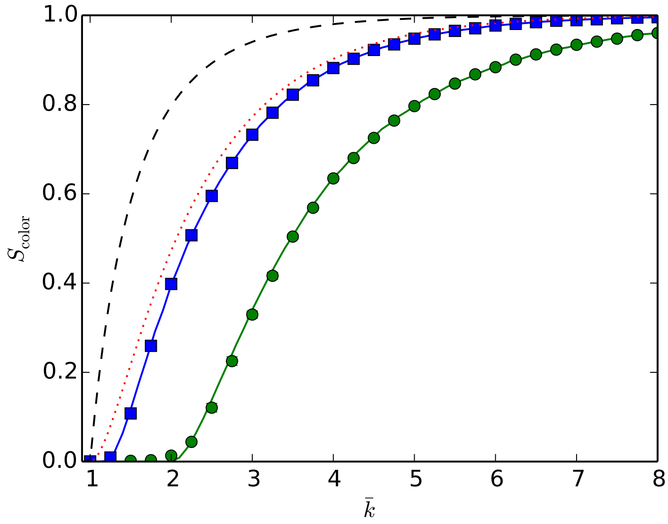

FIG. 2: Dependence of S_{color} on the average degree for different number of colors. Symbols show numerical results for networks of size $N = 1000$ (blue squares for $C = 10$ and green circles for $C = 2$ colors), the straight lines show the according analytical results. For comparison, the giant component size S is shown (black dashed). S_{color} is reduced due to two mechanisms: First, every node has to be connected to the giant component via two links. The according fraction of nodes $S_{\mathrm{color},\infty}$ is shown with a red dotted line. Second, increasing color frequencies further decrease S_{color}.

below one at $\bar{k} = 1/(1 - r_c)$, and as long as one $U_{\bar{c}}$ is one, equation 11 gives always zero. Therefor the critical value is

$$\bar{k}_{\mathrm{crit}} = 1/(1 - \max_c r_c). \tag{25}$$

In figure 3 upper left, analytical results for a highest color frequency $r_1 = 1/3$ are shown. The corresponding critical value is $\bar{k}_{\mathrm{crit}} = 1/(1 - \max_c r_c) = 3/2$ as expected. Different color distributions were used with different degeneration n_{deg} of the highest color frequency and $C = 10$ colors. We see the same critical value, but different critical exponents $\beta = n_{\mathrm{deg}}$ for $S_{\mathrm{color}} \propto (\bar{k} - \bar{k}_{\mathrm{crit}})^\beta$. This analytical result can be confirmed with numerical results, as shown on the bottom of the figure.

We can understand the critical exponent with expanding equation 11 using $U_{\bar{c}} = 1 - \varepsilon$ for the highest color frequency. We will show below using results from standard percolation that $\varepsilon \propto (\bar{k} - \bar{k}_{\mathrm{crit}})$. 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 $P_{\vec{\kappa}} \propto \varepsilon^{n_{\mathrm{deg}}}$ (unless $P_{\vec{\kappa}} = 0$ in the case $\sum_{c' \neq c} \kappa_{c'} = 0$ for any c). As equation 9 is therefore a superposition of either vanishing terms or terms with leading order $\varepsilon^{n_{\mathrm{deg}}}$, we have

$$\beta = n_{\mathrm{deg}}. \tag{26}$$

To complete the discussion of the critical behavior, we have to show that $1 - U_{\bar{c}} \propto (\bar{k} - \bar{k}_{\mathrm{crit}})$ 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 4 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 10 ($\phi_{\bar{c}} = 1 - r_c$), we find

$$U_{\bar{c}} = 1 - \frac{1 - u(\bar{k}\phi_{\bar{c}})}{1 - u(\bar{k})} \tag{27}$$

$$\approx 1 - \phi_{\bar{c}} \frac{\bar{k} - \bar{k}_{\mathrm{crit}}}{\bar{k} - 1} \tag{28}$$

which drops linearly from one in the critical region

$$0 < \bar{k} - \bar{k}_{\mathrm{crit}} \ll \bar{k}_{\mathrm{crit}} - 1. \tag{29}$$

In figure 3 we have $\bar{k}_{\mathrm{crit}} - 1 = 1/2$, and critical behavior up to about $\bar{k} - \bar{k}_{\mathrm{crit}} = 1/10$.

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 11 and the definition

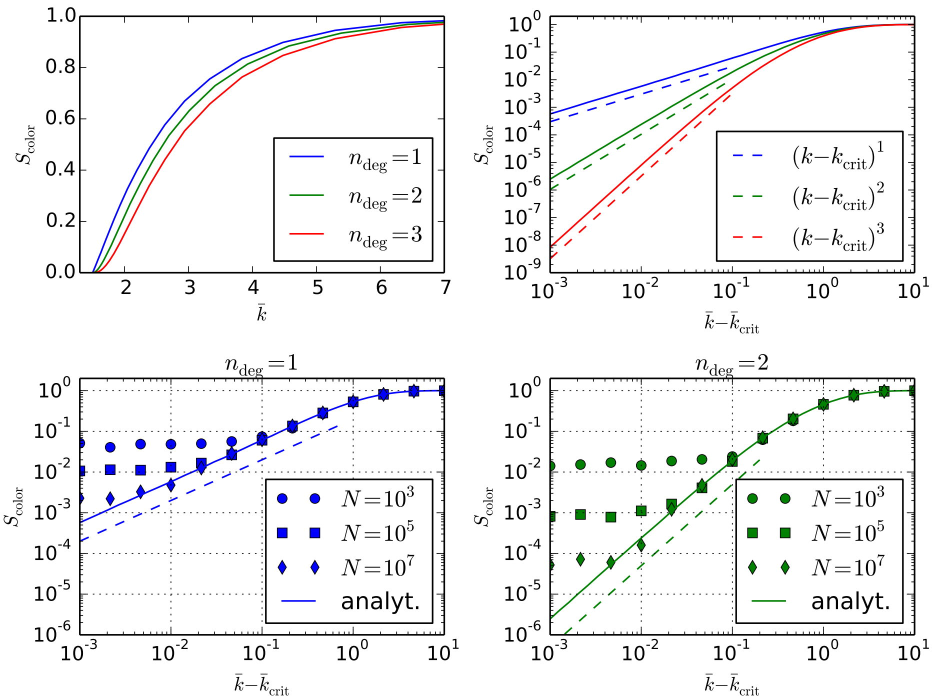

FIG. 3: For heterogeneous color distributions r_c, the highest frequency determines the behavior. On the upper left, analytical results for Poisson graphs are shown for different color distributions with $\max_c r_c = 1/3$ and $C = 10$, all with the same $\bar{k}_{\mathrm{crit}} = 1/(1 - \max_c r_c) = 3/2$. The critical exponent is determined by the degeneration of the highest color frequency. On the upper right, the same results are shown with a log-log plot confirming $\beta = n_{\mathrm{deg}}$. This can be understood with an expansion of equation 9. On the bottom, the analytical results are compared to numerical results which converge to the expected critical behavior with system size.

$\varepsilon = \bar{k} - \bar{k}_2$ 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

$$\bar{k} - \bar{k}_2 = \bar{k}_2 - \bar{k}_1. \tag{30}$$

Exceeding the critical value $\bar{k}_{\mathrm{crit}} = \bar{k}_2$ 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 (compare to the blue squares and line in figure 2 with $r_c = 1/10$ for all colors). There is a deviation from $S_{\mathrm{color},\infty}$ due to $S_{\mathrm{color}} = 0$ below $\bar{k}_{\mathrm{crit}}$, 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 $S_{\mathrm{color},\infty}$ 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 $S_{\mathrm{color},\infty}$.

B. Broad degree distribution

In figure 4, results for graphs with broad degree distributions with $p_k = nk^{-\alpha}$ are shown. n is a normalization constant. We see a strong reduction of $S_{\mathrm{color},\infty}$ compared to S, while the number of colors ($r_c = 1/C$) plays a minor role. Numerical results are averages over 50 networks of size $N = 10\,000$. For evaluating equation 9, 1000 events where sampled for every k.

For broad degree distributions, color distributions can show an additional type of heterogeneity, as a dependence of frequencies on the degree of a node can strongly influence the behavior. We used two colors, where the first color has a frequency of $\tilde{r}_{1,k} = 1$ for all degrees $k \geq k_{\mathrm{step}}$ larger than a certain k_{step}. These nodes have a probability of $\gamma = \sum_{k=k_{\mathrm{step}}}^{\infty} p_k$. Accordingly $\tilde{r}_{2,k} = 0$ for $k \geq k_{\mathrm{step}}$, and probabilities for smaller degrees are chosen as $\tilde{r}_{c,k} = 1/2$.

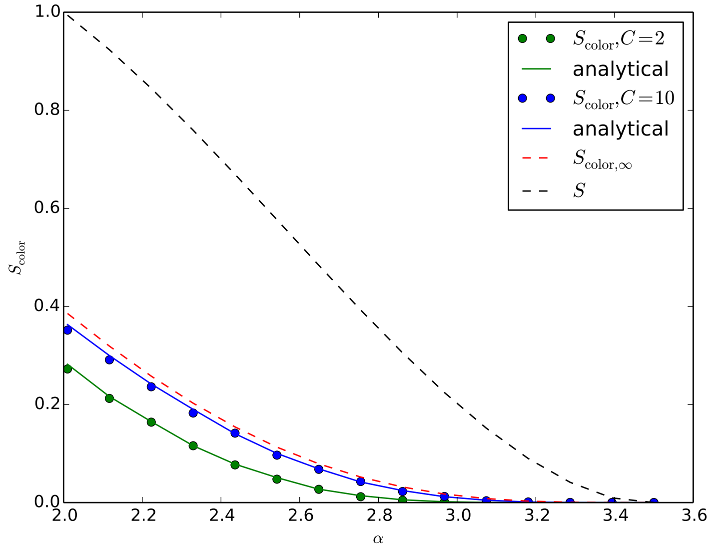

FIG. 4: The same as in figure 2 for scale free degree distributions.

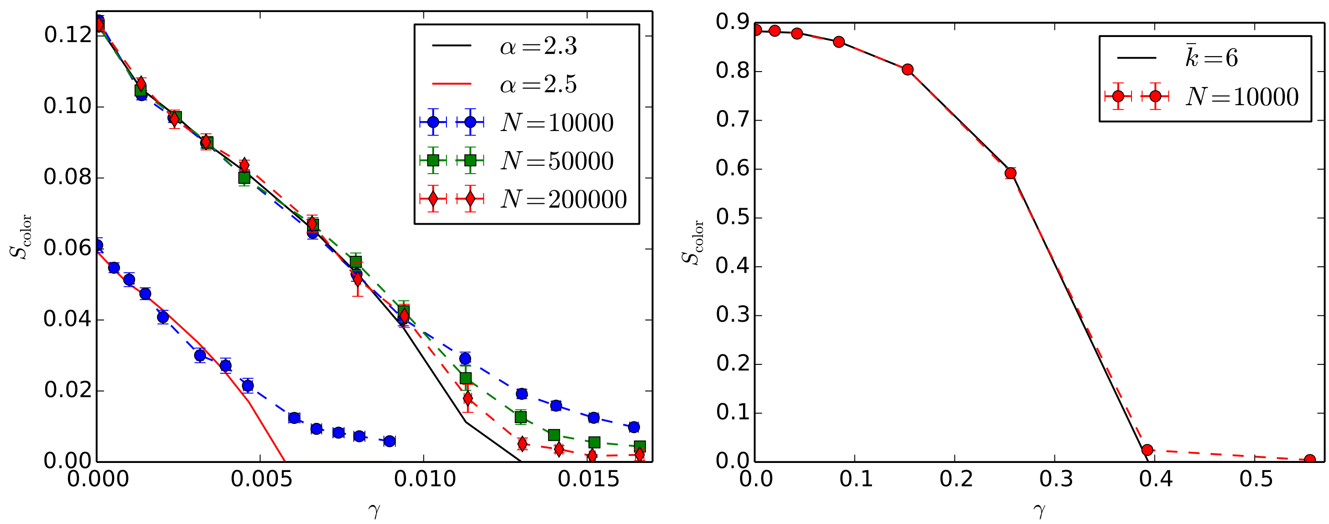

FIG. 5: Left: On graphs with broad degree distribution, S_{color} drops fast, if a fraction γ of nodes with the highest degree is restricted to one color, while the nodes with smaller degree can have one of two colors. This is shown for networks with $\alpha = 2.3$ and $\alpha = 2.5$ having $N = 10000$ with symbols. Analytical results need the modified version of equation 9 as described in section II D. Results are shown with straight lines. Without diversity on the hubs, nodes cannot communicate in the desired way. Right: For comparison results on a Poisson graph are shown, where a large fraction of nodes has to be of one color to hinder communication.

Figure 5 on the left shows results for an ensemble with $\alpha = 2.3$ and $\alpha = 2.5$. The analytical results for $\alpha = 2.3$ show that already for a portion of $\gamma = 1.4\%$ of the largest nodes occupied by the first color exclusively, S_{color} vanishes. These results were calculated as described in section II D. Numerical results confirm this behavior, but show finite size effects. On the right of the figure, results for Poisson graphs are shown. We see that for narrow degree distributions the effect of dominance of one color on the nodes of highest degrees is much smaller. Therefore, graphs with narrow degree distributions are much more robust for providing connection over color avoiding paths.

C. Application: Network of autonomous systems

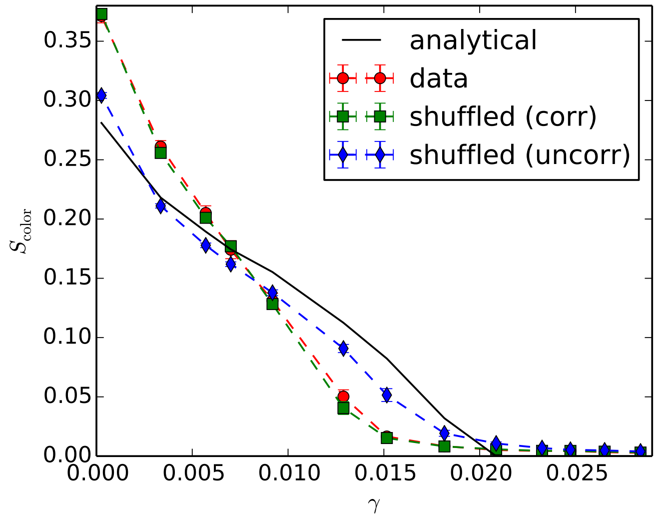

FIG. 6: Red circles show results for the network of autonomous systems, where colors are distributed in the same way as in figure 5. Averages were taken over 10 color distributions. Analytical results are shown with the black line and reproduce the results qualitatively. Deviations are due to degree-degree-correlations which are reserved in shuffled networks shown with green squares, while results with ignoring correlations are shown with blue diamonds.

The red circles in figure 6 show results for the autonomous systems network, where colors where distributed with degree-dependence over the nodes as described at the end of the last section. Averages where taken over 10 realizations of the color distributions. As expected from our results for scale free degree distributions, S_{color} drops to 0 even for small fraction γ which is exclusively of one color. That means that if there is no heterogeneity in the highly connected servers, it is not possible to avoid e.g. software versions. This is also interesting in the following sense: It is known that secret services try to store all decrypted data running through servers to decrypt it later. As this is connected to technical afford, the services will more likely monitor the large servers. Therefore it would be beneficial, if once using encryption, to sent parts of the message using small servers. Unfortunately, this seems to be impossible, the services only have to monitor a low percentage of servers to hinder alternative paths.

In order to assess the predictive power of our analytical method, we used a model ensemble with using the degree frequencies of the autonomous systems network as degree distribution p_k. Results for the according ensemble are shown with the black line. The qualitative behavior is represented well. To understand deviations, we compared to data from shuffled networks starting with the original data. Shuffling with ignoring degree-degree-correlations while only keeping the degree sequence gives results close to the analytical results (blue diamonds). Shuffling with also keeping degree-degree-correlations gives results close to the original network (green squares). Therefore, deviations between our theory and the data arise mainly due to degree-degree-correlations.

Appendix: Independence of components for Poisson graphs

In equation 9, the probabilities $1 - U_{\bar{c}}^{k'-k_c}$ for different colors may show dependencies among each other limiting the usability of the equation. In the following we discuss one particular case for Poisson graphs, in order to illustrate a sort of independence. Lets assume we only have one link connecting to a node in the giant component, and the node has color 3. Can the probability, that this link connects to $\mathcal{G}_{\bar{1}}$ and $\mathcal{G}_{\bar{2}}$ at the same time really be written as the product $(1 - U_{\bar{1}})(1 - U_{\bar{2}})$? As for Poisson graphs we have $S = 1 - u$, instead of discussing link probabilities u we can concentrate on node probabilities S.

We use the notation $S\left(\mathcal{Y}|\mathcal{X}\right)$ to denote the fraction of nodes in a set $\mathcal{Y}$ which is a subset of $\mathcal{X}$. With the set $\mathcal{N}$ of all nodes in the network we have $S_{\bar{c}} = S\left(\mathcal{G}_{\bar{c}}|\mathcal{N}\right)$. Clearly they must be dependent, as $S\left(\mathcal{G}_{\bar{1}} \cap \mathcal{G}_{\bar{2}} \cap \cdots \cap \mathcal{G}_{\bar{C}}|\mathcal{N}\right) = 0 \neq S\left(\mathcal{G}_{\bar{1}}|\mathcal{N}\right) \times \cdots \times S\left(\mathcal{G}_{\bar{C}}|\mathcal{N}\right)$. Lets call the set of all nodes with color c as $\mathcal{A}_c$, and the set of all nodes without this color as $\mathcal{A}_{\bar{c}}$.

Back to our case with the Poisson graph, we have for $c = 1, 2$

$$1 - U_{\bar{c}} = \frac{S_{\bar{c}}}{S(1 - r_c)} \tag{31}$$
$$= S\left(\mathcal{G}_{\bar{c}} \cap \mathcal{A}_{\bar{c}} | \mathcal{G} \cap \mathcal{A}_{\bar{c}}\right) \tag{32}$$
$$= S\left(\mathcal{G}_{\bar{c}} \cap \mathcal{A}_3 | \mathcal{G} \cap \mathcal{A}_3\right). \tag{33}$$

The last equation is intuitive, as the coloring of nodes is random, and it was tested numerically (results not shown). In order to test if the product $(1 - U_{\bar{1}})(1 - U_{\bar{2}})$ reflects the probability of connecting to $\mathcal{G}_{\bar{1}}$ and $\mathcal{G}_{\bar{2}}$ at the same time, we have to check if

$$S\left(\mathcal{G}_{\bar{1}} \cap \mathcal{G}_{\bar{2}} \cap \mathcal{A}_3 | \mathcal{G} \cap \mathcal{A}_3\right) = S\left(\mathcal{G}_{\bar{1}} \cap \mathcal{A}_3 | \mathcal{G} \cap \mathcal{A}_3\right) \times S\left(\mathcal{G}_{\bar{2}} \cap \mathcal{A}_3 | \mathcal{G} \cap \mathcal{A}_3\right) \tag{34}$$

holds. The comparison of the left hand side and the right hand side is shown in figure 7 for networks with $N = 1000$. For every number of colors $C = 3, 4, 5, 10$, ten network realizations were used. So we have tested numerically, that equation 9 is useful at least in this very simple case. This was to illustrate a sort of independence of the conditional probabilities for connecting to color-avoiding giant components. The generalization to many links and general degree distributions is missing so far.

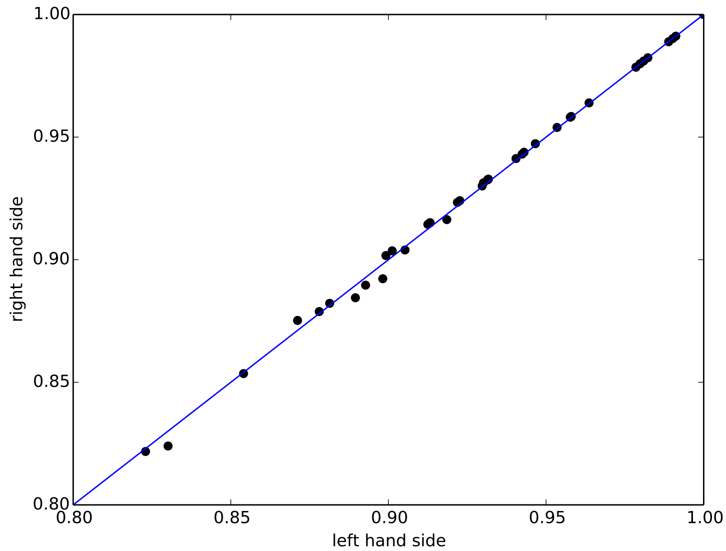

FIG. 7: Calculating both sides of equation 34 for Poisson networks with $N = 1000$ motivates the usage of the conditional probabilities $U_{\bar{c}}$ as independent quantities.